PRINCIPLES OF
VISION-ENABLED AUTONOMOUS FLIGHT

Jack Sanders-Reed

SPIE PRESS
Bellingham, Washington USA

Library of Congress Control Number: 2021944295

Published by
SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: books@spie.org
Web: www.spie.org

Cover images courtesy of Getty and Shutterstock.

Printed in the United States of America.
First Printing.
For updates to this book, visit http://spie.org and type "PM335" in the search field.

Table of Contents

Preface

This book is intended for anyone working in fields related to sensing for autonomous flight. This includes:

- Managers who may need to understand how and why vision systems are required for most autonomous flight scenarios.
- System engineers who must understand the strengths, limitations, and requirements for vision systems.
- Sensor system engineers who must select sensor suites, architectures, and algorithms.
- Regulators who must understand what can and cannot be expected from sensing systems.
- Students and faculty in aeronautical engineering and related fields (e.g., electrical engineering, physics, optics).

While this book can be read front to back, Chapters 1–3 deal with motivations and requirements, and are recommended for all readers. Chapter 4 (Sensors) and 5 (Architectures) are aimed at both managers and system engineers, while Chapter 6 (Algorithms) and the Appendix are aimed at engineers tasked with selecting and implementing various algorithms.

Although metric units are preferred in most applications, the aviation community commonly expresses altitude in feet, and speed in either miles per hour (mi/h) or Mach number. In this book we will prefer metric units but will employ the units that seem most likely to be intuitively understood by members of the community.

Jack Sanders-Reed
August 2021

Chapter 1

Introduction

"Don't worry, we've never had anyone go up and not come back down" –Flight instructor to new student pilot

This book addresses the following three basic questions related to vision-enabled autonomous flight (VEAF):

1. Why are vision systems fundamental and critical to VEAF?
2. What are the vision system tasks required for autonomous flight?
3. How can those tasks be approached?

In order to answer these fundamental questions, we not only identify the tasks to be performed but also use the existing gold standard, the human vision system, as a benchmark and then examine non-human vision system sensing technology.

Most vehicles (aircraft, automobiles, and ships) have historically relied on the human visual system and processing to develop a dynamic world model of the local environment in which the vehicle operates. This world model is the basis on which almost all control decisions are made. While other senses provide input (audio for the status of engines and for radio communication, and feel for vibrations and g-forces), vision is the overwhelmingly dominant source of input used to develop and maintain a current world model.

The role of sensing and development of a world model (whether human or machine) is shown in Fig. 1.1. The two drivers of action are the sensor-derived world model and the goal. Note that the databases in the upper path would be the equivalent of life history, experience, and knowledge for a human. Figure 1.1 also compares this model of sensing and world model to the traditional observe, orient, decide, and act (OODA) decision loop first described by US Air Force Colonel John Boyd [1].

When aircraft (and automobiles) were first introduced, all navigation, processing, control, and decision-making was performed by the human operator (pilot). Control surfaces to adjust attitude, throttle, and brakes were manipulated by the operator using direct mechanical linkages. As these

In the aerospace business, there is an F-106—in the National Museum of the US Air Force (Wright–Patterson Air Force Base, Dayton, Ohio)—that in 1970 made an uncrewed landing in a corn field [6] after the pilot had ejected. The F-106 had sufficient automation that the pilot was almost unnecessary! In this case, the world model existed at headquarters, and the airplane was simply given basic navigation and targeting data to operate on.

The point of the aforementioned two examples is to highlight that AI and autonomous aircraft capabilities have existed for years. The biggest stumbling block has been less about automation and AI than it is about sensing and understanding the local environment, building the dynamic world model in which the automation takes place. The constraint on Winograd's blocks world was that the system began with an accurate world model and the only changes were those made by the system, in rearranging the blocks, so it always had an up-to-date world model without a requirement for sensing changes introduced by external actors.

Much of the early work on automation (such as the Winograd blocks world) focused on the reasoning and decision aspects of autonomy while deferring the sensing portion. Whereas Fig. 1.1 relates sensing to ultimate actions, Fig. 1.3 provides a slightly different view of the engineering buildup from sensing, through perception to development of a world view, to ultimately provide input to an autonomous system. Any of these functions can be implemented with either traditional algorithms or via what goes by various terms such as AI / machine learning (ML), deep learning, or simply neural networks.

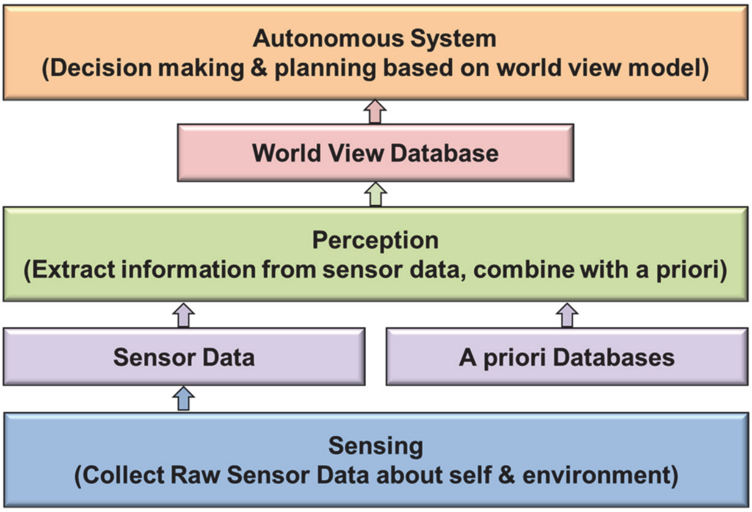

Figure 1.3 Autonomous behaviors rely on a world view that is derived from machine perception using input from the sensor data and *a priori* databases.

Chapter 2

Autonomous Flight Tasks

Takeoffs are optional. Landings are mandatory.

2.1 VEAF Tasks

The first step in understanding the role of vision systems in developing a world model for an autonomy engine is to identify the autonomy tasks that require a vision-system-derived world model. In the following we use the term "vision" or "vision system" to include both passive-imaging sensors [operating in the visible through long-wave infrared (LWIR) bands] as well as active systems such as 3D lidar or radar. The term "vision system" includes both the sensor and the associated sensor data-processing algorithms. The vision-system-based sensing and world-model-building phases of flight are

- Taxi
- Takeoff
- En-Route Navigation
- Obstacle Avoidance (Terrain and Airborne)
- Landing
- Formation Flight and Automated Air-to-Air Refueling (A3R)

An autonomous system must be able to perform these functions in all flying conditions, which includes day or night, and in various degraded visual environments (DVEs), such as rain, snow, fog/clouds, haze, and possibly smoke or dust.

Algorithms implementing any of the above functionality should include confidence factors for the solutions generated and in particular should be able to indicate when no solution is possible. Further, as part of the safety case, there should always be at least two independent solutions that can be compared for consistency.

Specific numerical requirements, such as field of regard (FOR) or detection range, will depend on the aircraft: speed, operational altitude and

Chapter 3

Sensing Requirements

"What are the facts? Again and again, what are the facts? And to how many significant digits? They are your only guide. If it can't be expressed in numbers, it's opinion." –Robert Heinlein

While the previous chapter established the basic autonomous flight capabilities that benefit from or require vision systems, and further provided a preliminary list of vision system functionality, in this chapter those capability requirements are narrowed down to specific sensing requirements: FOR, range, resolution and pixel counts, latency, and sensor phenomenology. As previously noted, an autonomous system must be able to perform these functions day or night, in DVEs, and must be able to identify when these functions cannot be performed.

In this chapter we will use some nominal values such as aircraft sizes, speeds, and decision heights to derive sensor parameters such as FOR and resolution. These are not prescriptive, absolute numbers. Instead, we are demonstrating a methodology to derive sensor parameters and generating some generic numbers, both of which can be used until more platform- and application-specific numbers can be generated.

The basic approach is to determine what must be detected (i.e., the dimensions) and how long it will take to recognize the threat and perform an appropriate action. We use DAA to illustrate the procedure. First, we determine what class of airspace we will be operating in and the speed and size of aircraft in that airspace. We also assume a reaction time to recognize the threat and take evasive action. In the following we will assume 10 s for a total recognition and maneuver time (this compares well with the cat-1 decision height time until touchdown for a 737), but the reader may wish to select a different reaction time based on the agility of their platform. Then we determine the minimum range at which we need to detect the other aircraft, assuming a worst-case, head-on closing scenario. If we need to estimate range from passive imagery, we then assume a minimum number of pixels across the target (5 px is a good number) to estimate the range based on the size of

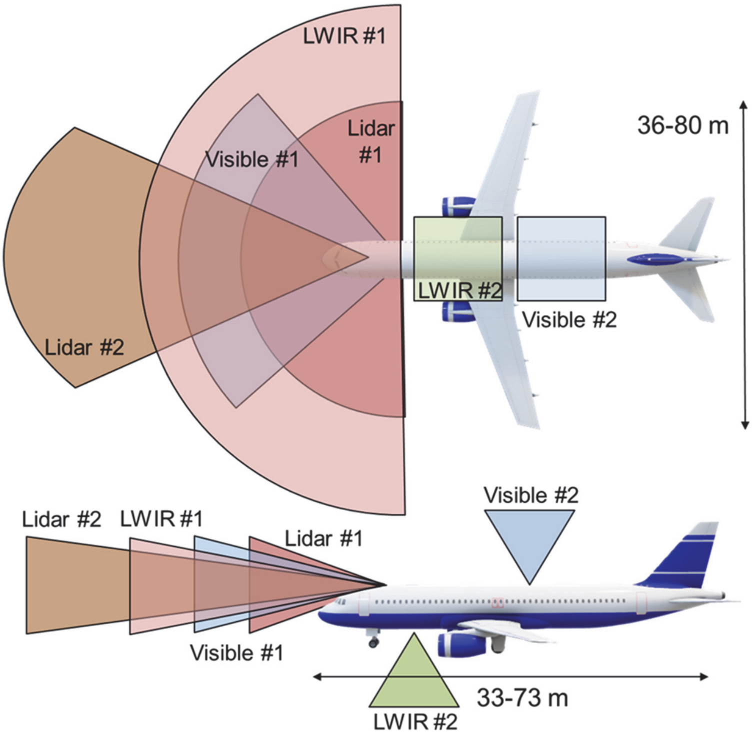

Figure 3.1 Example of a comprehensive sensor suite.

A comprehensive fixed wing sensor suite might consist of the following six sensors (Fig. 3.1):

- Lidar #1 (short range):
  - Forward-looking with a 180-deg × 25-deg FOR, 100-m range, and 0.25-mrad resolution. Generates 12,566 × 1,745 samples per scan.
  - Provides obstacle detection during taxi and gate operations.
- Visible sensor #1:
  - Forward-looking with a 90-deg × 25-deg FOR and 0.3-mrad resolution (5,236 px × 1,454 px).
  - Primary: Detect and interpret color airport lighting and read airport signage.
  - Secondary:
    - Provide detection of runway markings (edge, centerline, and thresholds).
    - Detect and interpret human airport-gate-personnel hand signs.
- Lidar #2 (long range):

Chapter 4

Sensing Systems

Optics is light work.

4.1 Human Vision

Since a human operator using the basic human visual system has from the beginning provided the world model, in which the human operates the vehicle, we can use the human visual system to generate an initial baseline for performance. Of course, as regulations develop for various functions, these functions may impose somewhat different (greater or lesser) demands on an autonomous vision system. The vehicle itself may limit certain aspects of the human vision system, such as the FOR. Similarly, we must recognize the limitations of the human visual system, including limited attention span.

FOR / FOV: The human can provide a 360-deg (4π steradian spherical to be precise) FOR, and at any moment, a 210-deg-wide by 150-deg-high binocular FOV [1]. The FOR in particular may be limited by the vehicle: in most transport aircraft, the pilot can only see in front of the aircraft, which drives certain rules of operation, such as the convention that an overtaking vehicle is responsible for avoiding a leading vehicle. However, in contrast to a transport aircraft with limited FOR, an automobile has 360-deg visibility (in a 2D plane) and hence has the responsibility to not move in front of a vehicle approaching from the rear, while a fighter aircraft is designed to have as close to 4π steradian visibility as possible.

Frame Rate and Latency: Broadcast video was established at 30 Hz (in the US; 25 Hz in Europe) based on the human visual response. In general, humans do not see variation at rates greater than 30 Hz (it is sometimes possible to see, especially the European 25-Hz electric light flicker out of the corner of the eye; the minimum temporal separation to distinguish two brief light pulses is 15–20 ms), and the eye can see momentary changes such as a flash or streak, which is of much shorter duration. In general, we can use 30 Hz as an upper limit on the video input rate from the human eye. This sets an upper limit on the required frame rates and latency for vision systems in

Chapter 5

Processing and Architectures

Experience is what you get when you are expecting something else.

You don't need a parachute to sky dive. You only need a parachute to sky dive more than once.

From a sensor perspective, a sensor generates raw data while subsequent processing extracts information from the raw data. However, at a system level, many people consider the sensor to include the front-end processing to perform both low-level functions such as flat fielding, image clean-up, greyscale management, compression, multisensor fusion, or stitching as required, and higher-level functions such as object detection, characterization, and tracking. This chapter provides a brief introduction to some of the sensor system issues a designer should consider.

5.1 Architecture

Architectures need to support several major functions:

- Safety: This is foundational and must be built in from the start. Some of the basic principles include redundancy, integrity, cross-checks, and confidence measures. Good architectures should have no single-point failures.
- Modular and upgradeable: Open systems with well-defined interfaces allow one to replace, upgrade, or add new sensors or algorithms as technology improves.
- Multiuse sensors: Within the constraints of cross-checks and redundancy, we wish to minimize the number of sensors required on the platform. An understanding of required FOVs, FORs, resolution, and when sensors are needed for each function can provide insights into how to minimize the total number of sensors.
- Off-nominal operations: The architecture should support contingencies for off-nominal operations, whether due to internal or external factors.

Chapter 6
Algorithms

What are the error bars? Everything is a probability distribution.

6.1 Terminology

Traditional algorithms, which we will refer to here as “conventional,” have existed as long as digital computers (longer actually since an algorithm is just a repeatable formula, such as a recipe), while machine learning (ML) has recently received high levels of interest. Unfortunately, significant confusion can arise over terminology. As such it is worth spending some time to clarify our terminology. In this book we will use the following:

- Conventional algorithms: These can typically be diagramed with a flow chart; they may include explicit mathematical formula but will almost always consist of a set of discrete computational steps to be performed. These will include many detection and tracking systems as well as significant portions of photogrammetry such as triangulation or MBPE. Conventional algorithms—when properly documented, implemented, and tested—can be certified under current flight safety regulations.
- Expert systems: These have traditionally been conventional algorithms coded to capture the knowledge of human experts, such as doctors for evaluating symptoms and creating a diagnosis. In many instances, these systems could be thought of as a decision tree incorporating a knowledge base and a rules engine to relate incoming data to the knowledge database. More advanced versions could use probabilistic decision trees and fuzzy logic. While these have traditionally been implemented as conventional algorithms, there is no reason they could not be implemented using neural networks and ML. Another example of an expert system would be a chess program, whether programmed explicitly with the rules of chess and various strategies, or trained as a neural net. For our applications, expert systems are probably minimal

6.6 Airborne Detect and Avoid

DAA requires that we detect other airborne objects and determine whether or not they are on a collision course with our platform. This involves a two-step process: (1) detect and track other airborne objects, and (2) determine if they are on a collision course with out platform.

Radar has generally been the preferred sensor since it can provide direct range measurements and it can penetrate clouds. However, passive-imaging sensors (in particular LWIR) are attractive for their generally lower cost compared with radar sensors and SWAP (making them particularly attractive, for small platforms that cannot tolerate the SWAP requirements of radar) and in some applications because they do not radiate. Radar and radar signal processing is generally outside the scope of this text, so we will confine our discussion to passive-imaging solutions.

Early (passive-imaging) DAA work focused on developing 4D trajectories for the detected objects and determining whether they would intersect the ownship trajectory. As discussed elsewhere, stereo- or two-sensor triangulation generally does not have sufficient range to generate useful trajectories. This left either 2D tracking augmented by a cued (laser) range finder or generating a synthetic baseline using ownship motion. There is a general rule for synthetic baseline triangulation, which states that you must be moving one derivative higher than the object you wish to range. Thus, for stationary objects on the ground, simply flying over at constant velocity is sufficient (leading to SFM algorithms), but for a target moving at constant velocity you must be accelerating. This led to schemes in which once a target was detected, the ownship had to execute a corkscrew maneuver to generate the acceleration required to range the constant-velocity target. Imagine all sorts of unmanned aircraft corkscrewing around the sky trying to passively range other aircraft.

The key insight was that for DAA we do not really care about the other aircraft trajectory per se. Instead, what we care about is time to impact [18]. Pilots judge a threat by whether or not it is moving in the FOV: a target which is observed to be moving across your FOV is not a collision threat. In contrast, an aircraft which is stationary in your FOV is either in formation flight with you, or it is on a collision course!

Consider two cases, resolved and unresolved targets: As soon as the object is resolved to more than a single PSF (i.e., has become at least a little bit larger than the PSF), we can begin to put absolute bounds on the range if we assume the largest object we will see is an Airbus A380 and the smallest is on the order of a Cessna 172 Skyhawk. Before it is resolved we can still make some estimates of range based on the time rate of change of intensity: if the percent change in intensity is rapid we conclude that the time to impact is less than if the percent change in intensity is small. This could be because the aircraft is far away but closing fast or because it is much closer but closing more slowly.

Chapter 7

Relevant Historical Aviation Accidents

O'Toole's commentary on Murphy's Law: Murphy was an optimist!

A crash occurs when you run out of options.

Returning to the theme of safety, it may be worthwhile to review a few famous aviation accidents, either related to failures of the vision system (pilot) or where the human vision system was a key component of recovery from another system failure. The purpose, of course, is to help ensure that any vision system developed for automation does not repeat these tragedies.

7.1 Midair Collision: Failure of DAA

The midair collision between a United Airlines Douglas DC-7 and a Trans World Airlines Lockheed L-1049 Super Constellation over Grand Canyon National Park on 30 June, 1956 [1,2] led to the creation of the FAA in 1958. At the time, it was not uncommon for aircraft flying near the Grand Canyon to make a detour to fly low over the Grand Canyon to provide passengers a view of this spectacular natural wonder from the air. The cause of the accident as described by the Civil Aeronautics Board (CAB) report was that the aircraft were both maneuvering around clouds and were unaware of each other, being outside controlled or monitored airspace, until moments before impact. All 128 people on both flights died in the subsequent crash. Both aircraft had departed Los Angeles International Airport at about the same time with one flight headed to Kansas City and the other to Chicago Midway.

The story, as related to this author, was that the pilots of the two aircraft, both departing the same place at about the same time, and knowing that the Grand Canyon was on their flight paths, had said, “See you over the Grand Canyon!”

The subsequent formation of the FAA reflected a common consensus among the airline industry, aircraft manufacturers, and the US government

Appendix

Triangulation Theory and Coordinate Transformations

A.1 Basic Triangulation Equations

The basic equations for two-camera triangulation are [1]

$$x_t = \frac{x_2 \tan\theta_1 - x_1 \tan\theta_2 + (y_1 - y_2)\tan\theta_1 \tan\theta_2}{(\tan\theta_1 - \tan\theta_2)} \tag{A1}$$

$$y_t = \frac{y_1 \tan\theta_1 - y_2 \tan\theta_2 + (x_2 - x_1) + x_2 - x_1}{(\tan\theta_1 - \tan\theta_2)} \tag{A2}$$

$$r_i = \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2} \tag{A3}$$

$$\langle z_t \rangle = \langle r_i \tan\varphi_i + z_i \rangle \tag{A4}$$

where $(x_1,\ y_1,\ z_1)$ is the location of sensor 1; $(x_2,\ y_2,\ z_2)$ is the location of sensor 2; θ_1 and θ_2 are the azimuth LOS to the target; φ_1 and φ_2 are the elevation LOS to the target for sensors 1 and 2, respectively; and r_1 and r_2 are the horizontal ranges to the target $(x,\ y)$ location. The target is located at $(x_t,\ y_t,\ z_t)$. The geometry is shown in Fig. A1.

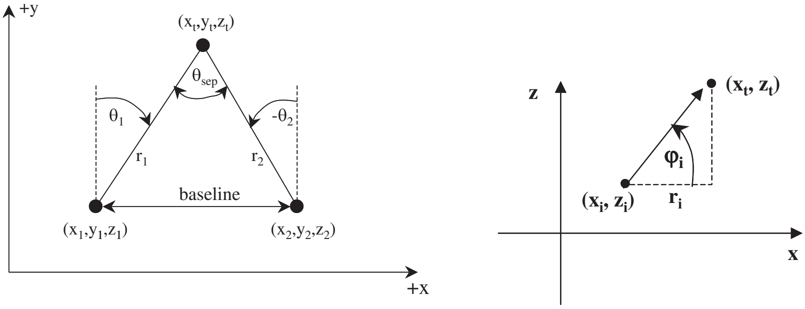

Figure A1 Triangulation geometry showing sensor 1, sensor 2, and target related by azimuth (θ) and elevation (φ) angles.

Jack Sanders-Reed spent thirty years developing and flight-testing advanced vision system technology, including novel sensors, distributed aperture sensor systems, intelligent agents for moving target detection, cable and wire detection, and automatic target recognition. As a Boeing Technical Fellow and then Chief Technologist for the Boeing Research & Technology, Avionics Systems Technology organization, Sanders-Reed led enterprise technology roadmap development and planning for machine perception and for mission and vehicle system architectures as well as supporting the sensors and apertures roadmap. He was the company lead for automated air-to-air refueling, working on vision system solutions for both boom/receptacle and probe/drogue refueling.

Sanders-Reed is internationally recognized for his more than 30 peer-reviewed and published conference papers and 6 patents, serving as conference chair for SPIE advanced vision system conferences (7 years), and as an invited lecturer at MIT as part of a high-speed imaging and analysis short course (18 years). He is also a Fellow of SPIE and co-author of the optics handbook *Photonics Rules of Thumb*. He is a developer of the Visual Fusion motion-analysis software package.

Sanders-Reed's career has included work in digital radiography and medical imaging (Picker x-ray), laser capillary ripple spectroscopy, and 5 years at MIT Lincoln Laboratory developing maximum-likelihood detection theory and multi-target tracking algorithms. He joined and helped grow a small business (SVS, Inc.) from 15 to 140 people prior to its purchase by Boeing in 2000. He holds a Ph.D. in physics from Case Western Reserve University and was top graduate from the Northeastern University High-Technology MBA program. He currently lives on a mountaintop at 7600 feet in a national forest.