

Autonomous drone cinematographer: Using artistic principles to create smooth, safe, occlusion-free trajectories for aerial filming

Rogerio Bonatti, Yanfu Zhang, Sanjiban Choudhury, Wenshan Wang, and Sebastian Scherer

Abstract Autonomous aerial cinematography has the potential to enable automatic capture of aesthetically pleasing videos without requiring human intervention, empowering individuals with the capability of high-end film studios. Current approaches either only handle off-line trajectory generation, or offer strategies that reason over short time horizons and simplistic representations for obstacles, which result in jerky movement and low real-life applicability. In this work we develop a method for aerial filming that is able to trade off shot smoothness, occlusion, and cinematography guidelines in a principled manner, even under noisy actor predictions. We present a novel algorithm for real-time covariant gradient descent that we use to efficiently find the desired trajectories by optimizing a set of cost functions. Experimental results show that our approach creates attractive shots, avoiding obstacles and occlusion 65 times over 1.25 hours of flight time, re-planning at 5Hz with a 10s time horizon. We robustly film human actors, cars and bicycles performing different motion among obstacles, using various shot types.

R. Bonatti (✉) · S. Choudhury · W. Wang · S. Scherer
Robotics Institute, Carnegie Mellon University, Pittsburgh, PA,
e-mail: rbonatti@cs.cmu.edu

S. Choudhury
e-mail: sanjibac@cs.cmu.edu

W. Wang
e-mail: wenshanw@andrew.cmu.edu

S. Scherer
e-mail: basti@cs.cmu.edu

Y. Zhang
Advanced Technology Research Division, Yamaha Motor Co., Ltd., Shizuoka, Japan
e-mail: zhangya@yamaha-motor.co.jp

1 Introduction

Aerial vehicles are revolutionizing the way both professional and amateur film makers capture shots of actors and landscapes, increasing the flexibility of narrative elements and allowing the composition of aerial viewpoints which are not feasible using traditional devices such as hand-held cameras and dollies. However, the use of drones for filming today is still extremely difficult due to several motion planning and human-computer interaction challenges.

Aerial cinematography incorporates objectives from different robotics areas. Similarly to high-speed flight [2, 21], the drone requires smooth and safe trajectories. Following literature in inspection and exploration, the vehicle reasons about viewpoints. In addition, temporal reasoning plays a major role when following a target, analogous to formation flight [32]. Cinematography also needs to consider artistic intent [1], using guidelines such as the rule of thirds, scale and relative angles [3].

Previous approaches in aerial filming do not address the complete problem in a sufficiently generalizable manner to be used in real-life scenarios. Off-line trajectory generation [8, 13, 28] cannot be used for most practical situations, and the on-line trajectory generation methods that have been proposed have limitations such as ignoring artistic objectives or only dealing with limited obstacle representations [22] (ignoring obstacles altogether in many cases).

Our key insight in this work is that this problem can be efficiently solved in real-time as a smooth trajectory optimization. Our contributions in this paper are threefold: (1) we formalize the aerial filming problem following cinematography guidelines for arbitrary types of shots and arbitrary obstacle shapes, (2) we present an efficient optimization method that exploits covariant gradients of the objective function for fast convergence, and (3) for over 1.25 hours of flight time while re-planning, we experimentally show robustness in real-world conditions with different types of shots and shot transitions, actor motions, and obstacle shapes. The supplementary video shows examples of trajectories: <https://youtu.be/QX73nBBwd28>.

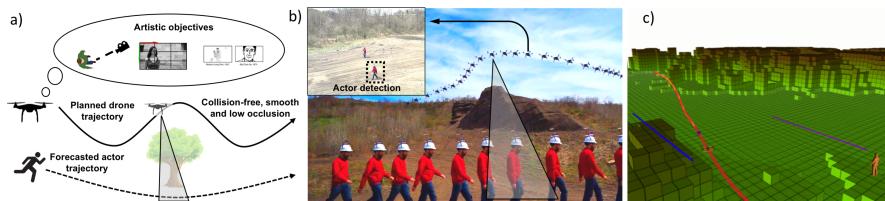


Fig. 1: Aerial cinematographer: a) The drone reasons about artistic guidelines, forecasts the actor’s motion and plans a smooth, collision-free trajectory while avoiding occlusion. b) Field results produce visually appealing images. The camera detects and tracks the actor, keeping him in the desired screen position. c) Motion planner view, with actor’s motion forecast (purple), desired cinematography guidelines (blue), and optimized trajectory (red).

2 Problem Formulation

The act of filming encompasses a broad spectrum of behaviors and definitions. For example, a single or multiple cameras can be filming a scene, which can be focused on a landscape, and/or contain one or multiple actors in the frame. In this work we focus on single-camera, single-actor scenarios, which are omnipresent in real-life application, and difficult to execute. Therefore, we have a quadrotor with trajectory ξ_q , coupled with a camera, that films one actor with trajectory ξ_a , where $\xi(t)$ maps time $t \in [0, t_f]$ to a configuration. At the end of the paper we also discuss how our definitions and algorithms could be extended to other filming scenarios.

Following literature in cinematography [1, 3], we identified a small set of camera positioning parameters that can define a large span of shot types (Figure 2). We define *static shots* as shots whose parameters remain static over time, independently of the motion of the actor in the environment, and *dynamic shots* as having time-dependant parameters.

While filming, a human finds appropriate camera movements based on a set of implicit and explicit costs. We define a generic framework for aerial filming, where the motion planner's objective is find a quadrotor's path ξ_q^* that minimizes a surrogate cost function J , that results in a smooth (J_{smooth}), collision-free (J_{obs}), occlusion-free (J_{occ}) trajectory that follows our artistic shot guidelines (J_{shot}) as closely as possible (Eq 1). The optimal trajectory minimizes the total cost J within a finite time horizon t_f , while belonging to the set of trajectories Ξ that obey the problem's boundary constraints (Eq 2).

$$J(\xi_q) = J_{smooth}(\xi_q) + \lambda_1 J_{obs}(\xi_q) + \lambda_2 J_{occ}(\xi_q, \xi_a) + \lambda_3 J_{shot}(\xi_q, \xi_a) \quad (1)$$

$$\xi_q^* = \arg \min_{\xi_q \in \Xi} J(\xi_q), \quad \forall t \in [0, t_f] \quad (2)$$

3 Related Work

Camera control in virtual cinematography is typically on through-the-lens control [4, 5, 10, 18] but disregards real-world limitations such as robot physics constraints and noisy motion predictions. When dealing with arbitrary real-life environments,

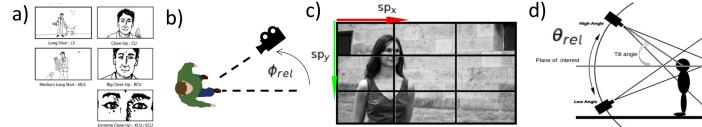


Fig. 2: Shot parameters, adapted from Bowen and Thompson [3]: a) shot scale ss corresponds to the size of the projection of the actor on the screen; b) line of action angle $\phi_{rel} \in [0, 2\pi]$; c) screen position of the actor projection $sp_x, sp_y \in [0, 1]$; d) tilt angle $\theta_{rel} \in [-\pi, \pi]$

voxel occupancy maps and truncated signed distance field (TSDF) [23] are common representations, and can supply distance and gradient of a point to nearest object surface. Aerial trajectory generation methods [20, 31] are typically designed for aggressive flight and rely on easy to evaluate objective or constraint functions. In general domains, we find techniques with more relaxed requirements [24, 29]. We build on CHOMP [25] due to its simple update rule that is amenable to new cost functions.

On aerial filming, we first expose works related to navigation using keyframes. Roberts and Hanrahan [28] generate off-line trajectories given potentially infeasible human-defined key-frames and Joubert et al. [13] provide a tool for interactive off-line design of camera trajectories. Similarly, Gebhardt et al. [9] improve a user’s time-parametrized trajectory, optimizing it for smooth motions. From user studies, Gebhardt et al. [9] show that smoothness is key to producing visually-appealing videos. Gebhardt et al. [8] also show a smooth trajectory optimization method for user-defined keyframes, including an intuitive user interface. Lino and Christie [17] analytically interpolate between viewpoints while maintaining shot quality. Galvane et al. [6] used this method to control quadrotors, but only in obstacle-free environments. Similarly, Joubert et al. [14] transition between shots for static actors while not colliding with them, but offer no solution to obstacle and occlusion avoidance. Lan et al. [15] also position the drone using keyframes, which defined by a user on the image itself, not on a global coordinate frame.

On a different line of work, Xie et al. [33] generates a set of candidate local camera movements to film landmarks using drones using visual composition rules, and combine the local trajectories into a global smooth camera sequence. Closest to our work, we find [7, 22]. Both methods use high-accuracy indoors motion-capture systems. Galvane et al. [7] use a search-based planner to find feasible trajectories; however, their system reacts to dynamic targets in a purely reactive manner, resulting in a non-smooth behavior. Nägeli et al. [22] apply MPC considering occlusion and safety. However, they plan for short time horizons, use simplistic elliptical representations for all obstacles, and use a high-accuracy motion-capture system. It is not clear if the black-box MPC solver that is used is amenable to other obstacle representations and noise in localization. In contrast, our method works for long time horizons, has a simple unconstrained update rule, operates on TSDFs, has small runtime onboard and can deal with noise in actor motion predictions.

4 Approach

Unlike previous works that operate either with high-accuracy indoor motion capture systems or precision RTK GPS outdoors, we use only conventional GPS, resulting in high noise for both drone localization and actor motion prediction. Therefore we decided to decouple the motion of the drone and the camera. The camera is

mounted on a 3-axis independent gimbal and can place the actor on the correct screen position by visual detection, despite errors in the drone's position. By decoupling camera movement, the trajectory to be optimized becomes $\xi_q(t) = [x_q(t) \ y_q(t) \ z_q(t)]^T$, assuming that the drone's orientation ψ_q points towards the actor at all times.

Designing differentiable cost functions for cinematography

We want trajectories which are smooth, safe, occlusion-free and that follow our artistic guidelines as closely as possible. Following the derivation seen in Section 3 of Zucker et al. [34], we define a parametrization-invariant smoothness cost that can be expressed as a quadratic function, and an obstacle avoidance cost based on a penalization c of the TSDF. We use obstacle avoidance for both the environment and for the dynamic actor, who for this purpose we represent as a moving sphere. In addition, we define two more cost functions specifically for cinematography:

Shot quality:

$$J_{\text{shot}}(\xi_q, \xi_{\text{shot}}) = \frac{1}{t_f} \frac{1}{2} \int_0^{t_f} \|\xi_q(t) - \xi_{\text{shot}}(t)\|^2 dt \approx \frac{1}{2(n-1)} \text{Tr}(\xi_q^T A_{\text{shot}} \xi_q + 2\xi_q^T b_{\text{shot}} + c_{\text{shot}}) \quad (3)$$

$$\nabla J_{\text{shot}}(\xi_q) = \frac{1}{n-1} (A_{\text{shot}} \xi_q + b_{\text{shot}}) \quad (4)$$

Written in a quadratic form, it measures the average squared distance between ξ_q and an ideal trajectory ξ_{shot} that only considers positioning via cinematography parameters. ξ_{shot} can be computed analytically: for each point $\xi_a(t)$ in the actor motion prediction, the drone position lies on a sphere centered at the actor with radius calculated via the shot scale, and angles given by ϕ_{rel} and θ_{rel} , as in Figure 2.

Occlusion avoidance:

$$J_{\text{occ}}(\xi_q, \xi_a) = \int_{t=0}^{t_f} \int_{\tau=0}^1 c(p(\tau)) \left\| \frac{d}{d\tau} p(\tau) \right\| d\tau \left\| \frac{d}{dt} \xi_q(t) \right\| dt, \quad (5)$$

Even though the concept of occlusion is binary, *i.e.* we either have or don't have visibility of the actor, a major contribution of our work is to define a differentiable cost that expresses a viewpoint's occlusion intensity for arbitrary obstacle shapes. Mathematically, we define occlusion as the integral of the TSDF cost c over a 2D manifold connecting both trajectories ξ_q and ξ_a . The manifold is built by connecting each drone-actor position pair in time using the path $p(\tau)$. We then derive the functional gradient:

$$\begin{aligned} \nabla J_{\text{occ}}(\xi_q, \xi_a)(t) = & \\ \int_{\tau=0}^1 \nabla c(p(\tau)) |L| |\dot{q}| & \left[I - (\hat{q} + \tau(\frac{\dot{a}}{|\dot{q}|} - \hat{q})) \hat{q}^T \right] - c(p(\tau)) |\dot{q}| \left[\hat{L}^T + \frac{\hat{L}^T \hat{L} \hat{q}^T}{|\dot{q}|} + |L| \kappa^T \right] d\tau, \end{aligned} \quad (6)$$

where: $q = \xi_q(t), \quad a = \xi_a(t), \quad p(\tau) = (1-\tau)q + \tau a, \quad \hat{v} = \frac{v}{|v|},$
 $\kappa = \frac{1}{|\dot{q}|^2} (I - \hat{q} \hat{q}^T) \ddot{q}, \quad L = a - q.$

Intuitively, the term $\nabla c(p(\tau))$ is related to variations of the gradient in space, and the term τ acts as a lever for the gradient. The term $c(p(\tau))$ is linked to changes in path length between camera and actor.

Covariant gradient and steepest descent method

Our objective is to minimize the functional $J(\xi_q)$ (Eq. 2). Following a first-order Taylor expansion around the current iteration i using gradient $\nabla J(\xi_{qi})$, we build Alg. 1, adapted from [25]. We follow a direction of steepest descent on the functional cost, using metric M , which compounds the quadratic terms coming from both J_{smooth} and J_{shot} . M can be seen as an approximation of the Hessian, and only needs to be inverted once, outside of the algorithm's main loop, making the optimization computationally efficient. We follow conventional stopping criteria for descent algorithms, and limit the maximum number of iterations. We use the solution to the previous planning problem concatenated with a straight line segment for future time as the initialization for the next optimization.

Algorithm 1: Optimize (ξ_q)

```

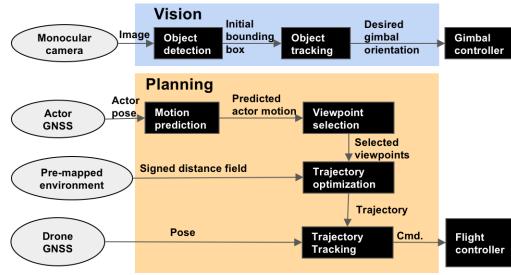
1  $M_{\text{inv}} \leftarrow (A_{\text{smooth}} + \lambda_3 A_{\text{shot}})^{-1};$ 
2 for  $i = 0, 1, \dots, i_{\max}$  do
3   if  $(\nabla J(\xi_{qi})^T M_{\text{inv}} \nabla J(\xi_{qi}))^2 / 2 < \epsilon_0$  or  $(J(\xi_{qi}) - J(\xi_{qi-1})) < \epsilon_1$  then
4     return  $\xi_{qi};$ 
5   end
6    $\xi_{qi+1} = \xi_{qi} - \frac{1}{\eta} M_{\text{inv}} \nabla J(\xi_{qi});$ 
7 end
8 return  $\xi_{qi};$ 
```

Gimbal control using the actor's image detection

Our system does not use the actor's state estimation coming from a noisy GPS to control the gimbal to keep the actor within the desired screen position. Instead, we only rely on image feedback. Using the monocular images from the camera, the detection module outputs a bounding box to initialize the tracking process at a higher frame rate. We then use a PD controller to keep the center of the bounding box in the desired screen position.

We selected state of the art algorithms with a good speed-accuracy tradeoff for the vision pipeline. We tested three algorithms: single shot detector (SSD) [19], Faster R-CNN [27], and YOLO2 [26]. Faster R-CNN performs the best in terms of precision-recall metrics while operating at a reasonable frame rate. We use MobileNet [12] for feature extraction due to its low memory usage and fast inference speed. The per-frame inference is around 300 ms, running on the a Nvidia TX2 computer. We train all models using a mixture of COCO [16] dataset and our own custom dataset, with a 1:10 ratio, and about 70,000 images in total. We limited the detection categories only to *person*, *car*, *bicycle*, and *motorcycle*, which commonly appear as actors in aerial filming. Finally for actor tracking, we use KCF [11] due to its real-time performance.

Fig. 3 System architecture. The vision subsystem controls the camera orientation using only the monocular image, independently of the planning subsystem. Planning uses the drone’s and actor’s current location, and the environment to generate trajectories for the flight controller.



5 Experiments

Systems and experimental setup

Our drone is the DJI M210 model, coupled with a NVIDIA TX2 computer for both vision and planning pipelines (Figure 3). The actor wears a Pixhawk PX4 module on a hat that sends his pose to the onboard computer via radio communication. A linear Kalman filter uses these pose estimates to infer his velocity, which is used to predict her trajectory (ξ_a) for the next 10 s. Using a point cloud map of the test site we pre-compute a TSDF map of the region of interest. Re-planning happens at 5 Hz with a 10 s horizon. To simulate the full pipeline and to decide on the relative weights between each cost function (Eq. 1), we built a ROS wrapper to test our software in a photo-realistic environment [30] (Fig. 4).

Experiments

(a) *Algorithm robustness*: We evaluated our algorithm performing different types of static and dynamic shots, following different types of actors: humans, cars and bicycles at different speeds and motion types. In total, we collected over 1.25 hours of flight time while re-planning and avoided obstacles and/or occlusion 65 times. The maximum velocity achieved during the tests was of 7.5 m/s. Figure 5 summarizes the most representative shots, which are also shown in the supplementary video.

(b) *Using occlusion cost function*: For the same shot type, we planned paths with and without the occlusion cost function (Figure 6). Our proposed occlusion cost significantly improves the aesthetics of the resulting image, keeping the actor on frame while avoiding obstacles.

(c) *Statistical analysis*: We evaluate our algorithm on randomized environments (Fig 7), and display results in Table 1.

Fig. 4 Photo-realistic simulator used to test the system. Third and first-person renderings shown on the left, and occupancy map with drone trajectory shown on the right.



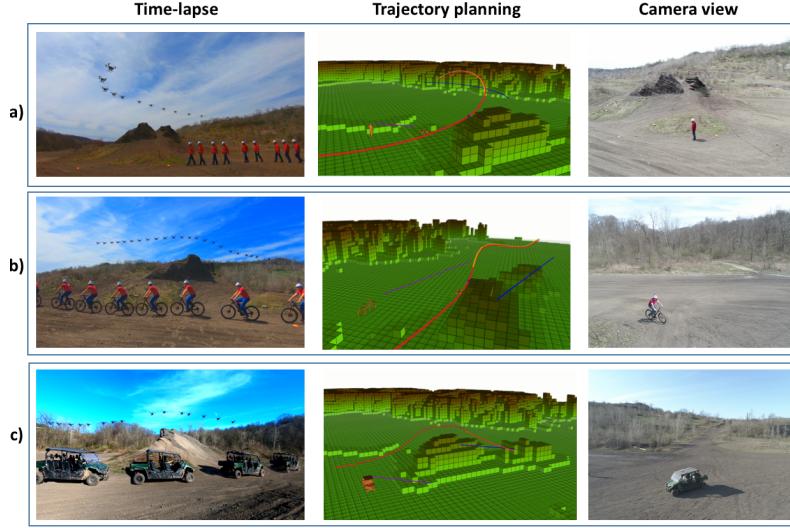


Fig. 5: Results: a) Circling shot around person, b) Side shot following biker, c) Side shot following vehicle. The planned trajectory (red) avoids colliding with and being occluded by the mountain, while remaining smooth even under high actor motion prediction noise. The actor's motion forecast is in purple, and the desired artistic shot is in blue.

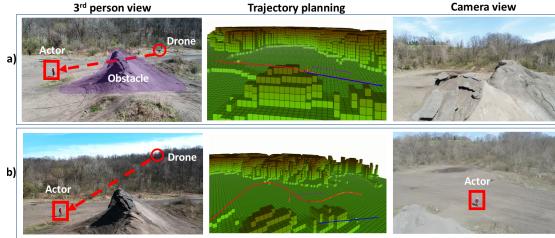
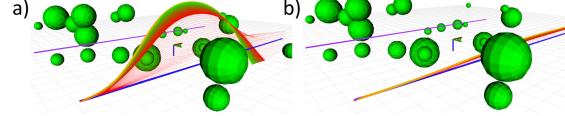


Fig. 6 Comparison of planning a) without and b) with occlusion cost function. The occlusion cost function significantly improves the quality of the camera image in comparison with pure obstacle avoidance, for same shot type.

Fig. 7 Randomized environment with obstacles to evaluate planner robustness. a) Solution including occlusion cost function, and b) Pure obstacle avoidance.



6 Conclusion and discussion

In this work we propose and validate a system for autonomous aerial cinematography. Our system is able to execute aesthetically pleasing shots, calculating trajectories that balance motion smoothness, occlusion, and cinematography guidelines. Our experimental results show the algorithm's robustness and speed in a real-life scenarios outdoors, even under noisy actor predictions. In addition, we show that the occlusion

Table 1: Evaluation of motion planner performance in the randomized environment from Fig 7. By adding the occlusion cost function we improve actor visibility over 10% in comparison with pure obstacle avoidance in environments with 40 spheres. However, by avoiding occlusion the planner also increases the average distance to the desired artistic trajectory. We generated 100 random configurations for each environment complexity level.

Success metric	Cost functions	Num. spheres in environment		
		1	20	40
Actor visibility along trajectory	$J_{occ} + J_{obs}$	$99.4 \pm 2.2\%$	$94.2 \pm 7.3\%$	$86.9 \pm 9.3\%$
	J_{obs}	$98.8 \pm 3.0\%$	$87.1 \pm 8.5\%$	$75.3 \pm 11.8\%$
Avg. dist. to ξ_{shot} , in m	$J_{occ} + J_{obs}$	0.4 ± 0.4	6.2 ± 11.2	10.7 ± 13.2
	J_{obs}	0.05 ± 0.1	0.3 ± 0.2	0.5 ± 0.3

cost function we introduced significantly improves the quality of the resulting image, and works for arbitrary obstacle shapes.

There are many key aspects that still need to be solved for us to create a fully autonomous filming system. In terms of actor localization, we are currently working on using only visual inputs to identify her position, with no need for GPS. In addition, one can incorporate an online-mapping module to the vehicle, which could be integrated with minimal changes to our algorithm, but was out of the scope of this work. We are also investigating techniques to automatically select the best artistic intent, *i.e.* type of shot, for a particular scenario, taking into account motion cues of the actor and obstacle configurations. In the longer term, we also plan to adapt our algorithm to multi-drone and multi-actor configurations by adding a cost function to penalize inter-drone sight.

Acknowledgements We thank Lentin Joseph, Aayush Ahuja, Delong Zhu, and Greg Armstrong for the assistance in field experiments and robot construction. Research presented in this paper was funded by Yamaha Motor Co., Ltd. .

References

- [1] Daniel Arjon. Grammar of the film language. 1976.
- [2] Andrew J Barry, Peter R Florence, and Russ Tedrake. High-speed autonomous obstacle avoidance with pushbroom stereo. *Journal of Field Robotics*, 35(1):52–68, 2018.
- [3] Christopher J Bowen and Roy Thompson. *Grammar of the Shot*. Taylor & Francis, 2013.
- [4] Marc Christie, Patrick Olivier, and Jean-Maria Normand. Camera control in computer graphics. In *Computer Graphics Forum*, volume 27, pages 2197–2218. Wiley Online Library, 2008.
- [5] Steven M Drucker and David Zeltzer. Intelligent camera control in a virtual environment. In *Graphics Interface*, pages 190–190. Citeseer, 1994.
- [6] Quentin Galvane, Julien Fleureau, Francois-Louis Tariolle, and Philippe Guillotel. Automated cinematography with unmanned aerial vehicles. *arXiv preprint arXiv:1712.04353*, 2017.
- [7] Quentin Galvane, Christophe Lino, Marc Christie, Julien Fleureau, Fabien Servant, and Philippe Guillotel. Directing cinematicographic drones. *arXiv preprint arXiv:1712.04216*, 2017.

- [8] Christoph Gebhardt, Benjamin Hepp, Tobias Nägeli, Stefan Stevšić, and Otmar Hilliges. Airways: Optimization-based planning of quadrotor trajectories according to high-level user goals. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2508–2519. ACM, 2016.
- [9] Christoph Gebhardt, Stefan Stevšić, and Otmar Hilliges. Optimizing for aesthetically pleasing quadrotor camera motion. 2018.
- [10] Michael Gleicher and Andrew Witkin. Through-the-lens camera control. In *ACM SIGGRAPH Computer Graphics*, volume 26, pages 331–340. ACM, 1992.
- [11] João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):583–596, 2015.
- [12] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
- [13] Niels Joubert, Mike Roberts, Anh Truong, Flora Berthouzoz, and Pat Hanrahan. An interactive tool for designing quadrotor camera shots. *ACM Transactions on Graphics (TOG)*, 34(6):238, 2015.
- [14] Niels Joubert, Dan B Goldman, Flora Berthouzoz, Mike Roberts, James A Landay, Pat Hanrahan, et al. Towards a drone cinematographer: Guiding quadrotor cameras using visual composition principles. *arXiv preprint arXiv:1610.01691*, 2016.
- [15] Ziquan Lan, Mohit Shridhar, David Hsu, and Shengdong Zhao. Xpose: Reinventing user interaction with flying cameras. In *Robotics: Science and Systems*, 2017.
- [16] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [17] Christophe Lino and Marc Christie. Intuitive and efficient camera control with the toric space. *ACM Transactions on Graphics (TOG)*, 34(4):82, 2015.
- [18] Christophe Lino, Marc Christie, Roberto Ranon, and William Bares. The director’s lens: an intelligent assistant for virtual cinematography. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 323–332. ACM, 2011.
- [19] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.
- [20] Daniel Mellinger and Vijay Kumar. Minimum snap trajectory generation and control for quadrotors. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 2520–2525. IEEE, 2011.
- [21] Kartik Mohta, Michael Watterson, Yash Mulgaonkar, Sikang Liu, Chao Qu, Anurag Makineni, Kelsey Saulnier, Ke Sun, Alex Zhu, Jeffrey Delmerico, et al. Fast, autonomous flight in gps-denied and cluttered environments. *Journal of Field Robotics*, 35(1):101–120, 2018.
- [22] Tobias Nägeli, Lukas Meier, Alexander Domahidi, Javier Alonso-Mora, and Otmar Hilliges. Real-time planning for automated multi-view drone cinematography. *ACM Transactions on Graphics (TOG)*, 36(4):132, 2017.
- [23] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE, 2011.
- [24] Nathan Ratliff, Matthew Zucker, J Andrew Bagnell, and Siddhartha Srinivasa. CHOMP: Gradient optimization techniques for efficient motion planning. In *Robotics and Automation, 2009. ICRA’09. IEEE International Conference on*, pages 489–494. IEEE, 2009.
- [25] Nathan D Ratliff, David Silver, and J Andrew Bagnell. Learning to search: Functional gradient techniques for imitation learning. *Autonomous Robots*, 27(1):25–53, 2009.
- [26] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. *arXiv preprint*, 2017.
- [27] Shaogang Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.
- [28] Mike Roberts and Pat Hanrahan. Generating dynamically feasible trajectories for quadrotor cameras. *ACM Transactions on Graphics (TOG)*, 35(4):61, 2016.
- [29] John Schulman, Jonathan Ho, Alex Lee, Ibrahim Awwal, Henry Bradlow, and Pieter Abbeel. Finding locally optimal, collision-free trajectories with sequential convex optimization. In *Robotics: Science and Systems*, volume 9, pages 1–10. Citeseer, 2013.
- [30] Shital Shah, Debadeepa Dey, Chris Lovett, and Ashish Kapoor. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. 2017. URL <https://arxiv.org/abs/1705.05065>.
- [31] David H Shim, H Jin Kim, and Shankar Sastry. Decentralized nonlinear model predictive control of multiple flying robots. In *Decision and control, 2003. Proceedings. 42nd IEEE conference on*, volume 4, pages 3621–3626. IEEE, 2003.
- [32] Matthew Turpin, Nathan Michael, and Vijay Kumar. Trajectory design and control for aggressive formation flight with quadrotors. *Autonomous Robots*, 33(1-2):143–156, 2012.
- [33] Ke Xie, Hao Yang, Shengqiu Huang, Dani Lischinski, Marc Christie, Kai Xu, Minglun Gong, Daniel Cohen-Or, and Hui Huang. Creating and chaining camera moves for quadrotor videography. *ACM Transactions on Graphics*, 37:14, 2018.
- [34] Matt Zucker, Nathan Ratliff, Anca D Dragan, Mihail Pivtoraiko, Matthew Klingensmith, Christopher M Dellin, J Andrew Bagnell, and Siddhartha S Srinivasa. CHOMP: Covariant hamiltonian optimization for motion planning. *The International Journal of Robotics Research*, 32(9-10):1164–1193, 2013.