

Contents

1	Introduction	1
1.1	A brief history of drones	1
1.2	Motivation	2
1.2.1	Ethical considerations	2
1.3	Context	2
1.4	Objectives	3
1.5	Structure	3
2	State of the art	4
2.1	Hardware	4
2.1.1	Multicopter systems	4
2.1.2	Sensors	5
2.1.3	Embedded computers	5
2.2	Computer Vision	5
2.3	Simultaneous Localization And Mapping	5
2.3.1	Localization	5
2.3.2	Mapping	5
2.4	Bundle Adjustment	5
3	Hardware and software architecture	6
3.1	Hardware	6
3.1.1	Sensors	6
3.2	Software	6
3.2.1	Parrot SDK	6
3.2.2	ROS	6
4	Localization	7
5	Mapping	8
5.1	Evaluation procedure	8
5.1.1	Experimental setup	9
5.2	Triangulation	9
5.2.1	Midpoint Method	10
5.2.2	Direct Linear Transformation	10
5.2.3	Optimal Correction	10
5.2.4	Comparison of triangulation methods	10
5.3	Bundle Adjustment	10
5.3.1	Convergence of the solver	11
5.4	Tuning the Bundle Adjustment	11
5.5	Map Initialization	11
6	Map Initialization	13

7 Simultaneous Localization and Mapping	14
7.0.1 Loop Closure	14
8 Conclusion	15
A Appendix Title	17

Acronyms

UAV Unmanned Aerial Vehicle.

Chapter 1

Introduction

1.1 A brief history of drones

UAVs (Unmanned Aerial Vehicles), more commonly called drones, are defined as flying vehicles without human operators on board. They can be remote-controlled, or controlled by on-board computers. The earliest recorded use of UAVs dates back to 1849, when Austria launched about 200 unmanned balloons armed with bombs against the city of Venice [6]. Due to unfavorable wind conditions, this attack failed, and the experiment was not repeated. The first functional UAVs were made towards the end of World War 1 and their use was, like the Austrian balloons, military. One example is the Kettering Bug (Figure 1.1), which was a torpedo with wings and a propeller developed by the US Army in 1918 [7].

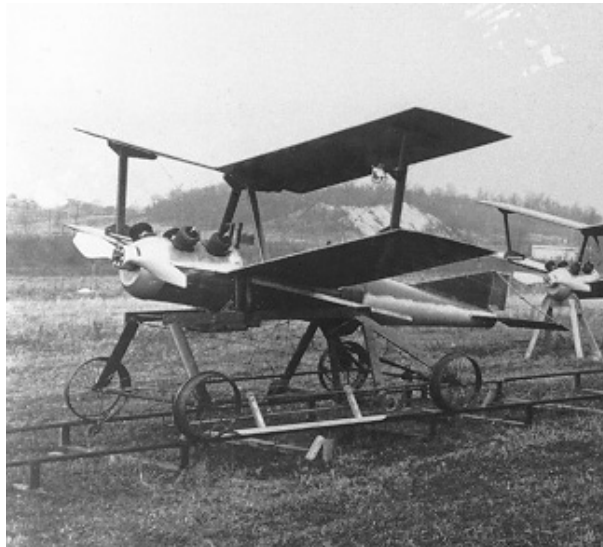


Figure 1.1: The Kettering Bug (1918)

Throughout the 20th century, UAVs became more and more sophisticated, and were used more and more, but always for military purposes. In the more recent years, civilian UAVs have started to appear on the markets and their number quickly exceeded that of military UAVs. In February 2017, the FAA (Federal Aviation Administration) of the United States estimated that around 1.1 million units were in use in the US alone, and expected that number to rise to 3.55 million by 2021 [4]. These civilian drones are very different from military drones, in both their form and their function: civilian drones are usually smaller, and use rotors to take off vertically. They are used in a wide variety of applications.

1.2 Motivation

The ability to remote-control small and agile flying objects over large distances through the air, and to bring them to previously inaccessible locations, makes many new things possible. With the increasingly lower prices and better performances of civilian UAVs, people keep finding more and more uses for these high-tech gadgets. Some examples of these applications are: crop monitoring in agriculture [1], delivery of mail or parcels, construction [2], cinematography, entertainment, or search and rescue operations. In all these applications, the more autonomous a drone is, the more efficient it will be at its task. One of the main challenges to achieve autonomy is for an UAV to be able to correctly identify its surroundings, and localize itself within them. In outdoor environments, GPS systems allow UAVs to know their position with great accuracy, but this is not possible in GPS-denied environments, such as indoors. The main subject of this thesis will be fully autonomous navigation by a quadcopter in a GPS-denied environment.

1.2.1 Ethical considerations

The new possibilities brought by drones also pose ethical questions about security and privacy. Even though this technology can improve people's quality of life, it also has the potential to diminish it. If drones start to be widely used commercially, we could reach a point where the sound nuisances that they cause seriously impacts people who live in densely populated areas. Also, they can make us feel less at home, knowing that we could be observed from the sky. For this reason, it is important to adopt strict regulations regarding the use of drones in public spaces. Fortunately, many countries are already adopting legislation in this direction.

1.3 Context

This thesis is part of a project at the UCL that spans over several years and several masters theses. This project was launched by professor Julien Hendrickx in the 2012-2013 academic year, and had as long-term goal to develop a program that would enable low-cost UAVs to navigate autonomously in indoor environments. This means creating a map of their environment, localizing themselves in this map, and avoiding obstacles during exploration, using only on-board sensors. Another goal is to allow several drones to collaborate to speed up exploration. Five theses have already been written on this subject, each taking the work of the previous a little further.

2012-2015: First three theses In each of the three academic years (2012-2013, 2013-2014, 2014-2015), one masters thesis on the subject of indoor navigation for autonomous low-cost drones was written. These masters theses formed the base of the future work. They implemented visual SLAM methods to allow drones to build a two-dimensional map based on keypoints (first red pucks, then visual landmarks that the drone detected from a textured field of view), and to localize itself within this map. During this time, inter-drone communication was also established, and was used to allow a drone to communicate the location of a target to another drone.

2015-2016: Recent work Last year, two groups of students simultaneously wrote theses on this subject. Before doing so, they joined forces to reimplement what had been done previously, but using the ROS interface, an interface to work with robots that would make many things simpler, and allow more flexibility (see section The work of the first group of students allowed a drone to search and follow a mobile target, and call a second drone to continue this task when its battery was low.

The second group of students extended the SLAM algorithm to allow to use a 3D map to localize the drone. Unfortunately, they did not implement triangulation to allow to project seen points

into 3D space, but rather made the assumption that all points were located on the ground when building the map. The end result was a drone capable of using a 3D map to localize itself, but not capable of building one from its observations.

1.4 Objectives

For my own thesis, my goal is to continue the work of last year's second group, to allow true 3-D SLAM: to build a 3D map based on observations by the monocular camera. To achieve this goal I will follow the following steps:

- Research the current state of the art for 3D Keyframe based monocular visual SLAM
- Implement a way to triangulate points based on observations
- Bundle Adjustment
- Dense reconstruction
- Obstacle Avoidance

1.5 Structure

Chapter 2

State of the art

2.1 Hardware

When talking about autonomous drone navigation, it is important to be aware of what the current hardware is capable of doing, and how we can expect it to evolve in the near future. We will talk about the three main aspects of this hardware: the multirotor systems themselves, with the different possible configurations, the sensors, and finally the embedded computers.

2.1.1 Multirotor systems

Multirotors, or multicopters, are defined as rotorcrafts with three or more rotors. Having more rotors enables them to maneuver in 3D space with fixed-pitch rotors, unlike helicopters, which have articulations at the bases of their rotors. The most common multirotors have 3, 4, 6, or 8 rotors, and are respectively called tricopters, quadcopters (or quadrotors), hexacopters, and octocopters. Having more rotors has the advantage of giving more agility, at the cost of more energy consumption, and therefore a shorter battery life. A free solid object in 3D space, such as a multicopter, has 6 degrees of freedom: 3 for translation and 3 for rotation. To be able to directly control each of these 6 degrees of freedom, it must be possible to give 6 independent controls to the drone. This means that tricopters and quadcopters are always under-actuated: they can't directly control all 6 degrees of freedom. For example, quadrotors whose rotors are all in the same plane (as is almost always the case), can only directly control their translational movement along the axis parallel to the rotation of their rotors, and their roll, pitch and yaw angles, so to control their position in the plane perpendicular to the direction of gravity, they have to first adapt their roll and pitch, so that the resulting force of gravity and the thrust of their motors points inside that plane. Most hexacopters also work this way, as their rotors are also often in the same plane. Some hexrotors, however, have tilted rotors, and are fully actuated [8].

Octocopters on the other hand, are always over-actuated. One example, the Omnicopter, developed at ETH Zurich [3] can perform a 360° rotation along any axis, and move in a straight line in any direction, which enables it to perform complex and precise manoeuvres. It is able to catch thrown ping-pong balls with a little net.

2.1.2 Sensors

2.1.3 Embedded computers

2.2 Computer Vision

2.3 Simultaneous Localization And Mapping

Simultaneous Localization and Mapping (SLAM) refers to the joint task of creating a map of a robot's surroundings, while also keeping track of the robot's location in this map. The word "robot" should be understood very broadly in this context, for example it could be a simple handheld camera. Because there are countless different types of robots that do SLAM, SLAM is also a very diverse field, with different algorithms for different kinds of sensors. Here, we will focus on monocular visual SLAM, which is SLAM where the main sensor is a monocular camera.

2.3.1 Localization

2.3.2 Mapping

2.4 Bundle Adjustment

Chapter 3

Hardware and software architecture

3.1 Hardware

For this project I used Parrot's AR.Drone 2.0. This quadrotor was commercialized in 2012 and is an updated version of the original AR.Drone that was launched in 2010. This drone was marketed as a high tech toy, and is designed to be controlled from a smartphone application (connected to the drone via Wi-Fi). A few augmented reality games are available for the AR.Drone, in which it can recognise some predefined tags using computer vision, and interact with abject or other drones with a tag. To encourage the creation of more games for their drones, Parrot has released an open SDK that allows to effectively reprogram the drones. This early release of an open SDK has made it quite popular in the scientific community to do research on autonomous flight. The drone consists of 4 rotors, each with their own electric motor and microcontroller, an internal computer with a 1GHz ARM Cortex A8 processor and 1GB DDR2 RAM at 200MHz, and various sensors.

3.1.1 Sensors

The AR.drone has the following sensors:

- 3 axis accelerometer with ± 50 mg accuracy
- 3 axis gyroscope with $\pm 2000^\circ/s$ accuracy
- Pressure sensor with ± 10 Pa accuracy
- 3 axis magnetometer with $\pm 6^\circ$ accuracy
- Ultrasound sensor (facing downwards)
- Frontal camera (HD 720p 30fps)
- Ventral camera (QVGA, 60 fps)

3.2 Software

3.2.1 Parrot SDK

The SDK released by Parrot allows to send commands and receive information from the drone. However, it does not allow access to the lowest-level parts of the drone. It is possible to send the drone commands to take off, land, emergency stop, hover, move in a certain direction, but not to directly control the command sent to the motors. Similarly,

3.2.2 ROS

Chapter 4

Localization

Chapter 5

Mapping

The task of mapping was the main challenge of this work. This task consists in placing recognizable features in a map, that can later be used as landmarks by the drone to estimate its own position. The main challenge in the 3D case, is that a point needs to be observed from at least 2 different positions to be mapped. The simplest approach to map a point, is to simply triangulate its position from two different views. We will begin by exploring this approach. We will find that although this does work reasonably well when we are certain of the position of the cameras, it does not when this position is uncertain. In addition, this method does not allow to take more than two views into account. To remedy these problems we will implement a bundle adjustment step, that allows to build a map that is globally consistent. Throughout this section, we will have to make design choices to try to obtain a method that is both fast enough to work in real time, and accurate enough for the drone to control its position. To evaluate these performances, we will perform a standardized test.

5.1 Evaluation procedure

The evaluation will happen in two phases: first the drone will initialize its map and then it will be moved to various known locations, and we will measure the accuracy of its position estimation. The setup used for this evaluation is illustrated on figure 5.1. During the initialization phase, we will try to emulate the way the drone would initialize its map just after taking off during a real flight mission (see more in section). First the drone is placed in a known position on a table (position A in figure 5.1). There, the drone is turned on, and it begins initializing its map. The drone is then successively placed in 3 other known locations (B, then C, then D), and at each of these locations, the drone is manually communicated its position. Every time this happens, the drone adds landmarks into the map, and optionally updates existing landmarks' position, or even removes some landmarks. In reality, the drone would not know its exact position when taking views at points B, C, and D (position A is defined as the origin), especially at the beginning, as it would have to rely on its IMU and other internal sensors to estimate its position. To emulate this, we implement a second type of test: the robustness test, where the position that is communicated to the drone is slightly different from its real position, at each of the uncertain points (B, C, and D).

In the second phase of the evaluation, the drone is again placed at different known locations. This time, no information is communicated to the drone from the outside, and the drone does not modify its map. The drone estimates its position using only its camera and the map that it built during the initialization phase. We compare the drone's estimated position with its real position to evaluate the quality of the map. The two quantities we will seek to optimize are the accuracy of the drone's position estimation, and the time taken to compute the map.

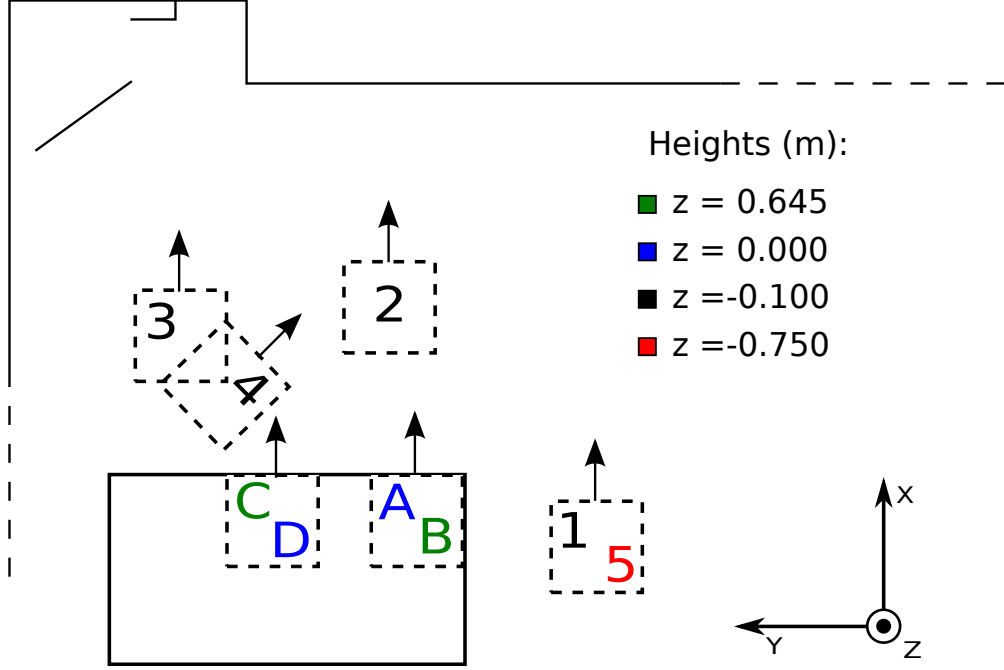


Figure 5.1: Different positions of the drone during the validation

5.1.1 Experimental setup

We will conduct two different finds of experiments: one to measure the accuracy, and one to measure the robustness of the mapping method. In both cases, we begin by placing the drone at the corner of a desk (position A on figure 5.1). After the drone has saved its view, it is placed on a stool exactly above its first position. Its exact pose is then manually communicated to the drone to simulate the measurements from its IMU and ultrasonic sensor. The drone then again takes a snapshot of its camera input, and matches points with the first view, and then adjusts its second pose and the position of the points through bundle adjustment. The drone and stool are then moved to the left (position B on figure 5.1), and again manually given its exact position, after which it takes another snapshot, matches points with the first two views, and readjusts the poses of the keyframes and the positions of the points through bundle adjustment. This entire process is repeated a third time when the drone is placed on the desk again at position D. The initialization of the map consists in making these four keyframes and adjusting their position and that of the landmarks 3 times. After this initialization is done, we place the drone at different locations, and measure how close it is to where it thinks it is. We evaluate the initialization based on how accurate its position estimation is, as well as on how much time the successive bundle adjustments took.

5.2 Triangulation

In the problem of triangulation, we try to find the 3D coordinates of a point from the 2D coordinates of the projection of this point on two images that were seen from different positions. We assume that the position of the camera taking these images is known exactly at both locations. If the camera positions is known exactly, and if the projection of the points into the image planes was perfect, then the two rays going from the camera centers, and through the images of the points would intersect at the location of the 3D point. In practice however, these two rays never cross exactly, so a method has to be found to find the best possible location of a 3D point from the pair of images.

5.2.1 Midpoint Method

The simplest solution would be to take the midpoint of the common perpendicular of the two rays. This method is intuitive to understand geometrically, and is quite easy to compute. In practice, however its results are not very good, as it does not correspond to an optimal value. The optimal solution would be to displace the pixels on both images until the resulting rays meet, keeping the displacement of the pixels as small as possible (in the least squared sense). Such a solution would give the maximum likelihood estimator of the position of the 3D points, under the assumption that the error of their projection on the image planes follows gaussian noise.

5.2.2 Direct Linear Transformation

5.2.3 Optimal Correction

There are several algorithms in the literature that triangulate the position of a point using optimal correction. The most popular one, proposed by Hartley and Sturm [5], computes the solution directly but requires finding the root of a 6th degree polynomial. Kanantani et. al.'s method [kanatani] finds a solution iteratively, but requires very few iterations to have an accurate solution, and in practice, is faster than the Hartley Sturm method. It also has better numerical properties, as unlike the Hartley-Sturm method, it does not have singularities at the epipoles.

5.2.4 Comparison of triangulation methods

Using the evaluation procedure described in section 5.1, we can compare these 3 triangulation methods.

	MDE (m)	MRE (rad)	TCT (s)	npts	avgtime
Midpoint Method	0.114	0.049	0.012	341	$3.52 \cdot 10^{-5}$
Direct Linear Transformation	0.360	0.127	0.013	341	$3.81 \cdot 10^{-5}$
Optimal Correction	0.097	0.039	0.024	341	$7.04 \cdot 10^{-5}$

5.3 Bundle Adjustment

Unfortunately the main source of error when mapping points is not inaccuracy of the camera, but uncertainty on the camera's position. This is bad, as a bad estimation of the camera's position will result in badly located landmarks, which in turn will result in a bad estimation of the camera position. In the long term, errors will accumulate, and the map will be completely distorted. Luckily, if we have enough point correspondences between two images, it is possible to deduce the relative displacement between the two images. This means that from a set of images, we can reconstruct a scene, without even needing an estimation of the position of the cameras that took the images. This is good news as it means that the images can give us some absolute information about the scene, not only relative to the drone. The problem of adjusting camera positions and 3D point locations in order to minimize the reprojection errors of the 3D points onto the image planes is known as bundle adjustment. As stated above, bundle adjustment has the advantage of being absolute with respect to the world, and so not having errors accumulate. Another advantage of bundle adjustment, is that it can easily take into account points that are seen by more than two cameras, which is not trivial for the triangulation techniques described above. The main disadvantage of bundle adjustment is that it is computationally heavy, so it is important to adapt it to be useable in real time.

To show the advantages of bundle adjustment, we begin by comparing it to simple triangulation,

in the case where the cameras' poses are known exactly, and in the cases where they are known approximatively, but subject to some noise.

	prec MDE (m)	prec MRE (rad)	rob MDE (m)	rob MRE (rad)
With bundle Adjustment	0.0842	0.0296	0.0907	0.0305
Without bundle Adjustment	0.1006	0.0354	0.0942	0.0366

5.3.1 Convergence of the solver

The first element we can tune is the convergence criterion of the solver that solves the bundle adjustment problem. We will stop either when the ration of the change in the objective function to the value of this function arrives below some threshold. This ensures that the criterion scales with the problem, which is important as the size of the problem can vary during operation (as the map grows, for example). The default value of the Ceres solver is 10^{-6} , but experimentally, we find that use a less severe threshold, to significantly speed up the computations, without impactig the quality of the results so much.

Convergence Threshold	MDE (m)	MRE (rad)	Time of Bundle Adjustment (s)
10	0.0975	0.0378	0.536
5	0.0967	0.0397	0.530
1	0.0989	0.0406	0.521
0.5	0.0970	0.0393	0.520
0.1	0.1065	0.0431	0.646
0.05	0.1385	0.0462	0.761
0.04	0.1380	0.0454	0.754
0.03	0.1575	0.0531	1.130
0.02	0.0793	0.0255	1.354
0.01	0.0722	0.0278	1.425
0.005	0.0765	0.0257	1.544
0.001	0.0879	0.0271	3.595
0.0001	0.0978	0.0330	5.207
0.00001	0.0916	0.0317	12.620
0.000001	0.0789	0.0307	13.232
0.0000001	0.0816	0.0229	15.453

5.4 Tuning the Bundle Adjustment

The main drawback of Bundle Adjustment is that it requires an iterative method to be solved and can take a lot of time, which of course is a limiting factor for a robot that builds a map in real time. Therefore it is important to optimize both the speed of the computations, and their precision. As is often the case, there will have to be a tradeoff between these two. To measure both the speed of the Bundle Adjustment and the accuracy of the map obtained, we will conduct some experiments using different parameters to initialize the map.

5.5 Map Initialization

A robot needs a map lo localize itself within it, but it needs to know its position to find the postion of surrounding objects and build a map. Because the mapping and localization tasks

are mutually dependent on one another, there needs to be a special procedure to build a map from nothing when it does not exist yet. Arbitrarily, we decide that the position of the drone when it starts flying is the origin (in all 6 degrees of freedom) of the map. However, with only a monocular camera, it is not possible to find the exact location of any visual features from one observation only, views from at least two different positions are needed to triangulate points.

To reach the position from which the drone will take a second view and triangulate points, the drone has to fly blindly. Blindly here means without using a map to localize itself visually, but the drone can still use its other sensors (IMU and ultrasonic sensor) to obtain an estimate of its position. Because the ultrasonic sensor is much more accurate than the IMU, and gives an absolute measure, we will mostly rely on this sensor to estimate the relative position from where we take the second view. Because the ultrasonic sensor only gives the distance from the bottom of the drone to the ground, the drone should fly straight up from its first position (the origin) to reach its second position.

Once in its second position, the drone can match seen keypoints from both views, and from its estimated position, triangulate those points to the map. However, the drone's estimation of its pose is prone to errors, especially as it only used its ultrasonic sensor and IMU. These errors can be corrected with the information from the cameras, because if we have matching observations, we can compute the fundamental matrix, and know exactly the displacement between the two views. Refining the poses of the views and the location of the landmarks simultaneously such as to minimize the reprojection error is a nonlinear optimization problem known as Bundle Adjustment. Using Bundle adjustment, we can refine the position of the second view, and use the ultrasonic sensor information to fix the scale. Another advantage of Bundle Adjustment is that it allows to take into account more than 2 views of a point.

Chapter 6

Map Initialization

Chapter 7

Simultaneous Localization and Mapping

7.0.1 Loop Closure

Chapter 8

Conclusion

Bibliography

- [1] Chris Anderson. “Agricultural Drones”. In: *MIT Technology Review* vol. 117 | no. 3 (2014).
- [2] “Bâtir avec des drones ?” In: <https://uclouvain.be/fr/sciencetoday/actualites/batir-avec-des-drones.html> (2016).
- [3] Dario Brescianini and Raffaello D’Andrea. “Design, Modeling and Control of an Omni-Directional Aerial Vehicle”. In: *IEEE International Conference on Robotics and Automation* (2016).
- [4] Marcus Chavers. “Consumer Drones By the Numbers in 2017 and Beyond”. In: *newsledge.com* (2017).
- [5] Richard I. Hartley and Peter Sturm. “Triangulation”. In: *CVIU - Computer Vision and Image Understanding* Vol. 68, No. 2 (1997).
- [6] Russell Naughton. *Remote Piloted Aerial Vehicles : An Anthology*. Tech. rep. Monash University, 2003.
- [7] Jimmy Stamp. “Unmanned Drones Have Been Around Since World War I”. In: *smithsonian.com* (2013).
- [8] Richard Voyles. *Dexterous Hexrotor Research*. <http://web.ics.purdue.edu/rvoyles/research.Hexrotor.html>.

Appendix A

Appendix Title