Autonomous UAV Cinematography

Ioannis Pitas
Ioannis Mademlis
pitas@csd.auth.gr
imademlis@csd.auth.gr
Department of Informatics, Aristotle University of Thessaloniki
Thessaloniki, Greece

ABSTRACT

The use of camera-equipped Unmanned Aerial Vehicles (UAVs, or "drones") for professional media production is already an exciting commercial reality. Currently available consumer UAVs for cinematography applications are equipped with high-end cameras and a degree of cognitive autonomy relying on artificial intelligence (AI). Current research promises to further exploit the potential of autonomous functionalities in the immediate future, resulting in portable flying robotic cameras with advanced intelligence concerning autonomous landing, subject detection/tracking, cinematic shot execution, 3D localization and environmental mapping, as well as autonomous obstacle avoidance combined with on-line motion re-planning. Disciplines driving this progress are computer vision, machine/deep learning and aerial robotics. This Tutorial emphasizes the definition and formalization of UAV cinematography aesthetic components, as well as the use of robotic planning/control methods for autonomously capturing them on footage, without the need for manual tele-operation. Additionally, it focuses on state-of-the-art Imitation Learning and Deep Reinforcement Learning approaches for automated UAV/camera control, path planning and cinematography planning, in the general context of "flying & filming".

CCS CONCEPTS

• **Computing methodologies → Neural networks**; *Reinforcement learning*; • **Applied computing → Media arts**; • **Computer systems organization → Robotics**.

KEYWORDS

autonomous drones; UAV cinematography; media production; artificial intelligence; intelligent shooting

ACM Reference Format:

Ioannis Pitas and Ioannis Mademlis. 2022. Autonomous UAV Cinematography. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3503161.3546970

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10.
https://doi.org/10.1145/3503161.3546970

1 INTRODUCTION

Employing camera-equipped Unmanned Aerial Vehicles (UAVs, or "drones") for professional media production has become mainstream during the past decade. The benefits include easy capture of stunning aerial footage, rapid deployment, access to difficult-to-reach places and innovative visual effects. However, efficient audiovisual coverage of a large area with multiple shooting targets/subjects requires a fleet of multiple UAVs, raising issues related to the cumbersome logistics of manual operation [14][15][13]. These problems can be solved by on-board cognitive autonomy functionalities, permitting automated capture of cinematographic footage with a reduced number of human operators.

Thus, the exciting emerging area of highly autonomous UAVs for media production is gradually coalescing under the common theme of drone intelligence oriented towards capturing aesthetically pleasing footage in dynamic environments. Lying at the crossroad of aerial robotics, aerial cinematography, machine learning and computer vision, the emerging field of autonomous UAV cinematography attempts to develop robust robotic UAV platforms for intelligent filming [7][19][18][21], that require minimal human intervention while ensuring safe operation and obeying artistic guidelines.

The main UAV cinematography concepts are the *Camera Motion Type* (CMT), describing the desired UAV trajectory relative to a (still or moving) physical target being filmed, and the *Framing Shot Type* (FST), referring to the percentage of video frame width/height covered by the target's image. Examples of CMTs include Orbit, Fly-By, Fly-Over, or Chase, while examples of FSTs include Close-Up, Medium Close-Up, Long Shot, etc. [16][11][10][9]. In professional UAV filming scenarios, of either scripted or live events, the director initially comes up with a *cinematography plan* in the form of a high-level shooting mission description. The shooting mission mainly consists in a sequence of target, CMT and FST assignments to available UAV-mounted cameras over time. Subsequently, during filming, the plan is implemented so as to capture the desired footage. Traditionally this is done manually, while in autonomous UAV cinematography it is performed by automatically translating the plan to a sequence of low-level UAV/camera actions.

Fully autonomous UAV filming systems require advanced artificial intelligence (AI) algorithms, ideally able to be executed on-drone using specialized computational hardware. Nowadays, Visual Simultaneous Localization and Mapping (SLAM), Deep Neural Networks (DNNs) and trajectory optimization methods are among the core AI algorithms employed for enabling autonomous UAVs in the media production domain. Jointly, they can provide the required cognitive infrastructure for geometrically and semantically mapping the flight environment using video inputs, as well as for planning the optimal vehicle path within this environment, in order to achieve aesthetic objectives and film the proper footage. A number of commercial cinematography-oriented UAVs exploiting such algorithms have emerged as market leaders.

This Tutorial presents the state-of-the-art and current research in this interdisciplinary subject, especially focusing on Deep Neural Networks. It concerns a very timely topic with high industrial potential, given that fully/semi-automated drones are slowly emerging as a viable alternative to manually tele-operated ones, thanks to recent advances in robotics and AI. The variety of the underlying functionalities that enable such autonomy make this a highly interdisciplinary topic of exceptional interest to many types of multimedia experts and technology providers.

2 TUTORIAL STRUCTURE

This Tutorial is composed of three lectures, which are described below.

2.1 Lecture 1: Introduction to UAV cinematography mission planning and control

This lecture (by Prof. Ioannis Pitas) presents a formalized overview of a multiple-UAV cinematography task. First, an audiovisual shooting mission is formally defined. The introduced audiovisual shooting definitions are encoded in mission planning commands, i.e., navigation and shooting action vocabulary, and their corresponding parameters. The UAV mission commands, as well as the hardware/software architecture required for manual/autonomous mission execution are described. The software infrastructure includes the planning modules, that assign, monitor and schedule different behaviors/tasks to the UAV fleet according to director and environmental requirements, and the control modules, which execute the planning mission by translating high-level commands to into desired UAV + camera configurations, producing commands for autopilot, camera and gimbal of the UAV fleet [17] [3] [2] [1].

2.2 Lecture 2: UAV cinematography

This lecture (by Prof. Ioannis Pitas) overviews various aesthetic and practical aspects of UAV cinematography for TV, cinema and media production. A vocabulary suitable for UAV shooting mission description is defined as well, containing various relevant CMTs, e.g., chasing and orbiting, as well as FSTs (e.g., close-up, long shot) [13] [16]. Additionally, UAV mission simulator architectures are presented, such as AirSim and Gazebo. Their use in various scenarios is detailed: pilot training, synthetic data generation for Deep Neural Network training [8] [22], as well as drone cinematography parameter selection based on drone output video quality [12]. Finally, examples of UAV mission simulations are presented.

2.3 Lecture 3: Deep learning for autonomous UAV cinematography

This lecture (by Dr. Ioannis Mademlis) overviews the current state-of-the-art in Deep Neural Networks for cinematography applications using autonomous UAVs. The use of object detectors/trackers for visual subject/target identification is first detailed. This is in fact a mature technology which is already being employed by commercial UAVs. Subsequently, emphasis will be placed on imitation (IL) and deep reinforcement learning (DRL) approaches for automated motion/camera control, path planning and cinematography planning, with the aim of capturing aesthetically pleasing footage. Such approaches are currently research prototypes. IL/DRL methods may rely on advanced robotics simulators or professional video datasets where the training of neural agents takes place; successful generalization to unknown real-world environments during actual deployment is not guaranteed. This disparity between training characteristics and test conditions in the real world is known as the "reality gap".

Examples of DRL research prototypes for automating UAV cinematography can be found in [20] and [4]. The first method attempts not to autonomously execute specific CMTs, but simply to capture frontal close-up shots of human targets. In the second method, the task does not involve low-level UAV control for CMT execution, but rather fully autonomous high-level, on-line cinematography planning that draws from a limited palette of rudimentary CMTs. Thus, the trained agent simply selects dynamically at each time instance the best current CMT, among the few supported ones. In general, DRL for autonomous UAV cinematography is an open area of research.

Non-RL deep neural architectures for learning UAV cinematography in a supervised manner have also been proposed. For example, [5] uses real, professional UAV video footage, where a Sequence-to-Sequence Convolutional Long Short-Term Memory network (ConvLSTM) learns to regress desired dense optical flows (OFs) from input dense OFs and CNN-derived semantic features extracted by each video frame. The output OFs are analytically translated into desired UAV/camera motion commands. Finally, the predicted motion commands over N immediate future time-steps are assembled into feasible UAV trajectories by exploiting any trajectory planning method. An alternative approach [6] employs an LSTM-based encoder-decoder neural architecture for temporal Sequence-to-Sequence prediction tasks, which learns to regress the desired next filming target appearance, at each time instance, from purely visual inputs under mild assumptions. Training also takes place on professional UAV footage, rendering this a variant of imitation learning.

3 CONCLUSIONS

This Tutorial presents high-level descriptions of autonomous UAV cinematography missions, details their practical aspects and discusses a formalized aesthetic vocabulary of UAV cinematography for professional media production. Subsequently, it focuses on the current state-of-the-art concerning deep learning algorithms employed for autonomous UAV cinematography. This is a very timely research area with high industry importance, where although significant progress has been made during the past few years, radical advances are expected in the near future.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911 (AI4Media).

REFERENCES

[1] A. Alcantara, J. Capitan, R. Cunha, and A. Ollero. 2021. Optimal trajectory planning for cinematography with multiple Unmanned Aerial Vehicles. *Robotics and Autonomous Systems* 140 (2021), 103778.

[2] A. Alcantara, J. Capitan, A. Torres-Gonzalez, R. Cunha, and A. Ollero. 2020. Autonomous Execution of Cinematographic Shots with Multiple Drones. *IEEE Access* (2020), 201300–201316.

[3] L. E. Caraballo, A. Montes-Romero, J. M. Diaz-Bañez, J. Capitan, A. Torres-Gonzalez, and A. Ollero. 2020. Autonomous Planning for Multiple Aerial Cinematographers. In *International Conference on Intelligent Robots and Systems (IROS)*.

[4] M. Gschwindt, E. Camci, R. Bonatti, W. Wang, E. Kayacan, and S. Scherer. 2019. Can a Robot Become a Movie Director? Learning Artistic Principles for Aerial Cinematography. *arXiv preprint arXiv:1904.02579* (2019).

[5] C. Huang, C.-E. Lin, Z. Yang, Y. Kong, P. Chen, X. Yang, and K.-T. Cheng. 2019. Learning to Film from Professional Human Motion Videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[6] C. Huang, Z. Yang, Y. Kong, P. Chen, X. Yang, and K.-T. T. Cheng. 2019. Learning to Capture a Film-Look Video with a Camera Drone. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.

[7] N. Joubert, D. B. Goldman, F. Berthouzoz, M. Roberts, J. A. Landay, and P. Hanrahan. 2016. Towards a Drone Cinematographer: Guiding Quadrotor Cameras using Visual Composition Principles. *arXiv preprint arXiv:1610.01691* (2016).

[8] E. Kakaletsis, C. Symeonidis, M. Tzelepi, I. Mademlis, A. Tefas, N. Nikolaidis, and I. Pitas. 2021. Computer Vision for Autonomous UAV Flight Safety: An Overview and a Vision-based Safe Landing Pipeline Example. *ACM Computing Surveys (CSUR)* 54, 9 (2021), 1–37.

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

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

[11] I. Karakostas, I. Mademlis, N. Nikolaidis, and I. Pitas. 2020. Shot Type Constraints in UAV Cinematography For Autonomous Target Tracking. *Information Sciences* 506 (2020), 273–294.

[12] I. Mademlis et al. 2022. A multiple-UAV architecture for autonomous media production. *Springer Multimedia Tools and Applications* (2022), 1–30.

[13] I. Mademlis, V. Mygdalis, N. Nikolaidis, M. Montagnuolo, F. Negro, A. Messina, and I. Pitas. 2019. High-level multiple-UAV cinematography tools for covering outdoor events. *IEEE Transactions on Broadcasting* 65, 3 (2019), 627–635.

[14] I. Mademlis, V. Mygdalis, N. Nikolaidis, and I. Pitas. 2018. Challenges in autonomous UAV cinematography: an overview. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*.

[15] I. Mademlis, N. Nikolaidis, A. Tefas, I. Pitas, T. Wagner, and A. Messina. 2018. Autonomous unmanned aerial vehicles filming in dynamic unstructured outdoor environments. *IEEE Signal Processing Magazine* 36 (2018), 147–153. Issue 1.

[16] I. Mademlis, N. Nikolaidis, A. Tefas, I. Pitas, T. Wagner, and A. Messina. 2019. Autonomous UAV Cinematography: A Tutorial and a Formalized Shot Type Taxonomy. *Comput. Surveys* 52, 5 (2019), 105.

[17] A. Montes-Romero, A. Torres-González, J. Capitán, M. Montagnuolo, S. Metta, F. Negro, A. Messina, and A. Ollero. 2020. Director Tools for Autonomous Media Production with a Team of Drones. *Applied Sciences* 10 (2020). Issue 4.

[18] T. Nägeli, J. Alonso-Mora, A. Domahidi, D. Rus, and O. Hilliges. 2017. Real-Time Motion Planning for Aerial Videography With Dynamic Obstacle Avoidance and Viewpoint Optimization. *IEEE Robotics and Automation Letters* 2, 3 (2017), 1696–1703.

[19] T. Nägeli, L. Meier, A. Domahidi, J. Alonso-Mora, and O. Hilliges. 2017. Real-time Planning for Automated Multi-view Drone Cinematography. *ACM Transactions on Graphics* 36, 4 (2017), 132:1–132:10.

[20] N. Passalis and A. Tefas. 2019. Deep reinforcement learning for controlling frontal person close-up shooting. *Neurocomputing* 335 (2019), 37–47.

[21] F. Patrona, I. Mademlis, A. Tefas, and I. Pitas. 2019. Computational UAV cinematography for intelligent shooting based on semantic visual analysis. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*.

[22] C. Symeonidis, P. Nousi, P. Tosidis, K. Tsampazis, N. Passalis, A. Tefas, and N. Nikolaidis. 2021. Efficient Realistic Data Generation Framework Leveraging Deep Learning-Based Human Digitization. In *Proceedings of the Engineering Applications of Neural Networks Conference*.