

SHIP STRUCTURE EXTRACTION IN ISAR IMAGE SEQUENCES BY A MARKOVIAN APPROACH

C. Benedek*, M. Martorella†

*DEVA Research Laboratory, MTA SZTAKI, Kende utca 13-17, H-1111, Budapest, Hungary, e-mail: bcsaba@sztaki.hu

†RaSS Center, CNIT, University of Pisa, Via Caruso 16, I-56122 Pisa, Italy, m.martorella@iet.unipi.it

Keywords: ISAR, Point Process, target detection.

Abstract

In this paper we propose a Multiframe Marked Point Process model of line segments and point groups for automatic ship structure extraction and target tracking in Inverse Synthetic Aperture Radar (ISAR) image sequences. For purposes 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. Evaluation is performed on real ISAR image sequences of ship targets.

1 Introduction

Identification and motion analysis of ship targets in airborne Inverse Synthetic Aperture Radar (ISAR) image sequences are key problems of Automatic Target Recognition (ATR) systems which utilize ISAR data [3, 5]. Remotely sensed ISAR images are able to provide valuable information for target classification and recognition in several difficult situations, where optical or SAR imaging techniques fail [6]. However, robust feature extraction and feature tracking in the ISAR frames are usually difficult tasks due to the high noise factors and low level of available details about the structure of the imaged targets (see Fig. 1(a)). 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. However, assuming that the targets have fixed size and structure; and small displacement is expected between consecutive time appearances, inter-frame information can be exploited to refine the detection procedure.

In this paper, we propose a robust Multi-Frame Marked Point process model [2, 4] for the target detection problem, which combines various features extracted from the ISAR images with prior geometric constraints about target shape persistency and smooth motion. Similarly to [2], we extract the central axis line segment and a few characteristic feature points from the imaged ship targets. However, new

data and prior features are involved in the process, and as an essential novelty, the target axis fitting and feature point positioning steps are jointly optimized, exploiting a strong mutual relationship between the structure elements.

2 Problem definition and notations

The input of the proposed algorithm is a sequence of 2D ISAR images, which contains a single ship target. A sample frame is visualized as a grayscale image in Fig. 1(a). Our primary aim is to measure relevant features of the objects, such as length or orientation. For this reason, we model the skeletons of the imaged targets by line segments in the proposed approach (Fig. 1(c)). Although the ISAR images provide only very limited information about the superstructures of the ships, we can often identify permanent bright points in the images, which can be tracked over the frames of the sequence (see Fig. 2(a)). These characteristic features are produced by stronger scatterer responses (such as containers or cabins) from the illuminated objects, and can be used for target identification.

Let us denote by S the pixel lattice of the images of the n frame-long ISAR image sequence and by $s \in S$ a single pixel. D refers to the global image data. We denote by u_t a target candidate in frame $t \in \{1, 2, \dots, n\}$. Each target's axis line segment is described by the $x(u)$ and $y(u)$ centre coordinates, $l(u)$ length and $\theta(u)$ orientation (see Fig 1(c)). In addition, $K(u) (\leq K_{\max})$ scatters can be assigned to the targets: $u \rightarrow (q_1, q_2, \dots, q_{K(u)})$, where each scatter 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. 2(c)). Let us denote by H the object space. The goal is to obtain a $\omega = \{u_1, u_2, \dots, u_n\} \in H^n$ target sequence, which we call configuration in the following.

3 Preprocessing

The first step of the proposed approach is background subtraction in the input ISAR images, which yields a binary foreground mask B_t at each frame $t = 1, \dots, n$. To decrease the number of false foreground pixels caused by speckle noise, we have proposed a Gaussian Markov Random Field model for foreground-background segmentation which can be optimized by efficient Graph-cut based algorithms.

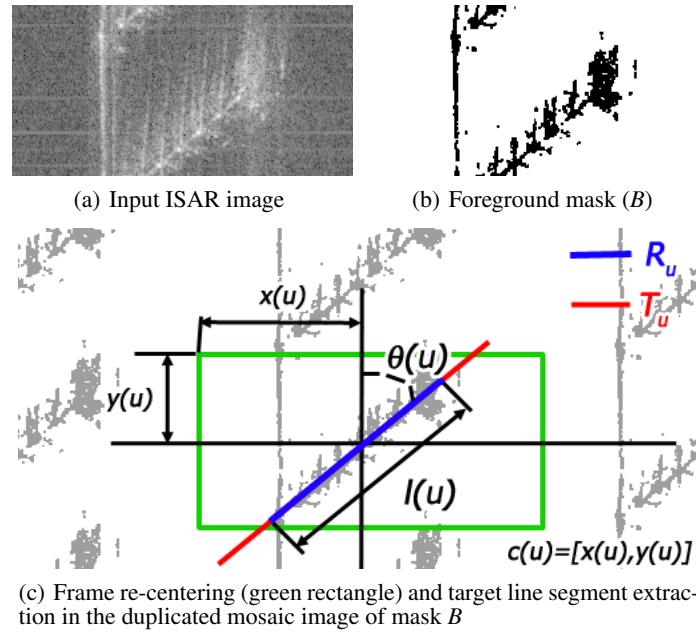


Figure 1: Target axis extraction and parameterization

3.1 Initial 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. 2(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. 1(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. 1(c)).

3.2 Scatter candidate set extraction and filtering

Permanent scatterers cause dominantly high amplitudes in the ISAR images; however, due to the presence of speckle, defocus and scatterer scintillation, the amplitudes may significantly vary over the consecutive frames, moreover we must expect notable differences between different scatterers of the same frame. For this reason, we cannot determine efficient global thresholds to extract all scatterers by simple magnitude comparison. Therefore, we extract first several scatter candidates, and utilize later the temporal persistency and the line-structure of the imaged targets to discriminate the real scatterers from the false candidates. A Local Maxima (LocMax) filter is used to extract the preliminary scatter candidates Fig. 2(b). As the results show, the real scatterers are efficiently detected, but the false alarm rate is high. Our initial scatter filtering step processes exploits two facts.

(i) For a given target candidate, we expect that the scatter candidates are “close” to the axis line (ii) 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. Based on the above assumptions, we can select a preliminary filtered scatter set, which can be observed in Fig 3(a). At this point the quality of the result is still notably poor compared to the expected output, which is similar to Fig 2(a): the filtering step does not select the right subset from the initial candidates of Fig. 2(b). The main reason for the failure is the inaccurate preliminary detection of the center line segment, which yields that false scatter candidates may align to the estimated axis. Therefore, we have proposed a solution, which combines information in parallel from the silhouette mask and the positions of the LocMax-scatter candidates. We exploit the fact that if we find a subset of the scatterers which fit a given line l , we can have a strong evidence the l is the axis line of the target For re-estimating the optimal line to the preliminary scatter candidates, we have used the RANSAC algorithm. After obtaining a re-estimated axis, we apply again the scatter filtering process, whose results are in Fig 3(b). We can observe a significant improvement in the 1st and 3rd frames; however, we can still find a false scatter (1st frame) and an erroneous result (2nd frame) which should be removed by further constraints and by considering temporal information.

4 Multiframe Marked Point Process Model

Following a Bayesian approach, we introduce a data-dependent Gibbs distribution on the configuration space as: $P_D(\omega) = 1/\zeta \cdot \exp(-\Phi_D(\omega))$ where ζ is a normalizing

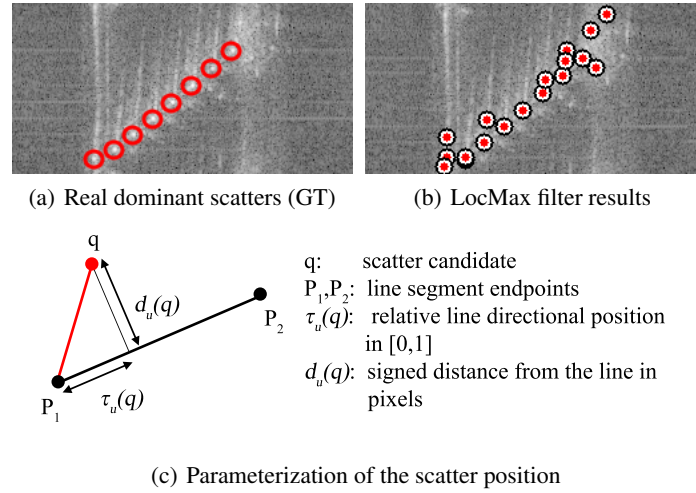


Figure 2: Dominant scatterer detection problem (a) highlighted true scatters, i.e. Ground Truth (GT), (b) LocMax filter result, (c) parameterization

constant and $\Phi_D(\omega)$ is the configuration energy:

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

In the above formula, $A_D(u_t) \in [-1, 1]$ is the data dependent unary potential and $I(u_t, \omega_t) \in [0, 1]$ is called the interaction potential, where $\omega_t = \{u_{t-Z}, \dots, u_t, \dots, u_{t+Z}\}$ is a sub-sequence of u_t 's Z -nearest neighbors. Parameter γ is a positive weighting factor between the two potential terms. We aim to find the Maximum Likelihood (ML) configuration estimate, obtained by minimizing $\Phi_D(\omega)$.

4.1 Definition of the Unary Potentials

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

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

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

For composing the *body-term* term, let us first denote by $L_u \subset S$ the set of pixels lying under the 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. 1(c)): $R_u = \{s \in L_u \mid d(s, [x(u), y(u)]) < l(u)/2\}$ and by $T_u \subset L_u \setminus R_u$ 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 the 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.

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

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

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

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

$$Q(f, d_0) = \begin{cases} \left(1 - \frac{f}{d_0}\right) & \text{if } f < d_0 \\ \exp(-0.1 \cdot f - d_0) - 1 & \text{if } f \geq d_0 \end{cases}$$

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

On the other hand, the *scatter-term* penalizes scatters, which are not located at local Maxima of the ISAR image:

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

$$\Psi(i, u) = \begin{cases} 0 & \text{if } q_i \text{ is a local max of the image} \\ 1 & \text{otherwise} \end{cases}$$

Parameters d_0 and d_Ψ are set by training samples [1].

4.2 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 static, we can usually observe 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 scatter positions and scatter 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 scatter number difference $I_{\#s}(u_t, \omega_t)$ and the median scatter alignment difference $I_{sd}(u_t, \omega_t)$.

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

$\delta_l, \delta_\theta, \delta_{\#s}, \delta_{sd}$ are positive and $\delta_l + \delta_\theta + \delta_{\#s} + \delta_{sd} = 1$. The first three sub-terms are calculated as the median values of the parameter differences between the actual and the nearby frames:

$$\begin{aligned} I_l(u_t, \omega_t) &= \min(\text{med}_l(t)/l_{\max}, 1) \\ I_\theta(u_t, \omega_t) &= \min(\text{med}_\theta(t)/\theta_{\max}, 1) \\ I_{\#s}(u_t, \omega_t) &= \min(\text{med}_K(t)/K_{\max}, 1) \end{aligned}$$

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

$$\text{med}_f(t) = \text{median}_{t-Z \leq i \leq t+Z} |f(u_t) - f(u_i)| \quad (1)$$

while l_{\max}, θ_{\max} and K_{\max} 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 scatter alignment difference feature $I_{sd}(u_t, \omega_t)$ evaluates the similarity of the relative scatter positions on the objects of close frames. First we define the target's scatter alignment vector in the following way:

$$\bar{\tau}(u) = (\tau_u(q_1), \tau_u(q_2), \dots, \tau_u(q_{K(u)}))$$

where - as defined in Sec. 2 - $\tau_u(q)$ is the line directional component of the q scatter's position to the axis of u .

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

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

Then, with using Equation (1), the scatter alignment difference term is obtained as:

$$I_{sd}(u_t, \omega_t) = \min(\text{med}_{sd}(t)/d_{\max}^{sd}, 1) \quad \text{where} \\ \text{med}_{sd}(t) = \text{median}_{t-Z \leq i \leq t+Z} \Theta(\bar{\tau}(u_t), \bar{\tau}(u_i))$$

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

5 Optimization

We have developed an iterative optimization algorithm to obtain an efficient target sequence considering the previously defined configuration energy function $\Phi_D(\omega)$.

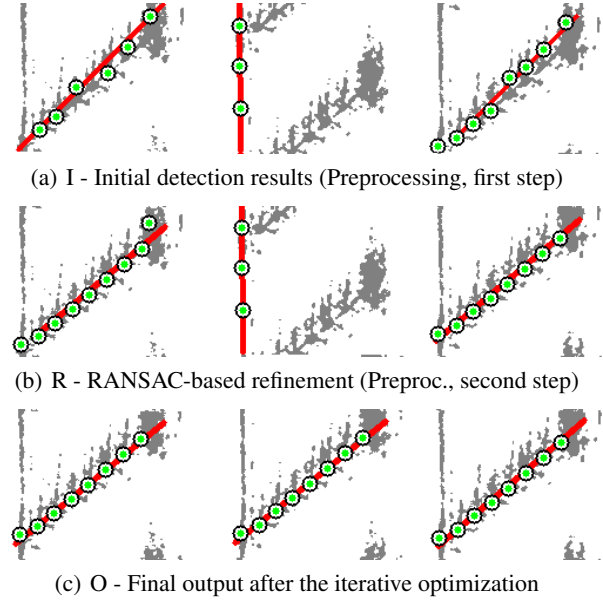


Figure 3: Detection results on three frames of SEQ1 after the steps of the workflow. The 8 scatters are correctly detected.

1. Execute the Initial Detection and the RANSAC-refinement steps (Sec. 3), and initialize the configuration with the obtained sequence (see also Fig. 3(b)): $\omega^{[0]} = \{u_1^{[0]}, u_2^{[0]}, \dots, u_n^{[0]}\}$, and set iteration counter $k = 0$, inverse temperature $\beta = \beta_0$, refinement parameter $\epsilon = \epsilon_0$ and boolean `STOP:=false`
2. Iterate the following steps while `STOP=false`
for each $t = 1, \dots, n$:

- $u := \text{Propose_RANDOM_OBJECT}(t)$
- Consider the ω^* configuration which could be would obtain if in $\omega^{[k]}$ we exchanged $u_t^{[k]}$ by u .
- Calculate the energy difference between $\omega^{[k]}$ and ω^* :

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

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

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

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`. GOTO step 2.

Step →	Normalized Axis Parameter Error			Scatter Det. (F-rate in %)			Average Scatter Pos. Error in pixels		
	I	R	O	I	R	O	I	R	O
SEQ1	0.32	0.28	0.05	44	83	98	13.1	3.4	0.2
SEQ2	0.11	0.08	0.02	88	94	99	1.3	0.9	0.5
SEQ3	0.09	0.11	0.06	88	84	94	3.1	4.2	1.5
SEQ4	0.08	0.06	0.04	93	93	96	2.8	2.5	2.1
SEQ5	0.21	0.16	0.09	93	94	96	0.7	0.6	0.4

Table 1: Quantitative evaluation results for the five test sequences. I/R/O refer to the different steps of the workflow (Initial, RANSAC and Optimized), similarly to Fig. 3

6 Experiments

We have tested our method on five airborne ISAR image sequences about different ship targets. In aggregate, the test data contains 123 evaluated ISAR frames (18 to 30 frames in each sequence) and 1014 true scatter appearances (8 or 9 scatters in each frame). For quantitative validation, we have manually generated Ground Truth (GT) data for both the axis segments and the scatter positions in all frames of all sequences. We have defined three types of error measures. The *Normalized Axis Parameter Error* is calculated as the sum of the center position and axis length errors normalized with the length of the GT target, and the angle error normalized by 90° . The *Scatter Detection Rate* is derived so that we count the number of true positive, false negative and false positive scatters (for a good match, the distance of a detected and a GT scatter should be below a threshold). Thereafter, we give the F-rate (harmonic mean of precision and recall) of the detection in percent. The third feature is the *Average Scatter Position Error*, which is measured in pixels. Table I shows the evaluation rates for the three steps of the workflow I/R/O (see also Fig. 3). We can observe that the proposed method can accurately deal with all the five test cases (SEQ1-SEQ5). The improvement from the Initial to the Optimized phase of the process is particularly significant in SEQ1 (the same sequence is shown in Fig. 3), which contains a difficult test case. The improvements are also remarkable in SEQ2-3, while SEQ4-5 contains easier scenarios where the initial detection is already notably efficient, thus the improvement is smaller.

Acknowledgements

This work was partially funded by the Array Passive ISAR adaptive processing (APIS) project of the European Defence Agency (EDA). The work of the first author was also supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the Hungarian Research Fund (Grant OTKA #101598).

References

- [1] 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.*, 34(1):33–50, 2012.
- [2] 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*, pages 3791–3794, Vancouver, Canada, 2011.
- [3] T. Cooke, M. Martorella, B. Haywood, and D. Gibbins. Use of 3D ship scatterer models from ISAR image sequences for target recognition. *Elsevier DSP*, 16:523–532, 2006.
- [4] X. Descombes, R. Minlos, and E. Zhizhina. Object extraction using a stochastic birth-and-death dynamics in continuum. *J. Math. Imaging and Vision*, 33:347–359, 2009.
- [5] D. Pastina and C. Spina. Multi-feature based automatic recognition of ship targets in ISAR. *IET Radar, Sonar Navigation*, 3(4):406–423, 2009.
- [6] 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*, 4(1):96–109, 2010.