

# Optimizing for Aesthetically Pleasing Quadrotor Camera Motion

CHRISTOPH GEBHARDT, STEFAN STEVŠIĆ, and OTMAR HILLIGES, AIT Lab, ETH Zürich

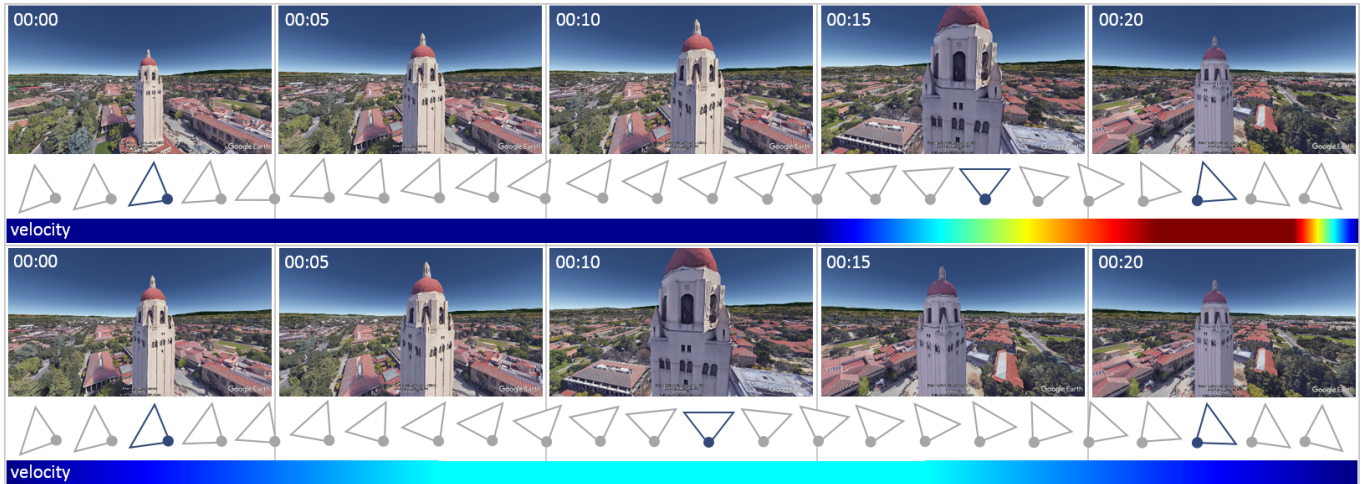


Fig. 1. Quadrotor camera tools generate trajectories based on user-specified keyframes in time and space. Reasoning about spatio-temporal distances is hard for users and can lead to visually unappealing results and fluctuating camera velocities. *Top row*: user-specified keyframes (blue) are positioned in time, such that the camera first moves too slow and then needs to accelerate drastically to reach the specified end-point. *Bottom row*: results of our method which automatically positions keyframes (blue) in time such that the camera moves smoothly over the entire trajectory (illustrative example).

In this paper we first contribute a large scale online study ( $N \approx 400$ ) to better understand aesthetic perception of aerial video. The results indicate that it is paramount to optimize smoothness of trajectories across all keyframes. However, for experts timing control remains an essential tool. Satisfying this dual goal is technically challenging because it requires giving up desirable properties in the optimization formulation. Second, informed by this study we propose a method that optimizes positional and temporal reference fit *jointly*. This allows to generate globally smooth trajectories, while retaining user control over reference timings. The formulation is posed as a variable, infinite horizon, contour-following algorithm. Finally, a comparative lab study indicates that our optimization scheme outperforms the state-of-the-art in terms of perceived usability and preference of resulting videos. For novices our method produces smoother and better looking results and also experts benefit from generated timings.

CCS Concepts: • **Computing methodologies** → **Computer graphics**; **Motion path planning**; **Robotic planning**; • **Computer systems organization** → External interfaces for robotics;

Additional Key Words and Phrases: computational design, aerial videography, quadrotor camera tools, trajectory optimization

Authors' address: Christoph Gebhardt, cgebhardt@inf.ethz.ch; Stefan Stevšić, stevsics@inf.ethz.ch; Otmar Hilliges, hilliges@inf.ethz.ch, AIT Lab, ETH Zürich.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

0730-0301/2018/8-ART90 \$15.00

<https://doi.org/10.1145/3197517.3201390>

## ACM Reference Format:

Christoph Gebhardt, Stefan Stevšić, and Otmar Hilliges. 2018. Optimizing for Aesthetically Pleasing Quadrotor Camera Motion. *ACM Trans. Graph.* 37, 4, Article 90 (August 2018), 11 pages. <https://doi.org/10.1145/3197517.3201390>

## 1 INTRODUCTION

Camera quadrotors have become a mainstream technology but fine-grained control of such camera drones for aerial videography is a high-dimensional and hence difficult task. In response several tools have been proposed to plan quadrotor shots by defining keyframes in virtual environments [Gebhardt et al. 2016; Gebhardt and Hilliges 2018; Joubert et al. 2015; Roberts and Hanrahan 2016]. This input is then used in an optimization algorithm to automatically generate quadrotor and camera trajectories. Intuitively, smooth camera motion is an obvious factor impacting the visual quality of a shot. This intuition alongside expert-feedback [Joubert et al. 2015] and literature on (aerial) cinematography [Arijon 1976; Audronis 2014; Hennessy 2015] forms the basis for most existing quadrotor tools. These take a spline representation, connecting user specified keyframes, and optimize higher derivatives of these splines, such as jerk.

Balasubramanian et al. [2015] define global smoothness as “a quality related to the continuity or non-intermittency of a movement, independent of its amplitude and duration”. However, because keyframe timings are kept fixed in current quadrotor camera optimization schemes [Gebhardt et al. 2016; Joubert et al. 2015], or close to the user input [Roberts and Hanrahan 2016], smooth motion can only be generated subject to these hard-constraints. This can cause strong variation of camera velocities across different trajectory segments and result in visually unpleasant videos.

Consider popular fly-by-shots, such as the one illustrated in Figure 1, where an object is filmed first from one direction and then gradually the camera yaws around its own z-axis by  $180^\circ$  as the quadrotor flies past the object until it is filmed from the opposing direction. To achieve visually pleasing footage both the quadrotor motion and the camera's angular velocity need to be smooth. Users generally struggle with this or similar problems in which they place the keyframes in the correct *spatial* location but too close (or too far) to each other *temporally* (see Figure 1, top and [video]). This is indeed a difficult task because keyframes are specified in 5D (3D position and camera pitch and yaw) and imagining the resulting translational and rotational velocities is cognitively demanding.

Although existing work provides UI tools (i.e. progress curves, timelines) to cope with this problem, it has been shown that users, especially novices, struggle to create smooth camera motion over a sequence of keyframes [Gebhardt and Hilliges 2018]. While optimizing for global smoothness may address this issue for novices, an interesting tension arises when looking at experienced users. Experts explicitly time the visual progression of a shot in order to achieve desired compositional effects [Joubert et al. 2015] (e.g. ease-in, ease-out behavior). Our first contribution is a large online study ( $N = 424$ ), highlighting this issue, where non-expert designed videos were rated more favorable when optimized for global smoothness while expert-designed videos were perceived as more pleasing with hard-constrained timings. To the best of our knowledge, this is the first study that provides empirical evidence for global smoothness indeed being important for the perception of aerial videography.

Embracing this dichotomy (of smoothness versus timing control), our second contribution is a trajectory optimization method that takes smoothness as primary objective and can re-distribute robot positions and camera angles in space-time. We propose the first algorithm in the area of quadrotor videography that treats keyframe timings and positions, and reference velocities as *soft-constraints*. This extends the state-of-the-art in that it allows users to trade off path-following fidelity with temporal fidelity. Such a formulation poses significant technical difficulties. Prior methods incorporate keyframe timings as *hard-constraints*, yielding a quadratic and hence convex optimization formulation (depending on the dynamical model), allowing for efficient implementation. In contrast, we formulated the quadrotor camera trajectory generation problem as a variable, infinite horizon, contour-following algorithm applicable to linear and non-linear quadrotor models. Our formulation has to discretize the model at each solver iteration according to the optimized trajectory end time. Although this formulation is no-longer convex, it is formulated as well-behaved non-convex problem and our implementation runs at interactive rates.

Finally, we show the benefit of our method compared to the state-of-the-art in a lab study in which we compare different variants of our method with [Gebhardt et al. 2016]. It can be shown that our method positively affects the usability of quadrotor camera tools and improves the visual quality of video shots for experts and non-experts. Both benefit from using an optimized timing initially, fine-tuning it according to their intention. In addition, the user study revealed that timing control does not need to be precise but is rather used to control camera velocity in order to create a certain compositional effect.

## 2 RELATED WORK

*Camera Control in Virtual Environments:* Camera placement [Lino and Christie 2012], path planning [Li and Cheng 2008; Yeh et al. 2011] and automated cinematography [Lino et al. 2011] have been studied extensively in virtual environments (VE), for a survey see [Christie et al. 2008]. These works share our goal of assisting users in the creation of camera motion (e.g., [Drucker and Zeltzer 1994; Lino and Christie 2015]). Nevertheless, it is important to consider that VEs are not limited by real-world physics and robot constraints, hence may yield trajectories that can not be flown by a quadrotor.

*Character Animation:* In character animation, a variety of methods exist which are capable of trading-off positional and temporal reference fit to optimize for smoother character motion. In [Liu et al. 2006], the authors specify constraints in warped time and then optimize the mapping between warped and actual time according to their objective function. For an original motion, [McCann et al. 2006] find the convex hull of all physically valid motions attainable via re-timing. Plausible new motions are then found by performing gradient descent and penalizing distance between possible solutions and the feasible hull. Like [Liu et al. 2006], our formulation is based on a time-free parameterization of a reference path. In contrast to the character animation methods, we adjust timings by optimizing the progress of the quadrotor camera on the reference according to a objective favoring smoothness. Unlike [McCann et al. 2006] our formulation does not require nested optimization.

*Trajectory Generation:* Trajectory generation for dynamical systems is a well studied problem in computer graphics [Geijtenbeek and Pronost 2012] and robotics [Betts 2009]. Approaches that encode the system dynamics as equality constraints to solve for the control inputs along a motion trajectory are referred to as spacetime constraints in graphics [Witkin and Kass 1988] and direct collocation in robotics [Betts 2009]. Used out-of-the box such approaches can lead to slow convergence time especially with long time horizons (cf. [Roberts and Hanrahan 2016]).

With the commoditization of quadrotors, the generation of drone trajectories shifted into the focus of research. Exploiting the differential flatness of quadrotors in the output space, [Mellinger and Kumar 2011] generated physically feasible minimal snap trajectories. Several methods exist for the generation of trajectories for aggressive quadrotor flight [Bry et al. 2015; Mellinger et al. 2012]. Traditionally, these methods convert a sequence of input positions into a time-dependent reference and based on a dynamical model generate a trajectory which follows this reference. For [Bry et al. 2015; Mellinger and Kumar 2011], time optimization is done in a cascaded manner where an approximated gradient descent for keyframe timings is calculated based on the original optimization problem. These formulations suffer from very long runtimes as the original problem needs to be called once for each keyframe to calculate the gradient approximation. In contrast, our method optimizes keyframe timings and trajectory jointly reducing optimization runtime and allowing to trade-off temporal and positional fit. In [Mellinger et al. 2012], sequentially composed controllers are used to optimize the timing of a trajectory such that physical limits are not violated given desired

feed-forward terms. Our work does not only ensure physical feasibility but is also capable of generating trajectories with different dynamics (smooth and more aggressive) for the same spatial input.

*Computational Support of Aerial Videography:* A number of tools for the planning of aerial videography exists. Commercially available applications and consumer-grade drones often place waypoints on a 2D map [APM 2016; DJI 2016; Technology 2016] or allow to interactively control the quadrotor’s camera as it tracks a pre-determined path [3D Robotics 2015]. These tools generally do not provide means to ensure feasibility of the resulting plans and do not consider aesthetic or usability objectives in the video composition task. The planning of physically feasible quadrotor camera trajectories has recently received a lot of attention. Such tools allow for planning of aerial shots in 3D virtual environments [Gebhardt et al. 2016; Gebhardt and Hilliges 2018; Joubert et al. 2015; Roberts and Hanrahan 2016] and employ optimization to ensure that both aesthetic objectives and robot modeling constraints are considered.

In [Joubert et al. 2015] and [Gebhardt et al. 2016], users specify keyframes in time and space. These are incorporated as hard-constraints into an objective function. Solving for the trajectory only optimizes camera dynamics and positions. This causes the generation of locally smooth camera motion (between keyframes) but can lead to varying velocities across keyframes. Joubert et al. [2015] detect violations of the robot model constraints. However, correcting these violations is offloaded to the user. In contrast, by generating timings or incorporating them as soft-constraints our optimization returns the closest feasible fit of the user-specified inputs, subject to our robot model, and generates globally smooth quadrotor camera trajectories. [Gebhardt and Hilliges 2018] re-optimizes keyframe timings in a cascaded optimization scheme. Here an approximated gradient on the keyframe times produced by the optimization formulation of [Gebhardt et al. 2016] is calculated and used to improve visual smoothness. However, this approach is relatively slow and the paper reports that users therefore did not make significant use of it in the evaluation. In contrast, our method runs at interactive rates optimizing trajectories of different duration within seconds (avg. 2.4 s). Roberts and Hanrahan [2016] take physically infeasible trajectories and compute the closest possible feasible trajectory by re-timing the trajectories subject to a non-linear quadrotor model. In contrast, we prevent trajectories from becoming infeasible at optimization time. Although the method of [Roberts and Hanrahan 2016] theoretically can be used to adjust timings based on a jerk minimization objective, our method can also trade-off the positional fit of a reference path in order to achieve even smoother motion.

Recently, several works address the generation of quadrotor camera trajectories in real-time to record dynamic scenes. [Galvane et al. 2016; Joubert et al. 2016] plan camera motion in a lower dimensional subspace to attain real-time performance. Using a Model Predictive Controller (MPC), [Nägeli et al. 2017] optimizes cinematographic constraints, such as visibility and position on the screen, subject to robot constraints for a single quadrotor. [Nägeli et al. 2017] extends this work for multiple drones and allows actor-driven tracking on a geometric path. Focusing on dynamic scenes, this work does not cover the global planning aspects of aerial videography.

*Online Path Planning:* Approaches that address trajectory optimization and path following have been proposed in the control theory literature. They allow for optimal reference following given real world influences. Methods like MPC [Faulwasser et al. 2009] optimize the reference path and the actuator inputs simultaneously based on the system state. MPC has been successfully used for the real-time generation of quadrotor trajectories [Mueller and D’Andrea 2013]. Nevertheless, [Aguilar et al. 2008] show that the tracking error for following timed-trajectories can be larger than for following a geometric path only. Motivated by this observation Model Predictive Contouring Control (MPCC) [Lam et al. 2013] has been proposed to follow a time-free reference, optimizing system control inputs for time-optimal progress. MPCC approaches have been successfully applied in industrial contouring [Lam et al. 2013] and RC racing [Liniger et al. 2014]. Recently, [Nägeli et al. 2017] extended the MPCC-framework to allow for real-time path following in 3D space with quadrotors. We propose a trajectory generation method that is conceptually related to MPCC formulations in that it optimizes timings for a quadrotor camera trajectory based on a time-optimal path-following objective. Our formulation treats keyframes, user specified reference timings and velocities as well as smoothness across the entire trajectory jointly in a soft-constrained formulation and allows users to produce aesthetically more pleasing videos.

### 3 METHOD

We propose a new method to generate *globally smooth* quadrotor camera trajectories. Our aim is to allow even novice users to design complex shots without having to explicitly reason about 5D spatio-temporal distances. Our central hypothesis is that smoothness across the entire trajectory matters and hence is the main objective of our optimization formulation. We first introduce the model of the system dynamics in Section 3.1 and discuss our optimization formulation in Section 3.2-3.5. See Appendix A for a table of notations.

#### 3.1 Dynamical Model

We use the approximated quadrotor camera model of [Gebhardt et al. 2016]. This discrete first-order dynamical system is incorporated as equality constraint into our optimization problem:

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{A}\mathbf{x}_i + \mathbf{B}\mathbf{u}_i + \mathbf{g}, \quad \mathbf{u}_{\min} \leq \mathbf{u}_i \leq \mathbf{u}_{\max}, \\ \mathbf{x}_i &= [\mathbf{r}, \psi_q, \psi_g, \phi_g, \dot{\mathbf{r}}, \dot{\psi}_q, \dot{\psi}_g, \dot{\phi}_g, \ddot{\mathbf{r}}, \ddot{\psi}_q, \ddot{\psi}_g, \ddot{\phi}_g, \ddot{\mathbf{r}}, \ddot{\psi}_q, \ddot{\psi}_g, \ddot{\phi}_g]^T, \quad (1) \\ \mathbf{u}_i &= [\mathbf{F}, M_{\psi_q}, M_{\psi_g}, M_{\phi_g}]^T, \end{aligned}$$

where  $\mathbf{x}_i \in \mathbb{R}^{24}$  are the quadrotor camera states and  $\mathbf{u}_i \in \mathbb{R}^6$  are the inputs to the system at stage  $i$ . Furthermore,  $\mathbf{r} \in \mathbb{R}^3$  is the position of the quadrotor,  $\psi_q$  is the quadrotor’s yaw angle and  $\psi_g$  and  $\phi_g$  are the yaw and pitch angles of the camera gimbal. The matrix  $\mathbf{A} \in \mathbb{R}^{24 \times 24}$  propagates the state  $\mathbf{x}$  forward, the matrix  $\mathbf{B} \in \mathbb{R}^{24 \times 6}$  defines the effect of the input  $\mathbf{u}$  on the state and the vector  $\mathbf{g} \in \mathbb{R}^{24}$  that of gravity for one time-step.  $\mathbf{F}$  is the force acting on the quadrotor,  $M_{\psi_q}$  is the torque along its z-axis and  $M_{\psi_g}, M_{\phi_g}$  are torques acting on pitch and yaw of the gimbal.

Please note that our formulation is agnostic to the dynamical model of the quadrotor. We verified this by incorporating the non-linear model of [Nägeli et al. 2017]. Qualitatively this does not impact results, yet the computational cost increases (see Figure 5).

### 3.2 Variable Horizon

In space-time optimization, the horizon length is defined by dividing the timing of the last keyframe by the discretization step  $\Delta t$ . However, one key idea in our formulation is that we treat trajectories, at least initially, as time free. In particular, our method does not take timed keyframes as input and therefore traditional approaches to determining the horizon length are not applicable.

Taking inspiration from MPC literature [Michalska and Mayne 1993], we make the length of the horizon an optimization variable itself by adding the trajectory end time  $T$  into the state space of our model ( $\mathbf{x} = [\mathbf{x}, T]^T \in \mathbb{R}^{25}$  with  $\frac{\delta T}{\delta t} = 0$ ). This has implications for the dynamical model. At each iteration of the solver we adjust the discretization step  $\Delta t = \frac{T}{N}$ . Here  $N$  is the number of stages in the horizon spanning the entire trajectory. The forward propagation matrices  $A$  and  $B$  are also recalculated based on the current  $\Delta t$ .

### 3.3 Reference Tracking Metric

We require a time-free parameterization of the reference to optimize the timing of keyframes. We use a chord length parameterized, piecewise cubic polynomial spline in hermite form (PCHIP) to interpolate the user-defined keyframes [Fritsch and Carlson 1980]. The resulting chord length parameter  $\theta$  describes progress on the spatial reference path defined as  $\mathbf{f}_d(\theta) = [\mathbf{r}_d(\theta), \psi_d(\theta), \phi_d(\theta)] \in \mathbb{R}^5$ . To prevent sudden changes of the progress parameter, we add  $\theta$  into our model and formulate its dynamics with the following linear discrete system equation:

$$\Theta_{i+1} = C\Theta_i + Dv_i, \quad 0 \leq v_i \leq v_{max}, \quad (2)$$

where  $\Theta_i = [\theta_i, \dot{\theta}_i]$  is the state and  $v_i$  is the input of  $\theta$  at step  $i$  and  $C \in \mathbb{R}^{2 \times 2}$ ,  $D \in \mathbb{R}^{2 \times 1}$  are the discrete system matrices. Intuitively,  $v_i$  approximates the quadrotor's acceleration as  $\theta$  is an approximation of the trajectory length.

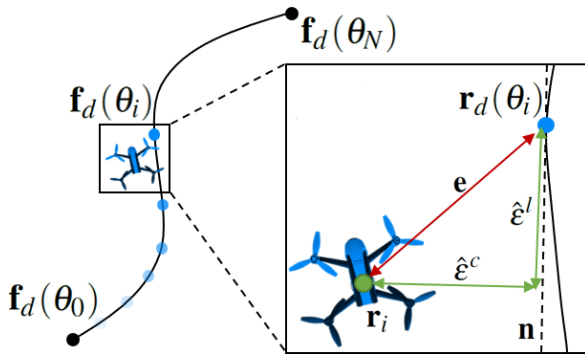


Fig. 2. The position (x,y,z) and orientation (yaw, pitch) over time are a function of path progress  $\mathbf{f}_d(\theta)$ . Inset: to advance along the path we optimize for smooth progress  $\theta$  via minimization of lag  $\epsilon^l$  and contour  $\epsilon^c$  error.

With this extension of the dynamic model in place, we now formulate an objective to minimize the error between the desired quadrotor position  $\mathbf{r}_d(\theta)$  and the current quadrotor position  $\mathbf{r}$ . With respect to the time optimization we want the quadrotor to follow  $\mathbf{r}_d(\theta_i)$  as closely as possible in time (no lag) but allow deviations from its contour for smoother motion. This distinction is not possible when

minimizing the 2-norm distance to the reference point. For this reason, we differentiate between a lag  $\epsilon^l$  and a contour error  $\epsilon^c$  similar to MPCC approaches (e.g., [Lam et al. 2013]). We approximate the true error from the spline by using the 3D-space approximation of lag  $\epsilon^l$  and a contour error  $\epsilon^c$  of [Nägeli et al. 2017] (see Figure 2, inset). The approximated lag error is defined as,

$$\hat{\epsilon}^l(\theta, \mathbf{r}_i) = \mathbf{e}^T \mathbf{n} \quad (3)$$

where  $\mathbf{e} = \mathbf{r}_d(\theta) - \mathbf{r}_i$  is the relative vector between desired and actual positions and  $\mathbf{n} = \frac{\dot{\mathbf{r}}_d(\theta)}{\|\dot{\mathbf{r}}_d(\theta)\|}$  is the normalized tangent vector of  $\mathbf{r}_d(\theta)$  at  $\theta$ . The resulting contour error approximation is given by:

$$\hat{\epsilon}^c(\theta, \mathbf{r}_i) = \|\mathbf{e} - \hat{\epsilon}^l(\theta, \mathbf{r}_i)\mathbf{n}\|. \quad (4)$$

Both error terms are then inserted into the cost term,

$$c^p(\theta, \mathbf{r}_i) = \begin{bmatrix} \hat{\epsilon}^l(\theta) \\ \hat{\epsilon}^c(\theta) \end{bmatrix}^T Q \begin{bmatrix} \hat{\epsilon}^l(\theta) \\ \hat{\epsilon}^c(\theta) \end{bmatrix}, \quad (5)$$

where  $Q$  is a diagonal positive definite weight matrix. Minimizing  $c^p$  will move the quadrotor along the user defined spatial reference.

Our experiments have shown that distinguishing between lag and contour error is important for the temporal aspects of the optimization. Trajectories generated by minimizing  $\|\mathbf{e}\|^2$ , depending on the weighting of the term, either lag behind the temporal reference or cannot trade-off positional fit for smoother motion. With appropriate weights for lag and contour error this behavior is avoided.

To give users fine grained control over the target framing we follow user-specified viewing angles in an analogous fashion. To attain the camera yaw and pitch we minimize the 2-norm discrepancy between desired and actual orientation of the quadrotor and camera gimbal. Given by the following cost terms:

$$c^\psi(\theta, \psi_{q,i}, \psi_{g,i}) = \|\psi_d(\theta) - (\psi_{q,i} + \psi_{g,i})\|^2 \quad (6)$$

$$c^\phi(\theta, \phi_{q,i}, \phi_{g,i}) = \|\phi_d(\theta) - \phi_{g,i}\|^2, \quad (7)$$

where  $\psi_{q,i}, \psi_{g,i}, \phi_{q,i}, \phi_{g,i}$  are the current yaw and pitch angles. Furthermore, we preprocess every keyframe by adding a multiple of  $2\pi$  to yaw and pitch such that the absolute distance to the respective angle of the previous keyframe has the smallest value.

### 3.4 Smooth Progress

For the camera to smoothly follow the path, we need to ensure that  $\theta$  progresses. By specifying an initial  $\theta_0$  and demanding  $\theta$  to reach the end of the trajectory in the terminal state  $\theta_N$ , the progress of  $\theta$  can be forced with an implicit cost term. We simply penalize the trajectory end time by minimizing the state space variable  $T$ ,

$$c^{end}(T) = T. \quad (8)$$

Minimizing the end-time can be interpreted as optimizing trajectories to be as short as possible temporally (while respecting smoothness and limits of the robot model). This forces  $\theta$  to make progress such that the terminal state  $\theta_N$  is reached within time  $T^1$ .

To ensure that the generated motion for the quadrotor is smooth, we introduce a cost term on the model's jerk states,

$$c^j(\ddot{\mathbf{r}}, \ddot{\psi}_q, \ddot{\psi}_g, \ddot{\phi}_g) = \|\ddot{\mathbf{j}}\|^2, \quad (9)$$

<sup>1</sup>This also prevents solutions of infinitely long trajectories in time where adding steps with  $\mathbf{u}_i \approx 0$  is free wrt. to Eq. (9)).



where  $\mathbf{j}_i = [\ddot{\mathbf{r}}, \ddot{\psi}_q, \ddot{\psi}_g, \ddot{\phi}_g]^T$  is jerk and angular jerk. We minimize jerk since it provides a commonly used metric to quantify smoothness [Hogan 1984] and is known to be a decisive factor for the aesthetic perception of motion [Bronner and Shippen 2015]. This cost term again implicitly effects  $\theta$  by only allowing it to progress such that the quadrotor motion following the reference path  $\mathbf{f}_d(\theta)$  is smooth according to (9). This is illustrated in Figure 2, left. The blue dot ( $\theta_i$ ) progresses on the reference path such that the generated motion of the quadrotor following  $\mathbf{f}_d(\theta_i)$  is smooth.

To still be able to specify the temporal length of a video shot with this formulation, we define the cost term,

$$c^{len}(N, \Delta t) = \|t_{len} - T\|^2, \quad (10)$$

where we minimize the 2-norm discrepancy between the trajectory end time  $T$  and a user-specified video length  $t_{len}$ . In case a trajectory is optimized for Eq. (10), the weight for Eq. (8) is set to zero.

### 3.5 Optimization Problem

We construct our overall objective function by linearly combining the cost terms from Eq. (5), (6), (7), (8), (9), (10) and a 2-norm minimization of  $v$ . The final cost is:

$$J_i = w_p c^p(\theta_i, \mathbf{r}_i) + w_\psi c^\psi(\theta_i, \psi_{q,i}, \psi_{g,i}) + w_\phi c^\phi(\theta_i, \phi_{g,i}) \quad (11) \\ + w_j c^j(\ddot{\mathbf{r}}, \ddot{\psi}_q, \ddot{\psi}_g, \ddot{\phi}_g) + w_{end} c^{end}(T) + w_{len} c^{len}(N, \Delta t) + w_v \|v\|^2,$$

where the scalar weight parameters  $w_p, w_\psi, w_\phi, w_j, w_{end}, w_{len}, w_v > 0$  are adjusted for a good trade-off between positional fit and smoothness. The final optimization problem is then:

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{u}, \Theta, v}{\text{minimize}} \quad \sum_{i=0}^N J_i & (12) \\ & \text{subject to } \mathbf{x}_0 = \mathbf{k}_0 & \text{(initial state)} \\ & \quad \Theta_0 = \mathbf{0} & \text{(initial progress)} \\ & \quad \Theta_N = \mathbf{L} & \text{(terminal progress)} \\ & \quad \mathbf{x}_{i+1} = \mathbf{A}\mathbf{x}_i + \mathbf{B}\mathbf{u}_i + \mathbf{g} & \text{(dynamical model)} \\ & \quad \Theta_{i+1} = \mathbf{C}\Theta_i + Dv_i & \text{(progress model)} \\ & \quad \mathbf{x}_{min} \leq \mathbf{x}_i \leq \mathbf{x}_{max}, & \text{(state bounds)} \\ & \quad \mathbf{u}_{min} \leq \mathbf{u}_i \leq \mathbf{u}_{max}, & \text{(input limits)} \\ & \quad 0 \leq \Theta_i \leq \Theta_{max} & \text{(progress bounds)} \\ & \quad 0 \leq v_i \leq v_{max} & \text{(progress input limits),} \end{aligned}$$

where  $J_i$  is quadratic in  $\mathbf{x}, \mathbf{u}, v$  and linear in  $\Theta$ . When flying a generated trajectory we follow the optimized positional trajectory  $\mathbf{r}$  with a standard LQR-controller and use velocity and accelerations states of  $\mathbf{x}$  as feed-forward terms.

## 4 IMPLEMENTATION

We implemented the above optimization problem with MATLAB and solve it with the FORCES Pro software [Domahidi and Jerez 2017] which generates fast solver code, exploiting the special structure in the non-linear program. We set the horizon length of our problem to be  $N = 60$ . The solver requires a continuous path parametrization. To attain a description of the reference spline across the piecewise sections of the PCHIP spline, we need to locally approximate it. Therefore, we implemented an iterative programming scheme able

to generate trajectories at interactive rates. For further details on the IP-scheme and the empirically derived weights of the optimization problem, we refer the interested reader to Appendix B.

## 5 TECHNICAL EVALUATION

To evaluate the efficacy of our method in creating smooth camera motion even on problematic (for hard-constrained methods) inputs, we designed a challenging shot and generated two trajectories, one with [Gebhardt et al. 2016] and the other one with our method.

Figure 3 plots the resulting positions in  $x, y, z$  and the corresponding camera angles. Our method adjusts the timing of the original keyframes ( $k_2$ - $k_4$ ) to attain smoother motion over time. This is visible when comparing the  $x$ -dimension of ours and [Gebhardt et al. 2016]. The need to trade-off timing and spatial location is illustrated by the orientation plot (Figure 3, bottom). The keyframes  $k_2, k_3$  have been moved very close to each other which would cause excessive yaw velocities since the quadrotor would need to perform a  $180^\circ$  turn. Since our method trades-off the positional fit it generates smooth motion also for the camera orientation.

We also conducted a qualitative comparison by recording different videos with the same consumer grade drone. The quadrotor followed trajectories generated with our method and with [Gebhardt et al. 2016] using the same input. Figure 4 shows resulting video frames and jerk profiles (also see [video]). Although the timing of keyframes was improved for smoothness, our method still generates trajectories with lower magnitudes of positional jerk and less variation in angular jerk.

To assess quantitatively that our method generates smoother camera motion, we compare the averaged squared jerk per horizon stage of user-designed trajectories generated with our method, with [Joubert et al. 2015] and with [Gebhardt et al. 2016]. Figure 5 shows

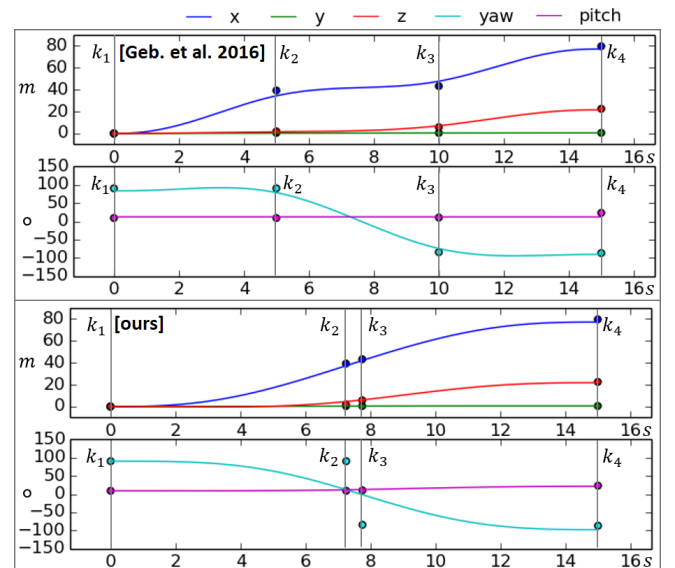


Fig. 3. Position ( $x, y, z$  in  $m$ ) and orientation (yaw, pitch in  $^\circ$ ) over time (in  $s$ ) for the same user-specified keyframes ( $k_1$ - $k_4$ ) for [Gebhardt et al. 2016] (top) and our method (bottom).

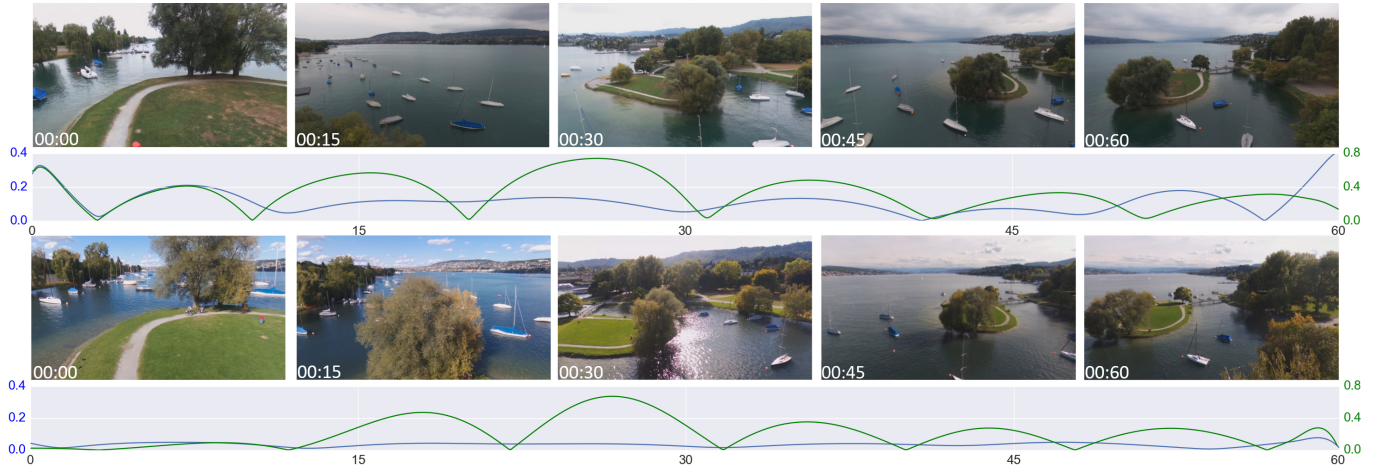


Fig. 4. Qualitative comparison of video frames as well as **jerk** (in  $\frac{m}{s^3}$ ) and **angular jerk** (in  $\frac{^\circ}{s^3}$ ) profiles of two trajectories generated with [Gebhardt et al. 2016] (top row) and our method (bottom row).

lower jerk and angular jerk values for our optimization scheme compared to both baseline methods, across all trajectories.

Finally, we evaluate the optimization runtime of our method. Therefore, we generated trajectories from the studies of [Gebhardt and Hilliges 2018; Joubert et al. 2015] using the approximated linear quadrotor model of Sec. 3.1 and the non-linear model of [Nägeli et al. 2017]. We measured runtime on a standard desktop machine (Intel Core i7 4GHz CPU, Forces Pro NLP-solver). The computation time for the trajectories are shown in Figure 5. In average, it took 2.41 s (SD = 2.50 s) to generate a trajectory with the linear model and 14.79 s (SD = 15.50 s) with the non-linear model.

## 6 PERCEPTUAL STUDY

Our technical evaluation shows that the proposed method generates smoother trajectories. However, it has not been validated that the trajectories generated with our method result in aesthetically more pleasing video. To this end, we conduct an online survey comparing videos which follow user-specified timings, generated with the methods of [Gebhardt et al. 2016; Joubert et al. 2015], with videos

generated by our method. Therefore, we compare user-designed trajectories from prior work [Gebhardt and Hilliges 2018; Joubert et al. 2015]. For each question we take the user-specified keyframes of the original trajectory and generated a time-optimized trajectory of the same temporal duration (via Equation 10) using our method. We then render videos for the original and time-optimized trajectory using Google Earth (based on GPS-coordinates and camera angles). The two resulting videos are placed side-by-side, randomly assigned to the left or right, and participants state which video they prefer on a forced alternative choice 5-point Likert scale. The five responses are: "shot on the left side looks much more pleasing", "shot on the left side looks more pleasing", "both the same", "shot on the right side looks more pleasing", and "shot on the right side looks much more pleasing". Each participant had to compare 14 videos.

### 6.1 Results

In total, 424 participants answered the online survey. Assuming equidistant intervals, we mapped survey responses onto a scale from -2 to 2, where negative values mean that the original, timed video is aesthetically more pleasing, 0 indicates no difference and a positive value indicates a more aesthetically pleasing time-optimized video. In order to attain interval data, our samples are build by taking the mean of the Likert-type results of the expert and non-expert designed videos per participant. Visual inspection of residual plots did not reveal any obvious deviations from normality.

Evaluating *all* responses of the survey, we try to attain a mean which compensates random participant effects. Therefore, we construct a linear mixed model using the participant as random intercept, the video as fixed intercept and specify a contrast such that the additionally introduced fixed-effect intercept represents the adjusted mean of the data. This overall mean has a positive value with a high confidence (see Figure 6). A type III ANOVA showed that there is a significant difference between zero (no effect) and the adjusted mean ( $F(1, 856) = 515.4, p < 0.001$ ). Unpacking this result further, we distinguish between videos that have been designed by

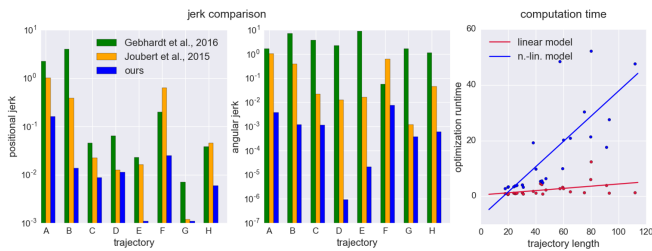


Fig. 5. Left: comparing avg. squared jerk (in  $\frac{m^2}{s^3}$ ) and angular jerk (in  $\frac{^\circ^2}{s^3}$ ) per horizon stage of different trajectories for our method, [Gebhardt et al. 2016] and [Joubert et al. 2015] (note that latter uses a different model). Right: our method's optimization runtime for different trajectories is plotted against their temporal length (both in sec). We differentiate between using a linear and a non-linear quadrotor model for trajectory generation.

*non-expert* users (data from [Gebhardt and Hilliges 2018]) and *expert* users (i.e. cinematographers, data from [Joubert et al. 2015]). Analyzing the results for significance, we perform a one sample t-test on the averaged Likert ratings for *expert*- and *non-expert*-designed videos. The effect of both conditions and their confidence intervals are shown in Figure 6. While they are significant for both conditions (*expert*:  $t(423), p < 0.001$ ; *non-expert*:  $t(423), p < 0.001$ ), the effect is positive and amplified for *non-expert* designed videos and negative for *expert* designed videos.

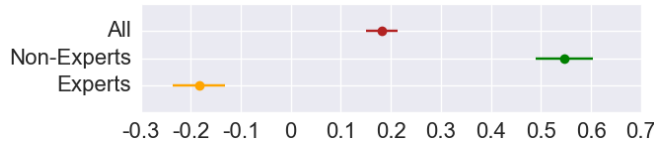


Fig. 6. Mean and 95% confidence interval of the effect of optimization scheme on the preference of *all*, *non-expert* designed and *expert* designed videos. Positive values indicate a preference for videos generated with our method. Negative values indicate that videos generated with the methods of [Gebhardt et al. 2016; Joubert et al. 2015] are preferred.

## 6.2 Discussion

The perceptual study provides strong evidence that our method has a positive effect on the aesthetic perception of aerial videos. Furthermore, it has shown that this effect is even stronger for videos by *non-experts*. This supports our hypothesis that *non-experts* benefit from generating trajectories according to global smoothness as main criteria. Looking at *expert* created videos the picture is different. These videos were rated as more pleasant when generated with methods which respect user-specified timings. This can be explained by the fact that *experts* explicitly leverage shot timings to create particular compositional effects. Optimizing for global smoothness removes this intention from the result. However, the significant positive effect of our method on *all* responses and a larger effect size for the positive effect of *non-expert*- compared to the negative effect of *expert* designed videos indicate that smooth motion is a more important factor for the aesthetic perception of aerial videos than timing. This suggests that users, especially experts, could benefit from a problem formulation which allows for soft-constrained instead of hard-constrained timings. In this way, users could still employ shot timings to create compositional effects, while the optimization scheme generates trajectories trading-off user-specified timings and global smoothness.

Based on these results, we formulate three requirements for quadrotor camera generation schemes: 1) smoothness should be the primary objective of quadrotor camera trajectory generation, 2) methods should auto generate or adjust keyframe timings to better support non-experts, 3) while providing tools for experts to specify soft-constrained timings. The proposed method already full-fills requirement 1) and 2). In the next section, we propose how our method can be extended such that *all* requirements are met.

## 7 METHOD EXTENSIONS

Recognizing the need to provide both global smoothness and explicit user control over camera timings, we present two method extensions

to control camera motion: an approach based on "classic" keyframe timings and a further approach based on velocity profiles.

### 7.1 Keyframe Timings

We augment our objective function with an additional term for soft-constraint keyframe timings. The original formulation does not allow for the setting of timing references based on horizon stages: due to the variable horizon we lack a fixed mapping between time and stage. To be able to map timings with the spatial reference, we use the  $\theta$ -parameterization of the reference spline. Reference timings hence need to be specified strictly monotonically increasing in  $\theta$ . Based on the reference timings and the corresponding  $\theta$ -values we interpolate a spline through these points, which results in timing reference function  $t_d(\theta)$  which can be followed analogously to spatial references by minimizing the cost,

$$c^t(\theta, i, \Delta t) = \|t_d(\theta) - (i * \Delta t)\|^2, \quad (13)$$

where  $i$  is the current stage of the horizon and  $\Delta t$  is the current discretization of the model. We add this cost into (11) and assign a weight to specify its importance  $w_t$ . By setting the value of  $w_t$  to a very large number, quasi hard-constrained keyframes are attainable.

### 7.2 Reference Velocities

The above extension enables mimicry of timing control in prior methods. However, the actual purpose of specifying camera timings in a video is to control or change camera velocity to achieve a desired effect (recall the fly-by example). Since determining the timing of the shot explicitly is difficult, we propose a way for users to directly specify camera velocities. We extend the formulation of our method to accept reference velocities as input. Again, we use the  $\theta$ -parameterization to assign velocities to the reference spline  $\mathbf{f}_d$ . To minimize the difference between the velocity of the quadrotor and the user-specified velocity profile  $v_d(\theta)$ , we specify the cost,

$$c^v(\theta, i, \Delta t) = \|v_d(\theta) - \dot{\mathbf{r}}_i^T \mathbf{n}\|^2, \quad (14)$$

where we project the current velocity of the quadrotor  $\dot{\mathbf{r}}_i$  on the normalized tangent vector of the positional reference function  $\mathbf{n}$ . We add this cost term and a weight  $w_v$  to (11).

## 8 USER STUDY

To understand whether our final method has the potential to improve the usability of quadrotor camera tools, whether soft-constrained timing methods produce videos of similar perceived aesthetics then hard-constrained timing methods and whether experts can benefit from our method, we conduct an additional user study. In this study, we compare different variants of our method with the method of [Gebhardt et al. 2016] as representative for quadrotor camera optimization schemes which use hard-constrained keyframe timings.

### 8.1 Experimental Design

*User Interface:* In our experiment we used the tool of [Gebhardt and Hilliges 2018] and extended the UI with a toolbar. This toolbar contains a slider to specify  $w_p$  (see Equation 11) and depending on the condition, a progress curve or a velocity profile. A progress curve allows for the editing of the relative progress on a trajectory



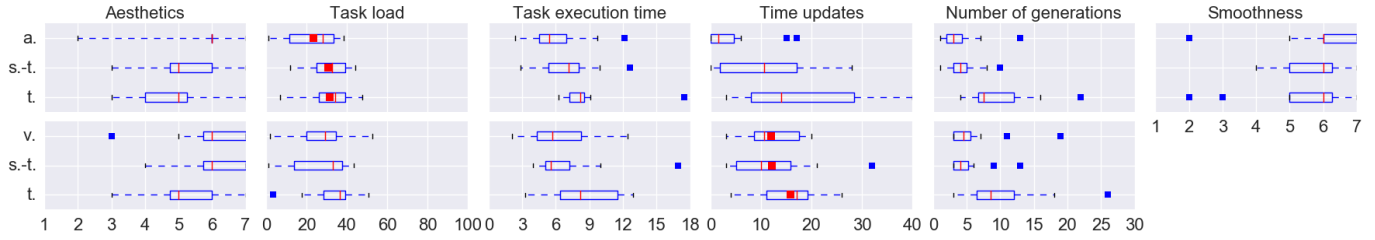


Fig. 7. Boxplots for the results of the user study (T1 on upper row, T2 on lower row). The investigated conditions are *auto* (a.), *velocity* (v.), *soft-timed* (s.-t.), and *timed* (t.). Task execution time is depicted in minutes. Aesthetics, smoothness and task load are shown in the respective scales of the questionnaire items. Time updates and number of generations are counts. In case the data of a measure is normally distributed the mean is displayed (red box).

over time (see Figure 8, a). A velocity profile enables editing of the camera speed over the progress on the trajectory (see Figure 8, b).

**Experimental conditions:** We investigate four different conditions: 1) In *timed*, participants work with the optimization method of [Gebhardt et al. 2016] and a progress curve (see Figure 8, a). 2) *Soft-timed* uses our optimizer and the progress curve. Participants can decide whether they want to specify keyframe timings (see Equation 13) or use the auto-generated timings. 3) In *auto* participants work with our optimization and the keyframe timings it provides. They can choose to fix the end time of a trajectory (see Equation 10). 4) *Velocity* uses our method and a velocity profile (see Figure 8, b). Participants can decide whether they want to specify camera velocity (see Equation 13) or use the auto-generated speed.

**Tasks:** The study comprises two tasks: 1) Participants were asked to design a free-form video of a building in a virtual environment (T1). We asked participants to keep the spatial trajectory as similar as possible across conditions whereas the dynamics of camera motion were allowed to differ. They performed the task in the conditions *timed*, *soft-timed* and *auto*. 2) Participants were asked to faithfully reproduce an aerial video shot with varying camera velocity (T2). Participants should try to reproduce camera path and dynamics of the reference video. This task was performed with the conditions *timed*, *soft-timed* and *velocity* to investigate the level of control over timing afforded by the different conditions. We use a within-subjects design and counterbalance order of conditions within a task to compensate for learning effects.

**Procedure:** Participants were introduced to the system and the four conditions and were given time to practice using the tool in a

tutorial tasks. Participants then solved T1 and T2 in the respective conditions. Tasks were completed when participants reported to be satisfied with the designed shot (T1) or the similarity to the reference (T2). For each task and condition participants completed the NASA-TLX and a questionnaire on levels of satisfaction with the result. Finally, a short exit interview was conducted. A session took on average approximately 70 min (introduction  $\approx$  9 min, tutorial  $\approx$  7 min, T1  $\approx$  22 min, T2  $\approx$  23 min).

**Participants:** We recruited 12 participants (5 female, 7 male). We purposely included 3 experts: an avid hobby quadrotor videographer, a professional videographer, experimenting with quadrotor videography in his free-time, and a professional quadrotor videographer. The remaining participants reported no experience in aerial or normal photo- or videography.

## 8.2 Results

We analyze the effect of the conditions on the usability of the tool and the aesthetics of the resulting videos. For significance testing, we ran a one-way ANOVA if the normality assumption holds and a Kruskal-Wallis test when it is violated. Analyzing the data of experts and non-experts separately, we found no significant differences in results and thus will not differentiate between them in this section.

**Usability.** To assess the effect of our method on the usability of the tool, we asked participants to fill out NASA-TLX and collected interaction logs (e.g. task execution time). In T1, *auto* has the lowest median in terms of task load, followed by *soft-timed* and *timed* (see Figure 7). This ranking remains the same for all interaction logging measures of T1 (see task execution time (TET), time updates and number of generations). Although there is no significant effect of conditions in T1 on task load ( $F(2, 33) = 1.78, p = 0.18$ ), the other measures do differ significantly (task execution time:  $H(2) = 7.38, p < 0.03$ ; time updates:  $H(2) = 10.45, p < 0.01$ ; number of generations:  $H(2) = 13.93, p < 0.01$ ). Pairwise comparison indicates that for TET and time updates *auto* is significantly different to *timed* (TET:  $p < 0.01$ ; time updates:  $p < 0.01$ ). For number of generations, *auto* and *soft-timed* significantly differ to *timed* (*auto-timed*:  $p < 0.01$ , *soft-timed-timed*:  $p < 0.05$ ). Other differences are not significant. *Auto* automatically generates timings and thereby camera velocities. This explains the condition's first rank in terms of task load and interaction logs as it simplifies the task drastically. For T2, *velocity* and *soft-timed* yield

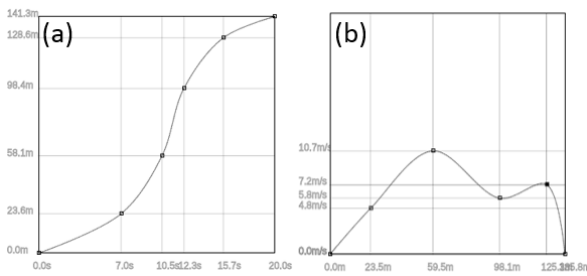


Fig. 8. Progress curve (a) and velocity profile (b).



a lower task load compared to *timed*, indicating a slight advantage of our method in terms of usability (differences are not significant:  $H(2) = 1.73, p = 0.42$ ). This ranking is confirmed by interaction logs where *soft-timed* and *velocity* perform similar and are followed by *timed*. The number of generations between conditions differs statistically significantly ( $H(2) = 8.04, p < 0.02$ ). A pairwise comparison indicates that *velocity* and *soft-timed* significantly differ from *timed* (*velocity-timed*:  $p < 0.03$ ; *soft-timed-timed*:  $p < 0.01$ ). Other differences are not significant (TET:  $H(2) = 2.49, p = 0.29$ ; time updates:  $F(2, 33) = 1.16, p = 0.33$ ).

**Aesthetics.** We are also interested in participants' perceived difference in aesthetics of the generated videos. We asked participants in both tasks to assess the visual quality of the video they designed on a scale ranging from 1 (not at all pleasing) to 7 (very pleasing, see Figure 7). Although differences are not significant (aesthetics in T1:  $H(2) = 4.694, p = 0.096$ ; aesthetics in T2:  $H(2) = 3.589, p = 0.166$ ), the variants of our method are perceived to produce aesthetically more pleasing videos in both tasks. For T1, we also asked users to rate the smoothness of videos on a scale from 1 (non-smooth) to 7 (very smooth). Figure 7 summarizes the results which do not differ significantly between conditions ( $H(2) = 1.828, p = 0.401$ ).

### 8.3 Discussion

Despite the small sample size of our experiment, the results indicate a positive effect of our method on both, the perceived aesthetics of results and the usability of the tool. *Auto* caused the lowest task load among conditions and participants were satisfied with the generated results. Although *soft-timed* and *timed* allow to specify the timing of a shot in the same manner (using the progress-curve or the timeline), *soft-timed* performed better than *timed* in terms of task load (T2) and aesthetics (T1 and T2). We think that this preference can be explained by two factors. First, participants in *soft-timed* generally used a workflow in which they initially take generated timings and then adjust keyframe times to create a desired visual. This workflow was implemented by experts but also by non-experts (if they used keyframe timings at all). Second, in *soft-timed* keyframe timings are specified as soft-constraints, allowing the optimizer to trade-off the temporal fit for a smoother trajectory. This makes *soft-timed* more forgiving than *timed* wrt to the space-time-ratio in-between keyframes, reducing adjustments participants had to do in order to solve a task in this condition (see time/velocity updates and no. of generations in Figure 7). In addition, soft-constrained timings allow the optimizer to still generate feasible trajectories even if the underlying user input would not yield a feasible result.

The preference for soft-constrained keyframe timings is also an indication for our general assumption that timing control is not used to precisely specify the time a camera should be at a certain position. Instead users employ timing to specify the velocity of the camera along the path. This is also suggested by looking at the results of the *velocity* condition. In T2, it performed similar to *soft-timed* and better than *timed* for task load and aesthetics, indicating that specifying camera dynamics via a velocity profile is an intuitive alternative for providing keyframe timings.

## 9 CONCLUSION

In this paper, we addressed the dichotomy of smoothness and timing control in current quadrotor camera tools. According to design requirements in literature [Joubert et al. 2015] their optimizers incorporate keyframes timing as hard constraints, providing precise timing control. A recent study [Gebhardt and Hilliges 2018] has shown that this causes users to struggle when specifying smooth camera motion over an entire trajectory. The current optimization formulations need to have matching distances between the 5 dimensions of a keyframe (position, yaw and pitch of camera angle) with its temporal position. This poses a particular hard interaction problem for users, especially novices. In this paper, we propose a method which generates smooth quadrotor camera trajectories by taking keyframes only specified in space and optimizing their timings. We formulated this non-linear problem as a variable horizon trajectory optimization scheme which is capable of temporally optimizing positional references.

In a large-scale online survey we compared videos generated with our method to videos generated with [Gebhardt et al. 2016] and [Joubert et al. 2015]. The results indicate a general preference for videos generated according to a global smoothness objective, but also highlight that videos of experts are aesthetically more pleasing when provided timing control. Based on these insights, we extend our method such that users can specify keyframe timings as soft-constraints but still globally smooth trajectories are attained. In addition, we allow users to specify camera reference velocities set as soft-constraints in the optimization.

We test the efficacy and usability of our optimization in a comparative user study (baseline is [Gebhardt et al. 2016]). The results indicate that our method positively affects the usability of quadrotor camera tools and improves the visual quality of video shots for experts and non-experts. Both benefit from using an optimized timing initially and having the possibility of fine-tuning it according to their intention. In addition, the user study revealed that timing control does not need to be precise but is rather used to control camera velocity in order to create a desired compositional effect.

## ACKNOWLEDGMENTS

We thank Yi Hao Ng for his work in the exploratory phase of the project, Chat Wacharamanatham for helping with the statistical analysis of the perceptual study and Velko Vechev for providing the video voice-over. We are also grateful for the valuable feedback of Tobias Nægeli on problem formulation and implementation. This work was funded in parts by the Swiss National Science Foundation (UFO 200021L\_153644).

## REFERENCES

- 3D Robotics. 2015. 3DR Solo. (2015). Retrieved September 13, 2016 from <http://3drobotics.com/solo>
- A. Pedro Aguiar, Joao P. Hespanha, and Petar V. Kokotovic. 2008. Performance limitations in reference tracking and path following for nonlinear systems. *Automatica* 44, 3 (2008), 598 – 610. <https://doi.org/10.1016/j.automatica.2007.06.030>
- APM. 2016. APM Autopilot Suite. (2016). Retrieved September 13, 2016 from <http://ardupilot.com>
- Daniel Arijon. 1976. Grammar of the film language. (1976).
- Ty Audronis. 2014. How to Get Cinematic Drone Shots. (2014). Retrieved August 29, 2017 from <https://www.videomaker.com/article/c6/17123-how-to-get-cinematic-drone-shots>

- Sivakumar Balasubramanian, Alejandro Melendez-Calderon, Agnes Roby-Brami, and Etienne Burdet. 2015. On the analysis of movement smoothness. *Journal of neuroengineering and rehabilitation* 12, 1 (2015), 112. <https://doi.org/10.1186/10.1186/s12984-015-0090-9>
- John T. Betts. 2009. *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming* (2nd ed.). Cambridge University Press, New York, NY, USA.
- Shaw Bronner and James Shippen. 2015. Biomechanical metrics of aesthetic perception in dance. *Experimental Brain Research* 233, 12 (01 Dec 2015), 3565–3581. <https://doi.org/10.1007/s00221-015-4424-4>
- Adam Bry, Charles Richter, Abraham Bachrach, and Nicholas Roy. 2015. Aggressive flight of fixed-wing and quadrotor aircraft in dense indoor environments. *The International Journal of Robotics Research* 34, 7 (2015), 969–1002. <https://doi.org/10.1177/0278364914558129>
- Marc Christie, Patrick Olivier, and Jean-Marie Normand. 2008. Camera Control in Computer Graphics. *Computer Graphics Forum* 27, 8 (Dec. 2008), 2197–2218. <https://doi.org/10.1145/1665817.1665820>
- DJI. 2016. PC Ground Station. (2016). Retrieved September 13, 2016 from <http://www.dji.com/pc-ground-station>
- Alexander Domahidi and Juan Jerez. 2017. FORCES Pro: code generation for embedded optimization. (2017). Retrieved September 4, 2017 from <https://www.embotech.com/FORCES-Pro>
- Steven M. Drucker and David Zeltzer. 1994. Intelligent Camera Control in a Virtual Environment. In *In Proceedings of Graphics Interface '94*. 190–199.
- T. Faulwasser, B. Kern, and R. Findeisen. 2009. Model predictive path-following for constrained nonlinear systems. In *Proceedings of the 48th IEEE Conference on Decision and Control (CDC)*. 8642–8647. <https://doi.org/10.1109/CDC.2009.5399744>
- F. N. Fritsch and R. E. Carlson. 1980. Monotone Piecewise Cubic Interpolation. *SIAM J. Numer. Anal.* 17, 2 (1980), 238–246. <https://doi.org/10.1137/0717021> arXiv:<https://doi.org/10.1137/0717021>
- Q. Galvane, J. Fleureau, F. L. Tariolle, and P. Guillotel. 2016. Automated Cinematography with Unmanned Aerial Vehicles. In *Proceedings of the Eurographics Workshop on Intelligent Cinematography and Editing (WICED '16)*. Eurographics Association, Goslar Germany, Germany, 23–30. <https://doi.org/10.2312/wiced.20161097>
- Christoph Gebhardt, Benjamin Hepp, Tobias Nägeli, Stefan Stevšić, and Otmar Hilliges. 2016. 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 (CHI '16)*. ACM, New York, NY, USA, 2508–2519. <https://doi.org/10.1145/2858036.2858353>
- Christoph Gebhardt and Otmar Hilliges. 2018. WYFIWYG: Investigating Effective User Support in Aerial Videography. (2018). arXiv:1801.05972
- T. Geijtenbeek and N. Pronost. 2012. Interactive Character Animation Using Simulated Physics: A State-of-the-Art Review. *Computer Graphics Forum* 31, 8 (2012), 2492–2515. <https://doi.org/10.1111/j.1467-8659.2012.03189.x>
- John Hennessy. 2015. 13 Powerful Tips to Improve Your Aerial Cinematography. (2015). Retrieved August 29, 2017 from <https://skytango.com/13-powerful-tips-to-improve-your-aerial-cinematography/>
- Neville Hogan. 1984. Adaptive control of mechanical impedance by coactivation of antagonist muscles. *IEEE Trans. Automat. Control* 29, 8 (1984), 681–690.
- Niels Joubert, Dan B Goldman, Floraine Berthouzoz, Mike Roberts, James A Landay, Pat Hanrahan, et al. 2016. Towards a Drone Cinematographer: Guiding Quadrotor Cameras using Visual Composition Principles. (2016). arXiv:1610.01691
- Niels Joubert, Mike Roberts, Anh Truong, Floraine Berthouzoz, and Pat Hanrahan. 2015. An Interactive Tool for Designing Quadrotor Camera Shots. *ACM Trans. Graph.* 34, 6, Article 238, 11 pages. <https://doi.org/10.1145/2816795.2818106>
- Denise Lam, Chris Manzie, and Malcolm C. Good. 2013. Multi-axis model predictive contouring control. *Internat. J. Control* 86, 8 (2013), 1410–1424. <https://doi.org/10.1080/00207179.2013.770170>
- Tsai-Yen Li and Chung-Chiang Cheng. 2008. *Real-Time Camera Planning for Navigation in Virtual Environments*. Springer Berlin Heidelberg, Berlin, Heidelberg, 118–129. [https://doi.org/10.1007/978-3-540-85412-8\\_11](https://doi.org/10.1007/978-3-540-85412-8_11)
- Alexander Liniger, Alexander Domahidi, and Manfred Morari. 2014. Optimization-based autonomous racing of 1:43 scale RC cars. *Optimal Control Applications and Methods* (2014). <https://doi.org/10.1002/oca.2123>
- Christophe Lino and Marc Christie. 2012. Efficient Composition for Virtual Camera Control. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '12)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 65–70. <https://doi.org/10.1145/1409060.1409068>
- Christophe Lino and Marc Christie. 2015. Intuitive and Efficient Camera Control with the Toric Space. *ACM Trans. Graph.* 34, 4, Article 82 (July 2015), 12 pages. <https://doi.org/10.1145/2766965>
- Christophe Lino, Marc Christie, Roberto Ranon, and William Bares. 2011. The Director's Lens: An Intelligent Assistant for Virtual Cinematography. In *Proceedings of the 19th ACM International Conference on Multimedia (MM '11)*. ACM, New York, NY, USA, 323–332. <https://doi.org/10.1145/2072298.2072341>
- C. Karen Liu, Aaron Hertzmann, and Zoran Popović. 2006. Composition of Complex Optimal Multi-character Motions. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '06)*. Eurographics Association, Aire-la-Ville, Switzerland, 215–222. <http://dl.acm.org/citation.cfm?id=1218064.1218093>
- James McCann, Nancy S Pollard, and Siddhartha Srinivasa. 2006. Physics-based motion retiming. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Eurographics Association, 205–214.
- D. Mellinger and V. Kumar. 2011. Minimum snap trajectory generation and control for quadrotors. In *2011 IEEE International Conference on Robotics and Automation*. 2520–2525. <https://doi.org/10.1109/ICRA.2011.5980409>
- D Mellinger, N Michael, and V Kumar. 2012. Trajectory Generation and Control for Precise Aggressive Maneuvers with Quadrotors. *The International Journal of Robotics Research* 31, 5 (2012), 664–674. <https://doi.org/10.1177/0278364911434236>
- H. Michalska and D. Q. Mayne. 1993. Robust receding horizon control of constrained nonlinear systems. *IEEE Trans. Automat. Control* 38, 11 (Nov 1993), 1623–1633. <https://doi.org/10.1109/9.262032>
- M.W. Mueller and R. D'Andrea. 2013. A model predictive controller for quadcopter state interception. In *Proceedings of the European Control Conference (ECC)*, 2013. 1383–1389. [https://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6669415](https://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6669415)
- T. Naegeli, 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* PP, 99 (2017), 1–1. <https://doi.org/10.1109/LRA.2017.2665693>
- Tobias Nägeli, Lukas Meier, Alexander Domahidi, Javier Alonso-Mora, and Otmar Hilliges. 2017. Real-time Planning for Automated Multi-view Drone Cinematography. *ACM Trans. Graph.* 36, 4, Article 132 (July 2017), 10 pages. <https://doi.org/10.1145/3072959.3073712>
- Mike Roberts and Pat Hanrahan. 2016. Generating Dynamically Feasible Trajectories for Quadrotor Cameras. *ACM Trans. Graph.* 35, 4, Article 61 (July 2016), 11 pages. <https://doi.org/10.1145/2897824.2925980>
- VC Technology. 2016. Litchi Tool. (2016). Retrieved September 13, 2016 from <https://flylitchi.com/>
- Andrew Witkin and Michael Kass. 1988. Spacetime Constraints. In *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '88)*. ACM, New York, NY, USA, 159–168. <https://doi.org/10.1145/54852.378507>
- I-Cheng Yeh, Chao-Hung Lin, Hung-Jen Chien, and Tong-Yee Lee. 2011. Efficient camera path planning algorithm for human motion overview. *Computer Animation and Virtual Worlds* 22, 2-3 (2011), 239–250. <https://doi.org/10.1002/cav.398>

## A NOTATION

For completeness and reproducibility of our method we provide a summary of the notation used in the paper in Table 1.

Symbol	Description
$\mathbf{r}, \dot{\mathbf{r}}, \ddot{\mathbf{r}}, \dddot{\mathbf{r}}$	Quadrotor position, velocity, acceleration and jerk
$\psi_q, \dot{\psi}_q, \ddot{\psi}_q, \dddot{\psi}_q$	Quad. yaw and angular velocity/acceleration/jerk
$\psi_g, \dot{\psi}_g, \ddot{\psi}_g, \dddot{\psi}_g$	Gimbal yaw and angular velocity/acceleration/jerk
$\phi_g, \dot{\phi}_g, \ddot{\phi}_g, \dddot{\phi}_g$	Gimbal pitch and angular velocity/acceleration/jerk
$\mathbf{x}, \mathbf{u}$	Quadrotor states and inputs
$A, B$	System matrices of quadrotor
$g$	Gravity
$T$	Trajectory end time
$N$	Horizon length
$\theta$	Progress parameter
$\mathbf{f}_d(\theta)$	Reference spline ( $\mathbb{R}^5$ )
$\mathbf{r}_d(\theta)$	Positional reference ( $\mathbb{R}^3$ )
$\psi_d(\theta)$	Pitch reference
$\phi_d(\theta)$	Yaw reference
$t_d(\theta)$	Time reference
$v_d(\theta)$	Velocity reference
$\Theta, v$	Progress state and input
$C, D$	System matrices of progress
$\hat{\epsilon}^l, \hat{\epsilon}^c$	Approximate lag and contour error

Table 1. Summary of notation used in the body of the paper

## B IMPLEMENTATION DETAILS

In this section, we provide details on the weights we use in the objective function, the iterative programming scheme we implemented to attain a continuous path parametrization and its performance.

### B.1 Optimization Weights

The values for the weights of the objective function we used in the user study and the online survey are listed in Table 2.

Weight (layed on)	Value	
	Online survey	User study
$w_p$ (position)	1	[0.1, 10] (user-specified)
$Q$ (lag, contour err.)	$diag(2, 1)$	$diag(2, 1)$
$w_\psi$ (yaw)	1	[0.1, 10] (user-specified)
$w_\phi$ (pitch)	1	[0.1, 10] (user-specified)
$w_j$ (jerk)	10	100, 10 (if $w_t, w_v > 0$ )
$w_T$ (end-time)	0	1, 0 (if $w_t, w_{end} > 0$ )
$w_{end}$ (length in t.)	1	1
$w_t$ (timing)	0	100
$w_v$ (velocity)	0	100

Table 2. Values for weights used in Equation 11.

### B.2 Iterative Programming Scheme

The solver requires a continuous path parametrization. To attain a description of the reference spline even in-between the piecewise sections of the PCHIP-spline, we need to locally approximate it. Therefore, we implement an iterative programming scheme where we compute a quadratic approximation of the reference spline around the  $\theta_i$ -value of each stage in the horizon. This process is described in Figure 9. In the first iteration of the scheme we initialize all  $\theta_i$  to zero and fit the entire reference trajectory (black spline) with a single quadratic approximation (blue spline). By solving the optimization problem of Equation 12, the progression of  $\theta$ -values will be decided based on the quadratic approximation (yellow dots). For the next iterations, we always take the value of  $\theta_i$  from the last run of the solver, project it on the reference spline (green dots) and fit a local quadratic approximation (red splines). Based on these fits the progress of  $\theta$ -values again is optimized. We stop the optimization when the difference of the  $\theta$ -values for all stages in-between iterations is smaller than a pre-defined threshold,  $\theta_{last} - \theta_{current} < \delta$ .

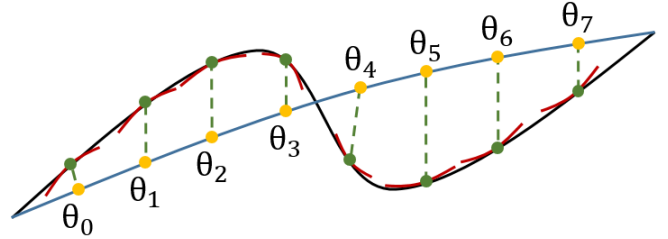


Fig. 9. Illustration of the iterative programming scheme, showing the reference path (black) as well as the quadratic approximations of the first (blue) and second (red) iteration with their respective  $\theta$ -values (yellow, green).