Estimating Ego-Motion in Panoramic Image Sequences with Inertial Measurements

Felix Schill, Robert Mahony and Peter Corke

Abstract This paper considers the problem of estimating the focus of expansion of optical flow fields from panoramic image sequences due to ego-motion of the camera. The focus of expansion provides a measurement of the direction of motion of the vehicle that is a key requirement for implementing obstacle avoidance algorithms. We propose a two stage approach to this problem. Firstly, external angular rotation measurements provided by an on-board inertial measurement unit are used to de-rotate the observed optic flow field. Then a robust statistical method is applied to provide an estimate of the focus of expansion as well as a selection of inlier data points associated with the hypothesis. This is followed by a least squares minimisation, utilising only the inlier data, that provides accurate estimates of residual angular rotation and focus of expansion of the flow. For the robust estimator we consider and compare RANSAC and a *k*-means algorithm. The least squares optimisation is solved using a geometric Newton algorithm. The approach in this paper does not require explicit features, and can be applied to patchy, sparse optic flow fields. The approach is demonstrated in simulations and on video data obtained from an aerial robot equipped with panoramic cameras.

1 Introduction

The estimation of the ego-motion of a camera from observation of a sequence of images is a classical problem in the computer vision literature. Algorithms based on

Felix Schill and Robert Mahony
Department of Engineering, Australian National University, ACT, 0200, AUSTRALIA, {firstname}.{lastname}@anu.edu.au

Peter Corke
CSIRO ICT Centre, Pullenvale, Queensland, AUSTRALIA. peter.corke@csiro.au
Acknowledgements: This research was supported by the Australian Research Council through discovery grant DP0880509.

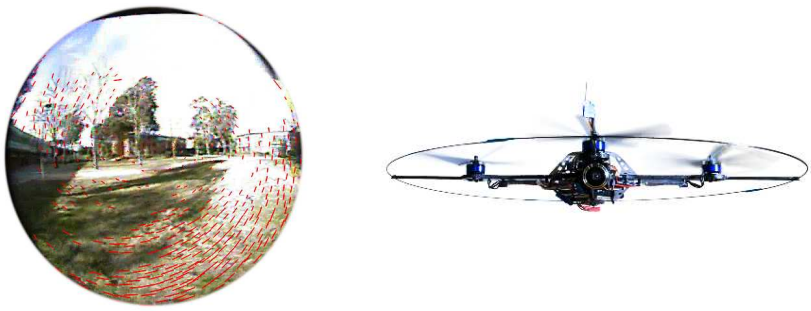

Fig. 1 Video image and optic flow (left) extracted from on-board video camera of the *Hummingbird* quad-rotor aerial robot (right)

eight or more point correspondences [18, 10] are well known, while for calibrated cameras algorithms exist for as few as five point correspondences [15, 22]. In the case where high accuracy is required the bundle adjustment method can be used [26]. In addition to the classical methods, there have been a wide range of other methods considered in the literature [14, 23, 12, 11, 3, 21].

It is well known that for an image sequence with a small field-of-view it is difficult to distinguish between translation and rotation around orthogonal axes [2, 7]. In addition, there is is often a natural bias in solving the instantaneous epipolar constraint, the most common approach to recovering instantaneous motion, associated with grouping of image points in one small area [7]. Using panoramic or catadioptric cameras with a large field-of-view can substantially overcome this issue [7]. Due to the inherent ambiguity in velocity there have been a number of studies based on qualitative analysis of ego-motion methods [5, 6, 25] that utilise panoramic cameras, however, these methods often use explicit search routines to determine the best motion fit and are computationally expensive. In recent work Lim and Barnes have developed methods to compute ego-motion from antipodal pairs of optic flow vectors [16, 17]. Almost all the literature in this area has been developed based on the assumption that the camera is the only sensor. In robotic applications, especially those involving aerial robots, there is almost always an inertial measurement unit (IMU) on the vehicle that can provide a substantially correct estimate of rotation over short periods. However, the vision system for such applications often has poor quality optics, and if the video signal is being transmitted to ground there are artifacts due to signal interference. The authors know of no prior work that addresses the specific issues associated with ego-motion extraction for such a situation.

In this paper, we propose an algorithm for extracting ego-motion from a panoramic image sequence where the angular velocity can be measured using a separate sensor. We are primarily motivated by applications in aerial robotics where the vehicle is equipped with a wide angle fish-eye (or catadioptric) lens and inertial sensors. The vision sequences obtained from such vehicles often contain large regions where there is insufficient texture to generate optic flow, for example, regions of sky, or re-

gions distant from the camera where the optics are of insufficient quality to generate good texture. This can make it difficult or impossible to find antipodal points. In addition, there are often extreme outliers in the optic flow field caused by errors in the optic flow algorithm induced by artifacts in the video due to signal interference and multi-path effects in video transmission. We propose a two stage approach. Firstly, optic flow is computed from the image sequence and then this flow is approximately de-rotated using the data from the gyroscopes (which are components of the vehicle's inertial measurement unit). This is achieved by subtracting the expected rotational optic flow (due to the measured angular velocity) from the measured optical flow. The resulting flow is almost entirely due to the translational ego-motion of the camera, except for errors and noise in IMU measurements and flow extraction, and has a simple structure that allows us to develop simple models to determine the unique focus of expansion (FoE), corresponding to the direction of motion of the camera. The focus of expansion estimate provides valuable information for object avoidance and vehicle guidance [20, 24].

We investigate two robust statistical methods, RANSAC and k-means, aimed at generating a reasonable hypothesis of the FoE of the flow and identify inlier and outlier optic flow measurements in the data. We then describe how an initial estimate of the focus of expansion can be refined in both translation and residual rotation by minimising a least squares cost based on the instantaneous epipolar condition posed on the sphere. We compute the geometric gradient and geometric Hessian and propose a Newton update step. The Newton minimisation is embedded in the RANSAC framework as the second model refinement step for each iteration. Given that the initial estimate provided by the first stage of the algorithm is moderately correct, this stage usually converges in at most three iterations. Moreover, the eigenvalues of the Hessian provide a measure of confidence in the estimate. A poor condition number for the Hessian indicates the likelihood of an unreliable estimate and the overall magnitude of eigenvalues is proportional to the distance scaled velocity of the vehicle. A common problem with Newton algorithms is poor convergence for badly chosen starting values. The presented problem is especially sensitive to rotation, however by using the rotation measurements from the onboard inertial sensors a reasonably close starting value can be computed.

The proposed algorithms are tested on synthetic data with outliers and noise, and demonstrated on video and inertial data obtained on a small aerial robotic vehicle.

2 Problem formulation

In this section, we introduce some notation and develop the cost function that will be used to refine ego-motion estimates on the sphere.

We are interested in applications where there is a wide field of view fish eye or catadioptric video camera moving through a static world environment. We assume that the camera frame rate is fast compared to the relative optical velocity of the observed scene, i.e. pixel displacements are sufficiently small that perspective changes

and lens distortions will not change significantly. Consequently the optic flow can be computed directly on the raw image sequence and the resulting flow vectors mapped back onto a spherical image plane based on a known calibration of the camera. The spherical optical flow field is denoted Φ and associates a flow vector $\Phi(\eta) \in T_\eta S^2$ in the tangent space of the sphere to a point $\eta \in S^2$ on the sphere. In practice, we are normally constrained to a sparse computation of optic flow, that is measurements at a finite number of points on the sphere indexed by $\{\eta_i\}$ for $i = 1, \ldots, n$, with n the number of optic flow vectors measured in a given reference image.

The optic flow can be split into translational and rotational components

$$\Phi(\eta) = \Psi(\eta) + \Theta(\eta). \tag{1}$$

Here $\Theta(\eta) := -\Omega \times \eta$, with $\Omega \in \{B\}$ the body-fixed frame angular velocity of the camera, is the contribution to the optic flow from the rotational motion of the camera, while the translation component of flow is given by $\Psi(\eta) := \frac{1}{\lambda(\eta)} \mathbb{P}_\eta v$, where $\mathbb{P}_\eta = (I_3 - \eta\eta^\top)$ is the projector onto $T_\eta S^2$ and $v \in \{B\}$ is the body fixed frame translational velocity of the vehicle.

In order to derive an optimisation-based method for identification of ego motion it is necessary to define a cost function. We propose to use a modified version of the instantaneous epipolar constraint. Let $\hat{w}$ denote the estimate of $w \in S^2$ of the true direction of motion of the vehicle. That is, set

$$w = \frac{v}{|v|}, \quad \text{for } v \neq 0,$$

and $\hat{w} \in S^2$ an estimate of $w \in S^2$. The direction of motion w is also the focus of expansion (FoE) of the translational flow field Ψ on the sphere.

For each individual optic flow vector $\Phi(\eta)$ measured at a point $\eta \in S^2$ the instantaneous epipolar constraint computed for estimates $\hat{w}$ and $\hat{\Omega}$ is

$$e_{\Phi(\eta)}(\hat{w}, \hat{\Omega}) := \langle \hat{w}, \big(\Phi(\eta) + \hat{\Omega} \times \eta\big) \times \eta \rangle. \tag{2}$$

Note that if $\hat{\Omega}$ is correct then $\Phi(\eta) + \hat{\Omega} \times \eta = \Psi(\eta)$ is the true translational optic flow. Taking the vector product of this with its base point η leads to a vector that is orthogonal to the direction of motion of the vehicle. Taking the inner product of the resulting vector with $\hat{w}$ is zero precisely when $\hat{w} = w$ is the true direction of motion. The instantaneous epipolar constraint is often written $\langle \hat{w} \times \eta, \big(\Phi(\eta) + \hat{\Omega} \times \eta\big) \rangle$, however, this can be transformed into (the negative of) Equation (2) using the properties of vector triple products and the form given above is more convenient for the gradient and Hessian computations undertaken later.

Since the optic flow is measured at a finite number of scattered points the cost considered is a sum

$$f(\hat{w}, \hat{\Omega}) := \sum_{i=1}^{n} e^2_\Phi(\eta)(\hat{w}, \hat{\Omega}) = \sum_{i=1}^{n} \langle \hat{w}, \big(\Phi(\eta) + \hat{\Omega} \times \eta\big) \times \eta \rangle^2 \tag{3}$$

It is clear that for ideal data the cost f is zero for the correct for $\hat{w} = w$ and $\hat{\Omega} = \Omega$. The cost f is a smooth function $f : S^2 \times \mathbb{R}^3 \to \mathbb{R}$ and can be optimised on this set using geometric concepts. A weakness of the cost proposed is that it is highly susceptible to perturbation by large magnitude outliers in the data. Optic flow algorithms often yield exactly this sort of error due to occasional mismatched point correspondences. Thus, direct minimisation of the cost f is likely to lead to poor ego-motion estimation. The next section applies robust statistical algorithms to overcome this issue.

3 Robust estimation of Focus of Expansion

In this section we present two robust statistical methods for providing an estimate of focus of expansion of the image sequence. The approach is directly based on the application domain and we assume that a measurement of angular velocity of the vehicle is available, for example through the inertial measurement unit that is mounted on the aerial vehicles that we consider. As both the camera and the inertial measurement unit are mounted on the vehicle we will use the vehicle coordinate frame for all calculations. Direction estimates can easily be transferred into a world frame by applying a rotation that can be obtained from a complementary attitude filter [19][4].

Using the measured angular velocity, $\Omega_y \in \{B\}$, the measured optic flow can be de-rotated.

$$\Psi_{\Omega_y}(\eta_i) := \Phi(\eta_i) - \Theta_{\Omega_y}(\eta_i) = \Phi(\eta_i) + \Omega_y \times \eta_i.$$

The resulting estimate, $\Psi_{\Omega_y}(\eta_i)$, of translational flow is only defined at measured flow points η_i.

Any two measurements of translational flow $\Psi_{\Omega_y}(\eta_i)$ and $\Psi_{\Omega_y}(\eta_j)$ can be used to generate a hypothesis for the focus of expansion of the flow field. Since the flow vectors must lie in a plane containing the flow vector and the base point η then the intersection of these two planes provides an estimate of the focus of expansion for the flow field. Thus,

$$\hat{w}(\Omega_y, \eta_i, \eta_j) := \frac{(\Psi_{\Omega_y}(\eta_i) \times \eta_i) \times (\Psi_{\Omega_y}(\eta_j) \times \eta_j)}{|(\Psi_{\Omega_y}(\eta_i) \times \eta_i) \times (\Psi_{\Omega_y}(\eta_j) \times \eta_j)|}. \tag{4}$$

This hypothesis is sign indeterminate so we introduce a corrected hypothesis $\hat{w}_c$ is

$$\hat{w}_c(\Omega_y, \eta_i, \eta_j) := \hat{w}(\Omega_y, \eta_i, \eta_j) \cdot \text{sign}(\langle \hat{w}, \Phi(\eta_1) \rangle) \tag{5}$$

There are potentially $n(n-1)/2$ (where n is the number of flow vectors measured) hypotheses that can be generated based on (4).

The k-means algorithm is applied by collecting a large number of of hypotheses and then clustering these hypotheses into classes. k-means clustering is a standard algorithm [9, 13] that is extensively used in the field of computer vision and we

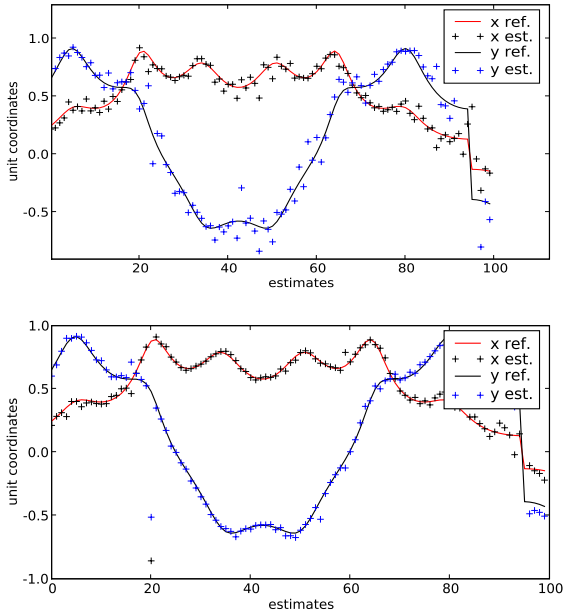

Fig. 2 Estimation results for *k*-means (top plot) and RANSAC (bottom plot), for synthetic flow, 30% outliers and 0.001 Gaussian noise (reference signal is the normalised true translation from the simulation) Note the significantly better performance of the RANSAC algorithm due to the Newton iteration step.

will not cover the details of the algorithm implementation. It is also possible to use different clustering algorithms. The distance between hypotheses for clustering purposes is the angle between two hypotheses. The estimate of the focus of expansion is provided by the centroid (the normalised mean) of the largest cluster of hypotheses, which also forms the inlier set. A Newton algorithm (described in Section 4) can be applied to the support set of the best cluster from the *k*-means algorithm to refine the best estimate, however it was found that this would often not converge if there are outliers present (see section 5). The described *k*-means algorithm can only generate an estimate for the focus of expansion, but not the residual rotation, and therefore relies on adequate removal of rotational flow.

To overcome these limitations, we merged the hypothesis generation of the *k*-means approach and a 6 DOF Newton model refinement into a more robust RANSAC framework. The RANSAC algorithm is based on consensus scoring of hypotheses and once again is a well known algorithm extensively used in computer vision applications [8, 10]. The algorithm is applied by randomly selecting pairs of flow vectors and generating hypotheses according to Equation 5 and using a normalised mean of the hypotheses generated. All flow vectors are then scored with regard to that hypothesis using the cost function (2) to determine the inlier or

consensus set, consisting of all flow vectors where $e_{\Phi(\eta)} < t$ (t is the acceptance threshold). The hypothesis along with its consensus set is then used to initialise the Newton iteration discussed in the next section and a refined estimate of both FoE and angular velocity is computed. Over several iterations, the refined estimate with the smallest residual is chosen as the best estimate from the algorithm.

4 Refining the motion estimate

In this section we present a geometric Newton algorithm that can be used to efficiently refine the estimates of ego-motion based on minimising the cost (3). The Newton algorithm requires a reasonable estimate of the local minima and identification of inlier flow vectors to provide a reliable estimate of ego-motion. The implementation of the Newton method also needs to respect the unit norm constraint on the focus of expansion estimate $\hat{w}$ in the optimisation problem. We achieve this by deriving the Newton algorithm with respect to the geometry of the constraint set. Details on geometric optimisation algorithms can be found in Absil *et al.* [1].

For the sake of simplifying notation we define

$$Z(\hat{\Omega}) := \left(\Phi(\eta) + \hat{\Omega} \times \eta\right) \times \eta. \tag{6}$$

The geometric gradient of f is an element of $T_{(\hat{w},\hat{\Omega})} S^2 \times \mathbb{R}^3$. It is obtained by differentiating $f(\hat{w},\hat{\Omega})$ in an arbitrary direction and then using the natural Riemannian metric to obtain a tangent vector;

$$\operatorname{grad} f(\hat{w},\hat{\Omega}) = \begin{pmatrix} \mathbb{P}_{\hat{w}} \left(\sum_{i=1}^{n} Z(\Omega) Z^{\top}(\Omega)\right) \hat{w} \\ -\sum_{i=1}^{n} \left((\hat{w}^{\top} Z(\Omega)) \mathbb{P}_{\eta_i} \hat{w}\right) \end{pmatrix}, \tag{7}$$

recalling that $\mathbb{P}_v = I_3 - vv^{\top}$ is the projection onto $T_v S^2$.

It is possible to consider a gradient descent method to optimise the cost function f defined in (3). However, due to the inherent nature of the data, the cost function is several orders of magnitude more sensitive to change in the angular velocity estimate than the FoE estimate, leading to very poor convergence of naive gradient descent algorithms. In practice, effective implementation of a gradient descent algorithm would require preconditioning of the gradient. Since an initial guess of the local minimum is available from the *k*-means or RANSAC algorithm (Section 3) it is possible to overcome this difficulty by using a Newton algorithm directly.

The geometric Hessian for f can be written

$$\operatorname{Hess} f(\hat{w},\hat{\Omega}) = \tag{8}$$

$$\begin{pmatrix} \mathbb{P}_{\hat{w}} \left(\sum_{i=1}^{n} Z(\Omega) Z(\Omega)^{\top}\right) \mathbb{P}_{\hat{w}} & -\mathbb{P}_{\hat{w}} \sum_{i=1}^{n} \left((\hat{w}^{\top} Z(\Omega)) \mathbb{P}_{\eta_i} + Z(\Omega) \hat{w}^{\top} \mathbb{P}_{\eta_i}\right) \\ -\sum_{i=1}^{n} \left((\hat{w}^{\top} Z(\Omega)) \mathbb{P}_{\eta_i} + \mathbb{P}_{\eta_i} \hat{w} Z(\Omega)^{\top}\right) \mathbb{P}_{\hat{w}} & \sum_{i=1}^{n} \left(\mathbb{P}_{\eta_i} \hat{w} \hat{w}^{\top} \mathbb{P}_{\eta_i}\right) \end{pmatrix}$$

The Hessian is written as an element of $\mathbb{R}^{6\times 6}$ due to the identification of tangent vectors in $T_{\hat{w}}S^2$ with elements of $\mathbb{R}^3$. However, the vector $\hat{w}$ is normal to the tangent space $T_{\hat{w}}S^2$ and it follows that $v_0 := (\hat{w}, 0) \in \mathbb{R}^6$ is a zero eigenvector of the Hessian $\mathrm{Hess}\, f$ in (8). The remaining five eigenvalues are associated with the quadratic structure of the cost f at the point $(\hat{w}, \hat{\Omega})$. Due to the zero eigenvalue the inverse Hessian in the Newton algorithm has to be implemented with a pseudo inverse routine. In addition, the new estimate must be re-normalised onto the sphere at each iteration of the Newton algorithm [1]. For initial conditions close to the minimum of f each iteration of the Newton algorithm provides an additional two orders of magnitude of accuracy. In practice, at most two or three iterations are sufficient for the purposes of our calculations given that a suitable initial condition is available.

As an additional advantage of applying the Newton algorithm, it is a straightforward exercise to compute the condition number of the Hessian as the ratio of the magnitudes of largest to smallest eigenvalues of the five meaningful eigenvalues of $\mathrm{Hess}\, f$ at the cost minimum. The condition number provides a reliability measure for the estimate of ego-motion of the system, a large condition number indicating that the minimisation is ill-conditioned. The eigenstructure of the Hessian can be used to identify directions of poor resolution of the ego-motion parameters.

5 Results

The combined algorithms were thoroughly tested with synthetically generated optical flow data, and on real video sequences obtained from a small-scale quad-rotor aerial vehicle. For the synthetic data, the true ego-motion of the vehicle is known and can serve as ground truth for comparisons. For the video sequences from the flying vehicle the inertial measurements from the on-board IMU were recorded. The measured rotations are used to de-rotate the spherical flow field. The trajectories of the vehicle can be compared qualitatively to the data obtained, however ground truth data was not available.

For the simulation tests the flow field was generated by creating a Gaussian-distributed point field (offset from origin: 18 along y-axis, sigma=10) for one-sided flow coverage. The offset of the point field simulates incomplete flow from a single camera, covering slightly less than half of the sphere. Tests for surrounding flow coverage where conducted on a Gaussian point field centered at the origin with the same variance.

Flow is created by translating and rotating the point field, and projecting start- and end position onto the sphere. A certain percentage of vectors (30%) is randomised to simulate large outliers, and Gaussian noise ($\sigma = 0.001$ or $\sigma = 0.002$) is added where appropriate in the simulations. Note that the Gaussian noise is applied directly to flow vectors on the unit sphere, which is independent of the pixel resolution of the camera. The value $\sigma = 0.002$ corresponds to 0.115 degrees on the sphere, which is approximately half a pixel for a PAL resolution camera with a 170 degrees field of view. The amount of outliers and noise approximately reflect the

One-sided flow

outliers, noise →	0%	0.0	0%	0.001	30%	0.0	30%	0.001	30%	0.002
k-means (translation)	0.061	0.03	2.1	1.8	1.2	1.1	3.1	2.50	5.6	4.9
RANSAC (translation)	0.001	0.001	1.0	0.9	0.5	0.003	2.2	1.6	7.7	3.4
k-means (with rot.)	15.9	15.4	15.2	14.0	17.9	16.0	20.0	16.0	17.6	13.7
RANSAC (with rot.)	0.74	0.002	6.0	0.9	19.6	13.7	12.8	10.2	13.2	12.7

Surrounding flow

outliers, noise →	0%	0.0	0%	0.001	30%	0.0	30%	0.001	30%	0.002
k-means (translation)	0.03	0.015	1.50	1.0	1.4	1.1	2.5	2.0	4.0	3.0
RANSAC (translation)	0.002	0.001	0.5	0.4	0.2	0.002	0.9	0.7	1.8	1.3
k-means (with rot.)	13.9	11.7	14.8	10.6	17.0	11.7	13.7	10.7	12.8	11.9
RANSAC (with rot.)	3.3	0.001	1.2	0.5	12.9	6.1	11.6	4.8	14.1	7.1

Fig. 3 Mean error (left value) and median error (right value) of FoE in degrees for synthetically generated flow with one-sided flow (upper table) and surrounding flow (lower table) coverage

distortions found in real optical flow. For simplicity, and independence of specific camera parameters, the simulation is unitless. The spherical camera is simulated by projecting scene points onto a unit sphere. In reality a fish-eye lens can be used, or other means of generating a spherical panoramic image.

The results can be seen in figure 3. Parameters for the *k*-means algorithm were k=20, and 70 randomly picked pairs. The RANSAC algorithm uses 8 iterations, a threshold $t = 0.05$, 20 vector pairs for the initial hypothesis. The Newton algorithm was run for four iterations.

A key observation found was that the performance of the *k*-means algorithm was not sufficiently reliable to use as the initialisation for the Newton algorithm. In particular, the segmentation of the image flow vectors was not sufficiently robust and the Newton iteration was often undertaken with some outliers that significantly disrupted the performance of the algorithm. As a consequence, the *k*-means results are presented without any refinement step while the RANSAC algorithm contains the Newton refinement as an integral part. The relative performance of the *k*-means (without Newton) versus the RANSAC (with Newton) is clearly seen in figure 2. This can also be seen in the results shown in Table 3. Nevertheless, in the absence of noise all algorithms perform well, even when flow can only be obtained from one side of the sphere. As noise increases, one-sided flow extraction becomes increasingly unstable (notably it is more affected by noise than by outliers). If residual rotation of up to 15 degrees/sec is present in the flow field (i.e. due to imperfect inertial measurements - see lines 3 and 4 in table 3), the estimation results worsen — again more pronounced in the one-sided case.

To test the real-world performance of the algorithm a panoramic PAL camera with a 170 degree field of view fisheye lens was mounted on a linear horizontal rail. The experiment was carried out indoors; most objects were within 2-4 m of the camera. The camera and lens are identical to those used on the quadrotor flying robot presented later. Optic flow is computed using the sparse implementation of pyramidal Lucas-Kanade (from OpenCV) on a 10 by 10 grid which is equally distributed

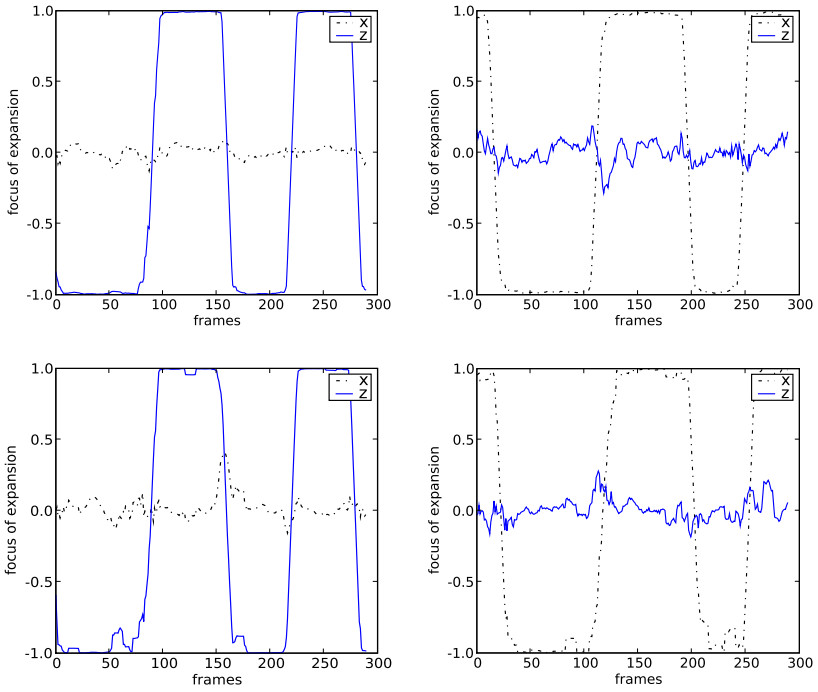

Fig. 4 FoE extraction for a camera mounted on a linear rail. The plots on the left show the results for the camera moving forwards and backwards along the optical axis (top: *k*-means, bottom: RANSAC), the plots on the right for the camera moving sideways perpendicular to the optical axis (again top: *k*-means, bottom: RANSAC).

across the image. Due to weight constraints on flying platforms the lens is of relatively low quality which results in significant blur towards the edges of the image. Despite the loss of texture it is still possible to extract usable optic flow from the periphery, albeit with slightly reduced confidence. The *k*-means and the RANSAC algorithm were applied to this data and the results are shown in figure 4. The camera did not rotate in this experiment, therefore it no IMU was used, and no derotation was applied. Both algorithms perform very well. Due to the absence of any rotation the *k*-means algorithm performs well, and the RANSAC algorithm does not improve the results.

The algorithms were also applied to video sequences that were collected from a small quad rotor flying vehicle (see figure 1). This electronically stabilised vehicle is equipped with a forward looking PAL camera, an IMU, and radio systems that transmit the real-time video images (25 fps) and inertial measurements to the base station. The IMU is a *ThinIMU micro* by Scinamics with 500 deg/sec full scale range, 3 g accelerometer and 12 bit resolution. Gyro drift after bias calibration is typically 1-2 degrees per minute. The IMU data is transmitted to the ground via a 2.4

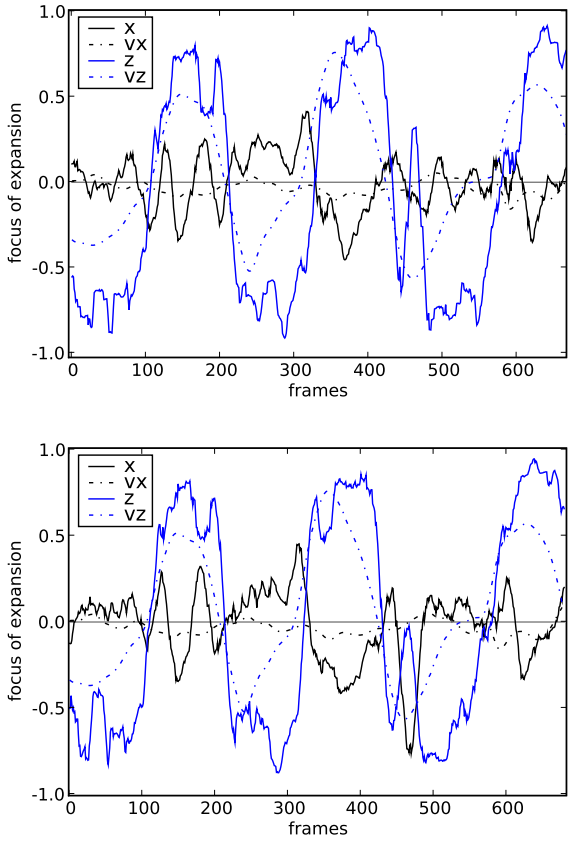

Fig. 5 FoE extraction for real world flyer scenarios, flying forwards and backwards along a straight line. The plot on the top shows the estimates from *k*-means, the lower plot shows the RANSAC results. The dashed lines are the approximate velocity data integrated from the inertial unit. Note that the FoE estimates are normalised, the velocity is not.

GHz *XBee* transmitter. Video is transmitted via a 5.8 GHz analog video link which offers good video quality at full resolution without compression artifacts and negligible latency. Video images are captured at 720×576 pixels by a PCI framegrabber card, but the captured images are downsampled to half resolution before computing optical flow which provided sufficient flow information for motion reconstruction. Efforts have been made to reduce all signal latencies to a minimum. Both signals (video and inertial measurements) were synchronised and recorded on the base station. Approximately 100 flow vectors were computed from each image. The first test sequence consists of the vehicle flying repeatedly forwards and backwards in a straight line for approximately 7-10 m. In the second test sequence the flight pattern is roughly circular (manually controlled). The flight tests were conducted outdoors.

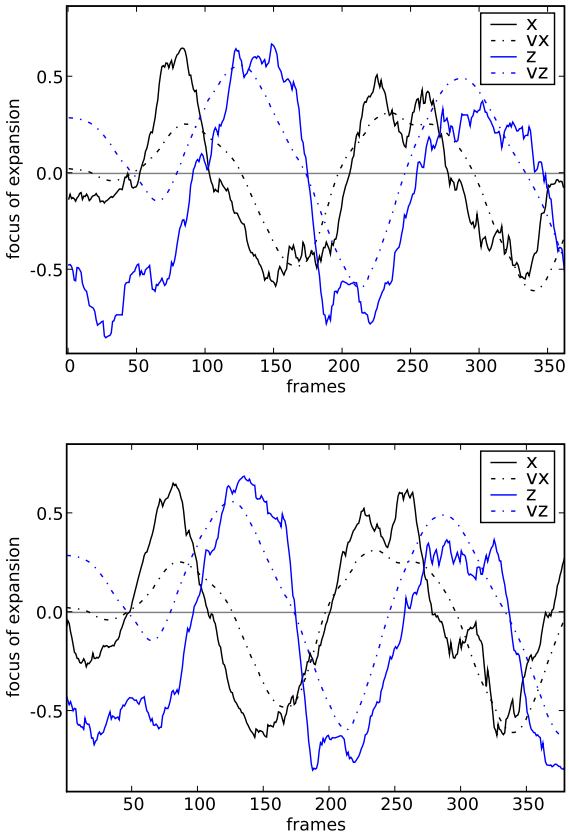

Fig. 6 The same as in figure 5, for the flyer following an approximate circle.

A few trees were within 2-3 m of the flight path (sequence 1 only) which produced strong translational flow due to the short range. Altitude above ground was approximately 1-2 m, and airspeed reached several meters per second. The results can be seen in figures 5 and 6. The *z*-axis denotes the forward direction of the flyer, the *x*-axis points to the right, and the *y*-axis points up.

Ground truth data was not available for the flight tests. However, for comparison the velocity estimate from the IMU is plotted as well. The velocity estimate was calculated by applying a complementary attitude filter with drift compensation, subtracting the estimated gravity vector from the accelerometer values, and integrating the residual acceleration. To avoid increasing drift errors a leaky integrator was used which pulls the velocity estimate back to zero over time. To reduce errors the IMU was carefully calibrated, and flights were kept short (20-30 seconds). Despite

the imperfections the estimates allow a qualitative comparison between two very different sensor modalities.

Figure 5 shows the FoE estimates sequence 1 (linear flight). The estimates for the sequence 2 (circular flight pattern) are shown in figure 6. The estimates are in vehicle (or camera) coordinates, thus, banking or pitching motions may result in deviations of the estimate from a straight line in world coordinates – however, such motions are small. A sliding window filter of 10 frames was applied to the plots. The plots show that for real data, both the *k*-means and RANSAC algorithm deliver reasonably good estimates of the direction of travel. The reduced performance compared to the initial experiment with a rail mounted camera can be attributed to the imperfect estimation of rotation, synchronisation errors between IMU and video, artifacts from the wireless video transmission (noise, occasional drop-outs), and larger accelerations and velocities. Significant differences between the performance of the two algorithms cannot be concluded as they are obscured by the higher level of noise. It is likely that the performance will improve significantly with more complete coverage of the sphere (i.e. two back-to-back cameras) — especially the RANSAC algorithm should benefit as rotation and translation become easier to separate. Future work will analyse the eigenvalues of the Hessian to measure the sensitivity and confidence for various flow vector distributions.

6 Conclusions

Two methods were presented for estimating the focus of expansion from sparse panoramic optical flow fields, namely a *k*-means clustering method, and a RANSAC framework using a Newton iteration for model fitting. We introduced a cost function, gradient and Hessian for estimation of direction of travel, and ego-rotation, which enables gradient descent methods and Newton methods for estimating ego-motion from spherical optic flow. The presented methods work with sparse, patchy flow fields, even if less than half the sphere is covered. Measurements from inertial sensors are used to provide a good initial value of rotation for the algorithms. The algorithms were evaluated and compared on synthetic data; it was found that the RANSAC algorithm performs better, but also that the *k*-means algorithm provides good results at much less computational cost. Tests on video and inertial measurements from a quad-rotor flying vehicle show that both algorithms can be applied to real data obtained from a single fish-eye camera, and provide a good estimate of the direction of travel. On real data the advantages of the RANSAC algorithm are negligible, as any possible improvements on estimation are masked by the higher sensitivity to noise. It is possible that the RANSAC algorithm as presented here will offer better performance is high quality cameras and well-calibrated quality optics are used; however this was not investigated. The presented algorithms can be used in robotics for obstacle avoidance — knowing the direction of travel relative to the world will indicate the point of potential impact. Divergence of optical flow can be computed in that direction to detect obstacles in the current path of the vehicles.

The presented cost function and estimation framework can also be used for visual servoing or visual odometry.

References

1. P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, New Jersey, USA., Princeton, NJ, January 2008. ISBN 978-0-691-13298-3.
2. G. Adiv. Inherent ambiguities in recovering 3-d motion and structure from a noisy flow field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:477–489, 1989.
3. A. Agrawal and R. Chellappa. Ego-motion estimation and 3d model refinement in scenes with varying illumination. In *IEEE Workshop on Motion and Video Computing, (MOTIONS 05)*, volume 2, pages 140–146, 2005.
4. Grant Baldwin, Rob Mahony, and Jochen Trumpf. A nonlinear observer for 6 dof pose estimation from inertial and bearing measurements. In *IEEE International Conference on Robotics and Automation (ICRA 09)*, 2009.
5. T. Brodsky, C. Fermuller, and Y. Aloimonos. Directions of motion fields are hardly ever ambiguous. *International Journal of Computer Vision*, 26(1):5–24, 1998.
6. C. Fermuller and Y. Aloimonos. Qualitative egomotion. *International Journal of Computer Vision*, 15:7–29, 1995.
7. C. Fermuller and Y. Aloimonos. Ambiguity in structure from motion: Sphere versus plane. *International Journal of Computer Vision*, 28(2):137–154, June 1998.
8. M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395, 1981.
9. J.A. Hartigan and M.A. Wong. Algorithm as 136: A k-means clustering algorithm. *Applied Statistics*, 28(1):100–108, 1979.
10. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
11. M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using region alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(3):268–272, 1997.
12. A. Jepson and D. Heeger. Subspace methods for recovering rigid motion i: algorithm and implementation. *International Journal of Computer Vision*, 7(2), 1992.
13. T. Kanungo, D.M. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A.Y. Wu. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:881–892, 2002.
14. J. Koenderink and A. van Doorn. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22(9):773–791, 1975.
15. H. Li and R. Hartley. Five-point motion estimation made easy. In *18th International Conference on Pattern Recognition (ICPR06)*, pages 630–633, 2006.
16. J. Lim and N. Barnes. Estimation of the epipole using optical flow at antipodal points. In *OMNIVIS*, 2007.
17. John Lim and Nick Barnes. Directions of egomotion from antipodal points. In *Proceedings of CVPR*, 2008.
18. H. Longuet-Higgins. A computer algorithm for reconstruction of a scene from two projections. *Nature*, 293:133–135, 1981.
19. R. Mahony, T. Hamel, and J.-M. Pflimlin. Complementary filter design on the special orthogonal group so(3). In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on*, pages 1477–1484, Dec. 2005.
20. Robert Mahony, Felix Schill, Peter Corke, and Yoong Siang Oh. A new framework for force feedback teleoperation of robotic vehicles based on optical flow. In *Proceedings IEEE International Conference on Robotics and Automation (ICRA)*, 2009.

21. A. Makadia, C. Geyer, S. Sastry, and K. Daniilidis. Radonbased structure from motion without correspondences. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*, June 2005.
22. D. Nister. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 756–770, 2004.
23. J.W. Roach and J. K. Aggarwal. Determining the movement of objects from a sequence of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(6):55–62, 1980.
24. Felix Schill, Robert Mahony, Peter Corke, and Luke Cole. Virtual force feedback teleoperation of the insectbot using optic flow. In *Proceedings of the Australasian Conference on Rotoics and Automation*, Canberra, Australia, December 2008.
25. C. Silva and J. Santos-Victor. Direct egomotion estimation. In *Procedings 13th International Conference on Pattern Recognition*, volume 1, pages 702–706, 1996.
26. B. Triggs, P.F. McLauchlan, R.I. Hartley, and A.W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *In Vision Algorithms 99*, pages 298–372, 1999.