Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 205 (2022) 20–27

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Military Communications and Information Systems (ICMCIS 2022)

Onboard Audio and Video Processing for Secure Detection, Localization, and Tracking in Counter-UAV Applications

Andrea Toma*, Niccolò Cecchinato, Carlo Drioli, Giuseppe Oliva, Giovanni Ferrin, Gianluigi Sechi, Gian Luca Foresti

Dept. of Mathematics Computer Science and Physics, University of Udine, Via delle Scienze 206, Udine 33100, Italy

Abstract

Nowadays, UAVs are of fundamental importance in numerous civil applications like search and rescue and military applications like monitoring and patrolling or counter-UAV where the remote UAV nodes collect sensor data. In the last case, flying UAVs collect environmental data to be used to contrast external attacks launched by adversary drones. However, due to the limited computing resources on board of the acquisition UAVs, most of the signal processing is still performed on a ground central unit where the sensor data is sent wirelessly. This poses serious security problems from malicious entities such as cyber attacks that exploit vulnerabilities at the application level. One possibility to reduce the risk is to concentrate part of the computing onboard of the remote nodes. In this context, we propose a framework where detection of nearby drones and their localization and tracking can be performed in real-time on the small computing devices mounted on board of the drones. Background subtraction is applied to the video frames for pre-processing with the objective of an on-board UAV detection using machine-vision algorithms. For the localization and tracking of the detected UAV, multi-channel acoustic signals are instead considered and DOA estimations are obtained through the MUSIC algorithm. In this work, the proposed idea is described in detail along with some experiments and, then, methods of effective implementation are provided.

---

* Corresponding author. Tel.: - ; fax: - .
E-mail address: andrea.toma@uniud.it .

10.1016/j.procs.2022.09.003

Keywords: video background subtraction; audio MUSIC localization; detection localization and tracking; UAV security; onboard multimedia processing

1. Introduction

The last few years have seen a quick increase in the development and utilization of unmanned aerial vehicles (UAVs) in military operations, and it is expected that this trend is going to continue in the future. UAVs introduced many advantages. In fact, they perform various tasks with good efficiency even in dangerous scenarios with reduced risks to humans [1]. For these reasons, this technology has been attracting growing attention in the military applications from defense to armed strikes and employed by a growing number of states worldwide [2]. After being employed as test dummies for aircraft, UAVs contributed to target training and as decoys [3]. Recently, UAVs saw employment in Intelligence, Surveillance and Reconnaissance (ISR) and they have also been armed for striking targets. However, in this work, we focus on unarmed military UAVs for patrolling and defense operational military applications. Specifically, the main application for which we conducted this work is counter-UAV [4] for anti-UAV monitoring strategies against illegal use of UAVs and external UAV attacks [5, 6] even in hostile environments. We aim at designing a proactive counter-UAV system to protect army tanks and patrols from aerial attacks launched by UAVs in non-protected urban areas[†]. In this framework, monitoring and analysis of the acoustic and video scene by using microphone arrays and cameras is undoubtedly of fundamental importance.

On the other hand, the utilization of UAVs still has some limitations that should be addressed. For example, the risks of military UAVs being subjected to cyber and electronic attacks are taken into account in [7]. From a computational point of view, due to the stringent size constraints, UAVs are normally equipped with low power computing devices with limited computational capacity. For this reason, most of the computation must be decentralized on a ground central unit (GCU) with higher computational power. This means that a suitable streaming network should be implemented where the data, acquired by the remote nodes through the sensors mounted on them, is transmitted to the GCU wirelessly. This exposes the system not only to the risk of cyber attacks from external entities but also to problems related to latency and delays due to the propagation of the signals compromising the real-time constraint of the application. This is why we consider moving part of the computation directly on the low power computing device onboard the UAVs.

To show this, an implementation of two methods is proposed in this work. The background subtraction applied to video frames and the MUltiple SIgnal Classification (MUSIC) algorithm applied to multi-channel acoustic signals. This is meant to be incorporated into the detection, Direction of Arrival- (DOA-) based localization, and tracking of flying UAVs. Background modeling or subtraction is an important task in video processing and analysis because it helps the detection phase of moving or incoming objects in a sequence of images. Either fixed or mobile cameras can be used. It then extracts information about the presence of possible intruders in the surroundings. In the literature, background subtraction for video processing is proposed in [8-15]. On the other hand, DOA-based localization and object tracking can be performed by the MUSIC algorithm. The application of this algorithm to audio signal processing in UAVs is investigated in [16-19].

In our research, multimedia data (both audio and video) is captured by onboard microphone arrays and cameras and processed in real-time by the two algorithms. Up to now, the video processing is already implemented on low-cost and small-size computing units onboard our UAVs, while the audio processing is still proposed research and it is on the way to be effectively implemented onboard UAVs following the methodologies discussed in this manuscript. Localization and tracking from acoustic sources like UAVs are important tasks in microphone array processing [20-22] and can be found in a number of different scenarios involving, for example, a team of drones [23]. Acoustic-only localization is especially challenging when audiovisual sensors are installed on small size UAVs [24-28] because of the nature and intensity of disturbances originated by the electrical engines and the propellers

---

[†] Italian MoD Project "A proactive counter-UAV system to protect army tanks and patrols in urban areas (Proactive Counter UAV)" (Grant Number: a2018-045)

[29, 30]. An effective solution is found by adding a video-based analysis to the acoustic data processing [31]. The video processing provides additional information about the surrounding scene to the audio-based algorithm, increasing in this way its accuracy.

To sum up, the motivation and main contribution that this work proposes can be summarized in this way. First of all, we consider processing onboard to increase the level of cyber security, reduce latency, and delays for real-time applications. Furthermore, processing of both audio and video components is proposed that can be fused to increase the level of accuracy and real-time effectiveness. The proposed approach can be implemented in real streaming architectures with low-cost sensors mounted on UAVs.

The remaining sections are organized as follows: the system architecture for the real-time detection, localization and tracking is introduced in Sec. II; the background subtraction method for video frames and experiments are discussed in Sec. III; in Sec. IV the MUSIC algorithm that provides the DOA estimation of multi-channel acoustic signals is discussed along with the experiments; Sec. V concludes the manuscript with directions for future work.

2. The System Architecture for Real-Time Detection, Localization and Tracking

The developed system is composed of a DJI MATRICE 100 drone equipped with audio-visual sensors and a single-board-computer (SBC) able to collect the multimedia data from the sensors and process and send it to the ground station for a deeper processing with high-complexity algorithms. Fig. 1 shows the MATRICE 100 drone to which a circular sensor frame is attached.

The sensor frame has been installed one meter away from the UAV, to reduce the ego noise captured by the microphones. As shown in Fig. 2, it hosts a compact 8-microphone uniform circular array (UCA) composed of four Semitron seMODADMP441 microphone modules, each one hosting a pair of micro electromechanical systems (MEMS) digital microphones in stereo configuration. The MEMS microphone has a flat frequency response from 60 Hz to 15 kHz [30].

At the center of the circular array a camera has been installed. The two peripherals are connected to a Raspberry Pi 4B placed on the other side of the frame with a LiPo battery pack. The multimedia data is acquired and pre-processed onboard. The pre-processed data is then sent to the ground station using a 5 GHz Wi-Fi connection based on IEEE 802.11ac Wave2 protocol. The wireless connection is managed by a TP-Link access-point (AP) EAP-225-outdoor with a MikroTik 15dBi 2x2 multiple-in and multiple-out (MIMO) sectorial antenna MTAS-5G-15D120 [4, 32]. The use of these sensors allows to autonomously locate in advance acoustic and visual sources on board the drone, and to send the pre-processed data to the ground station for a more accurate analysis. This is due to the calculation limitations of the on-board computer that, being installed in a drone, had to be small in size, not demanding of computational and energy resources.

Fig. 1. The flying MATRICE 100 UAV with the sensor frame [30]

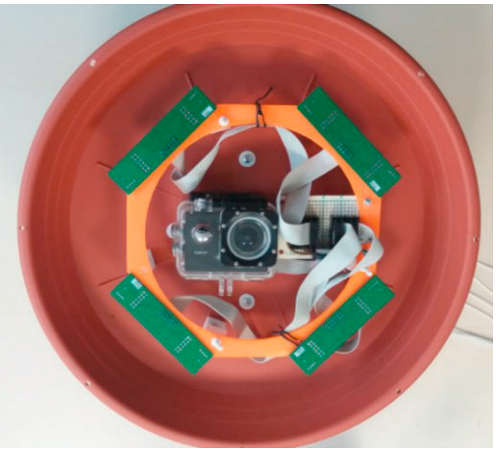

Fig. 2. The structure of the circular plastic frame [30]

3. Background Subtraction of Video Frames with Experiments

In this work, a real-time on-board background subtraction on the video frames has been implemented. This video processing technique was used for hovering drones whose function is that of surveillance and espionage of the underlying environment with particular application for detection of drones in the scene with a static background. This processing was performed on board, with edge computing fashion, in order to facilitate a future application of computer vision algorithms using AI accelerators on-board the SBC, as illustrated in our previous work in [32]. The background subtraction can help a motion-detection technique whose task is to detect all objects and entities moving in a scene. With a future use of an on-board classifier, the recognition and intelligent tracking of objects will make the system very autonomous and with advanced functions, as shown in the work of Seidaliyeva et al. in [13]. The video acquisition and processing software has been performed on the UAV's SBC and it has been written in Python using OpenCV (Open Source Computer Vision Library) for video processing. It is a cross-platform open-source software library used to develop computer vision algorithms in real-time applications. It also allows the user to view and process video sources, even in a complex way.

For background subtraction, the following steps have been followed:

1. acquisition on two different variables of the same flow
2. calculation of the subtraction between the two frames to identify the pixels that have undergone variation
3. application of the `cv2.COLOR_BGR2GRAY` filter to convert the video frames from a 3D color matrix (RGB) to a unidimensional matrix corresponding to a grayscale image

Fig. 3. Main steps of the background subtraction

4. application of the *GaussianBlur* filter to slightly blur the edges of the images and remove noise. This filter applies median value to central pixel within a kernel size
5. transformation of the grayscale blurred image into a binary image, according to a certain threshold
6. use of the dilate function to enlarge the white pixels in order to extend the identified area more evenly
7. old frame update with the current frame

The binary image transformation allows setting the value 255 (white color) for pixels that change their values in the next frame. The pixels that have not changed are set to 0 (black color).

The resulting processing is shown in Fig. 3 where it is possible to see the main processing steps. In the upper right corner, the original video frame is visible. On its right, there the frame subtraction is reported. In the lower right corner, it is possible to notice the threshold application and, on its right, the final processing where the changed pixels have been dilated in order to better identify the drone.

4. Music and DoA of Multi-Channel Acoustic Signals with Experiments

Let us consider N acoustic sources and an array with M microphones. Both the sources and the microphones are assumed to be with omnidirectional characteristics. In the free-field assumption, the discrete-time signal received by the m-th microphone can be modeled as:

$$x_m(k) = \sum_{n=1}^{N} \alpha_{nm} s_n(k - k_n - \tau_{nm}) + v_m(k) \tag{1}$$

where $\alpha_{n\,m}$ is the attenuation factor of the sound propagation that is inversely proportional to the distance from source n to microphone m, $s_n(k)$ are the unknown uncorrelated source signals, k_n is the propagation time from the unknown source n to the reference microphone in the array, $\tau_{n\,m}$ is the time difference of arrival between the m-th and the reference microphones for source n, and $v_m(k)$ is the additive noise signal at the microphone m and assumed to be uncorrelated with both the source signal and the noise at the other microphones. A detailed description can be found in [33].

Without entering in the details, the power pseudo-spectrum of the MUSIC beamformer output is given by:

$$P_{\mathrm{MUSIC}}(f,\tau) = \frac{1}{\mathbf{A}^H(f,\tau)\mathbf{G}(f)\mathbf{G}^H(f)\mathbf{A}(f,\tau)} \tag{2}$$

where $\boldsymbol{A}(f,\tau)$ is the steering vector corresponding to a given direction and $\boldsymbol{G}(f)$ is the matrix containing the eigenvectors corresponding to the noise subspace. MUSIC requires the analysis of eigenvalues for estimation of source number and it can be applied for DOA-based localization when $N \leq M$.

To mitigate the effect of incorrect response power estimation due to the variations of the SNR at each frequency and the geometric mean problem, a normalized arithmetic mean (NAM) is adopted. Specifically, to obtain a power spectrum in which each frequency gives the same contribution, a normalization on the power spectrum is implemented by imposing a constraint for the values to be in the range [0, 1]. The NAM can be written as:

$$P_{\mathrm{NAM}}(\tau) = \sum_{f=0}^{L-1} \frac{P(f,\tau)}{\max_{\tau'}[\mathbf{P}_{\tau'}(f)]} \tag{3}$$

where $P_{\tau'}(f)$ is the vector relative to the power for all the desired directions and $max[\cdot]$ denotes the maximum operator. NAM is effective when used in combination with MUSIC. From Eq. 3, the time of arrivals τ_n for the N acoustic sources can be estimated ($\hat{\tau}_n$) and, then, used to calculate the estimated DOAs of the source signals that are observed by the microphones. In [34] and [35], some variants of the MUSIC algorithm are applied to the localization of the noise emitted by a flying drone, and their performance is discussed.

This is still proposed research and it is on the way to be effectively implemented onboard the UAVs. However, to show the ongoing experiments with the real-time representation of estimated DOAs, we consider the 8-microphones UCA divided in two parts, each of them consisting of 4 microphones. This is obtained by separating the 8 signal channels into two groups consisting of 4 channels each. The estimated DOAs for each of the two groups can be represented through acoustic maps in terms of polar plot images as shown in Fig. 4. In this case, without loss of generality, the UCA collects the acoustic signals with two-dimensional acquisition capacity. Consequently, the DOA representation is related to the azimuth angle whose values, corresponding to the detected acoustic sources, are represented in the acoustic maps (horizontal axis).

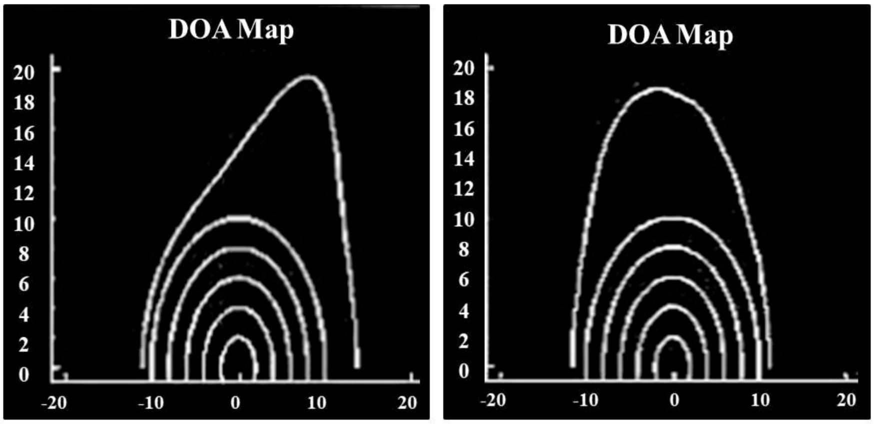

Fig. 4. Acoustic map representations of the estimated DOAs

In the same way, the acoustic maps can also be obtained when the two groups of 4 channels are merged together resulting in a single 8-microphones UCA processing.

5. Conclusions and Future work

An onboard audio and video processing approach in an UAV network for secure detection, localization, and tracking in counter-UAV applications was proposed. The objective is an increased level of cyber security and a reduced latency and delays for real-time applications. Additionally, processing of both audio and video signals can increase the level of accuracy and real-time effectiveness. Specifically, we saw the implementation of background subtraction applied to video frames and the MUSIC algorithm applied to multi-channel acoustic signals. The work in this manuscript proposed directions and guidelines for an onboard implementation of such a system. In particular, the proposed approach can be implemented in real streaming architectures with low-cost sensors mounted on UAVs. In future work, a fully and effective implementation of the two methods for both audio and video processing on low-cost and small-size processing units onboard the UAVs will be conducted in the same way as described in this manuscript. Merging of the two groups of 4 channels will allow a single 8-channel map with a better sound analysis. Microphone arrays with three-dimensional acquisition capacity will also be considered.

Acknowledgments

This research was partially supported by Italian MoD project a2018-045 “A proactive counter-UAV system to protect army tanks and patrols in urban areas” (Proactive Counter UAV), and by the ONRG project N62909-20-1-2075 “Target Re-Association for Autonomous Agents” (TRAAA).

References

[1] G. Udeanu, A. Dobrescu, and M. Oltean, “Unmanned Aerial Vehicle in Military Operations,” In: *Scientific Research and Education in the Air Force*, 18 (2016): 199–206, 10.19062/2247-3173.2016.18.1.26.

[2] G. B. A. Ronconi, T. J. Batista, V. Merola, “The Utilization of Unmanned Aerial Vehicles (UAV) for Military Action in Foreign Airspace,” In: *UFRGSMUN - UFRGS Model United Nations*, ISSN: 2318-3195, v.2 2014, p. 137–180.

[3] H. Wang, H. Cheng, and H. Hao, “The Use of Unmanned Aerial Vehicle in Military Operations,” In: *Long S., Dhillon B.S. (eds) Man-Machine-Environment System Engineering, MMESE 2020, Lecture Notes in Electrical Engineering*, vol 645, Springer, Singapore, https://doi.org/10.1007/978-981-15-6978-4_108.

[4] A. Toma, N. Cecchinato, C. Drioli, G.L. Foresti, G. Ferrin, “Towards drone recognition and localization from flying UAVs through processing of multi-channel acoustic and radio frequency signals: a deep learning approach,” In: *IST-190 Symposium on AI, ML and BD for Hybrid Military Operations (AI4HMO)*, 2021.

[5] Y. Kim, Y.G. Min, P.S. Hee, J. Wun-Cheol, S. Soonyong, and H. Tae-Wook, “The analysis of image acquisition method for anti-uav surveillance using cameras image,” In: *International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, pp. 549–554.

[6] H. Liu, Z. Wei, Y. Chen, J. Pan, L. Lin, and Y. Ren, “Drone detection based on an audio-assisted camera array,” In: *IEEE Third International Conference on Multimedia Big Data (BigMM)*, 2017, pp. 402–406.

[7] K. Hartmann and K. Giles, “UAV exploitation: A new domain for cyber power,” In: *8th International Conference on Cyber Conflict (CyCon)*, 2016, pp. 205–221, doi: 10.1109/CYCON.2016.7529436.

[8] F. Sanchez-Fernandez, P. Brunet and S. -M. Senouci, “Background subtraction for aerial surveillance conditions,” In: *2014 14th International Conference on Innovations for Community Services (I4CS)*, 2014, pp. 28–33, doi: 10.1109/I4CS.2014.6860549.

[9] R. Lingam, K. S. Kumar. “Statistically tuned Gaussian background subtraction technique for UAV videos,” In: *Sadhana*, 39 (2014): 785–808.

[10] J. Li, D. H. Ye, T. Chung, M. Kolsch, J. Wachs and C. Bouman, “Multi-target detection and tracking from a single camera in Unmanned Aerial Vehicles (UAVs),” In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 4992–4997, doi: 10.1109/IROS.2016.7759733.

[11] G. Tzanidou; P. Climent-Pérez; G. Hummel; M. Schmitt; P. Stütz; D. N. Monekosso; P. Remagnino, “Telemetry assisted frame registration and background subtraction in low-altitude UAV videos,” In: *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2015, pp. 1–6, doi: 10.1109/AVSS.2015.7301779.

[12] A. H. AlSaadi, I. Swamidoss, A. Almarzooqi, S. Sayadi, “Analysis of different background subtraction methods applied on drone imagery under various weather conditions in the UAE region,” In: *Proc.SPIE*, Sept. 2019, doi: https://doi.org/10.1117/12.2534644.

[13] U. Seidaliyeva, D. Akhmetov, L. Ilipbayeva, E. T. Matson, “Real-Time and Accurate Drone Detection in a Video with a Static Background,” In: *Sensors*, 2020, 20(14):3856, https://doi.org/10.3390/s20143856.

[14] F. Schilling, F. Schiano, and D. Floreano, “Vision-based flocking in outdoor environment,” In: *ArXiv, abs/2012.01245*, 2020.

[15] F. Tufano, C. V. Angelino, L. Cicala, “Visual Target Detection and Tracking in UAV EO/IR Videos by Moving Background Subtraction,” In: *Blanc-Talon J., Distante C., Philips W., Popescu D., Scheunders P. (eds) Advanced Concepts for Intelligent Vision Systems*, *ACIVS 2016*, *Lecture Notes in Computer Science*, vol 10016, Springer, Cham, https://doi.org/10.1007/978-3-319-48680-2_48.

[16] K. Okutani, T. Yoshida, K. Nakamura and K. Nakadai, “Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter,” In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 3288–3293, doi: 10.1109/IROS.2012.6385994.

[17] T. Yamada, K. Itoyama, K. Nishida and K. Nakadai, “Sound Source Tracking Using Integrated Direction Likelihood for Drones with Microphone Arrays,” In: *2021 IEEE/SICE International Symposium on System Integration (SII)*, 2021, pp. 394–399, doi: 10.1109/IEEECONF49454.2021.9382619.

[18] C. Yang, Z. Wu, X. Chang, X. Shi, J. Wo and Z. Shi, “DOA Estimation Using Amateur Drones Harmonic Acoustic Signals,” In: *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2018, pp. 587–591, doi: 10.1109/SAM.2018.8448797.

[19] K. Nakadai *et al*., “Development of microphone-array-embedded UAV for search and rescue task,” In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 5985–5990, doi: 10.1109/IROS.2017.8206494.

[20] C. Veibäck, M. A. Skoglund, F. Gustafsson, and G. Hendeby, “Sound source localization and reconstruction using a wearable microphone array and inertial sensors,” In: *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, 2020, pp. 1–8.

[21] T. Chen, Q. Huang, L. Zhang, and Y. Fang, “Direction of arrival estimation using distributed circular microphone arrays,” In: *2018 14th IEEE International Conference on Signal Processing (ICSP)*, 2018, pp. 182–185.

[22] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, “A survey of sound source localization methods in wireless acoustic sensor networks,” In: *Wireless Communications and Mobile Computing*, vol. 2017, pp. 1–24, 2017.

[23] M. Basiri, F. Schill, P. Lima, and D. Floreano, “On-board relative bearing estimation for teams of drones using sound,” In: *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 820–827, 2016.

[24] V. Didkovskyi, S. Kozeruk, and O. Korzhik, “Simple acoustic array for small uav detection,” In: *2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO)*, 2019, pp. 656–659.

[25] T. Yamada, K. Itoyama, K. Nishida, and K. Nakadai, “Sound source tracking by drones with microphone arrays,” In: *2020 IEEE/SICE International Symposium on System Integration (SII)*, 2020, pp. 796–801.

[26] C. Drioli, G. Giordano, D. Salvati, F. Blanchini, and G. L. Foresti, “Acoustic target tracking through a cluster of mobile agents,” In: *IEEE Transactions on Cybernetics*, pp. 1–14, 2019.

[27] S. Mischie and G. Găşpăresc, “On using respeaker mic array 2.0 for speech processing algorithms,” In: *2020 International Symposium on Electronics and Telecommunications (ISETC)*, 2020, pp. 1–4.

[28] M. B. Andra, B. P. A. Rohman, and T. Usagawa, “Feasibility evaluation for keyword spotting system using mini microphone array on uav,” In: *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 2264–2267.

[29] Y. Yamazaki, C. Premachandra, and C. J. Perea, “Audioprocessing-based human detection at disaster sites with unmanned aerial vehicle,” In: *IEEE Access*, vol. 8, pp. 101 398–101 405, 2020.

[30] D. Salvati, C. Drioli, G. Ferrin, and G. L. Foresti, “Acoustic Source Localization From Multirotor UAVs,” In: *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8618–8628, 2020, doi: 10.1109/TIE.2019.2949529.

[31] D. Salvati, C. Drioli, A. Gulli, G. L. Foresti, F. Fontana, and G. Ferrin, “Audiovisual active speaker localization and enhancement for multirotor micro aerial vehicles,” In: *Proceedings of 23rd International Congress on Acoustics (ICA)*, Aachen, Germany, Pages 3338–3345, 2019.

[32] A. Toma, N. Cecchinato, C. Drioli, G. L. Foresti and G. Ferrin, “CNN-based processing of radio frequency signals for augmenting acoustic source localization and enhancement in UAV security applications,” In: *2021 International Conference on Military Communication and Information Systems (ICMCIS)*, 2021, pp. 1–5, doi: 10.1109/ICMCIS52405.2021.9486424.

[33] D. Salvati, C. Drioli and G. L. Foresti, “Incoherent Frequency Fusion for Broadband Steered Response Power Algorithms in Noisy Environments,” In: *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 581–585, May 2014, doi: 10.1109/LSP.2014.2311164.

[34] D. Salvati, C. Drioli, G.L. Foresti, “Power method for robust diagonal unloading localization beamforming,”
In: *IEEE Signal Processing Letters*, 26 (5), art. no. 8676348, 2019, pp. 725–729.

[35] D. Salvati, C. Drioli, G. Ferrin and G. L. Foresti, “Beamforming-Based Acoustic Source Localization and Enhancement for Multirotor UAVs,” In: *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 987–991.