Visual-Inertial UAV Attitude Estimation Using Urban Scene Regularities

Myung Hwangbo and Takeo Kanade

Abstract— We present a drift-free attitude estimation method that uses image line segments for the correction of accumulated errors in integrated gyro rates when an unmanned aerial vehicle (UAV) operates in urban areas. Since man-made environments generally exhibit strong regularity in structure, a set of line segments that are either parallel or orthogonal to the gravitational direction can provide visual measurements for the absolute attitude from a calibrated camera.

Line segments are robustly classified with the assumption that a single vertical vanishing point or multiple horizontal vanishing points exist. In the fusion with gyro angles, we introduce a new Kalman update step that directly uses line segments rather than vanishing points. The simulation and experiment based on urban images at distant views are provided to demonstrate that our method can serve as a robust visual attitude sensor for aerial robot navigation.

I. INTRODUCTION

Attitude estimation is one fundamental step towards stabilizing an aircraft and achieving autonomous navigation of UAVs (Unmanned Aerial Vehicles). Research efforts for reliable and robust attitude (roll and pitch) estimation have been made using a variety of sensors such as IMU (Inertial Measurement Unit), magnetometer, thermal sensor, camera and GPS. In a low cost and small payload navigation system for UAVs, no individual sensor has been able to provide a sufficient level of information for the required accuracy. A combination of a complementary set of multiple sensors is, therefore, needed to improve stability and eliminate error accumulation over long periods of operation.

In recent years, vision sensors that enable natural scenes to act as an alternative attitude source have been preferred as a complementary measurement. Not only is a camera self-contained and passive like an inertial sensor but it is also interactive with its environment. Detection of a reliable world reference in images can provide absolute attitude measurement with a relatively constant level of accuracy. Horizon detection [1][2][3], for example, has been a prevalent method for vision-based attitude estimation for mid-altitude missions. Figure 1(a) shows the horizon, equal to a straight line separating the sky and the ground regions, which expresses camera attitude directly. For human vision, the horizon is commonly served as an attitude reference by pilots to control an aircraft. Unfortunately this method is not applicable at low altitudes since the horizon loses visibility due to the occlusion of its surroundings, as shown in Figure 1(b)-(c).

There still exists another visual cue to infer attitude from urban images. Since man-made environments generally

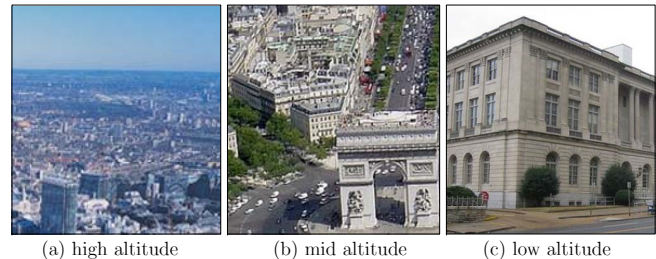(a) high altitude (b) mid altitude (c) low altitude

Fig. 1. Urban scenes at different altitudes: (a) at a high altitude, the horizon is clearly visible for obtaining the attitude. (b-c) at lower altitude, the horizon is occluded by man-made structures but it is possible to infer the invisible horizon from the vanishing points of the parallel edges on the structures.

exhibit strong regularity in structure, a camera will observe a number of lines which are either parallel or orthogonal to the gravity direction. Consequently, theses structures allow us to obtain a set of vanishing points, intersections of each parallel line bundle on an image, which have an explicit geometric relationship with the absolute attitude.

Using the advantage of this urban scene regularity, we propose a driftless attitude estimator that combines inertial gyro data and visual straight lines in a Kalman filter framework. Instantaneous relative rotation is traced by a tri-axial IMU and its unbounded error accumulation is then reset by gravity-related image line measurements. Each line also needs to be correctly associated with a vertical vanishing point or horizontal vanishing point since its corresponding Kalman update is different. Note that the scene of our interest is not limited to a close view of urban structuressuch as the facade of a building or indoor scene. We also consider a distant view of dense constructions like that in Figure 1(b), which is a more likely setting for fixed-wing UAV operations. These distant images are more challenging because an estimation method is required to endure a sparse set of short and noisy line segments and to reject outliers. In summary, our two main contributions are (i) the formulation of a Kalman filter system designed for image line measurements and (ii) a robust line classification scheme of vertical, horizontal and outlier edges.

This paper is organized as follows. In Section II, related work is discussed. Section III describes the relationship between attitude, vanishing points and the horizon in a projective space. The sensor fusion framework and line classification methods are presented in Section IV and V respectively. The evaluation results for our proposed estimator obtained by simulation and outdoor experiments are shown in Section VII Section VIII, respectively.

Myung Hwangbo and Takeo Kanade is with The Robotics Institute at Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA (myung@cs.cmu.edu, tk@cs.cmu.edu)

II. RELATED WORK

Attitude estimation for miniaturized UAVs has been typically studied in terms of estimation fusion of multiple low-cost sensors. Vast exploration has been done regarding fusion of IMU, GPS and magnetometer in a Kalman filter framework aimed at a full state of the vehicle including position and orientation. In the use of a single sensor, Taylor *et al* [4] used a tiny pair of infrared thermopiles to measure the absolute attitude from the temperature difference between the sky and the ground, but its accuracy and resolution are not reliable or precise at low attitudes. Euston *et al* [5] developed a nonlinear complementary filter that provides long-term drift free attitude estimates using only the IMU. They remove dynamic acceleration from raw accelerometer output using a simplified vehicle dynamic model and then combine the recovered gravitational direction with integrated gyro angles. Their simplified model is, however, not appropriate for highly aggressive maneuvers in a city canyon.

In vision-based methods, when the horizon is visible, it can provide an absolute camera attitude from images taken at mid and high attitudes [1] [2] [3] [6]. Ettinger *et al* [1] detected the horizon by extracting a straight line that separates the sky and the ground based on context differences of both regions, and Demonceaux *et al* [6] extended it to the detection of a curved horizon line in catadioptric images.

Alternatively, when the horizon is occluded and not visible, the vanishing points present in an urban scene have been analyzed to infer this invisible horizon [7] [8] [9]. For example, a set of mutually orthogonal vanishing points was found from parallel line bundles and sky detection identified the vertical vanishing point [8]. Bazin *et al* [9] estimated a current rotation matrix including the yaw angle by tracking vanishing points over a catadioptric image stream and computing the infinite homography on every frame. Antone and Teller [7] developed a batch process to recover the history of camera orientations from the bundle adjustment of vanishing points. Our approach in this paper is also based on vanishing point detection, however we use raw line measurements directly to refine the attitude and do not require line tracking. Fusing these with the IMU gyro angle, we compare each line segment with a current best attitude estimate. The ability to cope with occasional textureless images makes our method more tolerant to real urban UAV task scenarios.

Vanishing point detection problems have been explored to take advantage of the strong structural regularities found in man-made environments. Comprehensive evaluation between several approaches can be found in [10] and [11]. We represent vanishing points of a calibrated camera on a Gaussian sphere [12] [13] and the consistency measure based on Rother's work [14]. Our algorithm used in this paper is categorized as an edge-based, RANSAC-based, and non-iterative method.

III. ATTITUDE FROM VANISHING POINTS

For a pinhole camera, the perspective projection of parallel lines in a scene intersects at a single point on an image called the *vanishing point* (VP). A geometric relationship between
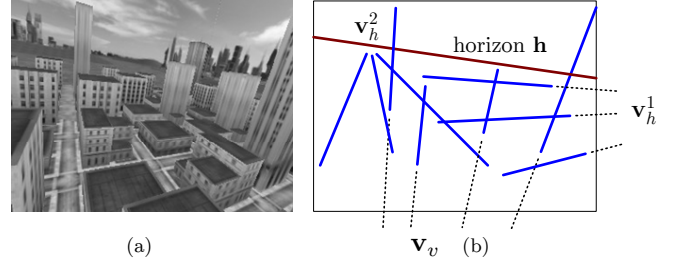

Fig. 2. Manhattan world example: sets of parallel lines converge at a single vertical VP ($\mathbf{v}_v$) and at horizontal VPs ($\mathbf{v}_h^1$, $\mathbf{v}_h^2$) on the image. The horizon **h**, equivalent to camera attitude (roll and pitch), can be derived from either $\mathbf{v}_v$ or a vanishing line connecting $\mathbf{v}_h^1$ and $\mathbf{v}_h^2$.
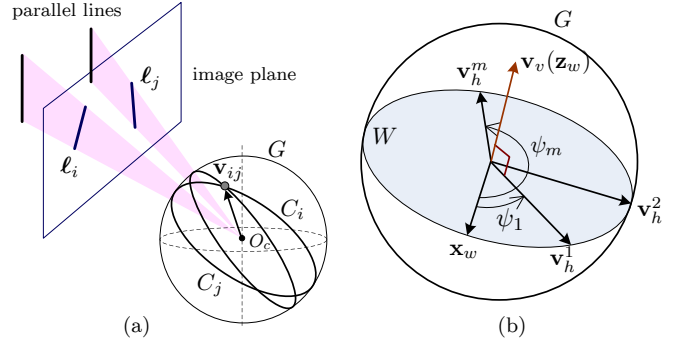

Fig. 3. (a) Gaussian sphere representation of parallel lines $\boldsymbol{\ell}_i$ and $\boldsymbol{\ell}_j$ meeting at a single point $\mathbf{v}_{ij}$ (b) Camera attitude indicated by either an antipodal point corresponding to the single vertical VP ($\mathbf{v}_v$), or the great circle **G** on which multiple horizontal VPs ($\mathbf{v}_h^1, \ldots, \mathbf{v}_h^k$) lie.

camera orientation, the horizon, and vanishing points along principle world coordinates has been well established in a Gaussian sphere [12] using 2D projective geometry when a camera calibration matrix **K** is given. All vanishing points, including VPs at infinity, can be equally considered in Gaussian sphere representation.

A. Gaussian sphere

The Gaussian sphere G in Figure 3(a) is a unit sphere located at the camera's optical center O_c. In the 2D projective space $\mathbf{P}^2$, an image line $\ell_i \in R^{3 \times 1}$ is represented as a normal vector of the great circle C_i in homogeneous coordinates. The intersection of the two parallel edges ℓ_i and ℓ_j is a vanishing point $\mathbf{v}_{ij}$, which is computed using the duality between the points and lines in the projective plane, *i.e.*, $\mathbf{v}_{ij} = \ell_i \times \ell_j$. $\mathbf{v}_{ij}$ is viewed as the direction to the corresponding 3D point at infinity.

In man-made structures, we can arguably claim that vertical edges are merged into a single vanishing point ($\mathbf{v}_v$) equivalent to the gravity direction $\mathbf{z}_w$ in the world coordinate. We call $\mathbf{v}_v$ the *vertical vanishing point*. On the other hand, horizontal edges (orthogonal to the gravity direction) are possibly grouped by multiple vanishing points residing in the world ground plane. We call them *horizontal vanishing points* $\mathbf{V}_h = (\mathbf{v}_h^1, \ldots, \mathbf{v}_h^m)$. Also, $\mathbf{v}_v$ and $\mathbf{V}_h$ are subject to the following geometric constraint in a calibrated camera:

$$\mathbf{v}_v^\top \mathbf{v}_h^i = 0 \quad i = 1, \ldots, m \tag{1}$$

Figure 2 illustrates each vanishing point corresponding to an orientation in the scene and vice versa. In general, unless the scene satisfies the Manhattan world assumption, the horizontal vanishing points in $\mathbf{V}_h$ are not necessarily orthogonal to each other. Therefore, in this paper, no constraints are made for $\mathbf{V}_h$ other than (1).

The vanishing points that lie on the same plane in a scene define a *vanishing line* in an image. The horizon $\mathbf{h}$ in Figure 2(b) is equal to the vanishing line that connects any two horizontal vanishing points ($\mathbf{v}_h^i$ and $\mathbf{v}_h^j$). In $\mathbf{P}^2$, line $\mathbf{h}$ is *dual* to point $\mathbf{v}_v$. This fact can be geometrically interpreted to mean that the horizon $\mathbf{h}$ is the projection of the world ground plane $\mathbf{W}$ of which the normal $\mathbf{z}_w$ is projected on $\mathbf{v}_v$.

$$\mathbf{h} = \mathbf{v}_h^i \times \mathbf{v}_h^j, \quad \mathbf{v}_v = \mathbf{v}_h^i \times \mathbf{v}_h^j, \quad (2)$$

The orientation of the great circle $\mathbf{W}$ of the Gaussian sphere in Figure 3(b) is the same as that of the world ground plane with respect to the camera. Therefore, if a relative camera pose with respect to an airplane is known, airplane attitude can be determined when either the $\mathbf{v}_v$ or at least two $\mathbf{v}_h$ are recovered from an image

B. Vertical vanishing point ($\mathbf{v}_v$)

Assume that a camera c is attached to an airplane b so that the camera's principle axis is aligned along the fuselage centerline. The $\mathbf{v}_v$ is a perspective projection of the world z-axis $\mathbf{z}_w^c$ with the camera matrix $\mathbf{P} = [\mathbf{I}\ \mathbf{0}]$. Camera translation has no effect on the projection of a point at infinity. The gravity direction $\mathbf{z}_w^c$ in the camera coordinate after coordinate transformations is,

$$\mathbf{v}_v = \mathbf{P}\,\mathbf{z}_w^c = \mathbf{R}_b^c\,\mathbf{R}_w^b\,\mathbf{z}_w^w = \begin{bmatrix} -s_\phi c_\theta \\ -c_\phi c_\theta \\ -s_\theta \end{bmatrix} \quad (3)$$

where

$$\mathbf{R}_b^w = \begin{bmatrix} c_\theta c_\psi & -c_\phi s_\psi + s_\phi s_\theta c_\psi & s_\phi s_\psi + c_\phi s_\theta c_\psi \\ c_\theta s_\psi & c_\phi c_\psi + s_\phi s_\theta s_\psi & -s_\phi c_\psi + c_\phi s_\theta s_\psi \\ -s_\theta & s_\phi c_\theta & c_\phi c_\theta \end{bmatrix}, \quad \mathbf{R}_b^c = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix}$$

Let $\mathbf{v}^* = (v_x, v_y)^\top$ be a vanishing point $\mathbf{v}$ on the image plane in the inhomogeneous coordinates. Then the location of the vertical vanishing point on an image is determined only by its attitude (ϕ, θ).

$$\mathbf{v}_v^* = \begin{bmatrix} \dfrac{x_w^c}{z_w^c}, & \dfrac{x_w^c}{z_w^c} \end{bmatrix}^\top = \begin{bmatrix} \dfrac{s_\phi c_\theta}{s_\theta}, & \dfrac{c_\phi c_\theta}{s_\theta} \end{bmatrix}^\top \quad (4)$$

Once $\mathbf{v}_v^*$ is found in an image, the attitude can be immediately recovered by

$$\phi = \text{atan2}\,(v_x, v_y), \quad \theta = \text{atan}\dfrac{1}{\sqrt{v_x^2 + v_y^2}} \quad (5)$$

The horizon $\mathbf{h}$ on the image is a line described by $\mathbf{v}_v$.

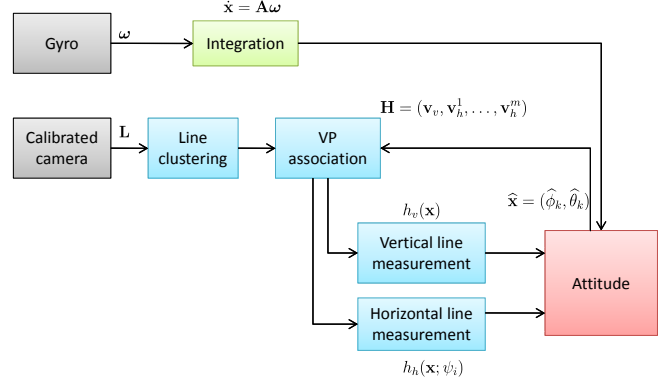$$(s_\phi c_\theta)x + (c_\phi c_\theta)y + s_\theta = 0 \quad (6)$$

Fig. 4. Sensor fusion of integrated gyro angle and image line measurements in the extended Kalman filter framework for realtime attitude estimation

C. Horizontal vanishing point ($\mathbf{v}_h$)

Without loss of generality, we can choose the projection of $\mathbf{x}_w^c$ as the horizontal vanishing point $\mathbf{v}_h$.

$$\mathbf{v}_h = \mathbf{P}\,\mathbf{x}_w^c = \mathbf{R}_b^c\,\mathbf{R}_w^b\,\mathbf{x}_w^w$$
$$= \begin{bmatrix} c_\phi s_\psi - s_\phi s_\theta c_\psi, & -(s_\phi s_\psi + c_\phi s_\theta c_\psi), & c_\theta c_\psi \end{bmatrix}^\top \quad (7)$$

$$\mathbf{v}_h^* = \begin{bmatrix} \dfrac{c_\phi s_\psi - s_\phi s_\theta c_\psi}{c_\theta c_\psi}, & \dfrac{-s_\phi s_\psi - c_\phi s_\theta c_\psi}{c_\theta c_\psi} \end{bmatrix}^\top \quad (8)$$

All the points in $\mathbf{V}_h$ are located along the horizon $\mathbf{h}$ and their locations are indicated by different yaw angles $(\psi_1, \ldots, \psi_m)$. See Figure 2(b) and Figure 3(b).

IV. SENSOR FUSION IN KALMAN FILTER

We follow the extended Kalman filter (EKF) framework to combine inertial and visual sensor for realtime attitude estimation. Figure 4 shows that the high-rate attitude prediction from inertial gyro output is updated by low-rate visual line measurements. One point to note is delayed measurement; vision processing may take a significant time until a designated measurement is available from a raw image. Our robust line classification takes up to 1 second when a lightweight embedded onboard system is used. If the measurement at time $t - k$ is available at time t, we simply reevaluate the state and its covariance from $t - k$ to t since the computation load of the process model is low.

A. Process Model

We use a kinematic process model that is comprised of the mechanization of a strap-down tri-axial gyroscope and a gyroscope bias error model, *i.e.* $(\mathbf{x}, \boldsymbol{\epsilon})^\top = (\phi, \theta, \epsilon_x, \epsilon_y, \epsilon_z)^\top$.

$$\dot{\mathbf{x}} = \begin{bmatrix} 1 & \sin\phi\tan\theta & \cos\phi\tan\theta \\ 0 & \cos\phi & -\sin\phi \end{bmatrix} (\boldsymbol{\omega}_g - \boldsymbol{\epsilon} + \mathbf{n}_g) \quad (9)$$

$$\dot{\boldsymbol{\epsilon}} = \boldsymbol{\epsilon} + \mathbf{n}_\epsilon \quad (10)$$

where $\boldsymbol{\omega}_g$ indicates a (scaled) raw gyro output, $\mathbf{n}_g$ and $\mathbf{n}_\epsilon$ are gyro measurement noise and gyro bias error noise, respectively. Note that the Coriolis effect is neglected and the gyro's scale and bias errors are modeled as single time-varing bias terms $\boldsymbol{\epsilon}$.

B. Measurement Update

All the detected line segments are first sent to the line classification algorithm in Section V to find vanishing points and to remove outliers. A set of valid lines $\mathbf{L}$ is then returned with the horizon measurement $\mathbf{H} = (\mathbf{v}_v, \mathbf{v}_h^1, \ldots, \mathbf{v}_h^m)$. Each line segment $\boldsymbol{\ell}_i \in \mathbf{L}$ is described by

$$\boldsymbol{\ell}_i = (\alpha_i, \mathbf{m}_i, s_i; \mathbf{v}_i \in \mathbf{H}) \quad \text{for } i = 1, \ldots, n \tag{11}$$

where $\boldsymbol{\ell}_i$ is parameterized as a midpoint $\mathbf{m}_i$, angle α_i, length s_i and associated vanishing point $\mathbf{v}_i$ in $\mathbf{H}$.

Figure 5 shows the expected line measurement at the current best attitude estimate $\widehat{\mathbf{x}}$. The observation model $g(\cdot)$ is an angle α of the line that connects the line midpoint $\mathbf{m}$ and an associated vanishing point $\mathbf{v}^*$ on the image. The vanishing point $\mathbf{v}^*$ is directly coupled with the attitude by either (4) or (8) depending on whether $\mathbf{v}^*$ is a vertical VP or one of horizontal VPs in $\mathbf{H}$. If the line is a vertical edge associated with $\mathbf{v}_v$, $\widehat{\mathbf{x}}$ is set to $\mathbf{v}^*$ using (4). On the other hand, if the line is a horizontal edge associated with $\mathbf{v}_h^k$, $\mathbf{v}^*$ is computed using (8) with an additional value ψ_k that specifies its position on the horizon $\mathbf{h}(\widehat{\mathbf{x}})$ in (6). ψ_k is estimated during the line classification procedure. At the midpoint $\mathbf{m}_i$ of each line segment $\boldsymbol{\ell}_i$, a line angle observation z_i of the true attitude is made as follows:

$$\widehat{\alpha}_i = \tan^{-1}\left(\frac{v_y^* - m_y}{v_x^* - m_x}\right) \tag{12}$$

$$= \begin{cases} g_v(\mathbf{v}^*, \mathbf{m}_i), \ \mathbf{v}^* = \mathbf{v}_v^*(\widehat{\mathbf{x}}) \text{ in (4)} & \text{if } \boldsymbol{\ell}_i \in \mathbf{v}_v \\ g_h(\mathbf{v}^*, \mathbf{m}_i), \ \mathbf{v}^* = \mathbf{v}_h^*(\widehat{\mathbf{x}}, \psi_k) \text{ in (8)} & \text{if } \boldsymbol{\ell}_i \in \mathbf{v}_h^k \end{cases}$$

$$z_i = \widehat{\alpha}_i + \mathbf{n}_\ell \qquad \text{for } i = 1, \ldots, n \tag{13}$$

where n_ℓ is a line angle noise. We set the variance of n_ℓ to be inversely proportional to the line length, $n_\ell = \lambda s_i^{-1}$, since a longer line segment is believed to include less noise resulting from quantization errors in the image processing step. The observation Jacobian $\mathbf{J}$ of the line angle is evaluated at $\widehat{\mathbf{x}}$ according to an associated vanishing point.

$$\mathbf{J}_v = \frac{\partial g_v}{\partial \mathbf{x}} = \left[\frac{-r^2 + r n_y}{d^2}, \ \left(\frac{1}{c_\theta^2}\right)\frac{r^2 n_x}{d^2}\right] \tag{14}$$

$$\mathbf{J}_h = \frac{\partial g_h}{\partial \mathbf{x}} = \left[\frac{-r^2 + (t_\psi/c_\theta)n_x - t_\theta n_y}{d^2}, \ \frac{n_x + t_\psi(-c_\theta + s_\theta n_y)}{(c_\theta d)^2}\right] \tag{15}$$

where r, d, n_x and n_y are defined in Figure 5.

The innovation in the Kalman update is an deviation angle e of a real measurement $\boldsymbol{\ell}_i$. This measurement update is sequentially repeated for all the line segments in $\mathbf{L}$ with different noise variance n_ℓ.

V. ROBUST LINE CLASSIFICATION

In most cases, line segments extracted from an image are not immediately usable for the Kalman update. Natural urban images may contain outliers, edges that are neither vertical nor horizontal in the gravitational direction. In addition, no knowledge is available about which vanishing point each line segment belongs to until a rigorous search for vanishing
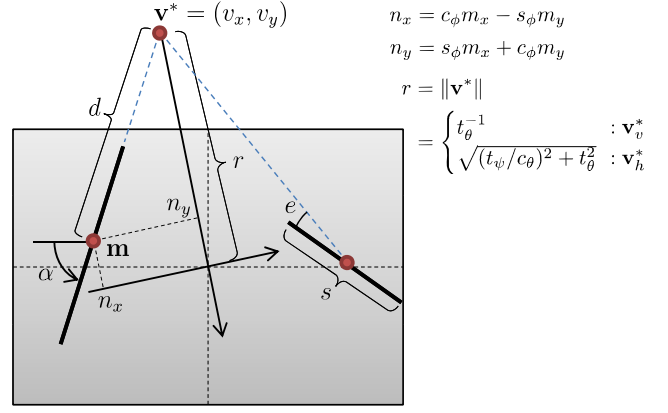

Fig. 5. The observation model $g(\cdot)$ of a line segment ℓ: The line angle α is obtained by connecting the midpoint m and the vanishing point $\mathbf{v}^*$ derived from the current attitude $\mathbf{x}$.

points is performed. Even though VPs are found, another analysis is needed to identify which VP is the vertical vanishing point or not if all VPs are horizontal.

We develop a RANSAC-based line clustering method to robustly find multiple vanishing points in the presence of outliers. The VP constraint (1) is also enforced to remove invalid vanishing points possibly driven by non-gravity lines.

A. Edge Detection and Line Fitting

A camera needs to be calibrated prior to use since we treat the camera as an angle sensor. We start by running the Canny edge detector on a raw image, and follow non-maximal suppression in order to obtain edges of one-pixel thickness. Edge points are then rectified and normalized using camera calibration matrix $\mathbf{K}$ and distortion factors. Each edge chain is recursively divided and fitted to a line segment with sub-pixel accuracy based on the thresholds of fitting error and minimum length (we use one pixel error and 10% of image size, respectively). For aerial images in our experiments, typically a minimum of twenty to several hundred line segments are obtained depending on the scene.

B. RANSAC-based Line Clustering

Assume the line detection procedure described above produces q line segments $\mathbf{L}_q = (\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_q)$. The first step of the line clustering is to partition $\mathbf{L}_q$ into p disjoint line groups in terms of vanishing point while p is unknown. Based on Rother's work [14], we employ a hypothesis-and-test scheme in the RANSAC, allowing us to avoid an exhaustive search $\mathcal{O}(q^3)$. Similar approaches can be found in [15] [16].

Our RANSAC procedure is as follows; At first, all the lines in $\mathbf{L}_q$ are marked as the unclassified group $\mathcal{U}$. One line pair $(\boldsymbol{\ell}_i, \boldsymbol{\ell}_j)$ is randomly sampled from $\mathcal{U}$ to create one vanishing point hypothesis, $\mathbf{v}_{ij} = \boldsymbol{\ell}_i \times \boldsymbol{\ell}_j$. This hypothesis is tested for all the lines in $\mathbf{L}_q$, given the score function (16), which is dependent on both the deviation angle e and the line length s in Figure 5. This score reflects that a vanishing point hypothesis with a minimum sum of angle errors weighted by line lengths is more likely to be a true one. The line segments

of a deviation angle larger than a user-defined threshold e_{th} are excluded when counting the score.

$$\text{Score}(\mathbf{v}; \mathbf{L}) = \sum_{i=1}^{q} f(\mathbf{v}, \boldsymbol{\ell}_i) \tag{16}$$

$$f(\mathbf{v}, \boldsymbol{\ell}_i) = \begin{cases} s_i \left(1 - \dfrac{e(\mathbf{v}, \boldsymbol{\ell}_i)}{e_{th}}\right) & \text{if } e < e_{th} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

where e is a deviation angle between a line segment and an ideal line connecting $\mathbf{v}_{ij}$ and the midpoint $\mathbf{m}$.

After a preset number of trials are repeated using different VPs, the best vanishing point hypothesis is chosen based on the score. Afterwards, a new line group $\mathcal{L}_1$ is created with the line members within e_{th} with respect to the chosen best hypothesis. The vanishing point $\overline{\mathbf{v}}_1$ which represents this line group $\mathcal{L}_1$ is recomputed as an optimal intersection point of all line members by maximizing (16) through nonlinear optimization. All the line members in $\mathcal{L}_1$ are excluded from $\mathcal{U}$ and the whole process is repeated until $\mathcal{U}$ is empty. Small line groups in which member size is less than the threshold are discarded since they contain outliers. Finally, a set of p vanishing points $\overline{\mathbf{V}} = (\overline{\mathbf{v}}_1, \ldots, \overline{\mathbf{v}}_p)$ and p line groups $\mathcal{L}_1, \ldots, \mathcal{L}_p$ can be obtained after the validation of our RANSAC procedure.

C. VP Association

The final step is to identify which vanishing point in $\overline{\mathbf{V}}$ corresponds to a single $\mathbf{v}_v$ and which others in $\overline{\mathbf{V}}$ correspond to possibly multiple $\mathbf{v}_h$s. We start this association process with the vertical VP prior $\widehat{\mathbf{v}}_v$ which is equivalent to the current best estimate $\widehat{\mathbf{x}}$ by (4). Under the assumption of a small estimation error, $\mathbf{v}_v$ is chosen as the closest point to $\widehat{\mathbf{v}}_v$ among $\overline{\mathbf{V}}$. Once $\mathbf{v}_v$ is determined, all the $\overline{\mathbf{v}}_i$s that satisfy the VP constraint (1) are selected as $\mathbf{V}_h$, *i.e.*,

$$\overline{\mathbf{V}} = (\overline{\mathbf{v}}_1, \ldots, \overline{\mathbf{v}}_p) \ \Rightarrow \ \mathbf{H} = (\mathbf{v}_v, \mathbf{v}_h^1, \ldots, \mathbf{v}_h^m) \subset \overline{\mathbf{V}} \tag{18}$$
$$s.t \ \ \mathbf{v}_v = \overline{\mathbf{v}}_{i^*} \ \text{ where } i^* = \arg\min_i |\widehat{\mathbf{v}}_v^\top \overline{\mathbf{v}}_i|,$$
$$|\mathbf{v}_v^\top \mathbf{v}_h^i| < \sin \delta_1, \ \ i = 1, \ldots, p \text{ and } m \leq p - 1$$

where δ_1 is a user-defined angle margin for selecting $\mathbf{v}_h$ since, in practice, the VP constraint (1) is not strictly satisfied due to image noise. Let $\mathbf{L}$ be a subset of $\mathbf{L}_q$ by selecting lines involved with $\mathbf{H}$ only. $\mathbf{H}$ could be refined via nonlinear optimization that minimizes both (1) and squared sum of angle deviation e of $\mathbf{L}$. Once we have done this, we call $\mathbf{H}$ *horizon measurement* which is an ordered and constrained subset of $\overline{\mathbf{V}}$. See Figure 3(a) for $\mathbf{H}$ on a Gaussian sphere.

Nonetheless, it is also possible that no vertical VP exists in $\overline{\mathbf{V}}$, for example, in cases a camera is looking down toward the ground. If the $\mathbf{x}_v$ corresponding to $\mathbf{v}_v$ that was found in (18) is located outside of an expected boundary in terms of Mahalanobis distance, $\mathbf{H}$ is searched again to obtain $\mathbf{v}_h$s without $\mathbf{v}_v$, *i.e.*,

If $D_m > D_{th}$,
$$\mathbf{H} = (\mathbf{v}_h^1, \ldots, \mathbf{v}_h^m) \subset \overline{\mathbf{V}} \ \ s.t \ \ |\widehat{\mathbf{v}}_v^\top \mathbf{v}_h^i| < \sin D_{th} \tag{19}$$
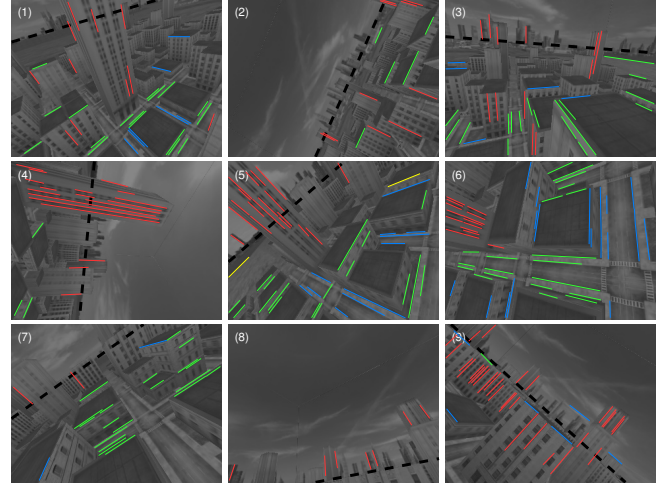

Fig. 6. (Simulated images) A Manhattan world is rendered in our graphical simulator. The ground-truth horizon is a dotted black line. Lines associated with $\mathbf{v}_v$ are in red and other lines associated with two $\mathbf{v}_h$s are in green and blue respectively.

where $D_m^2 = (\mathbf{x}_v - \widehat{\mathbf{x}}_v)^\top \mathbf{P}_\mathbf{x}^{-1} (\mathbf{x}_v - \widehat{\mathbf{x}}_v)$ and $\mathbf{x}_v = \mathbf{x}(\mathbf{v}_v)$ in (5). Finally, $\mathbf{H}$ and $\mathbf{L}$ are the inputs to the Kalman update in Section IV.

VI. DISCUSSION

The vertical vanishing point ($\mathbf{v}_v$) and the attitude $\mathbf{x} = (\phi, \theta)$ are identical entities. A pair of horizontal vanishing points $(\mathbf{v}_h^i, \mathbf{v}_h^j)$ also determines attitude. Therefore, $\mathbf{H}$ provides a different level of information regarding attitude, which can be categorized into the following three cases depending on its components:

- Redundant : $\mathbf{H}_1 = [\mathbf{v}_v, \mathbf{v}_h^1, \ldots, \mathbf{v}_h^m]$
- Full : $\mathbf{H}_2 = \mathbf{v}_v$ or $\mathbf{H}_3 = [\mathbf{v}_h^1, \ldots, \mathbf{v}_h^m]$
- Partial : $\mathbf{H}_4 = \mathbf{v}_h^1$

Redundant case $\mathbf{H}_1$ overdetermines the attitude from two causes which are $\mathbf{v}_v$ and $\mathbf{V}_h$. Full cases, $\mathbf{H}_2$ and $\mathbf{H}_3$, determine the attitude uniquely. Partial case $\mathbf{H}_4$ indicates that a single horizontal vanishing point cannot determine the attitude. However, $\mathbf{H}_4$ is still valid for the Kalman update because a Kalman filter enables the state to be estimated fully even from partial observation of the state. In simulation and experiments, we will compare the estimation performance in these measurement cases.

We use $\mathbf{L}$ rather than $\mathbf{H}$ in the Kalman update. Line segments in $\mathbf{L}$ is a lower-level and less-processed data from a raw image than vanishing points in $\mathbf{H}$. Therefore, $\mathbf{L}$ is more adequate for a Kalman filter because its noise distribution is closer to Gaussian, making it easier to model its variance than to model that of $\mathbf{H}$. Severe nonlinearity between VPs and the attitude causes a non-Gaussian noise in $\mathbf{H}$.

VII. SIMULATION

We developed a graphical simulator to test our visual-inertial attitude estimator during an urban operation. We use the airplane dynamic model [17] to simulate a small

TABLE I

Estimation comparison with partial, full and redundant cases of horizon measurement $\mathbf{H}$ (simulation results)

Sensors	Roll (°)	Pitch (°)	$\overline{P_\phi}$ (°)	$\overline{P_\theta}$ (°)
(a) gyro only	20.60 ± 17.89	2.44 ± 23.59	11.09	16.53
(b) gyro + $\mathbf{H}_4$	1.44 ± 2.55	0.44 ± 2.51	2.61	4.48
(c) gyro + $\mathbf{H}_3$	0.53 ± 1.47	-0.34 ± 1.50	0.81	1.35
(d) gyro + $\mathbf{H}_2$	0.46 ± 1.01	0.25 ± 0.93	0.75	1.07
(e) gyro + $\mathbf{H}_1$	0.30 ± 0.85	-0.25 ± 1.05	0.49	0.73

(See the section VI for the meaning of $\mathbf{H}_1, \ldots, \mathbf{H}_4$, $\overline{P_\phi}$ and $\overline{P_\theta}$ represent the means of pitch and roll error covariance.)

fixed-wing airplane with a 1m wing span. The IRRLICHT realtime rendering engine loads the 3D graphical model of a Manhattan world to emulate a $60°$ field-of-view camera as shown in Figure 6.

The simulation is focused on attitude estimation under continuously aggressive maneuvers. The airplane is driven manually by a computer-connected RC controller, which results in a large swing of pitch and roll angles up to $\pm 120°$ and $\pm 60°$ respectively. We sample a tri-axial gyroscope at 100Hz and intentionally set a large IMU variance ($\mathbf{n}_g$), 0.05 rad/s, in (10) in order to demonstrate the effectiveness of visual attitude sensing in short time intervals. The vision processing for producing line measurements $\mathbf{L}$ and horizon measurements $\mathbf{H}$ is identical to real experiment cases, except that a rendered image is captured. Figure 6 shows input image samples overlaying up to three line groups and the true horizon $\mathbf{h}$. The scale λ of the line measurement noise in (12) is set to 10.

Figure 7 shows estimation performance over a 30 second flight when all measurements in $\mathbf{L}$ are used, *i.e.*, case (e) in Table I. Estimation begins with an initial error $10°$ in both roll and pitch angles to see how fast the visual update reduces the error. The initial error diminishes nearly to zero in one visual update for the roll and in ten visual update for the pitch. Maximum estimation error is always bounded within $3°$ for both roll and pitch angles, and their variances are less than $2°$ on average. The greater occurrence of peaks in the pitch covariance demonstrates that the pitch is more sensitive to location perturbation of vanishing points than the roll in $\mathbf{J}_v$ (14) is. Each number of line measurements associated with $\mathbf{v}_v$ and $\mathbf{V}_h$ respectively is shown at the bottom of Figure 7. In most cases, the Kalman update is applied with redundant horizon measurement ($\mathbf{H}_1$) where both $\mathbf{v}_v$ and multiple $\mathbf{v}_h$ are observable from 30 line segments on average.

We compare these estimation performance with the partial, full and redundant cases of $\mathbf{H}$ as discussed in Section VI. We *intentionally* restrict the visual update to each case in Table I and irrelevant line measurements are abandoned. For example, Table I(e) is the result of using the line segments associated with the first $\mathbf{v}_h$ for attitude estimation. There is almost no difference between redundant and full cases ($\mathbf{H}_{1,2,3}$) in Table I(c)-(e). Even in the partial case ($\mathbf{H}_4$) in Table I(b) where only the first $\mathbf{v}_h$ is used in the Kalman update, no profound degrade in accuracy occurs, however, uncertainty increases significantly.
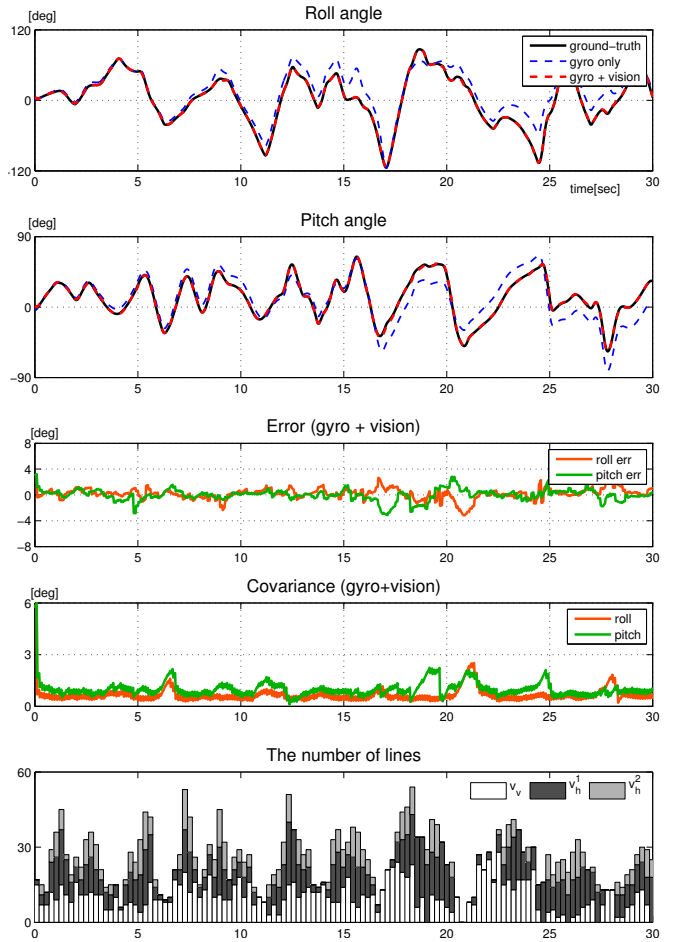

Fig. 7. Simulation results obtained during highly dynamic vehicle maneuvers and attitude estimation with redundant horizon measurement (gyro + $\mathbf{H}_1$): The estimation begins with an initial error $10°$. The performance measures are summarized in Table I(e). (Bottom) Each quantity of lines associated with $\mathbf{v}_v$ and the first two $\mathbf{v}_h$s respectively. In most cases, the horizon measurement is redundant as both $\mathbf{v}_v$ and multiple $\mathbf{v}_h$s are conceivable with sufficient amount of parallel lines.

VIII. EXPERIMENTS

A. Visual-inertial system

Figure 8 shows an RC model airplane of 1.2m wingspan named *SuperCub*. This customized airplane offers additional payload of up to 350g. The vehicle is equipped with an *O-Navi* MEMS-based IMU, a $70°$ field-of-view *Sentech* CCD camera, a 1GHz *VIA* pico-ITX computer and two LiPo batteries. The camera is attached with $30°$ pitch down to detect more visual features near the ground and captures a 320×240 low resolution image for realtime processing. The horizon measurement $\mathbf{H}$ can be obtained onboard at 5Hz from raw images. The IMU is sampled at 100Hz using a 11-bit A/D converter.

B. Outdoor experiments

We conducted outdoor flight experiments over a football field surrounded by low-profile buildings and a stand. The image examples of our urban scene are shown in Figure 11. Building edges and lane markers on the football field are
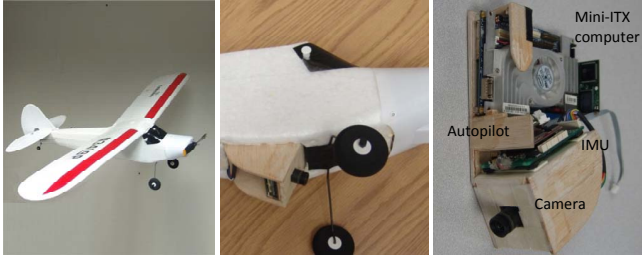
Fig. 8. A small RC model airplane of 1.2m wing-span equipped with a MEMS-based IMU, CCD camera and onboard computer.

TABLE II

Estimation comparison with partial, full and redundant cases of horizon measurement **H** (outdoor experiment)

measurement	roll (°)	pitch (°)	$\overline{P_\phi}$ (°)	$\overline{P_\theta}$ (°)
gyro only	2.47 ± 6.14	-0.24 ± 5.12	-	-
gyro + $\mathbf{H}_4$	0.68 ± 2.21	-0.25 ± 1.93	1.58	3.05
gyro + $\mathbf{H}_1$	0.61 ± 1.52	-0.49 ± 1.78	0.83	1.26

(See the section VI for the meaning of $\mathbf{H}_1, \ldots, \mathbf{H}_4$, $\overline{P_\phi}$ and $\overline{P_\theta}$ represent the means of pitch and roll error covariance.)

Fig. 9. Outdoor experimental results of inertial-visual attitude estimation (gyro + $\mathbf{H}_1$). The ground-truth is given by manual line fit of the horizon on the images.

Fig. 10. (Top) Each amount of lines associated with $\mathbf{v}_v$ and first two $\mathbf{v}_h$s respectively. Only a single $\mathbf{v}_h$ is detected between 9 to 10 second and makes a sharp increase of the pitch covariance in Figure 9(d). (Bottom) The average number of lines at each image is around 30.

detected as main line features. The camera and the IMU are calibrated in advance to obtain intrinsic parameters of both sensors and their relative orientation. We intentionally increase the IMU variance to 0.03 rad/s for the same reason in the simulation. Note that the ground-truth attitude of experimental data is obtained by drawing a manual line fit of the horizon on each image. Sky-ground boundaries and building structures enable us to draw on intuitive perception about the true horizon even with occlusion.

Compared with the simulations, in Figure 10 the number and the portion of vertical edges among line measurements are much smaller prior to 9 sec because no slender buildings are present around the test area. The number of lines increases greatly after 11 sec when the camera starts to see the football stand.

In the first half of Figure 9, even a small number of lines is sufficient for correcting the attitude without any error drift. Overall performance shows that attitude error is always bounded within 5° and its covariance is less than 1° when $\mathbf{v}_v$ or two $\mathbf{v}_h$s are observable, while the gyro inertial estimation accumulates the error. Especially when only a single $\mathbf{v}_h$ is available, corresponding with the moment of 9 to 10 sec in Figure 9, the pitch covariance increases rapidly, up to 3° even though its error remains small.

Figure 11 shows outdoor image samples, classified line segments and current attitude estimate on a Gaussian sphere. In the last column, current attitude estimate $\hat{\mathbf{x}}$ is plotted as entities such as pole axis and great circle on a Gaussian sphere. The horizon measurements **H**, a list of vanishing points detected by our RANSAC-based procedure, are plotted together to visually manifest how large the deviation between the visual measurements and attitude estimates. Figure 11(5) shows the case $\mathbf{H}_4$ where only a single $\mathbf{v}_h$ is
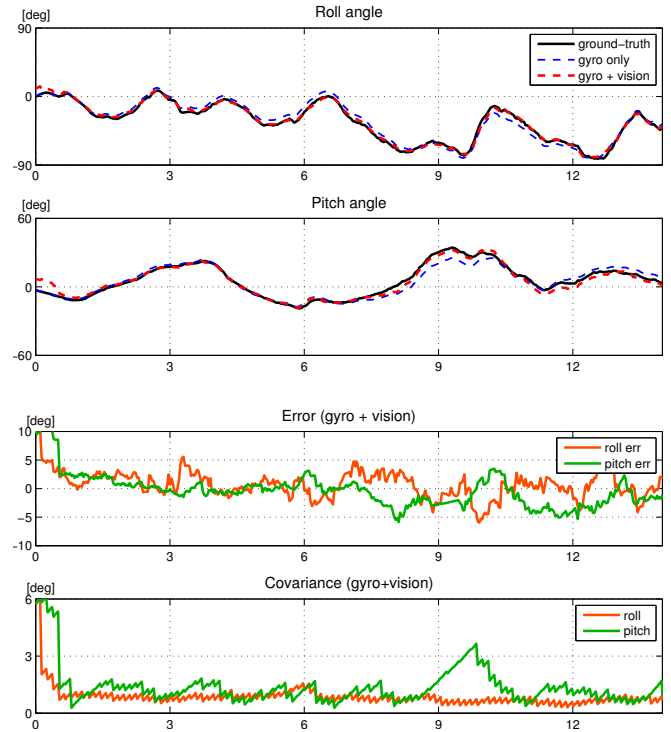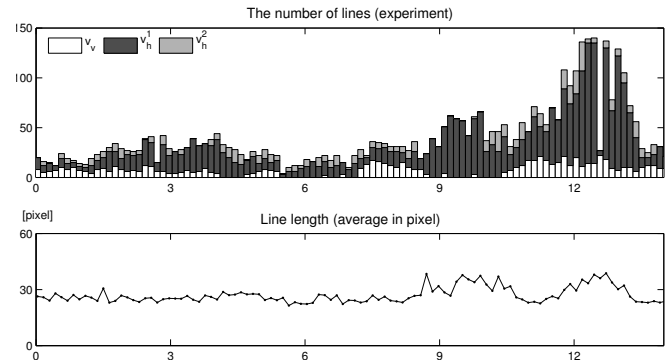
found and Figure 11(6) shows a case $\mathbf{H}_3$ where two $\mathbf{v}_h$s without $\mathbf{v}_v$ are found.

The average number of line segments per image is around 30 in Figure 10, while the average length is around 30 pixels in a 320×240 image. One drawback of our line extraction method, based on Canny edge detector, is that edge chain discontinuity caused by the noise or the collision with other chains can leave visibly long edges vulnerable to break into short pieces. Nonetheless, a collection of short line segments can still provide sufficient information about vanishing points in the scene to bound attitude error.

IX. Conclusion

We demonstrate that drift-free attitude estimation can be achieved using line regularities that exist in urban scenes mainly from man-made structures. Since attitude is equivalent to the vertical vanishing point in an image, attitude error in IMU gyro can be corrected by using gravitational line segments, once their vanishing points are identified and validated by the geometric constraint.

We also present the measurement update procedure which uses validated line segments in an extended Kalman filter framework. Lines are robustly classified into groups associated with vanishing points by the RANSAC method. Our simulation and experiments show that the attitude error is bounded within $3°$ even during aggressive vehicle maneuvers.

References

[1] S. M. Ettinger, M. C. Nechyba, P. G. Ifju, and M. Waszak, "Vision-guided flight stability and control for micro air vehicles," in *Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, 2002, pp. 2134–2140.

[2] S. Todorovic, M. C. Nechyba, and P. G. Ifju, "Sky/ground modeling for autonomous MAV flight," in *Proc. IEEE Int'l Conf. on Robotics and Automation*, May 2003, pp. 1422–1427.

[3] T. Cornall, G. Egan, and A. Price, "Aircraft attitude estimation from horizon video," *Electronic Letters*, vol. 42, no. 12, 2006.

[4] B. Taylor, C. Bil, and S. Watkins, "Horizon sensing attitude stabilisation: A VMC autopilot," in *18th International UAV Systems Conference*, 2003.

[5] M. Euston, P. Coote, R. Mahony, J. Kim, and T. Hamel, "A complementary filter for attitude estimation of a fixed-wing uav," in *Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, Sept 2008, pp. 340–345.

[6] C. Demonceaux, P. Vasseur, and C. Pégard, "Omnidirectional vision on UAV for attitude computation," in *Proc. IEEE Int'l Conf. on Robotics and Automation*, 2006, pp. 2842–2847.

[7] M. E. Antone and S. Teller, "Automatic recovery of relative camera rotations for urban scenes," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Sept 2000.

[8] C. Demonceaux, P. Vasseur, and C. Pégard, "UAV attitude computation by omnidirectional vision in urban environment," in *Proc. IEEE Int'l Conf. on Robotics and Automation*, 2007.

[9] J.-C. Bazin, I. Kweon, C. Demonceaux, and P. Vasseur, "UAV attitude estimation by vanishing points in catadioptric images," in *Proc. IEEE Int'l Conf. on Robotics and Automation*, 2008.

[10] P. Denis, J. H. Elder, and F. J. Estrada, "Efficient edge-based methods for estimating manhattan frames in urban imagery," in *European Conf. on Computer Vision*, 2008, pp. 197–210.

[11] J. A. Shufelt, "Performance evaluation and analysis of vanishing point detection techniques," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 3, pp. 282–288, 1999.

[12] S. T. Barnard, "Interpreting perspective images," *Artificial Intelligence*, vol. 21, pp. 435–462, 1984.

[13] R. T. Collins and R. S. Weiss, "Vanishing point calculation as a statistical inference on the unit sphere," in *Int'l Conf. on Computer Vision*, 1990, pp. 400–403.

[14] C. Rother, "A new approach for vanishing point detection in architectural environments," *Journal Image and Vision Computing, Special Issue on BMVC 2000*, vol. 20, no. 9–10, pp. 647–656, 2002.

[15] D. G. Aguilera, J. G. Lahoz, and J. F. Codes, "A new method for vanishing points detection in 3D reconstruction from a single view," in *Proceedings of the ISPRS Commission V*, 2005, pp. 197–210.

[16] H. Wildenauer and M. Vincze, "Vanishing point detection in complex man-made worlds," in *IEEE Int'l Conf. on Image Analysis and Processing*, 2007, pp. 615–622.

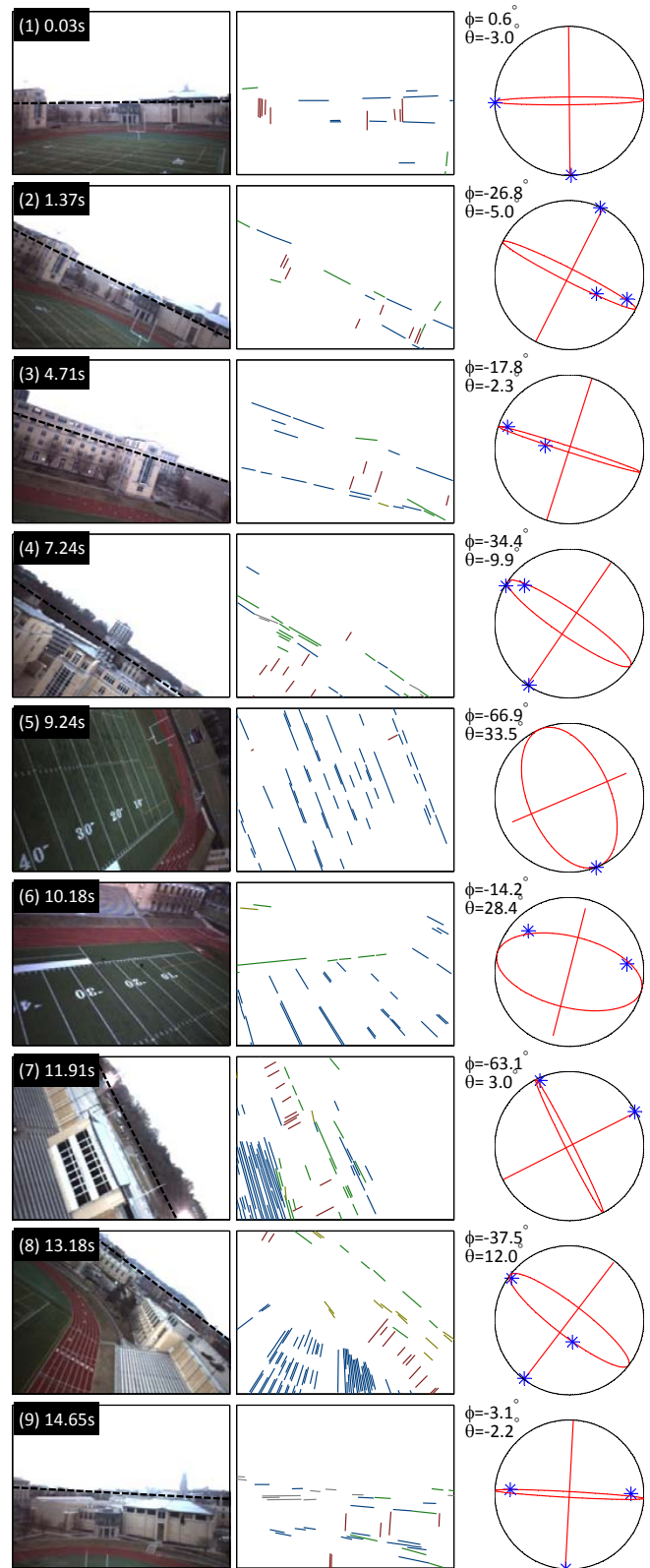[17] R. W. Beard, *An Introduction to Autonomous Miniature Air Vehicles*. Unpublished, 2006.

Fig. 11. (Left column) Outdoor images taken during one turn around over a football field. (Middle column) Line segments grouped by corresponding vanishing points are in different colors respectively after outliers are rejected. (Right column) Attitude estimate and vanishing points on the Gaussian sphere: the pole axis and the great circle indicate the current attitude estimate $\hat{\mathbf{x}}$ before Kalman update.