High-level Multiple-UAV Cinematography Tools for Covering Outdoor Events

Ioannis Mademlis*, Vasileios Mygdalis*, Nikos Nikolaidis*, Maurizio Montagnuolo†, Fulvio Negro†, Alberto Messina† and Ioannis Pitas*
*Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
‡Radiotelevisione Italiana (RAI) Centre for Research and Technological Innovation, Torino, IT

***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, this paper presents an overview of the state-of-the-art in this area, along with a brief review of current commercial UAV technologies and legal restrictions on their deployment. A novel taxonomy of UAV cinematography visual building blocks, in the context of filming outdoor events where targets (e.g., athletes) must be actively followed, is additionally proposed. Such a taxonomy is necessary for progress in intelligent/autonomous UAV shooting, which has the potential of addressing current technology challenges. Subsequently, the concepts and advantages inherent in multiple-UAV cinematography are introduced. The core of multiple-UAV cinematography consists in identifying different combinations of multiple single-UAV camera motion types, assembled in meaningful sequences. Finally, based on the defined UAV/camera motion types, tools for managing a partially autonomous, multiple-UAV fleet from the director's point of view are presented. Although the overall focus is on cinematic coverage of sports events, the majority of our contributions also apply in different scenarios such as movies/TV production, newsgathering or advertising.**

Keywords*—UAV cinematography, media production, drone swarm, shot types, intelligent shooting

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.

However, at the current technology level, several challenges arise. Minimal battery life/flight time (typically, less than 25 minutes), limited payload and disturbing sound issues, as well as safety considerations, are all factors interacting with cinematography planning. Legal requirements in most countries, such as a demand for direct line-of-sight between the pilot and the UAV at all times, or restrictions on flight above human crowds, complicate shooting. As a result, strict manual mission planning and designation of tight flight corridors burden the director and limit the creative potential. In case multiple UAVs are employed, synchronization and collision avoidance issues, as well as a need for each UAV to avoid entering the field-of-view of the others (*FoV avoidance*), further perplex the situation. Many of the above problems can be alleviated by automating the technical/non-creative aspects of UAV filming, through intelligent/autonomous UAV shooting software. However, the related research is still in its early stages and is plagued by a lack of standardization in UAV cinematography, which severely limits the available creative possibilities.

Following early preliminary work [2], [3], [4], [5], [6], [7], [8], this paper is an overview of the current situation in UAV cinematography, with an emphasis on filming non-scripted, outdoor events where targets (e.g., athletes) must be actively followed. This is deemed as the most complex application, due to highly unpredictable circumstances and a typically large area to be covered. First, we present the state-of-the-art on the use of drones in media production and broadcasting, along with a brief review of relevant UAV technologies and legal restrictions on their deployment. Up to now, commercial autonomous drone shooting applications are simplistic and mostly support a single UAV. As a first step towards remedying this situation, a newly developed UAV shot type taxonomy is proposed, emphasizing the capture of moving targets. The limitations of single-UAV shooting and the advantages of employing a fleet of multiple drones are discussed from a media production perspective. Finally, based on the developed UAV shot type taxonomy, tools for managing a multiple-UAV fleet from the director's point of view are presented. Thus, we attempt to streamline the process of creating/editing/tracking UAV shooting missions, while exploiting unplanned/opportunistic shooting potentials. In the context of this paper, a UAV shooting mission consists in a meaningfully assembled sequence of the proposed multiple-UAV/camera motion types. Multiple-UAV cinematography is further detailed through example scenarios.

II. State-of-the-art

Despite a number of legal restrictions, more and more live events in sport and entertainment are produced with the support of UAVs. Drone footage adds value to the filming of concerts and music festivals, as it gives viewers a realistic view of the size and dynamics of the event, without having to use highly expensive equipment such as helicopters or cranes.

A. Drones in media production

The use of UAVs in movie productions became common in the past years, as an efficient way to save time and money. Drones are substituting both dollies (static cranes) and helicopters. They allow, for instance, pan shots starting inside a building and moving up to an altitude of one hundred meters. Drones equipped with cameras can be used for shooting inside and outside buildings. They are suitable for filming landscapes from different angles, as well as (car) chases in action movies. Literal birds' eye views, as well as 360-degree views of persons and objects, significantly improve the viewers' experience. Probably the most well-known examples of the use of drones in movie productions is *James Bond - Skyfall* (2012) and *The Wolf of Wall Street* (2013), where they have been used to capture motorbike chasing scenes and building overview scenes, that would otherwise require a helicopter to be shot.

Drones also offer broadcasters an attractive option for the live coverage of sports, weather-related events and natural disasters, as well as coverage of events in areas particularly difficult to reach (e.g., war zones). News agencies have begun to embrace drones for aerial newsgathering, e.g., for covering hurricane destruction areas [9]. In advertising, many firms use drones for special effects, while production companies specialised on sport events employ UAVs not only for obtaining captivating footage, but also as part of the general coverage itself (e.g., in major international sporting events, such as the Olympics [10] [11] or the "World Rally Championship"), with several UAV pilots involved in the production process. According to the WRC organizers, UAVs allow close-ups from short distances and low altitudes without having to fly above cars or spectators.

B. Intelligent UAV cinematography

Algorithms have been presented recently [12] [13] that automate the flight process in single-UAV shooting missions, typically by pre-computing motor command and camera rotation sequences, so as footage with desired properties to be captured. Care is taken for cinematographic principles to be obeyed, but the emphasis is on static shots and environments. Related software systems [14] [15] require temporally ordered, example, desired "key-frames" pre-specified by the director on a 3D reconstruction of the scene, as a rudimentary cinematography plan.

Coming from a different research line, the method in [16] modifies a manually designed UAV trajectory with regard to vehicle physical limits, so as to produce an optimized variant guaranteed to be feasible, while respecting the visual content intended to be captured. A more general approach is presented in [17], where custom high-level user goals are taken into account (e.g., encoding cinematography goals).

Little effort has been expended on automating multiple-UAV shooting. Besides simply calculating optimal number of drones so as to maximize target coverage [18], a more advanced method has been presented in [19]. It is an on-line real-time planning algorithm that jointly optimizes feasible trajectories and control inputs for multiple UAVs filming a dynamic, indoor scene with FoV avoidance, by processing user-specified aesthetic objectives and high-level cinematography plans. It extends a previous, single-UAV method [20] that only optimized local trajectory segments.

III. UAV Technologies and Regulations

UAV cinematography relies on high-end drone and camera technologies that have been successfully commercialized on a large scale during the past decade. This Section briefly overviews these technologies, including the recent trend of panoramic cameras, as well as regulations and legal restrictions on their deployment.

A. Commercial Cinematography UAVs and Flight Regulations

In commercial UAV systems oriented to cinematography applications, flight time is typically restricted to 25-30 minutes on ideal conditions, while performing minimal UAV control operations. Flight time is determined by the employed battery capacity and UAV frame energy consumption. However, increasing battery capacity does not necessarily lead to increased flight time, due to the weight implications introduced to the UAV frame and engines as well (large batteries require large UAV frames for fitting, which in turn require larger engines). In terms of energy consumption, the following UAV operation ordering may be defined, from the least to most battery-intensive: camera operations (gimbal rotations, zoom), flying down, flying horizontally/hovering, flying up. In general, engine operations dominate the energy-related behavior, with camera operations being relatively negligible.

In terms of physical UAV specifications, maximum air speed ranges between 55-75km/h while ascend speed (acceleration) ranges between 3-5m/s. In fact, exploitable drone speed and acceleration may be less, taking account inertia effects and capturing undesirable motion blur. Moreover, high speed maneuvering is very costly in terms of battery consumption and perhaps requires challenging UAV piloting skills. Exploitable air speed and ascend/descend speed of UAVs are important when framing motion types, (e.g., Chase, Section IV,B) especially when applied to fast targets, while maintaining a static shot type (e.g., Medium Close Up, Section IV.A) at the same time. Thankfully, lack of UAV flying speed/acceleration can be compensated by shooting at a distance, exploiting a combination of a fast rotating stable gimbal, carrying a stable camera lens. Physical drone specifications are expected to be improving over time at a fast pace, while UAVs remain an emerging technology.

UAV operational limitations may also be imposed by legal restrictions. For generic applications, UAVs are typically classified into different categories, depending on their weight.

Adjusting/replacing components, may impact the category classification. For instance, UAVs exceeding 2kg of weight may be required to carry emergency parachutes in some countries [21]. Flying UAVs exceeding 15kg of weight might require special license or even be prohibited [22]. To this end, commercial cinematography UAVs typically do not exceed such limits. Maximum drone flight altitude is typically restricted to 400ft or 500ft (120m or 150m) within several European countries [21] [22] [23] [24] [25]. Visual Line-of-Sight (LOS) should be maintained by the licensed pilot of the UAV, either physically, or using visual aids (e.g., VR-goggles), while the horizontal distance between the drone and the pilot may be limited to specific meters (e.g., 500m). In addition, due to safety considerations, outdoor UAV flight in most countries is restricted above congested areas, crowds of people and airports, leading to permissible flight zones delineated by law ("geofencing"). Another important issue is that flight restrictions vary over different countries, while professional pilot licenses and insurance policies may not be internationally valid.

Thankfully, current aerial path planning algorithms are able to deal with complex dynamic and kinematic constraints in real-time, resulting in nearly-optimal collision-free paths being computed on-line [26]. In any case, considering the restrictions imposed by UAV specifications and flight regulations, intelligent UAV shooting should prioritize gimbal/camera control over vehicle control [27], [28].

B. *UAV Panoramic Cameras*

During the past years, commercial panoramic cameras designed especially for UAVs have started to appear, typically based on multiple-camera rigs. Drone cinematography seems a particularly effective use-case for 360° footage, since combining panoramic capture with the agile, aerial point-of-view offered by a camera-equipped UAV can potentially produce highly impressive visual results. Additionally, recent state-of-the-art drones for cinematography applications can optionally transmit live panoramic footage to a VR head-mounted display that supports head-tracking. This is performed wirelessly and at a low latency, thus enabling the end-user to realistically experience a dynamic scene from a uniquely immersive, first-person, aerial standpoint.

Video stitching algorithms [29], [30], [31], [32] constitute a basic building block of panoramic camera technology. In general, they traditionally rely on image stitching methods, composed of successive keypoint detection (SIFT [33] keypoints are typically employed), image alignment (single or double homography estimation is common), calibration and blending stages. Alignment is performed either by appropriately cutting and joining multiple images seamlessly, or by properly warping images. Video extensions of such methods may exploit inter-frame content redundancy to avoid frame-by-frame image stitching, or attempt to reconstruct a common virtual camera path in space so as to simultaneously stitch and stabilize videos in-software. Additionally, the spatiotemporal nature of the video input permits the introduction of temporal smoothness constraints during stitching computations, as well as the exploitation of dense optical flow maps for facilitating the process.

UAV panoramic cameras are an exceptional use-case for video stitching algorithms. Using a temporally synchronized, fixed multiple-camera rig allows for a constant, pre-computed homography, thus simplifying computations and ensuring real-time operation. Although high speeds and intense vibrations that are common in UAV footage capture complicate the situation, relevant products are mature enough for mass market adoption as of late 2018. Top-of-the-line panoramic UAV cameras currently offer a 4K-6K resolution at 30 fps (alternatively, FullHD-3K resolution at 60 fps), with full 360° horizontal coverage, 180° (or more) vertical coverage and no blind spots.

This is achieved by combining a temporally synchronized multiple-camera rig, composed of precisely localized cameras with overlapping fields-of-view, a camera stabilization system and advanced live image/video stitching software. When only two cameras are used, fisheye lenses are typically employed and the resulting image distortion is handled in software. On the downside, live VR capabilities come with a steep monetary overhead and possibly reduced video resolution, while a multiple-camera rig can significantly increase the UAV weight. Obviously, this may be a critical issue when flying drones with limited payloads and/or limited battery capacity.

Current usage of panoramic cameras for general broadcasting is questionable due to the limited dynamics of the shooting one can obtain and the relatively low resolution of each view. Nevertheless, they can be fruitfully used for documenting places and using the resulting images for subsequent editing and composition. Mounting panoramic cameras on UAVs, that have the ability to shoot targets unfilmable with conventional means (helicopters, ground cameras), can potentially bring innovation to the narration. Good examples would be shooting caves, or jungles. Furthermore, the possibility of a drone to fly at relatively low heights makes it viable for innovative mobile shooting otherwise impossible with helicopters (which must fly at a certain minimal height).

IV. Single-UAV Cinematography

The various shot types in UAV cinematography can be described using two complementary criteria: the UAV/camera motion trajectory and the framing shot type. Each camera motion can be successfully combined with a subset of the possible framing shot types, according to the directorial specifications, in order to achieve a pleasant visual result.

A. *UAV Framing Shot Types*

The framing shot types are primarily defined by the relative size of the main subject/target being filmed (if any) to the video frame size. The following framing shot types can be defined [34]:

- Extreme Long Shot (ELS)
- Very Long Shot (VLS)
- Long Shot (LS)
- Medium Shot (MS)
- Medium Close-Up (MCU)
- Close-Up (CU)

- Two-Shot/Three-Shot (2S/3S)
- Over-The-Shoulder (OTS)

These correspond almost exactly to traditional cinematography framing shot types, except Over-the-Shoulder (OTS) which is defined here slightly differently: the main target is clearly visible at a fairly short distance from the camera, covering about $\frac{1}{3}$ of the video frame, while a secondary target is visible at the edge of the video frame and at a much shorter distance from the camera. OTS can be regarded as a variant of the Two-Shot, with the main target being filmed as in a Long Shot or Medium Shot and with the secondary target being filmed as in a Close-Up or a Medium Close-Up. Either the main or the secondary target can be a geographical landmark (e.g., a historical monument).

B. UAV Camera Motion Shot Types

Several standard types of UAV/camera motion trajectories have emerged since the popularization of UAVs. As in the case of framing shot types, most of them are derived/adapted from the ones found in traditional ground and aerial cinematography. For outdoor event coverage, the most important ones are the motion types that are relative to a (still or moving) target being filmed, in contrast to camera motion types that emphasize capturing the scene. UAV camera motion types involving actual target filming are briefly surveyed below, with illustrations shown in Figure 1.

- *Moving Aerial Pan with Moving Target* (MAPMT) and *Moving Aerial Tilt with Moving Target* (MATMT): Camera motion types, where the camera gimbal rotates slowly (mainly with respect to the yaw axis for MATMT, and the pitch axis for MATMT) so as to always keep the linearly moving target properly framed, while the UAV is slowly flying at a steady trajectory with constant velocity. The target and the UAV trajectory projections onto the ground plane are approximately perpendicular (MAPMT) or parallel (MATMT) to each other. MAPMT and MATMT fit well with LS, MS, MCU, 2S/3S and OTS framing shot types.
- *Lateral Tracking Shot* (LTS) and *Vertical Tracking Shot* (VTS): Camera motion types where the camera gimbal remains static and the camera always focused on the linearly moving target [35], [36]. The camera axis is approximately perpendicular to the target trajectory and parallel to the ground plane. The UAV flies sideways/in parallel to the target, matching its speed if possible (LTS). VTS is a variant of LTS where the UAV flies exactly above the moving target, the drone camera stays focused on the target vertically, and the camera axis is perpendicular to the target trajectory. Both LTS and VTS fit well with VLS, LS, MS, MCU and 2S/3S framing shot types, while LTS also fits with OTS.
- *Pedestal/Elevator Shot With Target* (PST): A camera motion type where the UAV is slowly flying up or down, along the z-axis, with constant velocity [35], [36]. The camera gimbal rotates slowly (mainly along the pitch axis), so as to always keep the linearly moving target properly framed. The projections of the camera axis and of the target trajectory on the ground plane remain approximately parallel during shooting. PST fits well with LS, MS and 2S/3S framing shot types.
- *Reveal Shot* (RS): A camera motion type where the camera gimbal is static, with the target initially out of frame (e.g., hidden behind an obstacle) [36]. The UAV flies at a steady trajectory with constant velocity, until the target becomes fully visible. RS fits well with LS, MS and 2S/3S framing shot types.
- *Orbit* (ORBIT): A camera motion type, where the camera gimbal is slowly rotating, so as to always keep the still or linearly moving target properly framed, while the UAV (semi-)circles around the target and, simultaneously, follows the latter's linear trajectory (if any) [35], [36]. During shooting, the difference in altitude between the UAV and the target remains constant. ORBIT fits well with VLS, LS, MS, MCU, CU and 2S/3S framing shot types.
- *Fly-Over* (FLYOVER): A camera motion type where the camera gimbal is slowly rotating (mainly along the pitch axis), so as to always keep the still or linearly moving target properly framed. The UAV follows/intercepts the target from behind/from the front, flying parallel to its trajectory and with constant velocity, until passing over the target. Then, it keeps flying along the same trajectory for some time, with the camera still focusing on the receding target. FLYOVER fits well with LS, MS, MCU, CU and 2S/3S framing shot types.
- *Fly-By* (FLYBY): A camera motion type where the camera gimbal is slowly rotating, so as to always keep the still or linearly moving target properly framed, while the UAV follows/intercepts the target from behind/from the front and to the left/right, passes it by and keeps on flying at a linear trajectory with steady altitude [36]. The UAV and target trajectory projections onto the ground plane remain approximately parallel during shooting. FLYBY fits well with LS, MS, MCU, CU and 2S/3S framing shot types.
- *Chase/Follow Shot* (CHASE): A camera motion type where the camera gimbal remains static and the camera always focused on the target [36]. The UAV follows/leads the target from behind/from the front, at a steady trajectory, steady distance and matching its speed if possible. CHASE fits well with VLS, LS, MS, 2S/3S and OTS framing shot types.

V. Multiple-UAV Media Production

Today, UAV-assisted media production is based on a single UAV, leading to several limitations which are discussed below.

A. Limitations of single-UAV shooting

In off-line shooting with full post-production editing (i.e., for movies/TV), the single available UAV works at different time intervals to produce several takes, which are then processed and edited in the post-production phase. Simultaneous scene coverage from multiple viewpoints cannot be truly achieved and, thus, the editor has to work with less available

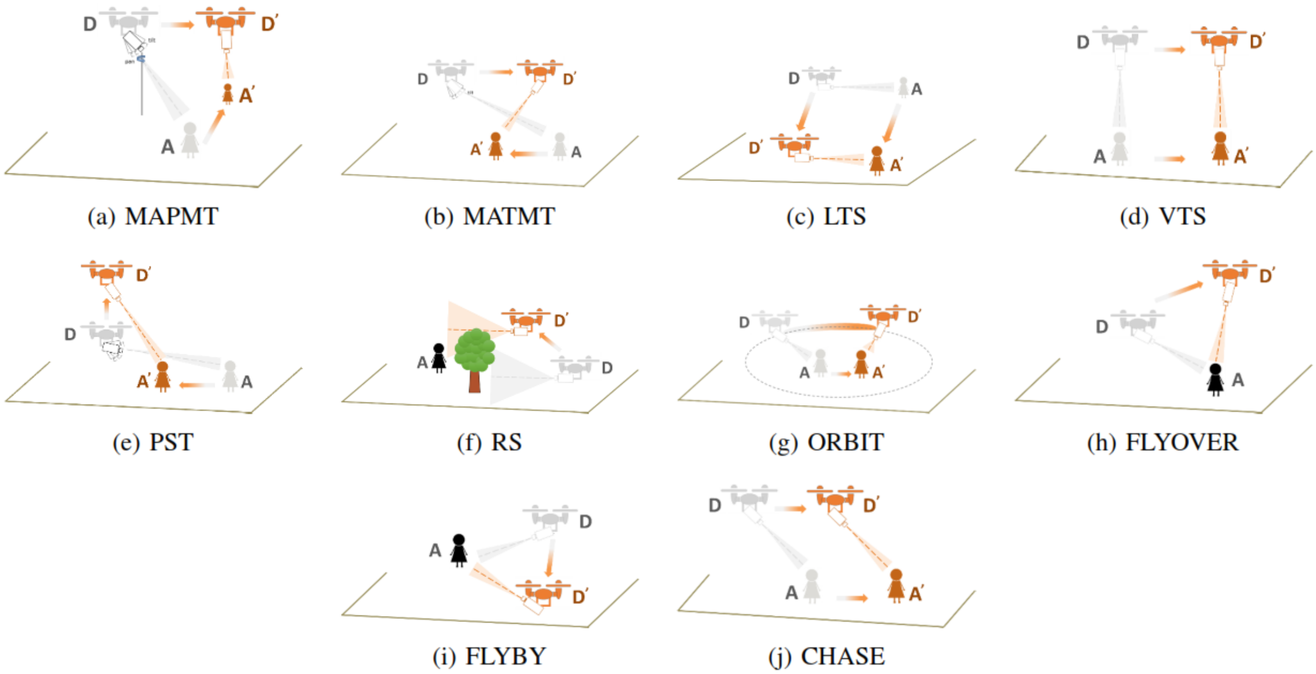

Fig. 1: Geometry of the described camera motion types.

raw footage. Alternatively, the scene may be re-shot from different angles. However, this leads to increased time and monetary costs, while it is also vulnerable to varying natural lighting conditions.

When filming live events for deferred broadcast and, thus, with potential post-production modifications, we can assume that each scene is different at any time, with no chance to recreate it from different angles if needed. Uninterrupted shooting of such a scene by a single UAV is possible but, at any given time instance, the scene is still observed from a single viewpoint. Therefore, the artistic possibilities are greatly reduced due to less available footage.

Single-UAV shooting of live events (i.e., for live TV programmes) with limited post-production, bears all disadvantages of shooting a live event with a single hand-held camera, i.e., mainly, no extensive editing/post-production can be performed. In this context, the constraint of a single-viewpoint per time instance is coupled with the viewer being directly exposed to “dead” time intervals (e.g., due to the need for a single UAV to travel between different shooting locations of interest, or for recharging).

B. Multiple-UAV cinematography

A fleet of UAVs may overcome the above issues, while additionally providing a broader range of angles, the ability to capture a scene overview from above simultaneously with regular medium/close-up takes, spatiotemporally extensive scene coverage by employing drone relays, as well as the possibility for novel cinematographic effects, by integrating footage from multiple UAVs.

Conceptually, multiple-UAV cinematography involves either a swarm of cooperating UAVs, or many individual drones simultaneously shooting the same event. This Section presents some interesting scenarios and specifies the difference between these approaches.

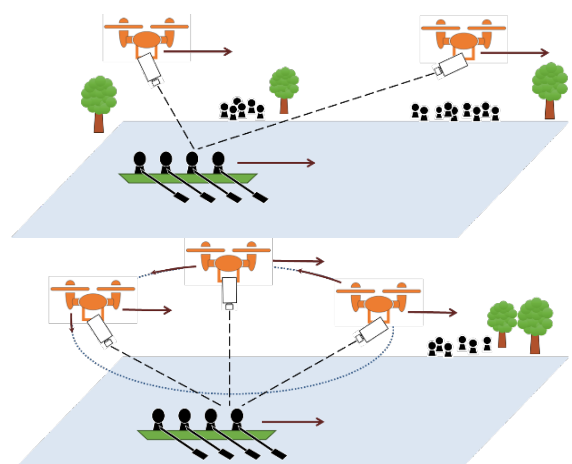

Fig. 2: Target following using two UAVs (top) and orbiting around target using three UAVs (bottom)

First, two UAVs following a moving target is the most

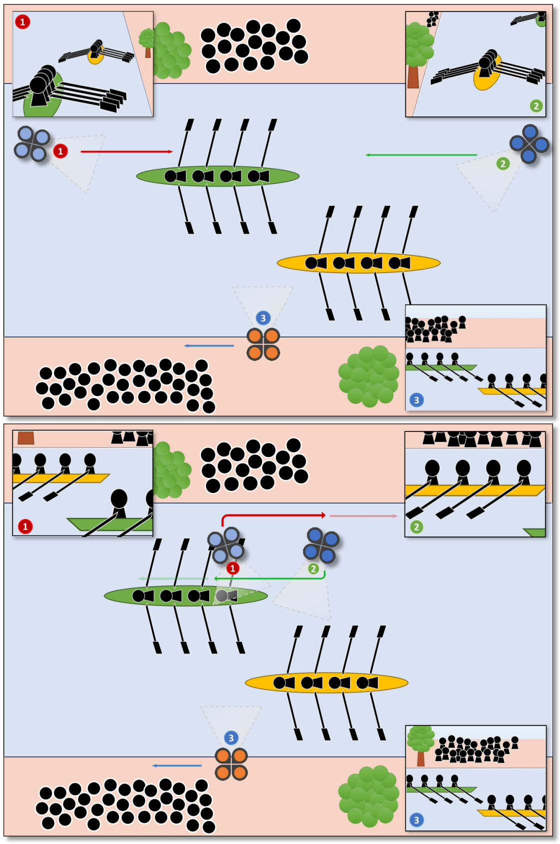

Fig. 3: The Dancing Drones scenario.

obvious way of employing multiple drones in outdoor events. One drone may follow the target from the front and the other one from behind, along the target motion trajectory. Their framing shot types may differ, so that the two drones provide complementary visual information. For example, the front UAV may shoot a CU, while the trailing drone may shoot a LS.

Second, three UAVs orbiting the same target with a phase difference and different framing shot types can be another interesting shooting type. Such configurations may lead to novel visual effects, e.g., rotational views of the moving target. Both scenarios are illustrated in Figure 2.

A significant technical difference between the two above scenarios is the following. In the former case, each drone executes a different task, however, its movement should be constrained by the movement of the other drones as well, e.g., in order to maintain similar distance to the target. In the latter case, all drones essentially receive the same command, i.e., perform orbiting, and their positioning must be controlled by formation-based methods. Therefore, both cases imply a level of intra-swarm cooperation and would be tricky to address using many isolated drones.

Third, a more complex 2-UAV scenario is presented, the "Dancing Drones", using an example rowboat race. At all stages, both drones should keep focus on the rowboats. Initially, both drones fly at the same altitude, starting from a fixed distance between one another, with the one in front of a target rowboat and the other one behind the target, so that the boat lies exactly halfway between them. Then, the first half of 2 opposite-direction Fly-Overs is executed, with the trailing UAV flying much faster and the two drones moving toward each other and the boat. Shortly before they would collide (ideally directly above the boat) they avoid hitting by moving in opposite directions, perpendicular to their previous ones, without slowing down, losing focus on the boat or changing altitude (MAPMT). Subsequently, both drones change again their flight direction to parallel with the boat's and continue flying, now performing the second half of 2 opposite-direction Fly-Bys. Finally, both drones slow down and reverse their flight directions, heading again towards each other. This procedure, illustrated in Figure 3, can be repeated until the race is finished.

As it can be observed, multiple-UAV scenarios can be viewed as constrained combinations of existing single-UAV shot types, essentially defining composite cinematography patterns. Multiple-UAV cinematography allows many such scenarios to be defined, opening an exciting new field for technical and industrial applications.

C. Opportunistic Shooting Exploiting Multiple-UAV Cinematography

In addition to devising complex camera motion types, multiple-UAV cinematography also enables enhanced opportunistic shooting capabilities. For example, unexpected/unplanned events might occur during live coverage (or even expected events at unknown time-instances), including e.g., leader break-away, crashes, falls, accidents. A number of these events may be automatically detected visually, by on-the-fly activity recognition [37], [38], [39], [40], [41], [42], or activity video summarization [43], [44], [45], [46], [47], [48], [49] systems. Alternatively, they can be manually annotated by a director that overviews the coverage. In addition to such events, chances for opportunistic shooting might arise that are related to artistic events, e.g., perfect lighting conditions to shoot a monument nearby. In order to capture such events, specific behaviors that consist of a combination of camera motion types (e.g., ORBIT around the fall) might be triggered and assigned to the drone fleet. An example of the above scenario is depicted in Figure 4.

VI. Media Production Tools for Multiple-UAV Cinematography

In this Section, we describe the necessary ground infrastructure that is required to integrate all the multiple-UAV cinematography advantages[1]. A set of tools is proposed, to be used by the director and his team in order to manage the UAV fleet from the editorial point of view, during both the preproduction and the production stage. These tools constitute

[1]Every flight and shooting mission is required to be in accordance with relevant flight rules and legislation, as well as respect safety and security implications. For the purpose of simplicity, this paper only focuses on the artistic perspective.

Fig. 4: Drone 1 detects leader breakaway and starts following him.

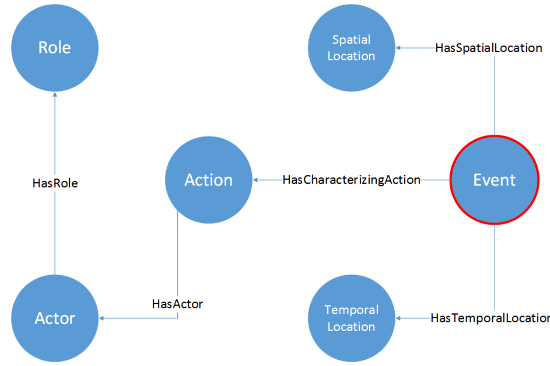

Fig. 5: Fundamental conceptualisation.

the "Director's Dashboard", a software interface that allows the director to pre-specify and manage a UAV shooting mission in the context of cinematographic requirements. As such, the UAV shooting mission is a semantically, spatially and temporally structured high-level definition of a mission plan for the UAV fleet, i.e., "what" is to be filmed, "when" and "how". The output of the Dashboard is assumed to be fed to a mission planning software, similar to the one proposed in [50], in order to be checked in terms of feasibility, and thereby produce low-level commands (flight path trajectories, camera control commands) for each actor (drone) in the UAV fleet. The Director's Dashboard interacts with the mission planning software, in order to manage the UAV fleet during production, and perhaps even modify the shooting mission upon execution.

At the core of the mission plan, is the definition of Events of editorial interest. Before shooting, a list of Events that have an editorial relevance in a media production process are explored. Such events may be defined as any real-world occurrence that is (at least partially) spatially and temporally localized, with a set of actors playing different roles in this action, e.g., the start of a race. The key point is that a Shooting Mission is always linked to Events and represents the (possibly planned) reaction to their occurrence. Therefore, an Event reaction maybe defined by a Shooting Action or more, i.e., a Shooting Action Sequence. The conceptual framework is depicted in Figure 5. The higher level process upstream of any Mission Production is that of Event Management, i.e., the process by which the Events are organized hierarchically and associated to Shooting Action Sequences to be executed in reaction to them.

A. Mission Preproduction

A Shooting Action Sequence is made up of an ordered sequence of (Editorial) Shooting Actions. Such a high-level Editorial Shooting Action may be subsequently translated into one or more executable low-level UAV commands by the mission planning software. A Shooting Action is characterized by its type and parameters (modeled as a separate class named Shooting Action Description). The type of a Shooting Action carries all the information regarding the geometric properties of the multiple-UAV formation that can be subject to parameterization during the Mission Preproduction (for example the

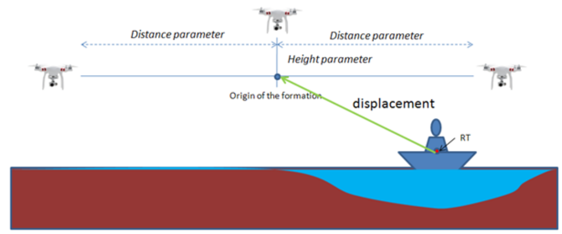

Fig. 6: Example of the relationship between the RT and the origin of the UAV formation.

angular speed of a circular shooting formation). The Shooting Action is normally aimed at having a reference target (RT) around which the formation and the shooting will initially take place. This aspect is tackled through several properties giving account of the position in space and speed of the formation origin, the geometric displacement of the RT from the formation origin and the relative azimuth of the formation axis w.r.t. the RT speed. If the RT is a fixed point, an absolute azimuth can be also provided. All these parameters are valid at the start of the Shooting Action (time $t = 0$). Figure 6 illustrates an example.

Variations of the tracking speed during the Shooting Action (for $0 < t < d$, where d is the duration of the shooting action) are modeled with the Tracking Variation class. The default situation is that there is no tracking variation, i.e., the constraint is that the formation speed must be equal (vectorially) to the RT speed, resulting in a situation where the formation is "locked" to the trajectory of the RT. Different situations can be modeled, specifying the type of functional variation of the tracking and possible parameters thereof. In general, the situation is that of Figure 7, where the vector $\mathbf{d}$ is the displacement between the origin of formation and the RT and $\mathbf{O} = \mathbf{r}(t)$ is the origin of a reference coordinate system $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ local to the RT w.r.t. which all vectors are defined over time. $\mathbf{O} = \mathbf{r}(t)$ is thus the RT trajectory and $\mathbf{p}(t)$ is a constant vector w.r.t. $\mathbf{O}$. Through the interface, the director has the possibility to define the tracking variation vectorial parameter $\mathbf{k}(t)$. Depending on the desired shooting effect, this tracking variation parameter can be $\mathbf{0}$ (null vector, formation

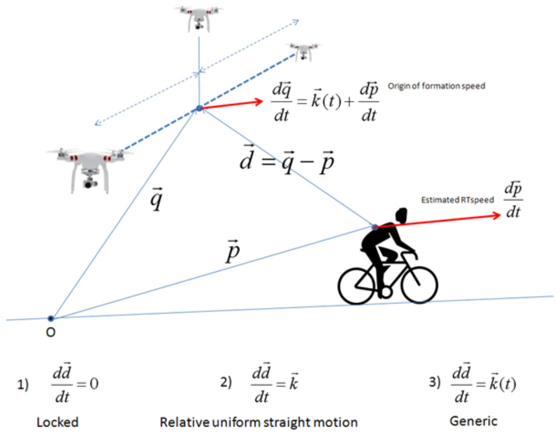

Fig. 7: Relative speeds.

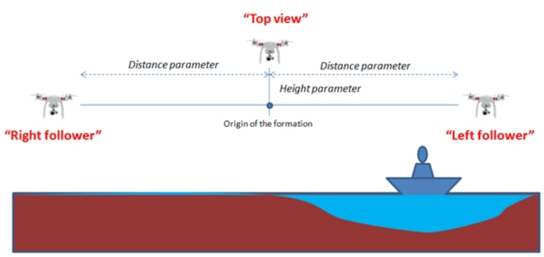

Fig. 8: An example of different Shooting Roles.

locked), a constant vector different than $\mathbf{0}$ (uniform straight motion relative to the RT trajectory) or a generic parametric vectorial function of time t. Thus, if $\mathbf{k}(t)$ is periodical, zero-average and bounded in $[0, T]$, it is ensured that the formation will not drift from the RT trajectory during the duration of the Shooting Action. The set of usable parametric functions for $\mathbf{k}(t)$ depends on the implemented UAV shot types.

If a prediction of the RT trajectory is available, following the equations of Figure 7, the director can have an estimation of the final positions as well. In fact, by simple integration:

$$\mathbf{q}(T) = \mathbf{q}(0) + \int_0^T \mathbf{k}(t)dt + \mathbf{p}(T) - \mathbf{p}(0) = \mathbf{q}(0) + \int_0^T \mathbf{k}(t)dt \tag{1}$$

where T is the Shooting Action duration. The final position of the origin of formation w.r.t. $\mathbf{O}(T)$ is the final position of the RT plus a displacement that depends on the tracking variation function. This information can be shown on the map and evaluated by the director, who can then recursively update the parameters until satisfaction. If the trajectory of the RT is available (e.g., the Shooting Action is planned along the road of a bicycle race or along a river or it is fixed), it can be also added by the director.

In each Shooting Action a set of Shooting Roles can be identified as part of its description. A Shooting Role is a role acted by one or more drones of the UAV team that has a specific editorial meaning. Figure 8 shows an example in which the three UAVs of the formation have three distinct roles (top view, right follower, left follower). Each role may have specific parameters. These different parameters are modeled with a specific distinct class Shooting Role Description. Shooting Action Description and Shooting Roles are linked through an association class carrying information about the composition of the roles output videos. This information, when present, represents the way in which the video streams of each of the drones playing a role in the Shooting Action should be handled from the compositional point of view during the Shooting Action. This includes both spatial and temporal information. The spatial information provides information about how to compose the multiple sources in a single picture, while the temporal information about when individual video sources are to be considered in the composition (i.e., when switching between one source and the other has to take place).

Each Shooting Role has a specific Shooting Target, which can be identified either by direct radio link (e.g., via RFID technology), or via visual examples (i.e., supported by some visual tracking/identification technology), or by role (e.g., "leader of a breakaway") or by trajectory. In the latter case, when the RT is considered as the target of a Shooting Role, the target trajectory coincides with the RT trajectory. It is important to stress that the RT is not necessarily a shooting target. For example, in a football match the RT can be the central point of the goal line of the guest team, but the Shooting Action associated to the occurrence of a goal can have the scorer of the goal, the goalkeeper and the audience as its target. Each Shooting Role can have its own Tracking Speed Variation as well (e.g., expressed as parametric functions of some geometrical parameters of the formation) which vectorially sums up with the possible formation Tracking Speed Variation. A RT can also be fixed in space (e.g., a certain fixed landmark), in which case the RT trajectory is a single point. All these data, where relevant, are input during the Mission Preproduction process.

B. Mission Production

The Mission Production process is made up of two temporally consecutive sub-processes: Mission Planning and Mission Execution. The latter depends on the first one. Mission Planning is the phase where all aspects related to a Shooting Mission, i.e., Shooting Action Sequences, Actors, Roles, Reference Targets, UAV formations, etc., can be defined and configured through the mission planning user interface. The Mission Planning process is broken down into a Mission Configuration process, involving GUI windows where the Mission Plan parameters are given as input to the system, and a Mission Simulation process, that simulates possible flight path trajectories which could be executed by the UAV fleet for that requested input.

A fundamental sub-process of Mission Planning is Mission Validation, i.e., the process by which an Editorial Shooting Mission is analyzed and checked in terms of feasibility. That is, depending on the introduced parameters (e.g., the number of available UAVs, the types of requested Shooting Actions etc.), the system decides to a priori approve or reject the

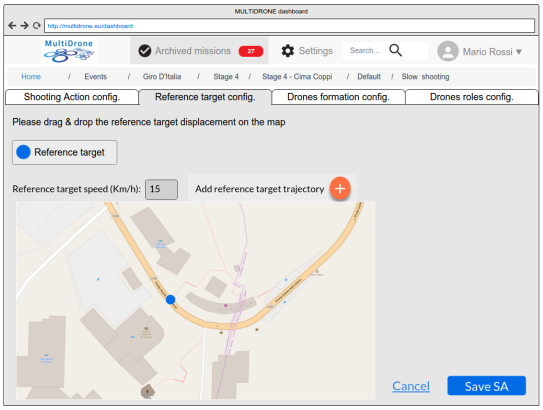

Fig. 9: Configuring the reference target in the Dashboard.

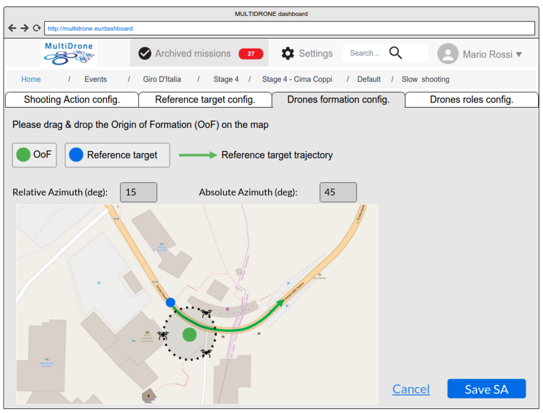

Fig. 10: Configuring drone formation in the Dashboard.

Shooting Mission. Therefore, only feasible Shooting Missions are allowed to be managed during Mission Execution. During Mission Planning, several Mission Missions could be defined, as main or backup plans.

Mission Execution is broken down into two distinct sub-processes: Mission Management and Shooting Management. Mission Management is the process by which the director takes final decisions about which among the several Shooting Missions associated to an Event and which among the planned Shooting Actions composing the selected Shooting Mission have to be executed. Shooting Management is performed by the cameraman who acts on camera parameters to optimize the quality of each drone's shooting. The director contributes to the Mission Planning and Mission Management processes. The cameramen are involved in the Shooting Management process. Example prototype GUI windows for managing target and fleet status parameters are shown in Figures 9, 10 and 11.

VII. Conclusions

Since the emerging field of autonomous UAV shooting is expected to revolutionize media production, this paper presents an overview of the area and proposes a taxonomy of cinematographic building blocks, in the context of filming outdoor

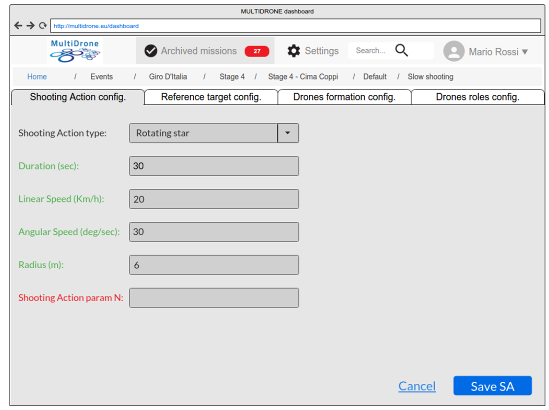

Fig. 11: Configuring a Shooting Action in the Dashboard.

events with moving targets. This taxonomy is then extended to formulate the concept of multiple-UAV cinematography, having important advantages over the current state-of-the-art. The above contributions are deemed as necessary for further progress in autonomous UAV shooting, e.g., by exploiting specific UAV shot type properties for optimizing energy consumption and scene coverage. Exploiting the proposed taxonomy, novel media production tools for managing a partially autonomous multiple-UAV fleet from the director's point of view are presented.

VIII. Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE).

References

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] NBC NEWS, "New drone video captures scale of Haiti hurricane damage," *TUGboat*, 2016.

[10] A. Charlton, "Sochi drone shooting olympic TV, not terrorists," *Associated Press*, 2014.

[11] K. Gallagher, "How drones powered Rio's olympic coverage," *The Simulyze Blog*, 2016.

[12] N. Joubert, J.L. E, D.B. Goldman, F. Berthouzoz, M. Roberts, J.A. Landay, and P. Hanrahan, "Towards a drone cinematographer: Guiding quadrotor cameras using visual composition principles," *ArXiv e-prints*, 2016.

[13] N. Joubert, M. Roberts, A. Truong, F. Berthouzoz, and P. Hanrahan, "An interactive tool for designing quadrotor camera shots," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, pp. 238, 2015.

[14] Amber Garage, "Skywand," https://skywand.com/.

[15] FreeSkies, "FreeSkies CoPilot," http://freeskies.co/.

[16] M. Roberts and P. Hanrahan, "Generating dynamically feasible trajectories for quadrotor cameras," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 61, 2016.

[17] C. Gebhardt, B. Hepp, T. Nägeli, S. Stevšić, and O. Hilliges, "Airways: Optimization-based planning of quadrotor trajectories according to high-level user goals," in *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 2016.

[18] A. Saeed, A. Abdelkader, M. Khan, A. Neishaboori, K.A Harras, and Amr. Mohamed, "On realistic target coverage by autonomous drones," *ArXiv e-prints*, 2017.

[19] T. Nägeli, L. Meier, A. Domahidi, J. Alonso-Mora, and O. Hilliges, "Real-time planning for automated multi-view drone cinematography," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 132:1–132:10, 2017.

[20] T. Nägeli, J. Alonso-Mora, A. Domahidi, D. Rus, and O. Hilliges, "Real-time motion planning for aerial videography with dynamic obstacle avoidance and viewpoint optimization," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1696–1703, 2017.

[21] FFD, "Nouveau guide DGAC "Activits Particulires" v1.2 (10 janvier 2017) Fdration Franaise de Drone (FFD)," *Fdration Franaise de Drone (FFD)*, Mar. 2017.

[22] "Verordnung zur Regelung des Betriebs von unbemannten Fluggerten," *Bundesgesetzblatt Teil I*, , no. 17, pp. 683, Apr. 2017.

[23] ANACOM, "ANACOM - Regulamento n. 1093/2016, de 14 de dezembro," .

[24] "Drones-Trabajos areos-Compaas o empresas-AESA-Agencia Estatal de Seguridad Area - Ministerio de Fomento," .

[25] "The Air Navigation Order 2016," .

[26] L. Yang, J. Qi, J. Xiao, and X. Yong, "A literature review of UAV 3d path planning," *Proceeding of the 11th World Congress on Intelligent Control and Automation*, pp. 2376–2381, June 2014.

[27] Q. Galvane, C. Lino, M. Christie, J. Fleureau, F. Servant, F. Tariolle, P. Guillotel, et al., "Directing cinematographic drones," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 3, pp. 34, 2018.

[28] K. Xie, H. Yang, S. Huang, D. Lischinski, M. Christie, K. Xu, M. Gong, D. Cohen-Or, and H. Huang, "Creating and chaining camera moves for qadrotor videography," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 88, 2018.

[29] R. Yuan, M. Liu, M. Hui, Y. Zhao, L. Dong, L. Kong, and Z. Cai, "Real-time video stitching based on ORB features and SVM," in *Applications of Digital Image Processing XLI*. International Society for Optics and Photonics, 2018.

[30] C. Xie, X. Zhang, H. Yang, L. Chen, and Z. Gao, "Video stitching based on optical flow," in *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2018.

[31] Q. Liu, X. Su, L. Zhang, and H. Huang, "Panoramic video stitching of dual cameras based on spatio-temporal seam optimization," *Multimedia Tools and Applications*, pp. 1–18, 2018.

[32] Y. Lu, B. Liu, W. Li, and N. Yu, "Fast video stitching for aerially captured HD videos," *International Journal of Image and Graphics*, vol. 18, no. 03, pp. 1850018, 2018.

[33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[34] B. Brown, *Cinematography: theory and practice: image making for cinematographers and directors*, Taylor & Francis, 2016.

[35] E. Cheng, *Aerial Photography and Videography Using Drones*, Peachpit Press, 2015.

[36] C. Smith, *The Photographer's Guide to Drones*, Rocky Nook, Inc., 2016.

[37] I. Mademlis, A. Iosifidis, A. Tefas, N. Nikolaidis, and I. Pitas, "Stereoscopic video description for human action recognition," in *Proceedings of the IEEE Series of Symposiums on Computational Intelligence (SSCI)*, 2014.

[38] I. Mademlis, A. Iosifidis, A. Tefas, N. Nikolaidis, and I. Pitas, "Exploiting stereoscopic disparity for augmenting human activity recognition performance," *Multimedia Tools and Applications*, vol. 75, no. 19, pp. 11641–11660, 2016.

[39] V. Mygdalis, A. Iosifidis, A. Tefas, and I. Pitas, "Graph embedded one-class classifiers for media data classification," *Pattern Recognition*, vol. 60, pp. 585–595, 2016.

[40] G. Varol, I. Laptev, and C. Schmid, "Long-term temporal convolutions for action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1510–1517, 2018.

[41] V. Mygdalis, A. Iosifidis, A. Tefas, and I. Pitas, "Semi-supervised subclass support vector data description for image and video classification," *Neurocomputing*, vol. 278, pp. 51–61, 2018.

[42] V. Mygdalis, A. Tefas, and I. Pitas, "Exploiting multiplex data relationships in support vector machines," *Pattern Recognition*, vol. 85, pp. 70–77, 2019.

[43] 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.

[44] 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.

[45] I. Mademlis, A. Tefas, N. Nikolaidis, and I. Pitas, "Compact video description and representation for automated summarization of human activities," in *Proceedings of the INNS Conference on Big Data*. Springer, 2016.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

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