SHOT TYPE FEASIBILITY IN AUTONOMOUS UAV CINEMATOGRAPHY

Iason Karakostas, Ioannis Mademlis*, Nikos Nikolaidis and Ioannis Pitas*

Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece

ABSTRACT

Aerial cinematography relying on camera-equipped umanned aerial vehicles (UAVs), or drones, has revolutionized media production during the past years. Autonomous UAV functionalities are already being employed to a degree, in a manner structured mainly around visual target tracking. From a cinematographic point of view, the desired shot type (i.e., Close-Up, Long Shot, etc.) is the most important factor affecting the artistic result. Achieving a specific shot type depends on the target-to-camera distance and the camera focal length. However, the interaction between UAV/camera motion trajectory (e.g., Orbit, Chase, etc.) and the visual tracker requirements constrains the range of feasible shot types at each time instance. In this paper, which extends previous work, these constraints are explored for a number of standard UAV/camera motion types, UAV shot types are classified and rules regarding shot feasibility over time are analytically derived. The proposed rules are evaluated in a realistic UAV simulation environment and achieve high performance, indicating possible benefits from their integration into an intelligent shooting system.

Index Terms— UAV cinematography, shot type, autonomous drones, target tracking

1. INTRODUCTION

The combination of Unmanned Aerial Vehicles (UAVs or "drones") and professional cameras has recently revolutionalized aerial cinematography in media production applications. UAVs are portable, able to access narrow spaces, implement novel visual effects and capture intriguing shots, at a low cost and with easy deployment. In professional production scenarios, at least two persons are needed to act in coordination for manual camera and UAV operation. Autonomous functionalities can facilitate their work and reduce the challenges of fully manual filming. Such capabilities in current commercial drones (e.g., the DJI Phantom IV Pro, or the more recent Skydio R1) are structured mainly around visual detection, tracking and active physical following of a specific target being shot, while relying on machine learning and computer vision modules (e.g., [1] [2] [3] [4] [5]).

The typical procedure in media production applications involves a director pre-specifying a cinematography plan, consisting in a temporal sequence of target assignments, camera motion types, shot types and framing compositions. Both the camera motion type (equivalent to UAV motion type, in our case) and the shot type are relative to a target being filmed, while achieving the requested shot type depends on the target-to-camera distance and the (adjustable) camera focal length f. If visual target tracking is involved, the maximum permissible focal length is necessarily constrained for a 2D visual tracker to operate properly. This is because the location (in pixel coordinates) of the target's Region-of-Interest (ROI) should differ no more than a threshold between successive video frames/time instances. This requirement places a constraint on the maximum target speed and on the maximum camera focal length, since a given 3D target displacement in the scene corresponds to a greater 2D ROI displacement (in pixels) at a greater zoom level. Estimating the maximum allowable f at each given circumstance is of utmost importance in cinematography applications, since it affects the currently permissible shot types.

Such a study was recently performed, in [6], assuming central framing composition (i.e., the selected target is always visible at the center of the video frame) and known 3D world positions of the UAV and the target. Industry-standard target-following UAV/camera motion types were geometrically modelled and, based on this modelling, the maximum permissible camera focal length for avoiding visual target tracking failure was analytically determined in the general case, as well as for a specific example camera motion type. That work was motivated by conclusions reached in preliminary relevant papers [7] [8] [9] [10] [11].

This paper is a follow-up work. It extends [8] and [6], first by explicitely deriving maximum focal length constraints for all modelled camera motion types. These constraints are subsequently exploited by proposing simple rules for determining shot feasibility at each time instance. To achieve this, useful UAV shot types are classified according to a ROI-to-video-frame ratio criterion. The described rule set is then empirically evaluated in a realistic UAV simulation environment and shown to achieve a very high rate of correct predictions. Incorporating shot type permissibility rules into media pro-

*These two authors contributed equally and are joint first authors.

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE).

duction automation software, such as intelligent UAV shooting algorithms [12] [13] [14], is expected to greatly enhance the robustness of autonomous cinematographic drones.

As in [6], the main underlying assumption is that the autonomous UAV operates in a consistent, global, Cartesian 3D map, where both vehicle and target position/velocity vector estimates are constantly provided. This can be easily achieved by combining RTK-GPS receivers [15] on both the UAV and the target, on-drone IMUs [16] and visual target localization methods [17]. Finally, the shooting camera is assumed to be suspended from a *gimbal*, allowing rapid, arbitrary camera rotation and attached to a fixed position of the UAV frame. In summary, this is a realistic UAV deployment setting, similar to the one presented in [18].

2. UAV CINEMATOGRAPHY MODELLING AND MAXIMAL FOCAL LENGTH CONSTRAINTS

In order to examine UAV shot type feasibility under constraints derived from 2D visual tracking requirements, the UAV/camera motion types should be formalized and geometrically modelled. Thus, it is assumed that, given a frame rate F, time t proceeds in discrete steps of $\frac{1}{F}$ seconds. The position vectors of the target and the UAV are denoted as $\tilde{\mathbf{p}}_t = [\tilde{p}_{t1}, \tilde{p}_{t2}, \tilde{p}_{t3}]^T$ and $\tilde{\mathbf{x}}_t = [\tilde{x}_{t1}, \tilde{x}_{t2}, \tilde{x}_{t3}]^T$, while the velocity vectors as $\tilde{\mathbf{u}}_t$ and $\tilde{\mathbf{v}}_t$, respectively, in a known fixed, orthonormal, right-handed World-Coordinate-System (WCS) with axes $\tilde{\mathbf{i}}$, $\tilde{\mathbf{j}}$, $\tilde{\mathbf{k}}$. Axis $\tilde{\mathbf{k}}$ is vertical to a local tangent plane (or "ground plane"). Additionally, at each time instance, a current, orthonormal, right-handed target-centered coordinate system (TCS), $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ is defined. Its origin lies on the current target position, its $\mathbf{k}$-axis is vertical to the ground plane and its $\mathbf{i}$-axis is the $\mathcal{L}_2$-normalized projection of the current target velocity vector onto the ground plane. In both coordinate systems, the $\mathbf{ij}$-plane is parallel to the ground plane and the $\mathbf{k}$-component is called "altitude". Vectors expressed in TCS are denoted without the tilde symbol (e.g. $\mathbf{p}$, $\mathbf{x}$). The 3D scene point at which the camera looks at time instance t is denoted by $\mathbf{l}_t$, while $\mathbf{o}_t = \mathbf{l}_t - \mathbf{x}_t$ is the LookAt vector (both in TCS). Below, it is assumed that $\mathbf{l}_t = \mathbf{p}_t$ at all times (central framing composition).

By implementing a geometrically modelled UAV/camera motion type and knowing the exact 3D target position (ignoring practical limitations, such as maximum drone speed, wlog), an autonomous UAV would be able to actively track and follow the desired target. However, sensor noise in 3D target position measurements and the unpredictability of the actual current target velocity (it may deviate from the predicted one by the unknown vector $\tilde{\mathbf{q}}_t = [\tilde{q}_{t1}, \tilde{q}_{t2}, \tilde{q}_{t3}]^T$) perplex the issue. Thus, the target ROI at time $t' = t + 1$ has to be explicitly localized via 2D visual tracking (in pixel coordinates), so that it can be exploited for 3D target position $\tilde{\mathbf{p}}_{t'}$ estimation and/or for adjusting the framing composition. Based on the above and working in TCS, the following general constraint for maximal focal length was derived in [6]:

$$f_{max} = \frac{R_{max} d_{t'} s_x s_y | E_1 + F \parallel \mathbf{x}_{t'} \parallel^2 |}{\sqrt{(s_x q_{t3} d_{t'}^2 - s_x x_{t'3} E_2)^2 + s_y^2 E_3^2 \parallel \mathbf{x}_{t'} \parallel^2}}, \quad (1)$$

where

$$E_1 = -q_{t1} x_{t'1} - q_{t2} x_{t'2} - q_{t3} x_{t'3}, \quad (2)$$

$$E_2 = q_{t1} x_{t'1} + q_{t2} x_{t'2}, \quad (3)$$

$$E_3 = q_{t2} x_{t'1} - q_{t1} x_{t'2}. \quad (4)$$

$$d_{t'} = \sqrt{x_{t'1}^2 + x_{t'2}^2} \quad (5)$$

Here, s_x and s_y denote the physical dimensions of a pixel. Whenever $\tilde{\mathbf{q}}_t$ is a non-zero vector and, therefore, prediction of $\tilde{\mathbf{p}}_{t+1}$ fails, the results of 2D visual tracking and actual $\tilde{\mathbf{p}}_{t+1}$ estimation must be employed for updating the target velocity vector and, hopefully, achieving a better prediction during the next time instance. Given that tracker behavior varies per algorithm, we simply assume a maximum search radius R_{max} (in pixels) defining the video frame region within which the tracked object ROI of time instance $t+1$ must lie, relatively to the video frame center, in order to permit successful tracking.

Additionally, by utilizing the above notation, six industry-standard target-tracking UAV/camera motion types were formalized and geometrically modelled in [8] and [6]: "Lateral Tracking Shot" (LTS), "Vertical Tracking Shot" (VTS), "Fly-Over", "Fly-By", "Chase" and "Orbit". Based on this modelling, Eq. (1) was adapted for the specific example cases of ORBIT and LTS in [8] and [6], respectively.

3. MAXIMUM FOCAL LENGTH IN SPECIFIC CAMERA MOTION TYPES

Here, we present the maximum focal length constraint adapted for VTS, FLYOVER, FLYBY and CHASE, for the first time. Their derivation follows from Eq. (1) and the descriptions of the various modelled camera motion types in [6]. These constraints will be employed in the next Section for the proposed shot type feasibility rule set.

Thus, in VTS, it holds that:

$$f_{max} = \frac{R_{max} F x_{t'3} s_x s_y}{\sqrt{s_y^2 q_{t1}^2 + s_x^2 q_{t2}^2}}. \quad (6)$$

Additionally, it holds that:

$$f_{max} = \frac{R_{max} d_{fb} s_x s_y | - E_{fb1} + F \parallel \mathbf{x}_{t+1} \parallel^2 |}{\sqrt{s_x^2 E_{fb1}^2 x_{t3}^2 + s_y^2 E_{fb2}^2 \parallel \mathbf{x}_{t+1} \parallel^2}}, \quad (7)$$

$$f_{max} = \frac{R_{max} d_{fo1} s_x s_y | - E_{fo1} + F \parallel \mathbf{x}_{t+1} \parallel^2 |}{\sqrt{s_x^2 E_{fo1}^2 x_{t3}^2 + s_y^2 E_{fo2}^2 \parallel \mathbf{x}_{t+1} \parallel^2}}, \quad (8)$$

for FLYBY and FLYOVER, respectively, where:

$$d_{fb} = \sqrt{(\frac{-x_{01}}{K} t + x_{01})^2 + x_{02}^2}, \quad (9)$$

$$E_{fb1} = q_{t1}(\frac{-x_{01}}{K}t + x_{01}) + q_{t2}x_{t2}, \quad (10)$$

$$E_{fb2} = q_{t2}(\frac{-x_{01}}{K}t + x_{01}) - q_{t1}x_{t2}, \quad (11)$$

$$d_{fo} = |(\frac{-x_{01}}{K}t + x_{01})|, \quad (12)$$

$$E_{fo1} = q_{t1}(\frac{-x_{01}}{K}t + x_{01}), \quad (13)$$

$$E_{fo2} = (q_{t2}(\frac{-x_{01}}{K}t + x_{01})). \quad (14)$$

For CHASE, it holds that:

$$f_{max} = \frac{R_{max}s_x s_y \phi_c| - F\phi_c^2 + x_{t1}q_{t1}|}{x_{t1}\sqrt{s_y^2\phi_c^2 q_{t2}^2 + s_x^2 x_{t3}^2 q_{t1}^2}}, \quad (15)$$

where

$$\phi_c = \sqrt{x_{t1}^2 + x_{t3}^2}. \quad (16)$$

4. SHOT TYPE FEASIBILITY

The desired shot type is mainly determined by the ratio of the ROI height to the video frame height (in pixels) c_s, a quantity that we refer to as "target video frame coverage". Below, we present a proposed list of common UAV shot types, adopted from traditional ground and aerial cinematography [19] [20] [21]. This classification scheme was reached after extensive visual inspection of professional UAV footage: *Extreme Long Shot* (ELS) is defined by $c_s < 5\%$, *Very Long Shot* (VLS) by $c_s \in [5, 20\%]$, *Long Shot* (LS) by $c_s \in [20, 40\%]$, *Medium Shot* (MS) by $c_s \in [40, 60\%]$, *Medium Close-Up* (MCU) by $c_s \in [60, 75\%]$, and *Close-Up* (CU) by $c_s > 75\%$.

Shot types and UAV/camera motion types are combined in the cinematography plan to produce the desired footage. In order to determine, at each time instance during shot execution, whether the pre-specified shot and camera motion type combination is currently feasible, the appropriate focal length f_s leading to the desired target video frame coverage must be calculated. f_s remains unchanged for camera motion types that retain a constant distance between the target and the camera (i.e., CHASE, VTS, LTS and ORBIT), otherwise it varies. In the latter case, keeping the coverage constant throughout the camera motion, by properly modifying f_s, will result in the cinematographic "dolly zoom" effect [19]. In general, a shot type can be achieved without risking 2D visual tracking failure, if the following relation holds:

$$f_s \leq f_{max} \quad (17)$$

In order to calculate the appropriate f_s for achieving the shot types described in Section 2 with respect to the desired UAV/camera motion type, we model the target as a sphere, with its center located at the TCS point $[0, 0, 0]^T$ and having constant radius R_t. Simple sphere-modelling allows us to consider its image on the video frame as a circle, with no perspective distortion when $\mathbf{l}_t = [0, 0, 0]^T$.

Below, the deviation vector $\mathbf{q}_t$ is assumed to be equal to $[0, 0, 0]^T$ for the desired f_s calculations. Thus, no target motion deviations are taken into consideration, since they do not significantly affect the resulting video frame coverage.

Determining the video frame coverage for every UAV motion type would normally include projecting the target sphere onto the video frame, finding the corresponding radius of the projected circle and computing the resulting coverage. This requires a search for the radius of the projected circle. The parameters determining the video frame coverage are the distance between UAV/camera and target, the camera focal length f and the physical target dimensions. Thus, wlog, instead of directly projecting the target onto the current image plane, we determine the video frame coverage as if the UAV/camera was positioned exactly above the target in an altitude equal to the actual distance between them. It is then trivial to find a 3D point being projected on the target image circle. The latter's radius is the distance between the projection of the above 3D point and the principal point. This projection can be obtained by utilizing the camera projection equations [22]:

$$x_d(t+1) = o_x - \frac{f}{s_x}\frac{\mathbf{r}_1^T(\mathbf{p}_{t+1} - \mathbf{x}_{t+1})}{\mathbf{r}_3^T(\mathbf{p}_{t+1} - \mathbf{x}_{t+1})}, \quad (18)$$

$$y_d(t+1) = o_y - \frac{f}{s_y}\frac{\mathbf{r}_2^T(\mathbf{p}_{t+1} - \mathbf{x}_{t+1})}{\mathbf{r}_3^T(\mathbf{p}_{t+1} - \mathbf{x}_{t+1})}, \quad (19)$$

in pixel coordinates, where $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ are the rows of the rotation matrix $\mathbf{R}$ that orients the camera gimbal according to the LookAt vector and o_x, o_y define the image center in pixel coordinates. The corresponding continuous coordinates of x_{im} and y_{im} on the image sensor are given by:

$$x_{im} = x_d s_x, \quad y_{im} = y_d s_y. \quad (20)$$

Thus, the video frame coverage percentage for the circular target ROI is given by:

$$c_s = \frac{2R_{im}}{H s_y}, \quad R_{im} = \sqrt{x_{im}^2 + y_{im}^2}. \quad (21)$$

where H is the height of the video frame in pixels and s_y the physical height of one pixel.

The above equations can be further simplified by defining R_{im} as the perspective projection of $\mathbf{p}_r = [R_t, 0, 0]^T$ (in TCS), where R_t is target radius, and by positioning the UAV/camera at $\mathbf{x}' = \mathbf{x}_{t+1} = [0, 0, z_d]^T$ where $z_d = \sqrt{x_{t'1}^2 + x_{t'2}^2 + x_{t'3}^2}$ is the distance between the target and the camera. Then, $y_{im} = 0$, thus, $R_{im} = x_{im}$ and:

$$x_{im} = \frac{1}{2}c_s H s_y \quad (22)$$

By utilizing Eqs. (18) and (20), and setting $o_x = 0$:

$$x_{im} = -f_s \frac{\mathbf{r}_1(\mathbf{p}_r - \mathbf{x}')}{\mathbf{r}_3(\mathbf{p}_r - \mathbf{x}')}. \quad (23)$$

Table 1. Mean evaluation results for the proposed shot feasibility rules over all motion types.

Shot type	F-measure	Precision	Recall
LS	0.992	0.997	0.987
MCU	0.943	0.900	0.991
CU	0.920	0.856	0.996
Mean	**0.946**	**0.909**	**0.990**

The rotation matrix in this case is described by:

$$\mathbf{R} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{24}$$

and the appropriate focal length can be obtained by:

$$f_s = \frac{c_s H s_y z_d}{2R_t}. \tag{25}$$

By exploiting Eqs. (17), (25) and the various camera motion type-specific variants of Eq. (1), the feasibility of a shot and motion type combination can be determined on-line at each time instance.

5. EMPIRICAL EVALUATION

In order to evaluate the presented shot feasibility rules under actual media production conditions, a realistic simulation was developed. To this end, AirSim [23] was employed, i.e., an open source, highly realistic UAV simulation environment (based on the Unreal 4 real-time 3D graphics engine). The setup of the realistic simulation involves a moving cyclist and a UAV equipped with a cinematographic camera which follows the motion types described in Section 2 and the focal length of the camera is adjusted so as to implement the LS, MCU, and CU shot types.

The UAV and target 3D positions can be almost instantly obtained using simulated RTK-GPS sensors in this environment, thus no 2D visual target tracking is needed in principle; the target may always be properly framed by orienting the gimbal according to the corresponding 3D LookAt vector. However, the Gaussian noise in the position measurements and the unpredictable target motion (unpredictable with regard to its deviations from uniform linear motion, i.e., due to non-negligible velocity deviation vector $\mathbf{q}_t$ at time instance t), make 2D visual target tracking a necessity and, therefore, impose constraints on the feasible shot types.

At each time instance t, the previous noisy 3D position of the target (from $t-1$) was employed to estimate its velocity. Naively assuming that the target will momentarily follow a uniform linear motion, we estimate its 3D position at $t' = t+1$ and adjust UAV trajectory and gimbal orientation, so that the desired central composition framing is maintained. Then, at time instance t', we compare the 2D projection of the estimated 3D target position with the 2D projection of the ground-truth 3D target position. If the distance of the two ROI center points is above the R_{max} limit, ground-truth tracking failure is assumed. This is then compared with the predictions of Eq. (17), regarding the current shot's feasibility, given the noisy 3D positions of the target and the UAV, the calculated target velocity and the estimated target position on the next video frame. Thus, true/false positive/negative prediction labels are computed for each time instance.

The velocity deviation vector $\mathbf{q}_t$ in Eq. (1) is simply calculated as the difference between the estimated target velocity at time instance $t-1$ and the actual target velocity at time instance t (distorted by noise). Therefore, temporally localized constant target acceleration is implicitely assumed in shot feasibility determination. Such an assumption is too strong to guide target position estimation itself, thus uniform linear motion is simply considered in that case, as previously described. This is reasonable if no constraints about near-future target trajectory are provided (e.g., a 3D spline modelling the road where the cyclist drives). However, it is not too strong to underpin shot feasibility analysis; it is acceptable to be overpessimistic about shot feasibility, but significantly more undesirable to lose central framing composition and/or induce 2D visual tracker failure, due to erroneous next target position prediction. Obviously, a different choice for estimating $\mathbf{q}_t$ implies a different target velocity deviation handling policy, making our model highly flexible.

The mean precision, recall and F-measure of the proposed rules over all camera motion types are depicted in Table 1, per shot type. For the evaluation, R_{max} was set to 128 pixels, UAV-to-target distance (in the camera motion types where it remains constant) was set to 30 meters, while the target-modelling sphere radious was set to 1 meter.

6. CONCLUSIONS

In this paper, previous work modelling industry-standard, cinematographic target-tracking UAV motion types and deriving a maximum focal length constraint for avoiding 2D visual tracking failure, was extended. Here, common UAV shot types have been classified quantitavely and the maximum focal length constraint has been explicitely adapted to all modelled camera motion types. Then, rules regarding shot feasibility over time are analytically derived and successfully evaluated in a realistic UAV simulation environment, achieving high prediction performance. The proposed rule set can be readily integrated into UAV intelligent shooting frameworks for adaptive UAV cinematography planning. Additionally, learning to predict a more informed vector $\mathbf{q}_t$ from visual data, given that the proposed formulas rely on assumed UAV velocity deviation at each time instance t, as well as tighter integration with the 2D visual tracker itself, are promising research avenues.

7. REFERENCES

[1] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[2] D. Triantafyllidou, P. Nousi, and A. Tefas, "Lightweight two-stream convolutional face detection," in *Proceedings of EURASIP European Signal Processing Conference (EUSIPCO)*, 2017.

[3] 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.

[4] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking," in *Proceedings of European Conference on Computer Vision (ECCV)*. Springer, 2016.

[5] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High-performance visual tracking with siamese region proposal network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 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. Mademlis, I. Mygdalis, C. Raptopoulou, N. Nikolaidis, N. Heise, T. Koch, J. Grunfeld, T. Wagner, A. Messina, F. Negro, S. Metta, and I. Pitas, "Overview of drone cinematography for sports filming," in *European Conference on Visual Media Production (CVMP) (short)*, 2017.

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

[9] 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.

[10] I. Mademlis, N. Nikolaidis, A. Tefas, I. Pitas, T. Wagner, and A. Messina, "Autonomous UAV filming in dynamic unstructured outdoor environments," *IEEE Signal Processing Magazine*, 2018, accepted for publication.

[11] F. Patrona, I. Mademlis, A. Tefas, and I. Pitas, "Computational UAV Cinematography for intelligent A/V shooting based on semantic visual analysis," in *IEEE International Workshop on Applications of Computer Vision (WACV) (submitted)*, 2019.

[12] 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.

[13] N. Joubert, 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 preprint arXiv:1610.01691*, 2016.

[14] 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.

[15] M. S. Grewal, L. R. Weill, and A. P. Andrews, *Global Positioning Systems, inertial navigation, and integration*, John Wiley & Sons, 2007.

[16] R. Mur-Artal and J. D. Tardós, "Visual-inertial monocular SLAM with map reuse," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 796–803, 2017.

[17] M. Monda, C. Woolsey, and C. Reddy, "Ground target localization and tracking in a riverine environment from a UAV with a gimbaled camera," in *Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit*, 2007.

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

[19] B. Brown, *Cinematography: Theory and Practice: Image Making for Cinematographers and Directors*, Focal Press, 3rd edition, 2016.

[20] E. Cheng, *Aerial Photography and Videography Using Drones*, Peachpit Press, 2016.

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

[22] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.

[23] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "AirSim: High-fidelity visual and physical simulation for autonomous vehicles," in *Proceedings of the Field and Service Robotics Conference*, 2017.