Moving Target Analysis in ISAR Image Sequences with a Multiframe Marked Point Process Model

Csaba Benedek *Member, IEEE* and Marco Martorella *Senior Member, IEEE*

Abstract—In this paper we propose a Multiframe Marked Point Process model of line segments and point groups for automatic target structure extraction and tracking in Inverse Synthetic Aperture Radar (ISAR) image sequences. For the purpose of dealing with scatterer scintillations and high speckle noise in the ISAR frames, we obtain the resulting target sequence by an iterative optimization process, which simultaneously considers the observed image data and various prior geometric interaction constraints between the target appearances in the consecutive frames. A detailed quantitative evaluation is performed on 8 real ISAR image sequences of different carrier ship and airplane targets, using a test database containing 545 manually annotated frames.

Index Terms—Marked Point Process, ISAR, target detection

I. INTRODUCTION

DETECTION and analysis of moving ship or airplane targets in airborne Inverse Synthetic Aperture Radar (ISAR) image sequences are key problems of Automatic Target Recognition (ATR) systems which utilize ISAR data. Remotely sensed ISAR images can provide valuable information for target classification and recognition in several difficult situations, where optical [1] or SAR imaging techniques fail [2], [3]. A number of ATR techniques based on sequences of ISAR images have been proposed in the literature. Some of them directly utilize the 2D ISAR frames [4], whereas others attempt a 3D image reconstruction before dealing with the classification problem [5], [6]. However, robust feature extraction and feature tracking in the ISAR images are usually difficult tasks due to the high noise factors and low level of available details about the structure of the imaged targets. In addition, due to the physical properties of the ISAR image formation process, even the neighboring frames of an ISAR sequence may have significantly different quality parameters in terms of noise or image focus. These artifacts can lead to significant detection errors in some low quality frames, which may mislead the classification and activity recognition modules of the ATR systems [4]. Some previous studies have proposed frame selection strategies to exclude low quality frames from the analysis. However, as pointed out in [4] extracting reliable features for frame selection may often fail. On the other hand, assuming that the target has a fixed size

and structure; and small displacement is expected between consecutive time appearances, inter-frame information can be exploited to refine the detection procedure. For this reason, our proposed system does not drop any frames of the input sequence, but it implements an approach where the detection result on the actual frame jointly depends on the current image data and the neighboring frame's target parameters.

Besides the length and axis line extraction of the target scatterer, another issue is to detect characteristic features of the objects which provide relevant information for the identification process. For this purpose, we identify permanent bright points in the imaged targets, which are produced by stronger scatterer responses from the illuminated objects. However, due to the presence of speckle, image defocus and scatterer scintillation, a significant number of missing and false scatterer-like artifacts appear in the individual frames, thus we focus on their elimination with spatio-temporal filtering constraints.

Target detection techniques in the literature may follow two different mainstreams. The *direct* methods [7] start with the extraction of primitives, such as blobs, edges or corners from the images, then they construct the objects from the primitives in a bottom-up approach. Although these methods can be computationally efficient, they may fail if the primitives cannot be reliably detected. On the other hand, *inverse* methods [8] assign a likelihood value to each possible object configuration and an optimization process attempts to find the configuration with the highest confidence. In this way, flexible object appearance models can be adopted, and it is also straightforward to incorporate prior information about shape and motion. Recently, Marked Point Processes (MPP) [9]–[12] have became widely adopted *inverse* methods in object recognition tasks, since they can efficiently model the noisy image based appearance and the geometry of a target using a joint configuration energy function. However, conventional MPP models deal with the extraction of static objects in single images [9] or in pairs of remotely sensed photos [10], [11]. Conversely, in the addressed scenario, a moving target must be followed across several frames. For this reason, we propose in this work a novel Multiframe MPP (F^mMPP) framework which simultaneously considers the consistency of the observed data and the fitted objects in the individual ISAR images, and also exploits interaction constraints between the object parameters in the consecutive frames of the sequence.

Optimization of MPP models is another critical issue. In the multiframe scenario, the dimension of the target sequence's parameter space may be particularly large, as it is proportional to the number of frames. This fact yields several local maxima of the likelihood function, which can mislead the optimization.

For this reason, in the proposed model we attempt to merge the advantages of both the direct and inverse approaches. First, we perform an initial detection using a *direct* model, which processes the sequence frame-by-frame. This step is quick, however, we must expect that the detector results in low quality frames are notably poor. The output of the direct detector provides the initial state of the F^mMPP optimization process, which yields the final output by iterations which consider inter-frame constraints regarding permanent structure and smooth target motion.

The workflow of the proposed method can be followed in Fig. II-B). In Sec. II we give a short overview on the related works in ISAR image formation and feature extraction. Sec. III presents the formal task definition and the notations, and Sec. IV deals with data preprocessing. We introduce in Sec. V the proposed F^mMPP model, and in Sec. VI the corresponding energy optimization algorithm. Issues of parameter settings are discussed in Sec. VII. In the experimental part (Sec. VIII), qualitative and quantitative results are provided on extraction and analysis of different ISAR sequences. Finally, concluding remarks are given in Sec. IX. Preliminary versions of the proposed model have been presented in [13], [14].

The contributions of the paper are twofold. On one hand, we introduce the general multiframe MPP framework, which provides a novel Bayesian tool for time sequence analysis in remotely sensed scenarios. Although in each application, the usable relevant image features and shape models depend on the imaging circumstances and the considered targets (e.g. large vessels versus small boats), in the proposed F^mMPP framework the data dependent and target specific terms can be modified in a flexible way, meanwhile the other model parts (prior interaction terms, optimization algorithm) can stay unchanged. On the other hand, we propose a concrete implementation of the F^mMPP method on the analysis of large carrier ships and airplanes from ISAR data, and perform a detailed quantitative validation on a real data set, which contains eight ISAR image sequences with 545 manually evaluated frames.

II. RELATED WORK

In this section a brief review of ISAR image formation and target's feature extraction will be presented to introduce the reader to concepts that will be developed in this paper and further motivate this work.

A. ISAR image formation

ISAR image formation algorithm can be interpreted as the solution of an inverse problem, as stated in pioneering work [2], [3]. Nevertheless, the simple ISAR formulation encounter a series of problems when dealing with real data, where assumptions made are not fully satisfied. To address real problems, modern ISAR imaging researchers have introduced a number of robust algorithms for image formation. Specifically, non-stationary ISAR signals are produced when targets undergo oscillating motions, such as in the case of ships, or when they maneuver during the radar dwell time, as it happens in many scenarios including maritime, ground and aerial targets. Time-Frequency Analysis based algorithms have been introduced to solve such problems as they are designed to handle non stationary signals [15]–[17]. Moreover, as the image quality in terms of resolution and focus is strongly related to the target's own motions, a time-window approach has been introduced to optimally and automatically select the data set time-series [18]. ISAR images are typically formed in the Range-Doppler (RD) domain, as the cross-range coordinate cannot be scaled in spatial coordinates if the target's rotation vector is not known or estimated. To solve the cross-range scaling problem, several attempts have been made to estimate the target's rotational component and consequentially rescale the ISAR image in spatial coordinate. Some examples can be found in [19], [20].

B. ISAR image projection plane

The ISAR image projection plane is the plane where the ISAR image is formed. It has been demonstrated that, under given assumptions, the transformation that maps any three-dimensional target onto the image domain is a simple projection. This fact greatly simplifies the interpretation of ISAR images, as projections are easy to understand and produce a result that is visually clear [2], [3]. Nevertheless, the unfortunate part is that the image plane orientation depends on the target motion, which is usually unpredictable. Therefore, the projection seen in the ISAR image may be hard to interpret when the projection axis is not known a priori. This clearly makes the problem of classifying or even recognizing targets from ISAR images a complicated task. Effort has been made to try to limit this problem either by estimating the time-window when a simple projection occurs, such as pure top or side views [19], or by trying to force such projections by suitably positioning the sensor [21]. Nevertheless, the problem of relating projections to 3D targets is still a problem that needs attention as it represents a crucial step in Automatic Target Classification (ATC) and Recognition (ATR).

C. Target's Feature Extraction from ISAR images

Most of the ATC/ATR systems that are based on radar images, make use of a two step approach to solve the problem: firstly features are extracted from the radar image and then they are fed to a classifier that decides based on comparing such features with those that have been previously stored in a database. The type and quantity of features that should be used in an ATC/ATR system is an open problem. Several papers have been written in the literature that show a number of approaches to select and extract features and the way they are used to classify targets [6], [7], [22]–[28]. Among the diverse approaches and set of features used, there are a few common aspects that seem to play an important role in practically all proposed classifiers. One such common aspect relates to obtaining an accurate estimation of the target's size (length, width, height) as it usually leads to an improved target classification. One of the main issues related to estimating the target's size is the visibility of scatterers at the edge of the target. As scatterers may appear and disappear in ISAR images based on shadowing effects or weak scattering mechanisms,
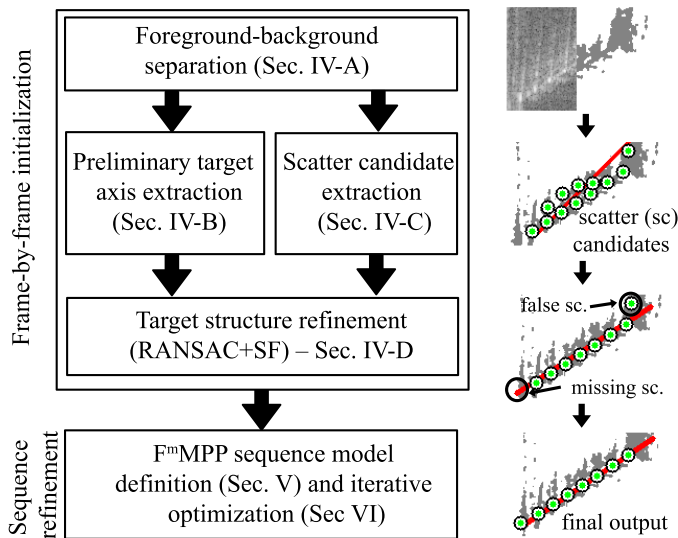
Fig. 1.   Complete workflow of the proposed method, with demonstrating output images for each step.

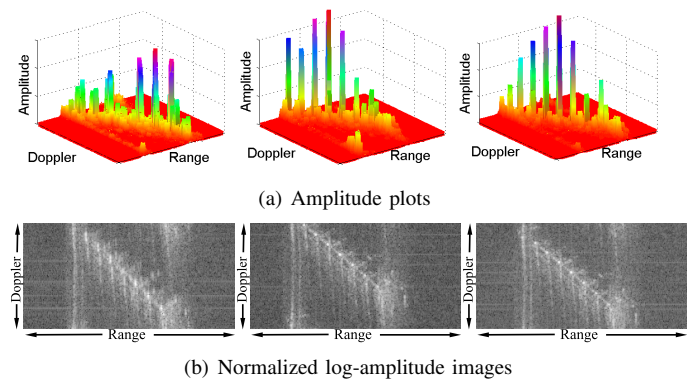(a) Amplitude plots

(b) Normalized log-amplitude images

Fig. 2.   Input demonstration: (a) amplitude plots of three nearby frames of an ISAR sequence in the range-doppler domain (all frames are displayed in the same amplitude scale) (b) normalized log-amplitude images of the same frames

the estimation of the target's size may be incorrectly performed [26]. Nevertheless, it has been shown that by observing a target for a longer period of time, scatterers typically appear and disappear from image frame to frame. Therefore, by using ISAR image sequences, such shortcomings may be overcome. Sequences of ISAR images have been used previously to improve both ISAR image formation (even allow reconstructing 3D ISAR images [5]) and target's classification and recognition [6]. Other important aspects related to classification is the resemblance between features extracted from the ISAR image under test and those present in the database. As scatterer's visibility strongly depends on the target's orientation with respect to the radar, a set of features should be related to such aspect angle to be effective when attempting to recognize the target. The aspect angle dependence of features is a problem that, with some limitation, can be reduced by employing aspect angle independent features [24]. It should be pointed out that such a claim on the aspect angle independence may be in some cases overstated. In a recent work [28], the concept of using *permanent scatterers* was introduced to capitalize the advantage of using scatterers that are visible for wide aspect angles. This allows reducing the data contained in the database as target's features should be available for a reduced number of aspect angles. Both the problem of target's size estimation and permanent scatterer selection have a common ground in the problem of scatterer's response variability in dependence of the target's aspect angle. In this work, a viable solution will be proposed that attempts at improving scatterer's position estimation both in terms of accuracy and robustness.

III. PROBLEM DEFINITION AND NOTATIONS

The input of the proposed algorithm is an n-frame long sequence of 2D ISAR data, imaged in the *Range-Doppler* domain, which contains a single *ship* (or *airplane*) target. Let us denote by S the joint pixel lattice of the images, and by $s \in S$ a single pixel. The amplitude of pixel s in frame

$t \in \{1, 2, \ldots, n\}$ is marked with $\xi_t(s)$. Since the observed $\xi_t(s)$ values may vary in a wide amplitude range (see Fig. 2(a)), for a more compact data representation we derive first images from the input maps by taking the logarithm of the observed amplitudes, thereafter we apply linear scaling for normalization:

$$g_t(s) = \frac{\log \xi_t(s) - \min_{r \in S} \log \xi_t(r)}{\max_{r \in S} \log \xi_t(r) - \min_{r \in S} \log \xi_t(r)}$$

Note that we replace the zero amplitude values with a small positive constant to avoid the calculation of $\log 0$. The images corresponding to the sample frames of Fig. 2(a) in the normalized log-amplitude domain are displayed in grayscale in Fig. 2(b). Apart from visualization, the logarithmic image representation suits well the widely adopted log-normal statistical models of ISAR target segmentation [29].

Our primary aim in this paper is to measure relevant features of the objects, such as length or orientation, which provide us information for target identification and behavior analysis. For this reason, we model the skeletons of the imaged targets by line segments in the proposed approach (Fig. 3(c)). Although as mentioned in Sec. II-C, the investigated ISAR images provide only very limited information about the superstructures of the targets, we can often identify stable bright points in the images, called *permanent scatterers* [28], which can be tracked over the frames of the sequence (see Fig. 4(a)). These characteristic features are produced by stronger scatterer responses (such as containers or cabins) from the illuminated objects, typically as a result of double or triple bounce effects, that are stable over larger aspect angle changes. Therefore permanent scatterers can be used for target identification.

In the following, we denote by u_t a target candidate in frame t. Each target's axis line segment is described by the $c(u) = [x(u), y(u)]$ center pixel $l(u)$ length and $\theta(u)$ orientation parameters (see Fig. 3(c)). In addition, an initially unknown $K(u)(\leq K_{\max})$ number of scatterers can be assigned to the targets, where each scatterer q_i is described in the target line segment's coordinate system by the relative line directional position, $\tau_u(q_i)$, and the signed distance, $d_u(q_i)$ from the center line of the parent object u (see Fig. 4(c)).
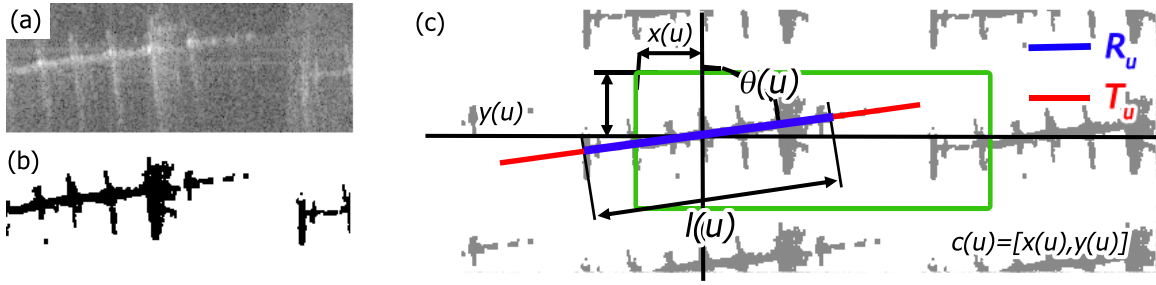
Fig. 3. Target representation in an ISAR image: (a) input image with a single *ship* object (b) binarized image (c) duplicated image and target fitting parameters. Original image border is shown by the green rectangle

(a) Real dominant scatterers (GT)   (b) LocMax filter results

P_1, P_2:	end points of the line segment
q:	scatter candidate
$\tau_u(q)$:	normalized line directional position (fit into the [0,1] interval)
$d_u(q)$:	signed distance from the P_1P_2 line segm. (measured in pixels)

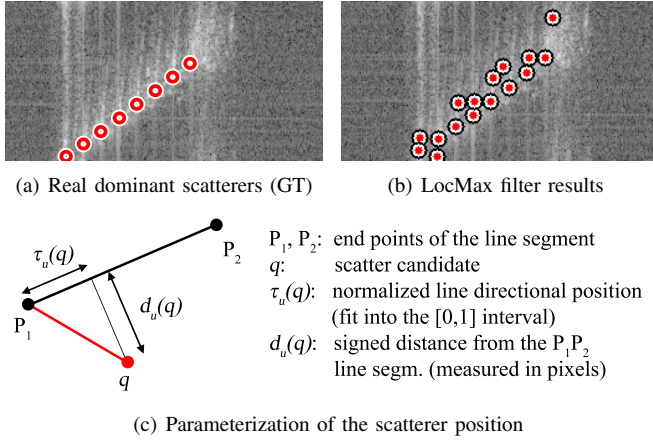(c) Parameterization of the scatterer position

Fig. 4. Dominant scatterer detection problem (a) highlighted true scatterers, i.e. Ground Truth (GT), (b) LocMax filter result, (c) parameterization

The goal is to obtain a $\omega = \{u_1, u_2, \ldots, u_n\}$ target sequence, which we call *configuration* in the following.

IV. DATA PREPROCESSING IN A DIRECT APPROACH

Data preprocessing is needed to prepare the data for subsequent analysis. It should be pointed out that some of the following methods rely on the fact that the image acquisition and therefore the system parameters are suitably tailored to handle certain type of targets. As an example, the range and Doppler extension of the imaged area should be larger than the range and Doppler occupation of the target. As the latter depends on the radar-target dynamics, this condition is not necessarily verified. Nevertheless, certain types of target and radar platforms typically produce predictable or at least bounded dynamics that can give direct information about the radar parameter settings that ensure that such a condition is satisfied. This fact allows us to define foreground-background segmentation in a straightforward way.

A. Foreground-background segmentation

In the first step, we segment the ISAR images into foreground and background classes by a binary Markov Random Field (MRF) model [1], [30] to decrease the spurious effects of speckle noise. The goal is to obtain a binary label map $B = \{b_s | s \in S\}$, where $b_s \in \{\text{fg}, \text{bg}\}$ labels correspond to the foreground and background classes, respectively. Assuming

that the ξ-amplitude values in both classes follow log-normal distributions [29], we model the $p_{\text{bg}}(s) = P(g_t(s)|b_s = \text{bg})$ and $p_{\text{fg}}(s) = P(g_t(s)|b_s = \text{fg})$ log-amplitude posterior probabilities by Gaussian densities. To experimentally validate this model, we have manually drawn foreground masks onto sample images, and investigated the $g_t(s)$ log-amplitude feature statistics in the foreground and background regions, respectively. Fig. 5 shows the $g_t(s)$ histograms of the two classes, as a result of evaluating 18 frames selected from a 245-frame-long ISAR sequence with uniform time intervals, and we can observe that the Gaussian approximation is valid.

To estimate the Gaussian distribution parameters of the foreground and background classes, one can choose either a supervised or an unsupervised approach. Supervised models need manual foreground-background evaluation of a few sample frames, however these key frames have to be carefully chosen, since the range of amplitudes may slightly fluctuate over the sequences. Unsupervised segmentation is a more convenient way from the point of view of the system operator, however, as shown in Fig. 5 the log-intensity domains of the two classes are usually significantly overlapping, thus involving prior knowledge in the process may be necessary. We have assumed having a prior estimation about the ratio of foreground areas compared to the image size, r_{fg}, which was a reasonable assumption regarding large vessels, since our targets have shown line segment structure, and the imaging step has intended to provide us spatially normalized images where the target is centered and the image is cropped so that it estimates a narrow bounding box of the target.

Thereafter, we have derived a preliminary foreground mask by thresholding the input frame followed by a pair of morphological closing and opening iterations, where the threshold corresponds to the $1 - r_{\text{fg}}$ value of the integral histogram. We often found the preliminary mask too coarse for object shape investigations, however, it proved to be appropriate for estimation of the $p_{\text{bg}}(s)$ and $p_{\text{fg}}(s)$ posterior probabilities, as the estimated Gaussian parameters differed only slightly from the supervised estimation results. Let us denote by $\mathbf{1}_s^{\text{fg}} \in \{0, 1\}$ the indicator function of the foreground class in a given segmentation, where $\mathbf{1}_s^{\text{fg}} = 1$ iff $b_s = \text{fg}$. We denote by $s \sim r$, if pixel s is in the 4-neighborhood of pixel r in the S lattice. The optimal foreground mask is derived through
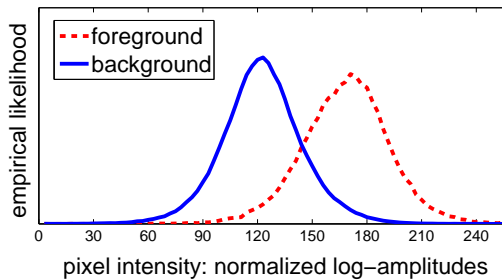
Fig. 5. Histogram of intensity values over a 18-frame long subsequence. Foreground and background regions are manually distinguished
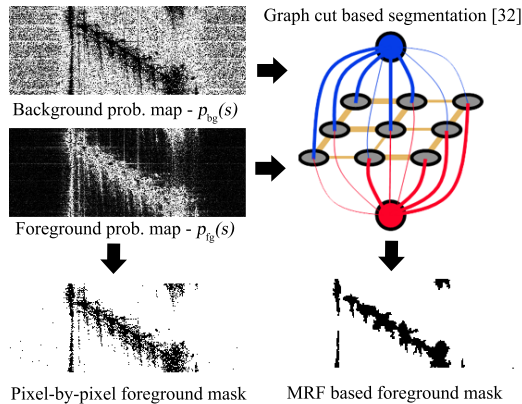

Fig. 6. Demonstration of the foreground-background segmentation. *Top left:* background and foreground probability maps (high probabilities indicated with greater intensities), *bottom left:* foreground mask through pixel-by-pixel maximum likelihood classification (only for reference), *top right:* sketch of graph-cut based MRF optimization [32], *bottom right:* foreground mask (B) by the proposed MRF model

minimizing the the following MRF energy [31] function:

$$B_{\text{opt}} = \underset{B \in 2^S}{\arg\min} \sum_{s \in S} \log \frac{p_{\text{fg}}(s)}{p_{\text{bg}}(s)} \cdot \mathbf{1}_s^{\text{fg}} + \qquad (1)$$

$$+ \sum_{r \sim s} \beta \left(\mathbf{1}_s^{\text{fg}} \cdot \mathbf{1}_r^{\text{fg}} + (1 - \mathbf{1}_s^{\text{fg}}) \cdot (1 - \mathbf{1}_r^{\text{fg}}) \right)$$

Since (1) belongs to the F^2 class of energy functions [31], efficient graph cut based optimization [32] can provide the optimal B mask, as demonstrated in Fig. 6. With also noting the time index, in the following we mark by $B_t(s) \in \{0,1\}$ the foreground mask value of pixel s in frame t.

B. Initial center alignment and line segment estimation

To get an initial estimation of the target axis segment, we detect first the axis line using the Hough transform of the foreground mask. At this point, we also have to deal with a problem which originates from the ISAR image synthesis module. The image formation process considers the images to be spatially periodic both in the horizontal and vertical directions, then, the imaging step estimates the target center, and attempts to crop the appropriate Rectangle of Interest (ROI) from this periodic image (a correctly cropped frame is in Fig. 4(a)). However, if the center of the ROI is erroneously

identified, the target line segment may 'break' into two (or four) pieces, which case appears in Fig. 3(a). Therefore, in the proposed image processing approach, we search for the longest foreground segment of the axis line in a duplicated mosaic image, which step also re-estimates the center of the input frame (see Fig. 3(c)).

C. Scatterer candidate set extraction

Permanent scatterers cause dominantly high amplitudes in the ISAR images; however, due to the presence of multiple scattering mechanisms within the same resolution cell and to defocussing effects, the amplitudes may significantly vary over the consecutive frames, moreover we must expect notable differences between different scatterers of the same frame, which effect is clearly demonstrated in Fig. 2(a). As a consequence, we cannot determine efficient global thresholds to extract all scatterers by simple magnitude comparison. Therefore focusing first on a high recall rate, we extract a large group of scatterer candidates, which may contain several false positives. Thereafter, we propose an iterative solution to discriminate the real scatterers from the false candidates, with utilizing the temporal persistence of the scatterer positions and the line-structure of the imaged targets.

In our implementation, the Local Maxima (LocMax) filter is applied to extract the Preliminary Scatterer Candidates (PSC), which operates on a $R \times R$ rectangular neighborhood around each pixel. We also use a foreground constraint: we only search for scatterers in the ISAR image regions labeled as 'fg' by the initial input binarization step. As the results in Fig. 4(b) show, the real scatterers are efficiently detected in this way, but the false alarm rate is high.

D. Scatterer filtering

The scatterer selection algorithm iterates various local moves, called *kernels*, in the object configuration space. In the following part of this section, we introduce two kernels and demonstrate their effects. Thereafter, the details of the complete spatio-temporal model and the iterative optimization process will be presented in Sec. V and VI.

The input of the *Scatterer Filtering* (SF) kernel is the actual estimation of the axis line segment and the PSC set. The kernel exploits two facts observed in cases of large carrier ships:

- For a given target candidate, we expect that the scatterer candidates are "close" to the axis line.
- The projection of two different scatterers to the axis line should not be "too close" to each other, as the later artifacts are mainly caused by multiple echoes from the same scatterer. The minimal distance of two real scatterer projections in the $\tau_u(q)$ domain is determined by a threshold parameter, T_τ, which varied between 0.05 and 0.07 (see also Fig. 7).

Based on the above assumptions, we select a filtered scatterer set by a sequential algorithm, which is detailed in Fig. 7. However, this filter is by nature very sensitive to the accuracy of the preceding axis estimation step. As shown in Fig. 10(a), applying the SF kernel directly for the output of the Hough-based axis detector results in notably weak classification

---

Algorithm 1: Scatterer Filtering (SF) Kernel

Input

- current estimation of the axis line segment of the target u: P_1P_2.
- set of scatterer candidates extracted by the locMax filter: $\{q_1, \ldots, q_K\}$

Steps of the algorithm

- calculate the $\tau_u(q)$ and $d_u(q)$ parameters of each q scatterer candidate w.r.t. the P_1P_2 segment
- deposit the scatterers in a queue, and sort the scatterers by their $|d_u(q)|$ value (absolute distance) in decreasing order
- For each q scatterer candidate taken in this order in the queue, determine, if there is another q^* in the queue where $|\tau_u(q) - \tau_u(q^*)| < T_\tau$. If we find such a q^*, remove q from the queue.
- Return all scatterers, which have not been removed by the above process

---

Fig. 7.   Pseudo code of the implemented Scatterer Filtering (SF) kernel

(expected scatterer configuration is similar here to the case in Fig 4(a)).

For the above reason, we have proposed a second move which can be applied for scenarios where a strong line-arrangement constraint is valid for the real scatterer configuration (like cases of large ships). We exploit here the fact that if we find a subset of the LocMax-scatterer candidates which fit a given line e_l, we can have a strong evidence the e_l is the axis line of the target. For re-estimating the optimal line to the preliminary scatterer candidates, we have used the RANSAC algorithm (we call this move the *RANSAC kernel* in the following). After obtaining a re-estimated axis, we apply again the Scatterer Filtering kernel: the result in the previous sequence part is demonstrated in Fig. 10(b). We can observe significant improvement in frames #19, #21 and #22; however, we can still find a false (#19) and a missing scatterer (#22), while the result in frame #20 is completely erroneous. It is also important to note, that the RANSAC kernel has a couple of limitations: it cannot be adopted, if there are only a few permanent scatterers on the target's image and the RANSAC-estimation may fail in cases of multiple duplicated scatterers in the PSC set which can form parallel lines. The later artifact can appear as a consequence of echoes in the imaging step. However, assuming that the target structure is fixed, and the position and orientation displacement is small between consecutive frames, temporal constraints can be exploited to refine the detector. Considering the previous observations, in the following sections we embed the previous deterministic kernels to a stochastic iterative framework, which enhances the detection considering the time sequence.

V. MULTIFRAME MARKED POINT PROCESS MODEL

In this section, we introduce the Multiframe Marked Point Process model, which enables to characterize whole target sequences instead of individual objects, through exploiting information from entity interactions. Following the classical Markovian approach, each target sample may only affect objects in its *neighboring frames* directly. This property limits

the number of interactions in the population and results in a compact description of the global sequence, which can be analyzed efficiently. In our model, we use a Z-radius frame neighborhood.

Let us denote by $\mathcal{D}$ the union of all image features derived from the input data. For characterizing a given ω target sequence considering $\mathcal{D}$, we introduce a non-homogenous data-dependent Gibbs distribution on the configuration space: $P_D(\omega) = \frac{1}{\zeta} \exp\left(-\Phi_D(\omega)\right)$ where ζ is a normalizing constant and $\Phi_D(\omega)$ is the configuration energy:

$$\Phi_D(\omega) = \sum_{t=1}^{n} A_D(u_t) + \gamma \cdot \sum_{t=1}^{n} I(u_t, \omega_t)$$

As it appears in the above formula, $\Phi_D(\omega)$ consists of a data dependent term, $A_D(u_t) \in [-1, 1]$ called the unary potential, and a prior term $I(u_t, \omega_t) \in [0, 1]$, called the interaction potential, where $\omega_t = \{u_{t-Z}, \ldots, u_t, \ldots, u_{t+Z}\}$ is a sub-sequence of u_t's $2Z$-nearest neighbors. Parameter γ is a positive weighting factor between the two potential terms. With the following definitions of the energy terms (Sec V-A and V-B) we attempt to ensure that the optimal sequence candidate exhibits the maximal likelihood, thus minimal $\Phi_D(\omega)$ energy. Thereafter, the optimal sequence can be obtained by minimizing $\Phi_D(\omega)$. Let H be the parameter space of an object u_t. We aim to find the Maximum Likelihood (ML) configuration estimate $\widehat{\omega}$, which is obtained as

$$\widehat{\omega} = \underset{\omega \in H^n}{\operatorname{argmax}} P_D(\omega) = \underset{\omega \in H^n}{\operatorname{argmin}} \Phi_D(\omega)$$

A. Definition of the Unary Potentials

The $A_D(u_t)$ unary potential characterizes a proposed object candidate in the tth frame depending on the local ISAR image data, but independently of other frames of the sequence. The unary potential is composed of two parts:

$$A_D(u_t) = \frac{1}{2}\left(A_D^{\mathrm{B}}(u_t) + A_D^{\mathrm{Sc}}(u_t)\right)$$

where $A_D^{\mathrm{B}}(u_t)$ is the *body-term* and is the $A_D^{\mathrm{Sc}}(u_t)$ *scatterer-term*.

For composing the data term, let us first denote by $L_u \subset S$ the set of pixels lying under the dilated line of u in the *duplicated* image. Let us denote by $R_u \subset L_u$ the pixels covered by the line segment u (see Fig. 3(c)): $R_u = \{s \in L_u \mid d(s, [x(u), y(u)]) < l(u)/2\}$ and by $T_u \subset L_u \backslash R_u$ the pixels of the L_u which lie outside the u segment but close enough to its endpoints.

The *body fitting feature*, $f_D(u)$ favors object candidates, where under the line segments (R_u) we find in majority foreground classified pixels in the B-mask of the actual frame, while the outside area T_u covers background regions:

$$f_D(u) = \frac{1}{\mathrm{Ar}\{R_u \cup T_u\}} \cdot \left(\sum_{s \in R_u} B(s) + \sum_{s \in T_u} 1 - B(s)\right),$$

where $\mathrm{Ar}\{.\}$ denotes area in pixels. Thereafter, the *body-term* of the unary potential of u is obtained as:

$$A_D^{B}(u) = Q(f_D(u), d_0),$$

where the following monotonously decreasing $Q(f, d_0)$ function is used:

$$Q(f, d_0) = \begin{cases} \left(1 - \dfrac{f}{d_0}\right) & \text{if } f < d_0 \\ \exp\left(-\dfrac{f - d_0}{10}\right) - 1 & \text{if } f \geq d_0 \end{cases}$$

d_0 is a parameter of the model, used as acceptance threshold for valid objects.

On the other hand, the *scatterer-term* penalizes scatterers that are not located at local Maxima of the ISAR image:

$$A_D^{\text{Sc}}(u) = Q\left(\frac{1}{K(u)} \cdot \sum_{i=1}^{K(u)} \Psi(i, u), \ d_\Psi\right),$$

where

$$\Psi(i, u) = \begin{cases} 0 & \text{if } q_i \text{ is in a local maximum} \\ & \text{in the input ISAR frame} \\ 1 & \text{otherwise} \end{cases}$$

Parameters d_0 and d_Ψ are set by training samples.

B. Definition of the Interaction Potentials

Interaction potentials are responsible for involving temporal information and prior geometric knowledge in the model. Since the observed object's structure can be considered rigid, we usually experience strong correlation between the target parameters in the consecutive frames. Since due to the imaging technique, the $c(u)$ center is not relevant regarding the real target position, we only penalize high differences between the $\theta(u)$ angle and $l(u)$ length parameters, and significant differences in the relative scatterer positions and scatterer numbers between close-in-time images of the sequence.

The prior interaction term is constructed as the weighted sum of four sub-terms: the median length difference $I_l(u_t, \omega_t)$, the median angle difference $I_\theta(u_t, \omega_t)$, the median scatterer number difference $I_{\#s}(u_t, \omega_t)$ and the median scatterer alignment difference $I_{\text{sd}}(u_t, \omega_t)$.

$$I(u_t, \omega_t) = \delta_l \cdot I_l(u_t, \omega_t) + \delta_\theta \cdot I_\theta(u_t, \omega_t) + \\ + \delta_{\#s} \cdot I_{\#s}(u_t, \omega_t) + \delta_{\text{sd}} \cdot I_{\text{sd}}(u_t, \omega_t) \quad (2)$$

where δ_l , δ_θ, $\delta_{\#s}$, δ_{sd} are positive and $\delta_l + \delta_\theta + \delta_{\#s} + \delta_{\text{sd}} = 1$.

The first three sub-terms are calculated as the median values of the parameter differences between the actual and the nearby frames:

$$I_l(u_t, \omega_t) = \min\left(\text{med}_l(t)/d_{\max}^l, 1\right)$$
$$I_\theta(u_t, \omega_t) = \min\left(\text{med}_\theta(t)/d_{\max}^\theta, 1\right)$$
$$I_{\#u}(u_t, \omega_t) = \min\left(\text{med}_K(t)/d_{\max}^K, 1\right)$$

where for target parameters $f \in \{l, \theta, K\}$:

$$\text{med}_f(t) = \underset{t-Z \geq i \geq t+Z}{\text{median}} |f(u_t) - f(u_i)| \quad (3)$$

while $d_{\max}^l$, $d_{\max}^\theta$ and $d_{\max}^K$ are normalizing constants. Note that median filtering proved to be more robust than averaging the difference values due to the presence of outlier frames with erroneously estimated objects.

The scatterer alignment difference feature $I_{\text{sd}}(u_t, \omega_t)$ evaluates the similarity of the relative scatterer positions on the objects of close frames. First we define the target's scatterer alignment vector in the following way:

$$\overline{\tau}(u) = \left(\tau_u(q_1), \tau_u(q_2), \ldots, \tau_u(q_{K(u)})\right)$$

where – as defined in Sec. III – $\tau_u(q)$ is the normalized line directional component of the q scatterer's projection to the axis of u.

Let u and v be objects of two different frames, which may have different numbers of scatterers. The difference between $\overline{\tau}(u)$ and $\overline{\tau}(v)$ is defined as:

$$\Theta\left(\overline{\tau}(u), \overline{\tau}(v)\right) = \frac{1}{2}\left(\frac{1}{K(u)} \sum_{i=1}^{K(u)} \min_{j \leq K(v)} |\tau_u(q_i) - \tau_v(q_j)| + \\ \frac{1}{K(v)} \sum_{j=1}^{K(v)} \min_{i \leq K(u)} |\tau_u(q_i) - \tau_v(q_j)|\right).$$

Then, with using (3), the scatterer alignment difference term is obtained as:

$$I_{\text{sd}}(u_t, \omega_t) = \min\left(\text{med}_{\text{sd}}(t)/d_{\max}^{\text{sd}}, 1\right) \text{ where} \\ \text{med}_{\text{sd}}(t) = \underset{t-Z \geq i \geq t+Z}{\text{median}} \Theta\left(\overline{\tau}(u_t), \overline{\tau}(u_i)\right).$$

For enabling efficient computation, we approximate the $\Theta\left(\overline{\tau}(u_t), \overline{\tau}(u_i)\right)$ feature with the calculation of the 1D distance transform map in a discretized domain of the $[0, 1]$ interval.

VI. OPTIMIZATION

In the literature, optimizing Marked Point Processes is usually performed with iterative stochastic algorithms, such as the Reversible Jump Markov Chain Monte Carlo (RJMCMC) sampler [33] or the Multiple Birth and Death (MBD) dynamics [9], [11], [12]. However, in contrast to the above MPP models, in the proposed system, we track a single object across several frames in the input sequence, where the geometrical constraints are strong between the expected object parameters in near ISAR images. For example, the length and orientation of the imaged target cannot change significantly within a short observation period. As a consequence, the weight of the prior geometric terms increase, and the iterative optimization process becomes sensitive to be stuck in local energy minima, where the configuration excellently fits the prior constraints (permanent structure and smooth motion), but it exhibits a poor match with respect to the data term. To handle the above difficulties, our proposed solution is initialized with the output of the preliminary detector of Sec. IV, which provides an initial configuration which is in most of the frames reasonably consistent with the input data. Thereafter, we proceed to an iterative refinement algorithm, which scans in each step the whole sequence, and attempts to replace the actual objects with more efficient ones considering the data and prior constraints in parallel. The two key points of this procedure are (i) the generation step of new object candidates and (ii) the evaluation step of the proposed objects w.r.t. the current configuration and the input data.

For object generation, we use two types of moves: the Perturbation Kernel and the RANSAC based birth kernel,

which are chosen randomly at each step of each iteration. The pseudo codes of the corresponding functions are shown in Fig. 8. The Perturbation kernel clones the actual object either from the current, or the previous or the next frame; and it adds zero mean Gaussian random values to the center position, length and orientation parameters. Finally, the scatterer positions are cloned from the object of the current frame and optionally additional scatterers are added or some scatterers are removed. The RANSAC based birth kernel has already been introduced in Sec. IV-D.

The whole process of the optimization algorithm is detailed in Fig. 9. We iterate object proposal and evaluation steps, which are followed by the possible replacements of the original objects versus newly generated ones. Let us assume that we are currently in the kth iteration of the process. To decide if we accept or decline the replacement of the object on the tth frame for the newly proposed object, u, we calculate first the energy difference $\Delta\Phi_\omega(u,t)$ between $\omega^{[k]}$, the original configuration before the kth iteration, and the configuration ω^* we would get from $\omega^{[k]}$ by replacing $u_t^{[k]}$ by u. It is important to note that to derive the energy difference we should only examine the objects in the Z-neighborhood of frame t and calculate the concerning unary and interaction potential terms. $\Delta\Phi_\omega(u,t) < 0$ means that the move results in decreasing global energy level. However, to prevent us from finishing the algorithm too early in a low quality local energy minimum, we embed the iterative process into a simulated annealing framework. In this way, as a function of the $\Delta\Phi_\omega(u,t)$ energy difference, we calculate a probability value of accepting the replacement move, and the decision is done by a random choice based on this probability. Regarding the cooling scheme, we have followed the implementation of [9].

VII. PARAMETER SETTINGS

We can divide the parameters of the proposed F^{m}MPP technique into three groups corresponding to the *data*-based target models, the *prior* sequence-level constraints and the *optimization*.

The parameters of the F^{m}MPP *data* (d_0 and d_ψ) and *prior* terms (r_{fg}, T_τ, d_{max}^l, d_{max}^θ d_{max}^K and d_{max}^{sd}) are set based on manually evaluated training data. We follow a supervised approach, since we have observed that using similar ISAR imaging conditions, we do not need to re-calibrate the model parameters for each sequence. For setting all of these co-efficients, one can take a Maximum Likelihood Estimator (MLE), details can be found in [34]. Further relevant prior term parameters are the sub-term weighting factors within the $I(u_t, \omega_t)$ interaction term, we used here uniform weights $\delta_l = \delta_\theta = \delta_{\#s} = \delta_{sd} = 0.25$.

Finally, to set the *optimization* parameters, we followed the guidelines provided in [9] and used $\delta_0 = 10000$, $\beta_0 = 20$ and geometric cooling factors $1/0.96$. Concerning the random object generation, the σ_x, σ_y, σ_θ and σ_l deviation factors depend on noise and the ranges of the corresponding parameter values. Higher σ-s result in more robust detection performance, however, they also decrease the speed of convergence.

Function Pool: Object generator functions

Variables

    n: number of frames of the sequence
    t: frame index

Notation

    $N(\mu, \sigma)$: Gaussian distribution with μ mean value and σ standard deviation.

Functions

Object u = **function** Propose_Obj_by_**PERTURBATION** (t)

- Pick up Δ randomly following the upcoming normal distributions:

    switch(t)
      case 1: $P(\Delta = 0) = 2/3$, $P(\Delta = +1) = 1/3$
      case n: $P(\Delta = 0) = 2/3$, $P(\Delta = -1) = 1/3$
      otherwise: $P(\Delta = -1) = 1/4$, $P(\Delta = 0) = 1/2$,
           $P(\Delta = +1) = 1/4$
    end-switch

- Generate a new object u and set its parameters randomly following the upcoming distributions:

    $x(u) \sim N(x(u_{t+\Delta}), \sigma_x)$, $y(u) \sim N(y(u_{t+i}), \sigma_y)$,
    $\theta(u) \sim N(\theta(u_{t+\Delta}), \sigma_\theta)$, $l(u) \sim N(l(u_{t+\Delta}), \sigma_l)$

    where σ_x, σ_y, σ_θ and σ_l are model parameters.

- Fill the scatterer vector of u by cloning the scatterer vector from u_t, and randomly add new scatterers from the *pre. scatterer candidate set* of frame $t + \Delta$, or delete some of the actual scatterers

return object u

Object u = **function** Propose_Obj_by_**RANSAC** (t)

- Generate a new object u
- Determine the axis line of u by applying RANSAC to the scatterer candidates of the tth frame.
- Estimate the endpoints of the axis line segment by morphology.
- Determine the scatterers for u by **Algorithm 1** of Fig. 7.

return object u

Fig. 8. Pseudo code of the Object Generation Kernels

TABLE I
MAIN PROPERTIES OF THE EIGHT ISAR IMAGE TEST SEQUENCES

Sequence name	Number of frames	Frame size (pix)	Tot. num. of scatterers	Avg axis length	Scat. per fr.
SHIP1	45	256×128	360	153.9	8
SHIP2	90	256×96	720	195.3	8
SHIP3	40	256×96	320	133.9	8
SHIP4	90	256×96	720	179.8	8
SHIP5	90	256×96	720	172.2	8
SHIP6	90	256×96	720	133.7	9
SHIP7	75	256×96	600	169.8	8
AIRPLN	25	128×128	NA	75.2	NA
Overall	520+25	–	4250	151.7	7.79

(a) Initial detection results (Preprocessing, first step)

(b) RANSAC-based refinement (Preproc., second step)

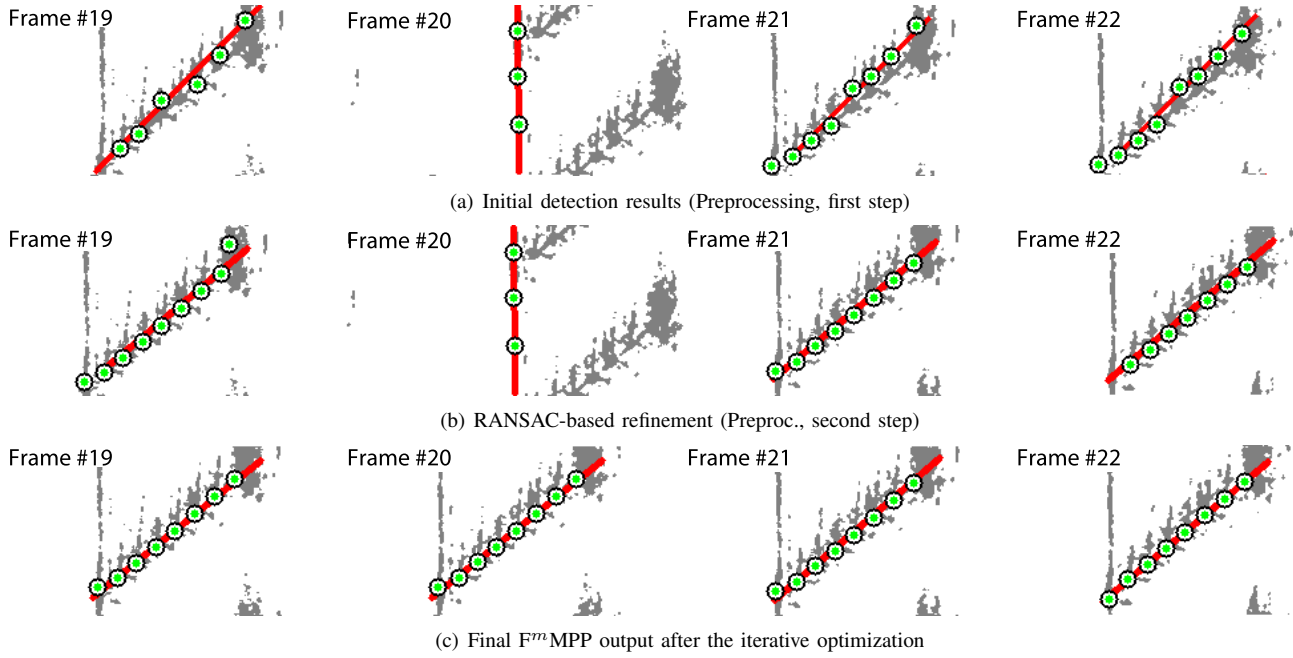(c) Final F^{m}MPP output after the iterative optimization

Fig. 10.   Center alignment and target line extraction results on Frames #19-22 of the `SHIP1` ISAR image sequence. Top: initial detection Middle: RANSAC re-estimation Bottom: detection with the proposed F^{m}MPP model.

Algorithm 2: Optimization of the configuration

Variables

> n: number of frames of the sequence
> k: iteration counter

Steps of the algorithm

1) Initialize the configuration with the output of the deterministic detector: $\omega^{[0]} = \{u_1^{[0]}, u_2^{[0]}, \ldots, u_n^{[0]}\}$, and set iteration counter $k = 0$, inverse temperature $\beta = \beta_0$, refinement parameter $\delta = \delta_0$ and boolean **STOP**:=false.

2) Iterate the following steps while **STOP**=false.

   for each $t = 1, \ldots, n$:
   - Pick up $\Psi_{\text{Birth}} \in \{\text{PERT}, \text{RANSAC}\}$ randomly
   - Generate a new object u so that:

     if $\Psi_{\text{Birth}} = \text{PERT}$:
         $u := \text{Propose\_Obj\_by\_}\textbf{PERTURBATION}(t)$
     if $\Psi_{\text{Birth}} = \text{RANSAC}$:
         $u := \text{Propose\_Obj\_by\_}\textbf{RANSAC}(t)$

   - Consider the ω^* configuration which could be obtained if in $\omega^{[k]}$ we exchanged $u_t^{[k]}$ by u.
   - Calculate the energy difference between and $\omega^{[k]}$ and ω^*:

$$\Delta\Phi_\omega(u,t) = \Phi_D(\omega^*) - \Phi_D(\omega^{[k]})$$

   - Calculate the $d_\omega(u)$ exchange rate as follows:

$$d_\omega(u) = \frac{\delta a_\omega(u)}{1 + \delta a_\omega(u)} \text{ with } a_\omega(u) = e^{-\beta \cdot \Delta\Phi_\omega(u)}$$

     and set

$$u_t^{[k+1]} = \begin{cases} u & \text{with probability } d_\omega(u) \\ u_t^{[k]} & \text{otherwise} \end{cases}$$

3) $k := k+1$, increase β and decrease δ with a geometric scheme.
4) If the process converged: **STOP**:=true.

Fig. 9.   Pseudo code of the preliminary scatterer filtering algorithm

---

VIII. Experimental results

A. Carrier ship sequence analysis

We have tested our method on seven airborne ISAR image sequences about different ship targets. The relevant properties of the test sets are summarized in Table I. In aggregate, the ship data set contains 520 evaluated ISAR frames (40 to 90 frames have been evaluated in each sequence) and 4250 true scatterer appearances (8 or 9 scatterers in each frame). For quantitative validation, we have manually created Ground Truth (GT) data for both the axis segments and the scatterer positions (for longer sequences we have evaluated the first 90 frames). For accurate GT generation, we have developed an accessory program with graphical user interface, which enables us to arbitrarily change the axis parameters, and add, shift or delete the individual scatterers in each frame, meanwhile the result appears immediately over the input image.

To consider different evaluation aspects, we have defined three types of error measures. The *Normalized Axis Parameter Error* (E_{AX}) is calculated as the sum of the x-y center position and axis length errors normalized with the length of the GT target, and the angle error normalized by $90°$:

$$E_{\text{AX}} = \frac{E_{\text{AX}}^x + E_{\text{AX}}^y + E_{\text{AX}}^l}{\frac{1}{n}\sum_{t=1}^n l(u_t^{\text{gt}})} + \frac{E_{\text{AX}}^\theta}{90°}$$

where the following subterms are calculated:

$$E_{\text{AX}}^x = \frac{1}{n}\sum_{t=1}^n ||x(u_t) - x(u_t^{\text{gt}})||, \ \ E_{\text{AX}}^y = \frac{1}{n}\sum_{t=1}^n ||y(u_t) - y(u_t^{\text{gt}})||$$

$$E_{\text{AX}}^l = \frac{1}{n}\sum_{t=1}^n ||l(u_t) - l(u_t^{\text{gt}})||, \ \ E_{\text{AX}}^\theta = \frac{1}{n}\sum_{t=1}^n \Delta\theta(u_t, u_t^{\text{gt}}))$$
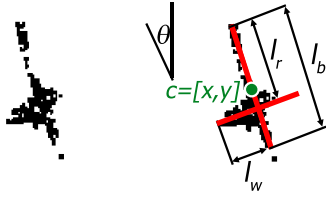
Fig. 11.   Airplane silhouette and the cross shaped fitted model

(a) Input sequence

(b) Initial detection

(c) Optimized detection

Fig. 12.   Airplane extraction: Comparing the results of the *initial* and the optimized F^mMPP detection in four sample frames from the `AIRPLN` sequence

Regarding E^θ_{AX}, we assume that $\theta(u_t), \theta(u_t^{\mathrm{gt}}) \in [0, 180°]$, and we use

$$\Delta\theta(u_t, u_t^{\mathrm{gt}}) = \min\left(|\theta(u_t) - \theta(u_t^{\mathrm{gt}})|, 180° - |\theta(u_t) - \theta(u_t^{\mathrm{gt}})|\right)$$

Table II summarizes the axis level detection rates (smaller error values are favored) for the three steps of the workflow which are displayed in Fig. 10: (a) *Initial* detection, (b) *RANSAC*-based refinement and (c) the final F^mMPP output after iterative optimization. We can observe that the errors decrease over the consecutive steps, and at the end of the process the summarized E_{AX} rate is between 3% and 7% in all sequences.

The *Scatterer Detection Rates* characterize the correctness of permanent scatterer identification. First, the corresponding detected and GT scatterers are automatically matched to each other by the Hungarian algorithm [35], which utilizes the $\tau(q)$ parameter of the scatterers for the assignment. A match is only considered valid if the distance of the assigned feature points is lower than a threshold. Thereafter, we count the number of true positive, false negative and false positive scatterers, which are listed in columns 3-5 of Table III.

The third feature is the *Average Scatterer Position Error* (E_{SP}), which is measured in pixels:

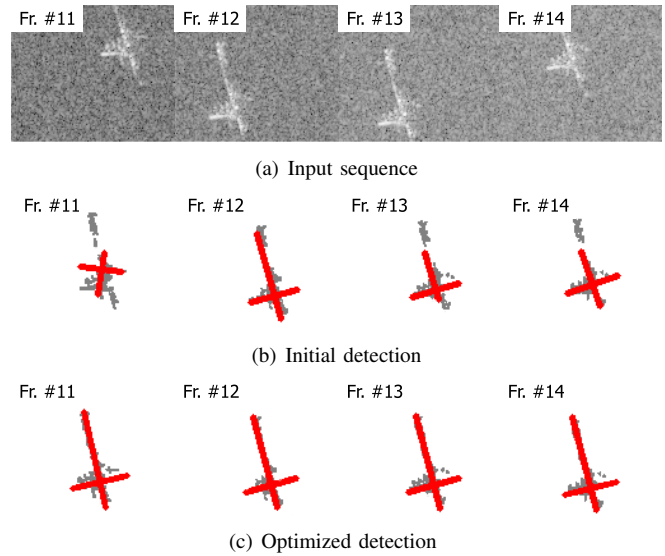$$E_{\mathrm{SP}} = \sum_{t=1}^{n} \frac{1}{\#\{i : m_i^t \neq 0\}} \sum_{i=1}^{K(u_t)} \mathbf{1}_{\{m_i^t \neq 0\}} \cdot \|q_i(u_t) - q_{m_i^t}(u_t^{\mathrm{gt}})\|$$

where m_i^t is the index of the GT scatterer matched to the ith detected scatterer of frame t, marking with $m_i^t = 0$ the unmatched scatterers. $\mathbf{1}_{\{.\}}$ refers to an indicator function and $\#\{.\}$ is the set cardinality. The measured E_{SP} values can be compared in the 6th column of Table III.

By examining the evaluation rates of Tables II and III, we can observe that the proposed method can accurately deal with all the seven test cases (`SHIP1`-`SHIP7`). The improvement between the outputs of the *Initial* and *Optimized F^mMPP* phases of the process is particularly significant in the `SHIP1` (shown in Fig. 10), `SHIP2` and `SHIP5` sequences, which contain difficult test cases. The developments are also remarkable in the `SHIP3`, `SHIP4` and `SHIP6` cases (see sample frames in Fig. 13), while the `SHIP7` sequence contains noisier images with several blurred frames, where the final error rates remain larger (see also the last row of Fig. 13).

B. Application to airplane detection

The proposed model can be generalized to analyze various targets in ISAR image sequences. In this section, we show a case study for airplane skeleton detection with the Multiframe MPP (F^mMPP) model. While some ships, such as carriers, in the ISAR image sequences can be approximated by line segments, airplanes appear as cross-like structures, where at least one of the wings can be clearly observed. Apart from the length and orientation of the body axis segment, the length of the wings and their connecting positions to the airplane body are also relevant shape parameters. For this reason, we use a cross shaped airplane model, as shown in Fig. 11. Parameters are the body center position $c = [x, y]$, body orientation θ, body length l_b, wing 'root' position l_r and wing length l_w.

Similarly to the ship detection procedure, the airplane extraction process consists of a coarse preliminary detection step, and the F^mMPP based iterative refinement step. The preliminary detection starts with the extraction of the body line, using the same Hough transform based technique as introduced for the ship detection process. Secondly, the initial l_r wing root position parameter is obtained with exhaustive search by histograming the silhouette pixels, which can be perpendicularly projected to the same points of the body line. In the F^mMPP based refinement stage the $A_D(u_t)$ data term is calculated in an analogous manner to the ship model, the difference is that the filling factors for the left and right wings are separately calculated, and their minimum (i.e. the better one) counts into the data term of the model. This later feature is necessary, since usually only one of the wings is fully visible in the ISAR data. Results of the airplane detection for 4 sample frames are demonstrated in Fig. 12 showing the output at the preliminary stage and after the F^mMPP optimization. Similar improvement can be observed to the ship detection scenarios. Note that it is also often possible to observe permanent scatterers in images of airplane targets. However, since airplane scatterers can appear both in the wings and in the body, their geometric alignment patterns may be more complex than in cases of the linear vessels.
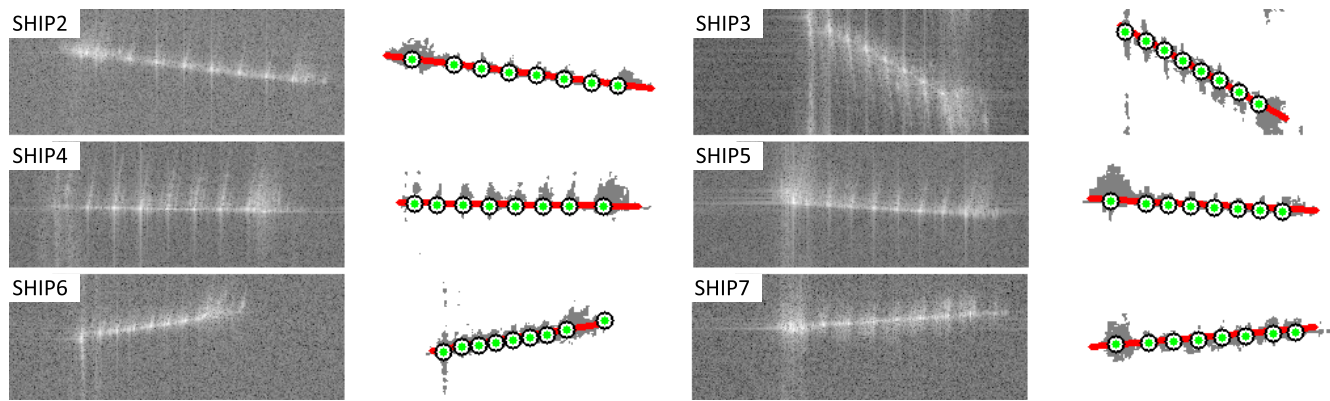
Fig. 13. Sample frames from the `SHIP2-SHIP7` data sets, and the corresponding detection results of the F^mMPP approach obtained by the optimization of the proposed ISAR sequence based model.

TABLE II
AXIS DETECTION - EVALUATION RESULTS FOR THE TEST SEQUENCES. E_{AX}^x, E_{AX}^y, E_{AX}^l AND E_{AX}^θ MEAN ERRORS ARE MEASURED IN PIXELS, THE NORMALIZED E_{AX} ERROR RATE IS EXPRESSED IN PERCENT (%).

Sequence	Step	E_{AX}^x	E_{AX}^y	E_{AX}^l	E_{AX}^θ	$\mathbf{E_{AX}}$ (%)
SHIP1	*Initial*	6.31	9.89	10.6	5.64	*23.6*
	RANSAC	5.11	5.69	9.11	2.18	*15.3*
	F^mMPP	**0.44**	**0.27**	**3.73**	**0.8**	**3.8**
SHIP2	*Initial*	5.85	1.72	13.11	1.51	*12.3*
	RANSAC	2.99	1.02	6.56	0.71	*6.2*
	F^mMPP	**0.47**	**0.17**	**4.29**	**0.58**	**3.2**
SHIP3	*Initial*	2.80	2.15	5.70	2.15	*10.3*
	RANSAC	1.65	1.33	4.92	1.52	*7.5*
	F^mMPP	0.33	0.30	2.65	0.90	**3.4**
SHIP4	*Initial*	2.37	0.83	5.96	0.58	*5.7*
	RANSAC	2.70	0.82	5.69	0.79	*6.0*
	F^mMPP	**0.64**	**0.06**	**4.37**	**0.38**	**3.2**
SHIP5	*Initial*	2.07	0.96	5.86	1.10	*6.4*
	RANSAC	1.43	0.47	3.50	0.86	*4.1*
	F^mMPP	**0.19**	**0.09**	**4.01**	**0.80**	**3.3**
SHIP6	*Initial*	2.33	1.54	3.71	1.96	*7.8*
	RANSAC	1.46	0.70	4.09	1.11	*5.9*
	F^mMPP	**0.01**	**0.07**	**3.20**	**0.50**	**3.0**
SHIP7	*Initial*	4.53	0.87	9.27	1.12	*9.9*
	RANSAC	3.32	0.72	9.21	0.75	*8.6*
	F^mMPP	**2.13**	**0.13**	**8.13**	**0.56**	**6.7**
AIRPL	*Initial*	1.68	6.16	16.32	2.56	*34.9*
	F^mMPP	**0.24**	**0.80**	**3.28**	**0.76**	**6.6**

TABLE III
EVALUATION OF SCATTERER DETECTION. NUMBER OF FALSE POSITIVE AND FALSE NEGATIVE SCATTERERS DETERMINE THE PRECISION AND RECALL FACTORS OF THE PROCESS, WHILE THE E_{SP} RATE SHOWS THE SCATTERER POSITIONING ACCURACY (LOW VALUES ARE PREFERRED)

Sequence	Step	Number of True/False Scatterers			E_{SP}
		True Pos	False Pos	False Neg	in pixel
SHIP1	*Initial*	249	117	111	*7.4*
	RANSAC	339	49	21	*2.2*
	F^mMPP	**349**	**10**	**11**	**0.5**
SHIP2	*Initial*	680	49	40	*4.8*
	RANSAC	703	23	17	*1.6*
	F^mMPP	**718**	**2**	**2**	**0.4**
SHIP3	*Initial*	301	33	19	*1.5*
	RANSAC	306	30	14	*1.6*
	F^mMPP	**311**	**22**	**9**	**1.0**
SHIP4	*Initial*	696	66	24	*1.1*
	RANSAC	699	64	21	*1.0*
	F^mMPP	**705**	**22**	**15**	**0.7**
SHIP5	*Initial*	691	69	29	*0.9*
	RANSAC	695	71	25	*0.6*
	F^mMPP	**707**	**29**	**13**	**0.3**
SHIP6	*Initial*	763	48	47	*0.9*
	RANSAC	763	49	47	*0.8*
	F^mMPP	**764**	**18**	**46**	**0.7**
SHIP7	*Initial*	562	61	38	*3.5*
	RANSAC	567	58	33	*2.9*
	F^mMPP	**559**	**37**	**41**	**2.5**

C. Computational complexity

Table IV lists the computational time requirements of the three consecutive steps of the proposed model (*Initial*, *RANSAC* and the iterative *Optimization*) on the `SHIP1-SHIP7` sequences using a standard desktop computer. The processing speed varies over the different test sets between 2 frames per second (fps) and 5fps, since the computational complexity depends on various factors, such as length of the sequence, image size, target length, quality of the initial detection step and number of scatterers. In most cases, the computational overload of the iterative optimization step is not significantly higher compared to the cost of the initialization and the RANSAC processes. Before running the F^mMPP optimization step, the method needs to collect a sequence of sample frames from the target. If online operation is required

TABLE IV
PROCESSING TIME CONCERNING THE THREE CONSECUTIVE STEPS (COLUMNS 2-4) AND OVERALL TIME REQUIREMENTS (COL. 5-6) REGARDING THE SEVEN SHIP SEQUENCES

Sequence name	Time req. of step (sec)			Overall time	Time req. per frame
	Init	Ransac	Opt		
SHIP1	5.5	3.7	13.6	22.9	0.50
SHIP2	13.4	10.7	14.0	38.1	0.42
SHIP3	4.6	3.3	6.4	14.3	0.36
SHIP4	7.7	4.2	9.4	21.3	0.24
SHIP5	13.4	5.4	5.4	24.1	0.27
SHIP6	8.2	5.4	4.7	18.3	0.20
SHIP7	13.4	8.8	4.7	26.9	0.36

(i.e. the detection cannot be delayed till a longer sequence part arrives), we can use a causal $\omega_t = \{u_{t-Z}, \ldots, u_{t-1}\}$ frame-neighborhood instead of the symmetric one of Sec. V.

IX. CONCLUSION AND FUTURE WORK

This paper has addressed the detection and characterization of large ship and airplane targets in ISAR image sequences using an energy minimization approach. We have proposed a robust joint model for axis extraction, feature point detection and tracking. We have shown that in case of noisy sequences, the introduced Multiframe Marked Point Process schema can significantly improve the results of frame-by-frame detection.

As future work, we aim to extend the proposed technique with different data models and target types, including small boats and various airplanes, considering both aerial and terrestrial radar systems. Another important issue will be to ensure the adaptivity of the algorithms through self-learning parameter estimation strategies, enabling fully automatic analysis of various targets. Finally we aim to test and evaluate the model in various target classification, recognition and behavior analysis tasks.

REFERENCES

[1] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multi-layer conditional mixed markov model," *IEEE Trans. Geosci. Remote Sensing*, vol. 47, no. 10, pp. 3416–3430, 2009.

[2] J. L. Walker, "Range-doppler imaging of rotating objects," *IEEE Trans. Aerospace and Electronic Systems*, vol. 16, pp. 23–52, 1980.

[3] D. A. Ausherman, A. Kozma, J. L. Walker, H. M. Jones, and E. C. Poggio, "Developments in radar imaging," *IEEE Trans. Aerospace and Electronic Systems*, vol. 20, pp. 363–400, 1984.

[4] A. Maki and K. Fukui, "Ship identification in sequential ISAR imagery," *Mach. Vision Appl.*, vol. 15, pp. 149–155, 2004.

[5] T. Cooke, "Ship 3D model estimation from an ISAR image sequence," in *Proc. IEEE Int. Radar Conf.*, 2003, pp. 36–41.

[6] T. Cooke, M. Martorella, B. Haywood, and D. Gibbins, "Use of 3D ship scatterer models from ISAR image sequences for target recognition," *Elsevier DSP*, vol. 16, pp. 523–532, 2006.

[7] D. Pastina and C. Spina, "Multi-feature based automatic recognition of ship targets in ISAR," *IET Radar, Sonar Navigation*, vol. 3, no. 4, pp. 406–423, 2009.

[8] X. Descombes and J. Zerubia, "Marked point processes in image analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 77–84, 2002.

[9] X. Descombes, R. Minlos, and E. Zhizhina, "Object extraction using a stochastic birth-and-death dynamics in continuum," *J. Math. Imaging and Vision*, vol. 33, pp. 347–359, 2009.

[10] C. Benedek, "Efficient building change detection in sparsely populated areas using coupled marked point processes," in *IEEE Geoscience and Remote Sensing Symposium*, Honolulu, Hawai, USA, 2010.

[11] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, 2012.

[12] Á. Utasi and C. Benedek, "A 3-D marked point process model for multi-view people detection," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, USA, 2011, pp. 3385–3392.

[13] C. Benedek and M. Martorella, "ISAR image sequence based automatic target recognition by using a multi-frame marked point process model," in *IEEE Geoscience and Remote Sensing Symposium*, Vancouver, Canada, 2011, pp. 3791–3794.

[14] C. Benedek and M. Martorella, "Ship structure extraction in ISAR image sequences by a Markovian approach," in *IET International Conference on Radar Systems*, Glasgow, UK, 2012.

[15] T. Thayaparan, L. Stankovic, C. Wernik, and M. Dakovic, "Real-time motion compensation, image formation and image enhancement of moving targets in ISAR and SAR using S-method based approach," *IET Signal Processing*, vol. 2, no. 3, pp. 247–264, 2008.

[16] V.C. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis*, Artech House, Norwood, MA, 2002.

[17] F. Berizzi, E.D. Mese, M. Diani, and M. Martorella, "High-resolution ISAR imaging of maneuvering targets by means of the range instantaneous Doppler technique: modeling and performance analysis," *IEEE Trans. Image Processing*, vol. 10, no. 12, pp. 1880–1890, 2001.

[18] M. Martorella and F. Berizzi, "Time windowing for highly focused isar image reconstruction," *IEEE Trans. Aerospace and Electronic Systems*, vol. 41, pp. 992–1007, 2005.

[19] D. Pastina and C. Spina, "Slope-based frame selection and scaling technique for ship ISAR imaging," *IET Signal Processing*, vol. 2, no. 3, pp. 265–276, 2008.

[20] M. Martorella, "A novel approach for ISAR image cross-range scaling," *IEEE Trans. Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 281–294, 2008.

[21] M. Martorella, "Optimal sensor positioning for ISAR imaging," in *IEEE Geoscience and Remote Sensing Symposium*, Honolulu, Hawai, USA, 2010.

[22] S. Musman, D. Kerr, and C. Bachmann, "Automatic recognition of ISAR ship images," *IEEE Trans. Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1392–1404, 1996.

[23] A.O. Knapskog, "Automatic classification of ships in isar images using wire-frame models," in *European Conf. Synthetic Aperture Radar (EUSAR)*, 2004, pp. 953–956.

[24] K-T Kim, D-K Seo, and H-T Kim, "Efficient classification of ISAR images," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 5, pp. 1611–1621, 2005.

[25] F. Rice, T. Cooke, and D. Gibbins, "Model based ISAR ship classification," *Digital Signal Processing*, vol. 16, no. 5, pp. 628–637, 2006.

[26] F.E. McFadden and S.A. Musman, "Optimizing ship length estimates from ISAR images," in *Proc. Int'l Joint Conf. on Neural Network*, 2000, vol. 1, pp. 163–168 vol.1.

[27] V. Zeljkovic, Q. Li, R. Vincelette, C. Tameze, and F. Liu, "Automatic algorithm for inverse synthetic aperture radar images recognition and classification," *IET Radar, Sonar Navigation*, vol. 4, no. 1, pp. 96–109, 2010.

[28] E. Giusti, M. Martorella, and A. Capria, "Polarimetrically-persistent-scatterer-vased automatic target recognition," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 11, Part 2, pp. 4588–4599, 2011.

[29] R.G. White and M.L. Williams, "Processing ISAR and spotlight SAR data to very high resolution," in *IEEE International Geoscience and Remote Sensing Symposium*, Hamburg, Germany, 1999, pp. 32–34.

[30] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[31] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, 2005.

[32] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.

[33] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny, "Structural approach for building reconstruction from a single DSM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 135–147, 2010.

[34] F. Chatelain, X. Descombes, and J. Zerubia, "Parameter estimation for marked point processes. application to object extraction from remote sensing images," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, Bonn, Germany, 2009, pp. 221–234.

[35] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistic Quarterly*, vol. 2, pp. 83–97, 1955.

Csaba Benedek received the M.Sc. degree in computer sciences in 2004 from the Budapest University of Technology and Economics (BME), and the Ph.D. degree in image processing in 2008 from the Pázmány Péter Catholic University, Budapest. Starting from October 2008, he worked for 12 months as a postdoctoral researcher with the Ariana Project Team at INRIA Sophia-Antipolis, France. He is currently a senior research fellow with the Distributed Events Analysis Research Laboratory, at the Computer and Automation Research Institute of the Hungarian Academy of Sciences, and an assistant professor with the Dept. of Electronic Technology at BME. Between 2010 and 2012, he was the Hungarian national project leader of the Array Passive ISAR adaptive processing (APIS) project funded by the European Defense Agency. His research interests include Bayesian image segmentation and object extraction, change detection, scene recognition and reconstruction from Lidar pointclouds and remotely sensed data analysis.

Marco Martorella (SM'09) received his Laurea degree (Bachelor+Masters) in Telecommunication Engineering in 1999 (cum laude) and his PhD in Remote Sensing in 2003, both at the University of Pisa. He is now an Associate Professor at the Department of Information Engineering of the University of Pisa where he lectures "Fundamentals of Radar" and "Digital Communications" and an external Professor at the University of Cape Town where he lectures "High Resolution and Imaging Radar" within the "Masters in Radar and Electronic Defence". He is a regular visiting Professor at the University of Adelaide and at the University of Queensland in Australia. He is author of more than hundred international journal and conference papers. He has presented several tutorials at international radar conferences and organized a special issue on Inverse Synthetic Aperture Radar for the *Journal of Applied Signal Processing*. He is a member of the IET Radar Sonar and Navigation Editorial Board, a senior member of the IEEE and a member of AFCEA. He is also chair of the NATO SET-196 on "Multichannel/Multistatic radar imaging of non-cooperative targets". He has been recipient of the 2008 Italy-Australia Award for young researchers and 2010 Best Reviewer for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. His research interests are mainly in the field of radar imaging, including passive, multichannel, multistatic and polarimetric radar imaging.