IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323096

VISION-BASED SCATTERING KEY-FRAME EXTRACTION FOR VIDEOSAR SUMMARIZATION

Ying Zhang[1], *Lichao Mou*[2,3], *Daiyin Zhu*[1], *Xiao Xiang Zhu*[2,3]

[1] Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, China
[2] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Germany
[3] Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Germany
E-mail: zhy1994@nuaa.edu.cn

ABSTRACT

Video synthetic aperture radar (VideoSAR) presents significant potential for improving the performance of information interpretation. Key frames represent the aspect-dependent electromagnetic energy, which frequently obscures other scattering physics dominated by specular returns. In this paper, we propose a vision-based background subtraction approach for capturing VideoSAR scattering key-frame information in simultaneously single-channel and single-pass configurations. The spatiotemporal key-frame extractor combines subaperture energy gradient with modified statistical and knowledge-based object tracker. It can robustly discriminate the scattering features of the alternation between transient persistence and disappearance. We evaluate the proposed method using several measured airborne data. Experimental results and performance comparison have demonstrated that the scattering key-frame extractor can achieve a high accuracy for VideoSAR summarization.

Index Terms—Video synthetic aperture radar, scattering key-frame, video summarization, computer vision

1. INTRODUCTION

Video monitoring of regional information is increasingly significant in remote sensing community [1]-[4]. Video synthetic aperture radar (VideoSAR), vividly described as *multiview microwave camera*, is independent of day, night, and weather. It represents a steady and reliable source supporting a wide range in in the aerospace surveillance.

Recently, techniques for information extraction from VideoSAR data have been investigated. A maximum-likelihood coherent change detection (CCD) product was applied to the monitoring of chemical explosion using multi-baseline polarimetric VideoSAR [5]. Owing to the shadow stabilization on the representation of vehicle physical dimensions and locations, Zhang *et al*. [6] proposed a novel VideoSAR low-rank plus sparse decomposition (LRSD) perspective for single-channel single-pass configuration to track the ground defocusing vehicles. Li *et al*. [7] employed the multichannel circular stripmap SAR imagery to track the ground target and estimate its motion parameters after clutter cancellation. The multi-channel technique was extended into VideoSAR mode by the along track interferometry. Nevertheless, multi-baseline and multi-channel configurations represent the limitation for the application of emergency response with stringent timing restriction not permitting the acquisition of repeat-pass and multi-sensor data [13], [15], [16].

Furthermore, massive VideoSAR information can be valuable resources if we can make full use of it to the region monitoring and intelligent analysis. But the large amounts of data have a great impact on hardware processing devices, so the video summarization technology that can extract effective information from quantities of video data has attracted interests and becomes a popular research topic. Aiming at the multitemporal VideoSAR sequences, electromagnetic (EM) scattering characteristic tends to be anisotropic. It is not easy to capture and recognize the scattering characteristic within a narrow subaperture due to the instantaneous "broadside flash" phenomena, especially on man-made objects. These create enormous challenges on vision-based VideoSAR summarization.

In this paper, a vision-based scheme for EM scattering key-frames extraction is proposed for VideoSAR summarization, which combines the subaperture energy gradient (SEG) with modified statistical and knowledge-based object tracker (MSAKBOT). The proposed SEG-MSAKBOT scheme can indicate the scattering states and key-frames of VideoSAR region. This paper is a further extensibility verification of the literature [11]. VideoSAR summaries can greatly help us to quickly understand and master the useful information. Besides, the key-frame extractor is built on the simultaneously single-channel and single-pass configurations, and can be used to break the limitation for the application of emergency response not permitting the acquisition of multi-channel or multi-pass data.

The rest of this paper is organized as follows: Section 2 describes the proposed methodology. Section 3 validates the methodology using several airborne VideoSAR data. Finally, Section 4 draws the conclusion.

2. METHODOLOGY

2.1 Scattering Key-Frames

Scattering key frames represent the VideoSAR content used for indexing, interactive browsing, semantic storage and information retrieval [11]. Since scattering characteristic is limited within those narrow subapertures, scattering key frames are the alternation nodes of anisotropic behavior and scattering energy. In terms of the received scattering signal, key frames are the intense fluctuation of EM energy over the building surface, i.e, varying scattering coefficients. Besides, in terms of visual effect in SAR image domain, they take in the format of varying grey levels.

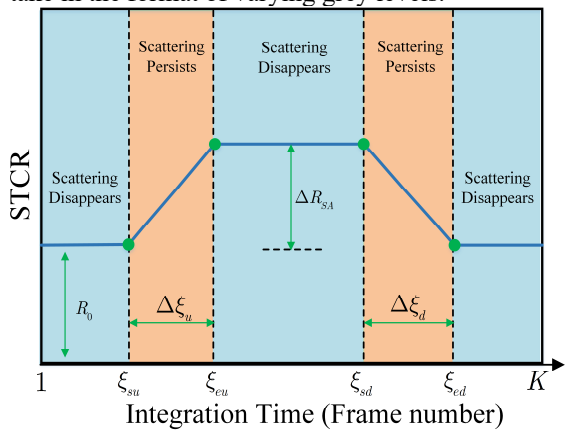


Fig. 1 Scattering key-frame representation.

Let $\mathbb{R}$ be the set of real-valued space. Consecutive VideoSAR frames could be expressed as $\{\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_K\} \subset \mathbb{R}^{m \times n}$, where m and n denote the number of pixels of azimuth and range directions of image $\mathbf{I}_k \in \mathbb{R}^{m \times n}$, for $k = 1, 2, ..., K$. Hence, the set of scattering key-frame of scattering target-to-clutter ratio (STCR) in Fig. 1 is defined as $\Xi = \mathcal{M}(\mathbf{I}) = \{\mathbf{I}_{\xi_{su}}, \mathbf{I}_{\xi_{eu}}, \mathbf{I}_{\xi_{sd}}, \mathbf{I}_{\xi_{ed}}\}$, where $\mathcal{M}(\cdot)$ denotes the scattering key-frames extractor, $\{\xi_{su}, \xi_{eu}, \xi_{sd}, \xi_{ed}\}$ are the different scattering key-frame positions associated with a response amplitude ΔR_{SA}. Scattering persistence time on rising and declining stages is calculated as $\Xi\{\Delta\xi_u, \Delta\xi_d\}$.

2.2 Key-Frame Extractor

Correlation analysis usually presents the direct relationship. Generally speaking, the energy $e(k)$ of the kth subaperture data can be written as:

$$\mathbf{E}_S = \{e(k) \mid e(k) = \mathrm{tr}(\mathbf{I}_k{}^T \mathbf{I}_k)\} \quad (1)$$

where $\mathrm{tr}(\cdot)$ is the trace of a matrix. SEG during the ε frames can be calculated as:

$$\mathbf{G}_{SE} = \left\{ g_{SE}(k) \mid g_{SE}(k) = \frac{d(e(k))}{dk} = \frac{1}{\varepsilon} \cdot [e(k+\varepsilon) - e(k)] \right\} \quad (2)$$

The SEG can determine the scattering trend and amplitude of the anisotropic target and enhance its robustness in VideoSAR sequences.

Next, MSAKBOT is designed. Let p denote a pixel point of frame $\mathbf{I}_k$. It is in the interest of an object-based selective update. When p belongs to the uncovered background, $\mathbf{B}(p)$ is the $\mathbf{I}(p)$ with a value v. Otherwise, the background is obtained with statistical $\mathbf{B}^S_{k+\Delta k}(p)$ on an element set:

$$S = \left\{ \mathbf{I}_k(p), \mathbf{I}_{k-\Delta k}(p), ..., \mathbf{I}_{k-q\Delta k}(p) \right\} \cup w \cdot \left\{ \mathbf{B}_k(p) \right\} \quad (3)$$

where the weight factor w is involved into q frames subsampled at a rate of $(1/\Delta k)$ to generate a stable background. Also, the statistical model $\mathbf{B}^S_{k+\Delta k}(p)$ is generated using the median method [8].

$$\mathbf{B}^S_{k+\Delta k}(p) = \arg\min_{i=1,2,...,t} \sum_{j=1}^{t} Distance(\Gamma_i, \Gamma_j) \quad \Gamma_i, \Gamma_j \in S \quad (4)$$

where the distance is a 2-D metric. The coarse VBS is obtained by taking scatterers' grey level into account.

$$\mathbf{BS}_k(p) = Distance(\mathbf{I}_k(p), \mathbf{B}_k(p)) \quad (5)$$

Scattering foreground $\mathbf{BS}_k$ indicates a varying grey-level information. It should be binarized via dual-threshold operation during the extreme intensity variations [8]-[9].

The undesired false alarms from moving shadows should take in account. Shadow detection results $\mathbf{SD}_k$ can be removed by the shadow-aided constraint OTSU method [10].

$$\mathbf{BS}_k = Distance(\mathbf{BS}_k, \mathbf{SD}_k) \quad (6)$$

As a coarse-to-fine key-frame extractor, further validation is to eliminate the outliers from small motions or 2-D defocus effect. Gradient is derived with respect to both spatial and temporal coordinates:

$$\frac{\partial \mathbf{I}_k(i,j)}{\partial(x,k)} = \mathbf{B}_k(i-1,j) - \mathbf{I}_k(i+1,j) \quad (7)$$

$$\frac{\partial \mathbf{I}_k(i,j)}{\partial(y,k)} = \mathbf{B}_k(i,j-1) - \mathbf{I}_k(i,j+1) \quad (8)$$

Where those past samples $\mathbf{I}_{k-\Delta k}$ have been approximated as $\mathbf{B}_k$. Next, the synthetic gradient can be generated:

$$\mathbf{G}_{BS} = \left\{ g_{BS}(i,j) \mid g_{BS}(i,j) = \sqrt{\left\| \frac{\partial \mathbf{I}_k(i,j)}{\partial(x,k)} \right\|^2 + \left\| \frac{\partial \mathbf{I}_k(i,j)}{\partial(y,k)} \right\|^2} \right\} \quad (9)$$

Temporal partial derivative offers a reliable result on false alarms with a low computational complexity. To estimate its coherence with background gradient $\mathbf{GB}_k$, gradient coherence $\mathbf{GC}_k$ and foreground coherence $\mathbf{FC}_k$ are considered to enhance its reliability.

Accordingly, the overall validation $\mathbf{OV}_k$ has been conducted, which combines the two coherence measurements.

$$\mathbf{OV}_k = \frac{1}{\sum_{(i,j)\in \mathbf{SO}_k} 1} \sum_{(i,j)\in \mathbf{SO}_k} \mathbf{GC}_k(i,j) \times \mathbf{FC}_k(i,j) \qquad (10)$$

Furthermore, the undesired false alarms have been suppressed by a coarse-to-fine threshold [12].

$$\mathbf{SF}_k = \begin{cases} 0 & if\ \mathbf{BS}_k < \mathbf{OV}_k \\ \mathbf{BS}_k & otherwise \end{cases} \qquad (11)$$

Finally, feature gradient $\mathbf{G}_{SF}$ is calculated by $\{\mathbf{SF}_k\}_{k=1}^{K}$ to extract the key-frame. Therefore, the estimated parameters $\Xi\{\xi_{su}, \xi_{eu}, \xi_{sd}, \xi_{ed}, \Delta\xi_u, \Delta\xi_d, \Delta R_{SA}\}$ are recorded as a STCR line by the SEG-MSAKBOT extractor.

3. EXPERIMENTAL RESULTS

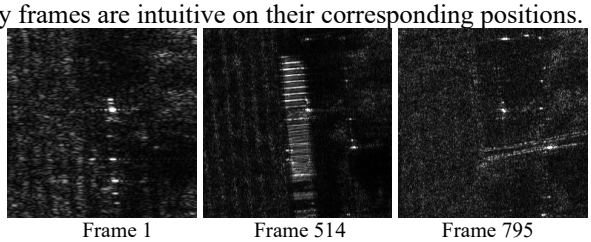
EM scattering key-frame extractor is tested on real X-band airborne data to demonstrate the validity. The relevant radar parameters can be referred to [11].

3.1 Evaluation Criteria

To quantitatively evaluate the SEG-MSAKBOT method, some criteria are selected to assess the performance, i.e., the mean parameter error (MPE), root mean squared error (RMSE), and the mean false alarm rate $MPfa$ [11].

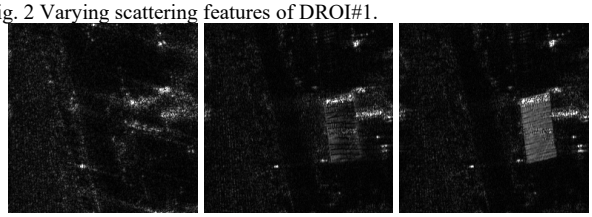
3.2 Key-Frame Extraction Results

Several dynamic region of interest (DROI) imaging results selected from the VideoSAR image sequence as shown in Figs. 2 (DROI#1), 3 (DROI#2), 4 (DROI#3), and 5 (DROI#4). Since the radar line of sight (LOS) exceeds the narrow angle extent ($\Delta\xi_u$ or $\Delta\xi_d$) of strong backscattering energy, scattering feature as a persist-to-disappear-to-persist phenomenon with a varying grey level will fluctuate within some persistence intervals. Fig. 6 presents the estimated STCR curves following by the radar LOS, and the scattering key frames are intuitive on their corresponding positions.



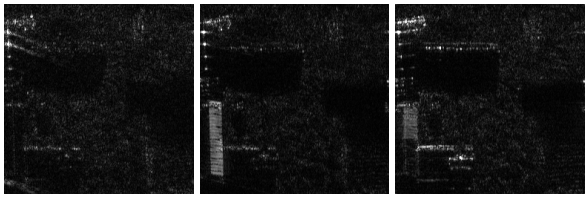
Frame 1 Frame 514 Frame 795

Fig. 2 Varying scattering features of DROI#1.



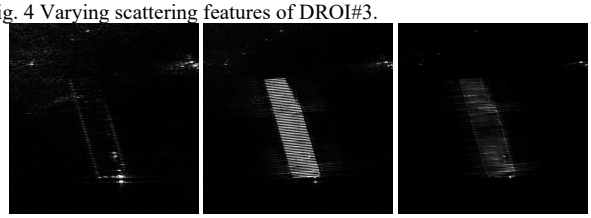
Frame 1 Frame 753 Frame 900

Fig. 3 Varying scattering features of DROI#2.



Frame 1 Frame 285 Frame 476

Fig. 4 Varying scattering features of DROI#3.



Frame 1 Frame 529 Frame 1169

Fig. 5 Varying scattering features of DROI#4.

Furthermore, as shown in Fig. 7, we evaluate the EM scattering feature extractor using the above-mentioned criteria. Both the MPE, and RMSE have been restricted below 2%. Accordingly, the SEG-MSAKBOT method ensures the estimation of scattering key-frame settle in an acceptable accuracy level. The extractor represents an effective clutter- and noise-suppression capability in terms of EM scattering key-frame, which exploits the spatiotemporal VideoSAR context information.

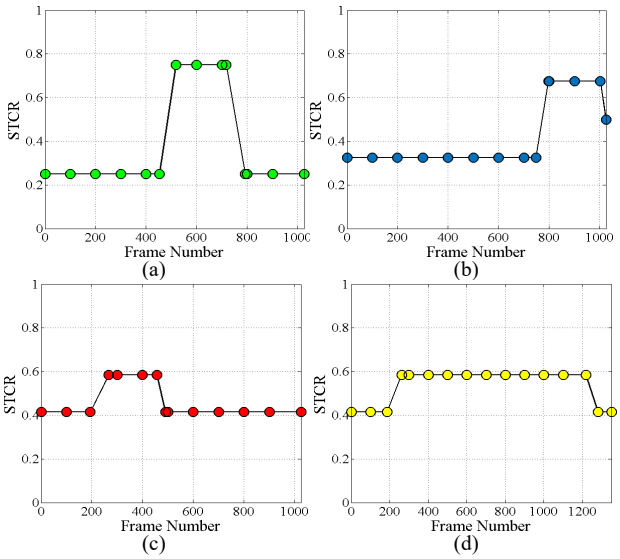
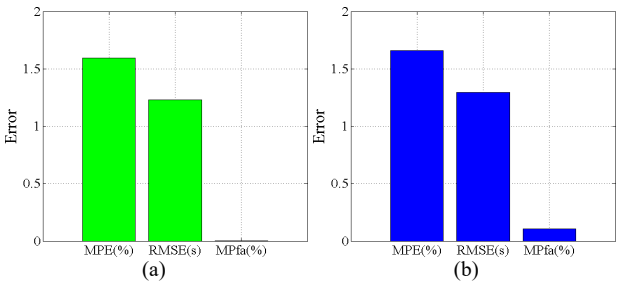


Fig. 6 Estimated STCR key-frames.
(a) DROI#1 (b) DROI#2 (c) DROI#3 (d) DROI#4



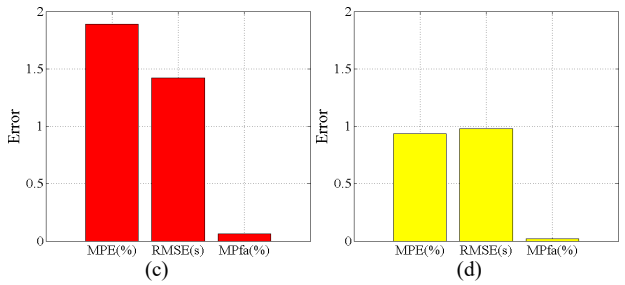


Fig. 7 Error analysis on the STCR key-frames.
(a) DROI#1 (b) DROI#2 (c) DROI#3 (d) DROI#4

3.3 Performance Comparison

Three state-of-the-art algorithms are selected: PBAS, ASOBS, Semi-Soft GoDec. As shown in Fig. 8, we can see that the SEG-MSAKBOT possesses the best MPE and RMSE on the four videos. The ASOBS and Semi-Soft GoDec algorithms are less robust to interference information. SEG-MSAKBOT algorithm accomplishes a high-precision key-frame extraction for video summarization.

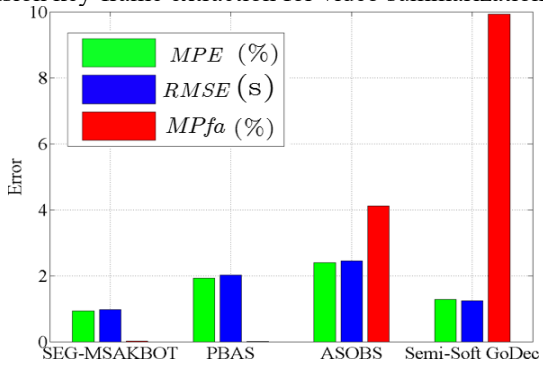


Fig. 8 Quantitative comparison on VideoSAR DROI#4 data.

3.4 SNR Influential Factor

To validate the robustness, we take DROI#4 as noise-free video while the noise is added with different signal-to-noise ratio (SNR) [14]. The results are presented in Fig. 9. With the SNR deterioration and target feature corruption, some parameters are fluctuating at a certain SNR value and the overall tendencies of errors and false alarm experience an increase, but still maintaining at lower level.

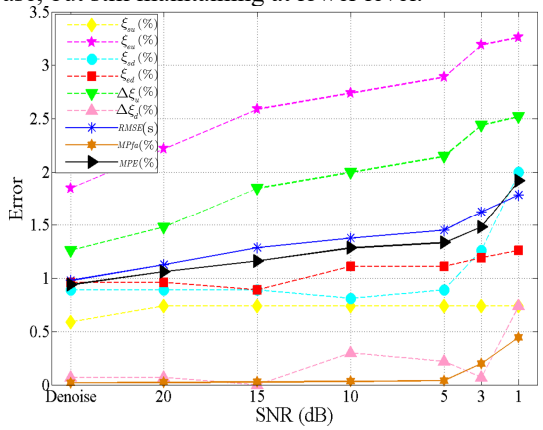


Fig. 9 Impacts of the SNR levels on estimated errors.

4. CONCLUSION

In this paper, a vision-based scattering key-frame extractor for VideoSAR summarization is proposed. It can indicate the scattering states and key-frame in a higher precision, particularly for simultaneously single-channel and single-pass configurations. Measured VideoSAR experiments are shown to confirm the effectiveness of the method.

In the future, with the development of deep learning [2], it is valuable in solving the issue of VideoSAR summarization.

REFERENCES

[1] J. N. Ash, E. Ertin, L. C. Potter, and E. G. Zelnio, "Wide-angle synthetic aperture radar imaging: models and algorithms for anisotropic scattering," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 16-26, Jul. 2014.

[2] X. X. Zhu *et al*., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8-36, Dec. 2017.

[3] B. Wu *et al*., "Resolution-based analysis for optimizing subaperture measurements in circular SAR imaging," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 12, pp. 2804-2811, Dec. 2018.

[4] Y. Zhang, D. Zhu, X. Mao, X. Yu, J., Zhang, and Y. Li. "Multi-rotors video synthetic aperture radar: system development and signal processing," *IEEE Aerosp. Electron. Syst. Mag.*, in press, 2020.

[5] D. A. Yocky, R. D. West, R. M. Riley, and T. M. Calloway. "Monitoring surface phenomena created by an underground chemical explosion using fully polarimetric VideoSAR," *IEEE Trans. Geosci. Remote Sens*., vol. 57, no. 5, pp. 2481-2493, May 2019.

[6] Y. Zhang, D. Zhu, P. Wang, G. Zhang and H. Leung, "Vision-based vehicle detection for VideoSAR surveillance using low-rank plus sparse three-term decomposition," *IEEE Trans. Veh. Technol.,* vol. 69, no. 5, pp. 4711-4726, May 2020.

[7] Y. Li, Y. Wang, B. Liu, S. Zhang, L. Nie and G. Bi, "A new motion parameter estimation and relocation scheme for airborne three-channel CSSAR-GMTI systems," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 4107-4120, Jun. 2019.

[8] S. Calderara, R. Melli, A. Prati, and R. Cucchiara, "Reliable background suppression for complex scenes," *ACM Int. Workshop on Video Surveillance and Sensor Networks*, Oct. 2006, pp. 211-214.

[9] A. Sobral, "BGSLibrary: An OpenCV C++ background subtraction library," in Proc. 9th Workshop Viso Comput. (WVC), Rio de Janeiro, Brazil, Jun. 2013, pp. 38-43.

[10] Y. Zhang and D. Zhu, "Height retrieval in postprocessing-based VideoSAR image sequence using shadow information," *IEEE Sensors J.*, vol. 18, no. 19, pp. 8108-8116, Oct. 2018.

[11] Y. Zhang, D. Zhu, H. Bi, et al, "Scattering key-frames extraction for comprehensive VideoSAR summarization: A spatiotemporal background subtraction perspective," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, July 2020. doi: 10.1109/TIM.2019.2953435.

[12] H. Yu, *et al*., "A coarse-to-fine model for rail surface defect detection," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 3, pp. 656-666, Mar. 2019.

[13] M. Y. Chua et al., "The maiden flight of Hinotori-C: The first C band full polarimetric circularly polarized synthetic aperture radar in the world," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 2, pp. 24-35, Feb. 2019.

[14] B. Taji, A. D. C. Chan and S. Shirmohammadi, "False alarm reduction in atrial fibrillation detection using deep belief networks," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 5, pp. 1124-1131, May 2018.

[15] M. Antoniou et al., "Passive SAR satellite constellation for near-persistent earth observation: Prospects and issues," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 33, no. 12, pp. 4-15, Dec. 2018.

[16] P. Addabbo et al., "Classification of Covariance Matrix Eigenvalues in Polarimetric SAR for Environmental Monitoring Applications," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 6, pp. 28-43, 2019.