UAV Motion Estimation using Hybrid Stereoscopic Vision

Damien Eynard[1] , Cedric Demonceaux[2] , Pascal Vasseur[3] , Vincent Fremont[4]
[1] MIS, Université de Picardie Jules Verne, Amiens.
[2] LITIS, Université de Rouen, Rouen.
[3] Le2I, Université de Bourgogne, Le Creusot.
[4] Heudiasyc, Université de Technologie de Compiègne, Compiègne.
[1]damien.eynard@u-picardie.fr, [2]pascal.vasseur@univ-rouen.fr,
[3]cedric.demonceaux@u-bourgogne.fr, [4]vincent.fremont@hds.utc.fr

Abstract

Motion and velocity are essential parameters for an Unmanned Aerial Vehicle (UAV) during critical maneuvers such as landing or take-off. In this paper, we present a hybrid stereoscopic rig made of a fisheye and a perspective cameras for motion estimation. The rotation and translation are estimated by a decoupling. The fisheye view contributes to determine the orientation and the attitude while the perspective view contributes to approximate the scale of the translation. Then, the calibrated stereo rig is used to estimate the altitude. While classical methods are generally based on feature matching between cameras, we propose in this paper an algorithm which tracks and exploits points in each view independently and filters the motion by Kalman filtering. Tracked points in each view are classed in two types: points located on the ground plane, which altitude is known and environment points which altitude is not known. Then, motion can be estimated robustly using the 2 points algorithm followed by a Kalman filter. We show that this approach is robust and accurate, and presents low sensitivity to noise by using the hybrid rig and Kalman filter.

1 Introduction

Unmanned Aerial Vehicles (UAVs) received a lot of attention during the last decade about command and onboard computer vision. Those topics aim to increase UAV autonomy that includes the capacity of performing maneuvers such as landing and take off, and autonomous flight. In this way, a fast, robust and accurate estimation of critical parameters such as altitude, attitude and velocities is required by the control loop.

Beside some sensors such as GPS or anemometers, vision is important and widely used for UAV. First, it can be used for other visual tasks. Next, cameras are compact, passive systems and therefore low energy consumers and can provide a great amount of information per second (up to 200Hz). Most methods for motion estimation use perspective cameras [15] [13] but this type of camera suffers of limited field of view, translation/rotation ambiguity and possibility of feature drop during the flight. Thus some works deals with hybrid systems. [2] propose a SfM method based on hybrid system made of fisheye and perspective cameras modeled by spheres, using an improved SIFT method to match features between cameras. Preprocessing of the improved SIFT requires more computation time than standard one which is not adapted for mixed views and real time navigation.

We proposed a hybrid vision system composed of a fisheye and perspective cameras. Omnidirectional systems possess a large field of view, are less sensitive to the blur of motion and avoid translation/rotation ambiguity while perspective camera keep the planar neighborhood and have a high and constant resolution. By combining such cameras, we exploit the advantages of each one.

To solve those failings, we propose to estimate navigation parameters (see fig.1) of an UAV using correspondence-free methods that satisfy the real time context:

- We presented in [5] a new and fast calibration method to calibrate n hybrid cameras.
- Attitude and orientation are estimated by [6, 7].
- We proposed in [8] to estimate the altitude and the ground plane segmentation by plane-sweeping knowing the homography between two views.
- Then we propose in this paper, by using informations issued from previous steps, to estimate the motion.

We assume that two types of points are tracked in each view separatly: environment points and ground points. Then, we propose to merge the two types of points. Finally, estimated motion is filtered by a Kalman filter to avoid perturbations due to noise and measurement errors.

Briefly, our approach presents several contributions. First, the system is able to estimate autonomously the motion without any other sensor and also provides attitude, altitude and the ground plane area. Next, we propose a correspondence-free approach which allows to treat images with different geometry (spherical and planar) and is particularly more robust than classical matching based stereoscopic approaches. Then, we demonstrate that merging planar points and 3D random environment points allows an accurate estimation of the translation. Finally, motion datas are smoothed by a Kalman filter, to get an estimation ready to be included in the control loop.

The organisation of the paper is as follows. Part II presents the modeling of our hybrid vision system. In part III, we propose to estimate and filter the motion of mixed views using and comparing methods based on environment points and planar points. Part IV is dedicated to experimental results on real sequences on a small UAV with a quantitative evaluation of the estimation of the translation.

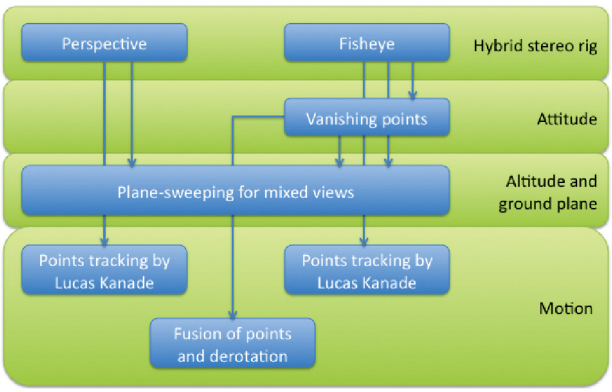

Figure 1. Estimation of UAV's parameters.

2 General Principle

2.1 Global Overview

We propose a mixed perspective/omnidirectional stereovision system able to estimate motion as well as altitude, attitude and the ground plane fig.1.

The benefit of large field of view cameras for UAVs has been already demonstrated in different works such as [9] for navigation or for attitude estimation [7]. The advantage of the omnidirectional sensor is the wide field of view while the drawbacks are the limited and non linear resolution of the image and the distortions. Advantages of fisheye in comparison with catadioptric cameras are their reduced sensitivity to vibrations and the suppression of the blind spot at the center of the image. On the other side, perspective cameras possess a good and constant resolution, low distortions but a limited field of view. The use of a mixed sensor allows to obtain the advantage of each sensor.

- First, by knowing the intrinsic parameters of the fisheye camera, a rectified equivalent perspective image could be recovered in order to perform for example, a feature matching. However, this approach requires different processing such as warping and interpolation which discard a real time performance.
- Some recent works propose the unitary sphere as unified space for central image processing and feature matching. However, as previously, this solution can not be implemented in real time and is not adapted to mixed views.
- Finally, we propose an approach which consists in tracking points in each view without stereo matching.

Since the motion is estimated from environment information we assume some points belong to the ground plane whose the segmentation and the altitude are known by method [8]. Having a pair of mixed cameras viewing the ground, where $\mathbf{R}_c$ and $\mathbf{t}_c$ estimated by calibration define the rigid transformation between two views, we will demonstrate that using both environment points mostly presented in the fisheye and ground points mostly presented in the perspective view let a better estimation of the motion at the metric scale (fig. 1).

2.2 Camera Models and Calibration

Althought fisheye lenses cannot be classified as single viewpoint sensor [1], we use the unitary sphere in order to model our camera [16]. Mei and Rives [14] have proposed a calibration method based on this spherical model. This model is particularly accurate and allows to model radial and tangential distortions of the fisheye lens. Mixed stereo calibration is obtained by [5].

3 Motion Estimation

Since altitude, attitude and ground plane segmentation are estimated by [8] [7], in this section we propose an algorithm based on tracked points in mixed views and Kalman filter to estimate and refine the translation $\mathbf{t}$ from two sets of 3D points: points located on the ground plane and point located randomly in the environment with unknown depth.

3.1 Motion of the stereo rig

In each image, tracking is performed by the method proposed in [4]. We define $\mathbf{x}_t$ a tracked point in the image acquired at the time t. Each point is related by a rotation $\mathbf{R}_t$ relatively to the world reference $\mathbf{X}_w$ estimated by the IMU (see eq.1).

$$\mathbf{x}_t = {}^t\mathbf{R}_w \mathbf{X}_w \Longleftrightarrow \mathbf{X}_w = {}^t\mathbf{R}_w^{-1} \mathbf{x}_t \tag{1}$$

Then, for a couple of points $(\mathbf{x}_t; \mathbf{x}_{t+1})$ that illustrate a motion during time t and $t+1$, we can express the point $\mathbf{x}_t$ from the frame t to the frame $t+1$ by eq.2. We get the rotation for a motion (eq.3).

$$\mathbf{x}_{t+1} = {}^{t+1}\mathbf{R}_w {}^t\mathbf{R}_w^{-1} \mathbf{x}_t \tag{2}$$

$$\mathbf{R}_{t+1} = {}^{t+1}\mathbf{R}_w {}^t\mathbf{R}_w^{-1} \tag{3}$$

Those points are related by a motion composed of the rotation $\mathbf{R}_{t+1}$ previously defined and the translation $\mathbf{t}_{t+1}$ (fig.2). The distance from the image point to the 3D point is defined by the altitude d (see fig.2) known by plane-sweeping in the case of a point belonging to the ground plane. We get eq.4.

$$\mathbf{x}_{t+1} = d\mathbf{R}_{t+1}\mathbf{x}_t + \mathbf{t}_{t+1} \tag{4}$$

3.2 Motion from planar points

As said previously, plane-sweeping estimates both altitude and the segmentation of the ground plane. By knowing those parameters, we know tracked points in the two views belonging to the ground plane with their depth (see fig.2). Then, the motion is estimated at the metric scale and defined as follow:

$$\mathbf{x}_{t+1} \times \mathbf{t}_{t+1} = -d(\mathbf{x}_{t+1} \times \mathbf{R}_{t+1}\mathbf{x}_t) \tag{5}$$

If the cross product of eq.5 is expressed as matrix representation we obtain the equation $\mathbf{A}\mathbf{t}_{t+1} = \mathbf{B}$ with $\mathbf{x}_t = (x_t, y_t, z_t)^T$, $\mathbf{A}$ is expressed in eq.6 and $\mathbf{B}$ expressed in eq.7:

$$\mathbf{A} = \begin{bmatrix} 0 & -x_{t+1} & x_{t+1} \\ z_{t+1} & 0 & -x_{t+1} \\ -y_{t+1} & x_{t+1} & 0 \end{bmatrix} \tag{6}$$

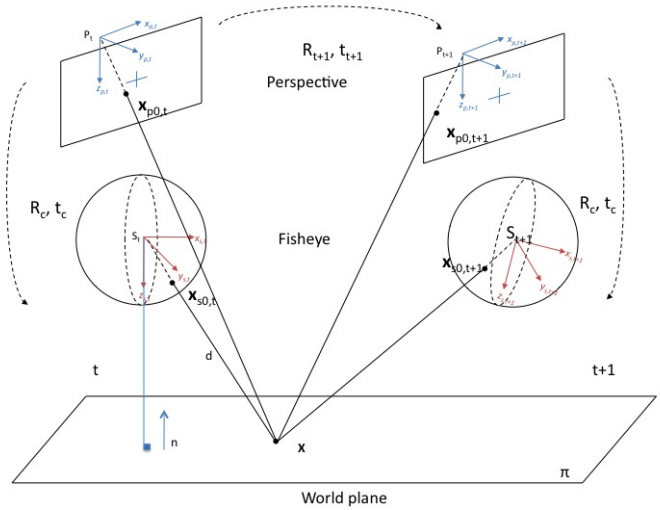

Figure 2. 3D points located on the ground plane and in the environment.

$$\mathbf{B} = -d(\mathbf{x}_{t+1} \times \mathbf{R}_{t+1}\mathbf{x}_t)^T \tag{7}$$

In case of estimation of translation with more than two points, we can easily obtain $\mathbf{t}_{t+1}$ by least square. Let us show how motion can be estimated in each view, then in mixed view. First, in perspective case, we get $\mathbf{A}_p$ and $\mathbf{B}_p$ matrix. Second, in the spherical case, we get $\mathbf{A}_s$ and $\mathbf{B}_s$ matrix. Then in mixed case, $\mathbf{A}_s$ and $\mathbf{A}_p$ are concatenated in $\mathbf{A}_m$ and $\mathbf{B}_s$ and $\mathbf{B}_p$ are concatenated in $\mathbf{B}_m$. For n perspective points and m spherical points, the size of $\mathbf{A}_p$ is $3n \times 3$, $\mathbf{A}_s$ is $3m \times 3$, $\mathbf{B}_p$ is $3n \times 1$, $\mathbf{B}_s$ is $3m \times 1$, $\mathbf{A}_m$ is $3(m+n) \times 3$ and $\mathbf{B}_m$ is $3(m+n) \times 1$.

3.3 Motion from environment points

In case of points located randomly in the environment, without any knowledge of their depth, we propose to extend the motion estimation proposed in spherical view by [3] to mixed views (see fig.2). The translation $\mathbf{t}_{t+1}$ is defined for two points as $(\mathbf{R}_{t+1}\mathbf{x}_t \times \mathbf{x}_{t+1})^T \cdot \mathbf{t}_{t+1} = 0$ and estimated up to scale. As said previously, perspective points and spherical points are concatenated to estimate the translation $\mathbf{t}_{t+1}$.

3.4 Fusion of 3D points from mixed views

From planar points, translation is estimated at the meter scale. However, the main drawback in case of 3D motion is the pixel projection noise sensitivity. From environment points, motion estimation has the advantage to be more robust to noise than on the plane but the estimation is performed up to scale. One of contribution of this paper is thus the combination of the two methods to increase both accuracy and robustness. The first method is defined by the eq.9 and the second by the eq.10 and $\mathbf{C}_m$ is defined by eq.8. By concatenating eq.9 and 10 we obtain eq.11 solved by least mean square.

$$\mathbf{C}_m = \begin{bmatrix} \mathbf{R}_{t/t+1}X_{s0,t} \times X_{s0,t+1} \\ \vdots \\ \mathbf{R}_{t/t+1}X_{sm,t} \times X_{sm,t+1} \\ \mathbf{R}_{t/t+1}X_{p0,t} \times X_{p0,t+1} \\ \vdots \\ \mathbf{R}_{t/t+1}X_{pn,t} \times X_{pn,t+1} \end{bmatrix} \tag{8}$$

$$\mathbf{A}_m T = \mathbf{B}_m \tag{9}$$

$$\mathbf{C}_m T = 0 \tag{10}$$

$$\begin{bmatrix} \mathbf{A}_m \\ \mathbf{C}_m \end{bmatrix} T = \begin{bmatrix} \mathbf{B}_m \\ 0 \end{bmatrix} \tag{11}$$

3.5 Kalman filtering

Once the motion is estimated at the millimeter scale, we observe some discontinuities and brutal variations. In order to reduce bad estimations and to refine the trajectory, we choose to use a linear Kalman filter [10]. The considered state is simply the translation vector of the ego-motion i.e. $\mathbf{x}_k = \mathbf{t}_{t+1}$, and is modeled as a linear Gaussian system given by the eq. 12:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{E}\mathbf{x}_k + \mathbf{v}_k \\ \mathbf{y}_k &= \mathbf{O}\mathbf{x}_k + \mathbf{w}_k \end{aligned} \tag{12}$$

where $\mathbf{E} = \mathbf{I}_3$ is the linear state transition model and $\mathbf{O} = \mathbf{I}_3$ is the observation model, assuming constant speed during a sample interval. The vectors $\mathbf{v}_k$ and $\mathbf{w}_k$ respectively correspond to the error model and the observation noise. They are supposed to be additive and white zero-mean Gaussian with used-defined covariance matrices. In order to keep the dynamic nature of the measurements, more uncertainty is given to the measures i.e. 10^6 and 10^2 for the model. From the above considerations, the Kalman filter consists in predicting the translation vector $\mathbf{t}_{t+1}$ and then obtaining a refined value using an update step when a new observation is available. We therefore obtain the translation vector of the ego-motion and its estimated accuracy, from all past observations up to the current time.

4 Results

We propose to estimate the motion by a linear method: the Least-mean squares (LS). This method is robust to Gaussian noise but sensitive to outliers [12]. Thus, outliers are rejected by the RANSAC method. Then, we tested our algorithm on a quadri-rotor with two uEye cameras with images processed offline. A Xsens IMU provides the attitude and the rotation of the motion while the plane-sweeping estimates the altitude. In each view, the number of tracked points is between 50 and 200. Then, we suppose that the perspective camera points to the ground plane which is known in the fisheye by the homography defined by plane-sweeping between two views [8]. The fig.3 presents the final 3D trajectory of the motion estimation while fig.4 presents the motion in each axis. It shows in red raw datas sensitive to noise and sudden changes while in blue datas are filtered and smoothed by Kalman.

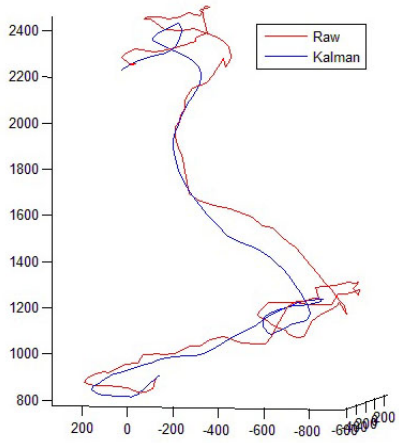

Figure 3. 3D trajectory of the UAV.

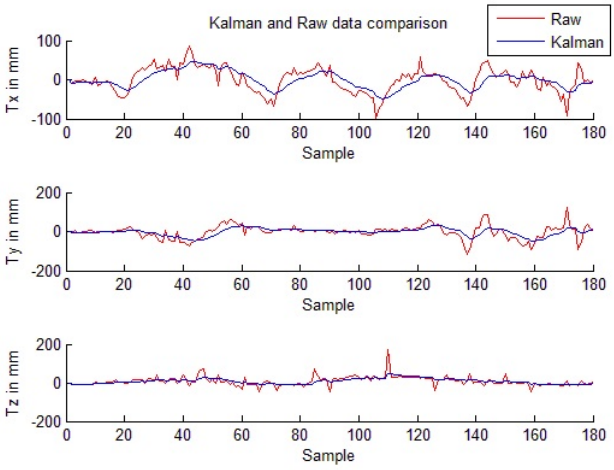

Figure 4. Raw datas vs. Kalman filtered datas.

The accuracy has been tested on a mobile robot with a translation along x-axis. The real distance is 2500mm and the estimated distance is 2352mm.

5 Conclusions and future works

We have presented in this paper a hybrid stereo system where points are tracked in each camera. Projections of the ground plane onto the mixed cameras are related by a homography which allows to estimate both altitude and ground plane using plane-sweeping. Once the altitude is estimated and the ground plane is segmented, tracked points on the ground provide the information of metric translation while environment points permit an accurate estimation of translation, up to scale. By combining points of mixed views, by merging two sets of points, translation is estimated accurately with the metric information and refined by Kalman filter.

Perspective of this work will be to get real time and embedded motion estimation.

6 Acknowledgments

This work is supported by the European FEDER and Région Picardie Project ALTO. Experimentations have been realized on UAV platform of Heudiasyc with cooperation of Luis-Rodolfo GARCIA-CARRILLO.

References

[1] S. Baker, and S. K. Nayar, "A Theory of Single-Viewpoint Catadioptric Image Formation," International Journal of Computer Vision,1999.

[2] Y. Bastanlar, A. Temizel, Y. Yardimci, "Effective Structure from Motion for Hybrid Camera Systems", International Conference on Pattern Recognition, 2010.

[3] J-C Bazin, I. Kweon, C. Demonceaux, P. Vasseur, "Motion Estimation by Decoupling Rotation and Translation in Catadioptric Vision", CVIU, 2009.

[4] J-Y. Bouguet,"Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm", In Intel Corporation Microprocessor Research Labs, 2000.

[5] G. Caron, D. Eynard. "Multiple Camera Types Simultaneous Stereo Calibration", In Proceedings of IEEE International Conference on Robotics and Automation, 2011.

[6] C. Demonceaux and P. Vasseur and C. Pégard, "Omnidirectional vision on UAV for attitude computation," In Proceedings of IEEE International Conference on Robotics and Automation, 2006.

[7] C. Demonceaux and P. Vasseur and C. Pégard, "UAV Attitude Computation by Omnidirectional Vision in Urban Environment," In Proceedings of IEEE International Conference on Robotics and Automation, 2007.

[8] D. Eynard, P. Vasseur, C. Demonceaux, V. Fremont, "UAV Altitude Estimation by Mixed Stereoscopic Vision", In Proceedings of IEEE International Conference on Robotics and Automation, 2010.

[9] S. Hrabar and G. Sukhatme, "Omnidirectional vision for an autonomous helicopter," In Proceedings of IEEE International Conference on Robotics and Automation, 2004.

[10] Rudolf Emil Kalman, "A new approach to linear filtering and prediction problems", Transactions of the ASME - Journal of Basic Engineering, 1960.

[11] Y. Ma, S. Soatto, J. Kosecka, S.S. Sastry, "An Invitation to 3D Vision", Springer, 2004.

[12] E. Malis, E. Marchand, "Méthodes robustes d'estimation pour la vision robotique", Journées nationales de la recherche en robotique, 2005.

[13] K. Kanatani, "Detection the motion of a planar surface by line & surface integrals", In Computer Vision, Graphics, and Image Processing, 1985.

[14] C. Mei and P. Rives, "Single View Point Omnidirectional Camera Calibration from Planar Grids," In Proceedings of IEEE International Conference on Robotics and Automation, 2007

[15] O. Shakernia, Y. Ma, T. J. Koo, S. Sastry, "Landing an Unmanned Air Vehicle: Vision based Motion Estimation and Nonlinear Control", Asian Journal of Control, 1999.

[16] X. Ying and Z. Hu, "Can We Consider Central Catadioptric Cameras and Fisheye Cameras within a Unified Imaging Model," In European Conference on Computer Vision, 2004.