UAV CINEMATOGRAPHY CONSTRAINTS IMPOSED BY VISUAL TARGET TRACKING

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

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

ABSTRACT

Camera-equipped drones have recently revolutionized aerial cinematography, allowing easy acquisition of impressive footage. Although they are currently manually operated, autonomous functionalities based on machine learning and computer vision are becoming popular. However, the emerging area of autonomous UAV filming has to face several challenges, especially when visually tracking fast and unpredictably moving targets. In the latter case, an important issue is how to determine the shot types that are achievable without risking failure of the 2D visual tracker. This paper studies the constraints imposed to cinematography decision-making during autonomous UAV shooting. It focuses on formalizing and geometrically modelling common target-following UAV motion types, in order to analytically determine the maximum permissible camera focal length (therefore, the range of feasible shot types) for avoiding visual target tracking failure.

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

1. INTRODUCTION

Vertical Take-off and Landing (VTOL) Unmanned Aerial Vehicles (UAVs or "drones") equipped with professional cameras have become an essential tool for a cinematographer during the past few years. They have been adopted in TV/movies production, newsgathering, advertisement and outdoor event coverage, since they can easily capture shots that would otherwise incur significantly higher production costs (e.g. the use of a helicopter or a crane). In many application scenarios there is a need for visual detection, tracking and active physical following of a specific target being filmed. In a typical manual operation scenario, this requires two different remote UAV operators (one for the camera and one for the drone) that must act in concert. Thus, a growing trend of increasing automation in drone functionalities has appeared, including machine learning and computer vision modules (e.g., [1] [2]), aiming to reduce the challenges arising from fully manual operation.

Typically, a director plans ahead a cinematography mission as a sequence of desired target assignments, shot types and UAV/camera motion types relative to the target. The shot type defines the zoom level and is regulated by the camera focal length f. In case autonomous visual target tracking is involved, the maximum 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 has not been previously performed, due to the lack of standardization and formalization in aesthetically meaningfull UAV/camera motion types. Incorporating shot type permissibility rules into media production automation software, such as intelligent UAV shooting algorithms [3] [4] [5], is expected to greatly enhance the robustness of autonomous drones. This paper, which extends preliminary work [6] [7], presents such a study. Industry-standard target-following UAV motion types have been geometrically modelled and the maximum permissible camera focal length for avoiding visual target tracking failure has been analytically determined.

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 employing Real Time Kinematic - Global Positioning System (RTK-GPS) receivers [8] on both the UAV and the target, as well as an on-drone Inertial Measurement Unit (IMU) [9]. However, purely visual target localization methods are also available [10]. Finally, the shooting camera is assumed to be suspended from a *gimbal* that allows rapid, arbitrary camera rotation around its yaw, pitch and roll axis. The gimbal is typically attached to a known, fixed position of the UAV vehicle frame.

2. UAV CINEMATOGRAPHY MODELLING

In cinematography, the shot type defines the percentage of the video frame covered by the target/subject being filmed

*These two authors contributed equally.

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

[11] and is mainly adjusted by the zoom level, therefore, by the camera focal length f. Camera motion type relative to the target is an equally important aspect of cinematography. After an extensive survey of professional UAV footage, we have identified five industry-standard target-tracking UAV/camera motion types. Below, they are formalized, and geometrically modelled. Additionally, the ORBIT UAV/camera motion type (not included here) has been modelled and examined in the preliminary study [7]. A more general derivation is presented in this paper, fully able to incorporate the case of ORBIT.

For reasons of simplicity, the employed modelling ignores the distinction between the drone and its mounted camera, since it is typically trivial to compute the 3D pose of the one given the other and gimbal feedback. It is assumed that, given a camera 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 $\tilde{\mathbf{u}}_t = [\tilde{u}_{t1}, \tilde{u}_{t2}, \tilde{u}_{t3}]^T$ and $\tilde{\mathbf{v}}_t = [\tilde{v}_{t1}, \tilde{v}_{t2}, \tilde{v}_{t3}]^T$, 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 case of a still target, the TCS $\mathbf{i}$-axis is defined as parallel to the projection of the vector $\tilde{\mathbf{p}}_0 - \tilde{\mathbf{x}}_0$ 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. $\mathbf{p}$, $\mathbf{x}$). Transforming between WCS and TCS is trivial, allowing for descriptions in the most suitable system for each motion type. 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, thus, the selected target point is visible at the center of the video frame.

Five target-tracking UAV/camera motion types, adopted from [6], have been modelled here.

1) *Lateral Tracking Shot (LTS)* [12] [13] and 2) *Vertical Tracking Shot (VTS)* are non-parametric camera motion types, where the camera gimbal does not rotate and the camera is locked on the moving target. In LTS, the camera axis is approximately perpendicular both to the target trajectory and to the WCS vertical axis vector $\tilde{\mathbf{k}}$, while the UAV flies sideways/in parallel to the target, matching its speed (if possible). In VTS, the camera axis is perpendicular to the target trajectory and the UAV flies exactly above the target, matching its speed (if possible). The base mathematical description for both is the following:

$$\tilde{\mathbf{v}}_t = \tilde{\mathbf{u}}_t \tag{1}$$

$$\tilde{\mathbf{o}}_t^T \tilde{\mathbf{u}}_t \approx 0, \quad \mathbf{x}_t = \mathbf{x}_{t-1}, \forall t. \tag{2}$$

Additionally, the following relations hold for LTS:

$$\mathbf{o}_t \times \mathbf{j} \approx \mathbf{0}, \quad x_{03} \approx 0, \tag{3}$$

while the following relations hold for VTS:

$$\mathbf{o}_t^T \mathbf{j} \approx 0, \quad x_{03} > 0. \tag{4}$$

3) *Fly-Over (FLYOVER)* and 4) *Fly-By (FLYBY)* [13]. They are parametric camera motion types, where the camera gimbal is rotating, so that the still or linearly moving target is always properly framed. The UAV intercepts the target from behind/from the front (and to the left/right, in the case of FLYBY), at a steady altitude (in TCS) with constant velocity, flies exactly above it/passes it by (for FLYOVER/FLYBY, respectively) and keeps on flying at a linear trajectory, with the camera still pointing at the receding target. The UAV and target trajectory projections onto the ground plane remain approximately parallel during shooting.

The common parameter that must be specified is K, i.e., the time (in seconds) until UAV is located exactly above the target (for FLYOVER), or until the distance between the target and the UAV is minimized (for FLYBY). Additionally, d, i.e., the length of the projection of that minimal distance vector onto the ground plane, must be specified for FLYBY.

The base mathematical description common to both camera motion types is the following:

$$\mathbf{v}_0 = [\frac{u_{01}K - x_{01}}{K}, 0, u_{03}]^T \tag{5}$$

$$\tilde{\mathbf{v}}_t = \tilde{\mathbf{v}}_{t-1}, \quad \tilde{\mathbf{u}}_t = \tilde{\mathbf{u}}_{t-1}, \forall t \tag{6}$$

$$\tilde{\mathbf{u}}_t = \tilde{\mathbf{u}}_{t-1}, \forall t \tag{7}$$

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{x}}_0 + \frac{t}{KT}(\tilde{\mathbf{x}}_{KT} - \tilde{\mathbf{x}}_0) \tag{8}$$

$$[\tilde{u}_{t1}, \tilde{u}_{t2}, 0]^T \times [\tilde{v}_{t1}, \tilde{v}_{t2}, 0]^T \approx \mathbf{0} \tag{9}$$

$$\mathbf{l}_t = \mathbf{p}_t \quad t \in [0, 2KT] \tag{10}$$

Additionally, the following hold for FLYOVER:

$$\tilde{\mathbf{x}}_{KT} = [\tilde{p}_{01} + \tilde{u}_{01}K, \tilde{p}_{02} + \tilde{u}_{02}K, \tilde{x}_{03} + \tilde{u}_{03}K]^T \tag{11}$$

$$x_{t2} \approx 0, \quad \mathbf{x}_t^T \mathbf{j} \approx 0, \forall t \tag{12}$$

and the following hold for FLYBY:

$$|x_{02}| = d > 0, \quad x_{t2} = x_{02}, \forall t \tag{13}$$

$$\mathbf{x}_{KT} = [0, x_{02}, x_{03}]^T \tag{14}$$

5) *Chase/Follow Shot (CHASE)* is a non-parametric camera motion type, where the camera gimbal does not rotate and the camera always points at the target [13]. The UAV follows/leads the target from behind/from the front, at a linear trajectory, constant distance while matching its speed, if possible. $\tilde{\mathbf{p}}_t$ refers to a varying target position in WCS. The mathematical description is the following:

$$\tilde{\mathbf{v}}_t \approx \tilde{\mathbf{u}}_t, \quad \mathbf{l}_t = \mathbf{p}_t \tag{15}$$

$$x_{t2} = x_{02} \approx 0, \quad \mathbf{x}_t = \mathbf{x}_{t-1}, \forall t \tag{16}$$

3. MAXIMUM FOCAL LENGTH CONSTRAINTS

In general, 2D visual tracking algorithms assume that the location of the target ROI center (in pixel coordinates) varies no more than a threshold between successive video frames. Different zoom levels, corresponding to different selected shot types and, therefore, different focal lengths, map a specific 3D target displacement to different 2D ROI displacements. Thus, it is important to determine the constraints on maximum focal length f imposed by the 2D visual tracker, since it affects the permissible shot framing types at each time instance.

Without loss of generality, we always consider time instance $t = 0$ and, thus, examine an entire shooting session as a sequence of repeated transitions between the "first" $(t = 0)$ and the "second" video frame $(t + 1 = 1)$, by proper manipulation of initial values of the variables per motion type. We also assume that the target ROI center is meant to be fixed at the image center for all video frames (this is called *central composition* in cinematography). Target position $\tilde{\mathbf{p}}_t$ is initially known and $\tilde{\mathbf{p}}_{t+1}$ can be predicted using the estimated target velocity vector $\tilde{\mathbf{u}}_t$ (i.e. $\tilde{\mathbf{p}}_{t+1} = \tilde{\mathbf{p}}_t + \tilde{\mathbf{u}}_t \frac{1}{F}$). If the prediction is accurate, the target ROI indeed remains at the center of the $(t + 1)$-th video frame.

In contrast, if the actual current target motion differs from the predicted one by the unknown velocity deviation vector $\tilde{\mathbf{q}}_t = [q_{t1}, q_{t2}, q_{t3}]^T$, the target ROI center position at time $t+1$ has to be explicitly localized via 2D visual tracking (in pixel coordinates), so that 3D target position $\tilde{\mathbf{p}}_{t+1}$ can be estimated.

Given that tracker behaviour 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, so as to permit successful tracking. Then, the maximum focal length so that there is no target tracking failure can be found by assuming that the expected position of the target in TCS is always $\mathbf{p}_{t+1} = [0, 0, 0]^T$. Additionally, $d_t = \sqrt{x_{t1}^2 + x_{t2}^2}$ is the horizontal distance between the target and the UAV on the $\mathbf{ij}$-plane at time instance t.

Based on the above and the camera projection equations [14], the following hold:

$$x_d(t+1) = o_x - \frac{f}{s_x} \frac{\mathbf{r}_1^T(\mathbf{p}_{t+1} - \mathbf{x}_{t+1})}{\mathbf{r}_3^T(\mathbf{p}_{t+1} - \mathbf{x}_{t+1})}, \tag{17}$$

$$y_d(t+1) = o_y - \frac{f}{s_y} \frac{\mathbf{r}_2^T(\mathbf{p}_{t+1} - \mathbf{x}_{t+1})}{\mathbf{r}_3^T(\mathbf{p}_{t+1} - \mathbf{x}_{t+1})}, \tag{18}$$

where $x_d(t+1)$, $y_d(t+1)$ are the target center pixel coordinates at the time instance $(t + 1)$, o_x,o_y define the image center in pixel coordinates and s_x, s_y denote the pixel size (in mm) along the horizontal and vertical directions. $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ refer, respectively, to the first, second and third row of the rotation matrix $\mathbf{R}$ that orients the camera gimbal according to the LookAt vector.

In general, the coordinate transform matrix from TCS to the camera coordinate system can be found by two rotations and one translation of the unit TCS vectors. The first rotation required is around the TCS $\mathbf{k}$-axis and the second one around the TCS $\mathbf{j}$-axis. Thus, $\mathbf{R}$ can be described as [15]:

$$\mathbf{R} = \begin{bmatrix} cos(\theta_z)cos(\theta_y) & -sin(\theta_z) & cos(\theta_z)sin(\theta_y) \\ sin(\theta_z)cos(\theta_y) & cos(\theta_z) & sin(\theta_z)sin(\theta_y) \\ -sin(\theta_y) & 0 & cos(\theta_y) \end{bmatrix}, \tag{19}$$

where θ_z and θ_y are the appropriate angles of rotation for $\mathbf{R}_z$ and $\mathbf{R}_y$ respectively. However, in most of the motion types, given that $\mathbf{R}$ is an orthogonal change-of-basis matrix and that the UAV does not fly exactly above the target, it is easier to obtain the rows of $\mathbf{R}$ as follows. Since the camera axis points directly at the target, the unit vector of the $\mathbf{k}$-axis for the Camera Coordinate System, i.e. $\mathbf{r}_3$, can be obtained from $\mathbf{x}_{t+1}$ as follows:

$$\mathbf{r}_3 = \left(\frac{-\mathbf{x}_{t+1}}{\parallel \mathbf{x}_{t+1} \parallel} \right)^T. \tag{20}$$

For motion types where the UAV is not exactly above the target, $\mathbf{r}_1$ is the cross product of $\mathbf{r}_3$ with the unit vector $\mathbf{k}$:

$$\mathbf{r}_1' = \left(\mathbf{k} \times \frac{-\mathbf{x}_{t+1}}{\parallel \mathbf{x}_{t+1} \parallel} \right)^T, \tag{21}$$

$$\mathbf{r}_1 = \frac{\mathbf{r}_1'}{\parallel \mathbf{r}_1' \parallel}. \tag{22}$$

Thus, $\mathbf{r}_2$ is given by the cross product $\mathbf{r}_3 \times \mathbf{r}_1$:

$$\mathbf{r}_2' = \left(\frac{-\mathbf{x}_{t+1}}{\parallel \mathbf{x}_{t+1} \parallel} \times \left(\mathbf{k} \times \frac{-\mathbf{x}_{t+1}}{\parallel \mathbf{x}_{t+1} \parallel} \right) \right)^T, \tag{23}$$

$$\mathbf{r}_2 = \frac{\mathbf{r}_2'}{\parallel \mathbf{r}_2' \parallel}. \tag{24}$$

Using the above and the limit constraint $R_{t+1} = R_{max}$, we derive the following equation:

$$R_{max} = \sqrt{(x_d(t+1) - o_x)^2 + (y_d(t+1) - o_y)^2}. \tag{25}$$

Assuming that $\mathbf{x}_{t'} = [x_{t'1}, x_{t'2}, x_{t'3}]^T$ and $\mathbf{p}_{t'} = [\frac{q_{t1}}{F}, \frac{q_{t2}}{F}, \frac{q_{t3}}{F}]^T$, where $t' = t+1$, Eq. (25) can be solved for f, to find the maximum focal length f_{max} for motion types having $d_{t+1} > 0$:

$$f_{max} = \frac{R_{max} d_{t'} s_x s_y |E_1 + F \parallel \mathbf{x}_{t'} \parallel^2 |}{\sqrt{(s_x q_{t3} d_{t'}^2 - s_x x_{t'3} E_2)^2 + s_y^2 E_3^2 \parallel \mathbf{x}_{t'} \parallel^2}}, \tag{26}$$

where

$$E_1 = -q_{t1} x_{t'1} - q_{t2} x_{t'2} - q_{t3} x_{t'3}, \tag{27}$$

$$E_2 = q_{t1} x_{t'1} + q_{t2} x_{t'2}, \tag{28}$$

$$E_3 = q_{t2} x_{t'1} - q_{t1} x_{t'2}. \tag{29}$$

Since most of the UAV motion types are not affected by target altitude changes between successive video frames,

which are less likely to happen than direction and speed changes, $\mathbf{p}_{t+1}$ can be expressed by:

$$\mathbf{p}_{t+1} = [\frac{q_{t1}}{F}, \frac{q_{t2}}{F}, 0]^T. \quad (30)$$

In this case, the maximum focal length is given by:

$$f_{max} = \frac{R_{max} d_{t'} s_x s_y | - E_2 + F \parallel \mathbf{x}_{t'} \parallel^2 |}{\sqrt{s_x^2 E_2^2 x_{t'3}^2 + s_y^2 E_3^2 \parallel \mathbf{x}_{t'} \parallel^2}}. \quad (31)$$

When the UAV/camera is located exactly above the target for the $(t+1)$-th video frame, i.e., $\mathbf{x}_{t+1} = [0, 0, x_{t'3}]^T$, matrix $\mathbf{R}$ cannot be derived as described before, since $\mathbf{r}_1 \times \mathbf{k} = 0$. In this special case, where $d_{t+1} = 0$, it is easier to calculate the rotation matrix using Eq. (19) where $\theta_z = 0$ and $\theta_y = 180^o$:

$$\mathbf{R} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \quad (32)$$

Then, the maximum focal length is given by:

$$f_{max} = \frac{R_{max} F x_{t'3} s_x s_y}{\sqrt{s_y^2 q_{t1}^2 + s_x^2 q_{t2}^2}}. \quad (33)$$

The above analysis can be applied separately to each UAV/camera motion type, so as to derive a formula for the maximum focal length in each case. This can be performed by substituting the relations describing each motion type, i.e., the relations governing $\mathbf{x}_{t+1}$, into Eq. (26), (31) or (33). Below, LTS is studied as an example.

4. LTS SIMULATIONS

In LTS, even small target altitude variations have a great aesthetic impact on picture framing. Therefore, our study was performed by considering $q_{t3} \neq 0$. The UAV position is given by $\mathbf{x}_{t+1} = [0, x_{t2}, 0]^T$. As $\mathbf{p}_{t+1} = [\frac{q_{t1}}{F}, \frac{q_{t2}}{F}, \frac{q_{t3}}{F}]^T$, Eq. (26) can now be rewritten as follows:

$$f_{max} = \frac{R_{max} s_x s_y | q_{t2} - F x_{t2} |}{\sqrt{s_y^2 q_{t1}^2 + s_x^2 q_{t3}^2}}. \quad (34)$$

Following [7] and expanding upon it, the UAV and target motion was simulated for various representative AV shooting scenarios. We studied 8 different cases of values for the deviation vector $\mathbf{q}_t$. In the first two cases the target linearly accelerates/decelerates respectively ($\mathbf{q}_{t1} = [5, 0, q_{t3}]^T$, $\mathbf{q}_{t2} = [-5, 0, q_{t3}]^T$). In the third and fourth cases, the target is moving along a different direction than the expected one ($\mathbf{q}_{t3} = [0, 5, q_{t3}]^T$, $\mathbf{q}_{t4} = [0, -5, q_{t3}]^T$) but remains on the TCS **j**-axis. In the remaining cases, the target is moving diagonally to the TCS axes ($\mathbf{q}_{t5} = [5, 5, q_{t3}]^T$, $\mathbf{q}_{t6} = [-5, -5, q_{t3}]^T$, $\mathbf{q}_{t7} = [-5, 5, q_{t3}]^T$, $\mathbf{q}_{t8} = [5, -5, q_{t3}]^T$). Velocity deviations are expressed in meters/second. Figure 1 depicts the expected against the actual position of the target in each case.

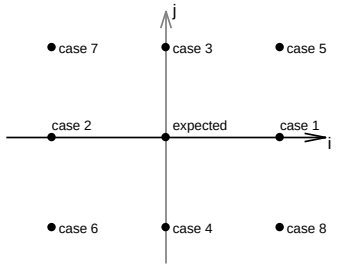

Fig. 1. The expected against the actual target position in the $(t+1)$-th frame, for the 8 simulated cases. The TCS axes are also depicted.

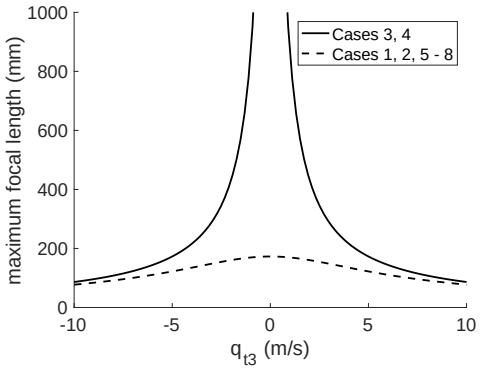

Fig. 2. Variation of f_{max} against q_{t3} for LTS.

The following simulation parameters have been used: maximum tracker search radius $R_{max} = 128$ pixels, pixel size $s_x = s_y = 0.009$ mm and video frame rate $F = 25$ fps. The simulation was performed for varying values of q_{t3}. The distance between the UAV and the target was chosen to be $\lambda = x_{t2} = 30m$. Simulation results fully comply with intuitive expectations and are shown in Figure 2. The obtained f_{max} values can easily be exploited for shot type feasibility testing, either on-line or in pre-production cinematography planning. Similar methodology may be followed for all motion types and various possible velocity deviation vectors (e.g., derived from target acceleration estimates).

5. CONCLUSIONS

In this paper five industry-standard target-tracking UAV/camera motion types have been formalized and geometrically modelled. This allows us to extract maximum focal length constraints for computer vision-assisted UAV physical target following. This is of utmost importance in cinematography applications, where the maximum focal length regulates the range of permissible shot types. The derived formulas can be readily employed as low-level rules in intelligent UAV shooting and cinematography planning systems.

6. REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] I. Tsingalis, A. Tefas, N. Nikolaidis, and I. Pitas, "Shot type characterization in 2D and 3D video content," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2014.

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

[13] C. Smith, *The Photographers Guide to Drones*, Rocky Nook, 2016.

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

[15] J. Angeles, *Fundamentals of robotic mechanical systems*, vol. 2, Springer, 2002.