A Tethered Unmanned Aircraft System for crowd dynamics and Unusual Pattern Extraction

Athanasios Douklias *
Institute of Communication & Computer Systems
Athens, Greece
0000-0002-5693-1969

Emmanuel Vasilopoulos
Institute of Communication & Computer Systems
Athens, Greece
0000-0002-0208-0337

Evangelos Maltezos
Institute of Communication & Computer Systems
Athens, Greece
0000-0002-3761-3631

Lazaros Karagiannidis
Institute of Communication & Computer Systems
Athens, Greece
0000-0002-9148-6258

Eleftherios Ouzounoglou
Institute of Communication & Computer Systems
Athens, Greece
0000-0002-5078-3248

Vasilis Nousis
Institute of Communication & Computer Systems
Athens, Greece
0000-0002-0116-8425

Angelos Amditis
Institute of Communication & Computer Systems
Athens, Greece
0000-0002-4089-1990

Abstract—Unusual pattern extraction and crowd dynamics and analysis has been receiving a constantly growing attention in the recent years contributing to safety, crisis management and situational awareness applications. Traditional video-based systems in this field that are exclusively monitored by human agents, are error-prone and present limited human performance. In this context, the design and implementation of an integrated tethered Unmanned Aircraft System (tUAS), namely INtelligentUxV Surveillance-Unusual Pattern Extraction (INUS-UPE) platform for crowd dynamics and unusual pattern extraction is presented. The INUS-UPE platform adopts a combination of key features and state-of-the-art technologies such real-time processing in the edge, video-based pattern analysis for object detection, multiple object tracking, and unusual pattern extraction, web-based user interface (UI) for visualization and possibility of information sharing related the extracted events to external systems such common operational pictures (COPs), Command & Control (C2) platforms, and Human Machine Interfaces (HMIs).

Keywords—unusual pattern, crowd dynamics, object detection, object tracking, deep learning, tethered drone, UAS

I. INTRODUCTION

Unusual pattern extraction and crowd dynamics and analysis has been receiving a constantly growing attention in the recent years. Many studies have been conducted to understand crowd patterns within a crowded scene [1] [2]. Besides the theoretical interest, a realistic modeling, understanding and monitoring of crowd pattern may lead to several societal benefits, e.g., improved design of buildings, aircraft, and ships or cost of transportation, pollution, and improving the population's quality of life, etc [3]. Furthermore, such approach can contribute to safety and security to smart cities [4] [5] or in surveillance zones to identify threats that compromise the safety of a crowd and infrastructures in several applications such civil protection, natural disasters, emergency situations and evacuation, optimized crisis management of a crowd during gathering events, minimization of risk particularly in highly crowded incidents, situational awareness, etc [3] [6]. There are various interpretations of situational awareness; nevertheless, the most prevalent is the one from [7], which categorizes it to the following main steps: (i) perception of the current situation, (ii) comprehension of said situation, and (iii) projection of the future condition.

Recent works seized the importance of exploiting crowd pattern analysis via video surveillance and human-computer interaction [8] [9]. In this context more and more cities and security agencies are setting up surveillance systems based on video-protection cameras [10]. Traditional video surveillance systems for situational awareness, that are exclusively monitored by human agents, are error-prone and present limited human performance by fatigue and waning concentration when continuously watching a video feed [1]. To this end, technological advancements in situational awareness apply several state-of-the-art technologies, such as machine/deep learning frameworks and advanced image processing and computer vision methods. In recent years, there are many approaches that have been proposed that provide object detection and tracking via cameras sensors, used in several applications such as self-driving vehicles, unusual pattern extraction during surveillance, search and rescue, etc. With the recent upsurge in the use of deep learning frameworks in computer vision there are several efficient object detection and object tracking algorithms that have been developed such as convolutional neural networks (CNNs) [11] and multiple object trackers (MOT) [12] respectively. Although the complexity of such architectures and high computational costs, they provide several advantages, compared to traditional methods, in terms of learning ability, high variability, and generalization [9].

Furthermore, recent trends in situational awareness and in the development of intelligent applications in Internet of Things (IoT) adopt modern technologies such edge computing that appears to decrease latency and computational processing as well as Unmanned Aircraft Systems-UAS (free flight or tethered) [13] [14] [15]. UAS have high mobility, are easily deployed, and capable of real-time monitoring by utilizing multi-sensor-based detection and remote sensing of objects [16]. Also, are considered as flexible solution, enabling to fly to places that are inconvenient or inaccessible to humans for data collection, having also the capability to acquiring fast and low costs, high-resolution images and videos. On the other side, tethered drones are able to provide long-term missions compared to free flight drone whose their total flight time is

under 1 hour [17]. A tethered UAS (tUAS) usually consists of one of two scenarios: (i) a drone (or else well known as Unmanned Aircraft Vehicle-UAV) that is kept powered by a ground station or terrestrial platform (that may be an Unmanned Ground Vehicle-UGV) by use of a cable that can transmit energy, data, and provide mechanical support between the air and ground operations and (ii) one or more free-flying drone attached to a load through a tether that serve only as a mechanical connection between the vehicles and the load [17].

II. Related Work

There are two main approaches for crowd dynamics depending on how objects are considered in relation to the crowd they belong to. Such approaches are the microscopic (or bottom-up) and the macroscopic (or top-down) [18]. In microscopic approaches the crowd is treated as a collection of objects. Objects' dynamics are studied individually, and afterwards the knowledge about these objects is used to infer information at crowd level. In macroscopic approaches the crowd is treated as a whole single entity. Independently of whether a microscopic or macroscopic approach is followed, the pipeline for crowd pattern analysis through a video streaming feed can be divided in the following four stages [12] [18]: (i) object detection stage, (ii) object tracking stage, (iii) feature extraction stage (to compute a set of features and metrics that describe the dynamics, topological structure and affective state of the crowd), and (iv) identification of crowd unusual pattern stage (to recognize particular patterns and/or unusual events in video sequences).

In the bibliography several terrestrial systems have been proposed for identifying unusual patterns in dense crowds. In [19] a new method for the detection and classification of unusual patterns in dense crowds has been proposed. A two stream network was proposed that uses crowd density heatmaps and optical flow information to detect panic and fight events. Additionally, in [20] a CNN-based approach was presented that efficiently detects and classifies if a video involves crowd unusual patterns. Their approach implemented a two-stream architecture using two separate 3D CNNs to accept a video and an optical flow stream as input to enhance the prediction performance.

Related to UAS, in [21] the authors provided a comprehensive review from a crowd monitoring and analysis perspective concluding the significance of the use of drones in public safety and security applications. In [6] a new testing procedure was introduced to detect the crowd features from UAS taken from different camera orientations and positions. The authors of [13] proposed a light-weight scheme of a fully-CNN for crowd detection in aerial images using spatial graphs and combining nimble computations and effectiveness. In [16] the authors discussed several component's information related to a real-time drone surveillance system for violent crowd patterns from UAS. The authors of [22] proposed a crowd flow detection method for video sequences shot by a UAS. Their method was based on a fully-CNN that learns to perform crowd clustering in order to detect the centroids of crowd-dense areas and track their movement in consecutive frames.

III. Our Approach

In this paper, we aim to contribute to the aforementioned growing research with the design and implementation of an integrated tUAS, namely INtelligentUxV Surveillance-Unusual Pattern Extraction (INUS-UPE) platform, that focuses on crowd dynamics and unusual pattern extraction for enhanced situational awareness applications. The INUS-UPE platform adopts a combination of key features and state-of-the-art technologies such: (i) a modular and flexible architecture with the potential to exploit several types of drones and camera sensors, (ii) real-time processing in the edge, (iii) a backend following the microservices architecture design paradigm, (iv) a video server from multiple video streaming, (v) a pattern analysis module that performs object detection, multiple object tracking via tracking-by-detection (TBD), and unusual pattern extraction, (vi) a geolocalization module, (vii) a user interface (UI) for parameter setting and drawing line/polygons for unusual pattern cases, for the display of message events and the geolocation of the detected/tracked objects as well as for the visualization of the raw and processed video, (viii) information sharing related the extracted events as well as video streaming to external systems, common operational pictures (COPs), Command & Control (C2) platforms, Human Machine Interfaces (HMIs), or other middleware brokers (e.g., Apache Kafka [23], etc.), (ix) adoption of a custom data model or other data model for enhanced situational awareness (e.g., e-CISE model). In figure 1, the INUS-UPE platform's architecture is depicted. The INUS-UPE platform consist of several hardware and software components described in detail in the next sections.

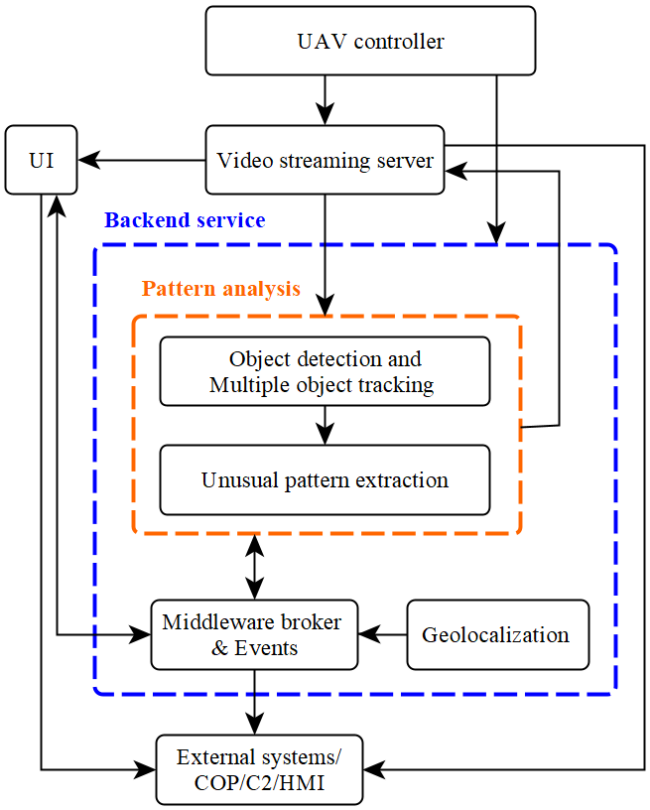

Fig. 1. INUS-UPE platform architecture

The INUS-UPE platform is controlled by authorized operators with related authorized access such: (i) a certified drone pilot who is responsible to operate the drone through a related remote controller, and (ii) the intelligence officer who operates a related workstation that performs in real-time the pattern analysis process in the edge and incorporates the UI. It should be noted that the related EASA's regulations are considered.

The backend of the system is divided into multiple loosely coupled modules that handle data acquisition, processing, presentation and integration with other systems. Each of these modules is developed in a separate dockerized service that communicates with the rest of the system via an MQTT broker or a WebRTC video server depending on the data exchanged (json/video). The aforementioned microservices are managed by a Kubernetes open source container orchestrator that oversees deployment and configuration.

IV. HARDWARE

The INUS-UPE platform consists of several hardware components as follows:

A. Drone

The DJI Mavic 3T EU Enterprise drone (figure 2-left) (Max Takeoff Weight: 1,050 g and dimensions when is unfolded without propellers: 347.5 mm × 283 mm ×107.7 mm) with GNNS (Galileo, GPS, BeiDou, GLONASS) and the related RTK module is utilized equipped both with RGB and thermal camera.

B. Tethered station

The used compatible tethered station is the V-line Pro (dimensions: 470 mm × 365 mm ×190 mm and weight: 17kg) (figure 2-right). Also, a portable power bank with capacity of 2016 Wh is utilized.

C. Intelligence officer's workstation

Is a computing device that performs in real-time the pattern analysis process at the edge. The intelligence officer's workstation is operated by the intelligent officer and it is equipped with a GPU that is capable to execute in real-time deep learning tasks, and to provide video streams. Also, incorporates the UI though which relevant information (raw and processed video streams, events, etc) are displayed.

All the aforementioned equipment is able to provide flexibility as it can be easily transferred and installed. Also, is able to provide either a continuous flight operation by constant charging (i.e., directly from the tethered station) or a long-tine flight (via the power bank) in inaccessible areas or areas with unavailable power supply.

Fig. 2. Utilized drone and tethered station

D. Drone controller

The DJI Mavic 3T Enterprise drone is controlled by the DJI RC Pro remote controller. It can achieve a communication range of up to 15 km with a 2T4R antenna system and is capable of sending a 1080p/60fps live feed with latency as low as 120 ms. More importantly is that has a mini-HDMI video output and supports the DJI Cloud API. The later enables integration with remote cloud services and retrieval of information regarding the drone's location, camera attitude, battery remaining, etc., which is utilized by the INUS-UPE to provide information to the system operator for the drone and aids in the associated localization.

V. SOFTWARE

The INUS-UPE platform consists of the following software components:

A. Video streaming server

The system utilizes an open source WebRTC based video server, Janus. The inclusion of a video server provides for an endpoint where internal components of the system and external systems can source video streams in a one to many distribution fashion. Video delivery through WebRTC leads to reduced latency in comparison to traditional web video distribution protocols such as DASH and HLS, owning to the peer to peer connections used to deliver the video content.

B. Paterns analysis module

1) Object detection: In the bibliography there are several approaches for object detection such the ones in [24]. Considering the object detection frameworks, there are various robust and efficient published deep learning networks, such YOLO in several versions [25] [26] and Single Shot MultiBox Detector (SSD) [27]. The INUS-UPE platform exploits the pre-trained version of YOLOv5l which is responsible for the object detection process.

2) Multiple object tracking via TBD: Recent developed and efficient TBD algorithms are the FairMOT [28], DeepSort [29], and ByteTrack [30]. It should be noted that, in TBD process, the performance of the detector and the data association algorithm affect the tracking accuracy and robustness. In any case, both object detection and object tracking algorithms are mainly influenced by the used cameras sensor (e.g., spectrum capabilities, pixel resolution, etc.), viewing perspective of the objects, and complexity of the scene, such as occlusions, shadows, light conditions, and scene background. The INUS-UPE platform exploits the ByteTrack tracker for the TBD process. It should be noted that the ByteTrack tracker does not implement a re-identification process, thus, the id tracking number of a tracked object will be remaining the same till the tracker failures. If the tracker fails (e.g., due to algorithm, occlusion, color appearance and illumination, background environment, etc.), a new id tracking number is assigned to the new detected object.

3) Unusual pattern extraction: In [18], some features referred that are crusial to be calculated for crowd pattern analysis/dynamics such: (i) direction (number of main directions of movement followed by the crowd), (ii) density (proximity of the objects in the crowd, determining how dense the crowd is), (iii) collectiveness or cohesion (degree of objects acting as a union in collective motions), (ix) valence (measures the positive and negative affect of the crowd), (x) arousal (monitors how calmed or excited the crowd is), and (xi) velocity (average speed at which objects or crowds are

moving). On the other hand, the nature of the unusual may be diverse and is not applies to all cases, thus the approach to consider this may differs according to the application and use case. In [18] the authors identified some types of unusual patterns such: (i) unusual appearance (occurs when a non-recognised object enters the scene), (ii) unusual action (understanding of the usual patterns of the objects in the scene, and the extraction of non-common ones) (iii) unusual position (unusual position of an object in the scene), and (iv) unusual movement (unexpected trajectory or velocity of an object or group in the scene).

It should be noted that due to the INUS-UPE platforms' modular architecture, the adoption of other existing or other new object detection or tracking algorithms is capable. Considering the aforementioned, the INUS-UPE platform adopts a feature extraction and a rule-based process for unusual pattern extraction for two cases under a situational awareness aspect. The first unusual pattern extraction case is an object that is in/out-of-bound i.e., from a defined line or polygon (drawn by the intelligence officer through the UI). Such case is under the type of unusual position pattern e.g., when a non-authorized object (s) enters a restricted/prohibited area, or a is detected on a dangerous zone. For the polygon paradigm, a 2D-array is initialized once with the same dimensions as a video frame, with values of 0. The polygon is used to update all the values of the array with any value other than 0, as well as the polygon's area. Finally, when an object(s) is detected with image coordinates x,y, the array is queried given the coordinates and if the value is not equal to 0, the object is in-bound. If the value is 0, the object is out-of-bound. Depending to the application, the related unusual pattern event is extracted when an object (s) is in/out-of-bound.

The second unusual pattern extraction case is an object which velocity is above a predefined threshold (defined by the intelligence officer through the UI). Such case is under the type of unusual movement or action e.g., when an object (or objects) move faster or slower; and direction, when predominant flows exist and the movement of an object deviates from these trends. Here, the object's velocity is calculated as pixels per frame. Considering a constant altitude and fixed position of the drone, a constant video resolution and a constant view angle, an unusual magnitude of velocity threshold can be derived after examining multiple video footage with an object that increases and decreases its velocity. Once the object's velocity surpasses that threshold, its movement is identified as unusual, and thus an unusual pattern event is extracted.

An example of the two considered use cases are depicted in fig. 3 and 4 using the free available datasets of OKUTAMA [31] and VisDrone [32].

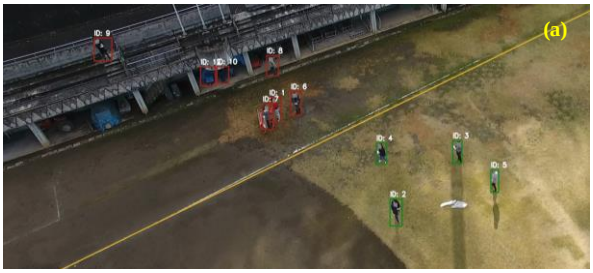

Fig. 3. Unusual pattern case #1: Objects that are out-of-bound (OKUTAMA dataset). Where: (i) the detected and tracked objects are with colored bounding boxes superimposed to the video frames a to e, (ii) the yellow line is the bound, and (iii) the green bounding boxes indicate that the objects are in-bound, while the red ones indicate that the objects are out-of-bound (and thus an unusual pattern event is extracted).

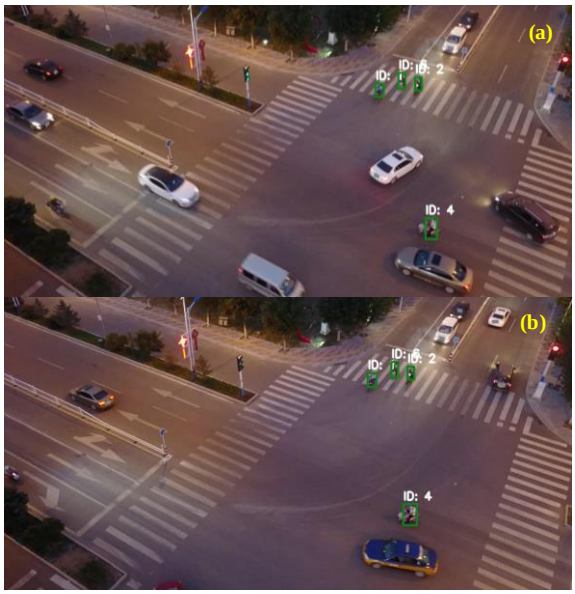

Fig. 4. Unusual pattern case #2: Objects which velocity is above a threshold (VisDrone dataset/ video frames a to c) and (OKUTAMA dataset /video frames d-f) . Where: (i) the detected and tracked objects are with colored bounding boxes superimposed to the video frames, and (ii) the green bounding boxes indicate that the object's velocity is not above the threshold while the red ones indicate that object's velocity is above the threshold (and thus an unusual pattern event is extracted).

C. Geolocalization module

All the detected objects are localized in an absolute reference frame (i.e., latitude and longitude in the WGS'84 system) by the geolocalization module. The module receives input from the MQTT broker regarding the position and orientation of the drone's camera and combines this information with the pixel coordinates of the centroid of each detected object, which are translated to angle offset from the center of the camera [33], and finally to estimate the related geolocation. The geolocalization information is embedded in the event messages generated by the system.

D. Event extraction

The INUS-UPE platform extracts three types of events, that is, object detection event, object tracking event, and unusual pattern extraction event. The supplementary information included to the related data model is the following: (i) image coordinates related the bounding box of the detected object (s), (ii) object detection and tracking confidence level (in percentage), (iii) id's of the tracked objects, (iv) geolocations of the detected/tracked objects and of the drone, (v) objects's velocity, and (vi) type of the unusual pattern extraction event.

VI. User Interface

In figure 5, a sample of the UI of the INUS-UPE platform is depicted. The UI provides in real-time the following capabilities to the intelligence officer: (i) activation of up to two assets (i.e., two drones, two (potential) terrestrial camera systems, or one drone and one terrestrial camera system) from which is able to visualize the related videos (and from a potential list of several available video streamings from different camera sensors) and to display telemetry drone data (height, battery, and communication signal power) or other data such sensors that potentially installed to the drone, (ii) drawing tool to draw the line or polygon directly to the visualized video as well as to provide parameter setting for the object in/out-of-bound and object with velocity above a threshold unusual pattern cases respectively, (iii) selection to visualize either only the raw video or the extracted information as superimposed layers related to the object detection (i.e., the bounding boxes), object tracking (i.e., object id's or potentially their tracking trajectories), the calculated velocity of the tracked objects, and drawn polygons or lines, (iv) map in which the geolocations of the detected/tracked objects and the geolocation of the drone are displayed, and (v) display information related the extracted events coming from the selected assets. It should be noted that the UI is web-based, thus, it is able to be shared to other external systems, C2 platforms, COPs, and HMIs.

Fig. 5. Sample of the UI of the INUS-UPE platform

VII. Conclusions

Recent technological advancements in safety, crisis management, and situational awareness applications adopt machine/deep learning frameworks, advanced image processing and computer vision, and crowd dynamics methods. In this context, this paper introduces an integrated tUAS, namely INUS-UPE platform, for video-based crowd dynamics and unusual pattern extraction that exploits several key features and state-of-the-art technologies. The related architecture and hardware and software components were described as well as representative experiments from free available datasets were performed. Due to the modular design and flexibility of the INUS-UPE platform, several extensions and enhancements can be made in future work for adaption to several applications and to adopt more types of unusual pattern cases or to include multiple objects of interest and types of them. Also, additional experiments will be performed in future work for several: (i) types of scenes via extended real-time flights, (ii) light conditions (e.g., night, day), (iii) types of camera sensors (e.g. RGB, thermal), and (iv) viewing perspectives.

Acknowledgment

This work is a part of the Flexi-cross project. This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101073879. Content reflects only the authors' view and the Research Executive Agency (REA)/European

Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar, and P.-A. Muller, "Recent trends in crowd analysis: A review," Machine Learning with Applications, vol. 4, p. 100023, Jun. 2021, doi: 10.1016/j.mlwa.2021.100023.

[2] E. Şengönül, R. Samet, Q. Abu Al-Haija, A. Alqahtani, B. Alturki, and A. A. Alsulami, "An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey," Applied Sciences, vol. 13, no. 8, p. 4956, Apr. 2023, doi: 10.3390/app13084956.

[3] N. Bellomo and L. Gibelli, Eds., Crowd Dynamics, Volume 3: Modeling and Social Applications in the Time of COVID-19. in Modeling and Simulation in Science, Engineering and Technology. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-91646-6.

[4] E. Maltezos et al., "Public safety in smart cities under the edge computing concept," in 2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), 2021, pp. 88–93. doi: 10.1109/MeditCom49071.2021.9647550.

[5] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22–32, Feb. 2014, doi: 10.1109/JIOT.2014.2306328.

[6] A. Almagbile, "Estimation of crowd density from UAVs images based on corner detection procedures and clustering analysis," Geo-spatial Information Science, vol. 22, no. 1, pp. 23–34, Jan. 2019, doi: 10.1080/10095020.2018.1539553.

[7] R. Geraldes et al., "UAV-Based Situational Awareness System Using Deep Learning," IEEE Access, vol. 7, pp. 122583–122594, 2019, doi: 10.1109/ACCESS.2019.2938249.

[8] [8] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan, "Crowded Scene Analysis: A Survey," IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 3, pp. 367–386, Mar. 2015, doi: 10.1109/TCSVT.2014.2358029.

[9] D. R. Patrikar and M. R. Parate, "Anomaly Detection using Edge Computing in Video Surveillance System: Review," arXiv:2107.02778 [cs], Aug. 2021, Accessed: Jan. 13, 2022. [Online]. Available: http://arxiv.org/abs/2107.02778

[10] F. Porikli et al., "Video surveillance: past, present, and now the future [DSP Forum]," IEEE Signal Process. Mag., vol. 30, no. 3, pp. 190–198, May 2013, doi: 10.1109/MSP.2013.2241312.

[11] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," Computational Intelligence and Neuroscience, vol. 2018, pp. 1–13, 2018, doi: 10.1155/2018/7068349.

[12] B. Tyagi, S. Nigam, and R. Singh, "A Review of Deep Learning Techniques for Crowd Behavior Analysis," Arch Computat Methods Eng, vol. 29, no. 7, pp. 5427–5455, Nov. 2022, doi: 10.1007/s11831-022-09772-1.

[13] G. Castellano, C. Castiello, C. Mencar, and G. Vessio, "Crowd Detection in Aerial Images Using Spatial Graphs and Fully-Convolutional Neural Networks," IEEE Access, vol. 8, pp. 64534–64544, 2020, doi: 10.1109/ACCESS.2020.2984768.

[14] A. Ramachandran and A. K. Sangaiah, "A review on object detection in unmanned aerial vehicle surveillance," International Journal of Cognitive Computing in Engineering, vol. 2, pp. 215–228, Jun. 2021, doi: 10.1016/j.ijcce.2021.11.005.

[15] R. Galliera and N. Suri, "Object Detection at the Edge: Off-the-shelf Deep Learning Capable Devices and Accelerators," Procedia Computer Science, vol. 205, pp. 239–248, 2022, doi: 10.1016/j.procs.2022.09.025.

[16] T. Simpson, "Real-Time Drone Surveillance System for Violent Crowd Behavior Unmanned Aircraft System (UAS) – Human Autonomy Teaming (HAT)," in 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA: IEEE, Oct. 2021, pp. 1–9. doi: 10.1109/DASC52595.2021.9594332.

[17] M. N. Marques, S. A. Magalhães, F. N. Dos Santos, and H. S. Mendonça, "Tethered Unmanned Aerial Vehicles—A Systematic Review," Robotics, vol. 12, no. 4, p. 117, Aug. 2023, doi: 10.3390/robotics12040117.

[18] F. Luque Sánchez, I. Hupont, S. Tabik, and F. Herrera, "Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects," Information Fusion, vol. 64, pp. 318–335, Dec. 2020, doi: 10.1016/j.inffus.2020.07.008.

[19] L. Lazaridis, A. Dimou, and P. Daras, "Abnormal Behavior Detection in Crowded Scenes Using Density Heatmaps and Optical Flow," in 2018 26th European Signal Processing Conference (EUSIPCO), Rome: IEEE, Sep. 2018, pp. 2060–2064. doi: 10.23919/EUSIPCO.2018.8553620.

[20] A. Mehmood, "Abnormal Behavior Detection in Uncrowded Videos with Two-Stream 3D Convolutional Neural Networks," Applied Sciences, vol. 11, no. 8, p. 3523, Apr. 2021, doi: 10.3390/app11083523.

[21] M. A. Husman et al., "Unmanned Aerial Vehicles for Crowd Monitoring and Analysis," Electronics, vol. 10, no. 23, p. 2974, Nov. 2021, doi: 10.3390/electronics10232974.

[22] G. Castellano, E. Cotardo, C. Mencar, and G. Vessio, "Density-based clustering with fully-convolutional networks for crowd flow detection from drones," Neurocomputing, vol. 526, pp. 169–179, Mar. 2023, doi: 10.1016/j.neucom.2023.01.059.

[23] "Apache Kafka," Apache Kafka. Accessed: Jan. 26, 2022. [Online]. Available: https://kafka.apache.org/

[24] C. Kim, M. Oghaz, J. Fajtl, V. Argyriou, and P. Remagnino, "A Comparison of Embedded Deep Learning Methods for Person Detection," presented at the International Conference on Computer Vision Theory and Applications, SCITEPRESS, Feb. 2019, pp. 459–465. doi: 10.5220/0007386304590465.

[25] "ultralytics/yolov5." Ultralytics, Jan. 13, 2022. Accessed: Jan. 13, 2022. [Online]. Available: https://github.com/ultralytics/yolov5

[26] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, doi: 10.48550/ARXIV.2207.02696.

[27] J. Martin et al., "Embedded Vision Intelligence for the Safety of Smart Cities," J. Imaging, vol. 8, no. 12, p. 326, Dec. 2022, doi: 10.3390/jimaging8120326.

[28] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, "FairMOT: On the Fairness of Detection and Re-identification in Multiple Object Tracking," Int J Comput Vis, vol. 129, no. 11, pp. 3069–3087, Nov. 2021, doi: 10.1007/s11263-021-01513-4.

[29] B. A. Griffin and J. J. Corso, "BubbleNets: Learning to Select the Guidance Frame in Video Object Segmentation by Deep Sorting Frames," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019, pp. 8906–8915. doi: 10.1109/CVPR.2019.00912.

[30] Y. Zhang et al., "ByteTrack: Multi-object Tracking by Associating Every Detection Box," in Computer Vision – ECCV 2022, vol. 13682, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds., in Lecture Notes in Computer Science, vol. 13682. , Cham: Springer Nature Switzerland, 2022, pp. 1–21. doi: 10.1007/978-3-031-20047-2_1.

[31] M. Barekatain et al., "Okutama-Action: An Aerial View Video Dataset for Concurrent Human Action Detection," in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jul. 2017, pp. 2153–2160. doi: 10.1109/CVPRW.2017.267.

[32] P. Zhu et al., "Detection and Tracking Meet Drones Challenge," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 11, pp. 7380–7399, Nov. 2022, doi: 10.1109/TPAMI.2021.3119563.

[33] A. Douklias, L. Karagiannidis, F. Misichroni, and A. Amditis, "Design and Implementation of a UAV-Based Airborne Computing Platform for Computer Vision and Machine Learning Applications," Sensors, vol. 22, no. 5, p. 2049, Mar. 2022, doi: 10.3390/s22052049.