# DEEPLOMATICS Project : Deep-Learning for the Real-time Multimodal Localization and Identification of small UAVs


**S. Hengy[1], E. Bavu[2], A. Matwyschuk[1], P. Naz[1], B. Fischer[1]**

1: French-German Research Institute of Saint-Louis
P.O. Box 70034, 68301 SAINT-LOUIS Cedex, France

sebastien.hengy@isl.eu

2: Le CNAM / LMSSC / EPN 04, 2 rue Conté, 75003, Paris, France

eric.bavu@lecnam.net



## *ABSTRACT*

*This paper will present the principle of the DEEPLOMATICS demonstration, based on a network of distributed acoustic modules and optronic sensors. During the initial detection phase, the fusion of the results of the acoustics sensors allow an approximate localisation of the intrusion. Then, this zone is visualised by an optronic system based on several wavelengths and two fields of view cameras which allows a precise and automatic detection of the UAV in the scene and then transmits a zoomed image of the intrusion to the operator. The acquired signals are exploited according to an original method developed in collaboration between ISL and CNAM laboratories. This method is based on remote artificial intelligence allowing improvement of the perception of threats and the operational situation.*

*To conclude, results of field trials will be presented to illustrate the capabilities of the DEEPLOMATICS demonstrator in term of complementarities between acoustics and optical sensors, and between the images provided by the various camera (Visible, SWIR, Active imaging …).*


## 1.0 INTRODUCTION

The acoustic detection and localization of different threats, including short duration events (artillery, gun shots) or low signature events (small UAVs) is particularly challenging in complex environments. Combining the acoustic sensors with other sensors, such as optical detection technologies, improves the overall detection capability. In the previous internal ISL project OASyS² (Optical and Acoustical System for Security and Surveillance) we have developed technological components for counter-UAV systems which can be combined with different technologies and subsystems [1]. Recently the project DEEPLOMATICS has used the basis studied previously and has investigated the use of up to date sensors and signal processing technics [2] radically improving system performance. New acoustic antennas using multiple MEMS microphones and exploiting AI concepts for efficient detection offer an affordable possibility to detect intrusions when used as peripheral detectors. These concepts may also be derived for the personal individual equipment in order to constitute the basis of a cooperative protection system for a group of soldiers.

### 1.1 Context

A detailed knowledge of the enemy's position, identification and localization of potential threats, as well as a continuous surveillance of strategic zones, for example in order to detect intrusions, are crucial for the successful completion of a mission. They not only provide the necessary information to select the best appropriate protection measures, but also help and optimise the own strategic decisions. Within the very wide range of different detection and sensing technologies, the ISL contributes with selected topics, providing complementary technical capabilities, to the overall goal of a comprehensive situational awareness.

## 1.2 General objectives of the DEEPLOMATICS project

Several microphone arrays are deployed on the area to be protected. Within its coverage area, each microphone array simultaneously localizes and identifies flying drones using a deep learning approach based on the Beam Learning network. Each array is attached to a local AI which processes spatial audio measurements in real-time, independently to the other units of the surveillance network.

A data fusion system refines the estimates provided by each of the AI-enhanced microphone arrays. This detected position is shared in real-time with an optronic system. Once this system has hooked its target, a Deep Learning tracking algorithm is used to allow an autonomous visual tracking and identification.

The optronic system is composed of various cameras (visible, thermal, and active imaging) mounted on a servo-turret. The active imaging system can capture scenes up to 1 km, and only captures objects within a given distance, which naturally excludes foreground and background from the image, and enhances the capabilities of computer vision.

The DEEPLOMATICS project combines benefits from acoustics and optronics to ensure real-time localization and identification of drones (Fig. 1), with a high precision (less than 7° of absolute 3D error, more than 90 % detection accuracy). The modular approach also allows to consider in the long term the addition of new capture systems such as electromagnetic radars.

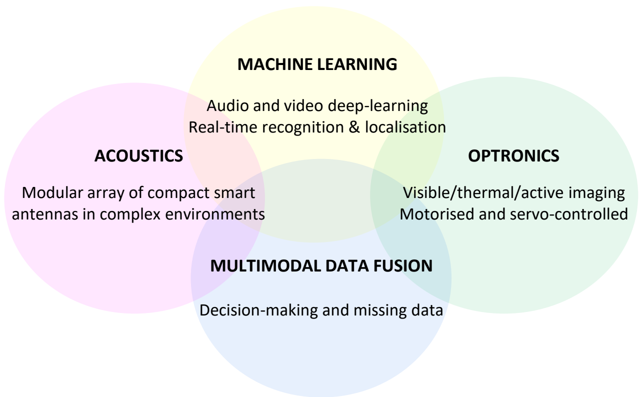



**Figure 1: Interdisciplinary and multimodal approach implemented**

# 2.0 ACOUSTIC DETECTION

## 2.1 Noises of small UAV – Measurement of acoustic data

In order to optimize the acoustical array, we have first measured the acoustical signals in an anechoic chamber in “free flight conditions” (Fig. 2). We have recorded the noise generated by a DJI Phantom 3 (quadcopter, 2 blades per propeller), a Parrot Bebop (quadcopter, 3 blades per propeller) and a Parrot Mars (quadcopter, 2 blades per propeller). …. The authors show clearly the contribution of the Blade Passing Frequency (BPF, 128 Hz) and harmonics, then the contribution of the electric motors at mid frequency (600 – 6000 Hz). The individual contribution of 1, 2 and 4 rotors at nominal rotational speed (around 6000 RPM) was also characterized (Fig. 3).

ISL organized also free field experiments with various acoustic arrays and UAVs in order to acquire acoustic signals in realistic conditions (real flights and various environmental conditions). This section gives a short overview of the experiments done and of the results obtained. These data have been used to characterize the physical content of the signals and to contribute to the acoustic database.

For the free field experiments, a Real Time Kinematics (RTK) system based on GNSS receivers (Global Navigation Satellite System) was implemented on a Base Station and an extra GNSS receiver on the UAV. This RTK System is used to collect ground truth data in a local “East North Up” (ENU) coordinate System. It was used in order to have precise localization of the UAV for post-processing analysis of the acoustic data. To compensate the extra weight of the GNSS/Receiver device we have removed the camera and corresponding gimbal.

Acoustic arrays using conventional metrological microphones as well as prototypes of new arrays using MEMs microphones have been deployed at different points of the site. Time segments of the signals recorded by the microphones of these arrays have been used in order to build the acoustic database that will be used for the classification study and for the localization study (for more details see paragraph 2.1).

**Figure 2: Illustration of the experimental set-up (anechoic chamber and open-field position)**

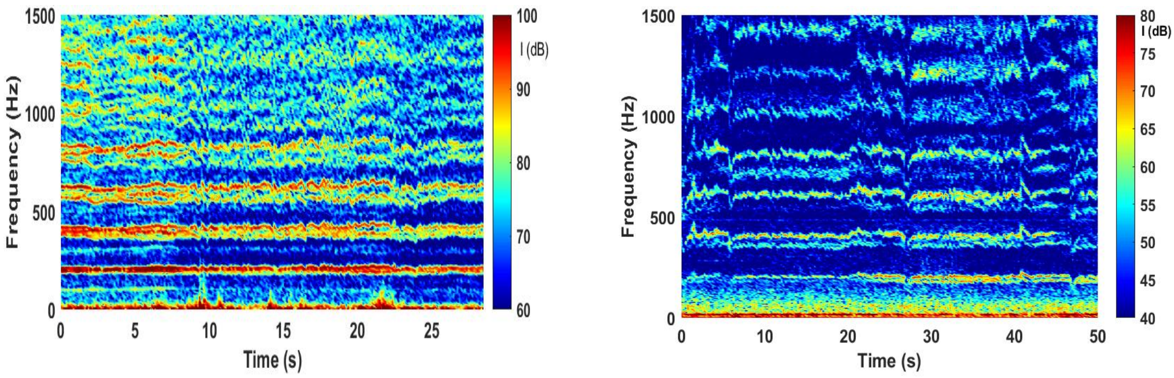



**Figure 3: Examples of acoustic spectrograms (Phantom 3 hovering and in a translation flight)**

## 2.1 Microphone array

From a technological point of view, the selected acoustic antennas are based on digital MEMS microphones (of the order of 20 elements), which have decisive advantages in terms of signal-to-noise ratio, miniaturisation, and densification of portable antennas. This relatively low-cost technology makes it possible to deploy a large number of sensors over large areas. Each module is linked by a USB (Universal Serial Bus) connection to a miniature and light central unit, carrying out the localisation and target recognition tasks locally, and integrating the deep neural networks pre-trained thanks to the databases created during the project.

**Figure 4: View of an acoustic sensor with its protective windshield and a local processing unit**

Each microphone array is attached to an electronic unit supporting the implementation of a deep neural network, trained for source localization and sound signature identification tasks. The neural network is a variant of the BeamLearning architecture [7] that we previously published for sound source localization. This variant of the network, Beamlearning-ID, has been specifically designed to simultaneously perform the recognition and localization tasks in real time [8]. The specialized AIs have been trained on a multi-channel dataset of acoustic signals from small UAVs in flight, under realistic conditions. These data acquisitions are augmented by a 3D spatializer. This augmentation will allow the neural network to respond as efficiently as possible to the localization and source identification tasks that will be performed simultaneously by the AI modules at the output of the compact microphone arrays.

Each intelligent acoustic module in the network will then transfer its recognition and location data to fuse the information and communicate it to the fusion and decision centre then to the optronic device.

## 2.2 IA training / databases

A multichannel dataset of multichannel audio data was built throughout the DEEPLOMATICS project to train the Beamlearning-ID network for drone localization and recognition. The acoustic signals recorded by the microphone arrays intrinsically convey information on the position of the acoustic source and its nature. The objective of the developed BeamLearning-ID deep network is to retrieve this information through supervised learning. Supervised learning requires a priori knowledge of this information. The audio data must therefore be annotated with the position and nature of the drone in flight.

The entire acoustic dataset is heavily annotated. To achieve this tedious task, a semi-automated process has been developed during the DEEPLOMATICS project. During the flight of the drones, a GPS-RTK beacon is mounted on the drone and allows to know the position of the drone in real time. In parallel, several ambisonic microphone arrays record the 3D sound scene. The GPS and acoustic data are then synchronized. Moreover, a 3D spatialization step of the sound scene can be implemented if the antenna used to record the sound scene on site is different from the one used for the inference unit using the BeamLearning-ID network. This 3D spatialization process has been developed to produce an automated annotation of the multichannel audio (with labels denoting the drone model, and its 3D position synchronized with audio data) [7].

During the DEEPLOMATICS project, various measurement campaigns allowed us to accumulate more than 34 hours of usable data of UAV flying data (simultaneous measurements of multichannel audio using high-order ambisonic microphone arrays and geo-referenced position using a high precision realtime kinematics (RTK) GPS carried by the flying drones).

Such a large and realistic database allows deep neural networks to extract hidden patterns in the observation data. The size of the dataset is obviously not the whole story. For the deep neural network to be effective, it is necessary to build a dataset with a large variability of data. This is the reason why supercomputers are now able to run algorithms exceeding human capabilities in the field of image recognition. Image recognition researchers are now looking for ways to generate realistic synthetic images to train neural networks where the data sets are not yet large enough. We have, for localization or acoustic recognition tasks, access to a tool that allows to lift this lock and to generate simulated, augmented, or modified databases, while respecting the realism and the physical validity of the 3D pressure field.

The CNAM/LMSSC has developed in the last few years a device that will allow to offer to localization and identification techniques by Deep Learning a flexibility and a realism not reached until now (Fig. 4).

Two tools developed and validated at the LMSSC are at our disposal to contribute to these objectives [6-7]. The first device allows the spatialized capture of the sound environment, used in the measurement campaigns, and the second allows the restitution of the three-dimensional field. These two devices allow us to render the 3D pressure fields of drones in flight on compact microphone arrays to train their individual artificial intelligence, even if the specific microphone arrays were not used for the on field experiments. One of the major advantages of this process is that we will also be able to "augment" the data captured during the measurement campaigns, by superimposing the 3D field of a large number of noisy environments, corresponding to potential locations for the installation of smart compact antennas (Fig. 4). These environments are also recorded by High-Order ambisonic (HOA) microphonic arrays.

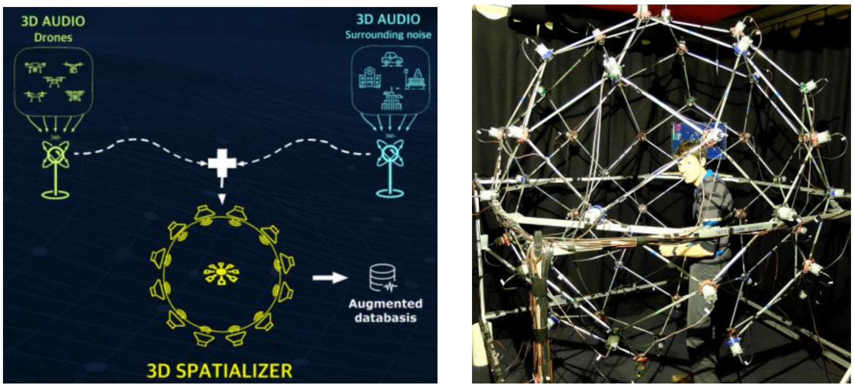



**Figure 5:** Schematic of the data augmentation strategy using the CNAM/LMSCC 3D spatializer

## 2.3 Acoustic distributed sensors

Results of the acoustic processing at the level of each microphones sensors are transmitted through a simplified C² (Command and Control) network to a Central PC (Personnal Computer) with a data fusion process (Fig. 6). The exchange of data is done using three types of frames adapted to the transfer of information and respecting the NMEA format defined for Deeplomatics (proprietary format):

– PABET : Proprietary Acoustic Bearing and Elevation to Target, for information transmission between acoustic sensors and fusion, 40 Hz

– PAOPT : Proprietary Actual OPTical Device orientation, for information transmission between optical sensor and fusion, 2 Hz

– PATLL : Proprietary Actual Target Latitude and Longitude, for information transmission between fusion and optical sensor, motorised tripod control, 5 Hz.

The real time visualisation through the Man-Machine Interface of the detection/localisation events is illustrated in fig. 7. The visualisation of the estimated trajectory as tracked during a test can be displayed after the test in a post-processing phase (Fig.8).

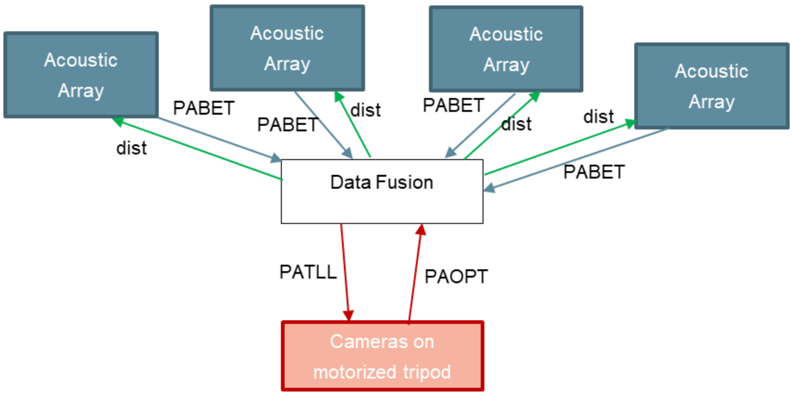



**Figure 6:** Concept of the fusion of the messages generated by the 4 acoustical sensors

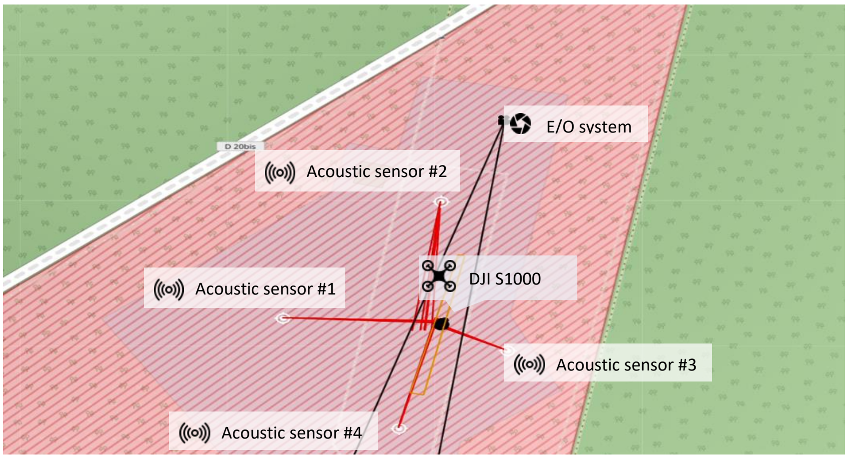



**Figure 7:** Man-Machine Interface with the acoustic arrays and the E/O system deployed during the experiments

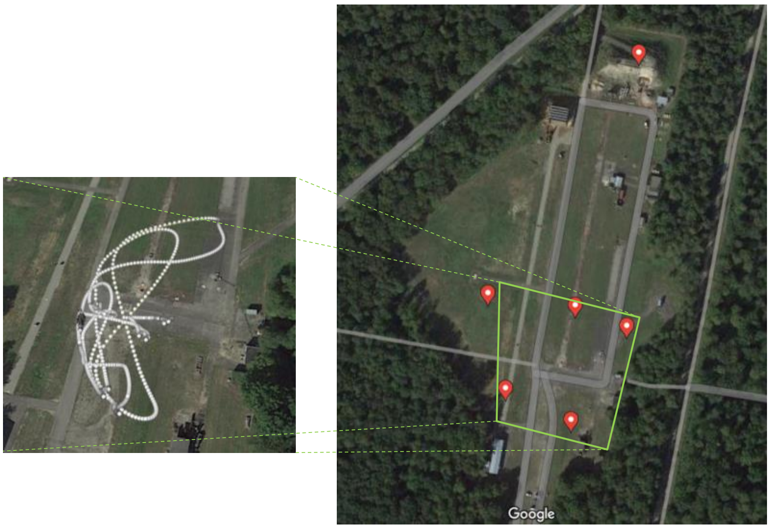



**Figure 8:** Tracking of a complex flight (case with 5 acoustic arrays)

# 3.0 ELECTRO-OPTICAL SENSORS

Electro-optical sensors such as CCD cameras are commonly used to record intensity images which are processed by automatic tracking algorithms or by a human operator in order to localize a flying object. Due to a very narrow field of view (FOV), optical sensors have to be precisely oriented in the direction of the object under investigation and, thus, rely on external sensor information, e.g. range and height information given by acoustics and/or radar.

In most cases, the tracking of a flying object with the sky as a background is not a problem: the contrast of such images is always very high and the system is able to track the target. However, when the drone is flying in front of a structured background such as trees or buildings, it can become very complicated to distinguish the object and a vision system capable of giving an image of the threat without its background could be of great help for tracking. This is precisely what a range-gated active imaging system can do: if the gate is thin enough and placed at the right distance, this technique is able to suppress the foreground and the background around the object (see in paragraph 3.2).

The electro-optical sensor has been also used to acquire images during filed tests organized by the NATO technical group SET-260 “Assessment of EO/IR Technologies for Detection of small UAVs in an Urban Environment” (Fig. 9). These data have been used to optimize the image processing algorithms.

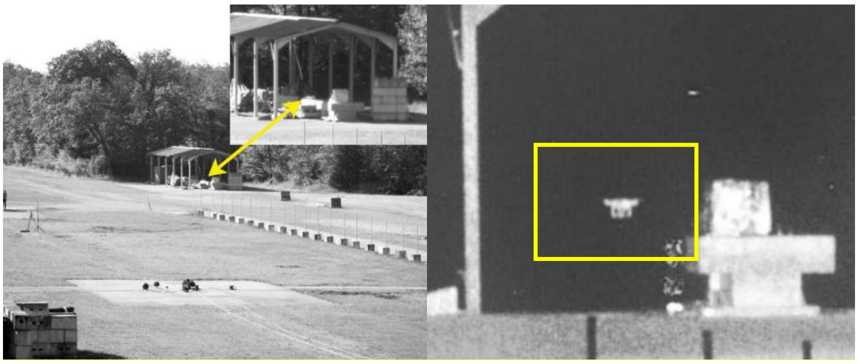
**Figure 9:** Tests in urban condition organized by the NATO S&T/SET-260 working group

## 3.1 Optronic turret with multiple cameras

In figure 10, it can be observed that the sensor consists of a colour camera associated with a range-gated active imaging system. The latter consists of a solid-state laser emitting light in the eye-safe short-wave infrared (SWIR) spectral range coupled with a SWIR camera. The laser beam is collimated in an optical fibre. At the fibre output, a specific collimator with lenses is installed to respect the desired divergence angle and to homogenize the laser beam. The imaging sensor is an InP/InGaAs EBCMOS array from INTEVAC (LIVAR M400) with a resolution of 640 × 480 pixels. To ensure a good resolution on the flying target, a 200 mm OPTEC SWIR lens is used. This lens provides an active imaging FOV of 1.97° × 2.46°. The system is placed on a motorized pan and tilt device capable of rapidly moving in the direction of the detected threat.

The use of short laser pulses and short exposure times on the sensor unit makes it possible to perform the range-gated imaging of the scene. Only the light which arrives at the sensor within a certain timing window contributes to the imaging process. In figure 11 the wide-angle colour image of the UAV and the narrow-FOV laser gated image of the same object can be observed. While the UAV is more or less invisible in front of the trees in the colour image, the contrast gain on the UAV with the active imaging system is clearly visible.


**Figure 10:** Acoustic array, E/O system deployed during the experiments



(a) (b)

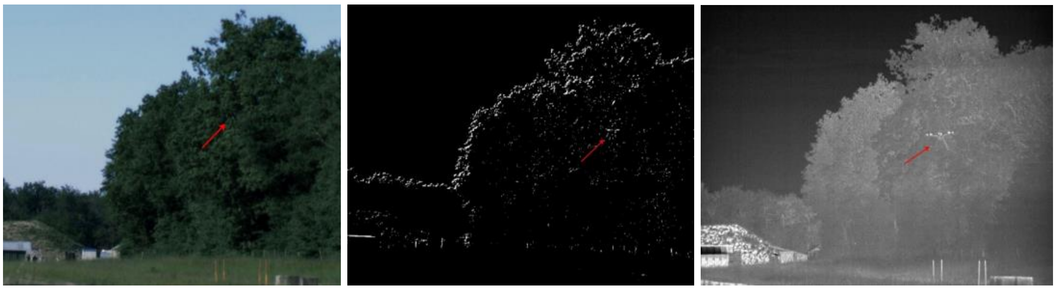
**Figure 11:** Example of the reconnaissance of the UAV in the image and of tracking process (a) in the wide-angle colour image and (b) the narrow-FOV laser gated image

## 3.2 Acquisition of images

As evocated previously, we have to deal with some challenging constrains such as:

- resolution limit
- detection with a structured background (trees, …)
- mound
- glare
- false alarm such as birds, reflection of the drone image on windows
- night flights ….
- multiples drones

The images recorded by the various cameras types are used to define how to of will allow us to overcome these constrains. As an illustration, examples of the vision capacity of this system are reported hereafter (Fig. 12 and 13).



(a) Visible image (b) Polarized image (+) (c) Thermic image (++)

**Figure 12:** Example of images with trees in the background

(a) Active imaging (b) Thermic image (=)

**Figure 13:** Example of images during the night

### 3.3 IA training / databases

The recognition and the tracking of the drone were performed from the images of the visible camera with the small field of view. The images of other cameras were used to counter difficult situations that could lose the tracking. The deep-learning has been validated after several terabytes of images. Across all the scenarios and the backgrounds used, the obtained results gave an average detection rate of 94%.

## 4.0 CONCLUSION

The principle of the DEEPLOMATICS Project studied in collaboration between ISL and CNAM laboratories was presented in this paper: the final demonstration in May 2022 has demonstrated that the combination of a network of distributed acoustic modules and optronic sensors is able to detect and localize small UAVs.

During the initial detection phase, the fusion of the results of the acoustics sensors allow an approximate localisation of the intrusion. The acquired signals are exploited according to an original method based on remote artificial intelligence allowing improvement of the perception of threats and the operational situation.

Real-time implementation of deep learning-based methods applied on raw acoustic data allows for simultaneous the detection, reconnaissance and also the localization in term of azimuth and elevation of continuous noise sources such of small UAV with the following characteristics:

- Angular error < 3°
- Detection rate > 95%
- Recognition rate > 85%

Data fusion of this information enhances the identification capability (95% recognition for DJI phantom 3, 99% for DJI S1000 after data fusion) and also allows transmission of a geo-referenced (latitude/longitude/altitude) information. The localization accuracy has been measured during 95 minutes of active drone tracking around the project's final field experiment, with a median radial error of 10.7 meters and a standard deviation of 15.6 meters. This information is then sent to the optronic pod equipped with visible/thermic/active imagery camera types for threat identification and refined tracking. It allows a precise and automatic detection of the UAV in the scene and then transmits a zoomed image of the intrusion to the operator.

The tests on the basis of non-cooperative drones validated:

- the robustness and performances of an original deep-learning approach,
- the acoustic / imagery complementarity,
- and the need for different and complementary optronic data (i.e. Thermal, Multispectral, Polarisation …) to improve the tracking of small UAVs.

Further studies should address more complex scenarios: several UAVs flying in different directions, urban areas ... with the aim of improving the resilience of the system.

**ACKNOWLEDGMENTS:**

The DEEPLOMATICS project has been funded by the DGA/AID Grant No. ANR-18-ASTR-0008. This study is done in cooperation between the ISL, the CNAM (LMSCC and Cedric teams) and the SME Roobost. Thanks to all the members who have participated to the project and in particular to the preparation and realisation of the field experiments:

- Éric Bavu, LMSSC, Cnam Paris, HESAM Université
- Hadrien Pujol, LMSSC, Cnam Paris, HESAM Université
- Alexandre Garcia, LMSSC, Cnam Paris, HESAM Université
- Christophe Langrenne, LMSSC, Cnam Paris, HESAM Université
- Sébastien Hengy, French-German Research Institute of Saint-Louis
- Oussama Rassy, French-German Research Institute of Saint-Louis
- Nicolas Thome, CEDRIC, Cnam Paris, HESAM Université
- Yannis Karmim, CEDRIC, Cnam Paris, HESAM Université
- Stéphane Schertzer, French-German Research Institute of Saint-Louis
- Alexis Matwyschuk, French-German Research Institute of Saint-Louis

## REFERENCES

[1] Christnacher F., Hengy S., Laurenzis M., et al., Optical and acoustical UAV detection, Electro-Optical Remote Sensing X 9988, SPIE Security+Defence, Edinburgh, UK, September 26-29, 2016

[2] Bavu E. et al, Deeplomatics: A deep-learning based multimodal approach for aerial drone detection and localization, 2nd e-symposium “Quiet Drones” on UAV/UAS Noise, 27th to 30th June 2022

[3] Naz P. et al., Outdoor field trials for the measurement of the acoustic signals of mini UAVs, Forum Acusticum 2000

[4] Hengy S. et al.: “Multimodal UAV detection: study of various intrusion scenarios”, Proc. of SPIE Vol. 10434, Oct. 5, 2017.

[5] Ramamonjy A.: “Développement de nouvelles méthodes de classification/localisation de signaux acoustiques appliquées aux véhicules aériens”, PhD thesis, CNAM Paris, 28 Mai 2019.

[6] Pujol, H., Bavu, E., Garcia, A., Langrenne C., Hengy S., Schertzer S., Matwyschuk A. (2022, April).

Deeplomatics : Localisation et reconnaissance acoustique de drones. 16ème Congrès Français d'Acoustique, CFA 2020.

[7] P. Lecomte, Ambisonie d'ordre élevé en trois dimensions: captation, transformations et décodage adaptatifs de champs sonores, Thèse de doctorat: Paris, CNAM, 2016.