2010 3rd International Congress on Image and Signal Processing (CISP2010)

Vision Aided Unmanned Aerial Vehicle Autonomy: An Overview

Yu-chi Liu [1,2]
1. Department of Automation; Tsinghua University;
Beijing, China, 100084;
2. Radar Academy;
Wuhan, Hubei Province, China, 430019

Qiong-hai Dai[1]
1. Department of Automation;
Tsinghua University;
Beijing, China, 100084

***Abstract*—Camera on-board Unmanned Aerial Vehicle (UAV) is not just a device to capture images and video. Computer vision helps UAV senses and recovers environmental information around it. In addition, computer vision techniques, such as stereo vision, optical flow fields etc. play very important roles in UAV autonomy. This paper investigates researches on computer vision applied in UAV autonomy. Visual servoing technique integrates visual methods and UAV flight control. Researches on this topic are classified into three catalogues: vision-based navigation, aerial surveillance and airborne visual Simultaneous Localization and Mapping (SLAM).**

Keywords-UAV autonomy; computer visison; visual servoing; visual navigation; aerial surveillance; airborne visual SLAM

I. INTRODUCTION

In recent years, Unmanned Aerial Vehicles (UAVs) have received great attention in many fields. Without consideration of carrying human, UAV can be designed into flexible ways, such as more powerful and agile in capability. Diverse UAVs appear from heavy to miniature in size. They take versatile duties in many fields, especially when human intervention is hard, dangerous, boring or expensive e.g. hazardous zone or forest-fire reconnaissance, traffic monitoring, disaster relief support, military operations etc [1][2].

Practically, UAV is controlled by a remote pilot or ground operator by means of satellite communications or data link. Typically, UAV relies on navigation instruments. Inertial Navigation System (INS) and Global Positioning System (GPS) are two popular instruments. INS acquires navigation data from accelerometers and gyroscopes. GPS receives navigation data from satellites. It is called GPS-aided inertial navigation. GPS became the main measure providing absolute position.

However, UAV may be out of outside operation and guidance. Satellite communication would be jammed during the war. UAV may fly beyond controlling range. GPS signal is not always reliable or sufficient, such as in low altitude, canyons, or urban areas, even indoor area for miniature UAV. What's more, Miniature UAV can't afford expensive and heavy traditional communication and navigation devices. UAVs are expected to hold enough autonomy to tackle situations mentioned above.

Visual sensors, such as electro-optic or infrared cameras, are cheap, portable, passive and abundant of information. Its natural connection to human driven tasks makes it to be the preferred option for world understanding. Computer vision can also provide UAVs with more information, which helps UAV to determine its pose and sense environmental conditions e.g. obstacles, terrain. With the aids of such information, UAVs are more capable of dealing with emergencies and to perform complex tasks, even on its own.

Recent researches have focused on UAV autonomy using computer vision. This paper give an overview of researches on vision aided UAV autonomy. Section II describes main computer vision techniques used in UAV visual servoing; section III introduces vision-based navigation, including autonomous landing, autonomous refueling and autonomous maneuvering; section IV introduces aerial surveillance with the aid of computer vision, including target tracking and co-operation of multi-UAVs; section V introduces airborne visual Simultaneous Localization and Mapping (SLAM); section VI gives the conclusion.

II. UAV VISUAL SERVOING

Computer vision techniques are very helpful for UAV autonomy, such as motion tracking, stereo vision etc. These techniques provide UAV with detailed pose and surrounding information. This paper doesn't explain such techniques. The interesting thing is how to link the information from vision sensor to UAV control. This section describes visual subsystem onboard and visual servoing method.

A. Visual Subsystem

Take one of platforms proposed by Computer Vision Group at UPM for example [2].

Generally, visual subsystem is compound of one or more cameras (e.g. binocular stereo vision), a servo controlled Pan Tilt platform, an onboard computer and a variety of cameras and visual sensors, including analog/digital cameras (Firewire, USB etc.). The system uses Gimbals to control camera's movement. Communication is based on a long-range wireless interface. Hardware and Software are designed to perform visual tasks involve optical flow, Hough transform, camera calibration, stereo vision to corner detection, visual servoing

National Basic Research Program of China (973 Program, No. 2010CB731800)

control implementation and Kalman filtering, among others. Scene information obtained from image processing and analysis provides data related to the camera's coordinate system. This information is useful for purposes of automatic camera control, but not for the attitude and position control of the UAV.

B. *Visual Servoing Method*

Visual servoing schemes enable to control robot motion by using visual features as feedback signal [3]. In many studies, camera is for providing pose estimation. Based on the position, other sensors such as GPS or inertial sensors can be used in the control law. The principal matter in vision control application is how to integrate information from vision sensors to UAV control routine.

There exist three methods of visual servoing: 3D, 2D and 2½D. 3D visual servoing techniques are for reconstruction of the target pose and called: position based visual servoing. It leads to a Cartesian motion planning problem and needs a perfect knowledge of the target geometric model. 2D visual servoing aims to control the dynamics of features in the image plane directly. Classical 2D methods suffer from high coupling dynamics between translation and rotational motion which makes the Cartesian trajectory uncontrollable. 2½D visual servoing consists of combining visual features obtained directly from the image, and features expressed in the Euclidean space. We take 2½D visual servoing to explain the basic method [21].

Let (X, Y, Z) denote three components of the translation, and (ϕ, θ, ψ) denote roll-pitch-yaw angles of the rotation. Then $P = (X, Y, Z, \phi, \theta, \psi)$ denote pose vector of UAV. Let $v_a = (v_{ax}, v_{ay}, v_{az}, \varpi_{ax}, \varpi_{ay}, \varpi_{az})$ denote UAV's instantaneous velocity, where $v_{a\cdot}$ and $\varpi_{a\cdot}$ respectively denote velocity component and angular velocity to axis. UAV dynamics model which links inputs and the aircraft motions have been provided by the French company Dassault Aviation, which can be approximated as following：

$$\dot{P} = v_{a\cdot} \tag{1}$$

Let S denote a visual feature. The relationship between S and the camera instantaneous velocity $v_c = (v_{cx}, v_{cy}, v_{cz}, \varpi_{cx}, \varpi_{cy}, \varpi_{cz})$ is defined by the interaction matrix L_s :

$$\dot{s} = L_s v_{c\cdot} \tag{2}$$

The camera is fixed on UAV. Respectively, ${}^cR_a$ and ${}^ct_a$ denote the rotation matrix and the translation vector between UAV frame and camera frame.

The link between the camera velocity screw (v_c) and the aircraft velocity screw (v_a):

$$v_c = {}^cS_a v_a \tag{3}$$

with:

$${}^cS_a = \begin{bmatrix} {}^cR_a & [{}^ct_a]\times {}^cR_a \\ 0_{3\times3} & {}^cR_a \end{bmatrix}.$$

Using the desired interaction matrix is a classical approximation in visual servoing when the considered displacements are small. Using this approximation, and recalling (1) 、(2)and (3) gives:

$$\dot{s} = L_s^{*}\,{}^cS_a \dot{P}\cdot \tag{4}$$

Now, we link the airplane dynamic model and the visual features model. Based on it, we can design a visual servoing control scheme, which is not described here.

III. VISION-BASED NAVIGATION

UAVs typically use GPS and inertial sensors combined to estimate its own state and form a navigation solution. However, in some circumstances, such as urban or low-altitude areas, GPS signal may be very weak or even lost. Under these situations, visual data can be used as an alternative or substitute to GPS measurements for the formulation of a navigation solution.

This section describes vision-based UAV navigation, including autonomous landing, autonomous refueling and autonomous maneuvering.

A. *Autonomous Landing*

A basic requirement is robust autonomous flight, for which autonomous landing is a crucial capability. It gives basic idea and method for UAV autonomy. According to flight characteristics, UAVs are classified into Vertical Take-Off and Landing (VTOL) UAVs and fixed-wing UAVs.

As to VTOL UAV vision-based landing, Sharp et al. [4] design a landing target with simple pattern, on which corner points can be easily detected and tracked. By tracking these corners to determine UAV's relative position to landing target using computer vision method.

Given the corner points, estimating the UAV state is an optimization problem. The equation relating a point in the landing pad coordinate frame to the image of that point in the camera frame is given by

$$\lambda_i x_i = APgq_i \tag{5}$$

where $\lambda_i \in \mathrm{IR}$ is an unknown scale, $x_i \in \mathrm{IR}^3$ is the homogeneous coordinates of the feature point in the image plane, $A \in \mathrm{IR}^{3\times3}$ is the camera calibration matrix, $P = [I \ \ 0] \in \mathrm{IR}^{3\times4}$ is the projection matrix, $g \in SE(3)$ is the homogeneous representation of the Euclidean motion between the landing pad coordinate frame and the camera-head frame, and $q_i \in \mathrm{IR}^4$ is the homogeneous representation of the point in the world. Given enough inputs from computer vision, it can be solved as a linear or nonlinear optimization.

During UAV state estimation, transformations between frames should be discussed. The geometry of the coordinate frames and Euclidean motions is involved in the vision-based state estimation problem, as shown in Figure 1: (a) Landing target, (b) Landing pad, (c) Camera head, (d) Camera base, (e) UAV, (f) Inertial frame. We use the notation $g_{ba} \in$ SE(3) to denote the Euclidean motion (translation and rotation) of coordinate frame b with respect to frame a. It describes how each coordinate frame transformation is measured. The transformations g_{ef} and g_{fb} are used only to evaluate the state estimates of the vision algorithm and are not used by the vision algorithm directly.

According to experimental test, the position estimates are within 5cm accuracy.

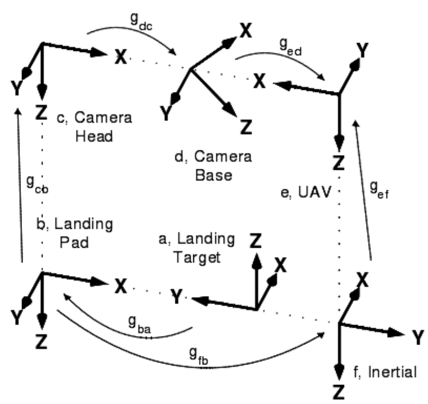

Figure 1. Geometry of the coordinate frames and Euclidean motions involved in the vision-based state estimation problem [4].

Saripalli et al. [5] took a further test of VTOL UAV landing on a moving target and got good result.

Fixed-wing UAV's auto landing is similar to that of VTOL in theory but more complicated in practice. Vision subsystem should recognize runway and keeps tracking on it during landing. Kalman filter is introduced to keep stability of tracking [10]. Horizon also needs detecting. According to runway and horizon, vision subsystem could estimate UAV's states [11], including location (*x,y,z*), attitude (*pitch, roll, yaw*) etc. Optical flow field method is also used to take state estimate [12].

B. Autonomous Refueling

The deployment of UAVs has been tested in overseas conflicts. People find that one of the biggest limitations of UAVs is their limited range. To enlarge their range, UAVs are expected capable of Autonomous Aerial Refueling (AAR). Currently, there are two ways for aerial refueling, i.e. refueling boom and "probe and drogue".

Very similar to auto landing, vision-based method for UAV keeping pose and position to flying tanker during docking and refueling receives great attention. AAR also needs considering some reference frames, such as UAV, tanker, camera frame etc. But AAR is much more sensitive and facing more subtle air disturbance. It requires 0.5 to 1.0 cm accuracy in the relative position.

A fixed number of visible optical markers are assumed to be available to help vision subsystem detect. However, temporary loss of visibility may occur because of hardware failures and/or physical interference. Fravolini et al. [13] proposed a specific docking control scheme featuring a fusion of GPS and MV distance measurements to tackle this problem.

Such studies are still under stage of simulation for its complexities [14].

C. Autonomous Maneuvering

Auto maneuvering is the extension of auto landing. Ideally, it means that UAV is capable of high level environment understanding and decision making e.g. defining its position, attitude estimation, obstacle detecting and avoidance, path planning etc. without outside instruction, guidance and intervention. Some studies related to UAV auto maneuvering considered different conditions, including GPS signal failure, unstructured or unknown flying zone etc.

Madison et al. [6] discussed miniature UAV's visual navigation in GPS-challenged environment, e.g. indoor. Vision subsystem geo-locates some landmarks while GPS provides accurate navigation. Once GPS is unavailable, vision subsystem geo-locates new landmarks with predefined landmarks. Using these landmarks, vision subsystem provides information for navigation. Tests show that vision-aided navigation drift is significantly lower than under inertial-only navigation.

NASA Ames Research Center runs a Precision Autonomous Landing Adaptive Control Experiment (PALACE) whose primary goal is to develop enabling technologies for autonomous landing of rotorcraft in non-cooperative environments. Non-cooperative sites are those areas that have no pre-positioned objects, beacons, or landing area markings that the vehicle may use to guide its descent. These sites may also contain unknown terrain features and obstacles that must be identified and avoided during the landing task. Non-cooperative sites of interest include building rooftops in an urban setting or rugged terrain in remote locations. GPS failure is also considered. Under support of PALACE, Hintze [7] used stereo vision to generate a surface plot of the landing area. Elevation Map, Hazard Map, Cost Map, which is the combined values of the angle and roughness map are generated based on Stereo Range Mapping. According to the 3D surface plot, UAV searches for a safe landing region. Then, monocular tracking provides position data to the navigation controller during descent.

Meingast [8] used vision-based system to recover the geometry and material properties of local terrain from a mounted stereo rig for the purposes of finding an optimal landing site. Dense elevation is estimated using a multi-frame planar parallax estimation developed by Irani et al. [9].

IV. AERIAL SURVEILLANCE

Equipped with small video cameras and transmitters, UAVs are quite suitable for surveillance tasks in areas either too remote or too dangerous to send human scouts. The main function is visual tracking. A more interesting matter is co-operate multi-UAVs for aerial surveillance.

A. Visual Tracking

Most of aerial surveillance is based on image processing algorithms and tracking techniques. Visual tracking includes feature, appearance and target tracking. Typical features are corner features or interest points. Appearance tracking uses a patch of pixels that corresponds to the object that wants to be tracked. Target tracking is based on object detection from motion analysis. One popular tracking method is Lukas–Kanade algorithm. RANSAC method for SIFT features are also effective.

Tracking method finds relation between video frames by a warping function that can be the optical flow or another model of motion. Table 1 summarizes some of the warping functions used and the degrees of freedom [2].

TABLE I. WARPING FUNCTIONS SUMMARY [2]

Name	Rule	D.O.F
Optical flow	$(x, y) + (t_x, t_y)$	2
Scale + translation	$(1 + s)((x, y) + (t_x, t_y))$	3
Scale + rotation + translation	$(1 + s)(R_{2x2}(x, y)^T + (t_x, t_y)^T)$	4
Affine	$\begin{pmatrix} 1+\mu_1 & \mu_3 & \mu_5 \\ \mu_2 & 1+\mu_4 & \mu_6 \end{pmatrix}$	6
Perspective	$\begin{pmatrix} \mu_1 & \mu_2 & \mu_3 \\ \mu_4 & \mu_5 & \mu_6 \\ \mu_7 & \mu_8 & 1 \end{pmatrix}$	8

Ludington [15] introduced a system demonstrates a ground target tracking architecture for mobile targets. The system has three components. The first is a particle filter, which estimates the position of the target within the incoming video frames. The second is a camera controller, which moves the field of view (FOV) to keep the target within the video frames. The third is a trajectory (waypoint) generator, which commands the GTMax to smoothly follow the target.

The visual tracking task is a state estimation problem. Its goal is to estimate the probability density of the position or state of the target within each incoming video frame using information from all the past frames. The optimal solution (in a mean-squared sense) is given by the Bayesian tracking filter. When the distribution is Gaussian and the target's dynamics are linear, the Bayesian solution becomes the well-known Kalman filter. However, if the linear, Gaussian assumption is violated, the distribution must be approximated by some other means.

Hegazy [16] provided a particle filtering framework based on Bayesian network for aerial surveillance and tracking over an urban environment to feed the military and police needs. The main functions are of scene analysis, target detection, position and following, and aerial mapping.

B. Co-operation of Multi-UAVs

Ryan et al. [17] thought the capabilities of small UAVs continue to grow with advances in wireless communications and computing power. Accordingly, research topics in cooperative UAV control include efficient computer vision for real-time navigation and networked computing and communication strategies for distributed control, as well as traditional aircraft-related topics such as collision avoidance and formation flight.

Part of the growing interest in UAVs stems from the increasing feasibility of these goals due to advances in small and efficient processors, cameras, and wireless networking.

Issues in cooperative UAV control can be concluded as follows [22]:

A. Aerial Surveillance and Tracking

B. Collision and Obstacle Avoidance

C. Formation Reconfiguration

D. High Level Control

E. Hardware and Communication

V. AIRBORNE VISUAL SLAM

Simultaneous localizations and map (SLAM) building is in essence the true autonomous navigation. The goal is to generate a map of the environment without any a prior knowledge, and to localize the vehicle within that map simultaneously, constraining drift rates of both the vehicle and the map. Without an initial map or an absolute localization means, it requires to concurrently solving the localization and mapping problems. For this purpose, vision is a powerful sensor, because it provides data from which stable features can be extracted and matched as the robot moves.

Compared with indoor, land and underwater situations, airborne visual SLAM however poses a more difficult problem. Being an airborne visual SLAM platform, UAV's high dynamic motion poses several challenges: re-observation of landmarks is difficult; feature extraction algorithms have to deal with large deviations in image brightness; highly non-linear process and observation models are generated, which corrupt the linearizing assumptions in the EKF; it requires high computational power. SLAM on an aerial vehicle means that the vehicle motion must be estimated in 6-DoF and the position of features in 3D. This adds to the size of the state vector from most ground vehicle applications where the vehicle motion and feature map is estimated in 2D.

The mathematical framework of the SLAM algorithm is based on an estimation process which, when given a kinematic/dynamic model of the vehicle and relative observations between the vehicle and landmarks, estimates the structure of the map and the vehicle's position and orientation within that map.

The problem is formulated using state variables to describe and model the system. The state of the system is described by the vector:

$$X = [x, s_1, s_2, s_3, ...] \tag{6}$$

where x denotes the state of the camera and si represents the state of each feature. Camera state has 12 variables. The First 6 variables represent the position of the vehicle in iteration k and in the previous iteration. The Next 6 variables, vector $[p, q, r]$, represent the rotation at iteration k and k-1. Rotation is expressed using Rodrigues' notation, which expresses a rotation around a vector with the direction of $\omega = [p, q, r]$ of an angle $\theta = \sqrt{p^2 + q^2 + r^2}$.

Generally, SLAM can be classified into monocular and stereo approaches. Bryson et al. [18] used monocular approach to take airborne visual SLAM without helps from other measurements. Alternatively, Nemra et al. [19] implied stereo approach.

Bryson et al. [20] also discussed problems of visual SLAM with co-operation of multi-UAVs. Two ways of co-operation are compared. One is an ideal, centralized architecture where the planning process occurs at a central planning node which designates control actions to each vehicle. The other is decentralized and coordinated, where each vehicle plans independently, based on the information that is shared between vehicles in the SLAM data fusion.

VI. Conclusions

UAV's autonomy depends on its ability on awareness of surrounding and decision making on its own. With Computer vision techniques, UAV can get rich outside information to determine its attitude and position, and to understand environment. Based on these, UAV can make right flight decision on its own. Furthermore, UAV can detect and track targets, and even cooperate with other UAVs for more demanding tasks. In the future, there is no man on UAV, there is no human's intervention on UAV, but there is a *vision-based autopilot* on UAV to take complex and hard missions.

References

[1] Office of the Secretary of Defense, USA, "Unmanned Aerial Vehicle (UAV) Roadmap 2005-2030," August 2005.

[2] P. Campoy, J. F. Correa, I. Mondragón, C. Martínez, M. Olivares, and L. Mejías et al., "Computer vision onboard UAVs for civilian tasks," Journal of Intelligent and Robotic Systems, 2009, vol. 54(1-3), pp. 105-135.

[3] S. Hutchinson, G. Hager, and P. I. Corke, "A tutorial on visual servo control," ITRA'96, 1996, vol.12(5), pp. 651-670.

[4] Sharp Courtney, O. Shakernia, and S. Shankar, "A vision system for landing an unmanned aerial vehicle," Proceedings of the IEEE Conference on Robotics & Automation, 2001. Seoul, Korea. IEEE Press, pp. 1720-1727.

[5] Srikanth Saripalli and Gaurav S. Sukhatme, "Landing on a moving target using an autonomous helicopter," Proceedings of the International Conference on Field and Service Robotics, 2003. Mt. Fuji. Japan. Springer Press, pp. 277-286.

[6] R.W. Madison, G.L. Andrews, P.A. DeBitetto, S.A. Rasmussen, and M. S. Bottkol, "Vision-aided navigation for small UAVs in GPS-challenged environments," Proceedings of AIAA Infotech at Aerospace Conference and Exhibit, 2007. Rohnert Park, CA, USA. AIAA Press, pp. 318-325.

[7] J. Hintze, "Autonomous landing of a rotary unmanned aerial vehicle in a non-cooperative environment using machine vision," Masters thesis, Brigham Young University, UK, 2004.

[8] M. Meingast, C. Geyer, and S. Sastry, "Vision based terrain recovery for landing unmanned aerial vehicles," Proceedings of 43rd IEEE Conference on Decision and Control, 2004. Atlantis, Paradise Island, Bahamas. IEEE Press, vol.2, pp. 1670-1675.

[9] M. Irani, P. Anandan, and D. Weinshall, "From reference frames to references planes: Multi-view parallax geometry and applications," Proceedings of European Conference on Computer Vision. 1998. Freiburg, Germany. Springer-Verlag, pp. 846-861.

[10] G. Salama, "Monocular and Binocular Visual Tracking," Ph.D. thesis, Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, 1999.

[11] Scott M. Ettinger, Michael C. Nechyba, and Peter G. Ifju, "Vision-Guided flight stability and control for micro air vehicles," Proceedings of IEEE International Conference on Intelligent Robots and Systems, 2002. Lausanne, Switzerland. IEEE Press, pp. 2134-2140.

[12] D. Dusha, W. Boles, and R. Walker, "Fixed-wing attitude estimation using computer vision based horizon detection," Proceedings of 12th Australian International Aerospace Congress, 2007. Melbourne Australia. WIT Press. pp. 1-19.

[13] M.L. Fravolini, A. Ficola, G. Campa, M.R. Napolitano, and B. Seanor, "Modeling and control issues for autonomous aerial refueling for UAVs using a probe-drogue refueling system," Aerospace Science and Technology, 2004, vol. 8(7), pp. 611-618.

[14] G. Campa, M. R. Napolitano, and M. L. Fravolini, "Simulation environment for machine vision based aerial refueling for UAVs," IEEE Transactions on Aerospace and Electronic Systems, 2009, vol. 45(1), pp. 138-151.

[15] B. Ludington, E. Johnson, and G. Vachtsevanos, "Augmenting UAV autonomy: vision-based navigation and target tracking for unmanned aerial vehicles," IEEE Robotics and Automation Magazine, 2006, vol. 13(3), pp. 63-71.

[16] T. Hegazy, B. Ludington, and G. Vachtsevanos, "Reconnaissance and surveillance in urban terrain with unmanned aerial vehicles," Proceedings of the 16th IFAC World Congress, 2005. Prague, Czech Republic. Elsevier, pp. 872-879.

[17] A. Ryan, M. Zennaro, A. Howell, R. Sengupta, and J. K. Hedrick, "An overview of emerging results in cooperative UAV control," Proceedings of IEEE Conference on Decision and Control, lantis, Paradise Island, Bahamas: IEEE Press, pp. 602-607.

[18] Mitch Bryson and S. Sukkarieh, "Building a robust implementation of bearing-only inertial SLAM for a UAV," Journal of Field Robotics, Special issue on SLAM in the field, 2007, vol. 24(2), pp. 113-143.

[19] K. Nemra and N.Aouf, "Robust airborne 3D visual simultaneous localization and mapping with observability and consistency analysis," Journal of Intelligent and Robotic Systems, 2009, vol. 55(4), pp. 345-376.

[20] Mitch Bryson and S. Sukkarieh, "Architectures for cooperative airborne simultaneous localisation and mapping," Journal of Intelligent and Robotic Systems, 2009, vol. 55(4), pp. 267-297.

[21] N. Metni and T. Hamel, "A UAV for bridge inspection: visual servoing control law with orientation limits," Automation in Construction, 2006, vol. 17(1), pp. 3-10.

[22] M. Bryson and S. Sukkarieh, "Co-operative localisation and mapping for multiple UAVs in unknown environments," Proceedings of IEEE/AIAA Aerospace Conference," 2007. Piscataway, USA. IEEE Press, pp. 229-241.