https://doi.org/10.2352/ISSN.2470-1173.2019.15.AVM-035

An Autonomous Drone Surveillance and Tracking Architecture

Eren Unlu; ISAE-SUPAERO; Toulouse, France
Emmaneul Zenou; ISAE-SUPAERO; Toulouse, France
Nicolas Riviere; ONERA; Toulouse, France
Paul-Edouard Dupouy; ONERA; Toulouse, France

Abstract

In this work, we present a computer vision and machine learning backed autonomous drone surveillance system, in order to protect critical locations. The system is composed of a wide angle, high resolution daylight camera and a relatively narrow angle thermal camera mounted on a rotating turret. The wide angle daylight camera allows the detection of flying intruders, as small as 20 pixels with a very low false alarm rate. The primary detection is based on YOLO convolutional neural network (CNN) rather than conventional background subtraction algorithms due its low false alarm rate performance. At the same time, the tracked flying objects are tracked by the rotating turret and classified by the narrow angle, zoomed thermal camera, where classification algorithm is also based on CNNs. The training of the algorithms is performed by artificial and augmented datasets due to scarcity of infrared videos of drones.

Introduction

Due to rapidly increasing accessibility of commercial UAVs, -publically known as *drones*-, great security and privacy violation risks appeared. These threats include risking aviation security by drones flying near airports, possible terrorist attacks with explosive payloads, illegal trafficking etc. We can give several examples around globe in recent years, such as the drone crash near White House [1], the protest during German chancellor's speech [2], multiple drones appearing around several French nuclear power plants alerting officials [3], a close collision risk avoided between a drone and a commercial airplane on LAX airport [4], unearthed illegal drug smuggling scheme by employing drones between Mexico and USA [5] and many more.

Several counter-measures in the market and academia offer autonomous detection, tracking and identification of drones, which is a primordial feature for continuous and efficient operation. The proposed systems use either RF signal detection (used for the communication between device and the ground operator) [6], acoustics [7], RADAR [8], LIDAR [9] or common passive optics (cameras) backed by computer vision algorithms [10].

One of the popular approaches for drone detection is to capture and intercept RF signals, where it is used for communication between the drone and the ground operator [11]. However, it misses the point that drones may be pre-programmed to fly to target without needing communication. Acoustics has been used also to detect drones by employing microphone arrays [12]. The aim is to classify specific sound of rotors of drones, however they fail to achieve high accuracy and operational range. Maximum range of audio assisted systems stay below 200-250 meters. Another disadvantage is the non-feasible nature of the system in urban or noisy environments such as airports.

Optics has been regarded as the most robust, reliable and efficient counter-measure in the community, with backing from state-of-the-art computer vision algorithms [13]. The tendency to include at least one optical element (RGB/infrared cameras) with computer vision algorithms can be observed in the market [14][15][16]. The recent breakthrough in object detection and recognition thanks to the deep learning algorithms has transfigured the perception of computer vision. Convolutional Neural Networks (CNNs) has already become the de facto approach for detection and recognition tasks [17][18]. In addition to developments in the field, decreasing cost per memory of GPU resources and increasing open source visual datasets via images and videos on internet has fueled this phenomenon. There are already various articles published in recent few years, proposing to use computer vision for autonomous drone surveillance task uses deep learning [19][20][21].

Due to reasons listed above, we have based our autonomous system on visual and thermal cameras backed with deep learning computer vision algorithms. The system is composed of a wide angle RGB camera which serves as a primary detector of possible aerial threats and a narrow angle thermal camera mounted on a rotating turret, which is used for further identification. Both for detection and classification purposes, we use YOLO deep learning architectures [22]. Due to limited GPU memory, we have investigated the performance of a lightweight detection architecture, where classification is handled by a separate architecture. These two architectures are trained to be used both for visual and infrared imagery, in an attempt to save resources. After the introduction of the system and the associated algorithms, we present experimental results obtained from various field tests and simulations.

System Overview

The system can be divided in to three instrumental parts as; a wide angle (16 mm focal length) high performance industrial RGB camera (2000x1700 pixels at approximately 25 FPS) placed on an adjustable static platform, a high performance, rapid rotating turret, where a narrow angle (41 mm focal length) infrared camera is mounted and lastly a Linux PC with a 2 GB Nvidia GPU which can run deep learning algorithms. The rotating turret and cameras can be seen in Figure 1.

Based on a modified lightweight YOLOv3 architecture, we detect the small intruders on the wide angle RGB camera's image plane (we refer it as *main image plane* further in the document). At this first stage, false alarms up to a degree is acceptable, where they are tracked and based on their movements and visual signatures they may be inspected by rotating the turret towards it and analyzed with the narrow angle thermal camera. Note that, even

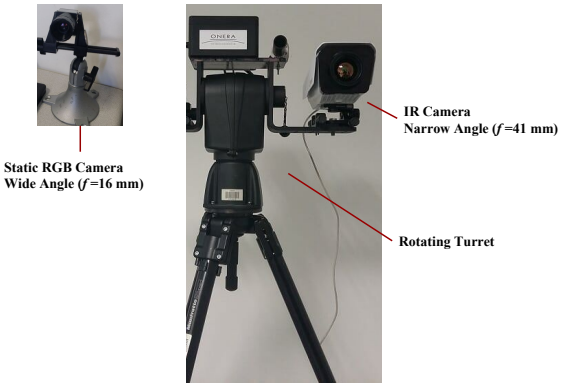


Figure 1. *The static wide angle (f = 16 mm) RGB camera with its adjustable platform and the rapid, versatile rotating turret with narrow angle (f = 41 mm) thermal camera.*

false alarms are acceptable it is crucial for system to minimize them for a proper and reliable operation. In addition to zoomed view, thermal camera has the advantage of higher background contrast, richer discriminatory visual features and operability under harsher weather conditions compared to RGB camera. As infrared cameras are lower in resolution, using them with a narrow angle lens in collaboration with a wide angle RGB camera is straightforward.

Detection on Main Image Plane

First purpose of our system is to be able to detect very small drones with substantially low false alarm rates on the main image plane. In order to achieve this we use a lightweight YOLOv3 architecture which requires a low GPU memory (approximately 1 GB) and operates with high FPS. YOLO algorithm operates in a different manner, more similar to Single Shot Detectors (SSDs), where a regression approach is followed. YOLO allows for combined detection and classification of the objects. However, we have observed that following a *divide-and-conquer* approach is more effective; where the detector and classifier architectures are separate. Especially, in the case of small object detection this effect is more pronounced. Based on our observations, discriminating objects require more wider and deeper, complex architectures; while detecting small objects with an acceptable false alarm rates does not. Note also that, detection is performed on a larger input image size (832x832 pixels), whilst we perform classification with 64x64 pixels of size. This is because, classification can be done only on the region of interest. Therefore, separating two processes in to two distinct architectures is profitable.

In this work, primarily we have investigated the performance of small aerial object detection with lighter convolutional neural networks, especially in width (e.g. less number of filters). The results are surprisingly good enough proving even drones as small as 8x8 pixels on 832x832 pixels image plane can be detected, with very low false alarm rates (virtually none in most of the cases) compared to conventional methods. As mentioned previously, the classification is separated, thus the lightweight detector captures not only drones, but birds, airplanes and other aerial vehicles/objects; however false alarms triggered by background clutters/objects are drastically reduced.

YOLO architectures third version, YOLOv3 has introduced the concept of upsampling, where in deeper layers, tensors are upsampled (interpolation) by two and feature maps are routed from the previous layers [23]. Different than other convolutional neural network detection methods, YOLO uses regression with predefined anchor boxes to estimate the position of objects. We have observed that, this allows it to capture semantic context better, especially in case of small drone detection.

The proposed lightweight architecture is derived directly from the YOLOv3 architecture, where number of layers is kept same. In other words, the depth of the architecture is same with the YOLOv3. However, as it can be seen from Figure 2, number of filters for all layers except the third upsampling layer, which is specifically responsible of small object detection are set to 16. By reducing number of filters drastically, we have saved a sizeable GPU memory. We have used this memory to use a two times larger input size (832x832). This can be seen as a trade-off between larger input size which permits much better detection of small object and the width of the architecture. Even with increased input size, the GPU memory of this detector is lower than 800 MBytes. This configuration is found to be the optimal based on our tests.

Tracking and Possible Threat Object (PTO) Concept

The objects detected by this lightweight architecture are immediately assigned as *tracks*. Each track is followed by a distinct Kalman Filter, which is derived from [24]. This framework also uses a Hungarian algorithm based detection to track assignment algorithm, where cost is the Euclidean distance between centroids of detection bounding boxes and tracks as in Figure 3. There are various reasons for the choice of this scheme. First, it favors linear movement, which we consider as an interesting flight path for a context (an object deliberately targeted towards a location), where small aerial objects are expected from the horizon. Of course drones may make non-linear movements, however an approaching target of interest shall have the priority. Note that, if the object is closer, its displacement between frames would be higher, as it appear larger. Thus, even a closer object is not doing a linear movement, it would be favored. Second, as only last T_{mov} frames are considered, this allows the tracking of new detections every time. Even a track is lost, it can be redetected. Even false alarm rates (counter background clutter and non-aerial objects) of our detector is very low, the presented scheme eliminates the favoring of false alarms in tracking.

We introduce a concept which is called Possible Threat Object (PTO). It is a track, that the algorithm decides to inspect with lower angle thermal camera by rotating the turret towards it. The calibration process, which refers to the task of determining the direction (pan and tilt) of the turret as a function of the position on the main image plane. Due to the limited space, this is out of scope of this paper. At an instance, there can be only one PTO. The tracks are chosen as PTO, inspected with the zoomed thermal camera for a while and designated to be a threatening drone or not, in the order with respect to prioritization method explained above. If a track is decided to be not a drone, it shall not be designated as PTO again, even it is still being tracked on main image plane.

A PTO is evaluated in periodic windows temporally, where

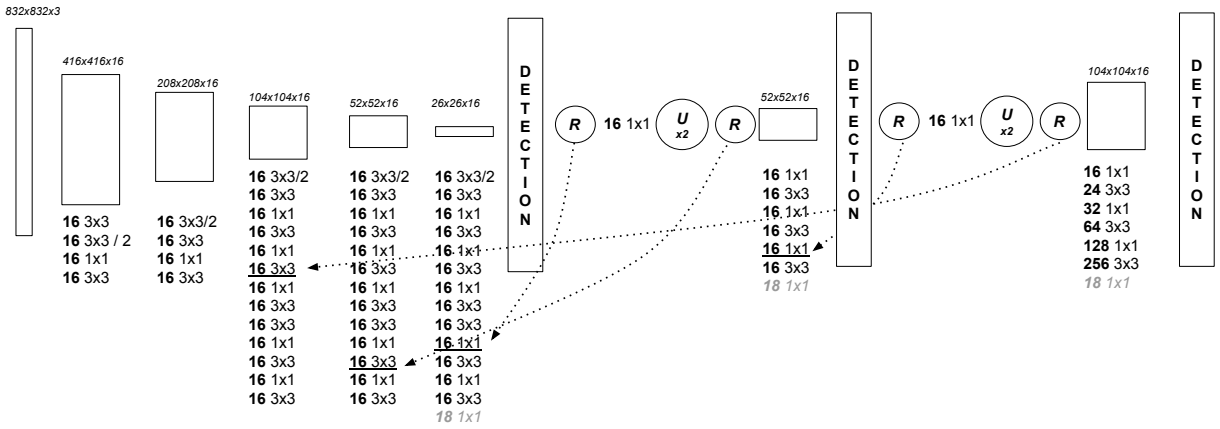


Figure 2. *Our lightweight, narrow version of the YOLOv3 architecture with 3 scale detection layer. Different than the default architecture, we prefer to use 832x832x3 input size to better detect the small drones. The circles with R denotes the route layer and the U denotes the upsampling layer. Note that, the previous layer filters, where the upsampled feature matrix is routed is underlined. Also, note that the layers where the positional dimensions of the feature matrix is divided in to two by proper strided max pooling are shown.*

each time unit is T^{PTO}_{window} frames long. The frames coming from the thermal camera is processed in parallel with a second lightweight YOLO detector, which is especially trained for infrared drone images. Note that, as the architectures are lightweight two of them can be loaded to the GPU (in addition to one classifier). Then these detected regions of interest are classified by the YOLO classifier. The classifier which has an input size of 64x64 pixels gives a confidence score (a scalar between 0 and 1) for each designated category for each detected bounding box in a thermal image. Then for all frames in a T^{PTO}_{window} frames long window, the maximum drone score among all detected and classified bounding boxes is taken s^{PTO}_t. The motivation behind this approach is the fact that, due to rapid motion of the zoomed thermal camera and the object, the object may not be present in the image plane. Also, degrading effects due to blurry frames caused by motion is also compensated. Another advantage of this scheme is the chance of evaluating different poses of the same object, where the maximum score among them shall give a more accurate result.

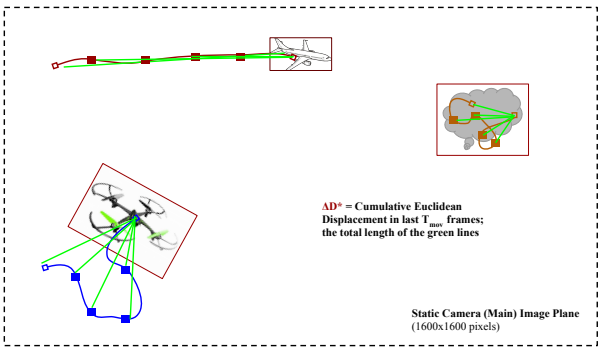


Figure 3. *Tracks are evaluated according to their movements in last T_{mov} frames. The total Euclidean distances between the last frame centroid and each of the last T_{mov} frames (total Euclidean displacement) is a measure of priority between existing tracks. This scheme favors the linear movement, while allowing the introduction of new tracks and checking for following tracks.*

The maximum scores of each time windows are averaged by a regular moving average filter as follows :

$$\sigma^{PTO}_t = \alpha\sigma^{PTO}_{t-1} + (1-\alpha)s^{PTO}_t \quad (1)$$

where α is a scalar determining the effect of the history. If the age of a PTO is larger than T^{PTO}_{min} frames and its σ^{PTO}_t is smaller than σ^{PTO}_{min}; the object is considered not to be a drone and it is unassigned as a PTO. Note that this object shall continue to be tracked on the main image plane, however it would not be assigned as PTO another time as it has been checked before. Then, if there are other candidate tracks, which has the highest maximum instantaneous euclidean distance based displacement is assigned as PTO, immediately.

Table 1. Parameters of the PTO based approach determined to be optimal based on our experiments.

Parameter	Value
T_{mov}	8 frames
T^{PTO}_{window}	5 frames
α	0.95
T^{PTO}_{min}	50 windows
σ^{PTO}_{min}	0.8

Classification and Artificial Thermal Image Dataset

As mentioned previously, classification is done by a separate YOLO architecture, which has an input shape of 64x64x1 pixels. The optimal configuration with respect to limited GPU memory budget is decided as in Figure 4. A novelty of our work is the utilization of RGB image datasets of drones, birds and commercial

aeroplanes for the CNN based detection and classification with infrared imagery. As it is not feasible to shoot thermal videos of all types of drones, birds etc. to generate a dataset and it is not possible to find open source datasets for thermal images, the idea to use RGB images efficiently is plausible. Figure 5.a shows the image of a drone taken by a thermal camera. As it can be observed, the thermal signature of a drone is quite uniform from mid-to-long distance. We have acquired 6000 images of various types of drones available in the market (also 6000 aeroplanes and birds), where their color is apparently darker than the sky background. Following this, the negatives of gray-scale of the images is taken, which mimics the thermal signature, where the background is highly dark and the object appears whitish. The main idea in this approach is that, as large number of variant samples are taken, even the exact values of thermal pixel intensities are deviant from the actual infrared images, the complex neural network can grasp the overall information required to classify objects.

Experimental Results

We have tested the performance of our system both in field operationally and with several videos. 6 shows the detections generated by our approach and several conventional detection algorithms, which produce high amount of false alarms due to background. In this document, we present the results for detection performance compared to cascaded HAAR detector and Mixture of

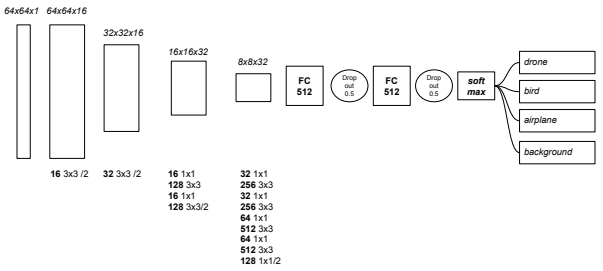


Figure 4. *YOLO classifier architecture for thermal image classification with 4 different categories.*

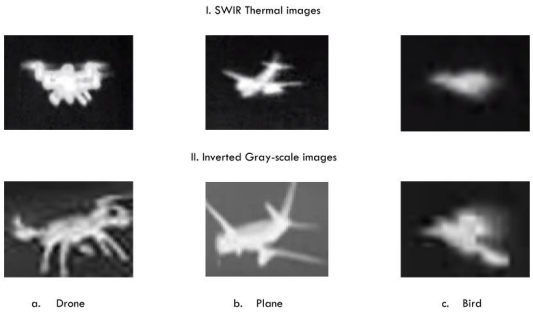


Figure 5. *The real thermal footages of drones, planes and birds (top) and inverted gray-scale images of drones, planes and birds where their color is darker than the sky background. With proper choice and image augmentation techniques, RGB datasets can be used to train deep learning algorithms to classify airborne targets.*

Table 2. Overall approximate true positive and false alarm rates of three different detectors, for different settings and environments

Method	True Positive	False Alarm
lightweight YOLO	0.91	0
cascaded Haar	0.95	0.42
GMM back. sub.	0.98	0.31

Gaussian background subtraction algorithm, for the case where the target drone is of size between 8x8 and 64x64 pixels. As it can be seen from 2, even though the hit rate of our YOLO based detector is slightly less, the false alarm issue is completely eliminated. Even though, detector fails to detect in certain frames, the Hungarian algorithm with distance as a parameter and Kalman filtering achieves the proper tracking of the target. By their intrinsic nature, cascaded detectors can only detect objects with predefined sizes, thus fail to produce precise bounding boxes in addition to their high false alarm rate. On the other hand, background subtraction algorithms completely fail in case of sudden illumination variations or jittering due to wind etc. In addition to these advantages, this approach only needs one hyper-parameter, the minimum confidence score, while the others require several ones, making optimal tuning harder.

Figure 7 shows the tracking of a successfully detected drone hovering on a marine environment, showing its preceding trail on previous frames. Note that any of the background objects

Figure 6. *True Positive and False Alarm Rates for small aerial object detection with a lightweight YOLO based detector and conventional methods. There exists 1 drone (green box at the bottom) and two birds (green boxes at top) in the scene. Detections by lightweight YOLO, cascaded HAAR and background subtraction are depicted with green, red and blue boxes, respectively. Note that, cascaded detector produces a lot of false alarms, whereas waves on sea causes false alarms for background subtraction. Footage taken from experiments by [25]*

cause false alarms despite containing complex shapes. This plausible outcome shall be rooted from the regression based nature of YOLO, which can capture intrinsic semantic information, even with employment of much fewer number of filters. Similar result can be observed in Figure 8.

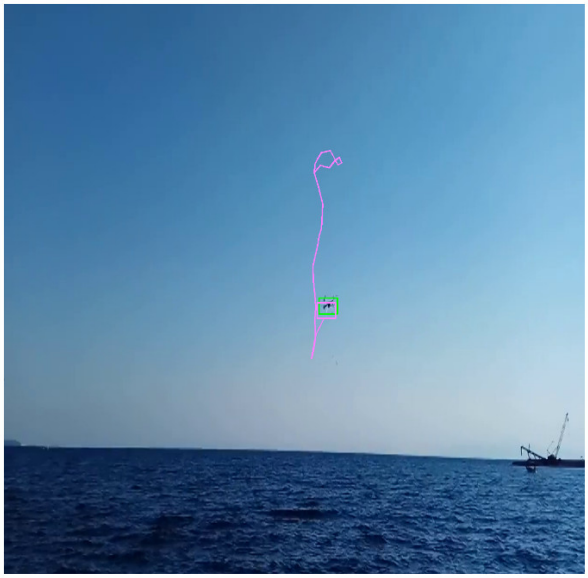


Figure 7. *The aerial object detection and tracking with Kalman filtering. Based on the movements of the target, the system is activated, turret is rotated towards it and the object is classified as drone or not with a zoomed daylight and thermal camera.*

Figure 8. *The aerial object detection and tracking with Kalman filtering. Based on the movements of the target, the system is activated, turret is rotated towards it and the object is classified as drone or not with a zoomed daylight and thermal camera.*

On the other hand, classification of thermal region of interests with four given object category has resulted with a true drone classification rate larger than 80% overall. Note that, as explained above our framework does not use categorical classification result directly, but process the drone confidence score. The reason behind the choice of four categories, rather than a drone versus others is due to the better performance. The aerial objects are highly similar, thus this approach gives a finer classification.

Conclusion

The outcomes of the work explained in this paper can be divided in to three as introduction of an integrated RGB and infrared camera based autonomous drone detection system incorporating an effective detection-tracking-identification policy, the study of lightweight YOLOv3 detectors' performance-memory trade-off and proposal of utilizing RGB images with image augmentation techniques for classifying thermal images. We have observed that with narrowing YOLO deep learning detector architectures, one can attain substantial performance, while keeping memory and computation time comparable to conventional methods, if classification and detection is separated. In the context of a sky background, even with high background clutter, YOLO's regression based approach allows detection of very small drones, with drastically lower false alarm rates.

References

[1] B. Jansen, "Drone crash at white house reveals security risks," *USA Today*, January 26, 2015.

[2] S. Gallagher, "German chancellors drone attack shows the threat of weaponized uavs," *Ars Technica*, 2013.

[3] A. Jouan, "Survols de centrales: un expert reconnu s inquiète," *http://www.lefigaro.fr/actualite-france/2014/11/25/01016-20141125ARTFIG00024-survols-de-centrales-un-expert-reconnu-s-inquiete.php*, 2014.

[4] J. Serna, "Lufthansa jet and drone nearly collide near lax," *LA Times*, March 19, 2016.

[5] S. Dinan, "Drones become latest tool drug cartels use to smuggle drugs into u.s." *https://www.washingtontimes.com/news/2017/aug/20/mexican-drug-cartels-using-drones-to-smuggle-heroi/*, August 20, 2017.

[6] P. Nguyen, M. Ravindranatha, A. Nguyen, R. Han, and T. Vu, "Investigating cost-effective rf-based detection of drones," in *Proceedings of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use*. ACM, 2016, pp. 17–22.

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

[8] A. Hommes, A. Shoykhetbrod, D. Noetel, S. Stanko, M. Laurenzis, S. Hengy, and F. Christnacher, "Detection of acoustic, electro-optical and radar signatures of small unmanned aerial vehicles," in *Target and Background Signatures II*, vol. 9997. International Society for Optics and Photonics, 2016, p. 999701.

[9] M. Hammer, M. Hebel, M. Laurenzis, and M. Arens, "Lidar-based detection and tracking of small uavs," in *Emerging*

Imaging and Sensing Technologies for Security and Defence III; and Unmanned Sensors, Systems, and Countermeasures, vol. 10799. International Society for Optics and Photonics, 2018, p. 107990S.

[10] T. Müller, "Robust drone detection for day/night counter-uav with static vis and swir cameras," in *Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VIII*, vol. 10190. International Society for Optics and Photonics, 2017, p. 1019018.

[11] S. R. Ganti and Y. Kim, "Implementation of detection and tracking mechanism for small uas," in *Unmanned Aircraft Systems (ICUAS), 2016 International Conference on*. IEEE, 2016, pp. 1254–1260.

[12] L. Hauzenberger and E. Holmberg Ohlsson, "Drone detection using audio analysis," 2015.

[13] S. Y. Nam and G. P. Joshi, "Unmanned aerial vehicle localization using distributed sensors," *International Journal of Distributed Sensor Networks*, vol. 13, no. 9, p. 1550147717732920, 2017.

[14] "How Droneshield works?" https://www.droneshield.com/how-droneshield-works/, 2018, accessed: 2018-10-22.

[15] "Introduction to Dedrones Airspace Security Platform," https://www.dedrone.com/webinars/introduction-to-dedrones-airspace-security-platform-11-28-2018, 11-28-2018, accessed: 2018-10-22.

[16] "Gryphon Skylight System. Detect, track and classify moving objects in your airspace." https://www.srcinc.com/pdf/Radars-and-Sensors-Gryphon-Skylight.pdf, 2018, accessed: 2018-10-22.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[18] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[19] A. Schumann, L. Sommer, J. Klatte, T. Schuchert, and J. Beyerer, "Deep cross-domain flying object classification for robust uav detection," in *Advanced Video and Signal Based Surveillance (AVSS), 2017 14th IEEE International Conference on*. IEEE, 2017, pp. 1–6.

[20] C. Aker and S. Kalkan, "Using deep networks for drone detection," *arXiv preprint arXiv:1706.05726*, 2017.

[21] M. Saqib, S. D. Khan, N. Sharma, and M. Blumenstein, "A study on detecting drones using deep convolutional neural networks," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2017, pp. 1–5.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[23] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[24] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," pp. 3645–3649, 2017.

[25] A. Coluccia, M. Ghenescu, T. Piatrik, G. De Cubber, A. Schumann, L. Sommer, J. Klatte, T. Schuchert, J. Beyerer, M. Farhadi *et al.*, "Drone-vs-bird detection challenge at ieee avss2017," in *Advanced Video and Signal Based Surveillance (AVSS), 2017 14th IEEE International Conference on*. IEEE, 2017, pp. 1–6.

Author Biography

Dr. Eren Unlu has acquired his B.Sc., M.Sc. and Ph.D. from Bilkent University, Institute EURECOM and CentraleSupelec with respectively, all on electrical engineering. He is currently working as a post-doctoral associate in ISAE-SUPAERO, France.

Dr. Emmanuel Zenou is graduate from ENS Paris Saclay and did his PhD in SUPAERO, Toulouse, France. He is Associate Professor at SUPAERO in Computer Vision and Data Analysis. After working in the field of Robotics, he works now mainly in the fields of Earth Observation and Space objects (asteroids, space debris)

Dr Nicolas Riviere works in the field of light scattering, laser imagery and LiDAR techniques. He is a research engineer at ONERA, a leading research center in French and European aerospace applications. Dr Riviere has extensive experience in leading scientific research projects, including at national and European levels.

Paul-Edouard Dupouy received his MSc in applied physics from the National Institute for Applied Sciences in 2009. He is currently at ONERA, the french aerospace lab, researching active imaging systems. He has experience in LiDARS, optical systems simulation and spectroscopy.

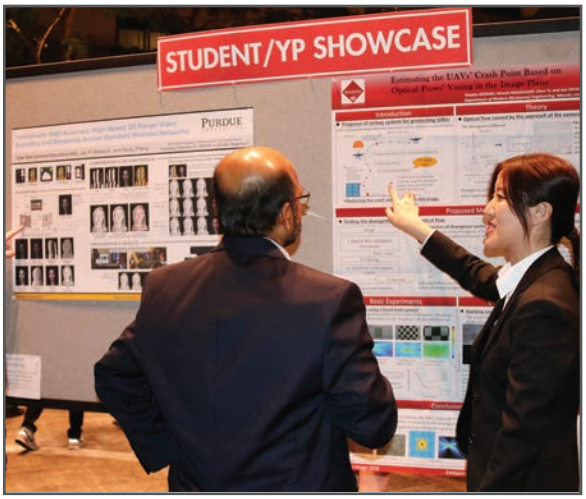
JOIN US AT THE NEXT EI!

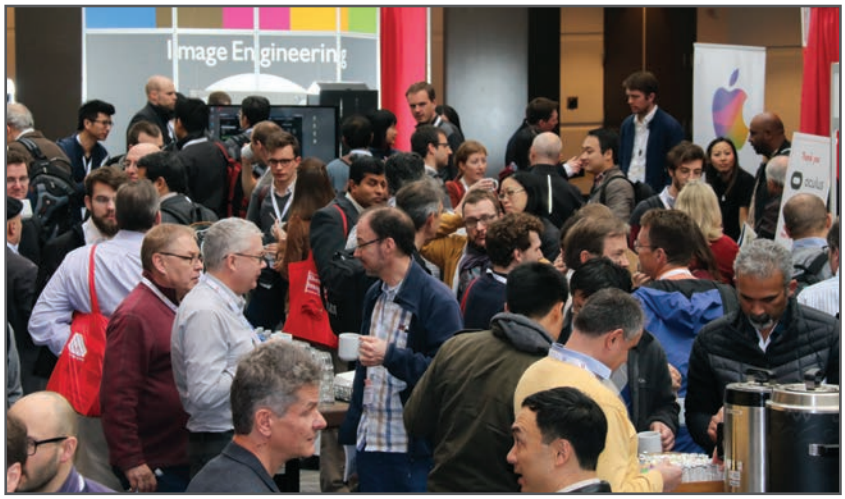
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org