Chapter 17
Visual Motion Tracking and Sensor Fusion for Ground-Based Kite Power Systems

Henrik Hesse, Max Polzin, Tony A. Wood and Roy S. Smith

Abstract An estimation approach is presented for kite power systems with ground-based actuation and generation. Line-based estimation of the kite state, including position and heading, limits the achievable cycle efficiency of such airborne wind energy systems due to significant estimation delay and line sag. We propose a filtering scheme to fuse onboard inertial measurements with ground-based line data for ground-based systems in pumping operation. Estimates are computed using an extended Kalman filtering scheme with a sensor-driven kinematic process model which propagates and corrects for inertial sensor biases. We further propose a visual motion tracking approach to extract estimates of the kite position from ground-based video streams. The approach combines accurate object detection with fast motion tracking to ensure long-term object tracking in real time. We present experimental results of the visual motion tracking and inertial sensor fusion on a ground-based kite power system in pumping operation and compare both methods to an existing estimation scheme based on line measurements.

17.1 Introduction

In this work we consider ground-based kite power systems as the ones developed in Switzerland within the Autonomous Airborne Wind Energy (A^2WE) project [3]. Ground-based airborne wind energy (AWE) systems feature ground-based steering of tethered wings through differential line lengths [4]. Since electrical power is also generated at the ground following a pumping cycle approach, most weight of the AWE system is contained to the ground. This approach allows the ground station (GS) to be constructed using mostly off-the-shelf components, reducing the cost and risk of development, and increasing reliability.

Henrik Hesse (✉) · Max Polzin · Tony A. Wood · Roy S. Smith
ETH Zurich, Automatic Control Laboratory, Physikstrasse 3, 8092 Zurich, Switzerland.
e-mail: henrik.hesse@glasgow.ac.uk

Automatic control approaches for AWE rely on the availability of estimated parameters describing the state of the kite, typically given by kite position and heading, for feedback control. Existing estimation schemes for ground-based generators, for example [4, 13], compute estimates of the kite state using measurements of the line length and line angles obtained at the GS. Estimators based on ground-based position measurements are effective if the kite and winch system can ensure sufficient line tension. However, when operating a two-phase generation cycle, we desire low line forces during the retraction phase to improve cycle efficiency. Additionally, when operating at long line lengths to reach higher altitudes, the aerodynamic forces generated by the kite may not sufficiently balance the weight and drag induced by the lines even during the traction phase. Both scenarios lead to situations where degradation of the estimate quality reduces controller performance and imposes a critical limitation on the usable cycle efficiency of ground-based kite power systems.

Visible effects of the decline in estimator performance are line sag and increased delay in the estimation of the kite heading. The latter has been shown to significantly affect the performance of tracking controllers [31]. To actively incorporate system delay in the control design, in [30, 31] the kite steering behavior is modeled as a delayed dynamical system. Identification of the involved model parameters is used in [30] in a predictive manner to account for the delay in the path generation and tracking steps.

In this work we focus instead on sensor fusion to reduce the estimation and overall system delay. Different sensor setups including ground-based line and onboard inertial and position measurements have been explored in [12] using a kinematic model for sensor fusion. Experimental demonstration on a ground-based kite system with relatively short and fixed line lengths showed the benefits of fusion of inertial sensors with line angle measurements. They further found that onboard position measurements from GPS are not usable in kite power applications due to the large accelerations and fast changes in direction inherent to such systems. A similar finding was suggested in [10] and an inertial navigation algorithm based on acceleration and gyroscope measurements is proposed for a fixed line-length system.

From the lessons learned in [10, 12] we develop a filtering scheme to fuse onboard inertial measurements with line data for a ground-based kite power system with pumping operation. In particular, we formulate an extended Kalman filter (EKF) based on a sensor-driven kinematic model to fuse onboard yaw-rate measurements with *delayed* position measurements to correct for the bias in low-cost inertial sensors. Limiting ourselves to one onboard measurement further ensures that we can establish a reliable downlink to the GS. The developed estimation scheme has been implemented and demonstrated in closed-loop pumping operation on the AWE system developed at Fachhochschule Nordwestschweiz (FHNW). As an alternative to noisy line measurement, we can also obtain position measurements from range sensors using ultra-wideband radios. In [21] we have developed a range-inertial estimator specifically for AWE applications based on range measurements between a transceiver fixed to the kite and a number of static range beacons scattered on the ground. Only approximate knowledge of the range beacon locations is required.

Experimental validation of any estimation algorithm however requires true knowledge of the kite dynamics. Video streams of moving kites can provide useful position information which is unaffected by tether dynamics, can be acquired in real time, and requires no transmission of data from kite to GS. Hence, in this work we further develop a visual motion tracking (VMT) approach which produces improved position estimates from ground-based video streams and can be used in post-processing to evaluate other estimation approaches. The developed VMT approach is demonstrated in this work for tethered wings and conventional soft kites.

The rest of this chapter is structured as follows. In Section 17.2 we describe the kite system and derive the model equation required for the sensor fusion scheme in Section 17.4. The underlying sensor configuration is defined in Section 17.3. In Section 17.5 we described the developed VMT approach. In Section 17.6 we present experimental results from the VMT and estimators implemented on a ground-based kite system.

17.2 System Description and Dynamics

In this work we consider a two-line AWE system with ground-based actuation and generation. The system has been developed at the FHNW as part of the A^2WE project [3] to focus on autonomous operation of kite power generators. The kite is actuated at the GS which contains a drum and motor for each line. The steering of the kite is achieved through differential line length.

This system can operate with variable tether lengths such that full power cycles can be flown using conventional soft kites or tethered wings. In such two-phase operation power is generated in the so-called *traction phase* where the kite is flown in the power zone on a figure-eight trajectory under high aerodynamic forces in cross-wind flight. Winching of the lines enables power generation during this phase and controls the forces exerted on the kite system. Once a maximum line length has been reached, the kite is guided to the side of the wind window where it can be stabilized and rewound under low aerodynamic forces. This second phase where the lines are reeled in at low tether force is referred to as *retraction phase*. When a minimum line length is reached the cycle is repeated leading to a net power generation.

17.2.1 Model Equations

We describe the motion of the kite in terms of three right-handed reference frames, as defined in Figure 17.1, where we follow the definitions as in [12] but denote unit vectors by $\mathbf{e}$ with subscripts indicating the axis. The inertial frame $\{G\} := (\mathbf{e}_{Gx}, \mathbf{e}_{Gy}, \mathbf{e}_{Gz})$ with its origin at the GS is defined such that $\mathbf{e}_{Gx}$ is parallel to the ground and the wind direction, $\mathbf{e}_{Gz}$ is pointing upwards, and $\mathbf{e}_{Gy}$ completes the right hand system. For a given line length r the kite position $\mathbf{p}^G$ is expressed in the inertial

frame $\{G\}$ in terms of the spherical coordinates $\phi \in [-\pi, \pi]$ and $\theta \in [0, \pi]$, as shown in Figure 17.1a, such that

$$\mathbf{p}^G = r \begin{bmatrix} \cos(\phi)\cos(\theta) \\ \sin(\phi)\cos(\theta) \\ \sin(\theta) \end{bmatrix}, \tag{17.1}$$

where we denote vectors as bold variables with superscripts indicating the reference system for projection of the vectors. We also define the kite position vector in spherical coordinates as

$$\mathbf{p}_{\theta\phi r} = \begin{bmatrix} \theta & \phi & r \end{bmatrix}^\top. \tag{17.2}$$

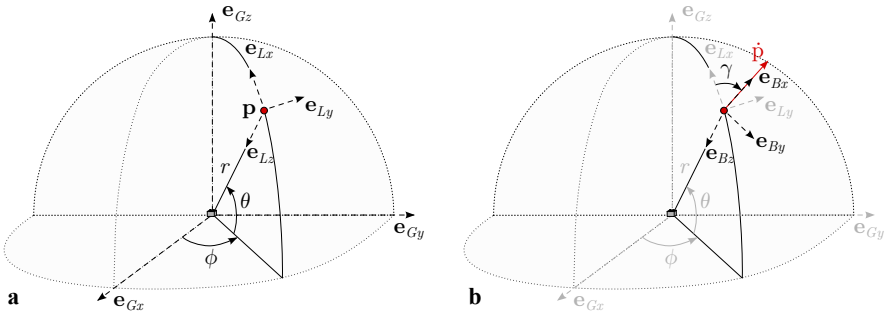

Fig. 17.1 Definition of reference frames: inertial frame $\{G\} = (\mathbf{e}_{Gx}, \mathbf{e}_{Gy}, \mathbf{e}_{Gz})$, local frame $\{L\} = (\mathbf{e}_{Lx}, \mathbf{e}_{Ly}, \mathbf{e}_{Lz})$, and body frame $\{B\} = (\mathbf{e}_{Bx}, \mathbf{e}_{By}, \mathbf{e}_{Bz})$. The wind direction is aligned with $\mathbf{e}_{Gx}$. **a** Kite position $\mathbf{p}$ in spherical components with azimuth angle ϕ, elevation angle θ, and line length r. **b** Definition of velocity vector orientation γ with velocity vector $\dot{\mathbf{p}}$ aligned with $\mathbf{e}_{Bx}$.

We can further define a *local* north-east-down (NED) coordinate system $\{L\} := (\mathbf{e}_{Lx}, \mathbf{e}_{Ly}, \mathbf{e}_{Lz})$ with its origin at the kite position $\mathbf{p}$. As shown in Figure 17.1a, $\mathbf{e}_{Lx}$ and $\mathbf{e}_{Ly}$ define the local tangent plane on a sphere with radius r with $\mathbf{e}_{Lz}$ pointing down from the kite towards the GS. A vector given in the inertial frame $\{G\}$ can be expressed in terms of the local frame $\{L\}$ using the transformation matrix C^{LG} as [12],

$$C^{LG} = \begin{bmatrix} -\cos(\phi)\sin(\theta) & -\sin(\phi)\sin(\theta) & \cos(\theta) \\ -\sin(\phi) & \cos(\phi) & 0 \\ -\cos(\phi)\cos(\theta) & -\sin(\phi)\cos(\theta) & -\sin(\theta) \end{bmatrix}, \tag{17.3}$$

such that for example $\mathbf{p}^L = C^{LG}\mathbf{p}^G$. The inverse transformation from frame $\{L\}$ to frame $\{G\}$ is given as $C^{GL} = \left(C^{LG}\right)^\top$. Note that the kite position, $\mathbf{p}^G$, defined in (17.1) can be extracted from C^{GL} as $\mathbf{p}^G = -rC^{GL}\mathbf{e}_{Lz}$.

At last, we define the *body* frame $\{B\} := (\mathbf{e}_{Bx}, \mathbf{e}_{By}, \mathbf{e}_{Bz})$, which is non-inertial, centered at the kite position, and fixed to the kite body, as shown in Figure 17.1b. We will use frame $\{B\}$ in this work to describe the orientation of the kite and hence assume that the axis $\mathbf{e}_{Bx}$ is always aligned with the kite velocity vector $\dot{\mathbf{p}}$. For the

derivation of the filtering equations, we further assume a rigid tether such that axis $\mathbf{e}_{Bz}$ coincides with $\mathbf{e}_{Lz}$, and $\mathbf{e}_{By}$ completes a right handed coordinate system. Hence, the transformation from the body frame $\{B\}$ to the local frame $\{L\}$ is given as

$$C^{LB} = \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{17.4}$$

where $\gamma \in [-\pi, \pi]$ is the *velocity vector orientation* introduced in [12, 13] as velocity angle. The notion of the velocity vector orientation has been demonstrated in recent publications as a crucial feedback variable to achieve successful autonomous flight during the traction phase [10, 13, 31] but also retraction phase [34]. Assuming small reeling speed, i.e. $\dot{r} \ll ||\dot{\mathbf{p}}||$, the velocity vector orientation is defined as

$$\gamma := \arctan_2\left(v_\phi, v_\theta\right) = \arctan_2\left(\cos(\theta)\,\dot{\phi}, \dot{\theta}\right). \tag{17.5}$$

where $\arctan_2\left(v_\phi, v_\theta\right) \in [-\pi, \pi]$ is the 4-quadrant arc tangent function and the kite velocity vector expressed in the $\{L\}$ frame is defined as

$$\mathbf{v}^L = \begin{bmatrix} v_\theta \\ v_\phi \\ v_r \end{bmatrix} = \begin{bmatrix} r\dot{\theta} \\ r\cos(\theta)\dot{\phi} \\ -\dot{r} \end{bmatrix}. \tag{17.6}$$

The velocity vector orientation can therefore be interpreted as the angle between the local north, $\mathbf{e}_{Lx}$, and the projection of the kite velocity vector onto the tangent plane of the wind window at the kite position. The velocity vector orientation is particularly suitable as feedback variable as it can be used to deduce the heading of the kite in a single scalar, e.g. for $\gamma = 0$ the kite moves upwards and for $\gamma = \pi/2$ parallel to the ground towards the left (as seen from the GS). More details on the derivation of the velocity vector orientation can be found in [13]. During retraction phases, v_θ can converge to zero and we use a regularized version of the velocity vector orientation, defined as [34]

$$\gamma_{\text{reg}} := \arctan_2\left(\cos(\theta)\dot{\phi} + c\sin(\phi - \phi_W), \dot{\theta} + c\sin\theta\cos(\phi - \phi_W)\right), \tag{17.7}$$

during retraction phases. A tuning parameter $c = 0.02$ was used in this work and ϕ_W can account for misalignment of the wind direction in the $(\mathbf{e}_{Gx}, \mathbf{e}_{Gy})$ plane.

Based on the definition of the kite velocity vector orientation in (17.5) we can derive the kinematic model equations which will be used to propagate the estimator states in Section 17.4.2. Similar to [10, 30] we describe the behavior of the kite as a unicycle on the (θ, ϕ)-plane with heading γ and tangent velocity $v_{\theta\phi} := \sqrt{v_\theta^2 + v_\phi^2}$ as

$$\dot{\theta} = \frac{v_{\theta\phi}}{r}\cos(\gamma), \tag{17.8a}$$

$$\dot{\phi} = \frac{v_{\theta\phi}}{r\cos(\theta)}\sin(\gamma), \tag{17.8b}$$

which we can extend with the tether reeling kinematics,

$$\dot{r} = v_r \,, \tag{17.8c}$$

to complete the model equations in this work.

17.3 Sensor Configuration and Modeling

The two-line, ground-based AWE system used in this work provides a range of sensors installed on the kite (inertial measurement unit) and the GS (line angles, line lengths, line forces, wind speed, wind direction, and video footage) as shown in Figure 17.2. A wind sensor, mounted roughly 5 m above the ground, provides measurements of the wind direction, $\tilde{\phi}_W$, in the $(\mathbf{e}_{Gx}, \mathbf{e}_{Gy})$ plane and wind speed, $\tilde{v}_W$, where $(\tilde{\bullet})$ denotes the noise-corrupted, unfiltered measurement of a variable for the remainder of this work. The different sensors, which are relevant to the sensor fusion and VMT approaches in Secs. 17.4 and 17.5, will be presented next together with the corresponding sensor modeling approach. The important sensor parameters are summarized in Table 17.1.

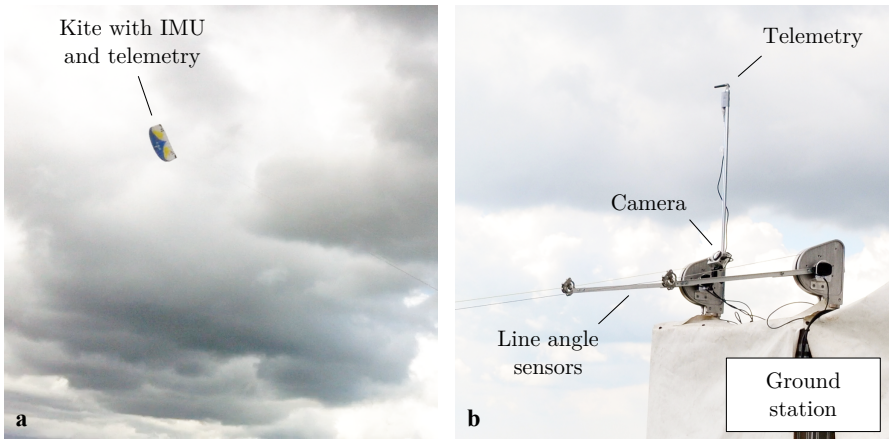

Fig. 17.2 Sensor configuration of the kite power system developed at FHNW. **a** Kite system. **b** Ground station.

17.3.1 Line Sensors

The GS is equipped with line angle sensors on both lines, as shown in Figure 17.2b, with two dedicated encoders on each lead-out sheave to measure azimuth and eleva-

Table 17.1 Overview of sensors with the corresponding variance $\sigma^2_{(\bullet)}$ of the sensor noise.

Sensor	Variable	Description	Properties
Line angle sensor at GS	$\tilde{\theta}$	Elevation angle	$\sigma^2_\theta = 8 \times 10^{-2}$
	$\tilde{\phi}$	Azimuth angle	$\sigma^2_\phi = 8 \times 10^{-2}$
Ground station motor	$\tilde{r}$	Line length	$\sigma^2_r = 10^{-3}$
	$\tilde{v}_r$	Reeling speed	$\sigma^2_{v_r} = 10^{-3}$
Inertial measurement unit	$\tilde{\omega}$	Yaw rate of kite	$\sigma^2_\omega = 10^{-4}$

tion angles, $\tilde{\theta}$ and $\tilde{\phi}$, respectively. In this work we only consider the measurements from the left line although weighting of left and right line angle measurements with the corresponding line force measurements can lead to improved estimates by favoring readings from the line with higher tension. Combined with line length measurements, $\tilde{r}$, obtained from the motors inside the GS, we model the line angle measurements to provide unbiased measurements of the kite position given in terms of spherical coordinates as

$$\tilde{\mathbf{p}}_{\theta\phi r} = \begin{bmatrix} \tilde{\theta} \\ \tilde{\phi} \\ \tilde{r} \end{bmatrix} = \begin{bmatrix} \theta \\ \phi \\ r \end{bmatrix} + \begin{bmatrix} \zeta_\theta \\ \zeta_\phi \\ \zeta_r \end{bmatrix} = \mathbf{p}_{\theta\phi r} + \zeta_{\theta\phi r}, \tag{17.9}$$

where $\zeta_{\{\theta,\phi,r\}}$ are the measurement noises which we assume to be independent, zero-mean Gaussian processes, i.e. $\zeta_{\{\theta,\phi,r\}} \sim \mathcal{N}(0, \sigma^2_{\{\theta,\phi,r\}})$. We have determined the variances $\sigma^2_{\{\theta,\phi,r\}}$ from experiments as stated in Table 17.1.

The assumption of unbiased line angle measurements in (17.9) is clearly questionable for ground-based AWE systems. Especially at long lines we can observe that tether dynamics significantly deteriorate the quality of ground-based measurements introducing bias and measurement delay [31].

17.3.2 Inertial Measurements

To improve the estimation of the kite heading we exploit additional onboard measurements from an inertial measurement unit (IMU) installed on the kite as depicted in Figure 17.2a. We use the Pixhawk Autopilot which is equipped with a Cortex M4 processor and several redundant sensors including a 3-axis accelerometer, 3-axis gyroscope, magnetometer, and barometer [22]. In this work we will focus on fusing the gyroscope measurements only as they are particularly relevant in the estimation of the kite velocity vector orientation [9].

The 3-axis gyroscope installed in the Pixhawk Autopilot has been set to operate at a bandwidth of 100 Hz providing measurements in the range of $\pm 2.75\pi$ rad/s.

In the temperature range of operation we can expect a bias of $\pm 3 \times 10^{-3}$ rad/s and a noise density of $\pm 5 \times 10^{-4}$ rad/(s$\sqrt{\text{Hz}}$). The yaw-rate measurement is streamed via telemetry to the GS, where it is synchronized with the line measurements. Since we only require gyroscope measurements for the developed estimator, a small Baud rate of 57,600 Hz could be selected to establish a reliable connection over 250 m at 100 Hz bandwidth.

Although we can calibrate the gyroscope to remove a static bias, it is common that low-cost inertial sensors tend to drift due to external factors. We will therefore model the gyroscope following a common approach in the literature, e.g. [19], which relates the true angular velocities ω^B to the gyroscope measurements, $\tilde{\omega}$, through

$$\tilde{\omega} = \omega^B + \mathbf{b}_\omega + \eta_\omega, \tag{17.10}$$

where ω^B indicates the angular velocity of the $\{B\}$ frame relative to the $\{G\}$ frame projected in the $\{B\}$ frame. The inertial measurements are corrupted by zero-mean Gaussian noise captured by the vector $\eta_\omega \sim \mathcal{N}(0, \sigma_\omega^2\, \mathrm{I}_{3\times 3})$. The non-static gyroscope bias is modeled as a continuous-time Gaussian random process,

$$\dot{\mathbf{b}}_\omega = \eta_b, \tag{17.11}$$

with $\eta_b \sim \mathcal{N}(0, \sigma_b^2\, \mathrm{I}_{3\times 3})$ with $\sigma_b^2 = 10^{-3}$ throughout this work.

Next, we relate the gyroscope measurements in (17.10) to our model equations in Section 17.2 to arrive at an expression for the velocity vector orientation, γ, defined in (17.5). For this we first link the rate of change of the orientation of the kite,

$$\dot{C}^{GB} = \frac{d}{dt}\left(C^{GL}C^{LB}\right) = \dot{C}^{GL}C^{LB} + C^{GL}\dot{C}^{LB}, \tag{17.12}$$

to the kite angular velocity, ω^B, as

$$\omega^B\times = C^{BG}\dot{C}^{GB} = C^{BG}\dot{C}^{GL}C^{LB} + C^{BL}\dot{C}^{LB}, \tag{17.13}$$

where we denote the skew-symmetric matrix of ω^B as $[\omega^B\times]$ with $[\omega^B\times]\mathbf{b} := \omega^B \times \mathbf{b}$ for $\mathbf{b} \in \mathbb{R}^3$. Hence, a gyroscope fixed to the kite will measure the angular velocities due to the motion on the sphere, $\dot{C}^{GL}$, and a change in heading, $\dot{C}^{LB}$. With

$$C^{BL}\dot{C}^{LB} = \begin{bmatrix} 0 & -\dot{\gamma} & 0 \\ \dot{\gamma} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{17.14}$$

we can directly relate the time-derivative of the velocity vector orientation, $\dot{\gamma}$, to the measured gyroscope outputs (17.10) as

$$\dot{\gamma} = \left(\omega^B\right)^\top \mathbf{e}_{Bz} + \dot{\phi}\sin(\theta) = \omega_{Bz} + \frac{v_{\theta\phi}}{r}\tan(\theta)\sin(\gamma), \tag{17.15}$$

where $\omega_{Bz} = \left(\omega^B\right)^\top \mathbf{e}_{Bz}$ is the measurable yaw rate of the kite. The effect of the additional term, $\dot{\phi}\sin(\theta)$, due to the motion on the sphere is illustrated (for a different definition of reference frames) in [9].

17.3.3 Visual Measurements

The GS is equipped with a GoPro HD Hero 2 video camera. The camera has been set to capture video streams with a resolution of 1280x960 pixel at 48 frames per second. Although the camera is equipped with a fisheye lens, extending the field of view to nearly π rad in a static setup, the field of view is not sufficient to capture the kite at all times. Especially during retractions the wind window is increased due to the reeling of the tethers. We therefore attached the camera to the lead-out sheave of the right line at the GS, as shown in Figure 17.2b. The lead-out sheave (and hence the camera) rotates with the azimuth angle, $\tilde{\phi}$, but is fixed in elevation.

The built-in fisheye lens introduces nonlinear distortions to the camera image. To successfully link the true kite position $\mathbf{p}^G$ in the inertial frame $\{G\}$ with a visual position measurement $\tilde{\mathbf{p}}^V$ from the video stream in the camera frame, $\{V\}$, we need to compensate for distortions (intrinsic calibration) and estimate the camera pose (extrinsic calibration). To undistort the video stream from the fisheye effect and calculate the intrinsic camera parameters, we used the calibration procedure from the Computer Vision System Toolbox™ in MATLAB® based on a pinhole camera with distortion coefficients for radial and tangential correction [17].

From the undistorted video stream, we extract the camera pose relative to the inertial frame given by the transformation matrix C^{GV} (extrinsic camera parameters). The orientation of the camera, C^{GV}, is computed using linear regression over the measured line angles and visually tracked positions at high line forces where we trust the line angle data. In each experiment, the calculated camera parameters do not change over time, except for an azimuth rotation of the lead-out sheaves which is measured at the GS and added to the tracked position estimate. Depth information is lost in position measurements extracted from a single video stream. However, since the kite is assumed to move on the sphere, we are able to recover position measurements $\tilde{\mathbf{p}}^V$ from the camera stream by taking the line length measurement, $\tilde{r}$, as depth information.

17.4 Filtering Schemes

In this section we present the filtering algorithms to obtain estimates of the feedback variables $\hat{\mathbf{p}}^G$ and $\hat{\gamma}$ where $(\hat{\bullet})$ will denote estimated variables in this work. The first approach presented in Section 17.4.1 uses only line measurements at the GS and provides a good starting point for simple kite power systems. It will also serve as a baseline estimator in this work. As line dynamics and filtering of noisy position

measurements can deteriorate the control performance, we introduce a sensor fusion algorithm based on onboard inertial measurements in Section 17.4.2.

17.4.1 Line-Angle-Based Estimation

Ground-based AWE systems typically provide basic measurements of azimuth/elevation angles and line length, $(\tilde{\theta}, \tilde{\phi}, \tilde{r})$, as described in Section 17.3.1. We therefore first present a line-based estimator following the approach in [12] based on an orientation-free kinematic process model. In this work, however, we describe the filter states directly in terms of the measurable spherical coordinates leading to the 6-dimensional state vector $\mathbf{x}_{\text{la}} = [\mathbf{p}_{\theta\phi r}^\top \ \dot{\mathbf{p}}_{\theta\phi r}^\top]^\top$ with the vector of spherical position coordinates, $\mathbf{p}_{\theta\phi r}$, defined in (17.2). The discrete line angle and length measurements, $\mathbf{z}_{\text{la}} = \tilde{\mathbf{p}}_{\theta\phi r}^\top$ are modeled as defined in (17.9). Under the assumption of decoupling the kite motion in azimuth and elevation directions, we can obtained the following simplified process model given as

$$\mathbf{x}_{\text{la}}^k = \begin{bmatrix} \mathrm{I}_{3\times3} & T_s\mathrm{I}_{3\times3} \\ 0_{3\times3} & \mathrm{I}_{3\times3} \end{bmatrix} \mathbf{x}_{\text{la}}^{k-1} + \eta_{\text{la}}^{k-1} \tag{17.16a}$$

$$\mathbf{z}_{\text{la}}^k = \begin{bmatrix} \mathrm{I}_{3\times3} & 0_{3\times3} \end{bmatrix} \mathbf{x}_{\text{la}}^k + \zeta_{\text{la}}^k, \tag{17.16b}$$

where we have discretized the process dynamics with the forward Euler method using a constant sampling rate, T_s. In (17.16) we denote identity and zeros matrices as I and 0, respectively, with subscripted dimensions. The process and measurement noises, η_{la} and ζ_{la}, respectively, are modeled as independent zero-mean Gaussian noise vectors. Writing the filter equations directly in the measurable spherical coordinates ensures that ζ_{la}, as defined in (17.9), captures the line angle and length sensor noises with the variances defined in Table 17.1.

We can then derive a steady-state *Kalman filter* [1] based on (17.16) to extract state estimates $\hat{\mathbf{x}}_{\text{la}}$. Since the underlying linear process model (17.16) is autonomous, the Kalman filter equations reduce to,

$$\hat{\mathbf{x}}_{\text{la}}^k = \begin{bmatrix} \mathrm{I}_{3\times3} & T_s\mathrm{I}_{3\times3} \\ 0_{3\times3} & \mathrm{I}_{3\times3} \end{bmatrix} \hat{\mathbf{x}}_{\text{la}}^{k-1} + K_{\text{la}} \left(\tilde{\mathbf{p}}_{\theta\phi r}^k - \hat{\mathbf{p}}_{\theta\phi r}^{k-1} - T_s \dot{\hat{\mathbf{p}}}_{\theta\phi r}^{k-1} \right), \tag{17.17}$$

where we have combined the prediction and measurement update steps in the steady-state Kalman filter [1]. The steady-state Kalman gain K_{la} can be precomputed using the variances of the measurement noise, ζ_{la} defined in Table 17.1. The estimated velocity vector orientation $\hat{\gamma}_{\text{la}}^k$, as defined in (17.5), is finally extracted at each time step from the estimated velocity vector, $\dot{\hat{\mathbf{p}}}_{\theta\phi r}^k$.

17.4.2 Sensor Fusion

Since position-based estimation of the velocity vector orientation, as presented in 17.4.1, involves differentiation of noisy position measurements, we require excessive filtering which introduces estimation delay in addition to the existing system delay. Hence, in this section we derive an estimator that uses onboard yaw-rate measurements, $\tilde{\omega}_{Bz}$, as modeled in Section 17.3.2.

Coupling (17.15) with the unicycle model in (17.8), we can derive a sensor-driven kinematic process model given in discrete time as

$$\begin{bmatrix} \gamma \\ v_{\theta\phi} \\ \theta \\ \phi \\ r \\ b_{\omega z} \end{bmatrix}^{k} = \begin{bmatrix} \gamma \\ v_{\theta\phi} \\ \theta \\ \phi \\ r \\ b_{\omega z} \end{bmatrix}^{k-1} + T_s \begin{bmatrix} v_{\theta\phi} r^{-1} \tan(\theta) \sin(\gamma) + \tilde{\omega}_{Bz} - b_{\omega z} \\ 0 \\ v_{\theta\phi} r^{-1} \cos(\gamma) \\ v_{\theta\phi} (r \cos(\theta))^{-1} \sin(\gamma) \\ \tilde{v}_r \\ 0 \end{bmatrix}^{k-1} + \eta^{k-1}, \tag{17.18}$$

with the state vector defined as,

$$\mathbf{x}_\omega = [\gamma \;\; v_{\theta\phi} \;\; \theta \;\; \phi \;\; r \;\; b_{\omega z}]^\top. \tag{17.19}$$

The time-varying gyroscope bias in yaw, $b_{\omega z} = (b_\omega)^\top \mathbf{e}_{Bz}$, is estimated as part of the filtering which is a common approach in estimation with low-cost IMUs [24]. The measured yaw rate, $\tilde{\omega}_{Bz}$, and reeling speed measured at the GS, $\tilde{v}_r$, are the inputs to the system which we can stack as $\mathbf{u}_\omega = [\tilde{\omega}_{Bz} \;\; \tilde{v}_r]^\top$. We further stack the individual process noise terms in $\eta = [\eta_\gamma \;\; \eta_v \;\; \eta_\theta \;\; \eta_\phi \;\; \eta_r \;\; \eta_{bz}]^\top$ and again use T_s as the sampling rate in the temporal forward Euler discretization. We can then write (17.18) as

$$\mathbf{x}_\omega^k = \mathbf{f}(\mathbf{x}_\omega^{k-1}, \mathbf{u}_\omega^{k-1}, \eta^{k-1}). \tag{17.20}$$

Estimation of the bias however requires an unbiased measurement. In ground-based AWE systems line angle measurements typically provide unbiased position measurements of the kite position during the traction phase [12]. We therefore define the measurement function, $\mathbf{z}_\omega^k$, to be the outputs of the process model (17.18) with

$$\mathbf{z}_\omega^k = \mathbf{h}(\mathbf{x}_\omega^k, \zeta^k) = \mathbf{p}_{\theta\phi r}^k + \zeta_{\theta\phi r}^k, \tag{17.21}$$

where $\zeta_{\theta\phi r}$ accounts for the noise in the line measurements as defined in Table 17.1.

17.4.2.1 Extended Kalman Filtering

Due to the nonlinear nature of the process dynamics (17.18) we implement an EKF [23], as outlined in Figure 17.3, to fuse the inertial measurements $\tilde{\omega}_{Bz}$ characterized in (17.10) with the line measurements $\tilde{\mathbf{p}}_{\theta\phi r}$ defined in (17.9). In Kalman

filtering an estimate is computed at each time step performing a *prior update* step to propagate the estimate and a *measurement update* step to correct the propagated estimate using line angle measurements.

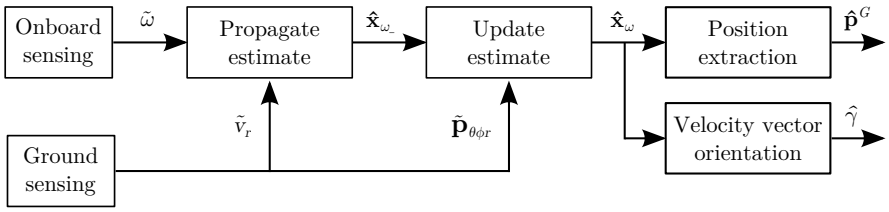

Fig. 17.3 Overview of sensor fusion scheme.

To arrive at the estimate propagation equations used in the prior update step, we define the estimate state vector, $\hat{\mathbf{x}}_{\omega}$, analogous to (17.19), and take the expectation of (17.18) to propagate the estimate $\hat{\mathbf{x}}_{\omega}^{k-1}$ based on the model inputs $\mathbf{u}_{\omega}^{k-1} = [\tilde{\omega}_{B_z}^{k-1} \;\; \tilde{v}_r^{k-1}]^\top$. The propagated estimate, $\hat{\mathbf{x}}_{\omega\_}^{k}$, is updated at each time step with the measurement $\mathbf{z}_{\omega}^{k}$ such that the resulting sensor fusion estimate is given as [23]

$$\hat{\mathbf{x}}_{\omega}^{k} = \hat{\mathbf{x}}_{\omega\_}^{k} + K_{\omega}^{k}\left(\mathbf{z}_{\omega}^{k} - \mathbf{h}(\hat{\mathbf{x}}_{\omega\_}^{k}, 0)\right), \tag{17.22}$$

where $\hat{\mathbf{x}}_{\omega\_}^{k}$ denotes the propagated state at k.

In the EKF scheme the Kalman gain K_{ω} is time varying due to the nonlinear process dynamics and is a function of the covariance matrices of the process and measurement noises given as $\eta \sim \mathcal{N}(0, Q)$ and $\zeta \sim \mathcal{N}(0, R)$, respectively. We can estimate $Q \in \mathbb{R}^{6\times6}$ and $R \in \mathbb{R}^{3\times3}$ from experiments, as done for R in Section 17.3.1, or tune them to effect the performance of the sensor fusion estimator. Unlike the line-based estimator in Section 17.4.1, we can directly extract estimates of the position vector and velocity vector orientation, $\hat{\mathbf{p}}^G$ and $\hat{\gamma}$, respectively, from the estimate state vector $\hat{\mathbf{x}}_{\omega}$.

In the absence of onboard measurements, the proposed estimator reduces to a line-based estimator, similar to the one in (17.16)-(17.17), but with a sound definition of the kinematic process model based on spherical coordinates and the definition of the velocity vector orientation in (17.5).

17.4.2.2 Extended Kalman Filtering with Delayed Measurements

The estimator in Section 17.4.2.1 was derived on the assumption that yaw rate, reeling speed and line measurements are obtained simultaneously. From experiments however we observe that line angle readings can be significantly delayed due to line dynamics, especially at low tether tension. Hence, in this section we want to account for a static delay τ in the line angle readings $\tilde{\theta}_{\tau}^{k} = \tilde{\theta}^{k-\tau}$ and $\tilde{\phi}_{\tau}^{k} = \tilde{\phi}^{k-\tau}$ during

the measurement update step in (17.22), where the subscript denotes the delay in measured variables.

We can augment the estimate state vector with the additional delay states,

$$\hat{\mathbf{x}}_{\omega,\tau} = [\hat{\gamma}\ v_{\theta\phi}\ \ \hat{\theta}\ \ \hat{\theta}_1\ \ \ldots\ \ \hat{\theta}_\tau\ \ \hat{\phi}\ \ \hat{\phi}_1\ \ \ldots\ \ \hat{\phi}_\tau\ \ \hat{r}\ \ \hat{b}_{\omega z}]^\top, \tag{17.23}$$

with $\hat{\mathbf{x}}_{\omega,\tau} \in \mathbb{R}^{6+2\tau}$ to arrive at the augmented estimate propagation equations given as

$$\hat{\mathbf{x}}^k_{\omega,\tau} = \hat{\mathbf{x}}^{k-1}_{\omega,\tau_-} + T_s \begin{bmatrix} \hat{v}_{\theta\phi}\hat{r}^{-1}\tan(\hat{\theta})\sin(\hat{\gamma}) - \hat{b}_{\omega z} \\ 0 \\ \hat{v}_{\theta\phi}\hat{r}^{-1}\cos(\hat{\gamma}) \\ 0_{\tau\times 1} \\ \hat{v}_{\theta\phi}(\hat{r}\cos(\hat{\theta}))^{-1}\sin(\hat{\gamma}) \\ 0_{(\tau+2)\times 1} \end{bmatrix}^{k-1} + T_s \mathbf{u}^{k-1}_\omega, \tag{17.24}$$

where $\mathbf{u}^{k-1}_\omega = [\tilde{\omega}^{k-1}_{B_z}\ \ \tilde{v}^{k-1}_r]^\top$ accounts again for the simultaneous measurements of the kite yaw rate and line reeling speed. The propagated augmented estimate is updated at time step k with the delayed measurements

$$\mathbf{z}^k_{\omega,\tau} = \left[\tilde{\theta}^k_\tau\ \ \tilde{\phi}^k_\tau\ \ \tilde{r}^k\right]^\top. \tag{17.25}$$

The introduction of additional delay states in (17.23) increases the system size and hence numerical cost of the filtering scheme. To facilitate real-time operation we therefore assume the line length measurement $\tilde{r}$ in (17.25) to be simultaneous. Based on the EKF scheme [23] and using the augmented estimate propagation model (17.24) and measurement model (17.25) we can obtain estimates where we actively account for the delay in the line angle measurements.

17.5 Visual Motion Tracking of Kites

Next we aim to extract the kite position from a video stream obtained from a camera at the GS as described in Section 17.3.3. Such visual position measurements are not affected by line dynamics and can help to assess characteristics of line dynamics in a typical pumping operation. Kite power systems, however, commonly operate at long line lengths leading to very small target sizes in a video stream. Additionally, we require from VMT applied to kite applications to be able to cope with:

- dynamic backgrounds due to camera motion,
- highly cluttered backgrounds and target occlusions due to clouds and sunlight,
- varying illumination and changing appearances over one pumping cycle, and
- high frame rates to track fast moving rigid wings.

In summary, tracking of kites in real time from ground-based video streams involves most challenges for modern motion trackers as described in [32]. Hence, despite the large development in motion tracking, as surveyed in [27], there still exists no

solution tailored to AWE systems. In this work we therefore adapt existing methodologies for motion tracking and object detection to develop a novel VMT approach. The algorithm developed here reaches the low computational complexity to achieve high frame rates while producing accurate tracking results. Long-term robust object tracking is accomplished in real time.

Localization of a target object in a sequence of consecutive frames is generally defined as *motion tracking*. From the performance evaluation of state-of-the-art motion trackers in [27], there currently exists no approach that achieves the high frame rates ($\geq$ 48 fps) and long-term tracking capabilities ($\geq$ 2 h) required for kite applications. Therefore, a simple classical motion tracker is not applicable in this scenario. Alternatively, *object detection* provides very accurate detection performance for long-term tracking. However, the numerical burden to achieve this increase in accuracy leads to low frame rates which are too low to achieve real-time tracking capabilities with object detection alone.

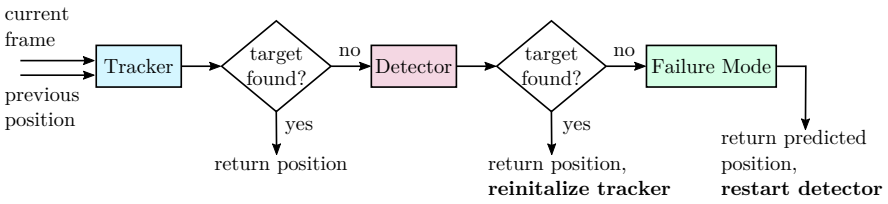

Fig. 17.4 Flowchart of the presented VMT approach.

By combining state-of-the-art algorithms for object detection and motion tracking, we overcome existing drawbacks in both and achieve low failure rates and good recovery capabilities. The approach followed in this work to couple the motion tracker [18] and object detector [7] is shown in Figure 17.4. To switch between tracker and detector we require a measure that we can use as a threshold. We have therefore extended the approach in [18], to return a quality measure for the localization performance of the tracker. Hence, based on the last known position, the tracker processes the current frame and locates the target. If the tracker succeeds with sufficient certainty, it returns the new target position. Otherwise, the detector assists by processing a sub-image of the current frame extracted at the last known kite position. The kite motion is assumed to be continuous and sufficiently slow such that the target is likely to be present in the sub-image. After successful target detection in the processed sub-image, the motion tracker is reinitialized. To address the situation when both the tracker and the detector fail to locate the target we have added three selectable *failure modes* to predict the most likely positions for reappearance: (a) remain at last tracked position, (b) predict new position based on a motion model and appropriate filtering, or (c) use an external source such as line angle measurements. Once the detector has successfully relocated the kite, the VMT algorithm continues with motion tracking.

To start the VMT, the motion tracker requires a single frame for initialization. The object detector can be initialized with an arbitrary number of labeled samples. During the initialization phase, we collect such training samples from the first frames of the video. Once enough training samples have been acquired, the detector is trained and the automatic tracking process starts. Next, we provide a brief overview of the implementation of the tracker and detector.

17.5.1 Motion Tracker

The implemented tracker is a modified version of a kernelized correlation filter [18] which has been shown to be fast while maintaining high tracking performance in recent benchmark tests [33]. Core of the tracker is a discriminating classifier obtained from a kernel ridge regression problem. We briefly summarize the method of [18] with our modifications in this section.

The tracker is limited to training data extracted from a single frame. We extract edges from a section of the video frame with the kite in the center to generate the feature vector, $\mathbf{x} \in [0,1]^n$, by using Sobel edge detection [20]. The extracted feature will be referred to as base training feature. It is the only training sample that is labeled to be positive. Permutations of the base sample corresponding to horizontal and vertical cyclic shifts of the frame section serve as negative samples and complete the training set,

$$\mathscr{X} := \left\{ P^{l-1}\mathbf{x} \mid l = 1,\ldots,n \right\},$$

where P is a shift generating permutation matrix. The classifier, $H_t(\cdot) = \mathbf{w}^\top \psi(\cdot)$, is obtained from a kernel ridge regression problem,

$$\mathbf{w} = \arg\min_{\mathbf{q}} \sum_i \left(\mathbf{q}^\top \psi(\mathbf{x}_i) - \mathbf{y}_i \right)^2 + \lambda \left\| \mathbf{q} \right\|^2, \tag{17.26}$$

where the regularized squared error between the training samples, $\mathbf{x}_i \in \mathscr{X}$, mapped to an implicit feature space by the kernel mapping, $\psi(\cdot) : [0,1]^n \to \mathbb{R}^m$, and their corresponding labels, $\mathbf{y}_i \in [0,1]$, is minimized. The ridge parameter, λ, penalizes over-fitting.

The optimization problem (17.26) is linear in the dual space and its solution can be written as a linear combination of the samples, $\mathbf{x}_i$, mapped to the feature space [25],

$$\mathbf{w} = \sum_i \alpha_i \psi(\mathbf{x}_i),$$

where the dual variables, α_i, are elements of the vector, α, obtained from the stacked label vector, $\mathbf{y}$, by

$$\alpha = (K + \lambda \mathrm{I}_{n\times n})^{-1}\mathbf{y}. \tag{17.27}$$

The elements of the kernel matrix, K, correspond to the inner products of all training samples in the feature space,

$$K_{ij} = \langle \psi(\mathbf{x_i}), \psi(\mathbf{x_j}) \rangle = \kappa(\mathbf{x}_i, \mathbf{x}_j),$$

and can be calculated by the kernel function, $\kappa(\mathbf{x}_i, \mathbf{x}_j)$, without instantiating any sample in the feature space. The interested reader is referred to [26] for more information on kernel methods.

For the particular structure of the training set, the kernel matrix becomes circulant. Exploiting the diagonalization property of the Fourier transform of any circulant matrix [16], the computation of (17.27) can be simplified to element-wise operations in the Fourier domain,

$$F\alpha = \frac{F\mathbf{y}}{F\mathbf{k}_{\mathbf{xx}} + \lambda}, \tag{17.28}$$

where F denotes the discrete Fourier transform (DFT) matrix of the unitary DFT, $\mathbf{k}_{\mathbf{xx}}$ is the first column of the kernel matrix K, and the fraction denotes element-wise division. For arbitrary samples, $\mathbf{x}_i, \mathbf{x}_j \in [0,1]^n$, and a linear kernel, $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$, the computation of the *kernel phase correlation*, $\mathbf{k}_{\mathbf{x}_i\mathbf{x}_j}$, can be computed efficiently in the Fourier domain as,

$$F\mathbf{k}_{\mathbf{x}_i,\mathbf{x}_j} = \frac{F^{-1}\mathbf{x}_i \odot F\mathbf{x}_j}{\left|F^{-1}\mathbf{x}_i \odot F\mathbf{x}_j\right|}, \tag{17.29}$$

where $\odot$ denotes the element-wise product. The magnitude normalization is an added modification from the derivation in [18]. We consider here a linear kernel for its simplicity but the method can by extended to more complex kernels potentially resulting in more discriminating classifiers.

In each frame a base candidate feature, $\mathbf{z} \in [0,1]^n$, is extracted given the previous position estimate. Similar to the training set, we consider a candidate set consisting of the base candidate feature, $\mathbf{z}$, and its relative shifts,

$$\mathscr{Z} := \left\{ P^{l-1}\mathbf{z} \mid l = 1, \ldots, n \right\}. \tag{17.30}$$

The new position estimate is given by the shift corresponding to the element of the candidate set that maximizes the classifier response,

$$\mathbf{z}^* = \arg\max_{\mathbf{s} \in \mathscr{Z}} H_t(\mathbf{s}).$$

The evaluation of the classifier, $H_t(\cdot)$, on all elements of the candidate set, $\mathscr{Z}$, is efficiently computed in the Fourier domain by,

$$F\mathbf{h}_t = F\mathbf{k}_{\mathbf{zx}} \odot F\alpha, \tag{17.31}$$

where $\mathbf{h}_t \in [0,1]^n$ is a vector of the stacked classifier outputs applied to the elements in $\mathscr{Z}$, $\mathbf{k_{zx}}$ denotes the kernel phase correlation between the base training feature, $\mathbf{x}$, and base candidate feature, $\mathbf{z}$. The values of the elements of $\mathbf{h}_t$ serve as similarity measures between the candidate features and the base training feature, $\mathbf{x}$, and allow to assess the quality of the location estimate. Note that given the normalization in (17.29) the output of the classifier $H_t(\cdot)$, is bound to be in $[0,1]$.

17.5.2 Object Detector

The object detector is based on the work described in [2, 7, 29] which has been shown to achieve high performances on benchmark tests in [11]. In this section, we illustrate the concept of the object detector by summarizing its individual components.

We consider labeled training samples manually collected from a video stream. Multiple feature channels are extracted from the data to train the detection classifier, $H_d(\cdot)$. The features consist of color information, the histogram of oriented gradients, and the gradient magnitude as suggested in [6]. The used classifier is a cascade of boosted decision trees introduced in [29]. Boosting methods are used in machine learning to construct strong classifiers by combining multiple weak classifiers. An illustrative example of boosting is shown in Figure 17.5. In particular, we consider the discrete AdaBoost method discussed in [15] with binary decision trees as weak classifiers. The boosted decision trees are coupled in a cascade with a constant rejection threshold as in [5, 8]. In such a cascade candidate samples are discarded if the sum of weak classifiers drops below a rejection threshold.

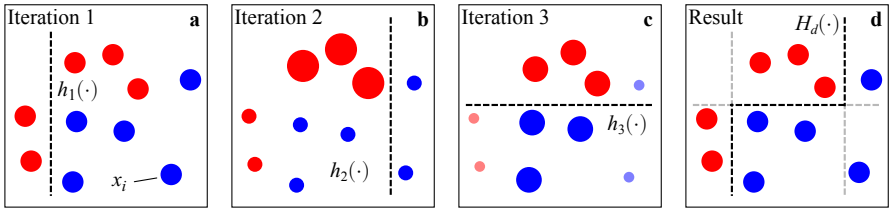

Fig. 17.5 Illustration of boosting [14]. In each iteration j a new weak classifier, $h_j(\cdot)$, (vertical or horizontal dashed line) is trained which minimizes the current classification error of the weighted positive (red), and negative (blue) samples. The resulting boosted classifier, $H_d(\cdot)$, is a weighted sum over all weak classifiers. **a** Equal weight is assigned to each feature image, x_i. **b** Weights of incorrectly classified samples are increased. **c** Subset of the heaviest samples is sufficient for optimal learning [2]. **d** Combination of all weak classifiers to form the boosted classifier, $H_d(\cdot)$.

The object detector is applied when the motion tracker fails to locate the target and is initiated at the last position that was tracked, as illustrated in Figure 17.6. The position is then obtained from the highest scoring detection. Successful detection initializes re-training of the tracker. If the detector fails to locate the target, a location

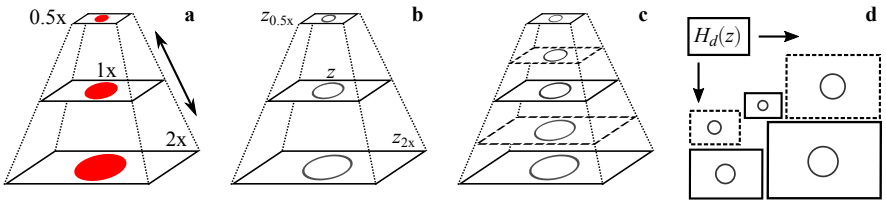

Fig. 17.6 Illustration of sliding window detection with aggregated channel features over an approximated feature pyramid. Approximating features is computationally less expensive then the extraction. Detection over a feature pyramid enables scale-invariant detections. The best detection of $H_d(z)$ is returned. **a** Resample extracted candidate patch at higher $(2x)$ and lower scales $(0.5x)$. **b** Extract feature images z, z_{2x}, $z_{0.5x}$ at all scales. **c** Approximate feature images at scales in between, $z_{0.75x}$, $z_{1.5x}$ [7]. **d** Run sliding window over all extracted and approximated feature images [2].

estimate is obtained from the failure mode. Note that limiting the detection area enables detection in real time.

17.6 Experimental Results

In this section we demonstrate VMT and the presented filtering approaches on experimental data obtained using the AWE platform developed at FHNW. In Section 17.6.1, we first compare VMT against (unfiltered) line angle measurements from experimental data of a tethered wing flight. This case aims to demonstrate the potential of VMT for rigid tethered wings with depower capabilities where line-based approaches would fail especially during retraction phases. Although the VMT approach has been implemented for high sampling rates of 100 Hz, due to hardware limitations it currently only provides vision-based estimates in post-processing. We therefore focused on a (real-time) sensor fusion approach which is evaluated in Section 17.6.2 using vision-based results as a reference solution.

17.6.1 Visual Motion Tracking of Tethered Wings

In this work we apply VMT as a tool to verify other estimation approaches for ground-based AWE systems. We have demonstrated the VMT approach for a range of kites in a series of visually challenging videos containing occlusions, camera motion, appearance changes, and long video duration. Here, we compare ground-based line angle measurements (Section 17.3.1) from experimental data of a rigid tethered wing flight to the positions tracked in video recordings from a ground-based fisheye camera (see Section 17.3.3). To obtain the tracked wing position from the video image we apply the VMT approach in Section 17.5. The tethered wing used

in this section consists of a 3 m^2 main wing and an elevator for depower during retractions.

Figure 17.7 shows a frame from the tracking results during the retraction phase of a small-scale tethered wing flight at 200 m line length. In Figures 17.8 and 17.9 the tracked image positions are compared against line-based position measurements. Note that the presented VMT approach provides 2D information and results are therefore presented in the (θ,ϕ)-plane only. We further see that VMT tends to lose the object regularly but the tracking failure is detected correctly and limited to short instances. As a result target loss is indicated correctly and the VMT returns no erroneous tracked positions.

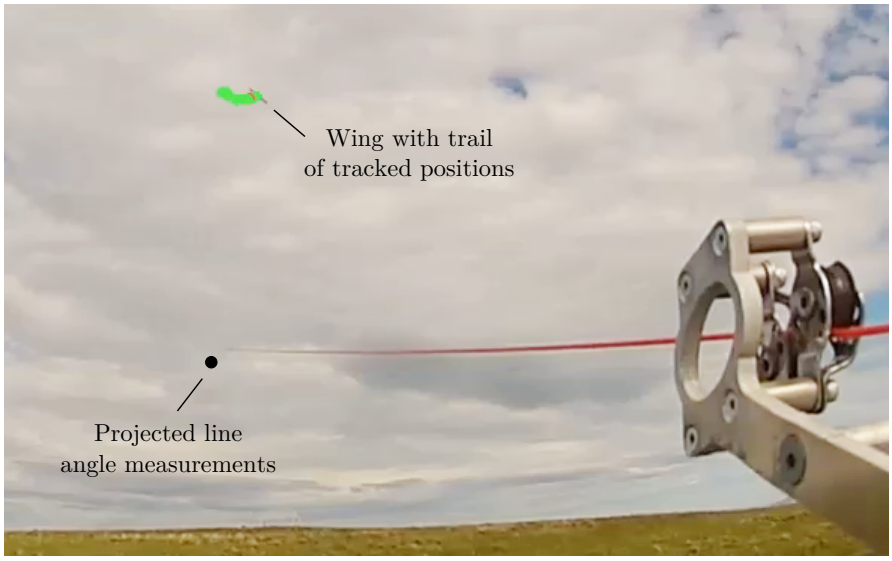

Fig. 17.7 Overlay of last 20 tracked positions of a tethered wing using VMT in green with the instantaneous projected line angle measurement in black.

The overlay in Figure 17.7 hints at the weakness of line-based position measurements during retractions where we desire low line forces to improve cycle efficiency. Drops in tether tension can significantly deteriorate the quality of line-based measurements leading to increased estimation delay and large errors due to line sag. From Figure 17.8 we can see that such events are not limited to the retraction phase but can also occur in turns flying a figure-eight trajectory. While the effect of line dynamics is less obvious in the azimuth measurements, we can clearly see large errors in elevation measurements of up to 0.1 rad even during the traction phase. This is further illustrated in Figure 17.9 which shows the tracked image positions over one figure-eight. The markers in Figure 17.9 also demonstrate the effect of delay in the line angle measurements. One would generally expect the delay to increase with

increasing line lengths [30] but Figure 17.8 clearly indicates variations of delay in the line-based elevation measurements.

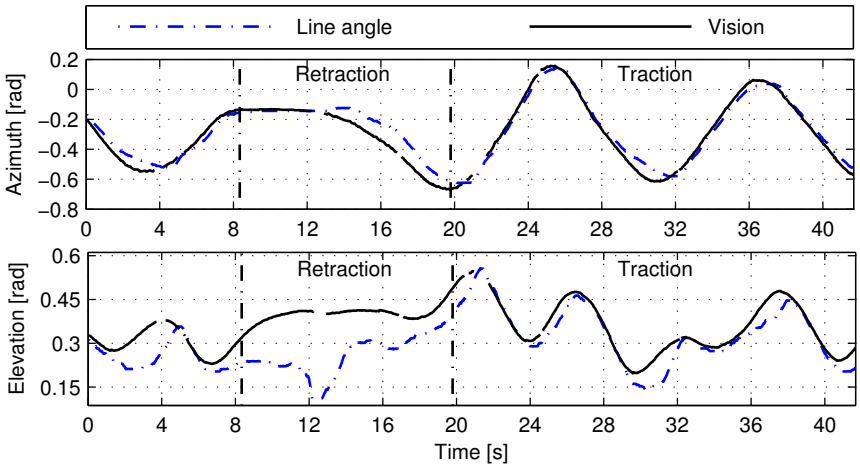

Fig. 17.8 Unfiltered position measurements of a tethered wing over a complete pumping cycle between 130m and 200m line lengths tracked at 24fps.

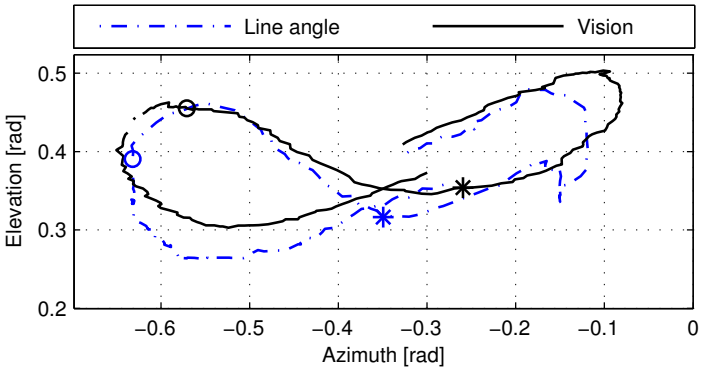

Fig. 17.9 Tracked image positions mapped in (θ, ϕ)-plane after calibration and correcting for the relative motion of the camera. Markers illustrate the instantaneous tracked positions at two different time instances.

In summary, line dynamics introduce varying lag, especially during low-line-tension scenarios such as uploop curves and retractions. This demonstrates the importance of effective winch control in ground-based AWE approaches to not only optimize the cycle power but also ensure sufficient estimation capabilities of the wing heading and position. The inertial sensor fusion approach, as demonstrated

next, can help to reduce the estimation delay allowing for lower line forces and hence improve system efficiency.

17.6.2 Inertial Sensor Fusion for Soft Kites

In this section we demonstrate the inertial sensor fusion for a HQ Apex III 10 m^2 kite at wind speeds of 3-6 m/s measured 5 m above ground. All estimators have been implemented on a Speedgoat Real-Time Target Machine [28] and demonstrated at 100 Hz in closed-loop operation on the ground-based FHNW system. The IMU data is streamed to the GS as described in Section 17.3.2. Note that the current FHNW system with two lines is designed for rigid wings with onboard depower capabilities. The soft kite in this experiment therefore remained fully powered during retractions and no significant line sag was observed. This test case nonetheless serves well to demonstrate the potential of limited onboard measurements in the proposed sensor fusion approach.

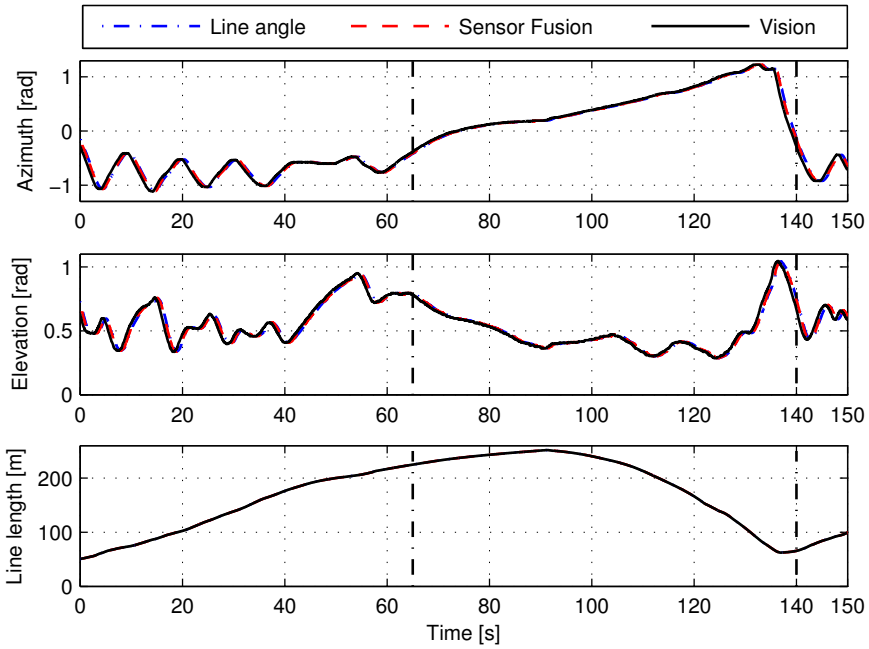

Fig. 17.10 Reference vision-based kite position estimate (solid) with sensor fusion (dashed) and line-angle (dotted) estimates over one pumping cycle. Retraction occurs between vertical lines.

We apply the VMT approach in Section 17.5 to obtain a reference solution that allows comparison of the different estimators presented in Section 17.4. An example

of the resulting position measurements over one full pumping cycle is presented in spherical coordinates in Figure 17.10, where the retraction phase (65-140 s) is indicated by vertical lines. From the video recording we can also compute estimates of the kite velocity vector orientation, $\hat{\gamma}$, as presented in Figure 17.11, where the regularized definition of the velocity vector orientation (17.7) was used during the retraction phase.

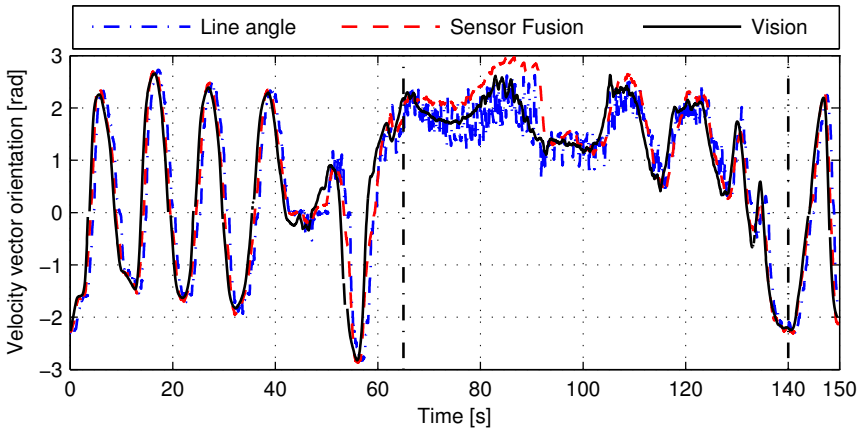

Fig. 17.11 Reference vision-based velocity vector orientation (solid) with sensor fusion (dashed) and line-angle (dotted) estimates over one pumping cycle. The regularized velocity vector orientation $\hat{\gamma}_{\text{reg}}$ is shown during retraction between vertical lines.

Figures 17.10 and 17.11 also show estimates of the position and velocity vector orientation obtained using sensor fusion (Section 17.4.2.2) and the line-based filtering scheme (Section 17.4.1) with the variances in Table 17.1. We tuned the covariance matrix of the process noise, Q, in the EKF equations of the sensor fusion approach to compromise between filter performance during traction and retraction phases. A delay of $\tau = 0.3$ s in the line angle measurements was assumed in (17.25).

The initial observation that no significant line sag was apparent during the experiment is confirmed in Figure 17.10. Since the kite could not be depowered during the retraction phase, line dynamics have no significant effect on the estimation of the kite position in this demonstration case. This is in sharp contrast to the results obtained with a tethered wing in Section 17.6.1 which requires no line tension to remain airborne. A delay of 0.6 s and 0.4 s is introduced in the estimation of the line angles using the line-based estimator and inertial sensor fusion, respectively.

During the traction phase a similar behavior can be observed in the estimation of the velocity vector orientation in Figure 17.11 where the line-based estimator and the sensor fusion results follow the vision-based reference with a slight delay. The exact estimation delay can be deduced from the root mean square (RMS) errors of line-based and sensor fusion estimates presented in Figure 17.12 for the traction phase ($t \leq 65$ s). The vision-based solution serves as reference to compute the re-

spective errors. By shifting the reference solution in the RMS error computation by the indicated delay, we can see that incorporating inertial measurements can reduce the estimation delay from 0.9 s to 0.2 s. This is a significant reduction in estimation delay compared to a common overall system delay of 1.5 s [31].

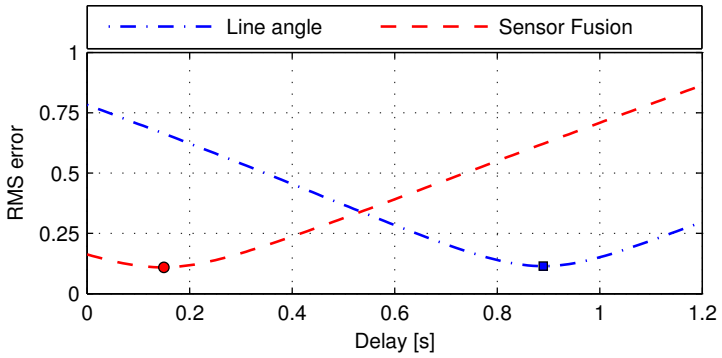

Fig. 17.12 RMS errors between the line-based and sensor fusion estimate of the velocity vector orientation $\hat{\gamma}$ during the traction phase. The vision-based solution shifted by the indicated delay is used as reference.

Note that tuning of the respective filter is a compromise between noisy feedback variables and estimation delay. We can generally see however that the yaw-rate measurements require less filtering, compared to line-based estimators, to arrive at a satisfactory estimate of the velocity vector orientation. A similar trend can be seen in Figure 17.11 during the retraction phase ($t > 65$ s) where the noise on line-based measurements is usually magnified due to the reduced kite velocity. Such oscillations combined with an increased system delay during this critical phase can significantly affect the performance of pumping cycle controllers, e.g. [34]. Sensor fusion in comparison provides smoother estimates closer to the vision-based reference solution. This is also reflected in Table 17.2 summarizing the RMS errors of both estimators during traction and retraction phases.

Table 17.2 RMS errors in the estimated velocity vector orientation, $\hat{\gamma}$, and position in (θ,ϕ)-plane given as $\hat{\mathbf{p}}_{\theta\phi} = \begin{bmatrix}\hat{\phi} & \hat{\theta}\end{bmatrix}^\top$ during traction and retraction phases. The vision-based solution is taken as reference.

Estimator	**Traction**		**Retraction**	
	$\hat{\gamma}$ (rad)	$\|\|\hat{\mathbf{p}}_{\theta\phi}\|\|$ (rad)	$\hat{\gamma}$ (rad)	$\|\|\hat{\mathbf{p}}_{\theta\phi}\|\|$ (rad)
Line angle	0.78	0.09	0.39	0.09
Sensor fusion	0.16	0.04	0.38	0.06

17.7 Conclusion

Automatic control of ground-based kite power systems commonly relies on ground-based measurements of line angles and lengths to estimate the kite position and heading for feedback control. In a two-phase operation line dynamics can significantly deteriorate the performance of flight controllers due to increased estimation delay and line sag. We have therefore proposed a sensor fusion scheme to reduce the estimation delay in ground-based kite power systems in pumping operation by fusing yaw-rate and line measurements. Limiting the extent of incorporated onboard information to one variable enables robust telemetry over long tether lengths demonstrated up to 280 m. The developed extended Kalman filtering scheme for sensor fusion has been demonstrated on experimental data to significantly reduce the estimation delay compared to an estimator relying solely on ground-based line angle and length measurements. Although the demonstration case was limited to soft kites, which showed little influence of line dynamics, we expect the strength of inertial-based sensor fusion to be exposed for longer line lengths.

We further developed a visual motion tracking approach to obtain reference solutions, or a ground truth, of the kite position and heading from ground-based video streams. Combining fast motion tracking with accurate object detection allows long-term tracking of kites in real time. In this work we demonstrated the visual motion tracking capabilities on experimental data of a tethered wing in pumping operation. The vision-based position estimates showed large discrepancies to ground-based line measurements during uploops and retractions when the tether forces are low. Line dynamics effectively limits the achievable cycle efficiency of rigid wings with ground-based estimation. The proposed sensor fusion can help to reduce the detrimental effects of line dynamics even in the presence of delayed line measurements.

The vision-based results revealed the complex nature of line dynamics leading to time-varying delay over a whole pumping cycle and even a single figure-eight. The current sensor fusion approach is however limited to static measurement delay of up to 50 time steps in real-time operation due to the augmentation of the filter state vector to incorporate the delay states. Future development of estimators for ground-based kite systems will therefore focus on fusion of spatially and temporally misaligned sensors in a computationally efficient manner.

We can further characterize the line dynamics as a time-varying bias which can be modeled or identified using the visual motion tracking results as a reference solution. Modeling of the line dynamics ensures that ground-based measurements can be fully exploited, but the proposed sensor fusion scheme could further benefit from additional onboard measurements. In particular, measurement of the kite velocity, which is a crucial quantity in the unicycle model employed in the developed extended Kalman filter, would improve the state estimation in the prediction step.

Acknowledgements This research was support by the Swiss National Science Foundation (Synergia) No. 141836 and the Swiss Commission for Technology and Innovation (CTI). We further acknowledge C. Houle, D. Aregger, and J. Heilmann from the Fachhochschule Nordwestschweiz (FHNW) for their test support and providing all the hardware used in the experiments. Develop-

ment of the hardware architecture enabling onboard measurements was done by M. Rudin and A. Millane (ETH Zurich). We are grateful for their support. The authors acknowledge the SpeedGoat Greengoat program.

References

1. Anderson, B. D. O., Moore, J. B.: Optimal Filtering. English. Dover Publications, Mineola, N.Y., USA (2005)
2. Appel, R., Fuchs, T., Dollár, P.: Quickly Boosting Decision Trees – Pruning Underachieving Features Early. In: International Conference on Machine Learning (ICML), Vol. 28, pp. 594–602, Atlanta, USA, June 2013
3. Autonomous Airborne Wind Energy Project (A^2WE). Accessed 31. October 2015. a2we.skpwiki.ch.
4. Bormann, A., Ranneberg, M., Kövesdi, P., Gebhardt, C., Skutnik, S.: Development of a Three-Line Ground-Actuated Airborne Wind Energy Converter. In: Airborne Wind Energy, Chap. 24, pp. 427–437. Springer, Berlin Heidelberg (2013). doi: 10.1007/978-3-642-39965-7_24
5. Bourdev, L., Brandt, J.: Robust object detection via soft cascade. In: Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 236–243, San Diego, CA, USA, June 2005. doi: 10.1109/CVPR.2005.310
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 886–893, San Diego, CA, USA (2005). doi: 10.1109/CVPR.2005.177
7. Dollár, P., Appel, R., Belongie, S., Perona, P.: Fast Feature Pyramids for Object Detection. **36**(8), 1532–1545 (2014). doi: 10.1109/TPAMI.2014.2300479
8. Dollár, P., Appel, R., Kienzle, W.: Crosstalk Cascades for Frame-Rate Pedestrian Detection. In: European Conference on Computer Vision, pp. 645–659, Florence, Italy (2012). doi: 10.1007/978-3-642-33709-3_46
9. Erhard, M., Strauch, H.: Sensors and navigation algorithms for flight control of tethered kites. In: European Control Conference (ECC), pp. 998–1003, Zurich, Switzerland (2013)
10. Erhard, M., Strauch, H.: Theory and Experimental Validation of a Simple Comprehensible Model of Tethered Kite Dynamics Used for Controller Design. In: Ahrens, U., Diehl, M., Schmehl, R. (eds.) Airborne Wind Energy, Chap. 8, pp. 141–165. Springer, Berlin Heidelberg (2013). doi: 10.1007/978-3-642-39965-7_8
11. Ess, A., Leibe, B., Gool, L. V.: Depth and appearance for mobile scene analysis. In: International Conference on Computer Vision (ICCV), pp. 1–8, Rio de Janeiro, Brazil (2007). doi: 10.1109/ICCV.2007.4409092
12. Fagiano, L., Huynh, K., Bamieh, B., Khammash, M.: On Sensor Fusion for Airborne Wind Energy Systems. IEEE Transactions on Control System Technology **22**(3), 930–943 (2014). doi: 10.1109/TCST.2013.2269865
13. Fagiano, L., Zgraggen, A. U., Morari, M., Khammash, M.: Automatic Crosswind Flight of Tethered Wings for Airborne Wind Energy: Modeling, Control Design, and Experimental Results. IEEE Transactions on Control System Technology **22**(4), 1433–1447 (2014). doi: 10.1109/TCST.2013.2279592
14. Freund, Y., Schapire, R. E.: Experiments with a New Boosting Algorithm. In: International Conference on Machine Learning (ICML), pp. 148–156, Bari, Italy, July 1996
15. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. The Annals of Statistics **28**(2), 337–407 (2000)
16. Gray, R. M.: Toeplitz and Circulant Matrices: A Review. Foundations and Trends in Communications and Information Theory **2**(3), 155–239 (2006). doi: 10.1561/0100000006

17. Heikkila, J., Silvén, O.: A four-step camera calibration procedure with implicit image correction. In: Conference on Computer Vision and Pattern Recognition, pp. 1106–1112, IEEE, San Juan, Puerto Rico, June 1997
18. Henriques, J. F., Caseiro, R., Martins, P., Batista, J.: High-Speed Tracking with Kernelized Correlation Filters. **37**(3), 583–596 (2015). doi: 10.1109/TPAMI.2014.2345390
19. Lefferts, E. J., Markley, F. L., Shuster, M. D.: Kalman filtering for spacecraft attitude estimation. Journal of Guidance, Control, and Dynamics **5**(5), 417–429 (1982)
20. MATLAB® Computer Vision System Toolbox™ Reference, MathWorks, Inc., Natick, MA, USA, 2015. http://mathworks.com/products/computer-vision
21. Millane, A., Wood, T. A., Hesse, H., Zgraggen, A. U., Smith, R. S.: Range-Inertial Estimation for Airborne Wind Energy. In: Conference on Decision and Control (CDC), pp. 455–460, Osaka, Japan, Dec 2015. doi: 10.1109/CDC.2015.7402242
22. Pixhawk Autopilot. Accessed 31. October 2015. https://pixhawk.org/modules/pixhawk.
23. Sabatini, A. M.: Kalman-filter-based orientation determination using inertial/magnetic sensors: Observability analysis and performance evaluation. Sensors **11**(10), 9182–9206 (2011)
24. Savage, P. G.: Strapdown Inertial Navigation Integration Algorithm Design Part 2: Velocity and Position Algorithms. Journal of Guidance, Control, and Dynamics **21**(2), 208–221 (1998)
25. Schölkopf, B., Herbrich, R., Smola, A. J.: A Generalized Representer Theorem. In: Computational Learning Theory, pp. 416–426. Springer-Verlag, Berlin, Germany (2001)
26. Shawe-Taylor, J., Cristianini, N.: Kernel methods for pattern analysis. Cambridge University Press (2004)
27. Smeulders, A. W. M., Chu, D. M., Cucchiara, R., Calderara, S., Dehghan, A., Shah, M.: Visual tracking: An experimental survey. **36**(7), 1442–1468 (2014). doi: 10.1109/TPAMI.2013.230
28. Speedgoat User Story: Efficiently harnessing wind power high above the ground using autonomous kites, Speedgoat GmbH, Liebefeld, Switzerland, 2015. https://www.speedgoat.ch/Portals/0/Content/UserStories/ethz_user_story.pdf
29. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 511–518, (2001). doi: 10.1109/CVPR.2001.990517
30. Wood, T. A., Hesse, H., Zgraggen, A. U., Smith, R. S.: Model-Based Flight Path Planning and Tracking for Tethered Wings. In: Conference on Decision and Control (CDC), pp. 6712–6717, Osaka, Japan, Dec 2015. doi: 10.1109/CDC.2015.7403276
31. Wood, T. A., Hesse, H., Zgraggen, A. U., Smith, R. S.: Model-Based Identification and Control of the Velocity Vector Orientation for Autonomous Kites. In: American Control Conference (ACC), 2015, pp. 2377–2382, Chicago, IL, USA, July 2015. doi: 10.1109/ACC.2015.7171088
32. Wu, Y., Lim, J., Yang, M. H.: Online object tracking: A benchmark. In: Conference on Computer Vision and Pattern Recognition, pp. 2411–2418, (2013). doi: 10.1109/CVPR.2013.312
33. Wu, Y., Lim, J., Yang, M.-H.: Object Tracking Benchmark. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 37, 9, pp. 1834–1848, (2015). doi: 10.1109/TPAMI.2014.2388226
34. Zgraggen, A. U., Fagiano, L., Morari, M.: Automatic Retraction and Full-Cycle Operation for a Class of Airborne Wind Energy Generators. IEEE Transactions on Control System Technology (2015). doi: 10.1109/TCST.2015.2452230