Indoor Exploration Using a µUAV and a Spherical Geometry Based Visual System

Tiago Caldeira[1], Lakmal Seneviratne[1,2], and Jorge Dias[1,3]

[1] Khalifa University for Science Technology and Research, UAE
[2] Center for Robotics Research from Kings College of London, UK
[3] Institute of Systems and Robotics, University of Coimbra, Portugal
tiago.caldeira@kustar.ac.ae

Abstract. This research presents a new vision system that explores a spherical geometry and will be provide innovative solutions for tracking, surveillance, navigation and mapping with micro Unmanned Aerial Vehicle (µUAV) in unknown indoor environments. The system will be used with µUAV in indoor environment and it is composed by twenty six cameras that are arranged in order to sample different parts of the visual sphere around the µUAV. This configuration allows that some of the cameras will have overlapped field of view. This system has been designed for the purpose of recovering ego-motion and structure from multiple video images, having a distributed omnidirectional field of view. We use the spherical geometry to extend the field of view, from one single direction to a single point of perspective, but with multiple views. This manuscript will prove that spherical geometric configuration has advantages when compared to stereo cameras for the estimation of the system's own motion and consequently the estimation of shape models from each camera. The preliminary field tests presented the theoretical potential of this system and the experimental results with the images acquired by 3 cameras.

Keywords: Vision, µUAV, Spherical Geometry, Indoor, Ego-motion, Trifocal tensor, Motion Flow.

1 Introduction

Mobile robots are important artifacts for the exploration of unknown areas, not only in hazard situations, but also as an extension of human capabilities. Those so called explorer robots can be developed to navigate in unleveled terrain. Both semi-autonomous and completely autonomous machines allowed a real time report of otherwise impossible places to visit. On the last decades researchers expanded their attention to others than Unmanned Ground Vehicles (UGV), for example, µUAV.

The most challenging part of 3D mapping is obtaining the full 6 Degrees of Freedom (DOF) of pose of the robot between each scan [1]. The vehicle position and orientation, is, in most cases, estimated with the combination of inertial sensors and GPS, but in indoor GPS denied environments tends to create an important barrier from the fully autonomous exploration [2]. To complete the autonomous exploration, problems like

L.M. Camarinha-Matos, S. Tomic, and P. Graça (Eds.): DoCEIS 2013, IFIP AICT 394, pp. 200–209, 2013.

localization and mapping unknown environments need to be solved. One way of achieving this is with the use of laser sensors or vision systems. Advances have been made in the use of vision sensors for target tracking and obstacle avoidance or to estimate vehicle pose [3], but the traditional stereo configuration does not provide enough information nor covers the totality of the space. In order to complete the visual information, inertial sensors were added to the system. The inertial sensed gravity [4][5] provides a vertical reference for the spherical vision system, enabling to establish an artificial horizon line and vertical features. Between each pair of cameras there is a rigid transformation that could be determined using a calibration process [6], getting the rectification of the stereo configuration and generate the epipolar constrains, and generating the optical flow [7].

In this paper we intend to demonstrate some of the advantages of the spherical geometry for multiple cameras complemented with inertial data in order to be successfully solving problems of tracking, surveillance, navigation and mapping.

In the next section it will be presented the relation with the Internet of Things (Section 2), followed by the research in UAV field and vision based controls (Section 3). The proposed solution and mathematical support (Section 4) will be complemented with some of the implementation details and the progress towards the full sensor developing (Section 5) that lead us to the conclusions and future work (Section 6).

2 Relationship to Internet of Things

When explaining the definition of "Internet of Things", Kevin Ashton [8] mention that "people have limited time, attention and accuracy – all of which means they are not very good at capturing data about things in the real world". This extension to the man capabilities was always present during the development of this project. In the early stages of development of this project, we intended to create a unique capturing device that not only capture video from all of its surroundings but also could be able to identify and track people or objects. By analyzing and categorizing moving objects, it will provide enough information for an organized database which could grow without human interaction. With multiple systems like this and if them are complemented with network connection, multiple nodes will act as smart and autonomous data collection spots, presenting content to the Internet of Things.

3 Related Work

In the last decade, many authors increased their research on the implementation of control algorithms of flying devices, mainly μUAV. Those devices have been used in multiple research goals, but with the goal of autonomous navigation or mapping an unknown environment. The first approaches on the UAV control were only focused on outdoor environments, where there was more space of maneuver and where GPS

signals were an important data to close the loop regarding the position and velocity of the UAV.

Samir Bouabdallah and Roland Siegwart [9] proposed an approach for full control of an UAV based on a quadrotor configuration. To achieve autonomy, the UAV needs a navigation system, and many authors proposed several solutions for indoor navigation systems [10][11][12]. Different methods have been used to not only navigate but also for map the environment. Active sensors like lasers or ultrasonic sensors where combined with visual elements [3][13]. The use of visual information for obstacles avoidance and simple navigation [14] was also object of research.

Grzonka et al. [15], already provided a solution to a fully autonomous indoor quadrotor based on LIDAR technology. Bachrach et al. [16] using a similar architecture to explore and map unstructured and unknown indoor environments. The unknown and unstructured environments were also the base for another research by Blöesch, Scaramuzza et al. [17], but using a visual approach to this problem.

4 Research Contribution and Innovation

In this paper we present our progress towards a new vision system that combines multiple cameras in a spherical geometry merged with inertial sensors.

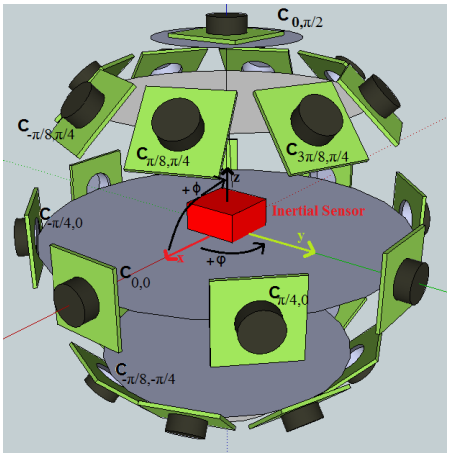

Fig. 1. Three-dimensional representation of the mechanical design with the cameras information and inertial sensor position

This system is composed by twenty six symmetrically distributed cameras with the same intrinsic parameters. As presented in Fig. 1, there are groups of 8 cameras that are equally distributed along the equatorial line and along the tropic lines. This is complemented with one camera in each pole. With this configuration, the angle

between any two consecutive cameras is always $\pi/4$ radians. The coordinates of each camera is always $C_{\varphi,\phi}$ where φ is the angle on z-axis and ϕ the angle on x-axis.

In the center of the sphere is a 6 DOF inertial sensor that will provide the system accelerations and orientations.

4.1 Geometry Relationships and Models

Using a spherical geometry allow a unique transformation between each pair of cameras. Even without a complete overlap between all cameras, it is possible to see all of the extension of the exterior space.

The center of projection of the camera $C_{\varphi,\phi}$ could be obtained with a rotation on x-axis $\mathbf{R}_x(\phi)$ and z-axis $\mathbf{R}_z(\varphi)$ from the frontal camera $C_{0,0}$ using (1).

$$^{C_{0,0}}\mathbf{R}_{C_{\varphi,\phi}} = \mathbf{R}_z(\varphi)\ \mathbf{R}_x(\phi) \tag{1}$$

where $\varphi = \frac{m\pi}{4} + \frac{n\pi}{8}$ and $\phi = \frac{n\pi}{4}$, $(\forall m \in M = \{-3,\ldots,0,\ldots,4\} \wedge \forall n \in N = \{-2,-1,0,1,2\})$

That way, the relation between a camera i ($C_{\varphi,\phi}$) with any other camera j ($C_{\varphi',\phi'}$) could be determined with (2).

$$^{C_{\varphi',\phi'}}R_{C_{\varphi,\phi}} = \mathbf{R}_z(\varphi' - \varphi)\,\mathbf{R}_x(\phi' - \phi) \tag{2}$$

By using developments on [18] and assuming that the intrinsic parameters among these two cameras are equal, the homography $^{j}\mathbf{H}_i$ between the images could be described in (3) where $\mathbf{K}$ represents the intrinsic parameters (the same in both cameras).

$$^{j}\mathbf{H}_i = \mathbf{K}\ {}^{j}\mathbf{R}_i\mathbf{K}^{-1} \tag{3}$$

That way, it is possible to calculate the epipolar constrain between the cameras (4).

$$e^{j} = \mathbf{P}'\begin{pmatrix}\vec{0}\\1\end{pmatrix} = \mathbf{K}\ {}^{j}t_i \tag{4}$$

The fundamental matrix, F, is obtained in (5).

$$^{j}\mathbf{F}_i = [e^{j}] \times \mathbf{K}\ {}^{j}\mathbf{R}_i\ \ \mathbf{K}^{-1} = [e^{j}] \times \mathbf{K}\,\mathbf{R}_z(\varphi' - \varphi)\,\mathbf{R}_x(\phi' - \phi)\ \ \mathbf{K}^{-1} \tag{5}$$

The unique setup of the geometry between the cameras allows the simplicity of the calculation not only between two, but also with each three cameras. The trifocal tensor [19] could be approached using the relationship of three corresponding lines. The lines $\mathbf{l_i}$, $\mathbf{l_j}$ and $\mathbf{l_k}$ are the projection of L (line in space) on each of the cameras i ($C_{\varphi,\phi}$), j ($C_{\varphi',\phi'}$) and k ($C_{\varphi'',\phi''}$) as represented in Fig. 2.

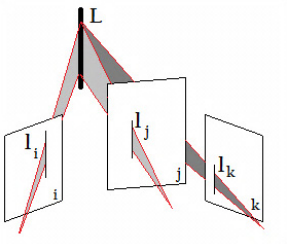

Fig. 2. Symbolic representation of three cameras representing the line L

Assuming the origin at camera I and using development of Hartley et al. [19], the cameras matrices for the three views would be $\mathbf{P}_i=[\mathbf{I}|\vec{0}]$, $\mathbf{P}_j=[\ {}^{j}\mathbf{H}_i|e^j]$, $\mathbf{P}_k=[\ {}^{k}\mathbf{H}_i|e^k]$, where ${}^{j}\mathbf{H}_i$ and ${}^{k}\mathbf{H}_i$ are the homography matrix between cameras j and i, and k and i respectively (calculated in (3)), while $e^j$and e^k are the epipolar constrain from camera j and k to i, as calculated in (4). Recovering the equation $l_a=\mathbf{l_j}^T\left(\ {}^{j}\mathbf{H}_{i_a}e^{kT}-e^j\ {}^{k}\mathbf{H}_{i_a}{}^T\right)\mathbf{l_k}$ where l_a represents the a coordinate of $\mathbf{l_i}$. Moreover, by definition, $l_a=\mathbf{l_j}^T\ \mathbf{T}_a\ \mathbf{l_k}$, what is equivalent to present the notation in (6).

$$\mathbf{T}_a=\left(\ {}^{j}\mathbf{H}_{i_a}\ e^{kT}-\ e^j\ {}^{k}\mathbf{H}_{i_a}{}^T\right) \tag{6}$$

Using the equations (2) and (3) that refers about the particular cameras configurations, the equation (6) could be rewrite as (7).

$$\mathbf{T}_a=[\mathbf{K}\,\mathbf{R}_z(\varphi'-\varphi)\,\mathbf{R}_x(\phi'-\phi)\mathbf{K}^{-1}]_a\ e^{kT}-\ e^j\,[\mathbf{K}\,\mathbf{R}_z(\varphi''-\varphi)\,\mathbf{R}_x(\phi''-\phi)\mathbf{K}^{-1}\,]_a{}^T \tag{7}$$

We could define the trifocal tensor $\mathbf{T}_i^{jk}$ between the cameras i, j and k, as being $\mathbf{T}_i^{jk}=[\mathbf{T}_x,\mathbf{T}_y,\mathbf{T}_z]$. This result could be used to extract the fundamental matrix between the cameras i and j in (8).

$${}^{j}\mathbf{F}_i=[e^j]\times[\mathbf{T}_x,\mathbf{T}_y,\mathbf{T}_z]e^k \tag{8}$$

This equation (8) shows that it is possible to recover the relationship between each subset of three cameras using the Trifocal tensor and the epipolar lines. Taking advantage of the unique relation between each cameras (pure rotations on x-axis and z-axis) the calculation will be simplified.

4.2 Motion Flow Properties in Image Plan

One very important part of our research is related with the movement of the entire system motion flow. In order to simplify the calculations we analyze the system using the camera $C_{0,0}$ as reference.

Assuming a three dimension motion has two components: translation and rotation.

If we define the projection $\vec{\boldsymbol{p}}=f\,\mathbf{P}/Z$, with p the image point, $\boldsymbol{P}=[X\ Y\ Z]^T$ the scene point with a depth Z, we could define the equation (9), where **V** is the world motion, **T** the translation and **ω** a vector that represents the rotation.

$$\mathbf{V}=-\mathbf{T}-\boldsymbol{\omega}\times\mathbf{P} \tag{9}$$

Assuming that between two consecutives frames the motion is small, we could derive the both sides of the projection equation. Using the previous equation (9) and from the image perspective analyzing its velocity projection components, we obtain the equations in (10).

$$\boldsymbol{v}=f\frac{Z\mathbf{V}-V_Z\mathbf{P}}{Z^2}=\begin{cases}v_x=\frac{T_Zx-T_xf}{Z}-\omega_yf+\omega_zy+\frac{\omega_xxy}{f}-\frac{\omega_yx^2}{f}\\ v_y=\frac{T_Zy-T_yf}{Z}+\omega_xf-\omega_zx-\frac{\omega_yxy}{f}+\frac{\omega_xy^2}{f}\end{cases}\Leftrightarrow$$

$$\Leftrightarrow\boldsymbol{v}=\frac{1}{Z}\begin{bmatrix}-f & 0 & x\\ 0 & -f & y\end{bmatrix}\mathbf{T}+\begin{bmatrix}\frac{xy}{f} & -f-\frac{x^2}{f} & y\\ f+\frac{y^2}{f} & -\frac{xy}{f} & -x\end{bmatrix}\boldsymbol{\omega} \tag{10}$$

From this equation (10) we see that is difficult to decouple the motion from rotations around x-axis or y-axis of the camera. Therefore, we use inertial sensors to confirm those. The relation between inertial sensor and the camera axis is represented on Fig. 3.

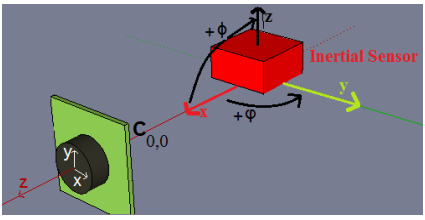

Fig. 3. Inertial and camera $C_{0,0}$ referential

4.3 Visual and Inertial Integration

Inertial information provided by accelerometer and gyroscopes are by themselves very important to obtain the attitude and motion parameters of the vision system.

However, the accelerometer provides the linear acceleration in the 3 axes of $\boldsymbol{a}$, sensing the gravity vector $\boldsymbol{g}$ summed with the visual system acceleration $\boldsymbol{a}_{SYSTEM}$. If the vision system is attached to the UAV ($\boldsymbol{a}_{SYSTEM}\equiv\boldsymbol{a}_{UAV}$), then the acceleration could be defined in (11).

$$\boldsymbol{a}=-\boldsymbol{g}+\boldsymbol{a}_{UAV} \tag{11}$$

This will mean that, if the system is motionless ($\boldsymbol{a}_{UAV} = 0$), only the gravity will be present, allowing the system to have a vertical reference from the gravity, and calculate the orientation n of the UAV using (12).

$$\hat{n} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = -\frac{\boldsymbol{g}}{\|\boldsymbol{g}\|} = \frac{1}{\sqrt{{a_x}^2 + {a_y}^2 + {a_z}^2}} \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} \tag{12}$$

The gyroscope measures the rotational acceleration in the 3 axes, and it is possible to separate the gravity from the sensed acceleration by performing the rotation update [4].

As the inertial sensor is fixed in the center of the sphere as soon as we determine the orientation $\hat{n}$ of the device, we could calculate the orientation $\hat{c}$ of every camera $\mathrm{C}_{\varphi,\phi}$, based on the rotation between the inertial system and the $\mathrm{C}_{0,0}$ in equation (13).

$$\hat{c} = \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} = {}^{IMU}\boldsymbol{R}_{\mathrm{C}_{0,0}} \quad {}^{\mathrm{C}_{0,0}}\boldsymbol{R}_{\mathrm{C}_{\varphi,\phi}} \quad \hat{n} \tag{13}$$

The ${}^{IMU}\boldsymbol{R}_{\mathrm{C}_{0,0}}$ could be easily determined using calibration [6]. Having the complete pose information of each camera, it is possible to predict location of the horizon line, from where it is also possible to easily identify vertical features on each image.

5 Implementation and Experimental Setup

In order to present the initial proof of concept, only one oct part of the sphere where designed. Using the configuration displayed in Fig. 4 it was possible to test different angles configurations.

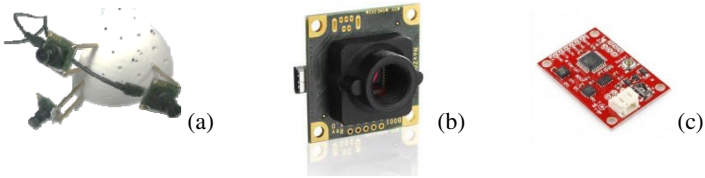

Fig. 4. Hardware used on the trials, (a) 3D printed version of part the Sphere(b) UI 1226LE-M camera from IDS-Imaging (c) 6DOF Inertial Measurement Unit

Each camera is connected through a serial connection link (USB) and could display at a 43 frames per second (fps) rate, value that decreases when multiple cameras are connected on the same bus. This frame rate is enough to use vision based tracking algorithms.

5.1 Experiments

The initial experiments proved the system concept and the correspondence between points in cameras were the expected. Images from the three cameras (see Fig. 5) were

captured and displayed simultaneous at a 25 fps rate. The inertial data captured was also validated in order to confirm the orientation of the system. The square chessboard was also important in order to calibrate the trifocal tensor.

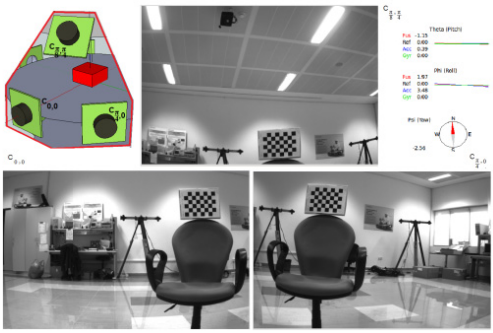

Fig. 5. Cameras Position, Attitude Control and respective Images - $C_{\frac{\pi}{8},\frac{\pi}{4}}$ (top) $C_{0,0}$ (left) and $C_{\frac{\pi}{4},0}$ (right)

After calibration (between each camera and with the inertial system) a set of images were collected from the three cameras (key frames selected and then grouped on Fig. 6) where two people move in opposite directions. With this set of images is possible to recover the motion on space of each person.

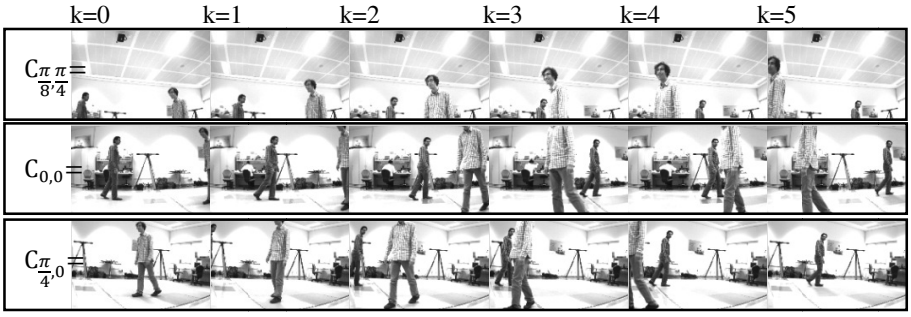

Fig. 6. Sequence with 6 frames from $C_{\frac{\pi}{8},\frac{\pi}{4}}$ (top) $C_{0,0}$ (middle) and $C_{\frac{\pi}{4},0}$ (bottom)

6 Conclusions and Future Work

We introduced a visual system based on multiple cameras with a spherical geometry with a complete field of view. This system simplifies the mathematical equations for correspondence and movement perception. The use of an orthogonal matrix (rotation matrix) in the homography and epipolar calculation will reduce the calculation time.

The first trials provided enough data to test basic vision algorithms and proved the capability and potential future of the system.

Other algorithms could be developed and proved based on this system, eventually aiming an autonomous μUAV to explore, navigate and map indoor environments.

References

1. Morris, W., Dryanovski, I., Xiao, J.: 3D Indoor Mapping for Micro-UAVs Using Hybrid Range Finders and Multi-Volume Occupancy Grids. In: Proc. of Robotics: Science and Systems, RSS (2010)
2. Ahrens, S., Levine, D., Andrews, G., How, J.P.: Vision-based guidance and control of a hovering vehicle in unknown, GPS-denied Environments. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2643–2648 (2009)
3. Achtelik, M., Bachrach, A., He, R., Prentice, S., Roy, N.: Stereo Vision and Laser Odometry for Autonomous Helicopters in GPS-Denied Indoor Environments. In: Proceedings of the SPIE Conference on Unmanned Systems Technology XI, Orlando, FL (2009)
4. Lobo, J.: Integration of Vision and Inertial Sensing. PhD Thesis, U. of Coimbra (2006)
5. Corke, P., Lobo, J., Dias, J.: An introduction to inertial and visual sensing. International Journal of Robotics Research, Special Issue 2nd Workshop on Integration of Vision and Inertial Sensors 26(6), 519–535 (2007)
6. Lobo, J., Dias, J.: Relative Pose Calibration Between Visual and Inertial Sensors. International Journal of Robotics Research, Special Issue 2nd Workshop on Integration of Vision and Inertial Sensors 26(6), 561–575 (2007)
7. Dias, J., Araújo, H., Paredes, C., Batista, J.: Optical Normal Flow Estimation on Log-polar Images. A solution for Real-Time Binocular Vision. Real-Time Imaging Journal 3, 213–228 (1997)
8. Ashton, K.: That 'Internet of Things' Thing. RFID Journal (July 22, 2009)
9. Bouabdallah, S., Siegwart, R.: Full Control of a Quadrotor. In: Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS (2007)
10. Michael Sobers Jr., D., Chowdhary, G., Johnson, E.N.: Indoor Navigation for Unmanned Aerial Vehicles. In: AIAA Guidance, Navigation, and Control Conference (2009)
11. Achtelika, M., Bachrach, A., He, R., Prentice, S., Roy, N.: Autonomous navigation and exploration of a quadrotor helicopter in GPS-denied indoor environments. In: Robotics: Science and Systems (2008)
12. Chowdhary, G., Michael Sobers Jr., D.: Integrated Guidance Navigation and Control for a Fully Autonomous Indoor UAS, Portland, Oregon (2011)
13. Morris, W., Dryanovski, I., Xiao, J.: 3D Indoor Mapping for Micro-UAVs Using Hybrid Range Finders and Multi-Volume Occupancy Grids. In: Proceedings of Robotics: Science and Systems, RSS (2010)
14. Ahrens, S., Levine, D., Andrews, G., How, J.P.: Vision-based guidance and control of a hovering vehicle in unknown, GPS-denied environments. In: Proceedings IEEE International Conference on Robotics and Automation, ICRA 2009, pp. 2643–2648 (2009)
15. Grzonka, S., Grisetti, G., Burgard, W.: A fully autonomous indoor quadrotor. IEEE Transactions on Robotics (99), 1–11 (2012)
16. Bachrach, A., He, R., Roy, N.: Autonomous Flight in Unknown Indoor Environments. Intl. Journal of Micro Air Vehicles, 217–228 (2009)

17. Blöesch, M., Weiss, S., Scaramuzza, D., Siegwart, R.: Vision Based MAV Navigation in Unknown and Unstructured Environments. In: IEEE International Conference on Robotics and Automation (2010)
18. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press (March 2004)
19. Hartley, R.I.: Lines and points in three views and the trifocal tensor. International Journal of Computer Vision 22(2), 125–140 (1996)