Communications for Autonomous Unmanned Aerial Vehicle Fleets in Outdoor Cinematography Applications

Ioannis Mademlis†, Paraskevi Nousi†, Cedric Le Barz⋆, Tiago Gonçalves⋆, Ioannis Pitas†

†Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
⋆Thales - Advanced studies department THERESIS, Palaiseau, France

***Abstract*—Camera-equipped UAVs (Unmanned Aerial Vehicles), or "drones", are a recent addition to standard audiovisual (A/V) shooting technologies. As drone cinematography is expected to further revolutionize media production, especially by employing a fleet of cooperating UAVs, this paper presents an overview of the relevant communication and data streaming challenges. Emphasis is given on partially autonomous UAV swarms for live filming of outdoor events. A proposed, specially designed multiple-UAV platform for live outdoor media production is then presented, along with its communication and data streaming modules. It includes a set of possible solutions to the aforementioned issues, employing off-the-shelf tools wherever possible. Additionally, the reasoning behind the choices made is explained, in the context of the proposed platform.**

***Index Terms*—UAV communications, data streaming, UAV fleet, media production, autonomous drones**

I. Introduction

Unmanned Aerial Vehicles (UAVs, or "drones") are a recent addition to the cinematographer's arsenal. By exploiting their ability to fly and/or hover, their small size and their agility, impressive video footage can be obtained that otherwise would have been impossible to acquire. Although UAV cinematography is expected to revolutionize A/V shooting, as Steadicam did back in the seventies [1], the topic has not yet been heavily researched and shooting is currently performed on a more or less ad-hoc basis. Employing drones in video production and broadcasting opens up numerous opportunities for new forms of content, enhanced viewer engagement and interactivity. It immensely facilitates flexibility in shot set up, while it provides the potential to adapt the shooting so as to cope with changing circumstances in wide area events. Additionally, the formation of dynamic panoramas or novel, multiview and 360-degree shots becomes easier.

Single-UAV shooting with a manually controlled drone is the norm in media production today, with a director/cinematographer, a pilot and a cameraman typically required for professional filming. However, in such a setting, each target may only be captured from a specific viewpoint/angle and with a specific framing shot type at any given time instance, limiting the cinematographer's artistic palette. Moreover, there can only be a single target at each time, restricting the scene coverage and resulting in a more static, less immersive visual result. Finally, the "dead" time intervals required for the UAV to travel from one point to another, in order to shoot from a different angle, aim at a different target, or return to the recharging platform, impede smooth and unobstructed filming.

Swarms/fleets of multiple UAVs, composed of many cooperating drones, are a viable option for overcoming the above limitations, by eliminating dead time intervals and maximizing scene coverage, since the participating drones may simultaneously view overlapping portions of space from different positions. Due to the possibly large number of fleet members, a degree of decisional and functional autonomy would significantly ease their control, by lightening the burden on human operators. Thus, cognitive UAV autonomy using artificial intelligence and robotics technologies would greatly enhance the appeal of UAV swarms in media production.

Several challenges arise at the current level of technology, especially in the case of employing a partially autonomous UAV swarm. These include battery, safety, sound, synchronization, privacy and legal issues. Communication and data streaming challenges are especially prominent among them. Following early preliminary work in the context of the EU-funded research and development project MULTIDRONE[1] [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], this paper overviews the challenges regarding networking and data streaming issues, while it also proposes a corresponding set of solutions based on employing off-the-shelf tools. Although MULTIDRONE concerns partially autonomous filming of sports events in extended outdoor environments (e.g., bicycle races), the determined challenges and solutions apply equally to a vast range of different applications.

II. General UAV Communication Challenges in Media Production

Infrastructure for communications and related issues is critical for successful deployment of UAV swarms in practical scenarios, especially in live event media coverage applications. Even in single-UAV missions it is challenging to stream high-resolution video (especially 4K UHD, i.e., the norm in

[1] https://multidrone.eu/

media production) down to a ground station with Quality-of-Service (QoS) guarantees, while simultaneously executing all of the previously described algorithms in real-time. Video acquisition, compression, synchronization and transmission are procedures easily implemented using professional cameras and open-source software, although the lack of media production-quality camera models with Camera Serial Interface (CSI) connectivity (allowing rapid and stable capture for reliable online processing) is an existing practical issue. However, they jointly consume significant processing power and energy, on a computing platform already strained in these resources. The issue cannot simply be solved by dedicated hardware, since the latter would come with additional energy consumption, monetary and weight overhead. Therefore, at the current stage of technology, a trade-off has to be made between the broadcast video resolution, the hardware cost and the level of vehicle cognitive autonomy.

In simpler, non-live coverage, i.e., when filming for deferred broadcast, or shooting a scripted sequence, on-the-fly video transmission is not required (video may simply be stored on-board and retrieved later). In fact, if all processing is performed on-board in a completely autonomous manner, there is not even need for networking. However, communications are required in all other cases, including the non-live single-UAV filming where a subset of the less critical algorithms previously described, e.g., crowd/landing site detection and high-level path planning, are executed on a computationally powerful ground station, at the cost of significant latency (at best, about one hundred milliseconds). In general, a private QoS-guaranteeing 4G/LTE infrastructure suffices for the task, given the high mobility of the UAVs and the possibly long distances that need to be covered in outdoor event filming. Traditional Wi-Fi is a less costly, suboptimal alternative with higher latency and significantly smaller range, while public LTE networks are not reliable due to the lack of a way to prioritize UAV communications over telephony. The main challenge lies in live broadcasting; even private LTE will not allow consistent 4K UHD video streaming, unavoidably leading to a fall back on FullHD resolution.

If a swarm of multiple cooperating UAVs is employed, additional issues arise. Most importantly, in live coverage, the available bandwidth may not be enough to support live FullHD video streaming from all drones concurrently, resulting in a hard upper limit on the number of drones (a simple linear relation exists between the required total bandwidth and the number of employed UAVs). Furthermore, if direct coordination between the drones themselves is required (so as to autonomously capture a multiple-UAV shot, to execute distributed variants of algorithms such as SLAM, or simply for redundancy/fault tolerance), then an intra-swarm Flying Ad Hoc Network (FANET) should be employed, supporting ad hoc routing and accounting for high node mobility, long distances and rapidly varying network topology. Despite recent advances, FANETs are not yet a mature technology; for actual deployment, either custom, optimized Wi-Fi extensions must be developed, or falling back to LTE infrastructure is unavoidable, at the cost of increased latency.

III. Communications and Data Streaming in MULTIDRONE

The MULTIDRONE platform is composed of multiple components, which are briefly described in this Section for clarity. Subsequently, the emphasis is given on describing MULTIDRONE communications in detail. Finally, the reasoning behind the choices made, as well as the related challenges in the context of MULTIDRONE, are provided.

A. The MULTIDRONE architecture

The proposed MULTIDRONE architecture includes a swarm of cooperating, camera-equipped UAVs and a central *Ground Station*. The Ground Station is employed by the Director and his team for pre-planning the shooting mission (using an appropriate GUI named "Dashboard"), for dynamic, autonomous mission re-planning and execution monitoring, as well as for semantic environment mapping concerning human crowds, since such areas constitute no-flight zones due to regulations and safety issues. Additionally, the Supervision Station is included on-ground, i.e., a GUI permitting a human operator to constantly monitor the status of the UAVs, so as to cancel the mission in case any security issues arise during the execution.

The UAVs are responsible for collectively executing the mission (mainly filming and physically following pre-specified moving targets, in an adaptive manner, so as to capture the desired shot types), as well as for gathering visual data to facilitate semantic mapping and target on-map localization. Multiple functionalities such as autonomous UAV localization and navigation, collision avoidance and camera control are implemented. Among other equipment, each UAV carries a PixHawk/PX4 Autopilot [12] (i.e., a popular low-level flight trajectory control system), an NVIDIA Jetson Tegra X2 computing board (containing a CPU and a GP-GPU), two cameras (a "navigation camera" and a "cinematographic camera") and a gimbal. The Jetson TX2 board is employed for real-time visual analysis of the captured video frames [13]. A range of human-centered visual analysis algorithms could be employed [14], [15], [16], [17], [18], [19], although the emphasis of MULTIDRONE is on object detection and tracking [20], [21], [22].

B. MULTIDRONE Communications Module

The majority of the communication exchanges between the UAVs and the Ground Station, including real-time live video streaming, are assured by an LTE system, composed of an LTE user equipment on-board and an LTE base station on-ground. Inter-UAV communications are assured by a WiFi mesh (WiFi). Each UAV carries onboard a dedicated *Communication* module that is responsible for:

- Acting as a default IP communication gateway/router to the ground and to other UAVs.
- Scheduling IP flows depending on applications precedence and assigned IP Quality of Service (QoS).

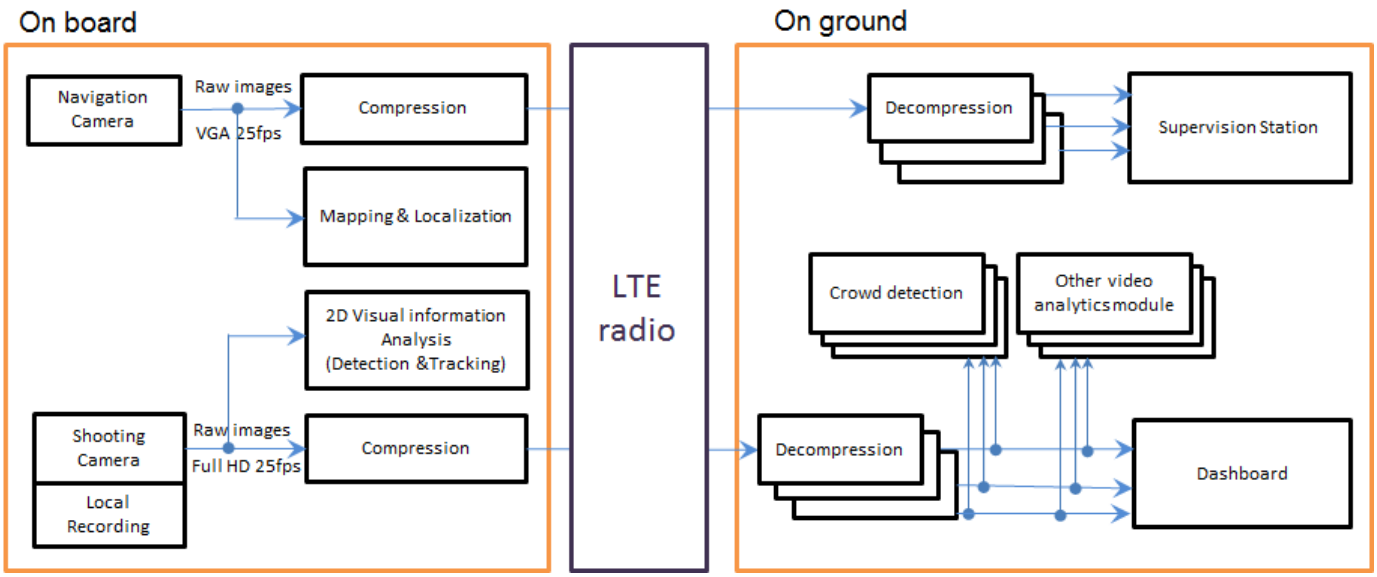

Fig. 1: Data flow for video streaming from the UAVs to the ground.

- Traffic shaping / admission control when congestion occurs.
- Authentication, encryption and other security-related mechanisms.

The Communication module can be considered as a default IP router for the rest of the system. As such, it exposes an Ethernet interface to the computers on-board the UAV and implements a full IP protocol stack. Since it is fully independent from the other modules in the architecture, it has its own hardware and own operating system (Linux OpenWRT). In addition, a separate *Video Streamer* module is necessary for video transmission and interacts heavily with the Communication module. For each UAV, two video streams are generated: one by the navigation camera (H.264 compressed, 4:2:0 chroma sub-sampling, 640x480 resolution), another one by the cinematographic camera (H.264 compressed, 4:2:2 chroma subsampling, 1920x1080 resolution, @25fps).

A Blackmagic Micro Cinema Camera with a motorised Panasonic x3 lens was selected as the cinematographic camera, supporting Full HD resolution. On the other hand, the navigation camera does not require FullHD, since its main purpose is simply to provide the Supervisor with good situational awareness. Compression takes place on-board the NVIDIA Jetson TX2 platform, which offers hardware-accelerated image/video compression. Video streams are then transmitted through the LTE radio network using the Real-time Transport Protocol (RTP). The RTP Control Protocol (RTCP) is also used for on-ground synchronization of video streams coming from different UAVs. The RTP packets hold a 32-bit RTP timestamp. Several consecutive RTP packets may have equal timestamps if they are (logically) generated at once, e.g., they belong to the same video frame. The Sender Report packet holds the correspondence between the RTP timestamp and the absolute 64-bit timestamp (system hour), that is broadcasted through the LTE network thanks to the Network Time Protocol (NTP).

Figure 1 depicts the data-flow for all video streams. It is assumed that 3 UAVs are connected to the Ground Station. The cinematographic camera video streams are transmitted to the Dashboard through the radio network. In parallel, the streams are also resized so that they can be processed on-board by the perception modules. The Video Streamer can process either the images coming from the cinematographic camera, or from the navigation camera, depending on the situation. On-ground, these streams will be uncompressed to be displayed on the Dashboard and, also, resized to be processed by any video analysis modules, such as the Crowd Detection module.

Regarding system scalability, increasing the number of UAVs only implies updating the configuration and hardware of the LTE modules. Of course, the required bandwidth should be manageable by the communication base station. Apart from that, there is no other major impact on the overall communication architecture.

Redundant RF communications are also provided for safety, through additional links. An example would be in case of LTE streaming failure. The navigation stream is sent to the pilot via RF in manual mode, thus an analog signal is required. RF can also handle the commands to control the gimbal and the camera from a transmitter. The Pixhawk may receive at the same time commands coming from the RF receiver and from the on-board computer, which received it from the Dashboard through the LTE.

C. Communication and Data Streaming Challenges in MULTIDRONE

Real-time video streaming of sports events requires consideration of cinematography aspects, e.g., detecting and tracking targets of interest, keeping them centered and zoomed in at

specific level, using multiple drones to get multiview shots, or letting drones explore various shot types (e.g., close up), while at the same time the UAV has to fly autonomously, avoid obstacles, obey regulations (e.g., no flight over human crowds, or above the preset altitude, etc.).

In such a setting, a great number of factors have to be considered before designing a communications and data streaming architecture. For instance, wireless communications may be weak and subject to failure (due to distance, obstacles, other wireless networks, etc.), while good quality video is massive in terms of Mbps required to transfer it (1 second of 720p 8-bit video requires 65.92 MBytes, which is prohibitive). Additionally, video compression must be used prior to streaming; H264 and H265 coding are great candidates, but they inevitably introduce delays during compression/decompression, as well as a drop in quality due to their lossy nature. Finally, the choice of network protocol stack to be employed is not straightforward. However, Real-time Transport Protocol (RTP) with User Datagram Protocol (UDP) is a good choice, since TCP (although also standardized for use with RTP) favors reliability instead of timeliness.

An important issue in the MULTIDRONE setting is synchronization, due to multiple UAVs sending video streams to the Ground Station (GS) concurrently and the GS providing the UAVs with feedback (e.g., navigation commands). Since low latency and clock synchronization are required, the NTP protocol was selected for ensuring that all participating devices use the same clock.

Visual analysis on the captured video frames introduces delays and, therefore, additional synchronization issues. Each transmitted video frame must be accompanied by metadata, including the results of visual analysis, the NTP timestamp corresponding to the moment it was captured, as well as UAV telemetry status, gimbal status and camera status at that moment. Such metadata can be sent as a separate stream, but then synchronization of metadata and video frames must take place at the receiver (the GS), which is problematic. Alternatively, they may be inserted into the stream, assuming they can survive the compression process (no watermarking). The best solution would probably be to insert metadata as RTP header extension.

GStreamer [23], an open source multimedia framework which offers bindings in multiple programming languages (including C/C++ and Python), can be employed for low-level handling of the above issues.

IV. Conclusions

In this paper, general communication and data streaming challenges when employing UAV swarms for media production have been presented. Additionally, a specially designed platform consisting of a partially autonomous UAV fleet and a central ground station for live filming of outdoor events is described, along with its communication and data streaming modules. It includes a set of possible solutions to the discussed issues, employing off-the-shelf tools wherever possible. Additionally, the reasoning behind the choices made is explained, in the context of the proposed platform.

V. Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE). This publication reflects the authors' views only. The European Commission is not responsible for any use that may be made of the information it contains.

References

[1] M. Goldman, "Drone implications," *SMPTE Newswatch*, 2016.

[2] O. Zachariadis, V. Mygdalis, I. Mademlis, N. Nikolaidis, and I. Pitas, "2D visual tracking for sports UAV cinematography applications," *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2017.

[3] I. Mademlis, V. Mygdalis, C. Raptopoulou, N. Nikolaidis, N. Heise, T. Koch, J. Grunfeld, T. Wagner, A. Messina, F. Negro, et al., "Overview of drone cinematography for sports filming," *European Conference on Visual Media Production (CVMP), short*, 2017.

[4] A. Torres-González, J. Capitán, R. Cunha, A. Ollero, and I. Mademlis, "A multidrone approach for autonomous cinematography planning," *Iberian Robotics Conference (ROBOT)*, 2017.

[5] I. Mademlis, V. Mygdalis, N. Nikolaidis, and I. Pitas, "Challenges in Autonomous UAV Cinematography: An Overview," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2018.

[6] I. Karakostas, I. Mademlis, N. Nikolaidis, and I. Pitas, "UAV cinematography constraints imposed by visual target tracking," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2018.

[7] I. Karakostas, I. Mademlis, N. Nikolaidis, and I. Pitas, "Shot type feasibility in autonomous UAV cinematography," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[8] I. Mademlis, N. Nikolaidis, A. Tefas, I. Pitas, T. Wagner, and A. Messina, "Autonomous unmanned aerial vehicles filming in dynamic unstructured outdoor environments," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 147–153, 2018.

[9] A. Messina, S. Metta, M. Montagnuolo, F. Negro, V. Mygdalis, I. Pitas, J. Capitán, A. Torres, S. Boyle, and D. Bull, "The future of media production through multi-drones' eyes," in *International Broadcasting Convention (IBC)*, 2018.

[10] I. Karakostas, I. Mademlis, N. Nikolaidis, and I. Pitas, "Shot type constraints in UAV cinematography for target tracking applications," *Information Sciences*, 2019, submitted.

[11] I. Mademlis, V. Mygdalis, N. Nikolaidis, M. Montagnuolo, F. Negro, A. Messina, and I. Pitas, "High-level multiple-UAV cinematography tools for covering outdoor events," *IEEE Transactions on Broadcasting*, 2019.

[12] L. Meier, P. Tanskanen, F. Fraundorfer, and M. Pollefeys, "Pixhawk: A system for autonomous flight using onboard computer vision," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011.

[13] P. Nousi, E. Patsiouras, A. Tefas, and I. Pitas, "Convolutional neural networks for visual information analysis with limited computing resources," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2018.

[14] I. Mademlis, A. Tefas, and I. Pitas, "A salient dictionary learning framework for activity video summarization via key-frame extraction," *Information Sciences*, vol. 432, pp. 319 – 331, 2018.

[15] I. Mademlis, A. Tefas, N. Nikolaidis, and I. Pitas, "Summarization of human activity videos via low-rank approximation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[16] I. Mademlis, A. Tefas, and I. Pitas, "Summarization of human activity videos using a salient dictionary," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2017.

[17] I. Mademlis, A. Tefas, and I. Pitas, "Regularized SVD-based video frame saliency for unsupervised activity video summarization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2018.

[18] I. Mademlis, A. Tefas, and I. Pitas, "Greedy salient dictionary learning for activity video summarization," in *Proceedings of the International Conference on MultiMedia Modeling (MMM)*. 2019, Springer.
[19] I. Mademlis, A. Tefas, and I. Pitas, "Greedy salient dictionary learning with optimal point reconstruction for activity video summarization," in *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2018.
[20] G. Symeonidis and A. Tefas, "Recurrent attention for deep neural object detection," in *Proceedings of the Hellenic Conference on Artificial Intelligence*. ACM, 2018.
[21] M. Tzelepi and A. Tefas, "Human crowd detection for drone flight safety using convolutional neural networks," in *Proceedings of the EURASIP Signal Processing Conference (EUSIPCO)*. IEEE, 2017.
[22] D. Triantafyllidou, P. Nousi, and A. Tefas, "Lightweight two-stream convolutional face detection," in *Proceedings of the EURASIP Signal Processing Conference (EUSIPCO)*. IEEE, 2017.
[23] "Gstreamer: Open source multimedia framework," .