TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

AUTOMATION & CONTROL INSTITUTE
INSTITUT FÜR AUTOMATISIERUNGS- & REGELUNGSTECHNIK

Design and Evaluation of a Deep Learning-based UAV Detection and Tracking System for Robotic Telescopes

DISSERTATION

Ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften (Dr.techn.)

unter der Leitung von
Univ.-Prof. Dr.sc.techn. Georg Schitter

eingereicht an der
Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik
Institut für Automatisierungs- und Regelungstechnik

von
Dipl.-Ing. Denis Ojdanić
[redacted]

Wien, im März 2024

Technische Universität Wien
Karlsplatz 13, 1040 Wien, Österreich

Acknowledgement

This thesis contains the results of the research work I have conducted over the past four years at the group for Advanced Mechatronic Systems of the Automation and Control Institute (ACIN), TU Wien. Successful and fruitful research is typically not done by individuals, but within a cooperative and fostering group environment consisting of researchers and supporters. For this reason I would like to take the opportunity to thank all the people who have supported me throughout my studies and who have played a crucial role in the successful completion of my doctoral thesis.

First of all I would like to thank my doctoral supervisor Univ.-Prof. Dipl.-Ing. Dr.sc.techn. Georg Schitter for welcoming me into his group and allowing me to work on interesting projects, which served as an inspiring source for my own research, as well as his help and guidance throughout the thesis. I would further like to thank Andreas for his invaluable insights and contributions towards my research. I am grateful to Christopher, with whom I shared an office and collaborated together on the research project Optofence II, who provided me thoughtful feedback, suggestions and support. Special thanks to Peter, who was a great support in the early months of my research career. I also want to extend my appreciation to all study assistants and students who contributed to the success of our research, especially Daniil, Benjamin and Niklas. I would also like to thank my colleagues Christian S., Christian H., Ernst, Hans, Nikolaus, Andras, Shingo, Thomas, Christoph B., Christoph Z., Martin F., Martin M., Daniel W., Johannes S., Johannes W., Daniel P., Bo, Mathias, Matthias, Alexander and Stefan.

Additional acknowledgment belongs to the industrial partner ASA Astrosysteme GmbH and their team, the BMLV, as well as the public funding agency FFG and the BMF, for making this research possible.

A special thank you goes to my parents and sister, who provided me the foundation for all my achievements. Finally, my heartfelt thanks goes to my dear wife Sarina, who is alway there for me and inspires me to be successful in live.

Abstract

Unmanned aerial vehicles (UAV)s have gained immense popularity over the past decade primarily due to their remarkable versatility and diverse fields of application. However, public availability and the ease of use has led to a soaring increase of UAV related incidents endangering people and critical infrastructure. In response, a significant effort is made in research and development of UAV detection, tracking and identification systems in order to facilitate an early threat assessment. Typical systems combine multiple sensors, such as radar, LiDAR, acoustics, radio frequency detection and electro-optics, with the latter being a key component as visual imagery is easily and conclusively interpretable for a human operator. A long operational range is crucial for a successful application of an optical system as it enables timely object identification. However, conventional camera-based solutions are limited with respect to the operational distance. Considering the fast speeds UAVs can achieve, current electro-optical systems leave very little reaction time to the operators once a clear image of the UAV is obtained up to the moment it reaches a prohibited or protected area.
Throughout this thesis the endeavour to increase the optical operational range, especially for the detection of micro and mini UAVs, is elaborated. For this purpose the usage of reflective telescopes is proposed, which enable long focal lengths and large apertures, while simultaneously retaining relatively lightweight system characteristics compared to camera lenses with similar specifications. However, the narrow field of view imposes challenging demands onto the system components to successfully track fast and agile UAVs. Hence, a sophisticated parallel architecture is implemented, which combines a deep learning-based object detector with a conventional object tracker through a dedicated transition strategy, to ensure fast and reliable object detection and tracking at 100 fps. Various algorithm combinations are trained, compared and tested on a custom UAV dataset, which is created throughout this thesis, to find the best performing combination. Furthermore, a high-speed automatic focus is implemented and

incorporated into the parallel architecture, to keep fast UAVs in focus. The telescope system is tested for its detection, tracking and focusing capabilities during multiple field tests, which include different types of UAVs and backgrounds resembling authentic scenarios. Finally, methods are explored to integrate a laser-based distance measurement for multiple objects into a telescope system.
This thesis strives to improve the optical detection range for small UAVs by implementing and demonstrating a telescope-based system capable of long range UAV detection. A holistic system is designed including all necessary individual components, which contribute to the success within the intended application. It has been demonstrated that a UAV with the size of 0.3 m in diameter can be detected and tracked up to distances of 4 km. Using a dedicated parallel software architecture, which facilitates a collaboration between an FRCNN object detector together with a MedianFlow tracker, detection and tracking of UAVs at 100 fps is enabled. The complementary implemented automatic focus keeps the UAV in focus down to distances of 150 m even at UAV speeds of 24 m/s. Finally, a prototypic laser system, which is integrated into a small telescope, measures the distance to four UAVs at 14 Hz per object to determine the 3D localization of the UAV.

Kurzfassung

Unbemannte Luftfahrzeuge (ULF) gewannen im letzten Jahrzehnt immens an Popularität aufgrund ihrer Vielseitigkeit und den diversen möglichen Anwendungsbereichen. Die einfache öffentliche Zugänglichkeit und die teils simple Steuerung der Drohnen führte zu einem Anstieg an Vorfällen, bei denen es durch den Einsatz von Drohnen zur Gefährdung der öffentlichen Sicherheit bzw. von kritischer Infrastruktur kam. Als Reaktion darauf wird erheblicher Aufwand in der Forschung und Entwicklung von Drohnen-Erkennungs-, Verfolgungs-, und Identifizierungssystemen unternommen, um eine frühzeitige Gefahrenerkennung und Bewertung einer herannahenden Bedrohnung aus der Luft zu ermöglichen. Typische Drohnendetektionssysteme kombinieren verschiedene Sensoren, wie Radar, LiDAR, Akustik, Funk und elektrooptische Detektion, wobei letzteres eine Schlüsselkomponente ist, da visuelle Bilder für einen menschlichen Bediener leicht und schlüssig interpretierbar sind. Eine große Reichweite ist für den erfolgreichen Einsatz eines optischen Systems entscheidend, da sie die rechtzeitige Identifizierung von Objekten ermöglicht. Konventionelle Lösungen sind jedoch limitiert in der Reichweite. Aufgrund der schnellen Drohnengeschwindigkeiten, bleibt beim Einsatz von aktuellen elektrooptischen Systemen nur eine kurze Zeitspanne zwischen der Aufnahme eines aufschlussreichen Bildes und dem Eintreffen der Drohne in einen geschützten Bereich, um entsprechende Gegenmaßnahmen einzuleiten.
Im Laufe dieser Arbeit wird ein System entwickelt, welches die Einsatzreichweite von optischen Systemen, die im Speziellen mikro und mini Drohnen detektieren sollen, deutlich erweitert. Für diesen Zweck werden reflektierende Teleskope eingesetzt, die durch ihre langen Brennweiten und großen Aperturen eine hohe Auflösung ermöglichen. Das relativ schmale Blickfeld des Systems bringt jedoch Herausforderungen beim Verfolgen von schnellen und agilen Drohnen mit sich. Daher wird eine parallele Architektur implementiert, welche eine Kollaboration zwischen einem Deep Learning basierten Objekt Detektor mit einem konventionellen Objekt Tracker ermöglicht, um mit bis zu 100 fps Drohnen zu detektieren und

zu verfolgen. Diverse Algorithmen werden auf einem eigenen Drohnendatensatz trainiert und verschiedene Kombinationen aus Detektoren und Trackern getestet. Des Weiteren ist ein proprietärer automatischer Fokus implementiert und in die parallele Architektur integriert worden, um eine scharfe Abbildung der Drohne zu gewährleisten. Die Detektions-, Tracking- und Fokussierfähigkeiten des Teleskopsystems werden durch umfangreiche Feldtests, welche diverse Drohnentypen und realitätsnahe Einsatzszenarien abdecken, getestet. Abschließend wird ein Prototyp entwickelt, der eine laserbasierte Distanzmessung in ein Teleskopsystem integriert, um die Distanz in Echtzeit von mehreren Objekten zu messen.
Ziel dieser Arbeit ist es, den optischen Erfassungsbereich für kleine Drohnen deutlich zu erhöhen, indem ein teleskopbasiertes System entwickelt und umfassend getestet wird. Ein ganzheitliches System wird entworfen, das alle notwendigen Einzelkomponenten integriert, die zum Erfolg der geplanten Anwendung beitragen. Es wird demonstriert, dass Drohnen mit einem Durchmesser von 0.3 m bis zu 4 km weit detektiert und verfolgt werden können. Die implementierte parallele Architektur ermöglicht eine Kollaboration zwischen einem FRCNN Objekt Detektor und einem MedianFlow Tracker, um Drohnen mit 100 fps zu verfolgen. Der eigens entworfene automatische Fokus hält heranfliegende Drohnen bis zu einer minimalen Distanz von 150 m im Fokus, selbst bei Drohnengeschwindigkeiten von 24 m/s. Ein prototypisches Lasersystem zur Distanzmessung ist zusätzlich in ein kleines Teleskopsystem integriert, welches die Distanz zu vier Objekten mit jeweils 14 Hz pro Objekt vermessen kann.

Contents

List of Figures

CHAPTER 1

Introduction

"*But optics sharp it needs, I ween, To see what is not to be seen.*" —John Trumbull

Unmanned aerial vehicles (UAV) have become increasingly popular and simultaneously widely available over the past decade [1]. The growing market is expected to reach EUR 10 billion by 2035 and EUR 15 billion by 2050 annually [2]. UAVs are versatile tools due to their ability to fly and to be remotely controlled, which makes it easy to reach inaccessible and even dangerous locations. Moreover, UAVs are often equipped with video cameras enabling them to capture footage and document an area from an airborne perspective. Finally, UAVs are capable of carrying payloads, which further increases the number of applicable operational scenarios. Some examples of applications covered by UAVs include leisure activities, like filming and videography [3], but also drone racing [4], land surveying [5], agriculture [6], search and rescue [7], inspection of buildings [8], wildfire monitoring [9] and many more. Apart from the civilian applications, UAVs are often used for military purposes [10]. While larger military UAVs are difficult to purchase and are typically restricted for export, small commercially available UAVs are readily available to the public. The drawback of the versatility, ease of use, the wide and relatively cost efficient availability, is the unauthorised use of UAVs, which is summarized by scenarios where legislations are violated, for example, by operating a UAV in restricted areas.
As of today, numerous documented incidents corroborate the threat posed by UAVs. Airports are especially vulnerable to the misuse of UAVs. An infamous incident is the 2018 shutdown of the London Gatwick airport for more than 24 hours due to nearby UAVs and in 2023 the airport had to delay its air traffic again for an hour due to a suspected UAV [11]. In 2020 the Frankfurt airport had to cease operation for 1.5 hours, which led to 9700 passengers being affected,

57 flights being rerouted and 72 cancelled [12]. In another encounter, a commercial jet almost collided with a UAV during approach at the Los Angeles International Airport [13]. Many more incidents, which have occurred in the past, reaffirm the vulnerability of air traffic with respect to misuse of UAVs [14–18].
Similarly, studies highlight the potential endangerment of nuclear power plants through unauthorised UAV usage in the vicinity in the form of distraction, reconnaissance or even as kinetic attacks [19]. In 2015 and in 2022 UAVs have been sighted close to nuclear power plants in France [20] and Sweden [21], respectively.
Besides the threat to critical infrastructure, the versatile nature of UAVs enables other illicit activities. In 2015 a UAV with traces of radioactive material was found on the roof of a governmental building in Japan [22]. In 2018 the president of Venezuela has been targeted by an assassination attempt involving the usage of UAVs [23]. Further illegal activities involve breaching of air exclusion zones, flying over crowds of people or the transportation of goods over state borders or into prisons [24–26].
Due to the alarming rate and diversity of UAV related incidents, authorities are trying to put forth regulations. The European Aviation Safety Agency (EASA) has created regulations against the illicit use of UAVs in the European Union. UAV operation is categorised into three groups according to the EASA regulation: 'open', 'specific' and 'certified' category [27]. Drones under the 'open' category have a maximum take off weight (MTOW) of below 25 kg, are permitted to a maximum flight height of 120 m and a visual line of sight (VLOS) has to be maintained between the operator and the UAV [27]. The risk emanating from this category is specified as low, however, intentional or unintentional unauthorised use of these drones can lead to hazardous situations around airports during take off and landing, but also at concerts or sporting events when operated around large assemblies of people for example. Authorised UAVs, which are performing a mission in the very low airspace like visual inspection or the transportation of goods, are within the 'specific' category and lastly, 'certified' drone operation is put on the same level as manned aviation [27].
Apart from categorisation of the drone operations, EASA labels the intention of unauthorised UAV usage into negligence, gross negligence and criminal/terroristic [27]. A negligence summarises incidents, where unaware UAV operators violate regulations or restricted areas with no intent of causing harm [27]. Actions of people who know the regulations and deliberately violate them disregarding any repercussions fall into the category of gross negligence [27]. An example are protesters who violate regulations using UAVs for their cause. Finally, criminal or terrorist activities involve deliberate and harmful applications of UAVs to disrupt or destroy infrastructure and endanger human beings [27].

Despite the regulations which are put in place, numerous UAV related incidents paired with the fact, that the commercial usage of UAVs is increasing, corroborate the necessity for appropriate UAV detection and tracking systems to promptly

assess potentially harmful activities and initiate counter measures. Early detection, meaning the initial perception of a UAV breaching the perimeter of a protected area and subsequently, reliable tracking to continuously monitor the UAV position for possible counter measures, is important.
Several methods based on different sensor technologies exist to detect and track UAVs over long distances. Radar [28], LiDAR [29], radio frequency (RF)-based sensors [30], acoustic microphone arrays [31] and electro-optics (EO) [32] are currently the main detection and tracking means. Typically, multiple sensors are combined to form a multispectral detection and tracking system, which aggregates the benefits of each sensor [33]. However, a key component in such a multispectral system is the optical sensor. Camera-based components allow to capture images of the surroundings and therefore, provide expressive and easily interpretable information. This facile and fast situational assessment is paramount for appropriate and timely measures against a potential threat posed by an incoming UAV.
Typical EO UAV detection sensors are equipped with zoom cameras, which features an adjustable field of view (FoV) to enable object detection and tracking [32]. For the surveillance of a wider area the cameras are attached to a pan and tilt mount. The major problem with optical sensors, which utilize lenses, is the limited detection range for small objects like commercially available UAVs, as lenses with large apertures and long focal lengths introduce aberrations and are disproportionally expensive. The operational range to detect and track micro and mini UAVs, with sizes down to 0.3 m in diameter, is limited to about 2 km, which leaves little time to assess the situation and prepare for counter measures once the image is captured of an incoming UAV [34].
To enable a timely situational assessment, the optical detection and tracking range for micro and mini UAVs has to be extended significantly. Longer ranges prolong the reaction time of the responsible personnel, once a visual image of an incoming threat is captured. For these long distances, long focal lengths and large apertures are required to ensure sufficient resolution. As UAVs are capable of achieving high speeds, real-time image processing is important, as it allows fast detection of UAVs within video images. A high sampling rate of the UAV position facilitates the control of the required high bandwidth pan-tilt camera mount system to keep the UAV within the FoV. Another challenge, which arises through the use of long focal lengths with larger apertures is the automatic focus, which, likewise, has to be fast and precise, to keep the UAV within focus and capture sharp and information rich images.
To enable optical long distance UAV detection and tracking, a system is needed, which automatically detects, tracks and focuses onto UAVs within video images at a high frame rate. The video images have to be captured by an optical system, which uses long focal lengths and large apertures, to guarantee sufficient resolution for long distance detection and tracking of small UAVs.

1.1 Scope of the thesis

This thesis deals with the design, implementation and experimental evaluation of a telescope-based UAV detection and tracking system. The prevalent goal is to extend the optical detection and tracking range for small commercially available UAVs to several kilometres. For this purpose, a complete telescope-based optical detection and tracking system is designed and developed including all essential components, ranging from the optical setup, the software for efficient detection and tracking, to the automatic focus. The proposed system is extensively evaluated during field tests, which incorporate realistic operational scenarios. An additional prototypic system is designed to demonstrate the potential of integrating a laser range measurement system into a telescope system for precise real-time 3D localization of multiple UAVs.
An overview of incumbent detection methodologies for UAVs, with a detailed review on optical techniques, is given in Chapter 2. In Chapter 3 a collection of the scientific journal and conference publications is listed, which form the core and main contribution of the thesis. Finally, in Chapter 4 a comprehensive discussion of the results followed by a summary and outlook, is given.

CHAPTER 2

State of the Art

This chapter provides an overview of the current state of the art of UAV detection and tracking techniques. First, a brief overview of relevant UAV classes, listing their most important properties, is given. Then, various sensing principles are elaborated in detail, which are used for the purpose of UAV detection and tracking. The main focus of the chapter covers the electro-optical (EO) sensing technology, where the relevant hardware and software algorithms are discussed to detect and track UAVs within camera frames. A brief overview about multispectral systems is presented, which combine multiple sensing techniques to form a holistic detection system. Finally, this chapter identifies open challenges and missing innovations to create a high performance long distance optical UAV detection and tracking system.

2.1 UAV classes and specification

The European Union and the European Union Aviation Safety Agency (EASA) issued regulations to standardise UAV operations across the EU member states [35]. These regulations categorise commercial UAVs based on their maximum takeoff weight (MTOW) and their operational risk. UAV operations are categorised into 'open', 'specific' and 'certified' category, whereas the first is not subject to any prior authorisation in contrast to the other two categories [27]. The 'open' category summarizes the main reference for commercially available UAVs.
The 'open' category is further divided in the subclasses A1 to A3 of UAV operation, which specify the proximity of operation to non-involved people or groups of people [35]. Apart from the operation, the UAV class is categorised into subclasses according to the MTOW of the UAV and various features, such as UAV speed

Table 2.1: Overview of the 'open' category showing the relation between UAV classes, MTOW and subcategory according to EASA regulation. The latter indicates where the UAV is allowed to be operated [35].

C-Class	MTOW	Subcategory
C0	$< 0.25\,\mathrm{kg}$	A1
C1	$< 0.9\,\mathrm{kg}$	A1
C2	$< 4\,\mathrm{kg}$	A2
C3	$< 25\,\mathrm{kg}$	A3
C4	$< 25\,\mathrm{kg}$	A4

and maximum altitude.

Irrespective of the categories listed in the regulation, UAVs are available in various shapes and propulsion systems as depicted in Fig. 2.1. UAVs can be shaped as small air planes in a 'fixed-wing' configuration and contain propellers or even jet engines as propulsion system. Gliders on the other hand do not have a propulsion system. Similarly to nature, flapping wings UAVs or Ornithopters are implemented, which propel themselves forward through flapping motions of their wings. Copter UAVs use propellers, like helicopters, and various designs, like the most common quad-copter or hexa-copter, exist. [36]

(a) Quadcopter [37].

(b) Fixed-wing [38].

(c) VTOL fixed-wing [39].

(d) Octocopter [40].

(e) Jet powered UAV [41].

(f) Ornithopter [42].

Figure 2.1: Examples of different UAV types with various propulsion systems.

This variety of designs makes UAVs very versatile tools for different applications. However, as the examples in Section 1 illustrate, these activities can also be of malicious nature. Hence, it is important to develop systems capable of detecting these small and fast objects. Various methods exist already to detect,

track and identify UAVs based on different sensor technologies. In the following section an overview of the most common and effective sensor types is given.

2.2 UAV detection and tracking technologies

Several sensor types exist to perform remote UAV detection. Detection involves spotting and discerning an object within the operational range, but not necessarily classifying or identifying it. Likewise, tracking can be defined as following an object over a certain amount of time. Classification, is the process of categorizing a detected or tracked object into a certain object class, whereas identification enables to uniquely discern objects, even within the same class. For the thesis the focus is put on detection and tracking, whereas an implicit classification is performed during the detection phase, as only UAVs are detected and other flying objects such as birds or other aircraft are treated as background. The main goal is an early threat detection and a subsequent tracking of a UAV, which allows a precise localization required for possible counter measures.

2.2.1 Radio frequency

One approach to localize UAVs is to passively detect the emitted radio frequency (RF) signals that are utilized for communication between the operator and UAV [43]. UAVs communicate via RF signals with a remote controller transmitting information like location, altitude, video images, battery level etc. and receiving control commands from the remote controller [44]. Typical frequency bands are 2.4 GHz and 5 GHz, with commercial detection systems operating in a much wider 100 MHz to 6 GHz bandwidth [45]. An RF sensor scans for signals unidirectionally allowing a 360° surveillance radius. An unidentified object can be detected up to a distance of 5 km [45, 46]. As an additional advantage, multiple UAVs can be detected simultaneously and localization of the controlling device is possible [47]. There are two methods to perform a localization, identification and classification once a signal is measured. The first method, decodes or hacks into the communication protocol between the UAV and the controller essentially granting access to detailed information by eavesdropping on GPS coordinates, altitude and speed data etc. The second method relies on the UAV sharing its information deliberately. DJI Aeroscope monitors location and registration number of DJI UAVs, which explicitly share this information, and the detection range by using dedicated RF sensors is given up distances of 50 km [47].
The second method, which is most commonly applied, is sensing passively for RF signals, comparing detected signatures to extensive databases for identification and utilizing triangulation to locate the UAV position [43]. Using this technique the study shows experimental results of detecting a DJI drone in about 100 m to 150 m with an angular uncertainty of about 12°, which enables localization of the

Figure 2.2: Passive RF-based UAV detection sensor by the company Dedrone [46].

UAV down to an error of about 12 m, when using two RF sensing modules and triangulation. As most commercial UAVs have unique RF signatures, an identification is performed by applying RF fingerprinting techniques, which discern different UAVs based on characteristics of the radio transmission signal like the spectrum of the signal or the amplitude envelope [48]. Combining appropriate signal analysis with machine learning facilitates classification and identification. Extracting, for example, the signal amplitude envelope with a subsequent principal component analysis and classification using a neural network results in an overall classification rate of 95 % in an indoor environment and can be also used outdoors [49]. By applying machine learning methods such as the neighbourhood component analysis [50], the k-nearest neighbour (KNN) algorithm after a short time Fourier transform [51] or the using of deep neural networks [52], UAV classification based on the received signal is possible. Similarly to the aforementioned analysis of the physical layer of the RF signal, eavesdropping on the Wi-Fi protocol, which is used for video streaming, allows detection and classification. Training and applying appropriate machine learning algorithms on the captured data, which consists of packet statistics like total number of packets, packet length, the inter-arrival time etc. facilitates classification [44]. Other than detecting UAVs, RF jammers are utilized to actively protect a perimeter against UAVs, by disrupting the control signal between the operator and the UAV [53].

RF detection techniques offer localization and classification capabilities of the UAV and localization of the operator, however, it can easily be bypassed by the adversarial side, by pre-programming the UAV flight path. Hence, no RF signals are transferred between controller and UAV, which, consequently, remains invisible to the detection system. Another drawback is the necessity of estimating the utilized communication specifications like the frequency band of the UAV. Using specialized RF communication technologies such as spread spectrum modulation

Table 2.2: Overview of the advantages and disadvantages of an RF-based UAV detection sensor.

Advantage	**Disadvantage**
Long range detection 5 km to 50 km depending on the RF detection method	Cannot detect autonomous UAVs, which do not communicate with a controller
Can localize UAV and operator accurately by decoding position from communication	False positives from detecting other RF signatures like Wi-Fi
Classification and identification possible	May require sophisticated signal processing
Wide area coverage and support for NLOS detection	Performance degrades in environments with obstacles
Not severely affected by weather and works during day and night-time	Vulnerable to jamming and spoofing
Passive method, hard to detect for adversary	Spread spectrum hopping and frequency modulation make decoding communication challenging

or frequency hopping imposes additional challenges onto the detector side, when trying to decode the communication [54]. The main advantages and disadvantages of the RF sensor, which are described within this section, are summarized in Table 2.2.
Two well known examples for RF-based UAV detection systems, which are commercially available, are DJI Aeroscope [47] and DedroneSensor RF-360 [46] with the latter being depicted in Fig. 2.2. The Dedrone sensor achieves RF-based detection up to 5 km according to its specifications.

2.2.2 Radar

Radar does not suffer from the aforementioned limitations and provides another method to detect UAVs. Typical radar systems emit radio frequency waves and measure the reflection of an obstructing object [55]. Based on this measurement, the distance, velocity and to a certain extent the type of an object can be determined [55]. A paramount parameter in this context is the radar cross section (RCS) [56], which is the equivalent of the object seen by radar and thus a measure on how detectable it is. Traditionally, radar systems are designed to detect large flying objects such as air planes with RCS larger than $1\,m^2$, which are flying at a high altitude and fast speeds with smooth trajectories. UAVs on the

Figure 2.3: Air and Ground Drone Surveillance Radar Ranger R8SS-3D by Teledyne FLIR [58].

other hand are very small compared to air planes and mainly consist of plastic, effectively reducing the RCS to the battery and the main board of the system [57]. Furthermore, UAVs fly relatively low, introducing ground clutter as an additional problem.

The usage of pulse-based radar systems is infeasible for detection of small UAVs due to limitations in bandwidth, power and ultimately the cost of the system [59]. A potential solution is the utilization of a cost effective W-band FMCW radar, as the millimeter wavelength is sensitive to small objects [60]. Furthermore, the Doppler-Effect can be exploited to estimate the UAV velocity and flight direction [60]. More expensive and complex MIMO radar systems like the holographic radar, combine an L-band 2-D antenna with signal processing to create a 3-D monitoring device capable of detecting small consumer UAVs [61]. Compared to conventional radar, which suffers from the trade-off between revisiting time and Doppler resolution, the 2-D antenna improves the achievable resolution [61]. Similarly, an X-band FMCW radar has been developed, which uses multiple simultaneous beams to continuously measure the scene and thus increase the range resolution especially for small and slow flying targets. This enables the detection of a DJI Phantom 4 up to 2 km [62]. Other studies claim longer detection distances of more than 3 km, however, a continuous track remains challenging [34]. Another crucial aspect for radar-based detection of small flying objects is the large number of false positives, as the small RCS of UAVs is ambiguous and confusable with birds or ground clutter. Using machine learning, targets are classified based on different collected features from the measured data such as flight trajectory, height, Doppler measurements, acceleration etc. [61]. The micro-Doppler effect

Table 2.3: Overview of the advantages and disadvantages of an radar-based UAV detection sensor.

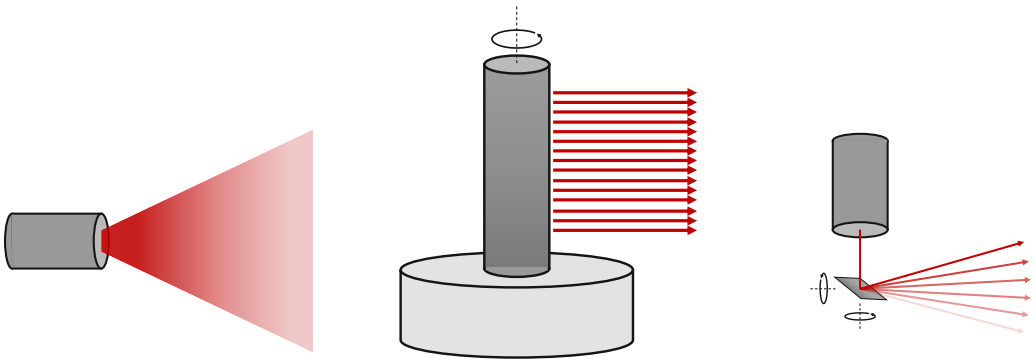
Advantage	**Disadvantage**
3 km to 5 km detection range	Classification and tracking is challenging
Accurate localization and velocity estimation	False positives, struggles to distinguish small UAVs from birds
Operation during day, night and adverse weather	Performance decreases with background clutter from e.g. trees and buildings
Wide area coverage	High power consumption
	Active measurement, can be detected by adversary

can also be exploited for the classification in combination with machine learning or deep learning algorithms. It provides information on the number and size of the rotor blades for example, which is valuable information for deducing the type of UAV or to distinguish a UAV from a bird [63–65].
While radar systems offer good object detection capabilities over relatively long distances and a wide areal coverage, classification, albeit possible, is challenging for small UAVs. A summary of the main advantages and disadvantages of a radar sensor is given in Table 2.3.
Some prominent incumbent companies providing commercially available radar systems for UAV detection include Robin Radar [66], Teledyne FLIR [58] and Aerial Armor [67], with the latter claiming a detection range of 5 km for UAVs according to the specification.

2.2.3 LiDAR

Similarly to radar, LiDAR or laser-based methods, use the time of flight (ToF) principle to obtain the distance to an object by measuring the time an electromagnetic wave takes to travel to an object and back [68]. Radar utilizes longer wavelength radio waves, whereas LiDAR transmits light waves. LiDAR systems, which are commonly applied for autonomous vehicular applications [69], have also been used for UAV detection and localization [70, 71].

Generally, LiDAR systems are classified into three different categories, flashing, scanning and rotating LiDAR as depicted in Fig. 2.4c. The first implements a diverging light source, which simultaneously illuminates the whole FoV. On the receiver end, a matrix of photodiodes measures the reflected signal to obtain the distance and a low resolution depth image [72, 73]. While flashing LiDAR can



(a) Flashing LiDAR. (b) Rotating LiDAR. (c) Scanning LiDAR.

Figure 2.4: Overview of different LiDAR methods. a) Flashing LiDAR illuminates a large FoV simultaneously. b) Rotating LiDAR uses an array of collimated laser beams to scan the surroundings by rotating the sensor. c) Scanning LiDAR uses tip-tilt mirrors to align the laser source for measurements.

achieve long operational distances of up to 1000 m, the resolution is low, which is a major impediment for the detection of small commercial UAVs [74]. As opposed to flashing LiDAR, rotating and scanning LiDAR employ a collimated laser beam spreading the laser light energy over a comparably small surface. The benefit is the ability to measure a much longer distance with high resolution, however, in a sequential and time consuming manner [75]. Rotating LiDAR has been applied to UAV detection [70], whereas the number of vertical sensors and thus the achievable resolution represents a paramount design parameter. Measurement distances of multiple kilometres have been achieved using scanning LiDAR systems [76, 77]. Exploiting the generated point cloud together with the signal image of a scanning LiDAR sensor, the classification rate can be improved [78]. While attempts are being made to reduce the point cloud density and therefore, increase the repetition rate of a FoV scan [79], the sequential scan of a larger FoV proves impracticable. A major challenge is the low scan rate for a large FoV, when trying to detect and track fast and small moving objects. To circumvent this drawback a combination with a camera system can increase measurement repetition rate. Combining ranging and camera systems has already be studied extensively, for example, to generate terrain maps with topography [80] or for atmospheric remote sensing [81, 82]. By benefiting from the camera image to localize the UAV, the distance to the UAV is measured with a laser range finder (LRF), which is attached to the camera mount in a bi-axial configuration and is aligned with the centre of the camera image. By tracking the UAV with this bi-axial system an accurate 3D position of the UAV is acquired [83]. To obtain a distance measurement using this configuration, the UAV has to be precisely centred within the camera image. However, the overall system bandwidth is limited by the bandwidth of the mount

Table 2.4: Overview of the advantages and disadvantages of an LiDAR-based UAV detection sensor.

Advantage	Disadvantage
Long range measurement possible >1000 m	Requires additional sensor to reduce the search area for long range measurements
Highly accurate distance and position measurement	Wide area coverage only up to very limited distances of a few hundreds of meters.
Low false positive rate	Emits a signal actively and can be detected by adversarial side
Works during day and night time	Depending on the wavelength, the performance might degrade in adverse weather conditions

actuating the camera and LRF. The main advantages and disadvantages of a LiDAR sensor for UAV detection are summarized in Table 2.4.

2.2.4 Acoustics

UAVs can be detected based on the emitted sounds during flight such as wind passing the wings, the sound of the engine and the propeller blades. Typically, the latter one is being used for detection and classification as the most dominant sound source. The sound is measured by acoustic sensors in the form of microphone arrays to cover a broad angular range. The microphone placement within these arrays can vary, for example, a spherical configuration for omnidirectional range coverage [84] or equal alignment into one direction for improved accuracy [85]. Furthermore, due to low cost of microphones, a single detection system may incorporate hundreds of microphones, which can collaborate in order to locate the sound emitting source [84, 85]. Tracking and trajectory estimation of a sound emitting object such as a UAV is performed by a meticulously calibrated microphone array, which enables delay-and-sum beam-forming [85]. Another option for a 3D localization of the UAV is the combination of multiple acoustic cameras [84].

Once a signal is recorded, the power spectral density is calculated and cross correlated with an a priori recorded database of signals for detection and classification [87]. To further improve the detection and classification accuracy of the sound-based system, machine learning approaches are incorporated. By combining certain feature extracting methods such as the Mel frequency or also the linear

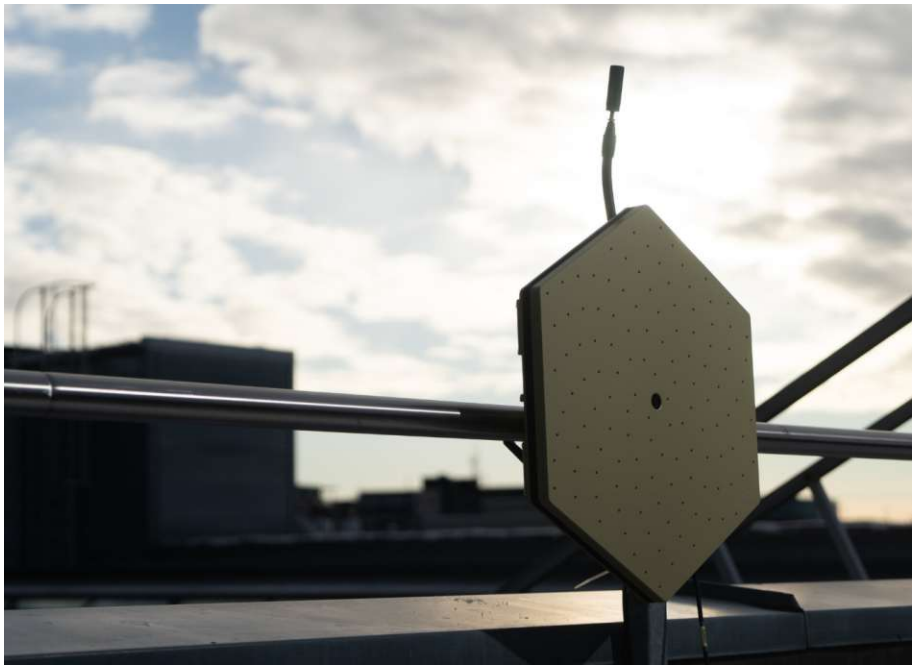


Figure 2.5: The Discovair G2 sensor consists of 128 microphones and an optical camera, which provides visualization of the targets based on the recorded audio signal [86].

predictive cepstral coefficients with a subsequent classification using a pre-trained support vector machine (SVM) eventuates improved results [88]. The training and test dataset consists of audio samples of different lengths and various ambient noises such as birds, thunderstorm, airplanes and UAVs [88]. Other methods combine the Mel frequency cepstral coefficients with K-neigherst-neighbor (KNN) algorithm for UAV classification [89]. More recently, the advances of deep learning proves to be a promising approach for the classification of UAVs based on recorded sound samples. Converting the measured audio signal into a spectrogram and feeding it into a convolutional neural network (CNN) or a recurrent neural network (RNN) for subsequent classification, which not only asserts if a UAV is present, but also the UAV type [90].
The advantages of acoustic sensor are the low cost for the microphones and the ability to operate in day and night conditions and without direct line of sight (LoS). Angular uncertainties for the target localization of about 4° can be achieved [91,92] and even classification is possible by using comprehensive databases and machine learning [88, 89]. The disadvantage of this technology is the short operational range of below 600 m [93], since it suffers greatly from ambient noise, which lies in a similar frequency band as the emitted UAV sound, and signal attenuation over long distances [94]. The main advantages and disadvantages of an acoustic sensor for UAV detection are summarized in Table 2.5.

In Fig. 2.5 an example of a commercially available acoustic UAV detection sensor Discovair G2 [86] is depicted.

Table 2.5: Overview of the advantages and disadvantages of an acoustic-based UAV detection sensor.

Advantage	**Disadvantage**
Passive detection method, hard to detect for adversary	Very short detection range below 600 m
Inexpensive and compact	Susceptible to environmental noise such as wind, traffic or other noises in a similar spectrum
Classification of UAVs possible	Prone to false positives due to similar noise sources
Operation possible at day, night-time and adverse weather conditons	

2.3 Electro-optics

EO systems use cameras to capture images of the surrounding area to detect, track and identify objects. The basic principle involves an object being illuminated passively by some ambient light source. The light reflected by the object itself is captured via an optical lens and focused onto a camera sensor to form an image as depicted in Fig. 2.6. The main advantage of EO detection is the captured image, which can either be analysed by advanced computer vision algorithms, but most importantly, it presents a meaningful piece of information for a human to interpret and decide for potential counter measures. EO systems are, however, susceptible to adverse weather conditions like rain or fog and typically reach their limits during night time. To some extent, EO systems can operate in these conditions, by selecting different imaging frequency bands or choosing active illumination, as discussed in more detail within this section. Some examples of commercially available camera-based UAV detection systems include SkyPatriot [95], AARTOS Camera [96] or the DeDrone PTZ cameras [97].
A brief summary of the main advantages and disadvantages of an EO sensor for UAV detection is given in Table 2.6. A detailed overview of EO UAV detection is given in this section, covering details ranging from the camera, the optics, the computer vision algorithms for automated object detection and tracking, the automatic focus and the mount used for pan-tilt motion of the whole system.

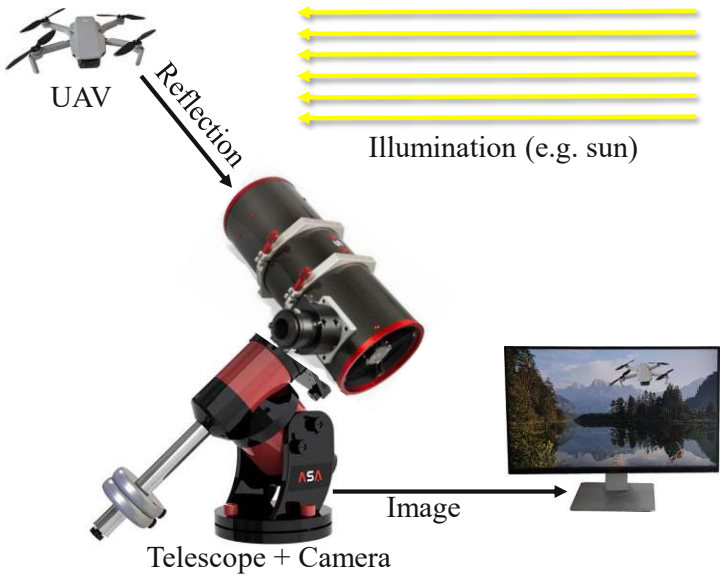


Figure 2.6: Overview of the image formation process.

2.3.1 Camera sensor

The basic principle of a camera sensor is to collect light photons, that hit the camera sensor and to encode them to form an image to be displayed. For this purpose, the camera sensor typically consists of a matrix of pixels, which capture incoming photons. The photons are converted into electrons and the percentage of electrons that originate from photons is denoted as quantum efficiency, which is also an indicator for the quality of the camera sensor [98]. The electrons are stored and collected within wells or pixels and later amplified into an analogue readable voltage [99]. Finally, an analogue digital converter (ADC) converts the information to a digital signal, which then forms the camera frame [99].

Types of camera sensors

Two main types of camera sensors exist: the charge-coupled device (CCD) [100] and the complementary metal-oxide semiconductor (CMOS) sensor [101]. Both sensors convert light in the form of photons to electrons. The fundamental difference between the two types is the readout electronics [99]. The CCD sensor reads the collected electrons from each pixel sequentially through a dedicated readout register [99]. After this phase, the electrons are converted into a voltage signal to be further converted to an analogue signal using an ADC as shown in Fig. 2.7a. This separation between pixel area and readout electronics alleviates the manufacturing process of the sensor. However, the sequential pixel readout is a bottleneck for reading the captured camera frames, which limits the achievable frame rates [99].

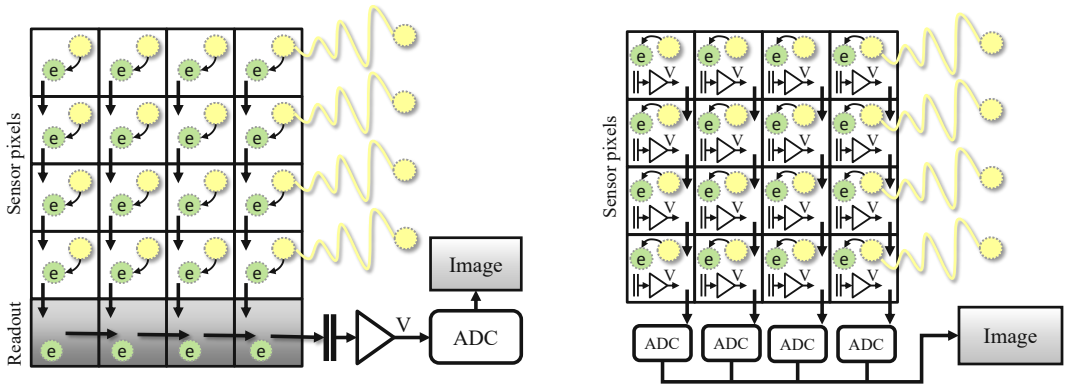
Table 2.6: Overview of the advantages and disadvantages of an EO-based UAV detection sensor.

Advantage	**Disadvantage**
Provides high-resolution images, enabling visual identification and payload analysis	Limited range of about 1 km to 2 km.
Capable of real-time monitoring	Limited during adverse weather conditions like rain or fog
Passive operation	Requires computational power for image processing
Enables visual confirmation also by human operators	Performance degrades at low-light and night, requires specialized cameras

A CMOS camera sensor contains more elaborate readout electronics [99]. Each pixel is equipped with its own capacitor and amplifier, which stores captured photons as analogue voltage [99]. Apart from the conversion per pixel, each pixel column shares one ADC, which reduces the data load per ADC significantly compared to the CCD counterpart as shown in Fig. 2.7b. The main advantage of CMOS sensors is the fast readout speed per frame, which are a consequence of the more complex readout electronics [99]. Likewise, CMOS sensors support partial readout of the camera sensor to reduce the sensor area responsible for image acquisition in order to enable higher frame rates [99]. Furthermore, compared to a CCD, the effect of blooming, which is characterized as localized overexposure within an image, can be better reduced by a CMOS sensor [99]. Compared to CCDs, however, especially front illuminated CMOS sensors have a lower sensitivity and dynamic range and therefore, inferior image quality. [99]. Depending on the desired needs, both CCD and CMOS sensors are used for the application of UAV detection and tracking [32, 102, 103].

Sensitivity, pixel size and resolution

Another key aspect when choosing a camera sensor is the sensitivity, which is characterized by the signal to noise (SNR) ratio of the camera [104]. The prevalent goal is to maximize the light or signal collecting ability of a sensor, while simultaneously minimizing noise. Regarding the camera sensor, two main attributes influence the sensor sensitivity, namely the QE and the sensor pixel size. As mentioned before, the former is the percentage of photons, which are converted into electrons within the sensor pixels. The QE is dependant on the wavelength of the incident light, but also on the type of camera sensor used as depicted in Fig. 2.8. The higher the QE, the more signal can be collected in a



(a) CCD camera sensor. (b) CMOS camera sensor.

Figure 2.7: Schematic illustration of CCD and CMOS camera sensors. The most prominent difference is the readout electronics, which is more sophisticated for the CMOS sensor.

shorter period of time.

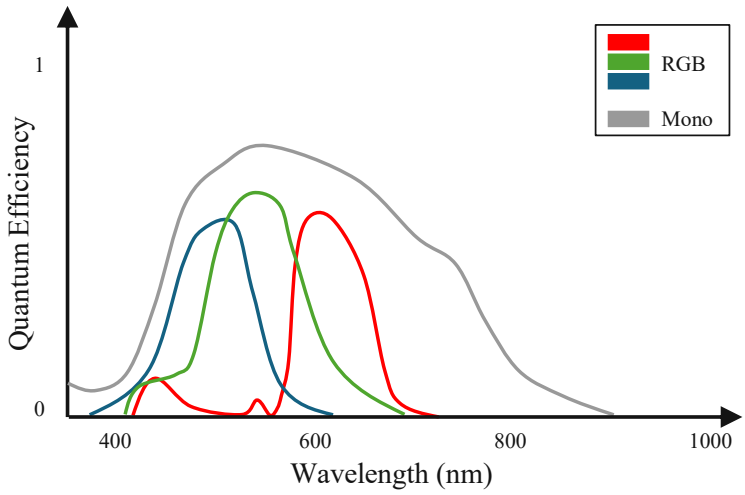


Figure 2.8: Typical spectra covered by RGB and mono camera sensors.

The second property influencing the sensor sensitivity is the pixel size. Larger pixels possess more area to collect photons [104]. A pixel size of 6 µm x 6 µm has a 9 times larger area than a pixel with the size of 2 µm x 2 µm and therefore, a 9 times better light collecting ability. Some cameras support, for example, binning, where multiple pixels are combined together to increase the camera sensitivity [105].
Types of noise, which degrade the image quality, are, for example, photon shot noise, noise during the ADC conversion into a digital signal, and dark current

noise [106]. The SNR of the imaging system can be improved by increasing the exposure time and thus, the time, the sensor pixels have to collect photons. The SNR also depends on the scene illumination, for example during daylight conditions more photons are available resulting in a higher SNR than during night time. The illumination of the scene can be increased, by a dedicated flash light, but for long range imaging applications, this remains a difficult method. Finally, approaches like frame averaging and software based de-noising algorithms can improve the SNR [107, 108].
Improving the camera sensitivity can be facilitated with a larger pixel size, however, there is a trade-off between the pixel size and the resolution. The resolution is composed of the pixel size and the magnification of the optical system, which focuses the incident light onto the camera sensor. By pairing a sensor with a small pixel size with optics of large magnification, a high resolution is achievable. However, this configuration results in a narrow FoV of the system and the small pixels impair the camera sensitivity. The theoretical limit for the achievable resolution is given by the Rayleigh criterion or the diffraction limit of the light when it passes through the optical aperture of the camera lens [109] given by

$$\theta = 1.22\frac{\lambda}{D}, \tag{2.1}$$

with λ is the wavelength of light, D is the aperture size and θ the smallest resolvable angle. Another crucial characteristic when choosing a pair of sensor and optics is the Nyquist sampling [110]. According to the theorem, to avoid aliasing, the pixel size should be at least half the size of the smallest detail the optics can resolve. Using the small angle approximation together with Eq. 2.1, the linear resolution lr on the sensor is given by

$$lr = 1.22\frac{\lambda f}{D}, \tag{2.2}$$

with f being the focal length of the optical system. Therefore, the pixel size p should be smaller than

$$p \leq 1.22\frac{\lambda f}{2D}. \tag{2.3}$$

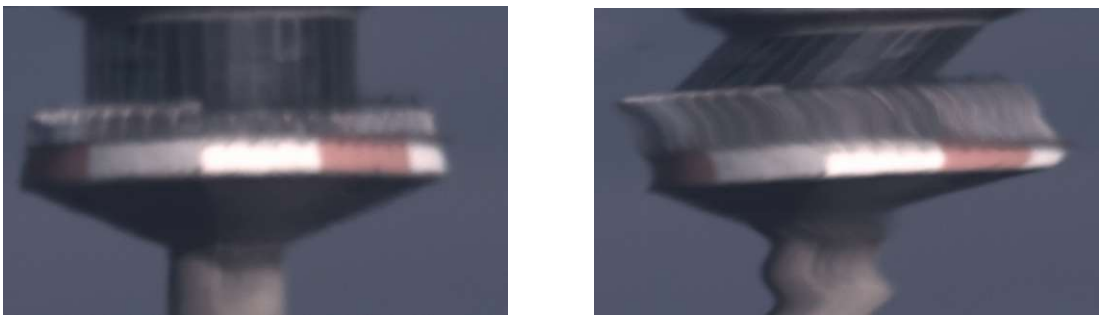
Hence, for a given optical system, a suitable camera shall be selected, which provides sufficient resolution and avoids undersampling. A telescope with an aperture size D of 0.1 m and a focal length f of 1.2 m should have pixels smaller than 3.9 μm x 3.9 μm to avoid undersampling using a wavelength λ of 530 nm, which is the peak response of a typical color camera for green light. As the telescope is not used in space, but on the surface of the earth, where atmospheric perturbations are present, the limitation of 3.9 μm per pixel can be increased.

Frame rate

For imaging applications, where it is desired to capture fast moving objects, the speed of the camera sensor is another major design aspect, as it impacts directly the maximum speed of an object passing through the FoV. The speed of the camera is denoted as frames per second (fps), which defines, how many images the camera captures per second. Modern CMOS cameras can achieve, for example, 500 fps across the full sensor [105]. Important factors influencing the frame rate are the exposure time, the camera sensor size, the bit depth of the pixels and eventually the interface between the camera and an image processing computer. Modern CMOS cameras support the selection of a reduced region of interest (ROI), which reduces the amount of sensor area to be read out to enable higher frame rates at the cost of a reduced FoV [99, 105].

Shutter

The shutter is responsible to control the time the sensor is open to collect light in order to expose and capture an image [111]. The traditional mechanical shutter uses curtains, which physically block the light from outside to reach the sensor [111]. A set of two curtains guarantees an even sensor exposure. One curtain opens the sensor by moving in one direction and the second curtain subsequently, closes the sensor by moving in the same direction [111]. The exposure time is equal to the time the sensor is open between the curtains [111].



(a) No vibrations during exposure. (b) Vibrations during exposure.

Figure 2.9: The effect of a rolling shutter when vibrations are present onto the camera system during the exposure of an image.

The electronic shutter times the exposure of the sensor by reading out the information stored within the pixel after a defined amount of time has passed [111]. As no moving parts are necessary, this method is much faster than the mechanic shutter and enables shorter exposure times, which is especially important for capturing fast moving objects in motion. However, as the read out speed of the sensor information is limited, it is performed sequentially. This introduces the

rolling shutter effect, as the lines of the sensor are read out one by one and hence, at slightly different points in time. In a highly dynamic and moving scene, this can lead to image distortion as the scenery changes within the time the individual lines of pixels are accessed and read out. Fig. 2.9 illustrates the effect of vibrations during the exposure of an image. A global shutter can mitigate this effect, whereas the whole sensor information is read out simultaneously, meaning all pixels are exposed at the same time and capture the same moment. While this approach eliminates the rolling shutter effect, it requires a more sophisticated sensor design and increases the cost of the camera [112].

Color space and spectrum

Cameras operate in different spectral bands including, for example, the visible (VIS) spectrum from 400 to 700 nm, the near infrared (NIR) between 800 nm and 2.5 µm as well as mid- or longwave (MIR) and (LIR) infrared above 2.5 µm also known as thermal [113, 114]. Fig. 2.10 summarizes the relevant spectral bands.

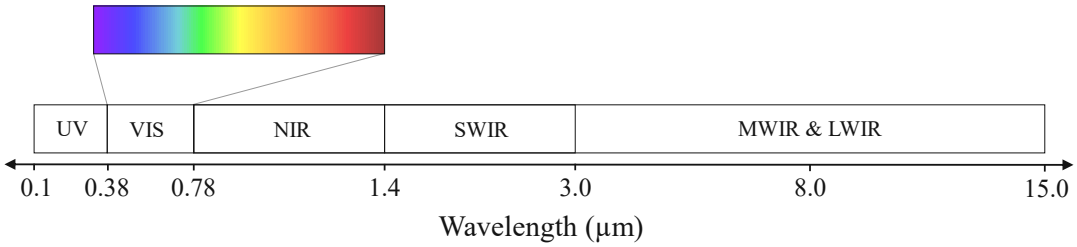


Figure 2.10: The light spectrum.

Compared to IR cameras, VIS cameras are comparably cheaper and also widely available. Moreover, VIS cameras offer higher resolution, capturing scenes in a greater detail. This is especially important, when trying to detect small objects in a long distance and potentially analyse an attached payload [115]. VIS cameras operate very well in daylight conditions, where a strong ambient illumination is present, and the spectrum resembles the human perception. Consequently, the captured images are easily interpreted by a human observer. Images can either be captured as color or greyscale. The former employs special color filters in front of the camera sensor combining essentially four pixels, typically consisting of two green, one blue and one red pixel, to one pixel [116] as depicted in Fig. 2.11. Compared to greyscale or mono cameras, which utilize all pixels as individual pixels, color cameras have therefore, a lower resolution and a lower quantum efficiency, as no additional filters are required [116]. However, by dividing the captured light into red, green and blue, more distinct information is captured to perform an algorithmic analysis of the image to detect objects. The operation of VIS cameras is limited in low light or low illumination conditions without an

active scene illumination and also during adverse weather conditions like fog.

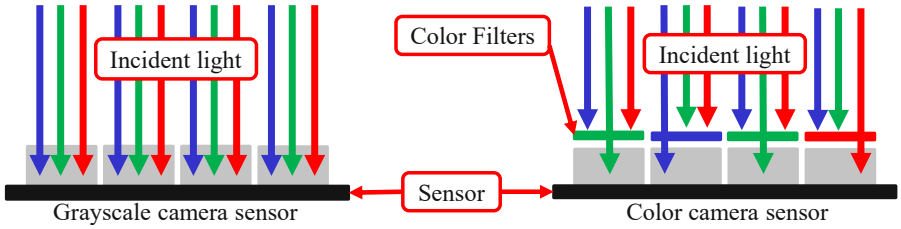


Figure 2.11: Schematic overview of mono and color camera sensors.

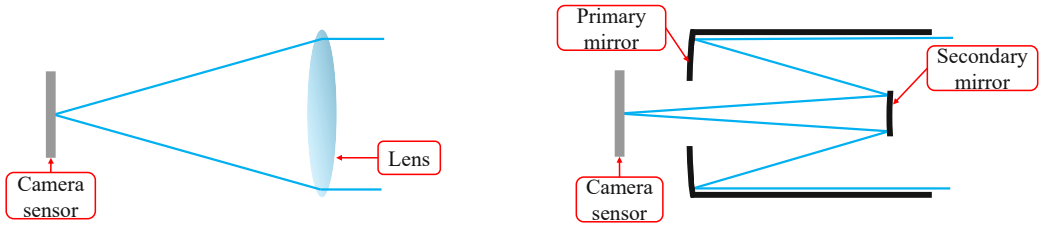
In these scenarios infrared cameras offer a solution. MIR and LIR cameras detect primarily the thermal emissions, while NIR cameras measure the reflected spectrum of the object itself [117]. For object detection under challenging lighting conditions, the effect of airglow, which is the weak emission of NIR radiation by the atmosphere, can be used for object detection [118]. Since imaging in the NIR range is much less influenced by Rayleigh scattering [119], NIR cameras offer an increased visibility in the presence of atmospheric water vapour or fog. Likewise, water has multiple absorption bands in the NIR spectrum [120], whereas many artificial materials are highly reflective, human made objects offer a potentially higher contrast.

2.3.2 Optics

The selection of an appropriate FoV is paramount for long range detection. The angular width of the FoV can be expressed as

$$FoV_w = 2\arctan(\frac{s_w}{2f}), \tag{2.4}$$

with s_w and f being the sensor width and the focal length. A wider FoV, e.g. by selecting a smaller f, allows the simultaneous observation of a broader area at the expense of a shorter operational distance, as the resolution of camera sensor is limited to a finite size. Selecting longer focal values for the optical system extends the operational range by reducing the FoV.
Conventional cameras are equipped with lenses, which consist of curved glasses to focus the light on the camera sensor. Camera lenses can be divided into two categories: prime and zoom lenses [121]. The prior type uses a fixed focal length and is more specialised for a specific purpose. Prime lenses offer larger apertures than zoom lenses in the same price category. Zoom lenses are more versatile, as the focal length is adaptable. The additional lenses within the optical system degrade the image quality and the aperture is not as large as with prime lenses.



(a) Refracting telescope. (b) Reflecting telescope.

Figure 2.12: Two telescope categories.

However, to image objects in a long distance with sufficient resolution, long focal lengths and large apertures are required, which makes lenses impractical as they get too large, heavy and expensive [122]. Furthermore, manufacturing lenses precisely at this scale is challenging.
Telescopes offer a cost efficient solution, while providing long focal lengths and larger apertures. Telescopes are divided into refracting and reflecting configurations [122]. Refracting telescopes, as depicted in Fig. 2.12a, use lenses to focus the incident light onto a camera sensor and have the same disadvantages as standard camera lenses. Reflecting telescopes employ curved mirrors to focus the light, as visible in Fig. 2.12b. In contrast to lenses, mirrors can be manufactured in a larger diameter enabling large apertures, while simultaneously being very thin and hence, light. Compared to lenses, telescopes with long focal lengths and larger apertures are less expensive and have less mass. Another drawback of lenses is the predisposition for chromatic aberrations, which occur as light passes through refracting elements [122]. Chromatic aberrations can be minimized through lens design, specialised lens elements and coatings, however, increasing the cost. Due to this reason, reflecting telescopes also support a wider spectrum of wavelengths, e.g. when utilizing the same optics with a VIS and a NIR camera.
A few common configurations for reflecting telescopes are the Newton, Cassegrain and Prime Focus telescope [122, 123]. The Newton telescope is a common configuration especially used by for amateur astronomers. It consists of a paraboloidal primary and flat secondary mirror and the simple design enables a low cost production [122]. A Cassegrain telescope consists of a concave primary and a convex secondary mirror, whereas the light is reflected from the secondary mirror through the center of the primary mirror towards the camera system. This configuration enables long focal lengths in a compact design [122]. The Schmidt Cassegrain telescope employs a spherical secondary mirror with a corrector plate to further reduce spherical aberrations [122]. A Ritchey-Chertien telescope has additional built in measures to reduce aberrations and is suited for professional astronomy [123]. A prime focus telescope omits the secondary mirror and places the camera in front of the primary mirror [124]. This configuration enables a wide FoV well suited for astronomical surveys. A variety of telescope types exist,

however, for the task of optical UAV detection compact telescopes are desired, which provide a wide FoV, but also a long focal length. A combination of Newton and a Cassegrain type telescope, or a Prime Focus and Cassegrain telescope can form a compact system covering short and long ranges.

2.3.3 Mount

Cameras equipped with appropriate optics, which enable a high magnification for a long operational range, have a narrow FoV. Consequently, a mount is required, to perform a pan and tilt motion to increase the coverage of the observed area for the system. Various approaches exits to facilitate the pan and tilt motion with a common configuration being the double gimbal system, which consists of either two direct drive or gear driven motors [125]. One possible design is the actuation of a camera by transmitting the torque generated by two stepper motors through a worm and a helical gear to the camera axes [126]. The Goddard Space Flight center uses direct driven pan-tilt cameras actuated by stepper motors to monitor external hardware tests [127]. Another option, which reduces unwanted moment of inertia and backlash, is the usage of spherical motors for the camera actuation [128].
As a long operational range is desired for UAV detection, cameras equipped with long focal lengths, for example, through the usage of telescopes, are necessary. Therefore, telescope mounts can be used for the application, as they offer the high precision positioning below single arcseconds. Most modern high performance telescope systems employ direct drive Permanent Magnet Synchronous Motors (PMSM), which provide high torque and enable high precision motion control [129].
A prerequisite to successfully track UAVs using a pan and tilt mount with a suitable motion control, is a high sampling rate of the UAV position [130]. Using for example an optical system with a focal length of 400 mm together with an 1 inch camera sensor, which has a width of 13.2 mm, to track a UAV flying at 250 km/h in distances down to 150 m, the camera would capture only 7 images of the UAV as it passes horizontally trough the FoV, if the camera is running at 100 fps. This highlights the importance of a fast computer vision pipeline, to be able to timely deliver commands to the mount to follow the UAV. The longer the distance between UAV and the optical detection system, the less strict the requirement becomes.

2.3.4 Computer vision

A camera sensor, paired with a lens or telescope, and a mount, which allows pan-tilt motion of the whole setup, are responsible to acquire an image of the target object and the surrounding area. The next necessary component is an appropriate algorithm, which can interpret the video images and detect flying

objects such as UAVs. Once the algorithm detects an object within the scenery, the pan-tilt mount is actuated to keep the object within the FoV of the camera. Algorithms used for this task are divided in two categories: object detection and tracking algorithms. The former localize objects on a per image bases, while the latter are typically initialized by a bounding box, which is either provided by a human or an object detection algorithm. Based on this initialization the object is tracked within the subsequent frames. The algorithms should detect and track UAVs not only with a high reliability, but also fast and at a high sampling rate, in order to ensure a low latency and high position update rate for the actuation of a mount to keep the UAV within the camera FoV.

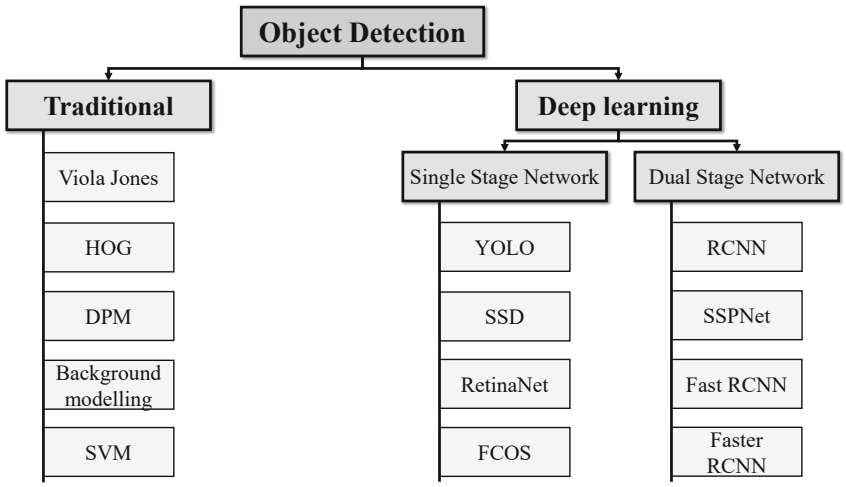


Figure 2.13: An overview of a few common object detection approaches separated into traditional and deep learning-based methods.

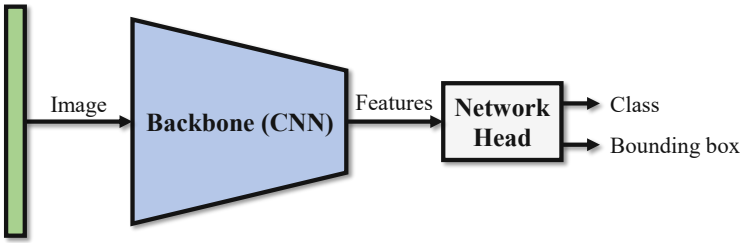
Object detection

For this thesis object detection is defined as detecting only the relevant object class within an image, which are UAVs, whereas all other similar object classes, such as birds or air planes, are treated as background. Hence, the term object detection involves implicitly a classification. Object detection within images may be carried out using a variety of approaches as depicted in Fig. 2.13. Traditional computer vision methods include sophisticated strategies to facilitate object detection. One of the earliest successful algorithms is the Viola Jones object detector, which is primarily implemented for the purpose of human face detection [131]. It employs a sliding window approach and runs an AdaBoost algorithm on Haar feature classifiers to detect human faces efficiently and at fast speed. Likewise, detection of humans is conducted by combining Histogram of Oriented Gradients (HOG) features with Support Vector Machine (SVM) classifiers [132]. The deformable parts model (DPM) is an extension of the HOG-based detector, which breaks an

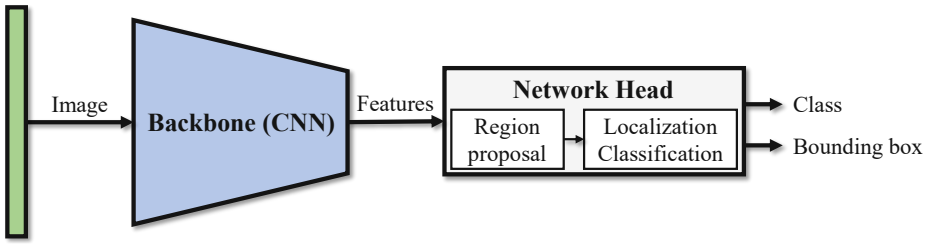
object like a human into different parts for individual detection [133]. Once the parts are detected, the whole object is integrated into a final integral detection. Other examples of traditional object detection methods involve the usage of morphological filters [134] and SVM classifiers [135] for the use case of sense and avoid applications between aircraft's and UAVs. Background modelling [136], where the goal is to separate the foreground pixels from background pixels to detect moving object within the scenery or optical flow [137], which is based on the Lucas-Kanade method [138], are further methods.
These traditional methods offer a relatively simple approach to calculate features, which are classified by e.g. a SVM. This non-complex approach is typically resource efficient and works well for simple scenes, however, in the presence of a more demanding scenario including variations in object size, illumination, occlusion or complex backgrounds, the performance of traditional methods deteriorates.

Since more than a decade deep learning-based approaches consistently outperform traditional methods and offer superior performances for object detection [139]. Convolutional neural networks (CNN) are trained on extensive datasets and learn to extract much more complex features in comparison to the traditionally handcrafted features, which can subsequently be used to form bounding boxes to localize objects and to classify them. These more sophisticated features, which capture simple patterns like corners and edges in early layers and more complex shapes in the deeper layers, enable the high accuracy of deep learning-based detectors. To learn extracting the correct features from a given image the networks are trained on a dataset. The training process requires a vast amount of annotated data and various datasets have been made publicly available for the purpose such as ImageNet [140], MS-COCO [141] or the Open Images dataset [142]. These large datasets include a variety of common classes, which are learned by the network to be detected. When training a network for a specific task, e.g. the detection of UAVs, a network can be trained from scratch on the specific dataset. Improved results can be achieved through the so called fine-tuning training approach, where algorithms are pre-trained on large publicly available general datasets such as ImageNet [140] or COCO [141] and then fine-tuned using a task specific dataset such as the Drone-vs-Bird dataset [143].

Region-based Convolutional Neural Network (RCNN) is one of the earlier successful dual-stage CNNs, which uses object proposals through selective search [144] and a CNN for feature extraction combined with SVM classifier [145] for object detection. Fig.2.14b illustrates the dual stage network architecture, whereas classification and localization are performed after a dedicated region proposal step. The speed of RCNN is limited, as it calculates features for all object proposals. Spatial pyramid pooling (SPPnet) network improves this inference speed, by calculating a single feature map for an image and extracting sub features for each proposal from the feature map [146]. Fast-RCNN combines the benefits from RCNN and SPPnet and calculates within a single CNN pass the features for the input image



(a) Single stage neural network.



(b) Dual stage neural network.

Figure 2.14: Overview of different network architectures. The backbone is used to extract features and can be implemented using a CNN. a) single stage neural networks perform classification and detection simultaneously within the network head. b) Dual stage networks perform classification and localization after a dedicated region proposal step.

and the associated region proposals. These are subsequently fed to a softmax classifier and a bounding box regressor [147]. Faster-RCNN (FRCNN) [148] adds a separate region proposal network (RPN), which shares convolutional layers with the detection network. The former generates region proposals, which may hold potential objects to be classified within the subsequent layers. This unified framework makes training simpler and more efficient. Furthermore, FRCNN used predefined anchor boxes to generate region proposals and Region of Interest Pooling for an efficient processing of the proposed regions. While offering high precision, dual stage networks require more computational resources, making them less suitable for real-time applications.

Single Shot Multibox detector (SSD) and You Only Look Once (YOLO) are, unlike FRCNN, single-stage object detectors, which do not use an RPN network for region proposals [149, 150]. Fig. 2.14 illustrates the difference between a single and a dual-stage neural network architecture. The single stage network performs classification and localization simultaneously within the network head. SSD and YOLO use default bounding boxes with various aspect ratios, which are then refined during inference to localize objects [149, 150]. As single-stage networks, SSD and YOLO are efficient and offer faster inference speeds at the cost of detection accuracy [149, 150].

Like SSD and YOLO, RetinaNet is a single-stage architecture, which introduces

Focal Loss to address the issue of class imbalance within object detection databases by down-weighing well-classified examples to train more on hard examples [151]. RetinaNet achieves improved detection results for the single-stage network architectures.
Fully Convolutional One-Stage (FCOS) is another single-stage fully convolutional neural network, which introduces the concept of center-ness to refine the localizations and does not rely on additional hyper-parameters like predefined anchor boxes [152].
Due to the success of deep learning-based methods, many studies are conducted using neural networks for the task of object detection [32, 153–157]. However, as a consequence of the large number of calculations, neural networks have high processing demands, which are usually handled through the usage of Graphics Processing Units (GPU)s to enable real-time performance [158].
For the success of an EO UAV detection and tracking system an algorithm is required, which is capable of automatically recognizing a UAV and monitoring its position over time. A suitable object detection algorithm can achieve both of these tasks. The state of the art shows, that deep learning algorithms prove to be the most accurate algorithms, however, imposing some additional hardware requirements due to the computational complexity of the involved models.

Object tracking

As opposed to object detection, where objects are detected individually for each frame, object tracking continuously follows an object over multiple video frames. As a first step, trackers need an initialization, which can be conducted by manually selecting a bounding box for the tracker to follow or by obtaining it through an object detector. The object tracking algorithm initializes itself based on this first information and learns features such as the size, position and appearance of the object. Utilizing this initialization, the tracker is capable to predict or find the object location in subsequent frames, constituting a track of the object. Object trackers can be categorised into short and long term trackers [159]. Upon loosing an object track, a short term tracker is not able to re-detect the object to continue tracking, but requires a new initialisation. Contrary, long term trackers are not re-initialized externally once a track is lost, but rather perform re-detection on their own.
Minimum Output Sum of Squared Error (MOSSE) [160] and Kernelized Correlation Filter (KCF) [161] are examples of short term object trackers, which match extracted features to an online learned model using correlation filters to track objects at high frame rates. Likewise, Channel and Spacial Reliability Tracker (CSRT) [162] uses correlation filters within the Fourier domain to improve the tracking accuracy at the cost of a lower inference speed. The MedianFlow [163] tracker is another well known high-speed tracker, which estimates an object position based on the optical flow trajectories. The tracking-learning-detection (TLD)

tracker combines, as the name suggest, three components to form a long term tracker. The tracker is initialized and is based on a MedianFlow tracker to follow an object. A learning component uses the initial information as well as subsequent frames to learn a dedicated detector online, which is then capable to detect the object in subsequent frames and correct the tracker if necessary [164].
Similarly to object detection, tracking is a vigorously researched field and in the last years an ever increasing trend of the usage of neural networks and CNNs is observed to solve this task [159, 165, 166]. The accurate tracking by overlap maximization (ATOM) tracker is well known tracker, which utilizes neural networks. Tracking is separated into a target estimation and classification phase, whereas, the latter distinctly separates the object from the background and the estimation locates the object with a bounding box [167]. Replacing the HoG features of correlation-based trackers by convolutional features can also increase the performance of and object tracker [168]. Other trackers, which use deep learning, facilitate tracking, for example, through bounding box regression for the prediction of the object position within the subsequent image [169] or by learning similarity functions using Siamese neural networks [170]. While achieving a high accuracy, deep learning-based trackers require more computational power and therefore, achieve lower frame rates compared to traditional trackers.
For EO UAV detection and tracking using a steerable mount, capturing the UAV position accurately at a high sampling rate is important to be able to keep the UAV within the FoV of the system even in the presence of illumination variations, background clutter, scale and speed variations. Deep learning-based trackers are capable of accurately tracking objects, however, require additional hardware resources, especially as they are combined with an object detection algorithm for initialization. Traditional object trackers can provide a relatively high accuracy over a short period of time, while being more hardware resource efficient.

Parallel detection and tracking

Deep learning-based object detectors typically provide the superior results, when examining the provided bounding boxes, however, these algorithms require a multitude of calculations, which are usually accelerated using GPUs. Nonetheless, this high demand for computational power has led to approaches combining a deep learning object detector with, for example, object tracking or estimation algorithms in parallel, which require much less computational power, to achieve higher frame rates. In an application like camera-based UAV detection, a narrow FoV enables long operational distances. As a consequence, the mount used for realigning the camera to keep the UAV within the FoV has to be highly dynamic. To achieve this, a high bandwidth is necessary, which requires a high frequency of position commands and hence, a high processing rate of the computer vision pipeline [130].
As mentioned, the utilization of estimation algorithms parallel to a deep learning

detection algorithm improves the achievable frame rates, for instance implemented in the SORT framework [171], or by combining YOLOv2 with a Kalman filter [172]. Also background subtraction algorithms are combined with Kalman filters to facilitate and ameliorate the tracking performance [173, 174]. Similarly for moving cameras in sense-and-avoid (SSA) applications, warped difference images with morphological filters and the usage of Particle filters improves the tracking performance [175].
Combinations with estimation algorithms enhance performance and achievable localization rates, however, estimators do not track the actual position, but, as the name suggests, estimate the position between consecutive detections. A dedicated tracking algorithm running in parallel to a detector localizes the object based on an internally learned model. The TLD tracker, mentioned in Section 2.3.4, combines tracking, learning and detection to improve the overall tracker performance [164]. However, the used detector only learns online on the image used for initialization and the subsequent tracked frames. Parallel execution of YOLOv3 [176] and SiamRPN [170] as an object tracker, enables 48 fps on a workstation equipped with an Intel i7-6800k CPU and a NVIDIA Geforce GTX 1080Ti GPU [177]. A similar concept benefits from the high accuracy of Siamese networks to verify upon request, whether a tracker is still tracking an object correctly or needs reinitialisation [178]. However, the majority of these methods relies either on detector or tracker, disregarding the possibility of a dedicated transition strategy, which benefits of a collaboration between the two.

Ground truth

For training and evaluation of computer vision algorithms, datasets are required, which contain an underlying ground truth. The ground truth denotes what is to be learned, detected or classified. In the scenario of object detection, this can be facilitated in various ways. The object of interest within an image may be either marked by a point, for example, representing the object center or top [179]. A more common approach is to mark the object with an encompassing bounding box [179], which is typically saved in the form of bounding box coordinates denoting the position and size within the image, either in pixel or relative to the image size [150]. Finally, the most tedious approach is to mark each individual pixel, which belongs to the object. While this is the most arduous and time-consuming approach to label a dataset, it enables algorithms to learn and precisely detect the parts of the image belonging to an object and to perform image segmentation [180].
Finally, apart from denoting the location of the object within an image, an additional class label allows for an algorithm to learn and differentiate between multiple classes of objects [179].

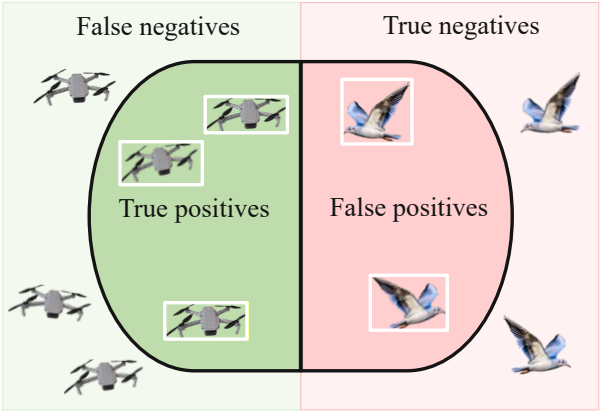


Figure 2.15: Illustration of the confusion matrix trying to detect all UAVs within the image.

Evaluation metrics

For object detection in images the most common metric used for evaluation is the mean average precision (mAP). This metric is derived from the precision [181]

$$Pr = \frac{TP}{TP + FP}, \tag{2.5}$$

and the recall [181],

$$Re = \frac{TP}{TP + FN}, \tag{2.6}$$

with TP denoting true positives, FP denoting false positives and FN denoting false negatives. Fig. 2.15 illustrates the confusion matrix for the case of UAV detection. The precision gives the percentage of correct predictions the algorithm makes out of all predictions, whereas recall represents the ratio between correct predictions out of all possible objects to find. A prediction is deemed to be correct and hence, a TP, if the intersection over union (IoU) of the predicted and the ground truth bounding box surpass a certain threshold, most commonly 0.5. The average precision [182]

$$AP = \int_{x=0}^{1} Pr(Re(x))dx, \tag{2.7}$$

is the area under the precision and recall curve, whereas x denotes different recall levels. Finally, the mean average precision (mAP) is given by [183]

$$mAP = \frac{1}{C} \sum_{i=1}^{C} AP_i, \tag{2.8}$$

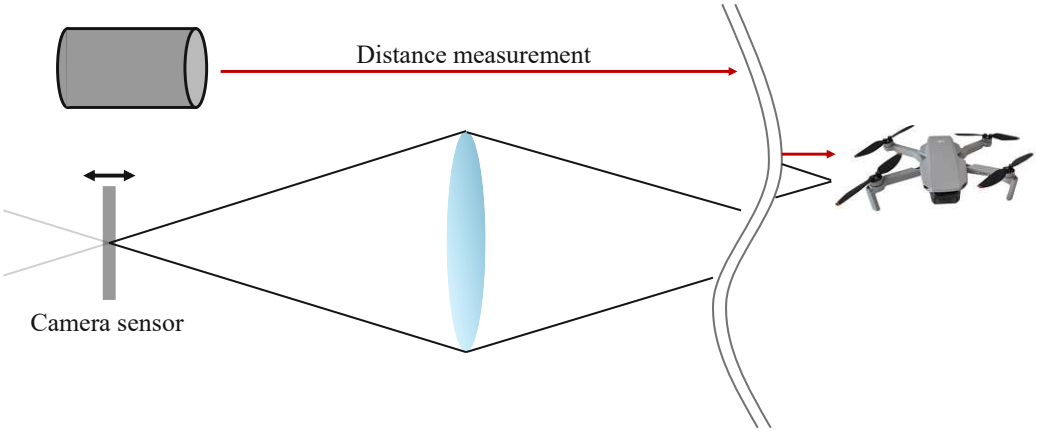
where C is the number of object classes. The AP and mAP are very meaningful evaluation tools, as they combine precision and recall over various IoU thresholds. For the evaluation the threshold may vary and be either evaluated for a fixed IoU threshold of, for example, 0.5 to form the mAP(0.5) or be the sum of all mAPs between the IoU thresholds of 0.5 to 0.95, which is denoted as mAP(0.5:0.95).
The center location offset (CLO) is another prevalent metric, which calculates the Euclidean distance between the center of the predicted and the ground truth bounding box [184]. The quality of a detection or track is quantified by the percentage of frames with a CLO at a certain threshold distance. The CLO however, disregards the proportion of the predicted bounding box with respect to the ground truth. The overlap score [182] or the accuracy [185] integrates this notion by incorporating the IoU for the evaluation. Similarly to Eq. 2.5 and 2.6, a tracking precision and recall can be defined and further, the tracking F-score

$$F(\tau) = \frac{2Pr(\tau)Re(\tau)}{Pr(\tau) + Re(\tau)}, \tag{2.9}$$

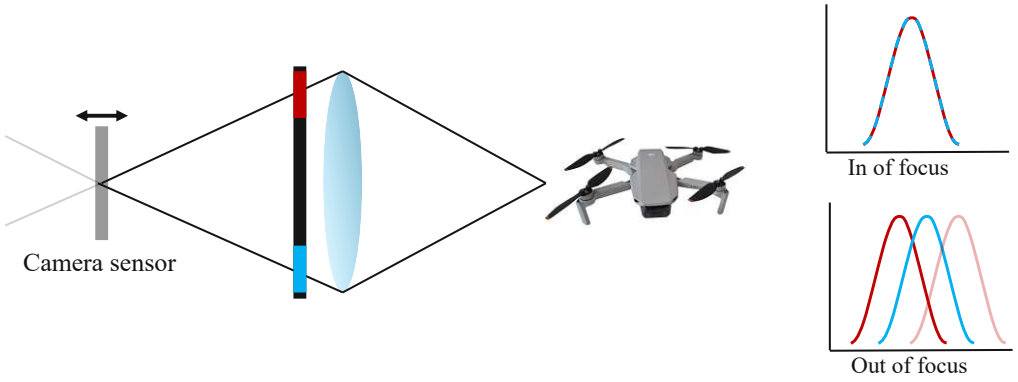
with τ being the prediction threshold, below which the detector or tracker output is ignored [186]. This metric represents the harmonic mean between precision and recall.
The robustness is a measure, which records how often a tracker loses an object during a track, whereas a failure is reached, once a minimum IoU threshold is surpassed [185]. By alternating the time point within the test video sequence or also the quality of initial bounding box, which may not be perfectly aligned with the ground truth, the temporal and spatial robustness are measured [187].
Another important property is the speed of the computer vision algorithms, which is measured by the achievable frame rate or the processing time between receiving a frame and outputting a detection or track. For the evaluation, typically a relative algorithm comparison is presented for a certain workstation to compare achievable frame rates [188]. To lessen the effect of the utilized hardware, the equivalent filter operations (EFO) quantify the algorithm speed [185].

2.3.5 Automatic focus

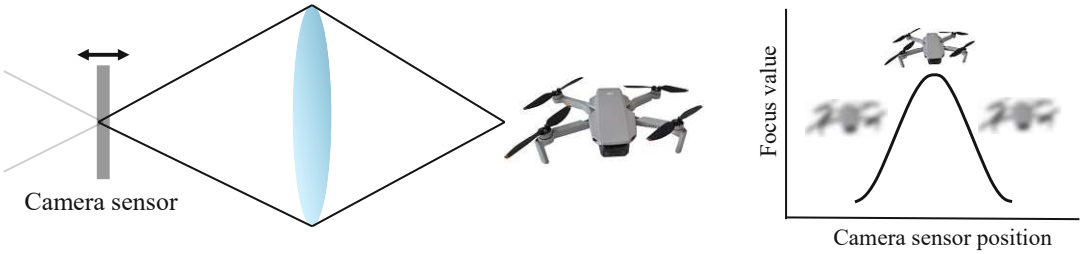
Long focal lengths paired with large apertures are necessary for the optical system to provide sufficient resolution to discern small objects over long distances. These optical requirements impose challenges onto the focus module of the optical system, as long focal lengths and large apertures lead to a narrow depth of field (DoF). For detection and tracking using computer vision, but also for human cross validation of the captured images, it is desired that the object of interest is imaged sharply.



(a) Active automatic focus.



(b) Phase-detect automatic focus.



(c) Contrast-based automatic focus.

Figure 2.16: Overview of various focus methods. a) shows the active focus, which determines the distance to the object though an additional active measurement. b) depicts the passive phase detect focus, which splits some light to additional sensors, to determine the necessary sensor position. c) the contrast-based focus, which adjusts the focus based on the contrast within the image.

Therefore, a highly performant automatic focus is necessary.
To find the optimal focus position two strategies can be employed: active and passive. The former requires a measurement of the distance to the object, for example, via time of flight (ToF) or ultrasonic sensors [189, 190], as depicted schematically in Fig. 2.16a. Once the distance is obtained, the exact position of the camera sensor with respect to the optics to obtain a focused image is determined via an optical model. Albeit being a very precise method, it requires additional hardware components and measurements to facilitate focusing capabilities.
In contrast to active measurement methods, a passive focus acquires all information necessary to find the optimal camera sensor position passively through the incident light. Phase detection is a frequently used approach, which splits the incoming light rays into two separate beams and through a comparison of the phase difference, the focus distance and direction is obtained [191]. The concept is illustrated in Fig. 2.16b. This enables a very precise and fast automatic focus, however, additional hardware, in the form of sensors and mirrors, and a precise alignment of the optical components are required.
A passive contrast-based auto focus does not depend on complex additional hardware components and determines the focus position solely using the incident light. For this purpose, the contrast within two camera images is calculated and compared, whereas the axial position of the camera with respect to the optics is adapted between capturing the two images as seen in Fig. 2.16c. Continuing this comparison and moving the camera in the direction, which maximizes the contrast, allows to iteratively find the best focus position [192]. Alternatively, using telescope systems, the camera position may remain steady, while the telescope mirrors are moved to adjust the focus [193]. Whereas the passive contrast-based focus is slower and less accurate than the aforementioned approaches, it is non-invasive, as no additional hardware is necessary and no light is diverted away from the camera sensor to additional focus sensors.
The contrast-based method requires firstly the calculation of the current focus measure or the contrast within the image and secondly an optimization algorithm, which moves the camera efficiently in the direction to maximize the contrast value and thus find the optimal focus. The contrast function is categorized into five groups: Gradient-, Laplace-, Wavelet-, Statistics- and Discrete Cosine Transform-based methods [194]. Gradient-based operators assume, that focused images contain more sharp edges than blurry images and therefore, calculate the focus measure through gradient calculations. The Tenengrad operator [195] and the Normalized Variance [196] are prominent examples for this group. The former has been used for telescope systems, when focusing onto terrestrial objects [197], while the latter achieves good results for astronomical observations [198]. Laplacian-based measures operate similarly by examining the sharpness of the image edges using the second derivative [194]. Focus measures based on the Variance of Laplacian prove efficient, when focusing onto small image patches [199]. Likewise, it is shown that the Discrete Wavelet Transform works well for small objects [200].

It is categorised in the group of Wavelet-based operators, which calculate the focus using discrete wavelet transform to characterise the spatial and frequency features of images [194]. The Discrete Cosine Transform operates on the same principle, but uses the discrete cosine transform [194]. Statistics-based methods exploit several image statistics such as texture descriptors to calculate the focus measure [194].
Once a focus or contrast value is calculated, an optimization algorithm moves the camera in the direction, which maximizes the focus value or contrast. The focus value, when plotted over the stage position, resembles a prominent peak with decreasing slopes at both sides, when moving away from the optimal focus position as visible in Fig. 2.16c. A straightforward way to find the peak is the Hill Climb algorithm [201], which compares two camera images captured at two different physical positions and then proceeds to move the camera in the direction of higher contrast at a fixed step size. Once the peak of the contrast value starts to decrease again, the camera is moved in the reversed direction with a smaller step size. Other approaches are the Fibonacci or Binary algorithms [192, 202], which move the camera with respect to the optics in a Fibonacci or Binary pattern. Modelling the curve and estimating the position on the curve, by applying curve fitting [203] or Gaussian fitting [204] is another approach, which reduces the number of required measurements.
Without having a precise active measurement of the distance between the UAV and the optical system, a passive automatic focus approach provides a solution, which enables capturing sharp images of a target. A radar system can actively provide a distance measurement, however, especially in close optical detection ranges, a precise focus is paramount. Moreover, the contrast-based passive focus offers an approach, which facilitates a focus with minimal additional hardware requirements. As the literature shows, various algorithms for the contrast measurement and the subsequent focus optimization exist, which can be applied for the task of focusing onto fast flying UAVs.

2.4 Multispectral detection

Each one of the presented sensor technologies and approaches offers individual advantages and disadvantages as summarized in Tables 2.2 to 2.6. In practice, a UAV detection and tracking system combines several sensors to form a multispectral system to account for all contingencies. An exemplary use case is an unidentified target detected by a radar system in a significant distance. As mentioned in Section 2.2.2, classification of small objects is challenging using radar information alone. An RF system could classify the UAV, however, if the UAV is flying autonomously without a dedicated communication link to an operator, it remains invisible to the RF system. Therefore, after this coarse radar localization, the information is used to find the target with the optical component

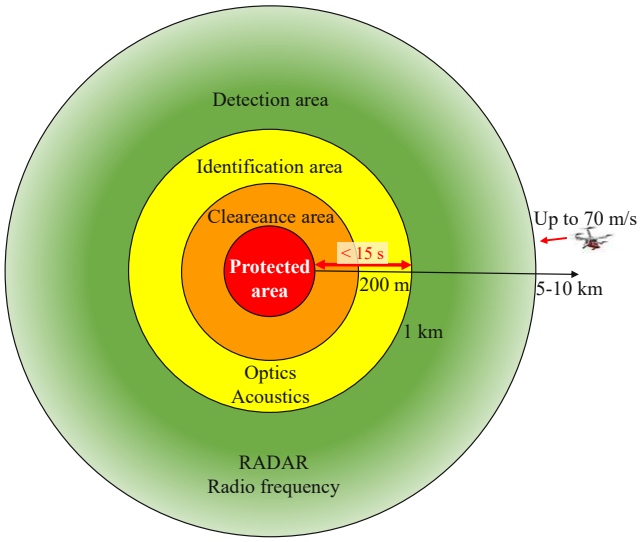


Figure 2.17: Example distances achieved by a multispectral state of the art UAV detection system. Radar and radio frequency can achieve long detection distances, however, if the latter one fails to perform a classification or identification, current acoustic and optical systems are limited to rather short ranges leaving a short time window for the preparation of counter measures.

of the multispectral system. The optical sensor has a narrow FoV to visually inspect the target over medium ranges up to 2 km [34]. The radar information is necessary to align the optical sensor with the object to capture images for classification and identification. If optical classification fails, the acoustic signals over very limited ranges of a couple of hundred meters can be used for identification. Figure 2.17 shows an exemplary multispectral system with the respective operation distances for each sensor technology. Ctrl+sky [33] and Ihtar [205] are two multispectral systems that combine several sensors for UAV detection. Both systems combine radar, optics and acoustics and dedicated jamming capabilities into a multi-sensoral detection system, while Ctrl+sky also adds RF detection sensors.

The main limitation of the state of the art is the relatively short operational range offered by current optical systems. As stated in Section 2.3, optical sensors offer the most conclusive information for classification, as visual material can easily be cross checked and understood by a human operator. Therefore, it is paramount to increase the optical detection distance in order to gain more time to prepare for appropriate countermeasures.

2.5 Research objectives and questions

The state of the art shows numerous UAV detection technologies, with optical sensors beeing a key component of any efficient UAV detection system. Current optical UAV detection sensors are mainly limited by the detection range for small sized objects, which is crucial for early threat assessment. In the following, open questions are addressed, which are investigated during the thesis in order to enable robust long range optical UAV detection and tracking.

Although many attempts have been made to create an optical long distance UAV detection system [32, 206, 207], the achievable ranges are limited to about 2 km for the detection of small objects such as micro and mini UAVs [34]. In order to substantially increase the optical detection range, long focal lengths and large apertures are necessary to enable high resolution imaging. This leads to the first research question:

Research Question 1:
Is it feasible to detect mini and micro UAVs with a diameter of 0.3 m in a distance of more than 2000 m with an optical UAV detection system?

While telescopes enable the necessary long focal lengths and large apertures for the application, a paramount design parameter to select a reasonable telescope aperture size is the knowledge of the required resolution to ensure reliable object detection using state of the art deep learning-based object detection algorithms. This leads to the second research question:

Research Question 2:
What optical resolution is required for a deep learning-based optical system to detect micro and mini UAVs of the size 0.3 m in diameter reliably?

The utilization of long focal lengths limits the FoV of the system and imposes challenging requirements on the computer vision pipeline, which is responsible for localizing a UAV within the camera frames. Fast image processing is necessary to be capable to localize a UAV crossing the telescope FoV at potentially fast speeds and a high localization rate of at least 100 fps is desirable for an underlying motion control of a pan and tilt device to follow the highly dynamic trajectories of UAVs [130]. The state of the art shows that parallel operation of detection and tracking algorithms can achieve high frame rates, however, a paramount feature during operation is the transition strategy between detector and tracker, hence, when the pan and tilt command of the mount should follow the detector or the tracker output. Therefore, the following question arises:

Research Question 3:
Which software architecture and transition strategy enable an efficient image-based detection and tracking of UAVs at frame rates of 100 fps?

Once a suitable architecture and transition strategy between detector and tracker are implemented, the question arises, which combination of algorithms achieve the best results for detection and tracking at high frame rates. Therefore, the following question is formulated:

Research Question 4:
Does a combination of a deep learning-based object detector and a traditional tracker exist, that enables a reliable detection and tracking of UAVs during daytime conditions?

Closely related to the analysis of different computer vision algorithms for an efficient UAV localization, a varying performance is expected for the usage of different color spaces of the input images for the detection using deep learning-based algorithms. This leads to the following research question:

Research Question 5:
Is the detection performance of deep learning-based detection algorithms maximized by using color or greyscale images as input during daytime conditions?

The usage of long focal lengths paired with large apertures imposes challenging requirements on the focus of the optical detection system. The focus is a paramount feature especially for long range applications, as out of focus objects might by completely invisible to the camera. Therefore, a fast and automatic focus module is desired, which is capable of keeping the UAV within focus, delivering a sharp image for the computer vision-based detection and tracking. The next question addresses this topic:

Research Question 6:
Is it feasible to design a high-speed automatic focus module for a telescope-based UAV detection system to keep a mini or micro UAV, flying up to its maximum speed, in focus?

Using a monocular telescope-based approach, a distance measurement or estimation to a tracked object is difficult. A precise measurement is typically facilitated via a dedicated laser ranging system. However, enabling real-time distance measurement of potentially multiple targets remains a challenge. This leads to the final research question:

Research Question 7:
Can a laser range finder be integrated to a telescope-based UAV detection system to enable real-time distance measurement of multiple UAVs?

CHAPTER 3

Contribution & Publications

This chapter gives an overview of the publications, which have been created during the thesis. In the following the references are listed:

[P1] D. Ojdanić, A. Sinn, C. Naverschnigg and G. Schitter, "Feasibility Analysis of Optical UAV Detection Over Long Distances Using Robotic Telescopes," in IEEE Transactions on Aerospace and Electronic Systems, vol. 59, no. 5, pp. 5148-5157, Oct. 2023
Journal Quartile: Q1 (2023)

[P2] D. Ojdanić, N. Paternoster, C. Naverschnigg, A. Sinn, and G. Schitter, "Evaluation of the required optical resolution for deep learning-based long-range UAV detection," In Pattern Recognition and Tracking XXXV, vol. 13040, pp. 78-84. SPIE, 2024

[P3] D. Ojdanić, C. Naverschnigg, A. Sinn, D. Zelinskyi, and G. Schitter, "Parallel Architecture for Low Latency UAV Detection and Tracking using Robotic Telescopes," in IEEE Transactions on Aerospace and Electronic Systems, vol. 60, no. 4, pp. 5515-5524, Aug. 2024
Journal Quartile: Q1 (2024)

[P4] D. Ojdanić, C. Naverschnigg, A. Sinn, and G. Schitter, "Algorithm evaluation for parallel detection and tracking of UAVs," In Optics, Photonics, and Digital Technologies for Imaging Applications VIII, vol. 12998, pp. 327-333. SPIE, 2024.

[P5] D. Ojdanić, C. Naverschnigg, A. Sinn, and G. Schitter, "Deep learning-based long-distance optical UAV detection: color versus grayscale," In Pattern Recognition and Tracking XXXIV, vol. 12527, pp. 80-84. SPIE, 2023

[P6] D. Ojdanić, D. Zelinskyi, C. Naverschnigg, A. Sinn, and G. Schitter, "High-speed telescope autofocus for UAV detection and tracking," Opt. Express 32, 7147-7157 (2024)
Journal Quartile: Q1 (2024)

[P7] D. Ojdanić, B. Gräf, A. Sinn, H. W. Yoo, and G. Schitter, "Camera-guided real-time laser ranging for multi-UAV distance measurement," Appl. Opt. 61, 9233-9240 (2022)
Journal Quartile: Q2 (2022)

The first six publications within this thesis deal with the analysis, implementation and evaluation of a telescope-based UAV detection and tracking system including all major components, such as the software architecture, necessary algorithms and the automatic focus [P1 - P6]. A further publication investigates an additional prototype of a possible optical design to integrate a laser range finder into a telescope system to measure the distance to multiple UAVs within the FoV in real-time.

Long range optical UAV detection is a paramount feature of any multispectral UAV detection system. Therefore, a feasibility analysis for a telescope-based UAV detection system, which can significantly extend the optical detection range, is conducted [P1]. Estimating the achievable resolution using the Rayleigh criterion [109], while taking into account the atmospheric disturbances, a trade off between system weight and required aperture size is found at 0.2 m to 0.3 m. Investigating further system parameters such as the FoV and the camera resolution leads to the conclusion that the best approach to cover a long operational range from 150 m up to 4000 m is a dual telescope approach. A telescope with a wide FoV for close to medium distances shall be paired with a telescope with along focal length for medium to long distances.
Another crucial aspect is the minimal required resolution in pixels for a deep learning-based object detector to reliably detect UAVs. For this purpose, a modified YOLOv4 architecture is trained on a custom UAV dataset and a mAP(0.5) of 86 % is achievable for UAVs down to a size of 15 x 15 pixels. Derived from this analysis, for the further work in this thesis an f/10 Meade Schmidt Cassegrain telescope with a focal length of 2540 mm and an f/1.3 ASA UWF300 telescope with a focal length of 390 mm are used to cover the whole optical operational range.

In an additional study, experiments are conducted to correlate the resolution of the optical system to the prediction probability of a deep learning object detector [P2]. For this purpose, FRCNN, pre-trained on the COCO dataset [141], is fine-tuned on a custom UAV dataset. Furthermore, a modified version of the USAF resolution chart is used to determine the resolution of the optical system. The modified resolution chart, which contains AruCo markers, enables

an automated evaluation of the resolution. Using a telescope with a focal length of 1325 mm and an aperture size of 102 mm together with the resolution chart and a UAV of 289 mm in diameter, a test dataset is created, which consists of images at various distances and backgrounds. Based on this test dataset, the resolution is correlated to the mAP(0.5) of FRCNN detecting a UAV in front of a clear and complex background. The results confirm the analysis presented in [P1].

The narrow FoV necessary to resolve objects over long distances imposes challenging requirements on the computer vision pipeline, which is responsible to localize a UAV within video images. A localization has to happen with low latency and at a high sampling rate to enable a suitable control strategy to actuate the telescope mount in order to follow the highly dynamic UAV trajectories. For this purpose a parallel architecture is developed, which combines a slow, but accurate deep learning-based object detector with a fast, but less accurate traditional object tracker [P3]. A collaboration between the two algorithms is facilitated via the prediction probability and the tracker reliably. This architecture enables detection and tracking of UAVs at 100 fps. Using the proposed architecture the probability of detecting a UAV twice, which is necessary to perform a trajectory prediction, is 78 % if the UAV is visible for 162 ms within the FoV compared to 29 % for a non-parallel approach. Additionally, extensive field tests have been performed with the telescope system, demonstrating detection capabilities of UAVs, which are 0.3 m in diameter, of up to 4000 m.

To find the best performing combinations of algorithms for the parallel architecture, an evaluation of different detection and tracking algorithms is necessary. Therefore, four state of the art deep learning object detectors and three traditional object trackers are investigated [P4]. The detectors are all pre-trained on the COCO dataset [141] and subsequently fine-tuned on a custom UAV dataset, which is used throughout this thesis. The dataset consists of 20500 images for training and another 25000 video images for testing, respectively. The comparison shows, that FRCNN combined with a MedianFlow tracker offer the best performance in terms of intersection over union and center location offset.

In order to maximize the detection performance of the object detector, a further evaluation is conducted, investigating whether color or greyscale images ensure an improved accuracy during daytime conditions [P5]. For this purpose, two models each are trained for different deep learning object detectors, namely color and greyscale models. For the latter, the UAV dataset is converted to greyscale images before the training process. This step emulates the usage of two identical cameras in terms of resolution and quantum efficiency. Across all detectors tested, the color models outperform the greyscale models indicating that the object detector achieves a higher performance, when working with RGB images due to the increased amount of information encoded within the image

during daytime conditions.

The consequences of using telescopes with long focal lengths and large apertures to ensure sufficient resolution for long range applications are challenging requirements on the automatic focus, which is responsible for maintaining a sharp image of a UAV. A passive contrast-based automatic focus is implemented using a stepper motor-based linear stage [P6]. A comparison of different contrast functions and search algorithms shows that the Tenengrad operator together with the Hill Climb search algorithm are sufficiently fast and accurate to ensure focus. Extensive field tests demonstrate the focus capability in ranges between 4500 m down to 150 m keeping UAVs in focus flying at speeds up to 24 m/s.

To enable a precise localization of UAVs in the 3D space, a prototype is implemented, which integrates a laser range finder into a telescope [P7]. The telescope and camera use computer vision to track UAVs. By exploiting the angular information obtained through the bounding boxes, the laser is steered with a fast steering mirror (FSM). An alignment between the laser and the angular information extracted from the computer vision algorithm is possible, as the laser is integrated into the optical path of the camera via a dicrotic mirror. This integration enables to point the laser towards a tracked UAV precisely, without the need to scan the whole FoV. Experiments in a laboratory environment show that the distance to four UAVs can be measured at a rate of 14 Hz per object, enabling a precise and fast 3D object localization.

Publication [P1]

Feasibility Analysis of Optical UAV Detection Over Long Distances Using Robotic Telescopes

Authors: D. Ojdanić, A. Sinn, C. Naverschnigg and G. Schitter

Reference:

[P1] D. Ojdanić, A. Sinn, C. Naverschnigg and G. Schitter, "Feasibility Analysis of Optical UAV Detection Over Long Distances Using Robotic Telescopes," in IEEE Transactions on Aerospace and Electronic Systems, vol. 59, no. 5, pp. 5148-5157, Oct. 2023

Journal Quartile: Q1 (2023)

Feasibility Analysis of Optical UAV Detection Over Long Distances Using Robotic Telescopes

Denis Ojdanić, Andreas Sinn, Christopher Naverschnigg, and Georg Schitter

***Abstract*—Substantial technological development has made Unmanned Aerial Vehicles (UAVs) more versatile, cheaper and accessible to the public in recent years. Alongside many positive effects and use cases, safety concerns are increasing as a plethora of incidents demonstrate the destructive potential of UAVs. To counteract this development and thus protect people and critical infrastructure, UAV detection, tracking and defence gains more and more research attention. Whereas many drone detection technologies are available and usually deployed in multi-spectral systems, optical detection and imaging of approaching objects are a key contribution to well-founded situational awareness. As reaction time is a crucial parameter for successful UAV defence, the operating distance of the optical detection system needs to be improved further. This paper presents the analysis, development and evaluation of a telescope-based UAV detection system. The system consists of a high precision mount and a telescope equipped with a camera. UAVs are detected in the captured video frames by the deep learning algorithm YOLOv4 using a modified architecture. The proposed system allows a significant increase in the optical detection range to more than 3 km under daylight conditions, extending the reaction time for counter UAV systems.**

***Index Terms*—UAV detection, long distance detection, telescopes, deep learning**

I. Introduction

OVER the past decade advances in technology led to a massive grow in popularity of unmanned aerial vehicles both in professional and private sector [1]. Easy public accessibility is accompanied by an increased risk of potential misuse, ranging from relatively harmless violations to severe hazardous situations. Extensive research is conducted on possible future threat scenarios and an alarming number of real live incidents has already occurred [2], [3]. An analysis on the threat UAVs pose to nuclear facilities concludes that UAVs are resourceful tools to cause a variety of perturbations in the form of distraction, reconnaissance or kinetic attacks [3]. Likewise, incidents around airports show a development for potentially dangerous situations [4]. A famous example is the shutdown of the London Gatwick airport for more than a day due to a nearby UAV in 2018 [5]. Similar to the mentioned scenarios, manifold other threats exists that have already taken place like the illegal smuggling of goods over state borders [6] or incidents near governmental buildings [7]. Summarizing, the stated examples and studies emphasise that early detection and identification of uncooperative UAVs is essential in order to protect people and critical infrastructure.

The authors are with the Automation and Control Institute (ACIN), Technische Universität Wien, 1040 Vienna. Further author information: (Send correspondence to Denis Ojdanić)

E-mail: ojdanic@acin.tuwien.ac.at, Tel.: +43 (0) 1 58801 376 515

Various technological solutions based on different sensors exist for the purpose of UAV detection and tracking. For example radio frequency (RF) signals, which are utilized for communication between the operator and drone, can be exploited to detect UAVs in a distance of up to $5\,\mathrm{km}$ [8], [9]. However, a major drawback of this technology is the inherent possibility for controlling the UAV without any communication for example using pre-programmed flight paths. In this scenario, the UAV remains completely invisible to the RF detection system.

RADAR does not suffer from the aforementioned limitation. It detects targets by emitting radio frequency waves and measuring the reflection of an obstructing object. The operational range can go up to several kilometres of distance depending on the size of the radar cross section of the object [10], [11]. Generally, a classification of the detected target is possible with the exploitation of the Doppler effect [12]. However, for objects like consumer drones, which have a small radar cross section, detection and identification becomes more challenging [11].

Acoustic sensors, in the form of microphone arrays, capture sound and detect UAVs by applying appropriate signal processing methods. Angular uncertainties for the target localization of about 4° can be achieved [13] and even identification is possible by using comprehensive databases [14]. The disadvantage of this technology is the small operational range of below $600\,\mathrm{m}$ [15], since it suffers greatly from ambient noise, which lies in a similar frequency band as the emitted UAV sound.

Finally, optical systems capture images with cameras using different spectral bands like the visible spectrum from 400 to $700\,\mathrm{nm}$, but also the near infrared as well as thermal signatures. Paramount for the optical system performance in terms of detection distance is the selection of the field of view (FoV) of the camera system. A wider FoV allows the observation of a broader area with the expense of a shorter achievable distance as the resolution of camera is limited to a finite size. Therefore, most optical UAV detection systems are based on cameras with relatively narrow FoVs, mounted on pan-tilt devices [16]. These devices allow the observation of a large area through realignment of the camera orientation, while providing acceptable resolution. Additionally, these systems can support optical zoom to adapt the FoV to the given task [17]. Current state of the art optical systems achieve a detection distance of about 1 to $2\,\mathrm{km}$ [18]. Whereas being susceptible to weather conditions like rain, snow and fog, the main advantage of optical systems over the previously mentioned methods is the inherent possibility to identify the

incoming target and additional payloads over long distances. This is a crucial benefit, allowing a differentiated situational assessment in order to prepare for appropriate countermeasures. For the task of target detection within the captured images or videos, optical systems rely on computer vision algorithms. Since 2012 deep learning based approaches, which utilize pre-trained neural networks, offer the best performances for object detection [19]. This trend is co-evolving with the accessibility of high performance computers, especially GPUs, for fast and parallel computing. The application of deep learning models based on RCNN [20], SSD [21] and YOLO [22] for UAV detection have been studied extensively in recent years, with studies suggesting that YOLO gives a good trade off in terms of accuracy and speed [23] [24]. Likewise, the usage of deep learning as the main method for UAV detection in systems equipped with pan-tilt-zoom cameras is investigated by various studies [16] [17] [25].
Each one of the presented sensor types offers individual advantages and disadvantages. In order to implement a functional UAV detection system the best advice is to combine several sensors to profit from the respective strengths. An exemplary use case is an unidentified object detected by a radar system in a significant distance. After this coarse localization, the optical component of the system is aligned towards the intruder to capture an image for identification purposes. Ctrl+sky [26] and Ihtar [27] are two multispectral systems that combine radar, RF, acoustics and cameras for UAV detection.

The limitation in the current state of the art is the operational range offered by optical systems. As stated above, optical sensors offer the most conclusive information for classification, as visual material can be classified by neural networks or even cross-checked and understood by a human operator. Therefore, it is paramount to increase the optical detection distance in order to gain more time to decide and prepare for appropriate countermeasures.

The contribution of this paper is the analysis, implementation and evaluation of a telescope-based optical detection system which is capable to detect UAVs over significantly larger distances and thus enabling early threat assessment. Following an analysis of the optical components, suitable hardware is selected to meet the requirements. Deep learning algorithms process the captured video frames and are used for UAV localization. Based on the extracted information, commands are provided to a mount controller to precisely actuate the telescope to keep the UAV within the FoV.
Section II presents an analysis on the most relevant design parameters to achieve long distance UAV tracking. In Section III the system is described together with details about the deep learning algorithms and control strategies in use. The performance of the proposed solution is evaluated in Section IV. Finally, a conclusion is given in Section V.

II. System Analysis

For a conclusive optical system analysis and design, first the expected UAV size and operation range are defined. The UAV classes examined within this work are commercially available drones which fall into the categories $C0$ to $C3$ according of European regulations [28]. Expressed in numbers, the smallest UAV sizes to be detected are about $0.3\,\mathrm{m}$ in diameter and the goal is to significantly extend the state of the art optical detection distances, which are currently around 1 to $2\,\mathrm{km}$ for daylight conditions [18]. In the following, an analysis of the most relevant design parameters and their impact on the system performance is given.

A. Resolution

One crucial parameter to distinguish objects in an image is the theoretically achievable resolution of the optical system. In optimal environmental conditions this value is limited by the diffraction limit which can be calculated using the Rayleigh criterion [29] given in Equation 1

$$\theta = 1.22\frac{\lambda}{D}, \tag{1}$$

where λ is the wavelength of light, D is the aperture size of the optical device and θ the smallest resolvable angle between two objects using the approximation $sin(\theta) \approx \theta$. A wavelength of $530\,\mathrm{nm}$ is selected in accordance to the peak response of typical color cameras for green light. Following this equation an increase of the telescope aperture is necessary to achieve ambitious requirements. However, in practice this is accompanied by two drawbacks. First, larger apertures need bigger telescope structures as seen in Fig. 1, where the weight of typical telescopes is depicted. A larger system mass increases the inertia, which inherently limits the achievable bandwidth for pan and tilt motions. Second, besides the fundamental limit described in Equation 1, the atmosphere further decreases the achievable resolution depending on the Fried parameter r_0 [30]. Local variations due to temperature, pressure and humidity differences affect the refracting index of air and the cumulation of these disturbances along the optical path results in beam spread, wander and intensity fluctuations [31]. Using Equations 2 and 3 to describe the variance of the beam tilt and spread [31]:

$$\alpha_{jit}^2 = 0.182\left(\frac{D}{r_0}\right)^{\frac{5}{3}}\left(\frac{\lambda}{D}\right)^2 \tag{2}$$

$$\sigma_{comp}^2 = 0.134\left(\frac{D}{r_0}\right)^{\frac{5}{3}}, \tag{3}$$

the long exposure Strehl ratio can be calculated [31]:

$$S = \frac{e^{\sigma_{comp}^2}}{1+\left(\frac{2.22\alpha_{jit}D}{\lambda}\right)^2}\frac{1-e^{\sigma_{comp}^2}}{1+\left(\frac{D}{r_0}\right)^2}. \tag{4}$$

Incorporating the atmospheric disturbances to the diffraction limit from Equation 1 results in Equation 5 [31],

$$\theta_{atm} = \theta\frac{Q}{S} \tag{5}$$

$$Q = \sqrt{\frac{e^{2\sigma_{comp}^2}}{1+\left(\frac{2.22\alpha_{jit}D}{\lambda}\right)^2}\cdot\frac{1-e^{\sigma_{comp}^2}}{1+\left(\frac{D}{r_0}\right)^2}}, \tag{6}$$

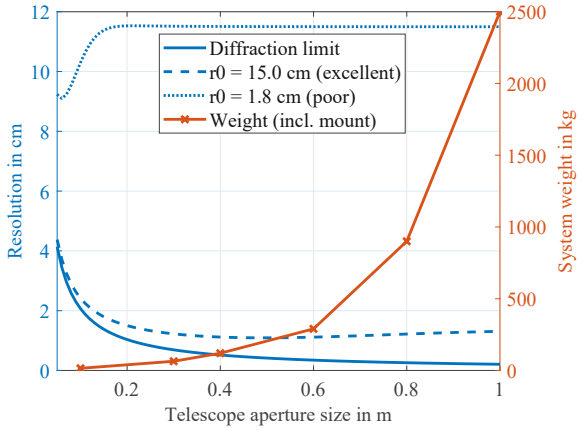


Figure 1: Theoretically achievable resolution of a telescope depending on the aperture size in a distance of 3.2 km. The resolution limited by diffraction (solid blue line) and atmosphere (dashed blue line) under poor (r_0 of 1.8 cm) and excellent (r_0 of 15 cm) conditions is depicted.

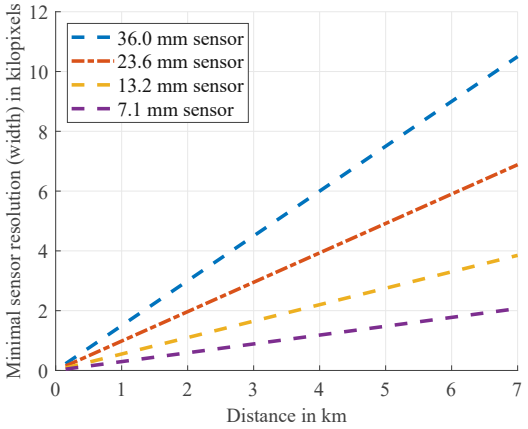


Figure 2: Minimum necessary resolution to obtain 15x15 pixels for an object of 0.3 m size at a given distance using a telescope with a focal length of 1200 mm.

with Q taken from [31]. Plotting the calculations for different aperture sizes and expected values for r_0 during daylight, Fig. 1 demonstrates that larger telescope diameters not necessarily increase the achievable resolution [32], [33]. Depending on environmental conditions smaller telescopes may perform similar as larger ones.

B. Field of view

Another paramount system property is the FoV. A large FoV allows simultaneous coverage of more area during the search for UAVs and an increased margin for tracking errors. However, increasing the observed area requires camera sensors with a higher number of pixels to ensure sufficient resolution. The width of the FoV is determined according to

$$FoV_{rad} = 2\arctan(\frac{s_w}{2f}), \tag{7}$$

with s_w being the camera sensor width and f the focal length of the telescope. A similar formula is applied for calculating the height. The achievable FoV in meter is given in Equation 8

$$FoV_m = 2d \cdot \tan(\frac{FoV_{rad}}{2}), \tag{8}$$

and is used to determine the necessary number of pixels needed to detect objects of different sizes for a given distance d. An exemplary calculation for a system with a typical focal length of 1200 mm and different sensors sizes demonstrates the requirements on the hardware of varying FoVs. An increased sensor size results in a larger FoV as per Equation 7. To detect and identify an object with a diameter of 0.3 m reliably, a minimum number of 15x15 pixels is needed, as discussed in Section IV. The red dash-dotted line in Fig. 2 shows that for the presented system a 23 mm wide APS-C sensor would require a 4k image resolution to detect an UAV reliably in 4 km of distance.

Ultimately the choice of a suitable FoV depends on the speed of the object, the frame rate and the expected detection distances. Assuming a camera frame rate of 60 FPS, Fig. 3 depicts the number of frames captured of a typical consumer UAV flying at a velocity of 31 m/s horizontally through the FoV, if the optical system is not moving. Referring back to the previous example system with a FoV of 1.1°, it is possible to capture 74 images of the UAV in a distance of 2 km with sufficient resolution to accurately detect the object as seen in Fig. 2. Additionally, the 74 captured frames provide enough margin for the object detection algorithm to detect the UAV

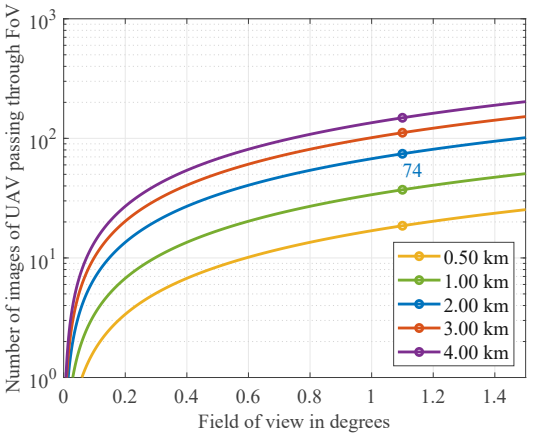


Figure 3: The number of images obtained of an UAV passing the camera FoV horizontally with a speed of 31 m/s at various distances. The camera is running with 60 FPS. For a FoV of 1.1°, 74 images are captured of an UAV passing at a distance of 2 km.

in one of the frames. Reducing the distance between telescope and UAV leads to a reduction of the number of captured images in such a scenario.
Therefore, to implement a system capable of observing long and short distances, a combination of multiple optical systems with different FoVs is necessary. Pairing a large FoV, which observes a broader area over shorter distances, with a narrow FoV, which offers high resolving capabilities for detection and identification over long distances, establishes a system with a wide and flexible operational range.

III. System Implementation

The main architecture of the system is depicted in Fig. 4. The system consists of a telescope and camera as the imaging devices and a commercially available mount to pan and tilt the optical components. The extraction of the UAV position from the video is performed on a PC running a deep learning algorithm. The UAV position is sent to an FPGA, which implements suitable control strategies for feedback control of the mount to keep the UAV within the FoV. In the following an overview of the most relevant components of the system is given.

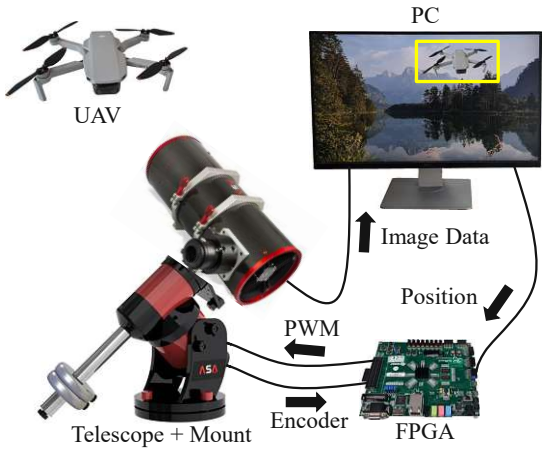


Figure 4: Overview of the system architecture consisting of the telescope and camera as the imaging system, a PC running computer vision algorithms and an FPGA implementing feedback control to actuate the telescope mount to keep the UAV within the FoV.

A. Optical design

For the optical system a Meade Schmidt Cassegrain telescope (LX200-ACF, Meade Acquisition Corp., Watsonville, USA) is selected, with a focal length of $2540\,\mathrm{mm}$ and an aperture size of $254\,\mathrm{mm}$. The choice for the telescope is justified by the calculated resolution using Equation 5 as seen in Fig. 1. A larger aperture, leading to an increased system weight, does not significantly improve the resolution. As an imaging device an ASI 385MC Cool (ZWO Company, Suzhou, China) camera is used, with a sensor width of $7.3\,\mathrm{mm}$, a sensor height of $4.1\,\mathrm{mm}$, and it provides up to 67 frames per second (FPS) at a resolution of 1936x1096 pixels. A FoV of $0.16°$ on the horizontal axis and $0.09°$ on the vertical axis is reached using the telescope and camera. As stated in Section II-B, to extend the optical detection distance a narrow FoV is necessary. A DJI Mini 2 with a size of $0.29\,\mathrm{m}$ x $0.05\,\mathrm{m}$ in a distance of $3\,\mathrm{km}$ for example covers roughly 65x11 pixels in an image, which is enough to perform a detection using current state of the art algorithms [16].

B. Object detection

After capturing a video stream with the camera, appropriate algorithmic methods for UAV localization within each video frame are needed. For this task a deep learning framework is selected and modified. To achieve acceptable results within short processing times, the algorithm runs on a PC equipped with a RTX 3080 GPU (Nvidia Corporation, Santa Clara, California, USA) with 10 GB of GPU RAM, an AMD Ryzen 3900 CPU (Advanced Micro Devices, Inc., Santa Clara, California, USA) and 32 GB of RAM.
A crucial perquisite to work with deep learning algorithms is an extensive and diverse dataset. A custom dataset is created consisting of six UAV multi-rotor models ranging in a diagonal size from $0.3\,\mathrm{m}$ to $2\,\mathrm{m}$. Video material of these UAV models is gathered during flight on different days, locations and backgrounds. From these videos, images are extracted every few seconds and manually labeled. Using data augmentation and synthetic data generation, the dataset is extended to a size of roughly 10.000 labeled images.

YOLOv4 is used as deep learning algorithm for the task of object detection as it provides accurate results while operating at a high frame rate [22], [34]. In order to meet the requirements of long operational distances for the proposed setup, the capability to detect small objects comprising of only a few pixels, is desirable. Therefore, a modified YOLO architecture is proposed based on the standard YOLO approach and the alterations are summarized in Fig. 5.
YOLOv4 uses CSPDarknet53 as a backbone, which is responsible for feature extraction and deriving from these features, detection is performed with the standard YOLO head [22]. In order to improve the detection performance for small objects, the backbone is modified. In the first step, the camera frames are resized to fit the network input image size. If the network resolution is low, an object covering only a few pixels in the camera frame will almost diminish after resizing, making detection more challenging. Hence, a simple possibility for improvement is the increase of the network input resolution, as the frame provided by the camera is resized to the network input size. Additionally, a rectangular input size of 800x512 is chosen to better fit the aspect ratio of the frames provided by the ASI camera. Increasing the resolution comes with a drawback of more memory consumption and lower achievable frame rates. To compensate for this, one might reduce the number of filters, however, a network with a smaller amount of filters learns less features and therefore suffers from poor detection and classification accuracies. A known strategy to

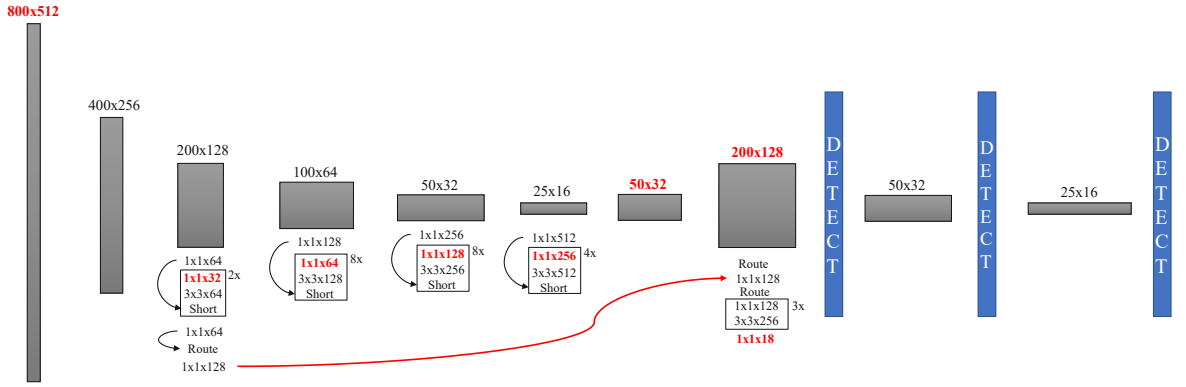


Figure 5: Proposed YOLO architecture. Marked in red are the changes made to the default architecture. Summarizing the adaptions, the network input resolution size is set to 800x512 and the number of filters is reduced in the residual blocks. Furthermore, routing from an earlier stage in the network is established as depicted by the red arrow, paired with appropriate upsampling to allow correct concatenation.

improve computation performance is the introduction of a 1x1xN block, before a 3x3xN convolution within the residual layers in order to reduce the number of filters while keeping as much information as possible [35]. In comparison to the default YOLOv4 configuration the value N, for residual layers is bisected to improve the frame rate and allow for higher input resolutions.

Similar to the residual layers, route layers are utilized within YOLOv4 to facilitate signal propagation through the deep network and to connect finer details observable in earlier layers with layers further back. This technique allows to preserve information from the earlier stages by avoiding the in-between processing [35]. For small object detection it is desirable to route from layers with a high resolution, thus the network is configured to route information from layer number 23 back to layer number 129, which brings information with the size of 208x152 to the later stages of the network [22]. To ensure mathematically correct concatenation the upsampling layers are set to 4 beforehand [22]. The modified route for small object detection is depicted in Fig. 5 by the red arrow.

Finally, the classical YOLO head consisting of three detection layers concludes the network architecture [22]. Each detection layer is responsible to detect objects of varying sizes based on anchor boxes learned from the training dataset. For the application to the custom UAV dataset, these anchor boxes are recalculated using the standard k-means clustering approach [22].

C. Control of the system

The control of the telescope mount is based on the system developed by Riel et al. [36]. For the current implementation a larger mount model, a DDM100 (ASA Astrosysteme GmbH, Neumarkt, Austria), is used. Additionally, the outer control loop is modified to a cascaded velocity (PI) and positional (PID) feedback loop to allow the system to react to the unpredictable motion of an approaching UAV. The velocity loop determines the current which serves as a reference input for an underlying field-oriented-control of the PMSMs motor of the DDM100 mount. The control loops are implemented on an FPGA (Avnet Inc., Phoenix, USA), and the communication between PC, which is connected to the camera and runs the deep learning algorithms for UAV detection, and FPGA is established via an UDP interface.

IV. Experiments and Results

For the experimental analysis and the evaluation of the performance in terms of detection distance, field tests are performed with the implemented setup depicted in Fig. 6. Additionally, the proposed YOLOv4 architecture is tested and compared to the default YOLOv4 configuration on a separate dataset consisting of labeled videos and images.

A. Resolution evaluation

To get an estimate of the resolving power of the presented system, experiments involving a modified version of the USAF resolution test chart are performed. Targets consisting of two white and three black beams are placed in a distance of 3.2 km to validate the imaging system as seen in Fig. 7. The width of the beams is reduced for each target to find the smallest resolvable pattern. To obtain a clear image, the focus is adjusted manually without dedicated auto focus algorithm. The experiment is performed under challenging conditions, as the sun is located behind the target due to the daytime. Considering the application of UAV detection, this scenario corresponds to the detection of an UAV, which is not directly illuminated by the sun and flying in front of a complex background rather than the plain sky. Fig. 8 shows the distribution of pixel intensities over the resolution target. The values are obtained by cropping the image around the target and calculating the mean of pixel intensities in the direction of the beams. Finally, the values are normalized to the corresponding maximal value for better comparison. The

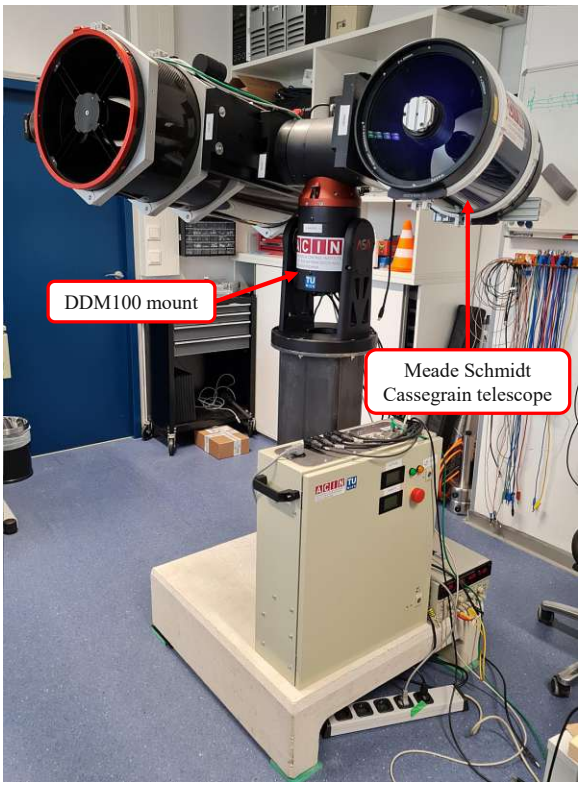


Figure 6: Implemented system with ASA DDM100 mount and two telescopes. For the evaluation of long distance UAV detection, only the Meade Schmit Cassegrain telescope is used.

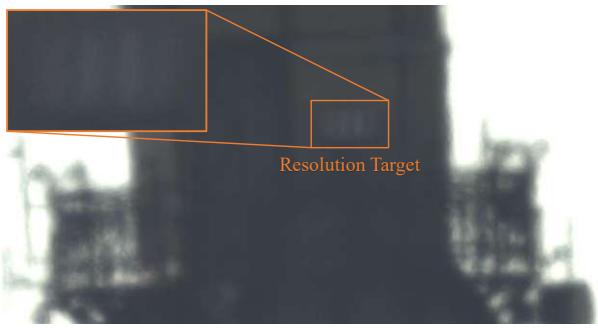


Figure 7: Example of a resolvable target with a beam width of 72 mm in a distance of 3.2 km under challenging conditions. In the top left the target, consisting of alternating black and white beams, is magnified. The distribution of the pixel value intensities is shown in Fig. 8.

two peaks representing the white beams are clearly visible in Fig. 8 for target beam widths of 89 and 72 mm. As the width of the black and white beams is reduced, the two peaks become smaller and not clearly distinguishable. The decision criterion to classify a target as resolvable is linked to the number of frames captured in the video that produce distinguishable

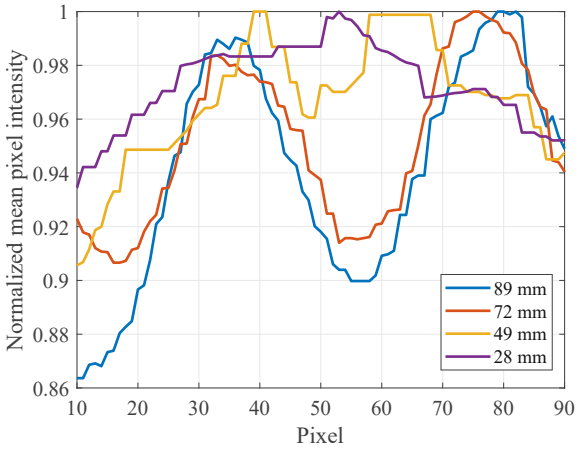


Figure 8: Mean pixel values obtained from resolution targets with various beam widths in a distance of 3.2 km. Higher pixel intensity values correspond to the white coloured beam and lower values to the black beam. Targets down to a beam width of 72 mm are distinguishable.

results as in Fig. 8 only a single observation can be shown. The 49 mm target from Fig- 8 represents the best image in a video sequence of roughly 2 minutes and produces a subjectively barely visible target, whereas the 72 mm bar width is constantly distinguishable throughout the whole video. Therefore, the experimentally obtained resolution is given with 72 mm in a distance of 3.2 km showing that detecting small UAVs over long distances is possible with the chosen optical setup even under challenging environmental conditions. Based on the measured resolution, the Fried parameter equals 2.9 cm, which represents a typical value for daylight observations [32], [33]. The experiments are repeated for the vertical axis of the camera sensor and similar results are obtained as the height and width ratio of the sensor size have a similar ratio to the horizontal and vertical number of pixels.

B. Evaluation of YOLOv4 architecture

For the test phase an additional dataset is needed along with corresponding labels, which have not been used for the training phase. Therefore, the dataset described in Section III is separated into two parts before the training process, namely a training and test set. The test set is further divided into three subsets according to the ground truth bounding box sizes marking the UAVs. The separation into less than 15x15 pixels, 15x15 to 50x50 pixels and larger than 50x50 pixels is necessary to evaluate the algorithm accuracy on different object sizes. It is important to note, that the test images are obtained from different video sequences than the training image set to guarantee that the evaluation is performed with previously unseen and completely new data.

The results of applying the standard and proposed YOLOv4 configurations to the test datasets are shown in Fig. 9. The proposed YOLOv4 architecture outperforms the standard con-

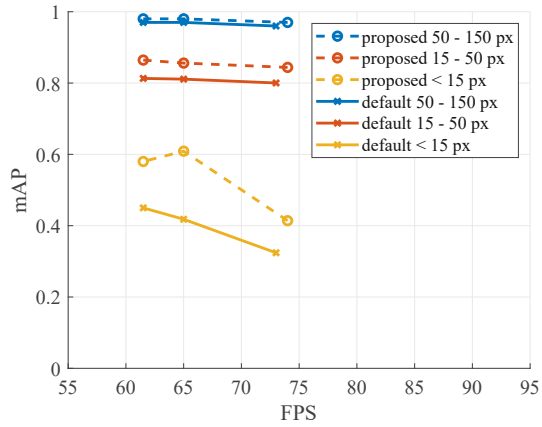


Figure 9: Mean average precision (mAP) for the three test datasets obtained with the default and the proposed YOLOv4 architecture. For small object sizes, the proposed architecture significantly outperforms the default architecture.

figuration around the operating frame rate of the used ASI camera considerably for object sizes below 15x15 pixels. At 65 FPS the proposed architecture achieves a mean average precision (mAP) of 0.61 compared to 0.42 of the default architecture. Likewise, for object sizes between 15x15 to 50x50 pixels the proposed model achieves a mAP of 0.86 compared to 0.81 of the default configuration. This indicates that a larger input resolution of the network, paired with a reduced number of filters, to save memory and increase execution speed, yields an improved performance. Introducing early routing within the network architecture combined with upsampling layers additionally improves the detection performance for small objects. For large objects both approaches leverage similar results. The measurements for different frame rates are obtained by altering the network input resolution showing the overall trend that higher resolutions produce better prediction capabilities. The false positive rate for both YOLOv4 architectures is neglectable when no UAV is located within the image. Table I shows the application of the models on the whole dataset with various intersection over union (IOU) thresholds for the average precision. As mentioned before, the improved precision is attributed to the success rate of the proposed model on small objects. Examining the average precision (AP) for different thresholds, the AP for an IOU threshold of 0.75 of the default architecture is significantly reduced compared to the proposed architecture. Considering the application of reliable long distance UAV detection, Table II shows the recall and precision values, when applying both YOLOv4 configurations to the three test datasets. For precision values close to 1, the respective recall indicates the amount of correctly detected UAVs, as each prediction is a correct prediction and the recall is calculated by the number of correct predictions divided by all possible ground truth objects. Therefore, it can be concluded, that the probability to detect an UAV with a size of 15x15 pixels using the proposed YOLOv4 architecture is

Figure 10: Example image of a detected drone with the bounding box. On the top left a magnification of the UAV is shown.

Table I: Average precision over different IOU thresholds (0.25 to 0.75) after application to whole test dataset. For strict thresholds of 0.75, the AP of the default architecture significantly drops compared to the proposed architecture.

Configuration	$AP_{0.25}$	$AP_{0.5}$	$AP_{0.75}$	FPS
Default YOLOv4	0.89	0.87	0.59	65
Proposed YOLOv4	0.93	0.91	0.70	65

Table II: Comparison of precision and recall results using an IOU threshold of 0.5. For precisions close to 1 recall values resemble the percentage of correctly detected UAVs in the absence of false positives.

	< 15 px		15 - 50 px		50 - 150 px	
Configuration	Prec.	Recall	Prec.	Recall	Prec.	Recall
Default YOLOv4	0.83	0.38	0.99	0.78	1.0	0.91
Proposed YOLOv4	1.0	0.58	0.99	0.82	1.0	0.98

58% and for UAVs larger than 15x15 pixels the probability increases to 82%.
In summary it is shown that the proposed system design paired with a modified deep learning algorithm extends the optical detection range significantly compared to conventional approaches, successfully demonstrating that UAVs are detected with a mAP of 86% in a distance of 3 km.

V. Conclusion

An UAV detection system is developed based on a telescope and camera for image acquisition, a precise mount for alignment of the optical components with the UAV, and a deep learning algorithm for UAV localization. Following an analysis of parameters influencing the system performance, suitable hardware components are selected to create a system that increases the optical detection range of small UAVs compared to the state of the art. Proper training and modification of the deep learning algorithm show that UAVs, covering between 15x15 to 50x50 pixels, are detected with a mAP of 86%. Paired with the proposed optical system the detection range for small unidentified UAVs is extended to more than 3 km under daylight conditions.

Future work will include the integration of a second telescope to add various FoVs for better situational awareness. Additionally, the impact of different light conditions on the achievable detection distances will be investigated.

Acknowledgments

This publication is funded by the Austrian defense research programme FORTE of the Federal Ministry of Agriculture, Regions and Tourism (BMLRT). The authors gratefully acknowledge the cooperation with ASA Astrosysteme GmbH and thank for their support and valuable expertise. Additionally, the support of the Municipal Department 31 of Vienna, Austria (Vienna Water, MA31) for providing access to the Favoriten Water Tower during measurements is gratefully acknowledged.

References

[1] C. F. Liew, D. DeLatte, N. Takeishi, and T. Yairi, "Recent developments in aerial robotics: A survey and prototypes overview," *ArXiv*, vol. abs/1711.10085, 2017.

[2] J. Pyrgies, "The UAVs threat to airport security: risk analysis and mitigation," *Journal of Airline and Airport Management*, vol. 9, no. 2, pp. 63–96, Dec. 2019.

[3] A. Solodov, A. Williams, S. A. Hanaei, and B. Goddard, "Analyzing the threat of unmanned aerial vehicles (UAV) to nuclear facilities," *Security Journal*, vol. 31, no. 1, pp. 305–324, Apr. 2017.

[4] "Air transportation safety investigation report a17q0162 - in-flight collision with drone," Sky Jet M.G. INC., Québec/Jean-Lesage International Airport, Quebec, Tech. Rep., Oct. 2017.

[5] BBC News, "Gatwick drone attack possible inside job, say police," 2010, accessed Sept 2021. [Online]. Available: https://www.bbc.com/news/uk-47919680

[6] B. Shields, "Air traffic control: How mexican cartels are utilizing drones to traffic narcotics into the united states," *Penn State J. Law Int. Aff.*, vol. 5, no. 1, p. 207, 2017.

[7] J. Bart, "Drone crash at white house reveals security risks," 2015, accessed Dez 2021. [Online]. Available: https://eu.usatoday.com/story/news/2015/01/26/drone-crash-secret-service-faa/22352857

[8] S. Yang, H. Qin, X. Liang, and T. Gulliver, "An improved unauthorized unmanned aerial vehicle detection algorithm using radiofrequency-based statistical fingerprint analysis," *Sensors*, vol. 19, no. 2, p. 274, Jan. 2019.

[9] Dedrone, "DedroneSensor RF-360," accessed Jan 2022. [Online]. Available: https://www.dedrone.com/products/hardware/rf-sensors/rf-360

[10] A. Duque De Quevedo, F. Ibanez Urzaiz, J. Gismero Menoyo, and A. Asensio Lopez, "Drone detection and RCS measurements with ubiquitous radar," in *2018 International Conference on Radar (RADAR)*. IEEE, Aug. 2018.

[11] G. Lykou, D. Moustakas, and D. Gritzalis, "Defending airports from UAS: A survey on cyber-attacks and counter-drone sensing technologies," *Sensors*, vol. 20, no. 12, p. 3537, Jun. 2020.

[12] K. Kang, J. Choi, B. Cho, J. Lee, and K. Kim, "Analysis of micro-doppler signatures of small UAVs based on doppler spectrum," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 5, pp. 3252–3267, Oct. 2021.

[13] V. Baron, S. Bouley, M. Muschinowski, J. Mars, and B. Nicolas, "Drone localization and identification using an acoustic array and supervised learning," in *Artificial Intelligence and Machine Learning in Defense Applications*. SPIE, Sep. 2019.

[14] A. Sedunov, D. Haddad, H. Salloum, A. Sutin, N. Sedunov, and A. Yakubovskiy, "Stevens drone detection acoustic system and experiments in acoustics UAV tracking," in *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*. IEEE, Nov. 2019.

[15] M. Benyamin and G. H. Goldman, "Acoustic detection and tracking of a class i UAS with a small tetrahedral microphone array," Tech. Rep., Sep. 2014.

[16] H. U. Unlu, P. S. Niehaus, D. Chirita, N. Evangeliou, and A. Tzes, "Deep learning-based visual tracking of UAVs using a PTZ camera system," in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, Oct. 2019.

[17] A. G. Haddad, M. A. Humais, N. Werghi, and A. Shoufan, "Long-range visual UAV detection and tracking system with threat level assessment," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, Oct. 2020.

[18] J. Farlik, M. Kratky, J. Casar, and V. Stary, "Multispectral detection of commercial unmanned aerial vehicles," *Sensors*, vol. 19, no. 7, p. 1517, Mar. 2019.

[19] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 261–318, Oct. 2019.

[20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2014.

[21] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Computer Vision – ECCV 2016*, 2016, pp. 21–37.

[22] A. Bochkovskiy, C. Wang, and H. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," *CoRR*, vol. abs/2004.10934, 2020.

[23] B. K. S. Isaac-Medina, M. Poyser, D. Organisciak, C. G. Willcocks, T. P. Breckon, and H. P. H. Shum, "Unmanned aerial vehicle visual detection and tracking using deep neural networks: A performance benchmark," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. IEEE, Oct. 2021.

[24] H. Liu, K. Fan, Q. Ouyang and N. Li, "Real-time small drones detection based on pruned YOLOv4," *Sensors*, vol. 21, no. 10, 2021.

[25] J. Park, D. H. Kim, Y. S. Shin, and S. Lee, "A comparison of convolutional object detectors for real-time drone tracking using a PTZ camera," in *2017 17th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, Oct. 2017.

[26] "ctrl+sky drone detection and neutralization system - datasheet," Advanced Protection Systems LLC, New Yersey, USA, Sydney, Australia, 2018, accessed Dez 2021. [Online]. Available: https://www.southerncrossdrones.com/download/aps-counter-drone-technology-sxd-.pdf

[27] Aselsan, "IHTAR anti-drone system - datasheet," ASELSAN A.S., Ankara, Türkiye, 2018.

[28] *Easy Access Rules for Unmanned Aircraft Systems (Regulation (EU) 2019/947 and Regulation (EU) 2019/945)*, European Union Std., 2019.

[29] Rayleigh, "XXXI. Investigations in optics, with special reference to the spectroscope," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 8, no. 49, pp. 261–274, Oct. 1879.

[30] D. L. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *Journal of the Optical Society of America*, vol. 56, no. 10, p. 1372, Oct. 1966.

[31] R. Tyson, *Principles of Adaptive Optics*. CRC Press, Sep. 2010.

[32] A. V. Sergeyev and M. C. Roggemann, "Monitoring the statistics of turbulence: Fried parameter estimation from the wavefront sensor measurements," *Applied Optics*, vol. 50, no. 20, pp. 3519–2528, Jul. 2011.

[33] T. Özışık and T. Ak, "First day-time seeing observations at the TÜBiTAK national observatory in turkey," *Astronomy & Astrophysics*, vol. 422, no. 3, pp. 1129–1133, Jul. 2004.

[34] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 13 029–13 038.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[36] T. Riel, A. Galffy, G. Janisch, D. Wertjanz, A. Sinn, C. Schwaer, and G. Schitter, "High performance motion control for optical satellite tracking systems," *Advances in Space Research*, vol. 65, no. 5, pp. 1333–1343, Mar. 2020.

Publication [P2]

Evaluation of the required optical resolution for deep learning-based long-range UAV detection

Authors: D. Ojdanić, N. Paternoster, C. Naverschnigg, A. Sinn, and G. Schitter

Reference:

[P2] D. Ojdanić, N. Paternoster, C. Naverschnigg, A. Sinn, and G. Schitter, "Evaluation of the required optical resolution for deep learning-based long-range UAV detection," In Pattern Recognition and Tracking XXXV, vol. 13040, pp. 78-84. SPIE, 2024

Evaluation of the required optical resolution for deep learning-based long-range UAV detection

Denis Ojdanić*, Niklas Paternoster*, Christopher Naverschnigg, Andreas Sinn, and Georg Schitter

Automation and Control Institute (ACIN), TU Wien, Gusshausstrasse 27-29, Vienna, Austria

1. ABSTRACT

This paper evaluates the required resolution of a telescope system experimentally to enable a reliable deep learning-based long-range UAV detection. FRCNN, a state-of-the-art deep learning object detector is fine-tuned for UAV detection with a custom dataset. A test dataset has been created of a small UAV in front of a clear and complex background at distances ranging from 500 m up to 2500 m using a telescope with a focal length of 1325 mm and an aperture of 102 mm. At each distance the resolution is measured with a modified version of the US Air Force resolution chart. The results show that a small UAV is detected with a mAP(0.5) of above 90 % in front of a complex background up to a distance of 1167 m given a minimum resolution of 9.3 mm or 8 µrad and up to 2222 m in front of a clear background given a minimum resolution of 38 mm or 17.1 µrad.

Keywords: Deep learning, resolution, telescope, UAV detection

2. INTRODUCTION

The trend to cost-effective small Unmanned Aerial Vehicles (UAVs) has resulted in a growing number of threatening and dangerous scenarios, posing heightened risks to critical infrastructure, including airports and power plants. On multiple occasions air traffic in major European cities such as Dublin,[1] Berlin,[2] and Nice[3] had to be temporarily halted or experienced significant delays due to the presence or even collisions with UAVs, underscoring the disruptive impact this technology can have on aviation operations. Small drones also have been illicitly utilized for illegal transportation of narcotics across national boundaries[4] and engaging in espionage, with a particular focus on the surveillance of critical infrastructure such as power plants and other strategically significant installations.[5] Therefore, UAV detection systems are crucial to identify the incoming threat and to prepare counter measures.

As a consequence, a lot of research and development is focusing on UAV detection. A typical UAV detection system combines different sensors including LiDAR,[6] RADAR,[7] acoustic sensors,[8] RF (radio frequency) based systems[9] and optical systems.[10] Optical detection relies on computer vision to detect and track an object. Detection is accomplished through the utilization of deep learning algorithms, which have consistently demonstrated their effectiveness over the years. State-of-the-art algorithms like SSD,[11] FRCNN,[12] and FCOS[13] have proven to be robust solutions for detection tasks and numerous studies utilize these algorithms for UAV detection.[14,15]

To obtain a high quality image, a suitable camera system has to be selected to capture the UAV with sufficient pixels covering the object for a reliable detection by a neural network.[10] Typically, cameras are used with a narrow field of view mounted on pan-tilt devices to enable surveillance of a large area.[16] Long focal lengths and large apertures, for example offered by telescopes, further increase the detection distance.[10] At long distances, the atmosphere has a greater influence on the image quality and limits the achievable optical resolution. In order to design and select suitable optical components to enable long distance UAV detection, an adequate experimental analysis of the necessary resolution is required. The optical resolution can be measured for example by using resolution test targets with the Contrast Transfer Function (CTF), or the application of the Modulation Transfer Function (MTF) through slit illumination.[17] Determining the achievable resolution of an optical detection system and correlating it with the detection probability of a neural network for UAV detection is a crucial

* Both authors contributed equally.
Further author information: Send correspondence to Denis Ojdanić
E-mail: ojdanic@acin.tuwien.ac.at, Telephone:+43 (0) 1 58801 376 520

aspect to determine the limits of the system at hand.
The contribution of this paper is the experimental evaluation of the required resolution for a telescope-based long distance UAV detection system to achieve reliable deep learning-based UAV detection.

3. METHODOLOGY

3.1 Telescope setup

The investigated system consists of a Celestron NexStar telescope (Celestron, USA) with a focal length of 1325 mm and an aperture of 102 mm. The camera used for image acquisition is an ASI 385 MC-Cool (ZWO Company, Suzhou, China) camera with a sensor diagonal of 8.37 mm and a quadratic pixel size of 3.75 µm. Fig. 1 shows the telescope system and camera on a tripod during field tests.

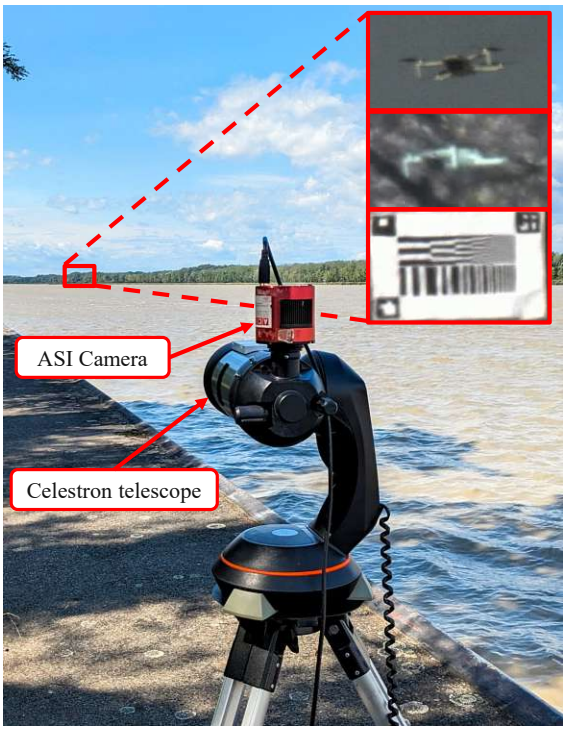
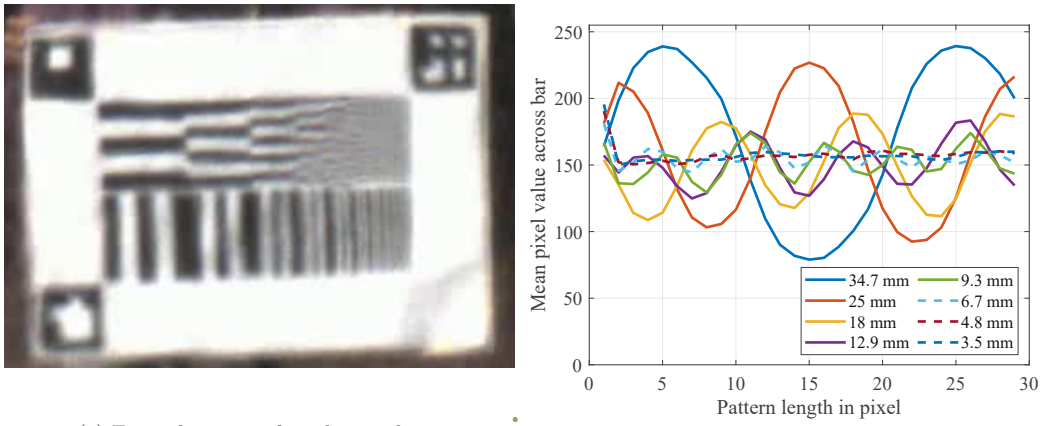


Figure 1: Schematic overview of the utilized measurement setup. In the distance the UAV and the resolution chart is illustrated.

3.2 Automated resolution measurement

An auxiliary software program is developed to automatically and consistently determine the optical resolution of the individual frames. A custom resolution chart, which enables automatic evaluation, consisting of black and white bars in different scales as seen in Fig. 2a, is created. The pattern is based on the 1951 United States Air Force (USAF) resolution chart.
The custom pattern consists of the mentioned bars, which are scaled down to a smaller size. The bars are plotted horizontally and vertically to determine the horizontal and vertical resolution of the optical system simultaneously. In order to automatically extract the achieved resolution, in addition to the horizontal and vertical bars, three ArUco markers[18] are added to the paper print as visible in the corners of Fig. 2a. Exploiting the rotation-invariant detection capabilities of ArUco markers, the corners of the resolution chart are automatically detected

and possible rotational errors are corrected. This alignment enables the individually scaled sub-patterns to be separated by a precise image crop. Moreover, the paper format and the scaling factor of the black and white bars is stored as an ID within the AruCo markers to enable automatic determination of the width of each bar. If the ArUco marker detection fails, the alignment and the selection of the region to be cropped can also be done manually. Once the sub-patterns are extracted, the mean pixel value along the length of the pattern bar is calculated. Plotting the mean value over the alternating black and white bars, will create a sine-like function as seen in Fig. 2b.



(a) Example image of resolution chart.

(b) Extracted pixel values.

Figure 2: a) The custom resolution chart including three AruCo markers, which facilitate a simple detection and orientation adjustment for extracting the individual sub-patterns. The resolution chart imaged through an optical device under test. b) The extracted pixel values for each horizontal sub-pattern. The dashed lines are not discernable according to the threshold marking the lowest achievable resolution at 9.3 mm for this example.

The optical resolution of a frame is ascertainable by the bar width of the smallest distinguishable bars. To determine a threshold for distinguish-ability, 100 images of varying resolutions are inspected subjectively to discern, which pixel value difference between black and white bars is still distinguishable to a human observer. In addition to the described procedure, a mean value for the resolution is calculated over an entire video consisting of multiple images, to consider the fluctuating influence of the atmosphere within the measurement.

3.3 Deep Learning-based object detection

To find a correlation between the optical resolution and the detectability of a UAV, the deep learning object detector FRCNN[12] is trained. For the training process approximately 14000 UAV images are used,[19] whereas this dataset is split into 91 % training and 9 % validation data. FRCNN is initialized using the weights pre-trained on the COCO dataset[20] and then fine-tuned on the custom UAV dataset using an RTX 3080 GPU (Nvidia Corporation, Santa Clara, California, USA) with 10 GB of GPU RAM for 30 epochs. The learning rate is set to 0.0015, with a weight decay of 0.0006 and a momentum of 0.8. For a stepwise reduction of the learning rate, it is multiplied by a factor of 0.1 after epoch 22 and 25. During the training process, the models are evaluated using the validation dataset and the best model according to the mean average precision (mAP) at a threshold of 0.5 is selected and used for the final experiments in Section 4. Fig. 3 shows the overview of the loss and the validation score during the training process. The best model is obtained at epoch 25 and is extracted to be used for inference.

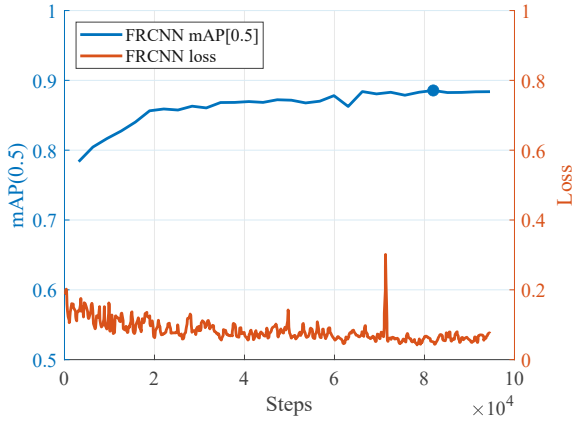


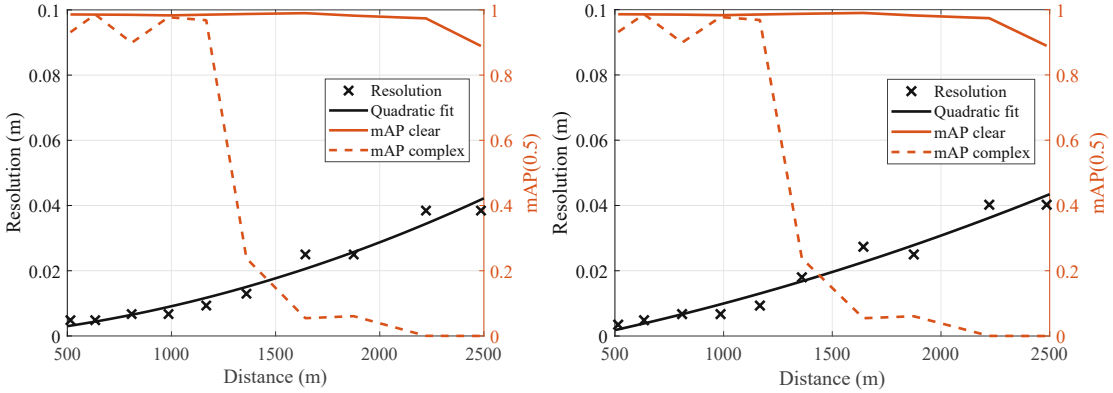
Figure 3: Overview of the training process showing the loss and the mAP(0.5) when applying the models onto the validation dataset. The blue circle marks the best performing model selected for inference.

4. EXPERIMENTS AND RESULTS

Figure 4: Example collection of images at various distances of the resolution chart and the DJI Mini 2 in front of a complex and a clear background. Note that the images are cropped around the relevant object.

For the experimental evaluation field tests are conducted using the presented optical system, the custom made resolution chart and a DJI Mini 2, which has a width of 289 mm. In order to experimentally validate the necessary optical resolution to detect UAVs at a certain distance, three measurements are performed per distance. A video is captured of the resolution chart in a distance x. At the same distance x, two more videos are captured of the DJI Mini 2 flying in front of a clear and complex background. In Fig. 4 example images are shown at various distances, ranging from approximately 500 m up to around 2500 m.
The results of the resolution chart measurements are evaluated according to the automated method described in Section 3.2. Within each video, the resolution chart is automatically detected via the ArUco markers or by manually selecting its corners. For longer distances, the ArUco markers are omitted to enable printing larger black and white bars. For these resolution charts, the region to be cropped is selected manually. The video

images are cropped to separate each sub-pattern of the resolution chart for evaluation. Evaluating a single image across all sub-pattern produces results as depicted in Fig. 2b. To consider the effect of the atmosphere onto the measurement, the mean of the obtained resolution is calculated over each video sequence, which last about 10 s at a camera frame rate of 60 fps.
To evaluate the results obtained by the deep learning algorithm, for each distance and background a 10 s video of the UAV is recorded at 60 fps. Every 6th image is extracted from these videos, which is about 100 images per background and distance, totalling to approximately 2000 images over all scenarios. These are manually labeled to obtain the ground truth for the subsequent evaluation. As an evaluation metric the mAP at a threshold of 0.5 is used.



(a) Horizontal resolution in meter.

(b) Vertical resolution in meter.

Figure 5: The resolution measured with the USAF resolution chart is shown over a distance from 500 m to 2500 m together with a quadratic fit of the data. Additionally, the mAP in front of a clear and complex background using a fine-tuned FRCNN object detector is depicted.

Fig. 5 shows the minimally distinguishable bar width in meter, which translates to the achievable resolution, and the mAP(0.5) over various distances. The results demonstrate that for a small UAV in front of a complex background, a minimal resolution of 9.3 mm or 8 µrad is necessary to enable UAV detection with a mAP(0.5) of above 90 % from a distance of 1167 m. For the case of a clear background the mAP starts to drop below 90 % at distances above 2222 m, where the measured resolution is 38 mm or 17.1 µrad. Presenting the results in terms of necessary number of black and white bars covering the size of the object to be detected, in front of a clear background it has to have at least the size of 7.5 times the achievable resolution bar width and for a complex background this factor increases to 31. It is expected that it is easier to discern objects in front of a clear background, as a higher contrast towards the constant and unchanging background allows the deep learning algorithm to detect the object better. In Fig. 4, the images of the UAV in distances above 1360 m illustrate the difficulty of detecting the UAV at low resolution in front of a complex background.
The presented measurements allow determination of system requirements of an optical system, when trying to detect UAVs of a certain size reliably. These numbers can be used to determine the size of the object, which can be reliably detected with a certain optical setup or also the required aperture when selecting an appropriate lens for a camera.
Summarizing the results, to achieve a mAP(0.5) of more than 90 % a minimal resolution of 8 µrad for complex and 17.1 µrad for clear background is necessary to detect UAVs of 289 mm in diameter reliably.

5. CONCLUSION

An experimental evaluation has been performed to correlate the optical resolution to the detection accuracy of a deep learning-based algorithm for the task of UAV detection. The results show, that for a detection of a UAV with a diameter of 289 mm with a mAP(0.5) of above 90 % using deep learning object detection, a minimum

resolution of 8 µrad for a complex and 17.1 µrad in front of a clear background is necessary. The presented results can be used as a basis to estimate the requirements for an optical system to detect objects of a given size in a certain distance.

ACKNOWLEDGMENTS

This publication is funded by the Austrian defence research programme FORTE of the Federal Ministry of Finance (BMF).
The authors gratefully acknowledge the cooperation with ASA Astrosysteme GmbH and thank for their support and valuable expertise.

REFERENCES

[1] Reuters, "Ireland vows to tackle drones after dublin airport shut six times," 2023. https://www.reuters.com/world/uk/ireland-vows-tackle-drones-after-dublin-airport-shut-six-times-2023-03-03/ [Accessed: (05.02.2024)].

[2] J. Klender, "Drone operator above tesla giga berlin spoils routine descent for passenger plane," 2022. https://www.teslarati.com/tesla-giga-berlin-drone-operator-berlin-brandenburg-airport-plane/ [Accessed: (05.02.2024)].

[3] J. Field, "Drone operator above tesla giga berlin spoils routine descent for passenger plane," 2023. https://aviationsourcenews.com/incident/emirates-a380-struck-by-drone-wing-damage-in-nice-france/ [Accessed: (05.02.2024)].

[4] S. Hollister, "Da tiny dji drone smuggled its own weight in drugs over the us border wall," 2021. https://www.theverge.com/2022/2/3/22916246/dji-mini-2-drone-smuggle-meth-us-mexico-border-wall [Accessed: (05.02.2024)].

[5] BBC, "Sweden drones: Sightings reported over nuclear plants and palace," 2022. https://www.bbc.com/news/world-europe-60035446 [Accessed: (05.02.2024)].

[6] S. Dogru and L. Marques, "Drone detection using sparse lidar measurements," *IEEE Robotics and Automation Letters* **7**(2), pp. 3062–3069, 2022.

[7] J. Drozdowicz, M. Wielgo, P. Samczynski, K. Kulpa, J. Krzonkalla, M. Mordzonek, M. Bryl, and Z. Jakielaszek, "35 GHz FMCW drone detection system," in *2016 17th International Radar Symposium (IRS)*, pp. 1–4, IEEE, 2016.

[8] J. Mezei, V. Fiaska, and A. Molnár, "Drone sound detection," in *16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, pp. 333–338, IEEE, 2015.

[9] P. Nguyen, M. Ravindranatha, A. Nguyen, R. Han, and T. Vu, "Investigating cost-effective rf-based detection of drones," in *Proceedings of the 2nd workshop on micro aerial vehicle networks, systems, and applications for civilian use*, pp. 17–22, 2016.

[10] D. Ojdanić, A. Sinn, C. Naverschnigg, and G. Schitter, "Feasibility analysis of optical uav detection over long distances using robotic telescopes," *IEEE Transactions on Aerospace and Electronic Systems* **59**(5), pp. 5148–5157, 2023.

[11] Wei, L., Anguelov, D., Erhan, D., et. al., "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*, pp. 21–37, Springer International Publishing, (Cham), 2016.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, **28**, Curran Associates, Inc., 2015.

[13] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9627–9636, 2019.

[14] Z. Chen, S. Huang, and D. Tao, "Context refinement for object detection," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 71–86, 2018.

[15] D. Ojdanić, C. Naverschnigg, A. Sinn, and G. Schitter, "Deep learning-based long-distance optical uav detection: color versus grayscale," in *Pattern Recognition and Tracking XXXIV*, **12527**, pp. 80–84, SPIE, 2023.

[16] H. U. Unlu, P. S. Niehaus, D. Chirita, N. Evangeliou, and A. Tzes, "Deep learning-based visual tracking of uavs using a ptz camera system," in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, **1**, pp. 638–644, 2019.

[17] C. Leung and T. Donnelly, "Measuring the spatial resolution of an optical system in an undergraduate optics laboratory," *American Journal of Physics* **85**(6), pp. 429–438, 2017.

[18] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition* **47**(6), pp. 2280–2292, 2014.

[19] D. Ojdanić, C. Naverschnigg, A. Sinn, D. Zelinskyi, and G. Schitter, "Parallel Architecture for Low Latency UAV Detection and Tracking using Robotic Telescopes," *IEEE Transactions on Aerospace and Electronic Systems* , submitted Jul 2023.

[20] Tsung-Yi Lin et. al., "Microsoft COCO: common objects in context," *CoRR* **abs/1405.0312**, 2014.

Publication [P3]

Parallel Architecture for Low Latency UAV Detection and Tracking using Robotic Telescopes

Authors: D. Ojdanić, C. Naverschnigg, A. Sinn, D. Zelinskyi, and G. Schitter

Reference:

[P3] D. Ojdanić, C. Naverschnigg, A. Sinn, D. Zelinskyi, and G. Schitter, "Parallel Architecture for Low Latency UAV Detection and Tracking using Robotic Telescopes," in IEEE Transactions on Aerospace and Electronic Systems, vol. 60, no. 4, pp. 5515-5524, Aug. 2024

Journal Quartile: Q1 (2024)

Parallel Architecture for Low Latency UAV Detection and Tracking using Robotic Telescopes

Denis Ojdanić, Christopher Naverschnigg, Andreas Sinn, Daniil Zelinskyi, and Georg Schitter

***Abstract*— This paper presents the implementation of a multi-threaded parallel architecture, which enables telescope-based optical UAV detection and tracking in real-time. For efficient image processing an accurate deep learning object detector is complemented in parallel by a fast object tracker. A transition strategy between detector and tracker is introduced based on the tracker reliability, which improves the object localization accuracy of the system. The deep learning algorithm initializes the tracker and in the subsequent frames the reliability of the tracker is compared to the confidence value of each newly detected object to determine whether a reinitialization is necessary. The implemented architecture successfully demonstrates the parallel combination of an FRCNN detector and a MEDIANFLOW tracker to achieve visual UAV detection and tracking at 100 fps. The proposed reliability-based strategy outperforms a purely detector and tracker-based strategy by 6% and 14% respectively in terms of intersection over union at a threshold of 0.5, in scenarios, when the target UAV is flying in front of a complex background. Additionally, the implemented parallel architecture increases the probability for a flight path estimation, which requires at least two localizations, by 49%, when compared to a non-parallel architecture. Field tests are conducted with the proposed architecture using a telescope system demonstrating UAV detection and tracking at 100 fps in distances up to 4000 m in front of a clear background.**

***Index Terms*— Deep learning, detection, parallel, tracking, UAV**

I. Introduction

The usage of unmanned aerial vehicles (UAV)s has seen an unprecedented growth in recent years due to their versatility and manifold utility [1]. Along with many positive operational scenarios, exploitation of the technology for malicious activities poses a major threat to public safety. Many incidents emphasise the enormous negative impact of UAVs including drone sightings in the vicinity to an airport in the UK in 2022 [2], dangerous situations close to nuclear facilities [3] and trafficking in and out of prisons or across state borders [4], [5]. The mentioned examples illustrate the dangerous potential of UAVs and show the necessity for UAV detection systems to enable timely reconnaissance in order to prepare appropriate defensive measures.

For the task of UAV detection different approaches exist including RADAR [6], [7], radio frequency [8], acoustic [9] and electro-optical detection [10]. Each of the mentioned methods have their benefits and drawbacks. Therefore, often multiple sensors are combined to a multispectral UAV detection system [11], [12]. However, most of these systems ultimately rely on electro-optical sensors to perform object classification, as visual images can easily be interpreted by human operators or advanced computer vision algorithms. To extend the operational range of an electro-optical system, a narrow field of view (FoV) and a large optical aperture are necessary, which can be achieved by using telescopes [10]. To increase the situational awareness of such a system to a larger area, dedicated mounts enable pan and tilt motion [13]. The typically narrow FoV of a few degrees coupled with high UAV velocities of more than $20\,\mathrm{m/s}$ require real-time computer vision based detection and tracking to maintain the UAV within the camera FoV.

Detection, for example in sense-and-avoid (SSA) scenarios between an aircraft and an UAV, is facilitated by using morphological filters [14] paired with SVM classifiers [15]. Other methods to detect moving objects are based on optical flow [16] or background modelling [17]. However, these traditional methods are often limited due to a high false alarm rate making a reliable detection difficult. Deep learning based approaches offer an ameliorated detection accuracy and extensive research has brought up a variety of algorithms suited for the task, such as YOLO [18], SSD [19], FRCNN [20] or Retinanet [21]. Consequently, a lot of recent research is conducted on deep learning based UAV detection [22], [23], [24]. Additionally, object tracking can be performed using deep learning by taking advantage of bounding box regression for the prediction of the object location within the next frame [25]. Siamase based trackers learn similarity functions between the desired target to track and the search regions [26]. The improved accuracy comes at the cost of an increased computational complexity, which limits the achievable frame rates of these methods. Object tracking as an autonomous task has been widely researched and a variety of non-deep learning solutions exist for this purpose, which require less computational effort. Minimum Output Sum of Squared Error (MOSSE) [27] and Kernelized Correlation Filter (KCF) [28] are examples of algorithms running at high frame rates. These trackers use a correlation filter to build a model of a selected object online and correlate extracted features to locate the object in

This publication is funded by the Austrian defence research programme FORTE of the Federal Ministry of Finance (BMF).
The authors gratefully acknowledge the cooperation with ASA Astrosysteme GmbH and thank for their support and valuable expertise.

The authors are with the Automation and Control Institute (ACIN), TU Wien, 1040 Vienna. Further author information: (Send correspondence to Denis Ojdanić) E-mail: ojdanic@acin.tuwien.ac.at, Tel.: +43 (0) 1 58801 376 520

consecutive frames. Channel and Spacial Reliability Tracker (CSRT) [29] offers an improved accuracy with a lower frame rate utilizing correlation filters calculated in the Fourier domain. Another example of a high-speed tracker is MEDIANFLOW [30], which uses the Lucas-Kanade method [31] and estimates an object position by examining the trajectories in future and past frames. As a prerequisite, an initialization is necessary, either by a human operator or dedicated detection algorithm.
To improve the detection and tracking accuracy, various strategies are explored to combine different algorithms. A common approach is the combination of a detection algorithm like background subtraction with a Kalman filter to ensure an improved tracking performance [32], [33]. For SAA applications detecting moving airborne objects on collision course is facilitated by extracting features from warped difference images with subsequent binarization and morphological filtration [34]. Ensuing creation of measurement vectors through examination of multiple frames enables object tracking via Particle filtering [34] or using Hidden Markov Model filters [35]. The SORT framework is an example of a combination of a deep learning object detector with a Kalman Filter to improve the achievable frame rates [36]. For SAA onboard of UAVs, a combination of YOLOv2 with estimators is used, which creates firm tracks by associating single frame detections over different frames in close proximity to form firm tracks to be then further processed by Kalman Filters [37]. Likewise, parallel execution on a multi-threaded system enables collaboration between trackers and detectors, whereas the decision for the final bounding box is determined by either trusting the detector, the tracker or alternating between the two. These methods, combining Tiny-YOLOv3 [38] with SiamRPN [26], allow frame rates of up to 48 fps on a workstation equipped with an Intel i7-6800k CPU and a NVIDIA Geforce GTX 1080Ti GPU [39]. Combining a traditional tracker with a parallel verifier enables to improve the performance on tracking failures by reinitializing the tracker with a trusted and verified localization [40]. While improving the overall tracking performance, these methods solely rely on the object detector, disregarding the input given by the trackers, which are usually very robust for a short number of frames after initialization. As a consequence, each miss-detection reinitializes the tracker and causes a tracking failure. To allow a collaboration between tracker and detector, a methodology is needed to determine, whether a reinitialization of the tracker is necessary or if the current track is more reliable than the detection.

The contribution of this paper is the implementation and experimental evaluation of a telescope-based system, capable of detecting and tracking UAVs reliably at 100 fps. A custom parallel architecture combines a slow and accurate deep learning object detector with a fast object tracker to enable a high sampling rate of the UAV position, which is necessary to precisely actuate the telescope mount. A transition strategy is proposed to further improve the collaboration between detector and tracker based on the detection probability and the tracker reliability. Field tests demonstrate the detection and tracking capabilities of the proposed system.
The remainder of the article is organized as follows. Section II offers a detailed description of the system architecture and methodology of the collaboration between a deep learning detector and a traditional object tracker. Section III describes the implemented system, the utilized hardware, the training dataset and specifies how the neural networks are trained. Section IV shows the experiments conducted and results obtained followed by concluding remarks in Section V.

II. System Description

In this section the proposed system architecture and the concept to enable efficient collaboration between object detection and tracking algorithm is presented. In order to combine two algorithms together with a camera to achieve high performance, a multi-threaded approach utilizing several CPU cores guarantees fast execution. Fig. 1 shows the proposed system architecture consisting of various threads running on different CPU cores. The communication between threads is implemented via shared buffer locations within the memory, which use mutual exclusion to prevent reading and writing to a buffer simultaneously. The camera writes the acquired frames to a double frame buffer, meaning it alternates between writing an image to two different buffer spots. As a frame contains a lot of data, writing and reading takes relatively long. To avoid a waiting and blocking behaviour, the double frame buffer enables simultaneous reading from one and writing to the other memory block. The detector and tracker access the double frame buffer to read new images, which they internally process to detect and track objects. The detector, as a sophisticated algorithm, requires more time to detect objects within a frame and thus only manages to process every fifth camera frame, which provides images at 100 fps. Upon detecting a UAV, the tracker is initialized, which has been idle up to this point. Once initialized, the tracker is capable of processing every camera frame and provides object localizations also on frames the detector does not process. The detector data is used to correct and, if necessary, reinitialize the tracker. Based on the timestamp of the frame, where the detection and, in parallel, the tracking is conducted, the mount controller sends the most recent localization as a pan and tilt command to the telescope mount.

A. Parallel Architecture

A. Reliability

The decision when to reinitialize the tracker, given a new detection, is based on the confidence of the detection

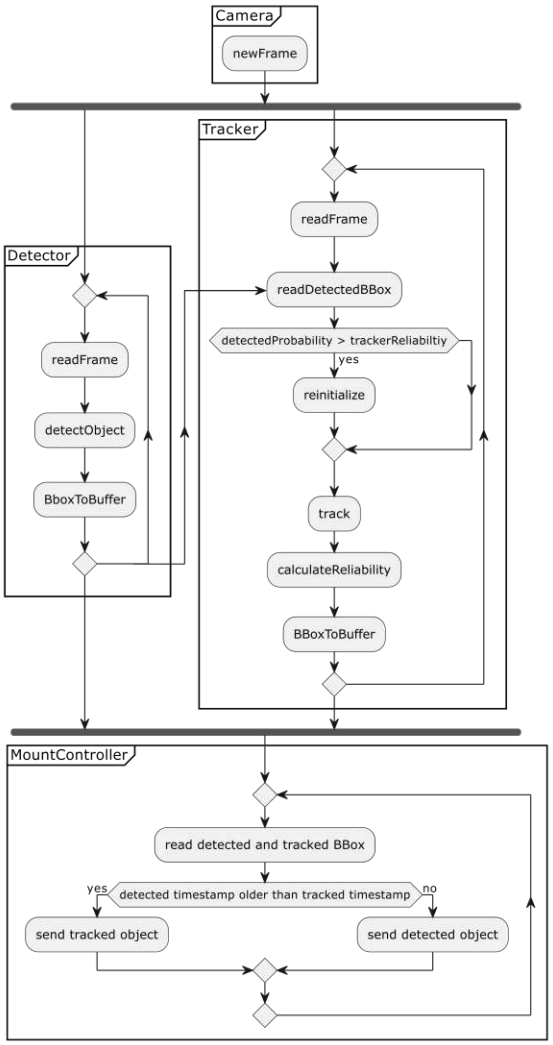


Fig. 1. Overview of the parallel architecture. The camera, which is attached to the telescope, detector, tracker and mount controller are each running on a separate thread and CPU. The camera provides frames to the detector and tracker and the latter one is initialized by a new detected object. Based on the most recent timestamp of the frame, where a tracked or detected object is found, the mount controller sends pan and tilt commands to the telescope mount.

and the reliability of the current track. Deep learning algorithms provide a confidence value for each bounding box predicted within an image, which is used to judge how certain the algorithm is about each detection. Classical object trackers like the mentioned KCF, MOSSE or MEDIANFLOW, do not provide such a value. To estimate a confidence for the tracker, the reliability metric is used, which can be interpreted as the probability that a tracker is correctly tracking a target n frames after the initialization. The reliability R is given by [41]

$$R = e^{-np}, \tag{1}$$

where p is the normalized failure rate. The failure rate for a tracker has to be determined a-priori in supervised manner and represents the track failures over time for a given number of frames.
By comparing the confidence reported for each detection by the deep learning algorithm with the pre-configured reliability value for the tracker, a mechanism is established, which allows collaboration between tracker and detector rather than consistently trusting either one of the two in any situation. Therefore, the tracker is being reinitialized, only if the confidence of a new detection is larger than the currently reported reliability. During the initialization of the tracker, the confidence reported by the detection is taken as a starting value for the reliability, which then degrades as the time passes according to Eq. 1.

III. System Implementation

The implemented system as shown in Fig. 2 consists of a Meade Schmidt Cassegrain telescope (LX200-ACF, Meade Acquisition Corp., Watsonville, USA) with a focal length of $2540\,\mathrm{mm}$, which is mounted on a DDM100 (ASA Astrosysteme GmbH, Neumarkt Austria). The system is equipped with the Moment CMOS scientific camera (Teledyne Photometrics, USA), which has a pixel size of $4.5\,\mu\mathrm{m}$ x $4.5\,\mu\mathrm{m}$ and is operated at a resolution of 1920 pixels x 1100 pixels at a frame rate of $100\,\mathrm{fps}$. The processing is done on a PC equipped with an RTX 3080 GPU (Nvidia Corporation, Santa Clara, California, USA) with 10 GB of GPU RAM, an AMD Ryzen 3900 CPU (Advanced Micro Devices, Inc., Santa Clara, California, USA) with 24 threads on 12 cores and 32 GB of RAM.

A. Object detection

For object detection region based convolutional neural network (FRCNN) [20], a state of the art deep learning object detection algorithm, is selected and trained, as it is one of the most accurate object detectors. As strategy for training, fine-tuning is applied, which starts the optimization process from already pre-trained network parameters. The network is initialized with weights pre-trained on the COCO dataset, which consists of more than 300.000 images with 80 object categories [42]. The dataset is chosen, because object classes within the dataset, like the airplane class, are similar to UAVs. Based on this initialization, FRCNN, equipped with a new detection head, is fine-tuned on the custom UAV dataset.
The dataset used for fine-tuning contains 18000 images, with approximately two thirds being taken from [10] with additional images of UAVs being added from field tests using the presented telescope and camera system. The remaining 6000 images are taken from the Drone vs Bird (DvB) dataset [43]. From this DvB dataset, which consists of multiple UAV and bird videos, a random selection of videos is set aside for experiments and from

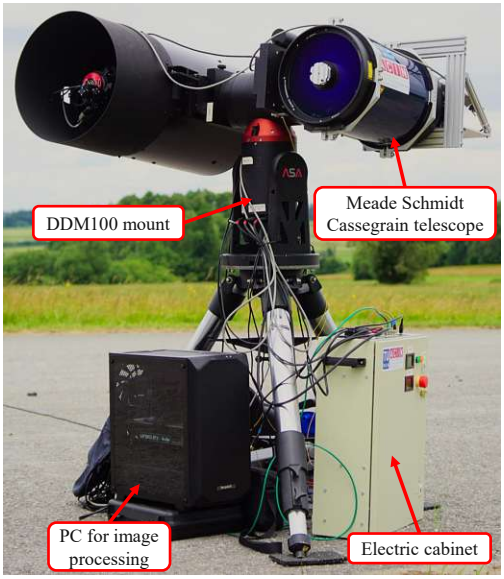


Fig. 2. The ASA DDM100 mount and Meade Schmidt Cassegrain telescope, used for target tracking and image acquisition [10].

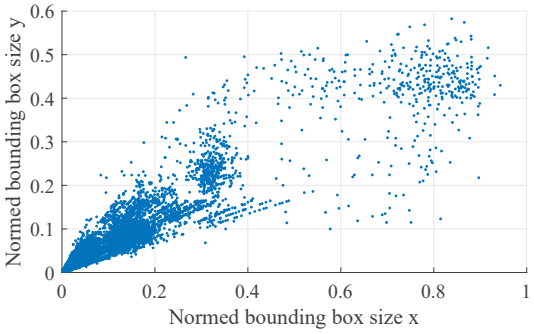


Fig. 3. Each point represents the size of a bounding box, showing the distribution of the bounding box sizes within the training dataset. The bounding boxes are normalized to the image width and height.

the remaining videos two images per second are extracted for fine-tuning to prevent over-fitting by adding numerous similar images to the training set. Fig. 3 shows the bounding box size normalized to the image width and height of the whole training dataset. Note, for the training and test dataset, only images and videos are selected, that contain a single UAV, as a different approach is necessary to track multiple UAVs with a narrow FoV [44].

FRCNN [20], pre-trained on the COCO-dataset, is fine-tuned on the custom training dataset, whereas about 8 % of the data is separated as a validation set and the remaining data as training set. The fine-tuning process is conducted for 40 epochs with a stepwise reduction

TABLE I
Parameters used for fine-tuning process of the selected object detection algorithm.

Algorithm	Learning rate	Weight decay	Momentum
FRCNN	0.0009	0.0007	0.9

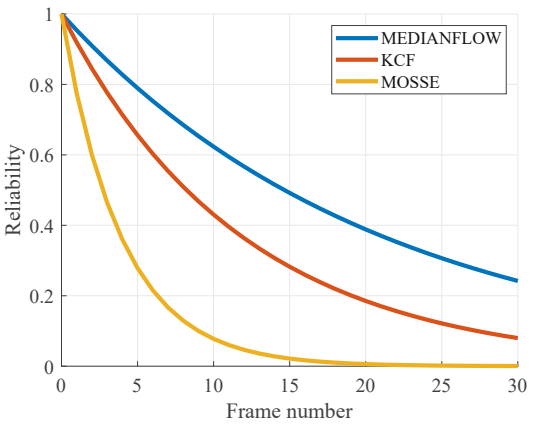
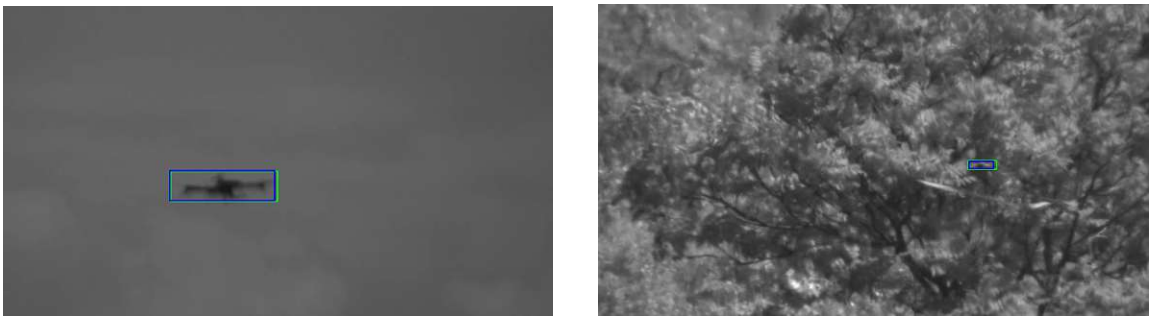


Fig. 4. The reliability of the selected trackers, which can be interpreted as the probability that the algorithm is still correctly tracking the object after a certain number of frames has passed.

of the learning rate by a factor of 0.1 after 30 and 35 epochs respectively. The remaining hyperparameters for fine-tuning are shown in Table I. Additional data augmentation via horizontal flipping of the images further extends the size of the dataset to prevent over-fitting [45]. The models are compared after each epoch and the best model according to the intersection over union (IOU) with an overlap threshold of 50 % is exported to be used for inference. During the training process the best performing model is exported at epoch 24 and achieved a mAP(0.5) of 88.8 % on the validation dataset.

B. Tracker selection

In order to configure the object tracker reliability according to Eq. 1, twelve videos, six with a clear and six with a complex background, containing 6.554 frames, are used. Each tracker is applied to the video sequences and the failure rate, meaning when the IOU between tracker output and ground truth label is below 10 %, is recorded [41]. Upon occurrence of such a track failure, the tracker is automatically reinitialized via the ground truth label to the next frame. The number of track failures is used to calculated the normalized track failure rate p and together with Eq. 1, Fig. 4 is obtained for the KCF, MEDIANFLOW and MOSSE trackers. This calculated reliability based on Eq. 1 is used within the reliability-based strategy to decide whether to reinitialize the tracker with a new detected object or not.



(a) (b)

Fig. 5. Two example images showing a drone in front of a "clear" (a) and "complex" (b) background. The green bounding box depicts the current detector output, while the blue bounding box represents the tracker.

TABLE II
Average time needed by each tracker for processing a single frame.

Algorithm	Processing time
MEDIANFLOW	6.7 ms
KCF	7.8 ms
MOSSE	1.1 ms

TABLE III
Achievable frame rate of each architecture approach.

Algorithm	Frame rate
Proposed approach	100 fps
Detector-based	100 fps
Tracker-based	102 fps
Detector-only	21 fps

Examining the time each tracker needs to process a frame in Table II, all three trackers prove to be suitable for the task, as the Moment camera is operated at a frame rate of 100 fps. However, the MEDIANFLOW tracker proves to be the most reliable from the tested trackers, as it maintains a higher reliability over time compared to the other trackers as seen in Fig. 4.

IV. Experiments and Results

For the experimental evaluation a test dataset is prepared apart from the mentioned training dataset. The dataset consists of videos taken from the Drone vs Bird challenge, videos captured with the presented telescope setup and simulated videos. The latter ones are generated by blending an image of a drone into a video and simulating its flight trajectory. The test video sequences are accounting for about 52.000 frames with a mean bounding box size of 132 x 55 pixels. The test dataset is categorized for the experiments into two different categories. *"Clear"* contains video sequences with images of a UAV in front of a mostly clear background, consisting of blue sky or an evenly overcast cloud cover as seen in Fig. 5a. The second category, *"complex"*, contains videos, where the UAV is in front of a complex background like trees, buildings or scattered and high-contrast clouds as in Fig. 5b. The experimental data only contains video sequences during daytime conditions.

A. Architecture evaluation

For the evaluation of the architecture the achievable frame rates, the IOU and the center location offset (CLO) are used as metrics [40]. The IOU metric gives a good estimate of how accurate the predicted bounding box represents the actual ground truth in terms of size and overlap and for the application an IOU of 0.5 is considered a successful object localization. The CLO on the other hand measures the Euclidean distance between the centers of the predicted and the ground truth bounding box. This metric shows how accurate the algorithms are locating the object, which is an important metric when trying to actuate and follow a UAV with a telescope-based system [40]. Using the IOU of 0.5 and the mean bounding box size of the test dataset, a CLO of 44 pixel is calculated and considered a successful object localization.
To evaluate the proposed reliability-based strategy, two additional transition strategies are implemented [39]. The first, a detector-based strategy, reinitializes the tracker on every single detection by the deep learning algorithm. The second, a tracker-based strategy, uses the detector for an initialization of the tracker and follows the tracker output until the tracker fails, which occurs, for example, when the correlation response of a track falls below a predefined threshold. Upon tracking failure, the detector reinitializes the tracker and the process continues. The proposed reliability-based strategy, as stated in Section II, combines tracker and detector via the reliability and confidence and therefore, decides whether a tracker reinitialization is necessary based on the reported probabilities. Apart from the transition strategies, a detector-only approach is also evaluated, which consists only of the deep learning detector without a parallel running object tracker.

The analysis of the parallel architecture shows the main advantage, which is the fast processing speed.

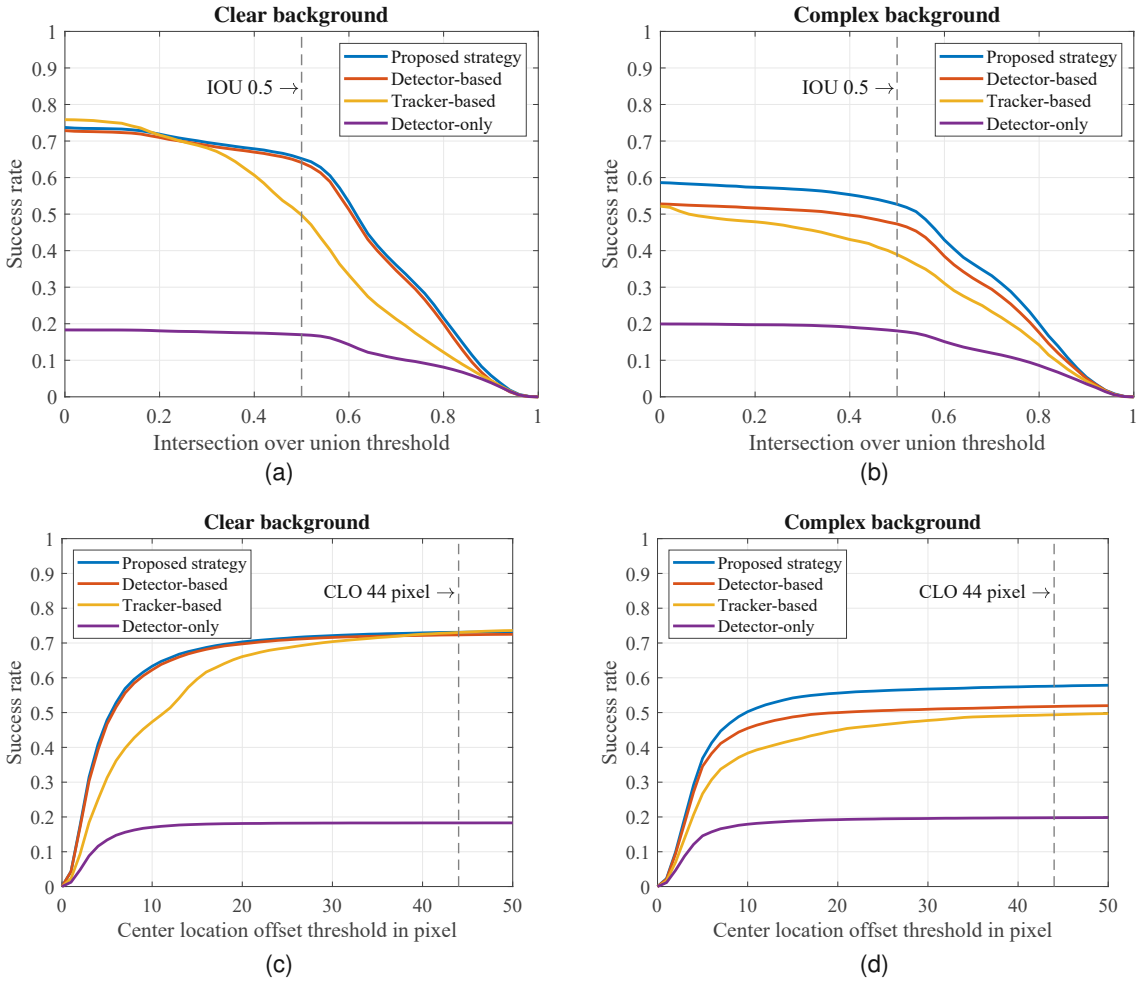


Fig. 6. IOU and CLO of applying the different transition strategies on the two test datasets at 100 fps with a "clear" and "complex" background. The success rate shows the percentage of correct detections and tracks for the corresponding metric threshold respectively.

Table III summarizes the frame rates of the different transition strategies. Using the detector-only strategy without any parallelization, a frame processing speed of 21 fps is achieved, which can still be considered as real-time. However, in an application like the tracking of fast and agile UAVs, it is desirable to acquire many position measurements of the target UAV in a short amount of time. The presented parallel architecture offers an improvement by a factor of 5, as the system is capable to provide the UAV position camera frames at 100 fps, which corresponds to the maximum frame rate of the Moment camera at the specified resolution. The tracker-based approach is negligibly faster than the other two methods, as it requires less tracker reinitializations and is therefore, limited by the speed of the camera.

The results of the application of the different transition strategies on the two test video datasets are depicted in Fig. 6 with the success rate showing the probability of a tracked or detected bounding box satisfying a given IOU or CLO threshold. To obtain these results, the test videos are fed frame by frame into the double frame buffer at a frame rate of 100 fps. As mentioned, FRCNN is used as the object detector, which initializes the MEDIANFLOW tracker and the results are evaluated based on both tracker and detector output. Evident for all investigated scenarios, the non-parallel detector-only approach achieves the worst results, as it does not process most of the frames due to the low achievable frame rate as seen in Table III. Fig. 6a and 6c depict the IOU and the CLO, when applied to the dataset with clear background. The results show, that the proposed and the detector-based strategy score almost equally, while the tracker-based solution performs poorly. The first two transition strategies achieve similar results, as the dataset containing the clear background leaves little room for the detector to make erroneous assumptions of potential UAV locations and therefore, when a new

TABLE IV
The IOU and CLO from Fig. 6 evaluated at an IOU threshold of 0.5 and a CLO threshold of 44 pixels. The latter threshold is determined by the mean bounding box size of the test dataset and an IOU of 0.5. The best results are displayed in bold.

	Proposed approach	Detector-based	Tracker-based	Detector-only
IOU clear	**0.65**	0.64	0.50	0.17
IOU complex	**0.53**	0.47	0.39	0.18
CLO clear	**0.73**	0.72	**0.73**	0.18
CLO complex	**0.58**	0.52	0.49	0.20

detected object is reported, the detector confidence is high. As a consequence, for both transition strategies, an almost equal amount of tracker reinitializations is reported. The tracker-based transition strategy underperforms, as it solely relies on the tracker, which might lose the drone and remain tracking some proportion of background incorrectly. As the detector does not correct the tracker until it fails, the resulting evaluations shows a degraded performance compared to the other transition strategies.

Fig. 6b and 6d depict the results of video sequences, where the UAV is flying in front of a complex background. Again, the tracker-based strategy is severely outperformed by the other transition strategies due to the same reasons as before. In contrast to the previous example, the proposed reliability-based strategy outperforms the detector-based one. In the case of a complex background, always trusting the detector results in a degraded performance, as every miss-detection causes a tracker reinitialization. Using the proposed reliability-based strategy, the tracker is only initialized via a detected object with a higher confidence than the current tracker reliability. This reduces the amount of tracker reinitializations, as detections with a lower confidence than the current tracker reliability, are ignored. Table IV evaluates the probability distributions shown in Fig. 6 at an IOU threshold of 0.5 and CLO threshold of 44 pixels. For a clear background, the proposed and the detector-based strategy score equally well both outperforming the tracker-based strategy in terms of IOU with about $14\,\%$. For a complex background the proposed reliability-based strategy outperforms the detector-based strategy in terms of IOU and CLO by $6\,\%$ and the tracker-based strategy by $14\,\%$ and $9\,\%$ respectively.

Finally, a major benefit of the parallel architecture is analysed, which is the improvement of the probability for a UAV flight path estimation, when compared to a non-parallel detector-only approach. A timely flight path estimation is necessary to reduce reaction times of the telescope system and enable correct pan and tilt motions of the mount to keep the UAV within the FoV of the telescope. Considering the worst case, a UAV (e.g. DJI Mavic 3) flying at maximum speed of $21\,\mathrm{m/s}$ horizontally through the telescope FoV remains visible only for a short amount of time depending on the distance to the telescope

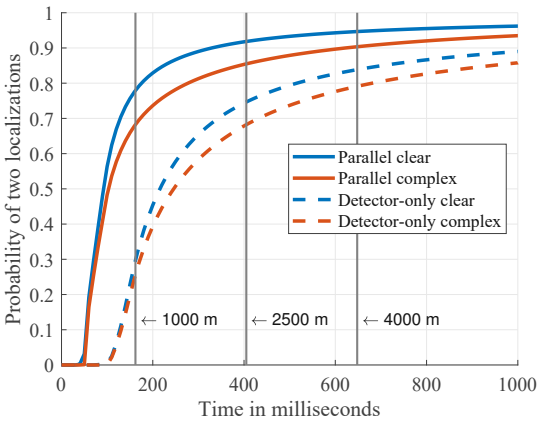


Fig. 7. The probability of two correct UAV localizations, which are necessary for a flight path estimation, within a given time span. The parallel architecture using the reliability-based strategy is compared to a non-parallel detector-only approach. The vertical lines show the minimum duration a DJI Mavic 3 remains within the FoV of the system at different distances.

system. In a distance of $4000\,\mathrm{m}$ the horizontal FoV of the Meade telescope is $13.6\,\mathrm{m}$, which means the UAV remains visible for $648\,\mathrm{ms}$. In a distance of $1000\,\mathrm{m}$ the FoV is reduced to $3.4\,\mathrm{m}$ and the time the UAV is visible is $162\,\mathrm{ms}$ as depicted in Fig. 7 by the vertical lines. Within this timespan the system should localize the UAV at least two times in order to estimate a flight path and keep the UAV within the FoV by appropriate pan and tilt motions of the telescope mount.

A prerequisite to determine the UAV flight path are at least two successful localizations in two frames. Therefore, the probability of two localization within a certain timespan is evaluated. For the evaluation, video sequences are fed at $100\,\mathrm{fps}$ into the architecture and a UAV localization is considered successful at an IOU threshold of 0.5 or larger. The probability is calculated as the number of video sequences, where two successful localizations are achieved, compared to the total number of video sequences.

Fig. 7 shows the results of comparing the reliability-based approach, which is a parallel architecture combining a detector and tracker, to a non-parallel and therefore less complex detector-only approach. In contrast to the previously introduced detector-based strategy, for the detector-only approach, no tracker is running in parallel and object localizations are determined solely by a detector. For both test datasets, ”clear” and ”complex”, the parallel architecture outperforms the detector-only architecture, in terms of localization probability. Considering the mentioned worst case example of a DJI Mavic 3 flying at maximum speed of $21\,\mathrm{m/s}$ horizontally through the telescope FoV, the vertical lines show the time the UAV remains within the FoV of the system for various distances. In a distance of $1000\,\mathrm{m}$ the UAV remains for $162\,\mathrm{ms}$ within the FoV

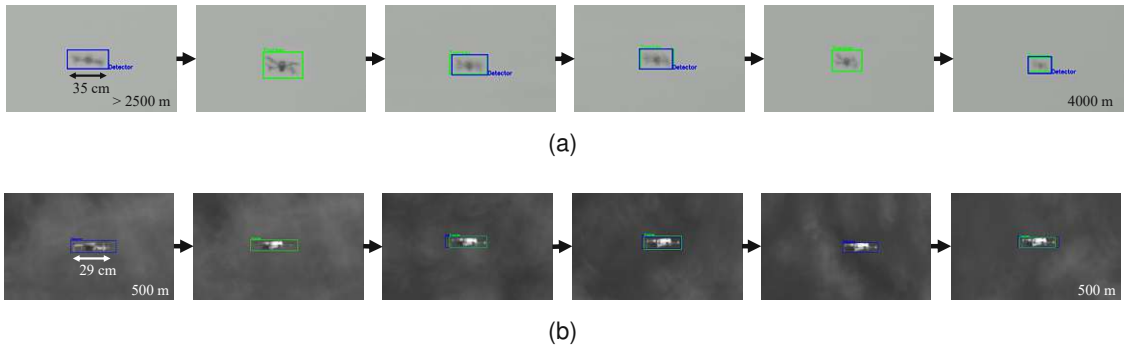


Fig. 8. Two example UAV tracks captured with the proposed telescope system showing a track of the DJI Mavic 3 in front of a clear background in distances between 2500 m and 4000 m. (a) and the DJI Mini 2 in front of a complex background in a distance of 500 m (b). The blue and green boxes visualize the detector and tracker bounding box respectively.

and the probability for two localizations using the parallel architecture is 78 % compared to 29 % of the detector-only approach for the "clear" dataset comparing the solid and dashed blue lines in Fig. 7. For longer distances, e.g. 4000 m, the UAV remains visible 648 ms and the probability increases to 95 % and 84 % for parallel and detector-only approach respectively for the "clear" case. For the "complex" test dataset, depicted by the solid and dashed red lines in Fig. 7, similarly, the parallel architecture achieves better results than the detector-only approach.

Apart from the evaluation using the test video sequences, field tests are conducted demonstrating the capabilities of the proposed architecture and the telescope system. The field tests are performed during daytime conditions in a rural area with mostly forest and meadows in the background, but also some buildings. For the tests a DJI Mavic 3 and a DJI Mini 2 are utilized as UAVs. The UAV is tracked in front of clear and complex background as depicted in Fig. 8 with the blue and green bounding boxes showing the detector and tracker output respectively. The UAVs are flying up to the maximum speed of 21 m/s for the DJI Mavic 3 and 16 m/s for the DJI Mini 2. During the field tests, the distance of the UAV with respect to the telescope system is determined through the UAVs internal global navigation satellite system (GNSS) module. Different dynamic flight trajectories are tested with an emphasis on tangential movement with respect to the telescope perimeter, as the focus is adjusted manually for the field tests. Tested trajectories include the UAV flying at maximum speed in horizontal and vertical direction, while being tracked by the telescope. A second scenario involves the UAV entering the FoV of the telescope with maximum speed to test the automatic detection and tracking of a new object within the FoV. Random flight trajectories are tested, whereas the pilot controls the UAV at will up to the maximum UAV speeds and accelerations. Finally, as depicted in Fig. 8a the UAV is gradually flying away from the system, while the focus is being adjusted manually, to determine the maximum detection distance of 4000 m.

In summary, the proposed reliability-based transition strategy outperforms the pure detector- and tracker-based strategies by 6 % and 14 % in terms of IOU in a scenario with a complex background. Furthermore, the implemented parallel architecture allows object detection and tracking at 100 fps, which improves the probability of two UAV localizations, and therefore, the probability of a flight path estimation. The proposed reliability-based parallel architecture improves the probability for two localizations by 49 %, when compared to a detector-only architecture, given a UAV is visible for at least 162 ms within the FoV of the telescope system.

V. Conclusion

A telescope-based UAV detection and tracking system has been developed, which combines FRCNN, an accurate deep learning algorithm, with MEDIANFLOW, a fast object tracker, to enable real-time UAV detection and tracking over very long distances. The two algorithms collaborate and the decision, if a reinitialization of the tracker is necessary is based on the detection probability and tracker reliability. The presented system allows to track UAVs at 100 fps and outperforms a detector- and tracker-based strategy by 6 % and 14 % in terms of IOU metric for complex backgrounds. Additionally, if the UAV is visible for at least 162 ms within the FoV of the camera, the parallel reliability-based architecture outperforms a non-parallel detector-only approach by 49 % for obtaining at least two localizations to enable flight path prediction. Furthermore, field tests have been conducted with the presented architecture and the telescope system demonstrating UAV detection and tracking capabilities up to a distance of 4000 m with a frame rate of 100 fps in front of a clear background. Future work will consist of integrating a second camera and telescope with a larger FoV to the system together with a corresponding detector and tracker to enable multi-FoV object detection and tracking.

REFERENCES

[1] C. F. Liew, D. DeLatte, N. Takeishi, and T. Yairi, "Recent developments in aerial robotics: A survey and prototypes overview," *ArXiv*, vol. abs/1711.10085, 2017.

[2] BBC News, "East midlands airport closes runway in new drone alert," 2022, accessed April 2022. [Online]. Available: https://www.bbc.com/news/uk-england-leicestershire-617694890

[3] C. Phillips and C. Gaffey, "Most French Nuclear Plants 'Should Be Shut Down' Over Drone Threat," Newsweek Magazine, 2015, accessed Feb 2022. [Online]. Available: http://europe.newsweek.com/most-french-nuclear-plants-should-be-shut-down-over-drone/-threat-309019

[4] S. Dinan, "Mexican drug cartels using drones to smuggle heroin, meth, cocaine into U.S." The Washington Times, 2015, accessed Feb 2022. [Online]. Available: https://www.washingtontimes.com/news/2017/aug/20/mexican-drug-cartels-using-drones-to-smuggle-heroi/

[5] "Charges over drone drug smuggling into prisons," BBC News, 2018, accessed Feb 2022. [Online]. Available: https://www.bbc.com/news/uk-england-43413134

[6] A. D. De Quevedo, F. I. Urzaiz, J. G. Menoyo, and A. A. Lopez, "Drone Detection and RCS Measurements with Ubiquitous Radar," in *2018 International Conference on Radar (RADAR)*, 2018, pp. 1–6.

[7] K. Kang, J. Choi, B. Cho, J. Lee, and K. Kim, "Analysis of micro-doppler signatures of small UAVs based on doppler spectrum," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 5, pp. 3252–3267, Oct. 2021.

[8] S. Yang, H. Qin, X. Liang, and T. Gulliver, "An Improved Unauthorized Unmanned Aerial Vehicle Detection Algorithm Using Radiofrequency-Based Statistical Fingerprint Analysis," *Sensors*, vol. 19, no. 2, p. 274, Jan. 2019.

[9] V. Baron, S. Bouley, M. Muschinowski, J. Mars, and B. Nicolas, "Drone localization and identification using an acoustic array and supervised learning," in *Artificial Intelligence and Machine Learning in Defense Applications*. 11169F SPIE, Sep. 2019.

[10] D. Ojdanić, A. Sinn, C. Naverschnigg, and G. Schitter, "Feasibility Analysis of Optical UAV Detection Over Long Distances Using Robotic Telescopes," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–10, 2023.

[11] Aselsan, "IHTAR anti-drone system - datasheet," ASELSAN A.S., Ankara, Türkiye, 2018.

[12] J. Farlik, M. Kratky, J. Casar, and V. Stary, "Multispectral Detection of Commercial Unmanned Aerial Vehicles," *Sensors*, vol. 19, no. 7, p. 1517, Mar. 2019.

[13] H. U. Unlu, P. S. Niehaus, D. Chirita, N. Evangeliou, and A. Tzes, "Deep learning-based visual tracking of UAVs using a PTZ camera system," in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, Oct. 2019.

[14] R. Carnie, R. Walker, and P. Corke, "Image processing algorithms for UAV "sense and avoid"," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.* IEEE, 2006.

[15] D. Dey, C. Geyer, S. Singh, and M. Digioia, "A cascaded method to detect aircraft in video imagery," *The International Journal of Robotics Research*, vol. 30, no. 12, pp. 1527–1540, aug 2011.

[16] J. W. McCandless, "Detection of aircraft in video sequences using a predictive optical flow algorithm," *Optical Engineering*, vol. 38, no. 3, p. 523, mar 1999.

[17] D. Avola, L. Cinque, G. L. Foresti, C. Massaroni, and D. Pannone, "A keypoint-based method for background modeling and foreground detection using a PTZ camera," *Pattern Recognition Letters*, vol. 96, pp. 96–105, sep 2017.

[18] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *https://arxiv.org/abs/2004.10934*, 2020.

[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016, pp. 21–37.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015.

[21] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[22] H. Liu, K. Fan, Q. Ouyang, and N. Li, "Real-Time Small Drones Detection Based on Pruned YOLOv4," *Sensors*, vol. 21, no. 10, p. 3374, may 2021.

[23] B. K. S. Isaac-Medina, M. Poyser, D. Organisciak, C. G. Willcocks, T. P. Breckon, and H. P. H. Shum, "Unmanned Aerial Vehicle Visual Detection and Tracking using Deep Neural Networks: A Performance Benchmark," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. IEEE, oct 2021.

[24] J. James, J. J. Ford, and T. L. Molloy, "Learning to Detect Aircraft for Long-Range Vision-Based Sense-and-Avoid Systems," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4383–4390, oct 2018.

[25] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking Without Bells and Whistles," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2019.

[26] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8971–8980.

[27] D. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2010.

[28] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, mar 2015.

[29] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4310–4318.

[30] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-Backward Error: Automatic Detection of Tracking Failures," in *2010 20th International Conference on Pattern Recognition*. IEEE, aug 2010.

[31] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision ," *IJCAI*, vol. 81, pp. 674–679, 04 1981.

[32] P. A. Prates, R. Mendonça, A. Lourenço, F. Marques, J. P. Matos-Carvalho, and J. Barata, "Vision-based uav detection and tracking using motion signatures," in *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*, 2018, pp. 482–487.

[33] J. Li, D. H. Ye, M. Kolsch, J. P. Wachs, and C. A. Bouman, "Fast and robust uav to uav detection and tracking from video," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 3, pp. 1519–1531, 2022.

[34] S. Huh, S. Cho, Y. Jung, and D. H. Shim, "Vision-based sense-and-avoid framework for unmanned aerial vehicles," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 4, pp. 3427–3439, 2015.

[35] L. Mejias, S. McNamara, J. Lai, and J. Ford, "Vision-based detection and tracking of aerial targets for uav collision avoidance," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 87–92.

[36] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, sep 2016.

[37] R. Opromolla and G. Fasano, "Visual-based obstacle detection and tracking, and conflict detection for small uas sense and avoid," *Aerospace Science and Technology*, vol. 119, p. 107167, 2021.
[38] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv:1804.02767*, 2018.
[39] D.-H. Lee, "CNN-based single object detection and tracking in videos and its application to drone detection," *Multimedia Tools and Applications*, vol. 80, no. 26-27, pp. 34 237–34 248, oct 2020.
[40] H. Fan and H. Ling, "Parallel Tracking and Verifying," *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 4130–4144, aug 2019.
[41] L. Cehovin, A. Leonardis, and M. Kristan, "Visual object tracking performance measures revisited," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1261–1274, 2016.
[42] Tsung-Yi Lin et. al., "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. [Online]. Available: http://arxiv.org/abs/1405.0312
[43] Angelo Coluccia et. al., "Drone vs. Bird Detection: Deep Learning Algorithms and Results from a Grand Challenge," *Sensors*, vol. 21, no. 8, p. 2824, apr 2021.
[44] N. R. Gans, G. Hu, and W. E. Dixon, "Keeping multiple objects in the field of view of a single ptz camera," in *2009 American Control Conference*, 2009, pp. 5259–5264.
[45] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, jul 2019.

Denis Ojdanić is a doctoral researcher at the Automation and Control Institute (ACIN), group for Advanced Mechatronic Systems (AMS) at TU Wien, Vienna, Austria. He received his MSc. degree in electrical engineering from TU Wien in 2019.
His primary research interests are object detection, tracking and identification using high-performance telescope systems.

Christopher Naverschnigg is a doctoral researcher at the Automation and Control Institute (ACIN), group for Advanced Mechatronic Systems (AMS) at TU Wien, Vienna, Austria. He received his MSc. degree in electrical engineering from TU Wien in 2021.
His primary research interests are modelling and contol of mechatronic systems, especially high-performance telescope systems.

Andreas Sinn is a doctoral researcher at the Automation and Control Institute (ACIN), group for Advanced Mechatronic Systems (AMS) at TU Wien, Vienna, Austria. He received his MSc. degree in electrical engineering from TU Wien in 2016.
His primary research interests are adaptive optics and system integration for optical ground stations, as well as high-performance telescope systems for object tracking and identification.

Daniil Zelinskyi received the B.Sc. degree in mechanical engineering from TU Wien, Vienna, Austria, in 2021. He is a student research assistant at the Automation and Control Institute (ACIN), a group for Advanced Mechatronic Systems, TU Wien. He is currently working toward the M.Sc. degree in electrical engineering at TU Wien. His primary research interests include optics and laser ranging using high-performance telescope systems.

Georg Schitter (Senior Member, IEEE) received an M.Sc. degree in electrical engineering from TU Graz, Austria in 2000, and an M.Sc. degree in information technology and Ph.D. degree in electrical engineering from ETH Zurich, Switzerland, in 2004. He is a Professor for Advanced Mechatronic Systems at the Automation and Control Institute (ACIN) of TU Wien, Vienna, Austria.
His primary research interests are on high-performance mechatronic systems, particularly for applications in the high-tech industry, scientific instrumentation, and mechatronic imaging systems, such as AFM, scanning laser and LIDAR systems, AR HUD, robotic telescope systems, adaptive optics, 3D printing, and lithography systems for semiconductor industry.
Dr. Schitter received the journal best paper award of IEEE/ASME Transactions on Mechatronics (2017), of the IFAC Mechatronics (2008-2010), of the Asian Journal of Control (2004–2005), and the 2013 IFAC Mechatronics Young Researcher Award. He served as an Associate Editor for IFAC Mechatronics, Control Engineering Practice, and for the IEEE TRANSACTIONS ON MECHATRONICS.

Publication [P4]

Algorithm evaluation for parallel detection and tracking of UAVs

Authors: D. Ojdanić, C. Naverschnigg, A. Sinn, and G. Schitter

Reference:

[P4] D. Ojdanić, C. Naverschnigg, A. Sinn, and G. Schitter, "Algorithm evaluation for parallel detection and tracking of UAVs," In Optics, Photonics, and Digital Technologies for Imaging Applications VIII, vol. 12998, pp. 327-333. SPIE, 2024.

Algorithm evaluation for parallel detection and tracking of UAVs

Denis Ojdanić, Christopher Naverschnigg, Andreas Sinn, and Georg Schitter

Automation and Control Institute (ACIN), TU Wien, Gusshausstrasse 27-29, Vienna, Austria

1. ABSTRACT

This paper presents the evaluation of object detectors and trackers within a parallel software architecture to enable long distance UAV detection and tracking in real-time using a telescope-based optical system. The architecture combines computationally expensive deep learning-based object detectors with traditional object trackers to achieve a detection and tracking rate of 100 fps. Four object detectors, FRCNN, SSD, Retinanet and FCOS, are fine-tuned on a custom UAV dataset and integrated together with three trackers, Medianflow, KCF and MOSSE, into a parallel software architecture. The evaluation is conducted on a separate set of test images and videos. The combination of FRCNN and Medianflow shows the best performance in terms of intersection over union and center location offset on the video test set, enabling detection and tracking of UAVs at 100 fps.

Keywords: Deep learning, detection, tracking, real-time, UAV

2. INTRODUCTION

The usage of unmanned aerial vehicles (UAVs) has increased drastically over the past decade due the relatively cheap availability and versatility of the technology.[1] However, safety concerns and safety related incidents have soared equally in the past, as numerous examples of near or actual collisions with commercial air planes in various countries demonstrate.[2,3] Another perilous situation includes a near collision with an air ambulance in a height of 400 m.[4] Similarly, UAVs offer a menacing potential to endanger critical infrastructure like nuclear power plants[5] and prove useful to smuggle drugs across state borders and into prisons.[6] As these examples illustrate the emanating danger posed by UAVs, deployment of appropriate detection systems is crucial to ensure timely threat reconnaissance.

Generally, UAV detection systems combine multiple sensors into holistic systems, which consist of RADAR,[7] LiDAR,[8] acoustics,[9] radio frequency (RF)[10] and electro-optics (EO) sensors.[11] The latter is a key component, as it allows a profound situational evaluation through visual imagery. EO systems have a relatively narrow field of view (FOV) camera, which are typically attached to a pan and tilt mount to enable monitoring a larger area. Utilization of telescope systems can further increase the operational range to multiple kilometres for small UAVs.[11] These EO systems rely on efficient computer vision algorithms to extract the UAV position within video frames to steer the mount and keep the UAV within the FoV.

Over the past decade deep learning based methods have proven to be the most promising approach to detect and classify objects in challenging imagery. Algorithms like YOLO,[12] SSD,[13] Retinanet,[14] FRCNN[15] or FCOS[16] are prominent examples for state of the art object detection. However, neural networks come at a high processing cost and therefore, are limited in the achievable frame rates. A high frame rate and the resulting high number of UAV localizations per second is an important prerequisite for a dynamic pan and tilt system, which has to keep up with the fast and agile movements of a UAV.[17] Combinations of neural networks with faster object tracking algorithms in a parallel manner strive to benefit from both, high accuracy of the neural networks and the fast processing speeds of tracking algorithms.[18] Fast tracking algorithms include Mosse,[19] KCF,[20] MedianFlow[21] or CSRT.[22] Combining neural networks with Kalman Filters offers improved frame rates too, however, compared to the trackers, a filter only estimates the position between two detections.[23] Various studies exist comparing object tracking and object detection algorithms for the use case of UAV detection.[24–26] To enable an efficient detection and tracking of UAVs at high frame rates, a thorough analysis and comparison of different detector

Further author information: Send correspondence to Denis Ojdanić
E-mail: ojdanic@acin.tuwien.ac.at, Telephone:+43 (0) 1 58801 376 520

and tracker combinations within a parallel architecture is required.
The contribution of this paper is the evaluation and comparison of various deep learning based object detectors and traditional object trackers within a parallel architecture to enable detection and tracking of UAVs at 100 fps. A high number of UAV localizations enables a dynamic actuation of a pan-tilt mount to track the fast and agile movements of the UAV.

3. OBJECT DETECTION AND TRACKING

To implement a system detecting UAVs at 100 fps a parallel software architecture is used, where the detector and tracker are running within separate threads.[27] The slow, but accurate deep learning based detector initializes a fast, but less accurate object tracker to achieve a high frame rate, while maintaining a high accuracy.[27] This combination of a deep learning object detector with a fast object tracker enables detection and tracking of objects at 100 fps. For object detection FRCNN, FCOS, RetinaNet and SSD are selected to be trained and evaluated as they cover both single and dual stage network architectures as well as anchor box and anchor box free approaches. For object tracking, MedianFlow, Mosse and KCF are used, as these algorithms have low processing demands, which makes them perfect candidates to be paired with a deep learning algorithm in parallel to increase the UAV localization frame rate.

3.1 Training

In order to train the four deep learning object detectors a pre-labelled training dataset consisting of approximately 18000 images is used.[27] The dataset contains images of different quad-copter UAVs in front of various backgrounds. The distribution of the bounding boxes is depicted in Fig. 1, which illustrates that the UAVs within the dataset cover a small pixel area with respect to the image size. Nevertheless, to further decrease the number of false positives during inference, about 8 % of pure background images, which do not contain any UAV are added to the dataset extending it to about 20500 images. These background images do contain other flying objects, like air planes or birds, in order to avoid erroneous detections during inference. 8 % of the 18000 images, which contain UAVs, serve as validation dataset and the remaining images with all the added background images are used for training.

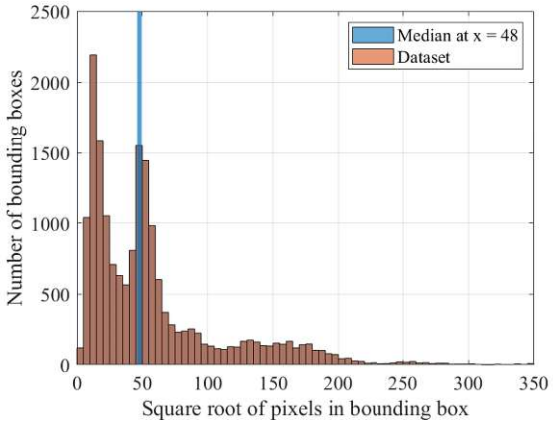


Figure 1: Distribution of the bounding box area in pixels within the dataset. Note that the pure background images are not depicted within this plot, as these images have no bounding box.

As a training strategy for the neural networks, fine-tuning is applied, whereas the networks are initialized with the weights trained on the COCO dataset.[28] The training and the experiments in Section 4 are conducted on an RTX 3080 (Nvidia Corporation, Santa Clara, California, USA) with 10 GB of GPU RAM. Additionally, the PC has an AMD Ryzen 3900 CPU (Advanced Micro Devices, Inc., Santa Clara, California, USA) with 12

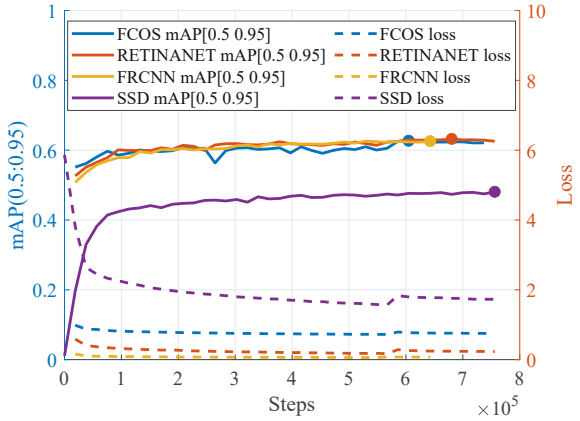


Figure 2: Overview of the training process showing the loss curves and the mAP(0.5:0.95) on the validation dataset for each deep learning object detector. The dots on the validation curve indicate the best performing model, which is later used for inference.

Table 1: Hyper-parameters for fine-tuning of the deep learning object detectors.

Algorithm	Learning rate	Weight decay	Momentum	Batch size
SSD	0.0001	0.0006	0.6	8
FCOS	0.0005	0.0007	0.8	4
FRCNN	0.0002	0.0007	0.8	2
RetinaNet	0.0003	0.0007	0.8	4

cores, 24 threads and 32 GB of RAM installed.
The algorithms are trained for approximately 40 epochs and the further hyper-parameters are summarized in Table 1. To diversify the dataset, data augmentation techniques are applied during the training process such as random horizontal flipping and color transformations of the images including greyscale transformation or changing the brightness, contrast, saturation and hue.[29] Furthermore, random cropping is applied after epoch 30, as activating the cropping earlier has led to an exploding loss. Fig. 2 depicts the loss and the mean average precision (mAP) on the validation dataset during the training process. A slight increase of the loss is visible for all models after random cropping is activated. The best model according to a mAP of 0.5 to 0.95 is saved during the training process and the dots on the validation curves represent the best performing models, which are selected for the experiments.

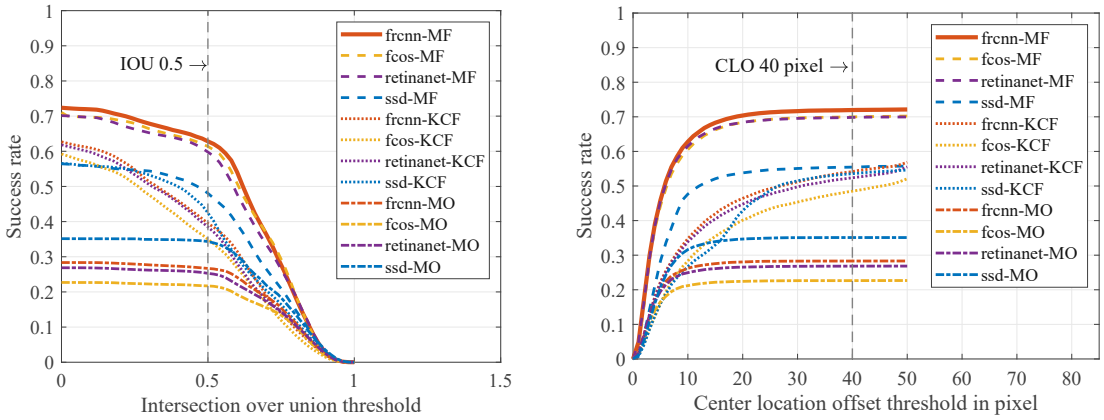
4. EXPERIMENTS AND RESULTS

For experimental evaluation a test video dataset of approximately 25000 video images of UAVs in front of different backgrounds is used.[11] Most of the images are captured using telescope systems. Additionally, of each video sequence every fifth image is extracted to form a test image dataset of 5000 images to analyse the mAP of the detectors.
Table 2 shows the results of applying the four deep learning object detectors onto the test image dataset. FRCNN scores the highest mAP(0.5:0.95) of 40.7 % closely followed by FCOS at 37.9 % and RetinaNet at 35.6 %. SSD, as the fastest and least complex algorithm, achieves 34.2 %.

To evaluate the combined results of detector and tracker running in parallel, the test videos serve as input at a frame rate of 100 fps. To achieve this detection and tracking frame rate, the slow object detectors are combined with fast object trackers.[27] To verify, whether the algorithms are localizing the object correctly, the detector and tracker output are merged based on their prediction probability.[27] The intersection over union (IOU) and

Table 2: mAP for the deep learning algorithms.

Algorithm	mAP(0.5)	mAP(0.5:0.95)
FRCNN	88.7 %	40.7 %
FCOS	89.6 %	37.9 %
RetinaNet	85.6 %	35.6 %
SSD	75.9 %	34.2 %



(a) IOU for all detector and tracker combinations.

(b) CLO for all detector and tracker combinations.

Figure 3: The IOU and CLO of all algorithm combinations on the test video dataset.

the center location offset (CLO) serve as evaluation metrics,[30] whereas an IOU of 0.5 or larger is considered as successful and and CLO of 40 pixels or less.

Fig. 3 depicts the results of applying all detector and tracker combinations to the video dataset and Table 3 summarizes the IOU at a threshold of 0.5 and the CLO at a pixel threshold of 40. The best combined performance is achieved by FRCNN and the MedianFlow tracker, which score an IOU of 62.8 % and an CLO of 72 % at a threshold of 0.5 and 40 pixels. The latter tracking algorithm also achieves the best results when compared to the other object trackers. Among the detectors, SSD scores the lowest results, when combined with MedianFlow, which ensues from the evaluation on the image test set as presented in Table 3. However, in combination with lower performing trackers like KCF or Mosse, it performs better compared to the other deep learning detectors. SSD consists of a simpler network architecture and therefore, processes approximately twice as many images per second as the other deep learning detectors. For the less reliable trackers, more frequent reinitialisations as facilitated by SSD, result in an higher combined performance compared to the other detector, when paired with Mosse or KCF. Finally, Fig. 4 shows an example video sequence of a detection and track using the parallel architecture with FRCNN and MedianFlow.

To summarize, among the analysed algorithms, FRCNN and MedianFlow achieve the best performance, when detecting and tracking UAVs at 100 fps scoring an IOU(0.5) of 62.8 %. This enables detection and tracking of UAVs flying at fast velocities in dynamic trajectories.

5. CONCLUSION

An algorithm analysis is presented for a real-time parallel software architecture, which combines accurate deep neural network detectors with traditional trackers to enable UAV localization at 100 fps. FRCNN, RetinaNet, SSD and FCOS are selected as detectors and fine-tuned for the task of UAV detection. The detectors together with three trackers, Medianflow, KCF and Mosse, are analysed on a separate test image and video dataset. An evaluation is performed to determine, which combination of neural network and object tracker yields the best

Table 3: The results of the IOU(0.5) and the CLO(40 pixels) for each algorithm combination. The results are sorted in a descending order according to the IOU(0.5).

Algorithm	IOU(0.5)	CLO(40 px))
FRCNN - MedianFlow	62.8 %	72.0 %
FCOS - MedianFlow	61.4 %	70.0 %
RetinaNet - MedianFlow	59.8 %	69.8 %
SSD - MedianFlow	48.1 %	55.5 %
SSD - KCF	42.5 %	53.4 %
FRCNN - KCF	39.5 %	54.0 %
RetinaNet - KCF	38.5 %	52.2 %
FCOS - KCF	35.0 %	48.3 %
SSD - Mosse	34.3 %	35.1 %
FRCNN - Mosse	26.7 %	28.3 %
RetinaNet - Mosse	25.3 %	26.8 %
FCOS - Mosse	21.7 %	22.6 %

(a) (b) (c) (d)

Figure 4: Image sequence of a detection and track of a UAV. The blue and green bounding boxes indicate the detector the tracker output.

results in terms of IOU at an threshold of 0.5 and the CLO at an pixel threshold of 40. FRCNN and MedianFlow prove to be the best performing algorithm combination by score an IOU of 62.8 % and a CLO of 72 % when detecting and tracking UAVs at a frame rate of 100 fps. By localizing UAVs at 100 fps, a dynamic control of the pan-tilt mount for the telescope is possible, which enables tracking and keeping fast and agile UAVs within the FoV.

ACKNOWLEDGMENTS

This publication is funded by the Austrian defence research programme FORTE of the Federal Ministry of Finance (BMF).
The authors gratefully acknowledge the cooperation with ASA Astrosysteme GmbH and thank for their support and valuable expertise.

REFERENCES

[1] Liew, C. F., DeLatte, D., Takeishi, N., and Yairi, T., "Recent developments in aerial robotics: A survey and prototypes overview," *ArXiv* **abs/1711.10085** (2017).
[2] "Drone collides with commercial aeroplane in canada." BBC News (Oct. 2017). Accessed Feb 2024.
[3] "Drone over windsor came close to british airways plane, report says." BBC News (Dec. 2023). Accessed Feb 2024.
[4] "Drone in 'near-miss' with air ambulance." BBC News (Sept. 2019). Accessed Feb 2024.
[5] Phillips, C. and Gaffey, C., "Most French Nuclear Plants 'Should Be Shut Down' Over Drone Threat." Newsweek Magazine (Feb. 2015). Accessed Feb 2022.
[6] Daly, M., "Cheap and They Don't Snitch: Drones Are the New Drug Mules." VICE (Jan. 2024). Accessed Feb 2022.

[7] Drozdowicz, J., Wielgo, M., Samczynski, P., Kulpa, K., Krzonkalla, J., Mordzonek, M., Bryl, M., and Jakielaszek, Z., "35 GHz FMCW drone detection system," in [*2016 17th International Radar Symposium (IRS)*], 1–4, IEEE (2016).

[8] Dogru, S. and Marques, L., "Drone detection using sparse lidar measurements," *IEEE Robotics and Automation Letters* **7**(2), 3062–3069 (2022).

[9] Mezei, J., Fiaska, V., and Molnár, A., "Drone sound detection," in [*16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*], 333–338, IEEE (2015).

[10] Nguyen, P., Ravindranatha, M., Nguyen, A., Han, R., and Vu, T., "Investigating cost-effective rf-based detection of drones," in [*Proceedings of the 2nd workshop on micro aerial vehicle networks, systems, and applications for civilian use*], 17–22 (2016).

[11] Ojdanić, D., Sinn, A., Naverschnigg, C., and Schitter, G., "Feasibility analysis of optical uav detection over long distances using robotic telescopes," *IEEE Transactions on Aerospace and Electronic Systems* **59**(5), 5148–5157 (2023).

[12] Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M., "Yolov4: Optimal speed and accuracy of object detection," (2020).

[13] Wei, L., Anguelov, D., Erhan, D., et. al., "SSD: Single Shot MultiBox Detector," in [*Computer Vision – ECCV 2016*], 21–37, Springer International Publishing, Cham (2016).

[14] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P., "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(2), 318–327 (2020).

[15] Ren, S., He, K., Girshick, R., and Sun, J., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in [*Advances in Neural Information Processing Systems*], **28**, Curran Associates, Inc. (2015).

[16] Tian, Z., Shen, C., Chen, H., and He, T., "Fcos: Fully convolutional one-stage object detection," (2019).

[17] Naverschnigg, C., Ojdanic, D., Sinn, A., and Schitter, G., "Analysis and Control of a Robotic Telescope System for High-Speed Small-UAV Tracking," *IEEE Transactions on Aerospace and Electronic Systems Magazine* (submitted 12.2023).

[18] Lee, D.-H., "CNN-based single object detection and tracking in videos and its application to drone detection," *Multimedia Tools and Applications* **80**, 34237–34248 (Oct 2020).

[19] Bolme, D., Beveridge, J. R., Draper, B. A., and Lui, Y. M., "Visual object tracking using adaptive correlation filters," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], IEEE (jun 2010).

[20] Henriques, J. F., Caseiro, R., Martins, P., and Batista, J., "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**, 583–596 (Mar. 2015).

[21] Kalal, Z., Mikolajczyk, K., and Matas, J., "Forward-Backward Error: Automatic Detection of Tracking Failures," in [*2010 20th International Conference on Pattern Recognition*], IEEE (aug 2010).

[22] Danelljan, M., Hager, G., Shahbaz Khan, F., and Felsberg, M., "Learning spatially regularized correlation filters for visual tracking," in [*Proceedings of the IEEE international conference on computer vision*], 4310–4318 (2015).

[23] Opromolla, R. and Fasano, G., "Visual-based obstacle detection and tracking, and conflict detection for small uas sense and avoid," *Aerospace Science and Technology* **119**, 107167 (2021).

[24] Park, J., Kim, D. H., Shin, Y. S., and Lee, S.-h., "A comparison of convolutional object detectors for real-time drone tracking using a ptz camera," in [*17th International Conference on Control, Automation and Systems (ICCAS)*], 696–699 (2017).

[25] Oh, H. M., Lee, H., and Kim, M. Y., "Comparing convolutional neural network(cnn) models for machine learning-based drone and bird classification of anti-drone system," in [*19th International Conference on Control, Automation and Systems (ICCAS)*], 87–90 (2019).

[26] Kristan, M., Leonardis, A., Matas, J., Felsberg, M., Pflugfelder, R., Kämäräinen, J.-K., Chang, H. J., Danelljan, M., Zajc, L. Č., Lukežič, A., et al., "The tenth visual object tracking vot2022 challenge results," in [*European Conference on Computer Vision*], 431–460 (2022).

[27] Ojdanić, D., Sinn, A., Naverschnigg, C., Zelinskyi, D., and Schitter, G., "Parallel Architecture for Low Latency UAV Detection and Tracking using Robotic Telescopes," *IEEE Transactions on Aerospace and Electronic Systems* (submitted 07.2023).

[28] Lin, T., Maire, M., Belongie S., et. al., "Microsoft COCO: Common Objects in Context," in [*Computer Vision–ECCV 2014, Proceedings, Part V 13*], 740–755, Springer (2014).

[29] Shorten, C. and Khoshgoftaar, T. M., "A survey on image data augmentation for deep learning," *Journal of big data* **6**(1), 1–48 (2019).

[30] Fan, H. and Ling, H., "Parallel Tracking and Verifying," *IEEE Transactions on Image Processing* **28**, 4130–4144 (aug 2019).

Publication [P5]

Deep learning-based long-distance optical UAV detection: color versus grayscale

Authors: D. Ojdanić, C. Naverschnigg, A. Sinn, and G. Schitter

Reference:

[P5] D. Ojdanić, C. Naverschnigg, A. Sinn, and G. Schitter, "Deep learning-based long-distance optical UAV detection: color versus grayscale," In Pattern Recognition and Tracking XXXIV, vol. 12527, pp. 80-84. SPIE, 2023

Deep Learning Based Long-Distance Optical UAV Detection: Color vs. Grayscale

Denis Ojdanić, Christopher Naverschnigg, Andreas Sinn, and Georg Schitter

Automation and Control Institute (ACIN), TU Wien, Gusshausstrasse 27-29, Vienna, Austria

ABSTRACT

This paper presents a comparison between grayscale and color based deep learning algorithms for long distance optical UAV detection using robotic telescope systems. Three deep learning object detection algorithms are trained with a custom dataset consisting of RGB images and the performance is evaluated against the same algorithms trained with the same dataset converted to grayscale. Network training from scratch and fine-tuning are evaluated. The results for all algorithms show that fine-tuning with RGB images maximizes the detection performance and scores about 5 % better in terms of mean average precision (mAP(0.5)) compared to fine-tuning on grayscale images.

Keywords: UAV detection, deep learning, color, grayscale

1. INTRODUCTION

Mini and micro drones, also known as unmanned aerial vehicles (UAVs), have gained massive popularity in recent years due to their versatility, ease of use and affordability. However, these characteristics paired with their small size, manoeuvrability and ability of carrying payload, make UAVs a modern safety hazard. UAVs have already been involved in several incidents including near collisions with air-planes,[1] unauthorized flyovers in the vicinity of nuclear power plants,[2] the illegal smuggling of goods over state boarders[3] and lately they play a significant role in modern warfare.[4] Given the diverse applicability of UAVs, which include potentially malicious utilization, and numerous accounts of incidents, research and development on proper detection and mitigation technologies is indispensable.

The detection of UAVs is an advanced research area with various approaches being investigated such as RADAR,[5] LiDAR,[6] radio frequency,[7] acoustics[8] and optics.[9] Most state of the art systems combine multiple of these technologies to benefit from the individual strengths forming a multispectral detection system.[10] Electro-optical systems are key components in all multispectral systems to ensure situational awareness, as captured images are easily interpreted. Therefore, systems equipped with cameras for object detection and classification are extensively researched. Optical systems are limited by their operational range and therefore, camera-based systems use narrow field of view (FoV) to extend the detection range. These devices are attached to mounts, which enable pan and tilt motion in order to observe a larger area and to track detected objects. Considering the selection of an appropriate camera, it is known that the achievable resolution, low light capabilities and quantum efficiency of monochrome sensors outperform color camera sensors due to the color filters used in the latter one.[11] Following this argumentation, the usage of monochrome cameras is advisable to improve detection distance.

However, state of the art object detection, which is usually facilitated via deep learning algorithms,[12] is mostly evolving around using color data as input, as the major publicly available datasets, e.g. COCO[13] or ImageNet,[14] which are often used for fine-tuning and transfer learning, suggest. Different studies show, that there is no consensus on whether color or grayscale images improve the object detection performance and analysis needs to be conducted for each specific field of application.[15,16]

The contribution of this paper is the evaluation of three state of the art deep learning based algorithms with respect to their UAV detection performance using either color or grayscale input images.

Further author information: Send correspondence to Denis Ojdanić
E-mail: ojdanic@acin.tuwien.ac.at, Telephone:+43 (0) 1 58801 376 520

2. UAV DETECTION

For UAV detection, three state of the art deep learning algorithms, FRCNN,[17] Retinanet[18] and SSD[19] are selected, which are predominantly used for efficient object detection. These three algorithms are trained once on color images and in a separate instance, on grayscale images, in order to evaluate, whether color or grayscale results in an improved deep learning based UAV detection performance. The main difference between color and grayscale images is the number of channels storing the data. For RGB images, an image contains three channels, e.g. red, green and blue with each channel storing the pixel values for the corresponding color. Grayscale images only contain one channel, which stores the grayscale information. Likewise, the filters of the input layers of deep learning algorithms contain three input channels for the color case. For grayscale images two network adaptions are suitable. First, the three input filter channels of the network can be reduced to one input channel to match the number of input image channels. Another possibility is to triplicate the grayscale input image to a three channel image, using the same input for each "color" channel. While the first approach is negligibly more computationally efficient, the second approach retains the same number of network weights, which allows a fair comparison between color and grayscale networks.

2.1 Dataset

The dataset used for training consists of 5000 images depicting different UAVs during daytime and clear weather conditions. About 50 % of the images are captured through various optical systems ranging from standard cameras to telescopes[9] and the remaining images are generated by blending cropped images of UAVs over different backgrounds. Therefore, the dataset contains images with various degrees of atmospheric blurring and backgrounds. Furthermore, other flying objects, like birds, are visible in the images. All images are captured or simulated by using color cameras and images. To transform the color data to grayscale the following relation is used[20]

$$Gray = 0.299R + 0.587G + 0.114B, \tag{1}$$

with R, G and B representing the red, green and blue channels of the color image. Due to this color to grayscale transformation, it is assumed for the training and test dataset, that properties, like resolution, quantum efficiency and low light performance of a color and mono camera are equal. Fig. 1 depicts the distribution of the bounding box sizes of the training dataset, which indicates, that the majority the UAVs within the dataset are small relative to the image size with the median bounding box size being approximately 85 pixels x 85 pixels.

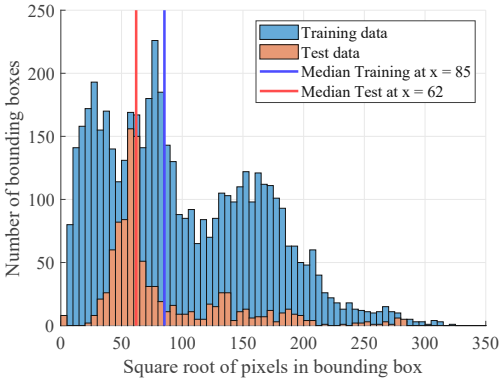


Figure 1. The histogram shows the distribution of the bounding box sizes within the training dataset with the median bounding box size being 85 x 85 pixels. Additionally, the distribution of the test dataset is depicted, which is used for evaluation in Section 3. The median bounding box size for the test dataset is 62 x 62 pixels.

2.2 Training procedure

To evaluate the algorithms two training approaches are applied. One method is training from scratch, where the algorithm weights are randomly initialized and optimized with the UAV dataset presented in Section 2.1. The second method is fine-tuning, whereas the algorithms are initialized with weights pre-trained on the COCO dataset. The COCO dataset[13] is selected, as it contains various classes, which are similar to the task of UAV detection, like "bird" or "plane". For both training procedures, the algorithms are trained for 30 epochs with a stepwise reduction of the learning rate. Table 1 shows the remaining parameters used for the training process. During the training process data augmentation is applied in the form of random flipping.

Table 1. The parameters used for training of each deep learning object detection algorithm.

Algorithm	Learning rate	Weight decay	Momentum
FRCNN	0.0004	0.0006	0.8
Retinanet	0.0009	0.0006	0.8
SSD	0.0007	0.0005	0.8

(a) (b)

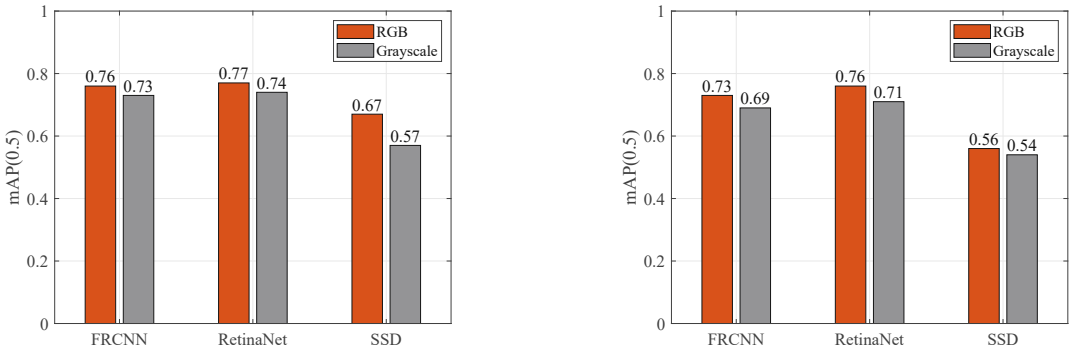
Figure 2. (a) An example color image showing a UAV. (b) The same image converted into grayscale according to Eq. 1.

3. EXPERIMENTS AND RESULTS

For the experimental analysis a test dataset is created, which consists of 1065 labeled images, which are captured during daytime and clear weather conditions. The bounding box size distribution of the test dataset is visible in Fig. 1. An example test image of a UAV is depicted in Fig. 2(a) and using Eq. 1, this image is converted to grayscale as shown in Fig. 2(b). As stated in Section 2.1, the color and test dataset appear to be captured by cameras with equal properties in terms of resolution, quantum efficiency and low light performance, as a consequence of using Eq. 1.

The results of applying the test dataset to the trained models are displayed in Fig. 3. As an evaluation metric the mean average precision (mAP) is used with an overlap threshold of 0.5 Comparing the two training methods, the fine-tuned models outperform the models, which are trained from scratch by an average of about 4 % in terms of mAP(0.5). It appears that the fine-tuned models learned to generalize better, due to being exposed to more data during the training process. Evaluating the results of comparing color to grayscale, the color models outperform the grayscale models by 5 % mAP(0.5) when fine-tuned and by 3 % when trained from scratch. For the case of fine-tuning, the difference between color and grayscale is larger, which appears to be a consequence of the COCO-dataset consisting of color images, which is advantageous for the later color fine-tuning phase.

Summarizing the results, fine-tuning as a training strategy achieves on average 4 % higher mAP(0.5) than training from scratch. Furthermore, the fine-tuned color models achieve the overall best performance, outperforming the fine-tuned grayscale models and the color models trained from scratch by 5 % respectively in terms of mAP.



(a) Fine-tuned models.

(b) Models trained from scratch.

Figure 3. (a) shows the mAP(0.5) for three object detection algorithms fine-tuned on RGB and on grayscale images. (b) depicts the results of the algorithms trained from scratch. The fine-tuned color models show the overall best performance and are therefore, the preferred training strategy.

4. CONCLUSION

Three state of the art deep learning object detection algorithms have been trained for the task of UAV detection and an evaluation has been performed whether color or grayscale images offer an advantage in terms of detection performance. The evaluation is performed using data captured by color cameras during daytime and clear weather conditions. The grayscale transformation implies that color and mono cameras both perform equally in terms of resolution, quantum efficiency and low light performance, when capturing the test data. Based on the evaluation it can be concluded, that models trained with color images outperform models trained with grayscale images for both chosen training strategies. Best overall results are achieved, when fine-tuning the models on color images, as it outperforms the fine-tuned grayscale models and the color models trained from scratch by about 5 % according to the mAP(0.5) respectively.
Future work will focus on a holistic system evaluation, which extends the presented algorithm analysis by an evaluation of different cameras and daytime conditions.

ACKNOWLEDGMENTS

This publication is funded by the Austrian defence research programme FORTE of the Federal Ministry of Finance (BMF).
The authors gratefully acknowledge the cooperation with ASA Astrosysteme GmbH and thank for their support and valuable expertise.

REFERENCES

[1] Serna, J., “Lufthansa jet and drone nearly collide near LAX.” Los Angeles Times (2016). Accessed March 2023.
[2] Phillips, C. and Gaffey, C., “Most French Nuclear Plants ’Should Be Shut Down’ Over Drone Threat.” Newsweek Magazine (Feb. 2015). Accessed Feb 2022.
[3] “Charges over drone drug smuggling into prisons.” BBC News (Mar. 2018). Accessed Feb 2022.
[4] Bi, Z., Chen, H., Hu, J., et. al., “Analysis of UAV Typical War Cases and Combat Assessment Research,” in [*2022 IEEE International Conference on Unmanned Systems (ICUS)*], 1449–1453 (2022).
[5] A. D. De Quevedo, F. I. Urzaiz, J. G. Menoyo, and A. A. Lopez, “Drone Detection and RCS Measurements with Ubiquitous Radar,” in [*2018 International Conference on Radar (RADAR)*], 1–6 (2018).
[6] Ojdanić, D., Gräf, B., Sinn, A., Yoo, H. W., and Schitter, G., “Camera-guided real-time laser ranging for multi-UAV distance measurement,” *Appl. Opt.* **61**, 9233–9240 (Nov 2022).

[7] Yang, S., Qin, H., Liang, X., and Gulliver, T., "An Improved Unauthorized Unmanned Aerial Vehicle Detection Algorithm Using Radiofrequency-Based Statistical Fingerprint Analysis," *Sensors* **19**, 274 (Jan 2019).

[8] Baron, V., Bouley, S., Muschinowski, M., Mars, J., and Nicolas, B., "Drone localization and identification using an acoustic array and supervised learning," in [*Artificial Intelligence and Machine Learning in Defense Applications*], 11169F SPIE (Sep 2019).

[9] Ojdanić, D., Sinn, A., Naverschnigg, C., and Schitter, G., "Feasibility Analysis of Optical UAV Detection Over Long Distances Using Robotic Telescopes," *IEEE Transactions on Aerospace and Electronic Systems* , 1–10 (2023).

[10] Farlik, J., Kratky, M., Casar, J., and Stary, V., "Multispectral Detection of Commercial Unmanned Aerial Vehicles," *Sensors* **19**, 1517 (Mar. 2019).

[11] Chakrabarti, A., Freeman, W. T., and Zickler, T., "Rethinking color cameras," in [*2014 IEEE International Conference on Computational Photography (ICCP)*], 1–8 (2014).

[12] Unlu, H. U., Niehaus, P. S., Chirita, D., Evangeliou, N., and Tzes, A., "Deep learning-based visual tracking of UAVs using a PTZ camera system," in [*IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*], IEEE (Oct. 2019).

[13] Lin, T., Maire, M., Belongie S., et. al., "Microsoft COCO: Common Objects in Context," in [*Computer Vision–ECCV 2014, Proceedings, Part V 13*], 740–755, Springer (2014).

[14] Deng, J., Dong, W., and Socher, R., et. al, "Imagenet: A large-scale hierarchical image database," in [*2009 IEEE Conference on Computer Vision and Pattern Recognition*], 248–255 (2009).

[15] Funt, B. and Zhu, L., "Does colour really matter? Evaluation via object classification," in [*Color and Imaging Conference*], **26**, 268–271, Society for Imaging Science and Technology (2018).

[16] Bui, H. M., Lech, M., Cheng, E., et. al., "Using grayscale images for object recognition with convolutional-recursive neural network," in [*2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*], 321–325 (2016).

[17] Ren, S., He, K., Girshick, R., and Sun, J., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in [*Advances in Neural Information Processing Systems*], **28**, Curran Associates, Inc. (2015).

[18] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P., "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42**(2), 318–327 (2020).

[19] Wei, L., Anguelov, D., Erhan, D., et. al., "SSD: Single Shot MultiBox Detector," in [*Computer Vision – ECCV 2016*], 21–37, Springer International Publishing, Cham (2016).

[20] Kanan, C. and Cottrell, G. W., "Color-to-grayscale: does the method matter in image recognition?," *PloS ONE* **7**(1), e29740 (2012).

Publication [P6]

High-speed telescope autofocus for UAV detection and tracking

Authors: D. Ojdanić, D. Zelinskyi, C. Naverschnigg, A. Sinn, and G. Schitter

Reference:

[P6] D. Ojdanić, D. Zelinskyi, C. Naverschnigg, A. Sinn, and G. Schitter, "High-speed telescope autofocus for UAV detection and tracking," Opt. Express 32, 7147-7157 (2024)

Journal Quartile: Q1 (2024)

High-Speed Telescope Autofocus for UAV Detection and Tracking

DENIS OJDANIĆ,* DANIIL ZELINSKYI, CHRISTOPHER NAVERSCHNIGG, ANDREAS SINN AND GEORG SCHITTER

Automation and Control Institute (ACIN), Technische Universität Wien, 1040 Vienna
**ojdanic@acin.tuwien.ac.at*

Abstract: This paper presents the analysis, implementation and experimental evaluation of a high-speed automatic focus module for a telescope-based UAV detection and tracking system. An existing optical drone detection system consisting of two telescopes and deep learning-based object detection is supplemented by suitable linear stages and passive focus algorithms to enable fast automatic focus adjustment. Field tests with the proposed system demonstrate that UAVs flying at speeds of up to 24 m/s towards the system are successfully tracked and kept in focus from more than 4500 m down to 150 m. Furthermore, different search functions and contrast measures are evaluated and it is shown that the Tenengrad operator combined with the Hill Climbing search function achieve the best performance for focusing on fast moving small UAVs.

1. Introduction

In recent years the necessity of optical detection systems, which are capable of detecting and tracking small, fast and agile unmanned aerial vehicles (UAVs) has become evident as numerous incidents illustrate the growing threat to public safety and infrastructure. A famous example is the shutdown of the London Gatwick airport for more than a day in 2018 due to a nearby UAV and in 2023 the airport had to cease operation for an hour again due to a suspected UAV [1]. In 2022 UAVs were sighted close to nuclear power plants and government buildings in Sweden [2]. Other incidents involve the smuggling over state boarders and prisons [3, 4] or the disruption of sporting events [5]. Early detection of illegally intruding UAVs to safety critical areas is paramount to prepare an appropriate response to a possibly hazardous situation.

UAV detection is generally performed in a multispectral manner, combining various sensor technologies, like RADAR [6], radio frequency [7], acoustics [8] and electro-optics [9], into a holistic system [10, 11]. An essential component of such a system is the electro-optical sensor, as it allows the most conclusive situational assessment through visual images. These optical sensors have a narrow field of view (FoV) and use mounts for pan and tilt movement [12]. Detection of UAVs is facilitated via computer vision algorithms, with convolutional neural networks, like YOLO [13], FRCNN [14] or Retinanet [15] outperforming conventional methods [16]. To achieve long detection ranges for small UAVs, the latest research incorporates telescopes, which allow larger apertures and thus, better resolution, to extend the operational range to more than 4 km [9]. The usage of optics with long focal lengths and large apertures result in a narrow depth of field (DoF), which increases the demand on a fast automatic focus especially when tracking high-speed UAVs.

The automatic focus is characterized into two categories: active and passive focusing methods. Active focus methods measure the distance to the object through time of flight or ultrasonic sensors to adjust the focus accordingly [17, 18]. The most common passive focus technique is phase detection, which divides the incoming light rays into pairs of beams and by comparing them, the focus direction and distance to the best focus position is determined [19]. Although this method is fast and does not require much computational power, additional sensors and optical elements as well as a precise alignment are necessary.

Passive contrast-based methods find the optimal focus position by comparing the contrast values of multiple images after shifting the camera sensor with respect to the optics in axial direction [20] or in the case of telescope systems, the mirrors with respect to each other [21]. Contrast functions can be divided into five major groups: Gradient-, Laplace-, Wavelet-, Statistics- and Discrete Cosine Transform based methods [22]. A well-known example of the gradient-based group is the Tenengrad operator [23], which has shown promising results, when applied using telescope systems and focusing on terrestrial objects like buildings [24]. Likewise, the Normalized Variance [25], also a gradient-based approach, achieves good results for astronomical observations [26]. To focus on very small image regions, as necessary for detection and tracking of small UAVs, the Variance of Laplacian [27] and the Discrete Wavelet Transform operator have proven to be good choices [28]. A variety of algorithms exist, which try to maximize the contrast value and thus the sharpness of the image, like the Hill Climb [29], Fibonacci or Binary algorithm [20, 30], curve fitting [31] or Gaussian fitting [32] etc.
Passive autofocus with contrast detection is frequently used in telescope-based imaging systems. However, the main applications are concentrated around focusing on large scale objects like buildings or for astronomical observations, which do not require a fast focus, but rather high precision [24, 26]. Passive high-speed autofocus methods for long focal lengths have been implemented for Time Delay Integration (TDI) cameras, which take continuous measurements at 250 Hz and estimate the maximum focus position through polynomials [33]. For the use case of telescope-based UAV detection and tracking, challenges arise through the movement of the UAV itself during the focusing phase, which introduces noise to the passive focus measurement. Furthermore, high UAV velocities require fast focusing speed to keep the UAV in focus.
The contribution of this paper is the analysis, implementation and evaluation of a fast passive autofocus module for a telescope-based UAV detection and tracking system. A high-speed linear stage is used to reposition the camera with respect to the telescope in accordance to the focus algorithm output. The article is organized as follows. Section 2 shows the analysis of the requirements for the autofocus module. In Section 3 the implemented hardware and software system is described in detail. The experiments and results are shown in Section 4 followed by a conclusion in Section 5.

2. System analysis

To derive the hardware specifications for the automatic focus, the requirements for a telescope-based UAV detection and tracking system are necessary. The operational range of the system is given from 150 m to 5000 m and it shall be possible to track for example a DJI Mavic 3, which has a diameter of 0.3 m and a maximum speed of 21 m/s. For reliable long distance detection of UAVs, at least 15 x 15 pixels covering the object are required [9]. Therefore, the focal length of the optical system should range from approximately 400 mm for closer distances to 2500 mm to resolve objects at long distances.

2.1. Travel range

To enable focusing over the entire operational range, either the camera has to be moved with respect to the telescope or the telescope mirrors have to be repositioned with respect to each other. For the analysis it is assumed, that the camera is moved. To calculate the required travel range necessary for the focus, the thin lens model [34] is used, where the image distance I is determined by

$$I = \frac{1}{\frac{1}{f} - \frac{1}{x}}, \tag{1}$$

with f being the focal length of the telescope and x is the distance to the object. The necessary

focus range range is calculated to 1 mm for a focal length of 400 mm and 41.3 mm for a focal length of 2500 mm.

2.2. Focus velocity

To keep fast UAVs in focus, a high focus velocity is crucial. Using Eq. 1 and assuming a challenging case, of a DJI Mavic 3 flying at 21 m/s towards the telescope system in the closest specified distance of 150 m, the required minimum focus velocity is 0.15 mm/s using a $f = 0.4$ m telescope and 6 mm/s using a $f = 2.5$ m telescope.

2.3. Focus accuracy

The required focus accuracy is deduced from the DoF, given by

$$DoF = \frac{2xDfc(x-f)}{D^2f^2 - c^2(x-f)^2}, \tag{2}$$

where D is the aperture, f is the focal length, c is the circle of confusion and x is the distance to the object [35]. Following guidelines of photography c is determined by $d/1500$, with d being the camera sensor diagonal [36]. Fig. 1 shows the DoF of two focal lengths and an aperture of 0.3 m, assuming an 1 inch camera sensor with a diagonal of 15.86 mm the resulting c is 0.01 mm. The minimal required focus accuracy is calculated by using the DoF of the closest distance of the object of 150 m. For a focal length of 0.4 m the DoF is 4 m, which translates to a necessary focus accuracy of 28 μm. Within these 4 m, the optical system captures a sharp image for a human observer [36]. For the longer focal length of 2500 mm the minimum accuracy is 285 μm.

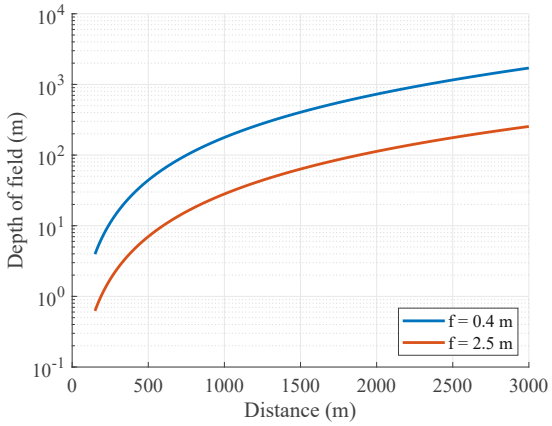


Fig. 1. DoF over distance to the observed object for the given focal lengths of the optical system assuming an 1 inch camera sensor.

3. System implementation

The implemented system is depicted in Fig. 2 and consists of two telescopes, an f/10 Meade Schmidt Cassegrain telescope (LX200-ACF, Meade Acquisition Corp., Watsonville, USA) with a focal length of 2.54 m and an ASA UWF300 telescope (ASA Astrosysteme GmbH, Neumarkt, Austria) with a focal length of 0.39 m and an aperture of 0.3 m. The telescopes are attached to a DDM100 mount (ASA Astrosysteme GmbH) to enable pan and tilt motion. To capture

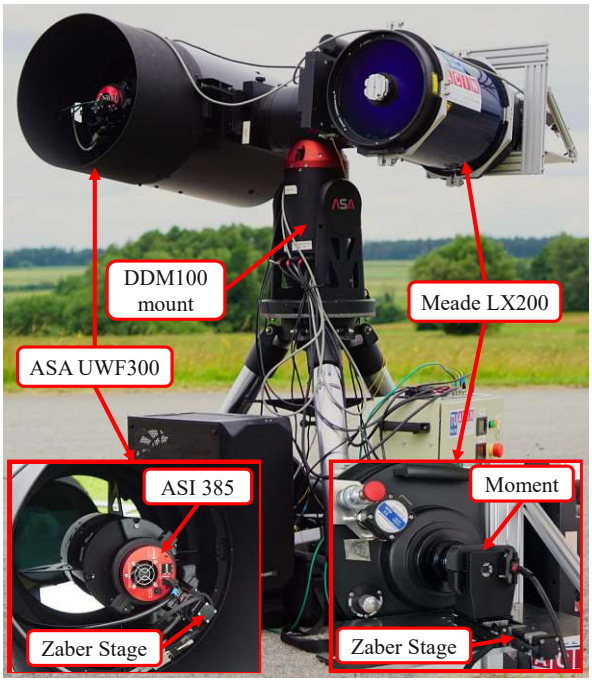


Fig. 2. The implemented dual telescope-system with the custom autofocus module consisting of a linear stage and an ASI camera for the ASA UWF 300 telescope and an additional stage for the Meade telescope, which adjusts the focus for the Moment camera.

images, the Meade telescope is equipped with a Moment scientific CMOS camera (Teledyne Photometrics, USA), with a sensor diagonal of 17.5 mm and an ASI 385 MC-Cool camera (ZWO Company, Suzhou, China) with a sensor diagonal of 8.4 mm is attached to the UWF300 telescope. Following the analysis presented in Section 2, two linear stages of the type LSM050B-E03T4 (Zaber Technologies Inc., Vancouver, Canada) together with corresponding controllers X-MCB1 are used to reposition the cameras with respect to the telescopes. The stage offers a travel range of 50.8 mm with a travel speed of 104 mm/s, while maintaining an accuracy of 3.5 μm. Both cameras are mounted to the stages using custom-made adapters. Implementing the passive focus by repositioning the cameras, rather than the telescope mirrors, is selected, as the latter approach is too slow for the intended application.

3.1. Software architecture

The system architecture is depicted in Fig. 3, whereas the software components, the camera, the object detector and the autofocus component are running in parallel [37]. The camera captures images, that are provided to the deep learning object detector, as developed and trained to detect UAVs in [37], and the autofocus component. For object detection, a fine-tuned FRCNN object detector is used, which is trained on a custom UAV dataset [37]. The autofocus component issues position commands to the stage driver, which controls the linear stage to position the camera with respect to the telescope.
The automatic focus consists of two phases. First, the operational range is scanned in z direction

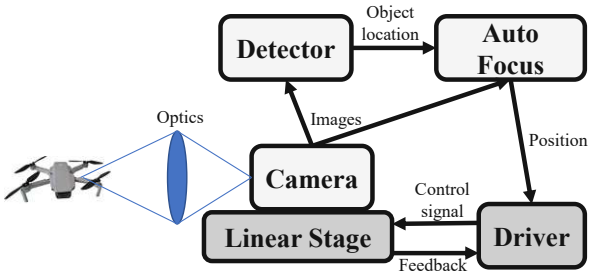


Fig. 3. Overview of the system architecture. The camera provides images to the object detector and autofocus algorithm, which send commands to the stage driver. The stage driver controls the linear stage, that physically moves the camera with respect to the optics to set the focus.

in the search for a UAV. Once a UAV is detected by the object detection algorithm and a bounding box shows the location of the UAV within an image, the focus optimization phase starts. During the optimization phase, the search algorithm tries to maximize the contrast value within the bounding box, by moving the camera with respect to the telescope and comparing the contrast of multiple images. For the calculation of the contrast, three functions are implemented and evaluated, namely the Tenengrad operator [23]

$$FV_{Tenengrad} = \sum_{n}^{N} \sum_{n}^{M} \nabla_x I(n,m)^2 + \nabla_y I(n,m)^2, \tag{3}$$

the Variance of Laplacian [27]

$$FV_{VarLap} = \frac{\sum_n^N \sum_n^M (\Delta I(n,m) - \overline{\Delta I})^2}{\overline{\Delta I}}, \tag{4}$$

and the Normalized Variance [25]

$$FV_{NVar} = \frac{\frac{1}{NM} \sum_n^N \sum_n^M (I(n,m) - \overline{I(n,m)})^2}{\overline{I}}, \tag{5}$$

where $I(n,m)$ is the image intensity of the pixel at the position n and m, N and M the overall size of the cropped image and $\overline{I}$ the average image intensity. Three search algorithms, which are responsible to maximize the contrast value, are selected and evaluated. The Hill Climb Algorithm [29] moves the camera stepwise into one direction, until the contrast value starts to decrease and then moves it in the other direction with a smaller step size. The Binary and the Fibonacci Algorithm [30] divide the search space in a binary and Fibonacci pattern respectively. The focus optimization is triggered periodically after 3 s to keep the UAV in focus, to ensure a continuous track.

4. Experiments and results

To evaluate the system, various field tests are performed with UAVs to analyse the contrast functions, the search algorithms and the required speed of the autofocus.

Table 1. Results of the evaluation of the contrast functions for both the UWF and the Meade telescope after each contrast function is normalized to 1. Furthermore the data shows the mean of 8 measurements per telescope. The best results are marked in bold.

Name	Sensitivity	Local maxima	Position error
Normalized Variance (UWF)	1.56	9.43	0.113 mm
Tenengrad (UWF)	**2.95**	2.43	**0.015 mm**
Variance of Laplacian (UWF)	1.34	**1.86**	0.027 mm
Normalized Variance (Meade)	1.96	8.12	0.55 mm
Tenengrad (Meade)	**5.16**	**2.75**	**0.23 mm**
Variance of Laplacian (Meade)	1.38	6.12	0.35 mm

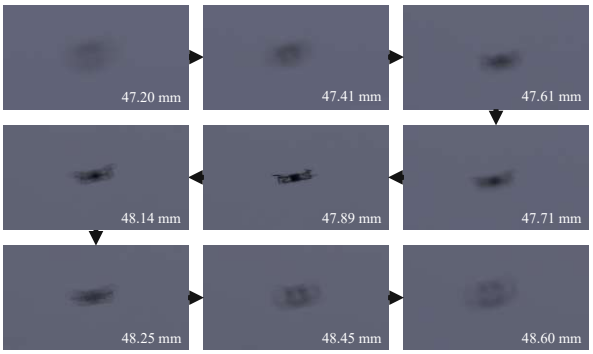


Fig. 4. A sequence of nine selected images from a linear stage scan to obtain the contrast measurement is depicted for the evaluation shown in Table 1. The white numbers within the images indicate the current stage position.

4.1. Contrast functions

To evaluate the contrast functions, a scan with the linear stage is performed, while filming for example a drone as seen in Fig. 4. Then, the contrast functions from Eq. 3 to Eq. 5 are applied to the images to obtain the contrast value per stage position and contrast function. The results of each function are normalized to 1 respectively as seen in Fig. 5.

The metrics used for the analysis are the sensitivity, the mean number of local maxima and the position error [26]. It is desirable to minimize the number of local maxima across the curve, as this reduces the probability of the search algorithm to stop focusing at a wrong position. Ideally, only one global maximum is present, which is the object to be tracked. The sensitivity of the curve is defined as the ratio between the maximum value divided by the mean of the contrast values, which are smaller than the mean of the contrast curve itself. A high sensitivity, therefore, manifests itself as a prominent peak. Finally, the position error shows the offset between the stage position of the maximum value of the measured contrast curve with respect to the manually selected best focus position.

Table 1 shows the mean value of 8 linear stage scan measurements performed with each telescope respectively. For both telescopes, the Normalized Variance has the worst performance in terms

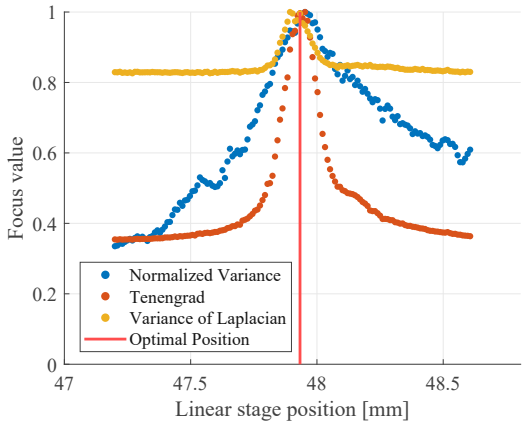


Fig. 5. Contrast values calculated using three different contrast measures for a linear stage scan, which contains a single UAV in the image in front of a clear sky.

Table 2. Results of approximately 450 tests of the search algorithms per telescope for the UWF and the Meade telescope. The values represent the mean, whereas the standard deviation is shown in brackets. The best results are marked in bold.

Name	Steps	Total distance (mm)	Time (s)	Accuracy (mm)
Hill (UWF)	**9.2 (3.1)**	**1.1 (0.6)**	**1.4 (0.5)**	-0.004 (0.1)
Binary (UWF)	25.3 (7.5)	3.8 (2.4)	3.4 (1.1)	-0.07 (0.3)
Fibonacci (UWF)	15.6 (3.3)	2.9 (1.4)	2.2 (0.8)	**-0.002 (0.2)**
Hill (Meade)	**7 (1.6)**	**30 (13)**	**2.4 (0.7)**	**-0.46 (1.4)**
Binary (Meade)	41.5 (11.5)	67.2 (40)	7.8 (2.5)	-0.58 (1.6)
Fibonacci (Meade)	23.3 (6.3)	44.5 (28.9)	4.8 (1.8)	-0.62 (3.1)

of all three evaluation metrics. The Tenengrad and Variance of Laplacian achieve similar results in terms of position error, however, Tenengrad shows a higher sensitivity and a lower number of maxima and is, therefore, found to the best performing contrast function for the task of focusing on small objects like UAVs. Furthermore, the position error is below the required accuracy as determined using Eq. 2 in Section 2.

4.2. Search algorithms

Another crucial aspect for a reliable contrast-based automatic focus is a deterministic optimization function to maximize the contrast value and thus find the best focus position. Fast moving UAVs paired with the large apertures and low DoF of the optical detection system, require a quick termination of the optimization algorithm.
As evaluation metrics the number of linear stage steps, the total distance travelled, the elapsed time from the start to end of the optimization and the accuracy, defined as the offset to the manually set best focus position, are selected. For the experiments, the UAV is maintaining a

(a) UAV flies towards the system.

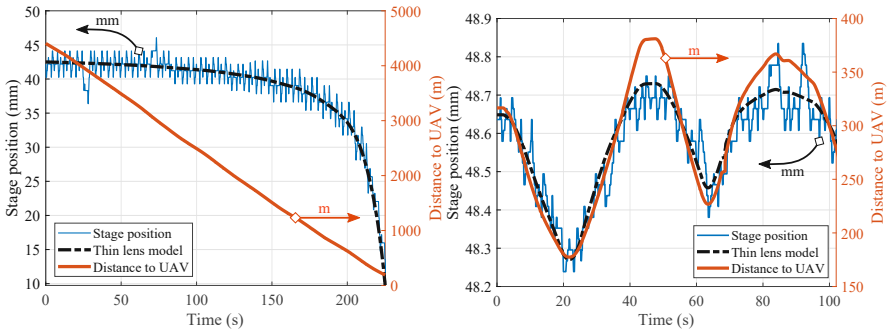
(b) UAV flies away from the system.

Fig. 6. Image sequence showing the DL-DoF test for the object detection algorithm. The initial focus is set manually, which is depicted on the left two images showing a distance to the UAV of 120 m. Then, the UAV flies towards 6a or away 6b from the system, while the focus is not adjusted. To obtain the DL-DoF, the distance is recorded, when the object detector fails, as seen in the right most images.

stationary position and the manually set optimal focus position is noted. Then, the image is defocused by offsetting the linear stage from the optimal focus position by up to 0.5 mm for the UWF and 5 mm for the Meade telescope. In the next step, the object detector is started, which triggers the search algorithm and thus the contrast maximization for the detected bounding box. During the evaluation of the search algorithms, the Tenengrad operator is used to obtain the contrast value.
Table 2 shows the results of the optimization functions for both telescopes, whereas each value represents the mean of approximately 450 tests and the standard deviation is shown in brackets. For both telescopes, the Hill Climbing optimization algorithm scores the best or close to best results according to all four metrics.

4.3. Speed and DoF

Finally, the speed and apparent deep learning DoF (DL-DoF) of the system is experimentally evaluated. The DoF calculated for the UWF telescope in a distance of 150 m is 4 m, which represents the depth of a completely sharp image. However, UAVs in slightly defocused images can still be detected using deep learning algorithms [9]. To experimentally evaluate the DL-DoF of the deep learning algorithm, a UAV is maintaining a certain distance to the system and the focus is set manually to obtain a sharp image. Then, without adjusting the focus, the UAV flies towards or away from the system, slowly going out of focus until the deep learning algorithm fails to detect the UAV as seen in Fig. 6. Using this method the DL-DoF of the UWF telescope for the most challenging case in a close distance of 120 m is determined to be approximately 75 m. The distances are obtained by extracting the GPS position from the UAV remote controller. Within this DL-DoF the UAV is quite defocused, however, still detectable by the deep learning algorithm. Therefore, the requirements towards the stage accuracy can be relaxed. As the measured DL-DoF represents the maximum DoF before the detector fails, for the calculation of the more tolerant accuracy a DL-DoF of 15 m is assumed to ensure robust detection of objects. Referring back to Eq. 2, the new more tolerant circle of confusion for object detection is calculated to be 0.06 mm



(a) Experiment using Meade telescope.

(b) Experiment using UWF telescope.

Fig. 7. Example data of successful continuous tracking and automatic focusing onto UAVs flying at speeds of up to 24 m/s towards the telescope system. The black dashed line shows the thin lens model, which fits well into the actual stage position. In Fig. 7a the non linear dependency between the UAV distance towards the linear stage position is visible, as defined by Eq. 1. The apparent noise of the stage position around the thin lens model is the optimization by the hill climb algorithm, which is periodically triggered to keep the UAV in focus as it moves.

Fig. 8. Example image sequence depicting a UAV flying at up to 15 m/s towards the telescope system, while the automatic focus keeps the UAV in focus.

resulting in a necessary positioning accuracy of 0.16 mm for the UWF and 1.3 mm for the Meade telescope. These more tolerant constraints approve the relatively large standard deviations of the accuracy measured in Table 2.

To evaluate the speed required by the focus module to keep up with the autofocus, experiments are performed with quad-copter and fixed wing UAVs. The DJI Mini 2 with a maximum speed of 16 m/s and a width of 289 mm and DJI Mavic 3 are used as quad-copter. The utilized fixed wing drone achieves a maximum speed of 41 m/s with a wingspan of 5 m. Fig. 7 shows two example flight trajectories together with some example images in Fig. 8, where a UAV is tracked continuously and the automatic focus is ensuring that the UAV stays in focus. During the tests the Tenengrad operator together with the Hill climbing search algorithm are used, whereas the autofocus optimization is triggered every 3 s. The maximum speed of the UAV flying towards the telescope system is recorded at 24 m/s resulting in a maximum stage velocity of 12 mm/s for the Meade telescope as shown in Fig. 7a. The necessary speed of the linear stage is higher than the calculated speed in Section 2, as the contrast based method requires multiple measurements at different stage positions to find the best focus. Also the UAV speeds during the tests are slightly higher than the assumed values in the system analysis. Fig. 7a confirms the calculated required travel range for the linear stage using the Meade telescope in Section 2.

In summary, by periodically running the automatic focus, continuous detection and tracking of UAVs flying a speeds of up to 24 m/s from more than 4500 m down to 150 m towards the

telescope system is possible, as the focus module ensures a sharp image.

5. Conclusion

An automatic focus module for a telescope system is presented, which enables it to keep small and fast moving UAVs in focus. Following the analysis and an integration of a suitable linear stage, it is experimentally validated that the Tenengrad contrast measure and the Hill Climb search algorithms show the best performance for the task of focusing fast and precisely onto small UAVs within a telescope image. Furthermore, it is demonstrated, that the designed focus module keeps a UAV, flying at a speed of 24 m/s, at a worst case distance of 150 m in focus when using the $f10$ Meade telescope with a focal length of 2540 mm, resulting in a stage velocity of up to 12 mm/s. This enables detection and continuous tracking of UAVs flying at high speeds towards the telescope system as the focus module ensures that the UAV remains in focus. Future work will be centred around an efficient search pattern, which includes the pan and tilt motion of the mount and the focus of the camera, to reduce the time needed until the object detector finds a UAV and the optimization phase is triggered. Furthermore, the effects of the bounding box size onto the focusing performance, especially in front of complex backgrounds will be investigated.

6. Funding

This publication is funded by the Austrian defence research programme FORTE of the Federal Ministry of Finance (BMF).

7. Acknowledgements

The authors gratefully acknowledge the cooperation with ASA Astrosysteme GmbH and thank for their support and valuable expertise.

8. Disclosures

The authors declare that there are no conflicts of interest related to this paper.

9. Data availability

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. "'Suspected drone' disrupts Gatwick Airport flights," BBC News (2023). Accessed Aug 2023.
2. "Sweden drones: Sightings reported over nuclear plants and palace," BBC News (2022). Accessed Aug 2023.
3. S. Dinan, "Mexican drug cartels using drones to smuggle heroin, meth, cocaine into U.S." The Washington Times (2015). Accessed Feb 2022.
4. "Charges over drone drug smuggling into prisons," BBC News (2018). Accessed Feb 2022.
5. "Drones crashing big sporting events, including U.S. Open, college football ," CNN US (2015). Accessed Aug 2023.
6. A. D. De Quevedo, F. I. Urzaiz, J. G. Menoyo, and A. A. Lopez, "Drone Detection and RCS Measurements with Ubiquitous Radar," in *International Conference on Radar (RADAR),* (2018), pp. 1–6.
7. S. Yang, H. Qin, X. Liang, and T. Gulliver, "An Improved Unauthorized Unmanned Aerial Vehicle Detection Algorithm Using Radiofrequency-Based Statistical Fingerprint Analysis," Sensors **19**, 274 (2019).
8. V. Baron, S. Bouley, M. Muschinowski, *et al.*, "Drone localization and identification using an acoustic array and supervised learning," in *Artificial Intelligence and Machine Learning in Defense Applications,* vol. 11169 (SPIE, 2019), pp. 129–137.
9. D. Ojdanić, A. Sinn, C. Naverschnigg, and G. Schitter, "Feasibility analysis of optical UAV detection over long distances using robotic telescopes," IEEE Trans. on Aerosp. Electron. Syst. **59**, 5148–5157 (2023).
10. Aselsan, "IHTAR anti-drone system - datasheet," ASELSAN A.S., Ankara, Türkiye (2018).
11. J. Farlik, M. Kratky, J. Casar, and V. Stary, "Multispectral Detection of Commercial Unmanned Aerial Vehicles," Sensors **19**, 1517 (2019).

12. H. U. Unlu, P. S. Niehaus, D. Chirita, *et al.*, "Deep learning-based visual tracking of UAVs using a PTZ camera system," in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society,* (IEEE, 2019), pp. 638–644.
13. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," https://arxiv.org/abs/2004.10934 (2020).
14. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems,* vol. 28 (2015).
15. T.-Y. Lin, P. Goyal, R. Girshick, *et al.*, "Focal loss for dense object detection," IEEE Trans. on Pattern Anal. Mach. Intell. **42**, 318–327 (2020).
16. L. Liu, W. Ouyang, X. Wang, *et al.*, "Deep learning for generic object detection: A survey," Int. J. Comput. Vis. **128**, 261–318 (2019).
17. N. L. Stauffer, "Active auto focus system improvement," (1983). US Patent 4,367,027.
18. Z. Zhou, C. Li, T. He, *et al.*, "Facile large-area autofocusing Raman mapping system for 2D material characterization," Opt. Express **26**, 9071–9080 (2018).
19. T. Sasakura, "Automatic focusing device using phase difference detection," (1999). US Patent 5,995,144.
20. Y. Yao, B. Abidi, N. Doggaz, and M. Abidi, "Evaluation of sharpness measures and search algorithms for the auto-focusing of high-magnification images," in *Visual Information Processing XV,* vol. 6246 (SPIE, 2006), pp. 132–143.
21. F. Bortoletto, C. Bonoli, D. Fantinel, D. Gardiol, C. Pernechele, "An active telescope secondary mirror control system," Rev. scientific instruments **70**, 2856–2860 (1999).
22. S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," Pattern Recognit. **46**, 1415–1432 (2013).
23. J. M. Tenenbaum, *Accommodation in computer vision* (Stanford University, 1971).
24. C. Yang, M. Chen, F. Zhou, *et al.*, "Accurate and rapid auto-focus methods based on image quality assessment for telescope observation," Appl. Sci. **10**, 658 (2020).
25. J. Vaquero, J. Pena, N. Malpica, *et al.*, "Evaluation of autofocus functions in molecular cytogenetic analysis," J. microscopy **188**, 264–272 (1997).
26. I. Helmy, F. Elnagahy, and A. Hamdy, "Focus measures assessment for astronomical images," in *International Conference on Innovative Trends in Communication and Computer Engineering (ITCE),* (IEEE, 2020), pp. 244–250.
27. J. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia, "Diatom autofocusing in brightfield microscopy: a comparative study," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000,* vol. 3 (2000), pp. 314–317 vol.3.
28. G. Yang and B. J. Nelson, "Wavelet-based autofocusing and unsupervised segmentation of microscopic images," in *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003),* vol. 3 (IEEE, 2003), pp. 2143–2148.
29. J. He, R. Zhou, and Z. Hong, "Modified fast climbing search auto-focus algorithm with adaptive step size searching technique for digital camera," IEEE Trans. on Consumer Electron. **49**, 257–262 (2003).
30. S. S. Rao, *Engineering optimization: theory and practice* (John Wiley & Sons, 2019).
31. Y. Xiong and S. Shafer, "Depth from focusing and defocusing," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,* (IEEE, 1993), pp. 68–73.
32. P. DiMeo, L. Sun, and X. Du, "Fast and accurate autofocus control using gaussian standard deviation and gradient-based binning," Opt. Express **29**, 19862–19878 (2021).
33. D. Wang, X. Ding, T. Zhang, and H. Kuang, "A fast auto-focusing technique for the long focal lens tdi ccd camera in remote sensing applications," Opt. & Laser Technol. **45**, 190–197 (2013).
34. A. Recknagel, *Elementarphysik (Elektrik Optik)* (P.E. Blank Verlag, 1953).
35. E. Krotkov, "Focusing," Int. J. Comput. Vis. **1**, 223–237 (1987).
36. H. Nasse, "Depth of field and bokeh," Carl Zeiss camera lens division report **1**, 4 (2010).
37. D. Ojdanić, C. Naveschnigg, A. Sinn, *et al.*, "Parallel Architecture for Low Latency UAV Detection and Tracking using Robotic Telescopes ," IEEE Trans. on Aerosp. Electron. Syst. (2023). Submitted Jul. 2023.

Publication [P7]

Camera-guided real-time laser ranging for multi-UAV distance measurement

Authors: D. Ojdanić, B. Gräf, A. Sinn, H. W. Yoo, and G. Schitter

Reference:

[P7] D. Ojdanić, B. Gräf, A. Sinn, H. W. Yoo, and G. Schitter, "Camera-guided real-time laser ranging for multi-UAV distance measurement," Appl. Opt. 61, 9233-9240 (2022)

Journal Quartile: Q2 (2022)

Research Article Applied Optics 1

Camera Guided Real-Time Laser Ranging for Multi-UAV Distance Measurement

DENIS OJDANIĆ[1,*], BENJAMIN GRÄF[1], ANDREAS SINN[1], HAN WOONG YOO[1], AND GEORG SCHITTER[1]

[1] The authors are with the Automation and Control Institute (ACIN), Technische Universität Wien, 1040 Vienna
[*] Corresponding author: ojdanic@acin.tuwien.ac.at

Compiled October 4, 2024

This paper presents the design and implementation of a scalable laser ranger finder (LRF) based prototype system, which enables distance measurement and precise localization of multiple UAVs in real-time. The system consists of a telescope and camera as the image acquisition components, supplemented by an LRF and a fast steering mirror (FSM) to obtain the distance measurement. By combining the optical path of the camera and the LRF through a dichroic mirror, the LRF is accurately aligned by the FSM based on the angular position of a UAV within the camera field of view. The implemented prototype successfully demonstrates distance measurements of up to 4 UAVs with a bandwidth of 14 Hz per object.

https://doi.org/10.1364/AO.470361

1. INTRODUCTION

The popularity of unmanned aerial vehicles (UAVs) has grown exponentially over the past several years [1]. Unfortunately, the commercial success is accompanied by a simultaneous growth of incidents in safety critical areas. Airports, for instance, are highly vulnerable to undetected UAVs as an occurrence around the Los Angeles International Airport shows, where a commercial jet almost collided with a UAV during approach [2]. In 2015 multiple UAVs were sighted in the proximity of several nuclear power plants in France [3] and studies on the threat posed by UAVs to the mentioned facilities highlight the necessity of deployment of appropriate technological solutions for detection and neutralization [4]. Besides the apparent threat to critical infrastructure, UAVs are versatile tools for the smuggling of goods across state boarders or in and out of prisons [5, 6]. Furthermore, recent developments in the control of multiple drones as swarms may revolutionize future threat scenarios as for example multiple UAVs flying in close formation increase the chance of reaching the desired destination through redundancy [7, 8]. Given the rising number of UAVs sold every year paired with the alarming amount of reported incidents, development and installation of UAV reconnaissance systems for precise UAV localization are paramount to prepare appropriate defensive countermeasures.

Various UAV detection and distance measurement methods have been researched extensively over the past decade. Radio frequency detection for example exploits the communication link between operator and the UAV to localize targets over distances of 5 km with accuracies of about 5° [9, 10]. Similar UAV localization accuracies over distances up to 600 m are achievable by the usage of acoustic microphone arrays recording the ambient sound and applying suitable signal analysis methods [11, 12]. While being able to detect multiple objects simultaneously, the achieved accuracies do not allow precise localization needed for directed countermeasures.
Electro-optical systems use cameras to capture images and reconnaissance of UAVs is performed by advanced computer vision algorithms [13] with operational ranges for typical systems going up to

2 km [14]. Estimations through a monocular vision approach are possible by applying for example pose estimation using a skeleton model of the UAV [15] or using deep neural networks [16], however the localization accuracy suffers greatly for long ranges [17]. Stereo vision provides better results for distance measurements and is used for autonomous driving, robotics and 3D building mapping [18]. For accurate measurements over long distances, the two cameras need to be displaced further away, which makes the system more complex.

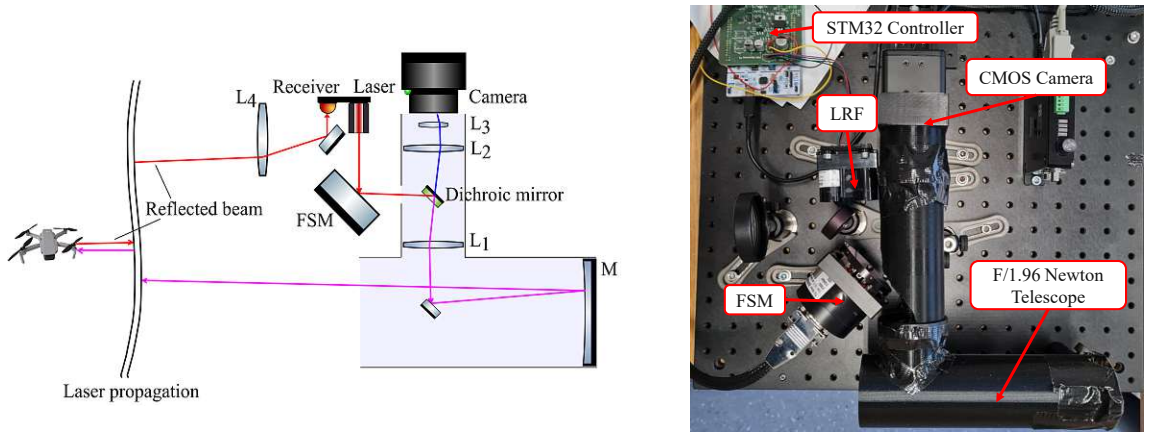
RADAR actively sends out a signal as radio wave and acquires the distance to an object by measuring the signal reflected by the target. It has been shown that RADAR is capable of detecting even small objects like consumer UAVs over distances of several kilometres [19]. Whereas detection is possible, object tracking is not reliable as it heavily depends on the objects radar cross section and the surrounding terrain [14].

Methods based on the time of flight (ToF) principle are emitting an electro-magnetic wave and measuring the reflected signal to infer the distance [20]. LiDAR, is one example for a ToF measurement, which is often used for autonomous vehicular applications [21] and has been applied to measure the distance to UAVs [22, 23]. Flashing LiDAR uses a diverging laser, which illuminates the whole field of view (FoV) and the reflected signal is measured by a matrix of photodiodes [24]. Flashing LiDAR cameras exist with operational ranges of up to 1000 m, nevertheless, the achievable resolution is low and does not allow to detect small objects like UAVs over long distances [25]. Rotating and scanning LiDAR are modifications to achieve longer distances by using collimated laser beams with little divergence and sequentially measuring the field of view (FoV) [26]. Studies have been conducted using rotating LiDAR to detect UAVs [22], concluding that the resolution depends on the number of vertical sensors in the receiver array. Scanning LiDAR extends the operational range to a few kilometres [27, 28], however, sequentially measuring a large FoV is time consuming and not suited to localize small and fast moving objects. The combination of ranging with optical imaging systems has been studied extensively to generate terrain maps with topography information [29] and for atmospheric remote sensing [30, 31]. To measure the distance to UAVs, bi-axial systems have been proposed consisting of a dome equipped with a camera and laser ranger finder (LRF)s aiming at the center of the camera FoV [32]. To obtain a measurement, the dome has to align the object within the center of the camera FoV, which limits the achievable bandwidth of the system making a real-time multi-object distance measurement infeasible. Generally, sensor fusion is applied to build holistic UAV detection and localization systems combining the strengths of various approaches, for example RADAR or acoustics to detect an object and electro-optics for the visual confirmation. [33–35] Nonetheless, an accurate long distance localization of multiple objects in real-time is still challenging with the described methods.

The contribution of this paper is the integration of a steerable LRF with a telescope system, which enables multi-UAV distance measurement by aiming the LRF with an FSM towards a UAV position extracted within a camera frame. The laser transmitter and camera optical path are combined by a dichroic mirror, merging the visual and the infrared light, allowing real-time alignment of the LRF. The article is organized as follows. Section 2 describes the proposed concept and implemented system. In Section 3 the required laser power is analysed and in Section 4 the distance measurement strategy is discussed in more detail. Finally, the experiments and results are shown in Section 5 followed by concluding remarks in Section 6.

2. SYSTEM DESCRIPTION

The main components of the system are the camera and telescope for image acquisition [36], the LRF for distance measurements and further optical elements for the alignment of visual and laser light paths. The key advantage of the proposed concept is the possibility to obtain the distance to an object by a single LRF measurement, as the laser beam is overlapping with the optical path of the visual light as indicated by the violet lines in Fig. 1a. The combination of the two light beams is facilitated by a dichroic mirror, which is transparent for wavelengths of the visible spectrum and reflective for infrared light. Objects, like UAVs, are extracted from camera frames by the usage of computer vision algorithms and based on the position within an image, the LRF beam direction is adjusted by a fast steering mirror (FSM). By applying this idea, no scanning of the FoV is required



(a) Schematic overview of the optical setup.

(b) Illustration of the implemented system.

Fig. 1. Overview of the optical setup consisting of a custom built Newtonian telescope and a commercially available CMOS camera for image acquisition. (a) Schematic overview: The LRF LiDAR-Lite v3 is used to measure the distance to a UAV and the LRF transmit beam is steered by an OIM102 FSM. The optical paths of the visual (blue) and infrared light (red) are merged by a dichroic mirror (violet). Table 1 shows details about the used components. (b) Image of the implemented small scale prototype: A STM32 Nucleo-64 controller board is responsible for interfacing and controlling the LRF and the FSM.

to find objects and the distance is obtained by a single LRF measurement. Finally, the proposed optical setup can be attached to a mount allowing pan and tilt motion to observe a 360° area, whereas simultaneous distance measurement is possible for objects within the field of view of the telescope [36]. The FSM based laser steering system allows fast acquisition and hence, distance measurements as soon as the object is within the field of view of the camera and detected by the computer vision algorithm. Rigidly mounted LRFs require a centering procedure of the telescope to enable the position measurement [32], which may be a time consuming procedure in case of multiple UAVs. Furthermore, the mount may be used to pan and tilt to another UAV or group of UAVs of interest.

A. System Components

To demonstrate the concept, a small scale prototype system is implemented as shown in Fig. 1b. The LiDAR-Lite v3 (Garmin, USA) is selected as an LRF, which has a maximum measurement range of 40 m and is operated at an update rate of 117 Hz. The laser beam has a wavelength of 905 nm at a beam divergence of 8 mrad. The peak power of the LRF is 1.3 W with a pulse duration of 500 ns. The communication to the LRF is established via an inter-

Table 1. Sizes and focal lengths of the optical lenses and mirrors used in the setup.

Name	Diameter	Focal length
Dichroic mirror	12.7 mm	
Secondary mirror	12.7 mm	
FSM	50.8 mm	
M	50.8 mm	100 mm
L_1	50.8 mm	100 mm
L_2	50.8 mm	100 mm
L_3	23 mm	50 mm
L_4	50.8 mm	125 mm

integrated circuit (I^2C) interface. For an accurate and quick laser alignment the OIM102 (Optics In Motion, USA) FSM is responsible, which has an mechanical range of $\pm$ 1.5° and a -3 dB bandwidth of 750 Hz for reference signals of 1.5 millidegrees. To control the FSM and the LRF, a custom build extension board is designed for a STM32 Nucleo-64 (MB1136) board (STMicroelectronics, Switzerland).

A CMOS camera (Ailipu Technology Inc., China) is used to capture images with a sensor width of 5.76 mm and height of 4 mm. The camera provides 24 frames per second (FPS) and is run with a resolution of 640 x 480 pixels. To focus the light on the camera sensor a Newtonian telescope with a focal length of 100 mm and a 2 inch aperture size is built. The focal ratio of f/1.96 is used as it allows a wide FoV for short ranges complementary to the achieved distances by the LRF. Finally, the combination of infrared and visual light is facilitated through a partly reflective and partly transmissive dichroic mirror. For the setup the DMSP805T (Thorlabs Inc, Germany) half inch short short-pass dichronic mirror is selected, which has a cut-off wavelength of 805 nm. Below the cutoff wavelength the mirror is transmissive and wavelengths above this threshold are reflected. The system is easily scalable to longer operational distances by selecting a more powerful long range LRF and an adequate telescope.

3. SYSTEM ANALYSIS

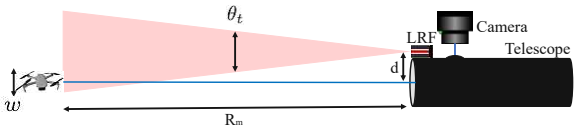
In this section, the influence of an offset between the optical paths between the camera and LRF are discussed justifying the proposed concept. Additionally, a model is presented for required laser power and the achievable measurement distances for the prototype system.

A. Laser to Camera Offset

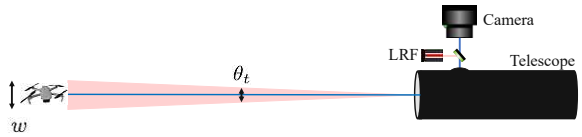
For the design of a system, which optically detects and subsequently measures the distance to a UAV using an LRF, the offset d between the path of the camera light and the laser is crucial. A simple solution is by placing the LRF close to the optical aperture [32] as depicted in Fig. 2a. The minimal distance R_m for which an object with a diameter of w is fully illuminated by a laser beam is given by

$$R_m = \frac{d + \frac{w}{2}}{tan(\theta_t)}, \tag{1}$$

where d is the LRF to camera offset of 0.1 m. Using the LiDAR-Lite v3 as an LRF of the small scale system with a beam divergence of 8 mrad, an object with a diagonal dimension of 0.05 m, corresponding to the surface area of the body of a front facing DJI Mini 2 excluding the rotors, will be fully illuminated in a distance of 15.6 m. Below this distance, the object is illuminated only partly, resulting in less



(a) Offset d between optical paths of camera and LRF.



(b) No offset between optical paths of camera and LRF.

Fig. 2. The optical path of the camera and the LRF are separated by a distance d (a). An object of size w is fully illuminated in distances greater than R_m. Below this distance, a correct measurement is not guaranteed, as most of the laser beam is missing the target. Reducing the beam divergence θ_t increases the distance R_m. To overcome these problems, the optical path of the camera and the LRF are merged (b). The distance to the UAV can be obtained by a single LRF measurement. Reducing the beam divergence θ_t has no negative effect on the operational range of the system and can even increase the measurement distance for small objects.

reflected laser power, which degrades the measurement quality. In a distance below 9.4 m the object is completely missed. For an upscaled system covering long measurement distances, similarly the camera to laser offset leads to a potential target miss. To enable measurements over the whole range, the epipolar line, being the blue line in Fig. 2a, has to be scanned at a cost of reduced system bandwidth, as multiple LRF measurements are necessary to find the object in the first place. Another approach is to combine the two optical paths of the laser beam and the camera as presented in Section 2 and in Fig. 2b, which automatically aligns the laser with the camera image. While the optical system of Fig. 2b is more complex and leads to an increased optical loss, the configuration allows high bandwidth, which is indefensible for real-time distance measurement.

B. Laser Power

A crucial design parameter for the ToF principle is the required laser peak power, as it directly influences the light intensity hitting the object and therefore forms the basis of the reflected signal. Using the typical rangefinder relation the received laser power

Table 2. Overview of the parameters for Eq. (2).

Symbol	Quantity
τ_a	Atmospheric transmission loss factor
τ_o	Optical transmission factor
D_R	Diameter of receiver optics
ρ_t	Target reflectivity
d_A	Effective surface area of the target
R	Distance between LRF and target
θ_R	Target surface angular dispersion
θ_t	Laser beam divergence

P_{det} can be determined by

$$P_{det} = \frac{\tau_a^2 \tau_o D_R^2 \rho_t d_A}{R^2 \theta_R (\theta_t R)^2} P_{peak}, \tag{2}$$

with the transmitted peak laser power P_{peak} and the remaining parameters from Table 2 [37]. The atmospheric transmission loss factor τ_a is given by

$$\tau_a = e^{-\gamma R}, \tag{3}$$

with γ being the atmospheric extinction coefficient [38]. For the presented calculation γ is set to 0.096 km^{-1}, which resembles excellent atmospheric conditions [38]. Using Eq. (2) together with the specifications in the datasheet of the LiDAR Lite v3, the maximally achievable measurement distance for various target distances and sizes can be inferred for the optical setup in Fig. 1b. As the LRF is intended to be incorporated into a system consisting of various optical elements as seen in Fig. 1a, a loss factor for the optical components τ_o and an increased receiver aperture size of 50.8 mm are considered. Fig. 3 shows the necessary peak power to measure the distance to different target sizes. Due to the noise of most detectors of LiDARs, the maximal detection distance in reality is influenced by environmental conditions such as ambient light and temperature [39]. The influence of the effective target surface area d_A is visible in Fig. 3 through the bend in the laser transmitter power curves. Above this bend, the required laser peak power increases more rapidly with the distance, as the laser beam cross section is larger than the target, which results in a loss of energy as part of

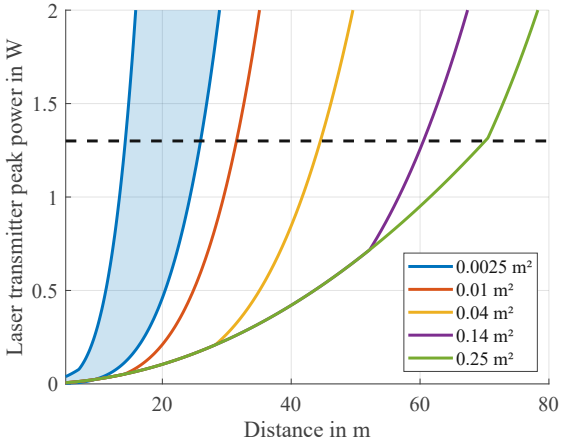


Fig. 3. The effect of different target sizes on the achievable measurement distance for the LRF of the small scale prototype system for a challenging target reflectivity of 30 % and an receiver aperture size of 50.8 mm. The dashed black line represents the peak laser power of 1.3 W according the datasheet. Additionally, the impact of the target material reflectance ρ is depicted in blue for an object of 0.05 m x 0.05 m and the reflectance ranging from 0.05 % to 0.55 %.

the laser light is not reflected by the target. In practice, the laser power cannot be simply increased for a long range detection considering saturation with high gain detectors at short distances and laser safety for human eye and skin [39, 40]. Therefore, keeping the laser beam cross section smaller than the effective surface area maximizes the amount of reflected laser light by the target and thus the achievable measurement distances. The dashed black line represents the LRF peak power of 1.3 W resulting in a theoretical maximum measurement distance of roughly 70 m for a target size of 0.25 m^2.

An important property for the analysis is the target reflectivity, as it has a great impact on the system performance. Depending on the type and colour of the plastic of the UAV a wide range of reflectivity ρ_t is applicable [41, 42]. For the presented model, a conservative reflectivity in the lower range of 5 % to 55 % is selected, as the properties of an intruding UAV is not known a priori. In Fig. 3 the shaded blue area shows an achievable measurement range between 14 m to 26 m for an object size of 0.05 m x 0.05 m for the investigated reflectivity range. As stated in Section 2, the small scale prototype system

is scalable to longer operational ranges by selecting a stronger laser and a lower beam divergence.

4. DISTANCE MEASUREMENT STRATEGY

The position of a UAV in a video frame is extracted by computer vision algorithms. As the goal is a proof of concept of the proposed design, the Kernelized Correlation Filter (KCF) tracker provided by OpenCV [43] is used, which is initialized manually to track a certain object. Based on the position provided by the algorithm, the LRF is accurately aligned by the FSM to obtain the distance to the object. Combining the object position within a frame with the measured distance, an exact horizontal localization of the objects relative position to the setup is calculated by

$$x = R \cdot tan(\alpha)\frac{f_x}{f_w}, \tag{4}$$

where α represents half of the angular horizontal FoV, R the distance to the object, f_x the offset of the object from the frame center in pixels in vertical direction and f_w the frame width in pixels. A similar calculation returns the y coordinate or the vertical position of the object replacing f_x with the offset in vertical direction f_y. The center of the coordinate system is situated at the telescope entry and by knowing the telescope position, the object can be localized precisely. In addition to knowing the precise object location, the object size can be determined based on the distance and the number of pixels covered by the object.

The benefit of the proposed concept is the possibility to align the laser to the UAV without the need of performing a scan with the laser to initially find the UAV. Nevertheless, as shown in the Section 5, some measurements might still miss the UAV due to non perfect bounding box localization by the computer vision algorithm and due to the complex shape of UAVs. Increasing the beam divergence, similar to flashing LiDAR, is an approach to deal with this issue, but this increases the required laser power according to Eq. (2) and is therefore not feasible. Another solution, at the expense of system bandwidth, is to perform a local scan of the area suggested by the bounding box of the computer vision algorithm using the FSM. The latter approach is selected, where multiple laser distance measurements are conducted inside the bounding box using a raster scanning trajectory. The obtained data points are clustered via the k-means clustering algorithm to extract the object in the foreground. The average distance of the foreground cluster is taken as the measured UAV distance.

Additionally, the setup allows to measure the distance to multiple objects in real-time by realigning the laser beam direction through tip and tilt motions of the FSM. When measuring the distance to an increasing number of objects, the system bandwidth, meaning the distance measurements performed per UAV, is reduced. After acquiring a frame and subsequently the positions through the computer vision algorithm, the bandwidth is comprised of the sequential FSM alignment and the LRF measurement. To improve the system performance the path between UAVs travelled by the FSM is optimized to reduce the movement of the mirror. To achieve this optimization, a brute force approach is chosen to calculate the shortest path between multiple UAVs.

5. EXPERIMENTS AND RESULTS

For the experimental analysis, indoor tests at room temperature are performed using targets of different sizes and materials to quantify the distance measurement capabilities of the small scale setup. Furthermore, experiments with UAVs are conducted to test the effect of small and complex shaped objects on achievable distances. Finally, an evaluation of the multi-object measurement performance is conducted.

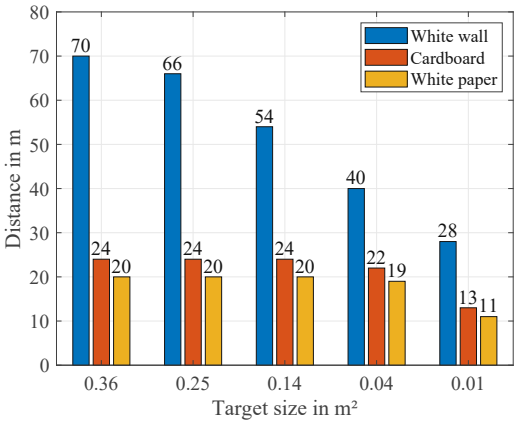


Fig. 4. Measured maximum distances for different target sizes and materials using the small scale prototype system.

A. Target size

The first part of the experiments examines the achievable distances of the small scale prototype system for varying target sizes, distances and materials to validate the presented model in Section 3. The evaluation setup is inspired by the LAS test station by Inframet, where the distance to targets of different sizes and reflectivity is measured [44]. In contrast to the proposed procedure, two smaller targets with the dimensions of 0.2 m x 0.2 m and 0.1 m x 0.1 m are added, which correspond to smaller UAV sizes like consumer drones. To obtain the maximum measurable distance, a target is used, which is larger than the beam diameter resulting in a maximum operation range of 70 m. The increase in achievable distance compared to the information provided in the datasheet is largely contributed to the added receiver optics L_4. The results for different target sizes and materials are presented in Fig. 4 and the impact of the decreasing object surface is evident for the white wall, as long distances are achievable due to its strong reflectivity. The laser beam area increases with the distance and as the target size is surpassed, less power is reflected and more transmit power is needed to further increase the operational range, which manifests itself as the steep increase in the required laser peak power curves visible in Fig. 3. To mitigate this problem, the laser beam divergence can be further reduced. Comparing the measurement results in Fig. 4 to the model in Fig. 3 the model shows slightly longer expected distances for example 70 m for a target size of 0.25 m^2 compared to the measured 66 m. To evaluate small and complex shaped objects, the maximal measurement distance to a DJI Mini 2 is measured at 15 m, which fits within the lower range of the expected distances as presented in Fig. 3 in the shaded blue area. The experiments show that the estimation of achievable distances according to the model in (2) corresponds to the measured results.

B. Multi-object measurement

In the following an analysis of the system bandwidth, which corresponds to the achievable frame rate, is given. The camera represents a fundamental limit achieving 24 FPS. Processing of a frame by computer vision algorithms to extract the object location adds delay, however, for multi target tracking parallel approaches can reduce the inference time. After obtaining the object position within a camera frame, the multi-object distance measurement is executed sequentially. The main components limiting the bandwidth are the FSM and the LRF. As stated in Section 2 the LRF has a bandwidth of 117 Hz or 8.5 ms between each measurement. The FSM has a -3 dB bandwidth of 750 Hz for small drive angles of a few millidegrees. To support coverage of the complete camera FoV, the mirror has to be operated in the full angular range of $\pm$ 1.5°, which reduces the bandwidth to 154 Hz or a settling time of 6.5 ms. To obtain the settling time, a step response is applied as an input signal to the FSM, which results in a 3° mirror step. The time between the application of the input signal and the first intersection with the step response is defined as settling time, as the FSM controller is tuned to have a small overshoot. The time of communication components responsible for sending position commands to the FSM is neglected as it is in the range of microseconds.

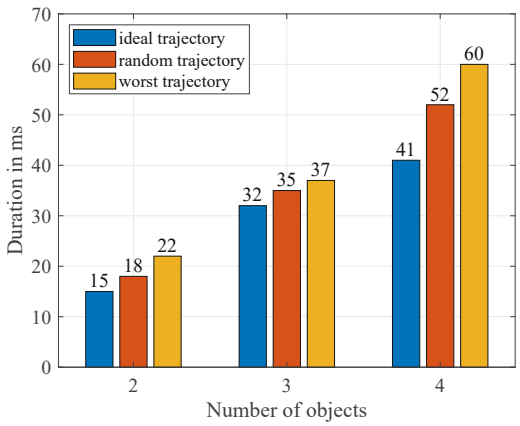


Fig. 5. Experimentally measured time required by the LRF and the FSM to measure the distance to multiple objects depending on the trajectory travelled by the FSM.

Fig. 5 shows the time needed by the LRF and the FSM to measure between 2 to 4 stationary objects comparing different trajectories taken by the FSM. The results indicate that a trajectory optimization already improves the performance when measuring the distance to two objects, as the path travelled by the FSM is reduced significantly compared to a random trajectory selection. The experimentally obtained bandwidth of the complete system, when applying trajectory optimization is 24 FPS for the measurement of one or two objects, 20 FPS for three

objects and 14 FPS for four objects, when using a single LRF measurement per object. Fig. 6 shows an example of a multi-object distance measurement. The blurry image is the result of a small dichroic mirror (cf. Fig. 1a) not covering the aperture L_1 of the camera and therefore creating two optical paths for the visual light, which passes through and beside the dichroic mirror. As the dichroic mirror has a different refractive index than the surrounding air, the two optical paths distort the resulting image quality. The image quality can be improved either by using a large dichroic mirror that fills up the entire aperture or by reducing the size of the optical setup of L_1 and L_2 to fit the dichroic mirror. For the proof of concept the image quality is sufficient and does not reduce the performance of the used KCF tracker.

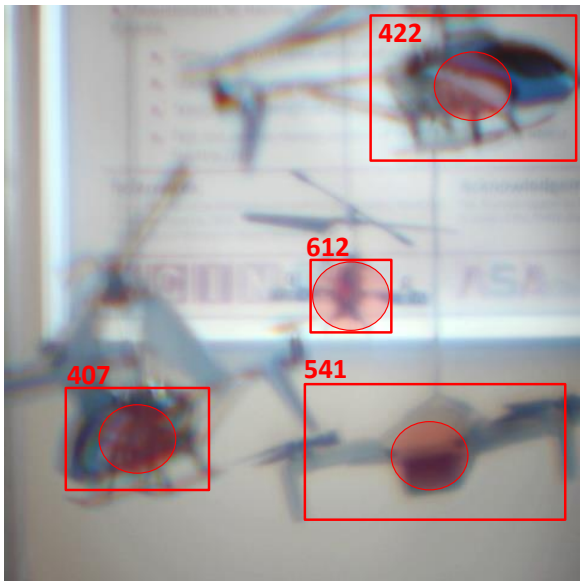


Fig. 6. Example of a multi-object distance measurement. The numbers on top of the bounding boxes indicates the distance to the objects in centimeter. The circle within the bounding boxes shows the size of the laser spot relative to the object size.

C. Local scanning

A crucial consideration for the measurement of small and complex shaped objects like UAVs is the inherent possibility of missing the target. A reason for a laser pulse miss is inaccurate localization provided by the computer vision algorithm combined with inadequate laser beam divergence. For single LRF measurements, the best approach

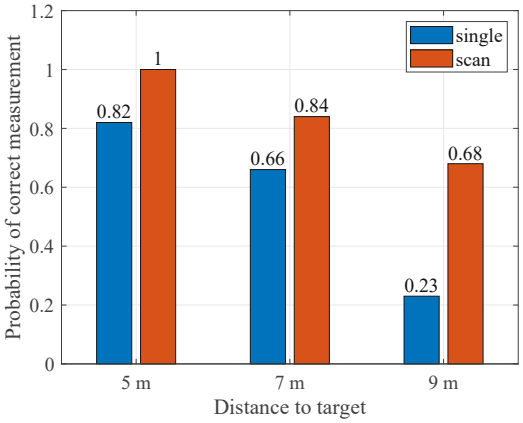


Fig. 7. Measured reliability of single LRF measurement compared to local scanning using 16 LRF measurements to obtain the distance. The bounding boxes are initialized with a random offset to the ideal ground truth bounding box. The random offset is in an interval between 1 and 0.5 regarding the intersection over union.

is to align the LRF with the bounding box center as depicted by the circles in Fig. 6. To reduce the number of miss measurements, the local scanning approach presented in Section 4 is used. For local scanning, a correct measurement is achieved, when the output of the clustering algorithm corresponds to the distance between UAV and the setup $\pm$ 0.1 m. Fig. 7 shows the results of comparing the single LRF measurement to the local scan, whereas a non ideal bounding box to object overlap is simulated. In object detection, an object is widely considered correctly detected, when the suggested bounding box overlaps with the ground truth bounding box by 50 % resulting in an intersection over union of 0.5 [45]. Therefore, for the evaluation, the tracker is initialized 50 times on a UAV with a random offset to the UAV center to simulate ideal and non-ideal tracker outputs. The offset between ground truth and initialized bounding box lies within 1 and 0.5 in terms of intersection over union and the reliability is calculated as the percentage of correct distance measurements. For the single measurement, a reduction in the measurement probability is observed for longer distances, which coincides with Fig. 3. If the bounding box is not accurately aligned with the target, not enough light is reflected to be detected by the LRF. The results clearly indicate an

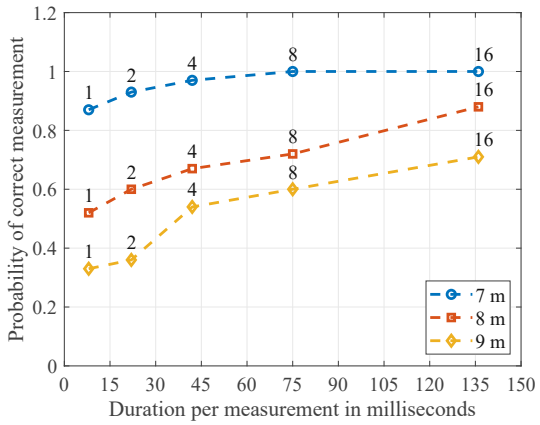


Fig. 8. Experimentally obtained probability of a correct distance inference, depending on the number of measurements performed by the local scanning. The markers denote 1, 2, 4, 8 and 16 measurements. The bounding box is initialized ideally on the object.

improved measurement reliability when using a local bounding box scan at the expense of a reduced system bandwidth. The impact on bandwidth is analysed in Fig. 8, where the probability of a correct measurement is plotted against the measurement duration. Different values for the duration are obtained by increasing the number of performed laser acquisitions from 1 to a maximum of 16 following a raster trajectory. Compared to Fig. 7 the one shot measurements score a higher probability, as the bounding box is initialized ideally on the UAV for this experiment. For longer distances, the probability of a correct measurement decreases, as the laser spot size is larger than the target object itself. By applying a local scanning measurement reliability can be increased.

The implemented small scale prototype system demonstrates real-time ranging of multiple UAVs enabling each second 14 distance measurements per object when measuring the distance to 4 objects. Furthermore, the probability to correctly measure the distance can be increased by introducing local scanning.

6. CONCLUSION

A scalable telescope based laser ranging system has been designed and developed that is capable to measure the distance to multiple UAVs using the positional information within a camera frame. The key concept is the combination of the optical paths of the visual light and the laser light, which allows fast laser to target alignment using an FSM. The laser beam is aligned based on information extracted by computer vision algorithms, making elaborate laser scanning of the whole FoV in search for a target redundant. This idea enables to perform a single LRF measurement to obtain the distance to a UAV, hence enabling multi UAV localization in real-time. With the proposed design and implemented prototype system a bandwidth of 14 Hz per object when measuring the distance to 4 objects and 20 Hz for 3 objects is achievable. The measurement reliability can be increased by introducing local scanning of the bounding box. Future work will focus on improving the measurement range, by incorporating a stronger laser with a smaller beam divergence and the integration of the system to a commercial telescope.

7. FUNDING

This publication has been funded by the Austrian defense research programme FORTE of the Federal Ministry of Agriculture, Regions and Tourism (BMLRT).

8. DISCLOSURES

The authors declare that there are no conflicts of interest related to this paper.

9. DATA AVAILABILITY

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. C. F. Liew, D. DeLatte, N. Takeishi, and T. Yairi, "Recent Developments in Aerial Robotics: A Survey and Prototypes Overview," ArXiv **abs/1711.10085** (2017).
2. J. Serna, "Lufthansa jet and drone nearly collide near LAX," Los Angeles Times (2016). Accessed Feb 2022.
3. C. Phillips and C. Gaffey, "Most French Nuclear Plants 'Should Be Shut Down' Over Drone Threat," Newsweek Magazine (2015). Accessed Feb 2022.
4. A. Solodov, A. Williams, S. A. Hanaei, and B. Goddard, "Analyzing the threat of unmanned aerial vehicles (UAV) to nuclear facilities," Security Journal **31**, 305–324 (2017).
5. S. Dinan, "Mexican drug cartels using drones to smuggle heroin, meth, cocaine into U.S." The Washington Times (2015). Accessed Feb 2022.
6. "Charges over drone drug smuggling into prisons," BBC News (2018). Accessed Feb 2022.

7. H. Duan, D. Zhang, Y. Shi, and Y. Deng, "Close formation flight of swarm unmanned aerial vehicles via metric-distance brain storm optimization," Memetic Comput. **10**, 369–381 (2018).
8. M. Gargalakos, "The role of unmanned aerial vehicles in military communications: application scenarios, current trends, and beyond," Journal of Defense Modeling and Simulation: Applications, Methodology, Technology. (2021).
9. S. Yang, H. Qin, X. Liang, and T. Gulliver, "An Improved Unauthorized Unmanned Aerial Vehicle Detection Algorithm Using Radiofrequency-Based Statistical Fingerprint Analysis," Sensors. **19**, 274 (2019).
10. W. Nie, Z.-C. Han, M. Zhou, L.-B. Xie, and Q. Jiang, "UAV detection and identification based on WiFi signal and RF fingerprint," IEEE Sensors J. **21**, 13540–13550 (2021).
11. V. Baron, S. Bouley, M. Muschinowski, J. Mars, and B. Nicolas, "Drone localization and identification using an acoustic array and supervised learning," in *Artificial Intelligence and Machine Learning in Defense Applications,* (11169F SPIE, 2019).
12. A. Sedunov, D. Haddad, H. Salloum, A. Sutin, N. Sedunov, and A. Yakubovskiy, "Stevens drone detection acoustic system and experiments in acoustics UAV tracking," in *2019 IEEE International Symposium on Technologies for Homeland Security (HST),* (IEEE, 2019).
13. S. Hu, G. H. Goldman, and C. C. Borel-Donohue, "Detection of unmanned aerial vehicles using a visible camera system," Appl. Opt. **56**, B214 (2017).
14. J. Farlik, M. Kratky, J. Casar, and V. Stary, "Multispectral Detection of Commercial Unmanned Aerial Vehicles," Sensors. **19**, 1517 (2019).
15. M. Rebert, "Prediction of micro-unmanned aerial vehicle flight behavior from two-dimensional intensity images," Opt. Eng. **58**, 1 (2019).
16. H. Jiang, L. Ding, Z. Sun, and R. Huang, "Unsupervised monocular depth perception: Focusing on moving objects," IEEE Sensors J. **21**, 27225–27237 (2021).
17. Z. Zhang, Y. Han, Y. Zhou, and M. Dai, "A novel absolute localization estimation of a target with monocular vision," Optik. **124**, 1218–1223 (2013).
18. A. Zaarane, I. Slimani, W. Al Okaishi, I. Atouf, and A. Hamdoun, "Distance measurement system for autonomous vehicles using stereo camera," Array. **5**, 100016 (2020).
19. A. D. De Quevedo, F. I. Urzaiz, J. G. Menoyo, and A. A. Lopez, "Drone Detection and RCS Measurements with Ubiquitous Radar," in *2018 International Conference on Radar (RADAR),* (2018), pp. 1–6.
20. T. Bosch, "Laser ranging: a critical review of usual techniques for distance measurement," Opt. Eng. **40**, 10 (2001).
21. X. Zhao, P. Sun, Z. Xu, H. Min, and H. Yu, "Fusion of 3D LiDAR and camera data for object detection in autonomous vehicle applications," IEEE Sensors J. **20**, 4901–4913 (2020).
22. M. Salhi and N. Boudriga, "Multi-Array Spherical LiDAR System for Drone Detection," in *2020 22nd International Conference on Transparent Optical Networks (ICTON),* (IEEE, 2020).
23. M. Hammer, M. Hebel, M. Laurenzis, and M. Arens, "LiDAR-based detection and tracking of small UAVs," in *Emerging Imaging and Sensing Technologies for Security and Defence III; and Unmanned Sensors, Systems, and Countermeasures,* vol. 10799 (2018), pp. 177 – 185.
24. G. Zhou, X. Zhou, J. Yang, Y. Tao, X. Nong, and O. Baysal, "Flash LiDAR sensor using fiber-coupled APDs," IEEE Sensors J. **15**, 4758–4768 (2015).
25. R. Horaud, M. Hansard, G. Evangelidis, and C. Ménier, "An overview of depth cameras and range scanners based on time-of-flight technologies," Mach. Vis. Appl. **27**, 1005–1020 (2016).
26. J. Martinez, A. Pequeno-Boter, A. Mandow, A. Garcia-Cerezo, and J. Morales, "Progress in Mini-Helicopter Tracking With a 3D Laser Range-Finder," IFAC Proc. Vol. **38**, 648–653 (2005).
27. A. Pawlikowska, R. Pilkington, K. Gordon, P. Hiskett, G. Buller, and R. Lamb, "Long-range 3D single-photon imaging LiDAR system," Proc. SPIE - Electro-Optical Remote. Sensing, Photonic Technol. Appl. VIII; Mil. Appl. Hyperspectral Imaging High Spatial Resolut. Sens. II **9250** (2014).
28. B. Kim, D. Khan, C. Bohak, W. Choi, H. Lee, and M. Kim, "V-RBNN Based Small Drone Detection in Augmented Datasets for 3D LADAR System," Sensors **18**, 3825 (2018).
29. A. G. Ledebuhr, J. F. Kordas, and I. T. Lewis, "HiRes camera and LiDAR ranging system for the Clementine mission," in *SPIE international symposium on aerospace/defense sensing and dual-use photonics,* (1995).
30. D. Joshi, D. Kumar, A. K. Maini, and R. C. Sharma, "Detection of biological warfare agents using ultra violet-laser induced fluorescence LIDAR," Spectrochimica Acta Part A: Mol. Biomol. Spectrosc. **112**, 446–456 (2013).
31. Z. Kong, T. Ma, Y. Cheng, Z. Zhang, Y. Li, K. Liu, and L. Mei, "Feasibility investigation of a monostatic imaging LiDAR with a parallel-placed image sensor for atmospheric remote sensing," J. Quant. Spectrosc. Radiat. Transf. **254**, 107212 (2020).
32. B.-H. Sheu, C.-C. Chiu, W.-T. Lu, C.-I. Huang, and W.-P. Chen, "Development of UAV tracing and coordinate detection method using a dual-axis rotary platform for an anti-UAV system," Appl. Sci. **9**, 2583 (2019).
33. A. Hommes, A. Shoykhetbrod, D. Noetel, S. Stanko, M. Laurenzis, S. Hengy, and F. Christnacher, "Detection of acoustic, electro-optical and RADAR signatures of small unmanned aerial vehicles," in *SPIE Proceedings,* (SPIE, 2016).
34. F. Christnacher, S. Hengy, M. Laurenzis, A. Matwyschuk, P. Naz, S. Schertzer, and G. Schmitt, "Optical and acoustical UAV detection," in *Electro-Optical Remote Sensing X,* (SPIE, 2016).
35. Aselsan, "IHTAR anti-drone system - datasheet," ASELSAN A.S., Ankara, Türkiye (2018).
36. D. Ojdanić, A. Sinn, C. Naveschnigg, and G. Schitter, "Feasibility Analysis of Optical UAV Detection Over Long Distances Using Robotic Telescopes ," IEEE Transactions on Aerosp. Electron. Syst. Submitted Mar. 2022.
37. R. D. Richmond and S. C. Cain, *Direct-Detection LADAR Systems* (SPIE Press, 2010).
38. J. Wojtanowski, M. Zygmunt, M. Kaszczuk, Z. Mierczyk, and M. Muzal, "Comparison of 905 nm and 1550 nm semiconductor laser rangefinders' performance deterioration due to adverse environmental conditions," Opto-Electronics Rev. **22** (2014).
39. S. Gundacker and A. Heering, "The silicon photomultiplier: fundamentals and applications of a modern solid-state photon detector," Phys. Medicine & Biol. **65**, 17TR01 (2020).
40. D. Bastos, P. P. Monteiro, A. S. Oliveira, and M. V. Drummond, "An overview of lidar requirements and techniques for autonomous driving," in *2021 Telecoms Conference (ConfTELE),* (IEEE, 2021), pp. 1–6.
41. H. Masoumi, S. S. Mohsen, and K. Zahra., "Identification and classification of plastic resins using near infrared reflectance spectroscopy," International Journal of Mechanical and Industrial Engineering. **6**, 213–220 (2012).
42. S. Zhu, H. Chen, M. Wang, X. Guo, Y. Lei, and G. Jin, "Plastic solid waste identification system based on near infrared spectroscopy in combination with support vector machine," Advanced Industrial and Engineering Polymer Research. **2**, 77–81 (2019).
43. "OpenCV Tracking API," Accessed: Feb. 2022.
44. Inframet, "LAS Test Station - inframet.com," Accessed: 2021-11-29.
45. R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP),* (IEEE, 2020).

CHAPTER 4

Conclusion and Future Work

Commercially available small UAVs pose a significant threat to critical infrastructure and public safety corroborating the utmost importance of further research and development on existing and new counter-UAV technologies. Optical sensors remain a key asset in any UAV detection system as visual imagery offers easily interpretable data of unidentified airborne objects. Within this thesis an optical UAV detection system has been developed that significantly increases the optical detection distance for micro and mini UAVs. Combining appropriate telescopes and cameras with an efficient software architecture, which is responsible to detect and track UAVs, and a custom high-speed autofocus solution, a holistic optical UAV detection system is presented. Additionally, potential optical configurations are investigated to enable integration of a laser range finder for an accurate real-time multi UAV 3D localization.
This last chapter elaborates the answers to the research questions raised in Section 2.5, followed by an outlook on future work.

4.1 Conclusion

To investigate solutions for long distance optical UAV detection using a robotic telescope system, seven research questions have been formulated in Section 2.5. In the following the questions are listed again with the corresponding answers drawn based on the results of this thesis.

Research Question 1:
Is it feasible to detect mini and micro UAVs with a diameter of 0.3 m in a distance of more than 2000 m with an optical UAV detection system?

Although optical UAV detection is a highly researched field over the past decade, current systems are limited through the usage of refractive camera lenses and aperture sizes. Reflective telescopes allow a solution to enable long focal lengths and large apertures, while maintaining comparably lightweight system characteristics. As thoroughly analysed and experimentally evaluated, feasibility for long range operation is proven using a telescope-based system, which can detect and track micro and mini UAVs of 0.3 m in diameter up to 4000 m away using an f/10 telescope with a focal length of 2540 mm [P1, P2, P3].

Research Question 2:
What optical resolution is required for a deep learning-based optical system to detect micro and mini UAVs of the size 0.3 m in diameter reliably?

Closely related to the previous research question the required resolution for a reliable deep learning-based optical UAV detection is experimentally evaluated [P2]. For a reliable detection in front of a clear background a minimal angular resolution of 17.1 µrad and for a complex background 8 µrad are necessary. Using Eq. 2.1, with the aperture of 102 mm of the telescope used during these experiments, the diffraction limit of the system is around 6 µrad. A reliable detection is defined as a mAP(0.5) of more than 90 %.

Research Question 3:
Which software architecture and transition strategy enable an efficient image-based detection and tracking of UAVs at frame rates of 100 fps?

A crucial aspect for a long range detection system is a low processing latency and a high sampling rate of the detection and tracking to enable suitable control strategies to actuate the telescope mount and follow the highly dynamic movements of the UAV. Therefore, reliable, fast and fully automatic object detection and tracking is necessary. In order to achieve high performance, deep learning and traditional algorithms are combined within a parallel architecture to achieve high frame rates, while maintaining high accuracy [P3]. A slow but accurate deep learning-based object detector is responsible to detect and, if necessary, correct a fast, but less reliable traditional object tracker, which is running in parallel on a different thread and CPU. A collaboration between the two algorithms is facilitated via a transition strategy, which is based on the detection probability and the tracker reliability. Using this transition strategy, an IOU at a threshold of 0.5 of 65 % for a clear and 53 % for a complex background is achieved compared to 17 % for a clear and 18 % for a non-parallel detector-only approach at a camera frame rate of 100 fps. Likewise, the probability of two consecutive detections for a UAV visible within the FoV of the system for a mere 162 ms is 78 % compared to 29 % for a non-parallel approach. The proposed parallel architecture ensures reliable detection and tracking of UAVs at 100 fps, which enables a highly dynamic

motion control of the telescope mount.

Research Question 4:
Does a combination of a deep learning-based object detector and a traditional tracker exist, that enables a reliable detection and tracking of UAVs during daytime conditions?

In continuation with the previous research question, the determination of a suitable pair of detection and tracking algorithms is a crucial aspect to maximize the performance of the parallel architecture. For this purpose, four state of the art deep learning object detection algorithms have been trained and evaluated on a custom dataset together with three traditional object trackers [P4]. The best performance in terms of IOU and CLO is achieved by FRCNN detector and Medianflow tracker. Among the investigated trackers, Medianflow is the most reliable, while enabling tracking at 100 fps. Combined with FRCNN, the most accurate, but also one of the slowest investigated object detectors achieving 21 fps on a PC with equipped with an RTX 3080 GPU, the parallel architecture enables reliable detection and tracking of UAVs at 100 fps.

Research Question 5:
Is the detection performance of deep learning-based detection algorithms maximized by using color or greyscale images as input during daytime conditions?

The optical system is highly dependant on the object detector, as it is the key component, which enables to localize a UAV and to initiate tracking and the actuation of the mount. The object detector localizes UAVs within captured images, whereas these images can have different underlying color representations. To maximize the detection performance during daytime conditions, a comparison between the most common color spaces, greyscale and RGB is performed [P5]. For training and testing, a color dataset is used captured during daytime conditions, which is converted to greyscale, to emulate an equal resolution and quantum efficiency between color and greyscale camera. In terms of mAP the algorithms trained on color images unanimously outperform the algorithms trained on greyscale images, as the three different channels within the images offer more distinctive information.

Research Question 6:
Is it feasible to design a high-speed automatic focus module for a telescope-based UAV detection system to keep a mini or micro UAV, flying up to its maximum speed, in focus?

As discussed in the first research question, a large aperture size is necessary to ensure sufficient resolution for long distance UAV detection. However, long focal

lengths with large apertures reduce the depth of field of the optical system and hence, impose challenging requirements on the automatic focus. In this thesis a custom high-speed automatic focus module is implemented, which augments the software architecture with additional parallel threads, used to calculate the focus value and the necessary focus position of the camera sensor with respect to the telescope [P6]. A passive contrast-based automatic focus is selected, as it only relies on the captured images and the camera position relative to the telescope. The camera is moved with respect to the telescope using a linear stage to adjust the focus. The contrast-based focus is integrated with minimal additional hardware compared to phase detect or active autofocus. The results of extensive field tests show that the usage of a passive contrast-based focus calculation using the Tenengrad operator together with the Hill Climb optimization, paired with a fast and accurate linear stage, keeps fast UAVs approaching at speeds of up to 24 m/s in close distances of down to 150 m in focus. To achieve this, the camera is moved by the linear stage with a velocity of up to 12 mm/s.

Research Question 7:
Can a laser range finder be integrated to a telescope-based UAV detection system to enable real-time distance measurement of multiple UAVs?

The implemented telescope system can detect and track the highly dynamic trajectories of UAVs through the reliable parallel architecture and the custom automatic focus module. However, a distance estimation to a detected UAV is challenging solely based on the captured images or the linear stage position of the automatic focus. Hence, an optical setup is investigated, which integrates a laser range finder into a telescope system for real-time UAV localization via the mount azimuth and altitude values, and a measured distance [P7]. A prototype system is designed, which consists of a telescope, camera, laser range finder and dichroic mirror, which enables a simple alignment of the laser with a fast steering mirror towards a detected UAV. It is demonstrated that up to 4 UAVs can be measured at 14 Hz per object using the proposed prototype. This allows real-time distance measurement and hence, an accurate 3D object localization.

4.2 Summary and Outlook

The goal of this thesis is to significantly increase the operational range of an optical UAV detection and tracking system for small commercially available UAVs during daytime conditions. This is achieved through the usage of high quality telescopes together with an efficient parallel software architecture integrating deep learning object detection and tracking algorithms. The implemented system enables detection and tracking of UAVs with a diameter of 0.3 m in distances between 150 m up to 4000 m at a frame rate of 100 fps [P1, P3]. A high-speed

automatic focus module ensures a sharp image is captured of the UAV flying at speeds of up to 24 m/s towards the system [P6]. Finally, an additional prototype is implemented showing a possible approach to integrate real-time laser distance measurement into a telescope system for an accurate multi UAV localization at 14 Hz per object for four objects [P7].

As this thesis focused on long range UAV detection mainly during daytime conditions, an interesting topic for further investigation is an evaluation of UAV detection during low light and night time scenarios, which may include the usage of either a different imaging spectrum [208] or active laser-based illumination [P7]. Likewise, a further increase of the detection range above 4000 m during daytime conditions remains a challenge, as atmospheric seeing degrades the image quality, reducing the signal produced by a UAV with the diameter of 0.3 m to a blurry dot within the camera image [P1 - P3]. However, application of super resolution algorithms to enhance the captured image [209] or also the tracking of the blurry dots with subsequently flight behaviour analysis may further increase the operational range.

Furthermore, the analysis of the computer vision algorithms presented within this thesis show the challenge of detecting and tracking UAVs below horizon or in front of a complex background [P2, P3]. Additional investigation is required to enable a more reliable tracking performance under such circumstances, for example, through the usage of video object detection [210] instead of image-based detection or also incorporating ensemble methods [211]. As UAV speeds are expected to increase in the future, the investigated real-time parallel architecture, which enables a sampling rate of 100 fps will be a crucial aspect to enable dynamic motion control of the telescope system to follow the fast and agile UAV movements [P3]. With the advent of deep learning dominating the field of computer vision, future computing systems combined with research and development on the algorithms themselves, will enable even higher frame rates with more accurate detection.

Another potentially intriguing aspect is the extension of the dataset, which has been created during the course of this thesis. The dataset consists of about 20500 images and 25000 videos images of UAVs [P4]. For future work, this can be extended by adding further airborne classes like birds, planes or helicopters.

Moreover, the integration and extensive evaluation of the proposed prototypic laser range system into the dual telescope system, developed within this thesis, is of interest [P7]. The laser range system can be used for measuring the distance to tracked UAVs for accurate localization, for implementing an active automatic focus, but also as a source of illumination in low light scenarios.

The presented system significantly extends the operational range for EO UAV detection systems, which enables early threat identification and prolonged reaction times. The system can be used in a stand-alone mode by adding wide FoV cameras for the initial object detection followed by a detailed high quality imaging using the presented system for classification and identification. However, the detection distances are limited by the wide FoV cameras. To fully exploit the long range

optical capabilities of the presented system, an integration with additional sensors offering a 360° coverage, such as radar or RF, is beneficial. A radar system, for example, delivers coordinates to an unidentified flying object in a long distance and by using the presented system, an image can be captured for subsequent classification and treat analysis. A seamless integration and forwarding of the captured data by the telescope towards a command and control software for the holistic multispectral UAV detection system via a dedicated interface is required. The captured data by the telescope, which includes a high quality image of the UAV and a precise localization using the azimuth and altitude coordinates together with the laser distance measurement, can serve as a basis to evaluate potential hazardous situations. An additional integration of counter measures, which have no recoil, such as jamming or spoofing, onto the telescope system itself might be possible to utilize the precise directional alignment towards the incoming UAV.

Bibliography

[1] Chun Fui Liew, Danielle DeLatte, Naoya Takeishi, and Takehisa Yairi. Recent Developments in Aerial Robotics: A Survey and Prototypes Overview. *ArXiv*, abs/1711.10085, 2017.

[2] JU SESAR. European drones outlook study unlocking the value for Europe. *Siebert, JU, Nov*, 2016.

[3] Ioannis Mademlis, Nikos Nikolaidis, Anastasios Tefas, Ioannis Pitas, Tilman Wagner, and Alberto Messina. Autonomous UAV cinematography: A tutorial and a formalized shot-type taxonomy. *ACM Computing Surveys (CSUR)*, 52(5):1–33, 2019.

[4] Rory Collean-Jones. Drone racing - sport of the future? BBC News, April 2016. `https://www.bbc.com/news/technology-36045577` Accessed: Mar 2024.

[5] Martin Peter Christiansen, Morten Stigaard Laursen, Rasmus Nyholm Jørgensen, Søren Skovsen, and René Gislum. Designing and testing a UAV mapping system for agricultural field surveying. *Sensors*, 17(12):2703, 2017.

[6] Dimosthenis C Tsouros, Stamatia Bibi, and Panagiotis G Sarigiannidis. A review on UAV-based applications for precision agriculture. *Information*, 10(11):349, 2019.

[7] David Erdos, Abraham Erdos, and Steve E Watkins. An experimental UAV system for search and rescue challenge. *IEEE Aerospace and Electronic Systems Magazine*, 28(5):32–37, 2013.

[8] D Mader, R Blaskow, P Westfeld, and C Weller. Potential of UAV-based laser scanner and multispectral camera data in building inspection. *The*

international archives of the photogrammetry, remote sensing and spatial information sciences, 41:1135–1142, 2016.

[9] Fatemeh Afghah, Abolfazl Razi, Jacob Chakareski, and Jonathan Ashdown. Wildfire monitoring in remote areas using autonomous unmanned aerial vehicles. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops*, pages 835–840. IEEE, 2019.

[10] Bi, Z., Chen, H., Hu, J., et. al. Analysis of UAV Typical War Cases and Combat Assessment Research. In *2022 IEEE International Conference on Unmanned Systems (ICUS)*, pages 1449–1453, 2022.

[11] 'Suspected drone' disrupts Gatwick Airport flights. BBC News, May 2023. `https://www.bbc.com/news/uk-england-sussex-65591023` Accessed Aug 2023.

[12] Flugdrohne: Frankfurt-Flughafen voruebergehend lahmgelegt. t-online, March 2020. `https://www.t-online.de/nachrichten/panorama/id_87443026/betrieb-eingestellt-flugdrohne-frankfurt-flughafen-voruebergehend-lahmgelegt.html` Accessed Mar 2024.

[13] Joseph Serna. Lufthansa jet and drone nearly collide near LAX. Los Angeles Times, March 2016. `https://www.latimes.com/local/lanow/la-me-ln-drone-near-miss-lax-20160318-story.html` Accessed Feb 2022.

[14] Drone collides with commercial aeroplane in canada. BBC News, October 2017. `https://www.bbc.com/news/technology-41635518` Accessed Feb 2024.

[15] Drone over windsor came close to british airways plane, report says. BBC News, December 2023. `https://www.bbc.com/news/uk-england-berkshire-67648285` Accessed Feb 2024.

[16] Reuters. Ireland vows to tackle drones after dublin airport shut six times, 2023. `https://www.reuters.com/world/uk/ireland-vows-tackle-drones-after-dublin-airport-shut-six-times-2023-03-03/` Accessed: (05.02.2024).

[17] Joey Klender. Drone operator above tesla giga berlin spoils routine descent for passenger plane, 2022. `https://www.teslarati.com/tesla-giga-berlin-drone-operator-berlin-brandenburg-airport-plane/` Accessed: (05.02.2024).

[18] James Field. Drone operator above tesla giga berlin spoils routine descent for passenger plane, 2023. `https://aviationsourcenews.com/incident/emirates-a380-struck-by-drone-wing-damage-in-nice-france/` Accessed: (05.02.2024).

[19] Alexander Solodov, Adam Williams, Sara Al Hanaei, and Braden Goddard. Analyzing the threat of unmanned aerial vehicles (UAV) to nuclear facilities. *Security Journal*, 31(1):305–324, Apr 2017.

[20] Catherine Phillips and Conor Gaffey. Most French Nuclear Plants 'Should Be Shut Down' Over Drone Threat. Newsweek Magazine, 2015. http://europe.newsweek.com/most-french-nuclear-plants-should-be-shut-down-over-drone-threat-309019. Accessed Feb 2022.

[21] Sweden drones: Sightings reported over nuclear plants and palace. BBC News, January 2022. https://www.bbc.com/news/world-europe-60035446 Accessed Aug 2023.

[22] Will Ripley. Drone with radioactive material found on Japanese Prime Minister's roof. CNN, 2015. https://edition.cnn.com/2015/04/22/asia/japan-prime-minister-rooftop-drone/index.html Accessed: (06.03.2024).

[23] Venezuela President Maduro survives 'drone assassination attempt'. BBC, 2018. https://www.bbc.com/news/world-latin-america-45073385 [Accessed: (06.03.2024)].

[24] Stephen Dinan. Mexican drug cartels using drones to smuggle heroin, meth, cocaine into U.S. The Washington Times, 2015. https://www.washingtontimes.com/news/2017/aug/20/mexican-drug-cartels-using-drones-to-smuggle-heroi/ Accessed Feb 2022.

[25] Charges over drone drug smuggling into prisons. BBC News, 2018. URL \url{https://www.bbc.com/news/uk-england-43413134}AccessedFeb2022.

[26] Sean Hollister. DA tiny DJI drone smuggled its own weight in drugs over the US border wall, 2021. https://www.theverge.com/2022/2/3/22916246/dji-mini-2-drone-smuggle-meth-us-mexico-border-wall [Accessed: (05.02.2024)].

[27] Drone incident management at aerodromes part 1. EASA - European Union Aviation Safety Agency, March 2021. https://www.easa.europa.eu/en/drone-incident-management-aerodromes-part-1 Accessed Mar 2024,.

[28] Jedrzej Drozdowicz, Maciej Wielgo, Piotr Samczynski, Krzysztof Kulpa, Jaroslaw Krzonkalla, Maj Mordzonek, Marcin Bryl, and Zbigniew Jakielaszek. 35 GHz FMCW drone detection system. In *2016 17th International Radar Symposium (IRS)*, pages 1–4. IEEE, 2016.

[29] Sedat Dogru and Lino Marques. Drone Detection Using Sparse LiDAR Measurements. *IEEE Robotics and Automation Letters*, 7(2):3062–3069, 2022.

[30] Phuc Nguyen, Mahesh Ravindranatha, Anh Nguyen, Richard Han, and Tam Vu. Investigating cost-effective rf-based detection of drones. In *Proceedings of the 2nd workshop on micro aerial vehicle networks, systems, and applications for civilian use*, pages 17–22, 2016.

[31] József Mezei, Viktor Fiaska, and András Molnár. Drone sound detection. In *16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*, pages 333–338. IEEE, 2015.

[32] Halil Utku Unlu, Phillip Stefan Niehaus, Daniel Chirita, Nikolaos Evangeliou, and Anthony Tzes. Deep learning-based visual tracking of UAVs using a PTZ camera system. In *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, Oct. 2019.

[33] ctrl+sky drone detection and neutralization system - datasheet. Advanced Protection Systems LLC, New Yersey, USA, 2018.

[34] Jan Farlik, Miroslav Kratky, Josef Casar, and Vadim Stary. Multispectral Detection of Commercial Unmanned Aerial Vehicles. *Sensors*, 19(7):1517, Mar. 2019.

[35] European Union Aviation Safety Agency (EASA). Easy access rules for unmanned aircraft systems (regulations (eu) 2019/947 and (eu) 2019/945), 2022.

[36] Bas Vergouw, Huub Nagel, Geert Bondt, and Bart Custers. Drone technology: Types, payloads, applications, frequency spectrum issues and future developments. *The future of drone use: Opportunities and threats from ethical and legal perspectives*, pages 21–45, 2016.

[37] DJI Mini 2. `https://store.dji.com/at/product/mini-2`. Accessed: 22.03.2024.

[38] Makeflyeasy striver mini binary 2100mm UAV Fixed Wing. `https://www.uavmodel.com/products/makeflyeasy-striver-mini-binary-2100mm-uav-fixed-wing?variant=41656279629978/`. Accessed: 22.03.2024.

[39] Makeflyeasy HERO 2180mm UAV VTOL. `https://www.uavmodel.com/collections/vtol/products/makeflyeasy-hero-2180mm-uav-vtol`. Accessed: 22.03.2024.

[40] Spreading Wings S1000. https://www-v1.dji.com/spreading-wings-s1000.html. Accessed: 22.03.2024.

[41] Do-DT55neo. https://www.airbus.com/en/products-services/defence/uas/uas-solutions/target-drone-systems. Accessed: 22.03.2024.

[42] Chinese university sets Guinness record in drone flight duration. https://www.globaltimes.cn/page/202210/1277794.shtml. Photo: Xinhua, Accessed: 22.03.2024.

[43] Phuc Nguyen, Taeho Kim, Jinpeng Miao, Daniel Hesselius, Erin Kenneally, Daniel Massey, Eric Frew, Richard Han, and Tam Vu. Towards rf-based localization of a drone and its controller. In *Proceedings of the 5th workshop on micro aerial vehicle networks, systems, and applications*, pages 21–26, 2019.

[44] Igor Bisio, Chiara Garibotto, Fabio Lavagetto, Andrea Sciarrone, and Sandro Zappatore. Unauthorized Amateur UAV Detection Based on WiFi Statistical Fingerprint Analysis. *IEEE Communications Magazine*, 56(4): 106–111, 2018.

[45] Shengying Yang, Huibin Qin, Xiaolin Liang, and Thomas Aaron Gulliver. An improved unauthorized unmanned aerial vehicle detection algorithm using radiofrequency-based statistical fingerprint analysis. *Sensors*, 19(2), 2019. ISSN 1424-8220.

[46] Dedrone. https://www.dedrone.com/products/hardware/rf-sensors/rf-360. Accessed: 10.01.2022.

[47] DJI Aeroscope. https://www.dji.com/at/aeroscope. Accessed: 08.11.2023.

[48] Martins Ezuma, Fatih Erden, Chethan Kumar Anjinappa, Ozgur Ozdemir, and Ismail Guvenc. Micro-UAV Detection and Classification from RF Fingerprints Using Machine Learning Techniques. In *2019 IEEE Aerospace Conference*, pages 1–13, 2019.

[49] Caidan Zhao, Caiyun Chen, Zhibiao Cai, Mingxian Shi, Xiaojiang Du, and Mohsen Guizani. Classification of Small UAVs Based on Auxiliary Classifier Wasserstein GANs. In *2018 IEEE Global Communications Conference (GLOBECOM)*, page 206, 2018.

[50] Martins Ezuma, Fatih Erden, Chethan Kumar Anjinappa, Ozgur Ozdemir, and Ismail Guvenc. Detection and Classification of UAVs Using RF Fingerprints in the Presence of Wi-Fi and Bluetooth Interference. *IEEE Open Journal of the Communications Society*, 1:60–76, 2020.

[51] Chengtao Xu, Bowen Chen, Yongxin Liu, Fengyu He, and Houbing Song. Rf fingerprint measurement for detecting multiple amateur drones based on stft and feature reduction. In *2020 Integrated Communications Navigation and Surveillance Conference (ICNS)*, pages 4G1–1–4G1–7, 2020.

[52] Sanjoy Basak, Sreeraj Rajendran, Sofie Pollin, and Bart Scheers. Drone classification from rf fingerprints using deep residual nets. In *International Conference on COMmunication Systems and NETworkS (COMSNETS)*, pages 548–555, 2021.

[53] Skylock. `https://www.skylock1.com/modular-components/counter-drone-systems/rf-detector-jammer/`. Accessed: 08.11.2023.

[54] Robert C Dixon. *Spread spectrum systems: with commercial applications*. John Wiley & Sons, Inc., 1994.

[55] Mark A Richards, Jim Scheer, William A Holm, and William L Melvin. Principles of modern radar. 2010.

[56] Hsueh-Jyh Li, Yean-Woei Kiang. *The Electrical Engineering Handbook*. Elsevier Inc., 2005.

[57] Tsai, Chin-Che and Chiang, Cheng-Tai and Liao, Wen-Jiao. Radar cross section measurement of unmanned aerial vehicles. In *IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM)*, pages 1–3, 2016.

[58] Teledyne FLIR. `https://www.flir.eu/products/ranger-r8ss-3d/`. Accessed: 05.03.2024.

[59] Angelo Coluccia, Gianluca Parisi, and Alessio Fascista. Detection and classification of multirotor drones in radar sensor networks: A review. *Sensors*, 20(15):4172, 2020.

[60] M. Caris, W. Johannes, S. Sieger, V. Port, and S. Stanko. Detection of small uas with w-band radar. In *18th International Radar Symposium (IRS)*, pages 1–6, 2017.

[61] Mohammed Jahangir and Chris Baker. Robust detection of micro-uas drones with l-band 3-d holographic radar. In *Sensor Signal Processing for Defence (SSPD)*, pages 1–5, 2016.

[62] De Quevedo, Allvaro Duque and Urzaiz, Fernando Ibanez and Menoyo, Javier Gismero and Lopez, Alberto Asensio. Drone detection and rcs measurements with ubiquitous radar. In *International Conference on Radar (RADAR)*, pages 1–6, 2018.

[63] Angelo Coluccia, Gianluca Parisi, and Alessio Fascista. Detection and classification of multirotor drones in radar sensor networks: A review. *Sensors*, 20(15), 2020. ISSN 1424-8220.

[64] Xin Guo, Chea Siang Ng, Erwin de Jong, and Adriaan B. Smits. Micro-Doppler Based Mini-UAV Detection with Low-Cost Distributed Radar in Dense Urban Environment. In *2019 16th European Radar Conference (EuRAD)*, pages 189–192, 2019.

[65] Samiur Rahman and Duncan A. Robertson. Radar micro-doppler signatures of drones and birds at k-band and w-band. *Scientific Reports*, 8(1), November 2018. ISSN 2045-2322.

[66] Robin Radar. `https://www.robinradar.com/drone-detection-radar`. Accessed: 05.03.2024.

[67] Aerial Armor - A Dedrone Company. `https://de.aerialarmor.com/drone-detection-equipment/drone-detection-radar-systems`. Accessed: 05.03.2024.

[68] Thierry Bosch. Laser ranging: a critical review of usual techniques for distance measurement. *Optical Engineering*, 40(1):10, Jan. 2001.

[69] Xiangmo Zhao, Pengpeng Sun, Zhigang Xu, Haigen Min, and Hongkai Yu. Fusion of 3D LiDAR and camera data for object detection in autonomous vehicle applications. *IEEE Sensors Journal*, 20(9):4901–4913, May 2020.

[70] Meriem Salhi and Noureddine Boudriga. Multi-Array Spherical LiDAR System for Drone Detection. In *2020 22nd International Conference on Transparent Optical Networks (ICTON)*, Jul. 2020.

[71] Marcus Hammer, Marcus Hebel, Martin Laurenzis, and Michael Arens. LiDAR-based detection and tracking of small UAVs. In *Emerging Imaging and Sensing Technologies for Security and Defence III; and Unmanned Sensors, Systems, and Countermeasures*, volume 10799, pages 177 – 185, 2018.

[72] Alain Quentel. *A scanning LiDAR for long range detection and tracking of UAVs*. PhD thesis, Normandie Université, 2021.

[73] Guoqing Zhou, Xiang Zhou, Jiazhi Yang, Yue Tao, Xueqin Nong, and Oktay Baysal. Flash LiDAR sensor using fiber-coupled APDs. *IEEE Sensors Journal*, 15(9):4758–4768, Sep. 2015.

[74] Radu Horaud, Miles Hansard, Georgios Evangelidis, and Clément Ménier. An overview of depth cameras and range scanners based on time-of-flight technologies. *Machine Vision and Applications*, 27(7):1005–1020, Jun. 2016.

[75] J.L. Martinez, A. Pequeno-Boter, A. Mandow, A. Garcia-Cerezo, and J. Morales. Progress in Mini-Helicopter Tracking With a 3D Laser Range-Finder. *IFAC Proceedings Volumes*, 38(1):648–653, 2005.

[76] Agata Pawlikowska, Roger Pilkington, Karen Gordon, Philip Hiskett, Gerald Buller, and Robert Lamb. Long-range 3D single-photon imaging LiDAR system. *Proceedings of SPIE - Electro-Optical Remote Sensing, Photonic Technologies, and Applications VIII; and Military Applications in Hyperspectral Imaging and High Spatial Resolution Sensing II*, 9250, Nov. 2014.

[77] Byeong Kim, Danish Khan, Ciril Bohak, Wonju Choi, Hyun Lee, and Min Kim. V-RBNN Based Small Drone Detection in Augmented Datasets for 3D LADAR System. *Sensors*, 18(11):3825, Nov. 2018.

[78] Ha Sier, Xianjia Yu, Iacopo Catalano, Jorge Pena Queralta, Zhuo Zou, and Tomi Westerlund. UAV tracking with LiDAR as a camera sensor in gnss-denied environments. In *2023 International Conference on Localization and GNSS (ICL-GNSS)*, pages 1–7. IEEE, 2023.

[79] Sedat Dogru and Lino Marques. Drone detection using sparse LiDAR measurements. *IEEE Robotics and Automation Letters*, 7(2):3062–3069, 2022.

[80] A G Ledebuhr, J F Kordas, and I T Lewis. HiRes camera and LiDAR ranging system for the Clementine mission. In *SPIE international symposium on aerospace/defense sensing and dual-use photonics*, 4 1995.

[81] Deepti Joshi, Deepak Kumar, Anil K. Maini, and Ramesh C. Sharma. Detection of biological warfare agents using ultra violet-laser induced fluorescence LIDAR. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 112:446–456, aug 2013.

[82] Zheng Kong, Teng Ma, Yuan Cheng, Zhen Zhang, Yichen Li, Kun Liu, and Liang Mei. Feasibility investigation of a monostatic imaging LiDAR with a parallel-placed image sensor for atmospheric remote sensing. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 254:107212, oct 2020.

[83] Bor-Horng Sheu, Chih-Cheng Chiu, Wei-Ting Lu, Chu-I Huang, and Wen-Ping Chen. Development of UAV tracing and coordinate detection method using a dual-axis rotary platform for an anti-UAV system. *Applied Sciences*, 9(13):2583, Jun. 2019.

[84] Joël Busset, Florian Perrodin, Peter Wellig, Beat Ott, Kurt Heutschi, Torben Rühl, and Thomas Nussbaumer. Detection and tracking of drones using advanced acoustic cameras. In *Unmanned/Unattended Sensors and Sensor Networks XI; and Advanced Free-Space Optical Communication Techniques*

and Applications, volume 9647, page 96470F. International Society for Optics and Photonics, SPIE, 2015.

[85] Ellen E. Case, Anne M. Zelnio, and Brian D. Rigling. Low-Cost Acoustic Array for Small UAV Detection and Tracking. In *IEEE National Aerospace and Electronics Conference*, pages 110–113, 2008.

[86] DISCOVAIR G2 - Acoustic CUAS. `https://www.sqhead.com/defense`. Accessed: 04.01.2024.

[87] Maksim Balakin, Anton Dvorak, and Daniil Kurylev. Real-time drone detection and recognition by acoustic fingerprint. In *2021 5th Scientific School Dynamics of Complex Networks and their Applications (DCNA)*, pages 44–45, 2021.

[88] Muhammad Zohaib Anwar, Zeeshan Kaleem, and Abbas Jamalipour. Machine learning inspired sound-based amateur drone detection for public safety applications. *IEEE Transactions on Vehicular Technology*, 68(3): 2526–2534, 2019.

[89] Zahoor Uddin, Muhammad Altaf, Muhammad Bilal, Lewis Nkenyereye, and Ali Kashif Bashir. Amateur drones detection: A machine learning approach utilizing the acoustic signals in the presence of strong interference. *Computer Communications*, 154:236–245, 2020. ISSN 0140-3664.

[90] Sara Al-Emadi, Abdulla Al-Ali, Amr Mohammad, and Abdulaziz Al-Ali. Audio based drone detection and identification using deep learning. In *15th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 459–464, 2019.

[91] Valentin Baron, Simon Bouley, Matthieu Muschinowski, JÃ©rÃ´me Mars, and Barbara Nicolas. Drone localization and identification using an acoustic array and supervised learning. In *Artificial Intelligence and Machine Learning in Defense Applications*, volume 11169, pages 129 – 137. SPIE, 2019.

[92] Alexander Sedunov, Darren Haddad, Hady Salloum, Alexander Sutin, Nikolay Sedunov, and Alexander Yakubovskiy. Stevens Drone Detection Acoustic System and Experiments in Acoustics UAV Tracking. In *IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–7, 2019.

[93] Minas Benyamin and Geoffrey H. Goldman. Acoustic detection and tracking of a class i uas with a small tetrahedral microphone array. 2014.

[94] Diana Tejera-Berengue, Fangfang Zhu-Zhou, Manuel Utrilla-Manso, Roberto Gil-Pita, and Manuel Rosa-Zurera. Analysis of Distance and Environmental Impact on UAV Acoustic Detection. *Electronics*, 13(3), 2024.

[95] SkyPatriot Sector. https://rinicom.com/skypatriot/. Accessed: 21.05.2024.

[96] AARTOS Camera. https://drone-detection-system.com/sensor-types-overview/detection-camera/. Accessed: 21.05.2024.

[97] Dedrone PTZ cameras. https://de.dedrone.com/products/drone-detection/extensions/ptz-cameras. Accessed: 21.05.2024.

[98] Tokio Ohta. *Solar-hydrogen energy systems: An authoritative review of water-splitting systems by solar beam and solar heat: Hydrogen production, storage and utilisation*. Elsevier, 2013.

[99] Dave Litwiller. Ccd vs. cmos. *Photonics spectra*, 35(1):154–158, 2001.

[100] John P. Janesick. *Charge-Coupled Devices for Imaging*. SPIE Press, 2001. ISBN 978-0-8194-3698-4.

[101] Albert Theuwissen. Cmos image sensors: State-of-the-art and future perspectives. In *ESSCIRC 2007 - 33rd European Solid-State Circuits Conference*, pages 21–27, 2007.

[102] Kohei Shimasaki, Nagahiro Fujiwara, Shaopeng Hu, Taku Senoo, and Idaku Ishii. High-frame-rate video-based multicopter tracking system using pixel-level short-time fourier transform. *Journal of Intelligent & Robotic Systems*, 103:1–24, 2021.

[103] Daisuke Matsuka and Masahiro Mimura. Surveillance system for multiple moving objects. *IEEJ Journal of Industry Applications*, 9(4):460–467, 2020.

[104] Ting Chen, Peter Bert Catrysse, Abbas El Gamal, and Brian A Wandell. How small should pixel size be? In *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, volume 3965, pages 451–459. SPIE, 2000.

[105] Teledyne Photometrics. Kinetix - the new category in scmos cameras datasheet. https://www.photometrics.com/wp-content/uploads/2023/03/Kinetix-Datasheet-Rev-A3-08032023.pdf, 2021.

[106] Mikhail Konnik and James Welsh. High-level numerical simulations of noise in ccd and cmos photosensors: review and tutorial. *arXiv preprint arXiv:1412.4031*, 2014.

[107] Rafael C Gonzalez. *Digital image processing*. Pearson education india, 2009.

[108] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

[109] Rayleigh. XXXI. Investigations in optics, with special reference to the spectroscope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 8(49):261–274, Oct. 1879.

[110] C.E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, jan 1949.

[111] Michael Schöberl, Siegfried Fößel, Hans Bloss, and André Kaup. Modeling of image shutters and motion blur in analog and digital camera systems. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 3457–3460. IEEE, 2009.

[112] M. Wany and G.P. Israel. Cmos image sensor with nmos-only global shutter and enhanced responsivity. *IEEE Transactions on Electron Devices*, 50(1): 57–62, 2003.

[113] DH Sliney. What is light? the visible spectrum and beyond. *Eye*, 30(2): 222–229, 2016.

[114] Maria Vasilopoulou, Azhar Fakharuddin, F Pelayo García de Arquer, Dimitra G Georgiadou, Hobeom Kim, Abd Rashid bin Mohd Yusoff, Feng Gao, Mohammad Khaja Nazeeruddin, Henk J Bolink, and Edward H Sargent. Advances in solution-processed near-infrared light-emitting diodes. *Nature Photonics*, 15(9):656–669, 2021.

[115] Yifan Yang, Qi Li, Chenwei Yang, Yannian Fu, Huajun Feng, Zhihai Xu, and Yueting Chen. Deep networks with detail enhancement for infrared image super-resolution. *IEEE Access*, 8:158690–158701, 2020.

[116] Ayan Chakrabarti, William T. Freeman, and Todd Zickler. Rethinking color cameras. In *2014 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8, 2014.

[117] Brandon Stark, Matthew McGee, and YangQuan Chen. Short wave infrared (swir) imaging systems using small unmanned aerial systems (suas). In *International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 495–501, 2015.

[118] Marc P. Hansen and Douglas S. Malchow. Overview of SWIR detectors, cameras, and applications. In Vladimir P. Vavilov and Douglas D. Burleigh, editors, *Thermosense XXX*, volume 6939, page 69390I. International Society for Optics and Photonics, SPIE, 2008.

[119] Ronald G. Driggers, Van Hodgkin, and Richard Vollmerhausen. What good is SWIR? Passive day comparison of VIS, NIR, and SWIR. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIV*, volume

8706, page 87060L. International Society for Optics and Photonics, SPIE, 2013.

[120] P.S. Thenkabail, J.G. Lyon, and A. Huete. *Hyperspectral Remote Sensing of Vegetation: Hyperspectral indices and image classifications for agriculture and vegetation*. Hyperspectral Remote Sensing of Vegetation. 2019. ISBN 9781138066250.

[121] MR Shortis, CJ Bellman, S Robson, GJ Johnston, and GW Johnson. Stability of zoom and fixed lenses used with digital slr cameras. In *Proceedings of the ISPRS Commission V Symposium of Image Engineering and Vision Metrology*, pages 285–290, 2006.

[122] Harrie GJ Rutten and Martin AM Van Venrooij. *Telescope optics: a comprehensive manual for amateur astronomers*. Willmann-Bell, 2002.

[123] Hiroshi Ohtani, Akira Uesugi, Yoshio Tomita, Michitoshi Yoshida, George Kosugi, Junichi Noumaru, Shoichi Araya, Kentaro Ohta, and Yoshihiro Mikayama. The 60cm ritchey-chrétien telescope and its instruments. *Memoirs of the Faculty of Science, Kyoto University. Series of physics, astrophysics, geophysics and chemistry*, 38(2):167–185, 1992.

[124] G. Grant Williams, Ed Olszewski, Michael P. Lesser, and James H. Burge. 90prime: a prime focus imager for the Steward Observatory 90-in. telescope. In Alan F. M. Moorwood and Masanori Iye, editors, *Ground-based Instrumentation for Astronomy*, volume 5492, pages 787 – 798. International Society for Optics and Photonics, SPIE, 2004.

[125] Jason Osborne, Gregory Hicks, and Robert Fuentes. Global analysis of the double-gimbal mechanism. *IEEE Control Systems Magazine*, 28(4):44–64, 2008.

[126] Artur Zawadzki and Marek Gorgon. Automatically controlled pan tilt smart camera with FPGA based image analysis system dedicated to real time tracking of a moving object. *Journal of Systems Architecture*, 61(10): 681–692, November 2015. ISSN 1383-7621.

[127] Yan Lui. Pan and tilt platform for a thermal vacuum video camera. In *24th Space Simulation Conference*, 2007.

[128] Shengxiong Wen, Yaowu Ding, Xuan Wu, and Kun Bai. Kinematic analysis and robust control of a spherical-motor-based visual tracking system. *IEEE/ASME Transactions on Mechatronics*, 28(4):1871–1879, 2023.

[129] Pablo Gutierrez. Standardization of direct drive servos in telescope applications. In *Large Ground-based Telescopes*, volume 4837, pages 325–335. SPIE, 2003.

[130] C. Naverschnigg, D. Ojdanic, A. Sinn and G. Schitter. Analysis and Control of a Robotic Telescope System for High-Speed Small-UAV Tracking. *IEEE Aerospace and Electronic Systems Magazine*, submitted Dez. 2023.

[131] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I, 2001.

[132] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1, 2005.

[133] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9): 1627–1645, 2009.

[134] R. Carnie, R. Walker, and P. Corke. Image processing algorithms for UAV "sense and avoid". In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.* IEEE, 2006.

[135] Debadeepta Dey, Christopher Geyer, Sanjiv Singh, and Matthew Digioia. A cascaded method to detect aircraft in video imagery. *The International Journal of Robotics Research*, 30(12):1527–1540, aug 2011.

[136] Danilo Avola, Luigi Cinque, Gian Luca Foresti, Cristiano Massaroni, and Daniele Pannone. A keypoint-based method for background modeling and foreground detection using a PTZ camera. *Pattern Recognition Letters*, 96: 96–105, sep 2017.

[137] Jeffrey W. McCandless. Detection of aircraft in video sequences using a predictive optical flow algorithm. *Optical Engineering*, 38(3):523, mar 1999.

[138] Bruce Lucas and Takeo Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision . *IJCAJ*, 81:674–679, 04 1981.

[139] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. *International Journal of Computer Vision*, 128(2):261–318, Oct. 2019.

[140] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[141] Lin, T., Maire, M., Belongie S., et. al. Microsoft COCO: Common Objects in Context. In *Computer Vision–ECCV 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[142] Ivan Krasin, Tom Duerig, Neil Alldrin, Vittorio Ferrari, Sami Abu-El-Haija, Alina Kuznetsova, Hassan Rom, Jasper Uijlings, Stefan Popov, Andreas Veit, et al. Openimages: A public dataset for large-scale multi-label and multi-class image classification. *Dataset available from https://github.com/openimages*, 2(3):18, 2017.

[143] Angelo Coluccia, Alessio Fascista, Arne Schumann, Lars Sommer, Anastasios Dimou, Dimitrios Zarpalas, Fatih Cagatay Akyon, Ogulcan Eryuksel, Kamil Anil Ozfuttu, Sinan Onur Altinuc, Fardad Dadboud, Vaibhav Patel, Varun Mehta, Miodrag Bolic, and Iraj Mantegh. Drone-vs-Bird Detection Challenge at IEEE AVSS2021. In *17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, nov 2021.

[144] Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. *International journal of computer vision*, 104:154–171, 2013.

[145] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[146] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916, 2015.

[147] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[148] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[149] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single Shot MultiBox Detector. In *Computer Vision – ECCV 2016*, pages 21–37. Springer International Publishing, Cham, 2016. ISBN 978-3-319-46448-0.

[150] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection, 2020.

[151] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327, 2020.

[152] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection, 2019.

[153] Hansen Liu, Kuangang Fan, Qinghua Ouyang, and Na Li. Real-Time Small Drones Detection Based on Pruned YOLOv4. *Sensors*, 21(10):3374, may 2021.

[154] Brian K. S. Isaac-Medina, Matt Poyser, Daniel Organisciak, Chris G. Willcocks, Toby P. Breckon, and Hubert P. H. Shum. Unmanned Aerial Vehicle Visual Detection and Tracking using Deep Neural Networks: A Performance Benchmark. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. IEEE, oct 2021.

[155] Jasmin James, Jason J. Ford, and Timothy L. Molloy. Learning to Detect Aircraft for Long-Range Vision-Based Sense-and-Avoid Systems. *IEEE Robotics and Automation Letters*, 3(4):4383–4390, oct 2018.

[156] Abdel Gafoor Haddad, Muhammad Ahmed Humais, Naoufel Werghi, and Abdulhadi Shoufan. Long-range visual UAV detection and tracking system with threat level assessment. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, Oct. 2020.

[157] Jihun Park, Dae Hoe Kim, Young Sook Shin, and Sangho Lee. A comparison of convolutional object detectors for real-time drone tracking using a PTZ camera. In *2017 17th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, Oct. 2017.

[158] Won Jeon, Gun Ko, Jiwon Lee, Hyunwuk Lee, Dongho Ha, and Won Woo Ro. Deep learning with gpus. In *Advances in Computers*, volume 122, pages 167–215. Elsevier, 2021.

[159] Matej Kristan, Jiri Matas, Ales Leonardis, Michael Felsberg, Roman Pflugfelder, Joni-Kristian Kamarainen, Luka Ë‡Cehovin Zajc, Ondrej Drbohlav, Alan Lukezic, Amanda Berg, et al. The seventh visual object tracking vot2019 challenge results. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, page 0, 2019.

[160] Dav Bolme, J. Ross Beveridge, Bruce A. Draper, and Yui Man Lui. Visual object tracking using adaptive correlation filters. In *2010 IEEE Computer*

Society Conference on Computer Vision and Pattern Recognition. IEEE, jun 2010.

[161] Joao F. Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. High-Speed Tracking with Kernelized Correlation Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):583–596, mar 2015.

[162] Martin Danelljan, Gustav Hager, Fahad Shahbaz Khan, and Michael Felsberg. Learning spatially regularized correlation filters for visual tracking. In *Proceedings of the IEEE international conference on computer vision*, pages 4310–4318, 2015.

[163] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Forward-Backward Error: Automatic Detection of Tracking Failures. In *2010 20th International Conference on Pattern Recognition*. IEEE, aug 2010.

[164] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-learning-detection. *IEEE transactions on pattern analysis and machine intelligence*, 34(7):1409–1422, 2011.

[165] Matej Kristan, Jiri Matas, Ales Leonardis, Michael Felsberg, Luka Cehovin, Gustavo Fernandez, Tomas Vojir, Gustav Hager, Georg Nebehay, and Roman Pflugfelder. The visual object tracking vot2015 challenge results. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 1–23, 2015.

[166] Matej Kristan, Jiří Matas, Aleš Leonardis, Michael Felsberg, Roman Pflugfelder, Joni-Kristian Kämäräinen, Hyung Jin Chang, Martin Danelljan, Luka Cehovin, Alan Lukežič, et al. The ninth visual object tracking vot2021 challenge results. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2711–2738, 2021.

[167] Martin Danelljan, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. Atom: Accurate tracking by overlap maximization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4660–4669, 2019.

[168] Chao Ma, Jia-Bin Huang, Xiaokang Yang, and Ming-Hsuan Yang. Hierarchical convolutional features for visual tracking. In *Proceedings of the IEEE international conference on computer vision*, pages 3074–3082, 2015.

[169] Philipp Bergmann, Tim Meinhardt, and Laura Leal-Taixe. Tracking Without Bells and Whistles. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, oct 2019.

[170] Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, and Xiaolin Hu. High performance visual tracking with siamese region proposal network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8971–8980, 2018.

[171] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, sep 2016.

[172] Roberto Opromolla and Giancarmine Fasano. Visual-based obstacle detection and tracking, and conflict detection for small uas sense and avoid. *Aerospace Science and Technology*, 119:107167, 2021. ISSN 1270-9638.

[173] Pedro Alexandre Prates, Ricardo MendonÃ§a, AndrÃ© LourenÃ§o, Francisco Marques, J. P. Matos-Carvalho, and JosÃ© Barata. Vision-based UAV detection and tracking using motion signatures. In *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*, pages 482–487, 2018.

[174] Jing Li, Dong Hye Ye, Mathias Kolsch, Juan P. Wachs, and Charles A. Bouman. Fast and Robust UAV to UAV Detection and Tracking From Video. *IEEE Transactions on Emerging Topics in Computing*, 10(3):1519–1531, 2022.

[175] Sungsik Huh, Sungwook Cho, Yeondeuk Jung, and David Hyunchul Shim. Vision-based sense-and-avoid framework for unmanned aerial vehicles. *IEEE Transactions on Aerospace and Electronic Systems*, 51(4):3427–3439, 2015.

[176] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv:1804.02767*, 2018.

[177] Dong-Hyun Lee. CNN-based single object detection and tracking in videos and its application to drone detection. *Multimedia Tools and Applications*, 80(26-27):34237–34248, oct 2020.

[178] Heng Fan and Haibin Ling. Parallel Tracking and Verifying. *IEEE Transactions on Image Processing*, 28(8):4130–4144, aug 2019.

[179] Tim Ellis. Performance metrics and methods for tracking in surveillance. In *Proceedings of the 3rd IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 02)*, pages 26–31, 2002.

[180] Kaiming He, Georgia Gkioxari, Piotr DollÃ¡r, and Ross Girshick. Mask r-cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.

[181] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240, 2006.

[182] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010.

[183] Rafael Padilla, Sergio L. Netto, and Eduardo A. B. da Silva. A survey on performance metrics for object-detection algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 237–242, 2020.

[184] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9): 1834–1848, 2015.

[185] Matej Kristan, Jiri Matas, Ales Leonardis, Michael Felsberg, Luka Cehovin, Gustavo Fernandez, Tomas Vojir, Gustav Hager, Georg Nebehay, and Roman Pflugfelder. The visual object tracking vot2015 challenge results. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 1–23, 2015.

[186] Alan Lukežič, Luka Čehovin Zajc, Tomáš Vojíř, Jiří Matas, and Matej Kristan. Now you see me: evaluating performance in long-term visual tracking. *arXiv preprint arXiv:1804.07056*, 2018.

[187] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2411–2418, 2013.

[188] Yang Liu, Peng Sun, Nickolas Wergeles, and Yi Shang. A survey and performance evaluation of deep learning methods for small object detection. *Expert Systems with Applications*, 172:114602, 2021.

[189] Norman L Stauffer. Active auto focus system improvement, jan 1983. US Patent 4,367,027.

[190] Zhifei Zhou, Chun Li, Tianyin He, Changyong Lan, Peihua Sun, You Zheng, Yi Yin, and Yong Liu. Facile large-area autofocusing Raman mapping system for 2D material characterization. *Opt. Express*, 26(7):9071–9080, Apr 2018.

[191] Takao Sasakura. Automatic focusing device using phase difference detection, nov 1999. US Patent 5,995,144.

[192] Yi Yao, Besma Abidi, Narjes Doggaz, and Mongi Abidi. Evaluation of sharpness measures and search algorithms for the auto-focusing of high-magnification images. In *Visual Information Processing XV*, volume 6246, pages 132–143. SPIE, 2006.

[193] F. Bortoletto, C. Bonoli, D. Fantinel, D. Gardiol, C. Pernechele. An active telescope secondary mirror control system. *Review of scientific instruments*, 70(6):2856–2860, 1999.

[194] Said Pertuz, Domenec Puig, and Miguel Angel Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432, 2013. ISSN 0031-3203.

[195] Jay Martin Tenenbaum. *Accommodation in computer vision*. Stanford University, 1971.

[196] JJ Vaquero, JM Pena, N Malpica, A Santos, C Oritz De Soloranzo, and F Del Pozo. Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of microscopy*, 188(3):264–272, 1997.

[197] Chunping Yang, Minhao Chen, Fangfang Zhou, Wei Li, and Zhenming Peng. Accurate and rapid auto-focus methods based on image quality assessment for telescope observation. *Applied Sciences*, 10(2):658, 2020.

[198] Islam Helmy, Farag Elnagahy, and Alaa Hamdy. Focus measures assessment for astronomical images. In *International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*, pages 244–250. IEEE, 2020.

[199] J.L. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia. Diatom autofocusing in brightfield microscopy: a comparative study. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 3, pages 314–317 vol.3, 2000.

[200] Ge Yang and Bradley J Nelson. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)*, volume 3, pages 2143–2148. IEEE, 2003.

[201] Jie He, Rongzhen Zhou, and Zhiliang Hong. Modified fast climbing search auto-focus algorithm with adaptive step size searching technique for digital camera. *IEEE Transactions on Consumer Electronics*, 49(2):257–262, 2003.

[202] Singiresu S Rao. *Engineering optimization: theory and practice*. John Wiley & Sons, 2019.

[203] Y. Xiong and S.A. Shafer. Depth from focusing and defocusing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 68–73. IEEE, 1993.

[204] Peter DiMeo, Lu Sun, and Xian Du. Fast and accurate autofocus control using gaussian standard deviation and gradient-based binning. *Opt. Express*, 29(13):19862–19878, Jun 2021.

[205] Aselsan. IHTAR anti-drone system - datasheet. ASELSAN A.S., Ankara, Türkiye, 2018.

[206] Abdel Gafoor Haddad, Muhammad Ahmed Humais, Naoufel Werghi, and Abdulhadi Shoufan. Long-Range Visual UAV Detection and Tracking System with Threat Level Assessment. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pages 638–643, 2020.

[207] Mrunalini Nalamati, Ankit Kapoor, Muhammed Saqib, Nabin Sharma, and Michael Blumenstein. Drone detection in long-range surveillance videos. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6, 2019.

[208] Petar AndraÅ¡i, Tomislav RadiÅ¡iÄ‡, Mario MuÅ¡tra, and Jurica IvoÅ¡eviÄ‡. Night-time Detection of UAVs using Thermal Infrared Camera. *Transportation Research Procedia*, 28:183–190, 2017. ISSN 2352-1465. INAIR 2017.

[209] Vasileios Magoulianitis, Dimitrios Ataloglou, Anastasios Dimou, Dimitrios Zarpalas, and Petros Daras. Does Deep Super-Resolution Enhance UAV Detection? In *16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6, 2019.

[210] Licheng Jiao, Ruohan Zhang, Fang Liu, Shuyuan Yang, Biao Hou, Lingling Li, and Xu Tang. New generation deep learning for video object detection: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8):3195–3215, 2022.

[211] Ángela Casado-García and Jónathan Heras. Ensemble methods for object detection. In *ECAI 2020*, volume 325, pages 2688–2695. IOS Press, 2020.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde. Die aus anderen Quellen oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.
Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder in ähnlicher Form in anderen Prüfungsverfahren vorgelegt.

Wien, im Juni 2024

Denis Ojdanić

Author Information and Publication List

Denis Ojdanić is PhD student at the Automation and Control Institute (ACIN) of TU Wien. He received his M.Sc. degree in Embedded Systems from TU Wien, Vienna, Austria, in 2019. His primary research interests are in the field of computer vision including object detection, tracking and identification using high-performance telescope systems.

Journals

D. Ojdanić, A. Sinn, C. Naverschnigg and G. Schitter, "Feasibility Analysis of Optical UAV Detection Over Long Distances Using Robotic Telescopes," in IEEE Transactions on Aerospace and Electronic Systems, vol. 59, no. 5, pp. 5148-5157, Oct. 2023

D. Ojdanić, C. Naverschnigg, A. Sinn, D. Zelinskyi, and G. Schitter, "Parallel Architecture for Low Latency UAV Detection and Tracking using Robotic Telescopes," in IEEE Transactions on Aerospace and Electronic Systems, vol. 60, no. 4, pp. 5515-5524, Aug. 2024

D. Ojdanić, D. Zelinskyi, C. Naverschnigg, A. Sinn, and G. Schitter, "High-speed telescope autofocus for UAV detection and tracking," Opt. Express 32, 7147-7157 (2024)

D. Ojdanić, B. Gräf, A. Sinn, H. W. Yoo, and G. Schitter, "Camera-guided real-time laser ranging for multi-UAV distance measurement," Appl. Opt. 61, 9233-9240 (2022)

Conferences

D. Ojdanić, C. Naverschnigg, A. Sinn, and G. Schitter, "Deep learning-based long-distance optical UAV detection: color versus grayscale," In Pattern Recognition and Tracking XXXIV, vol. 12527, pp. 80-84. SPIE, 2023

D. Ojdanić, N. Paternoster, C. Naverschnigg, A. Sinn, and G. Schitter, "Evaluation of the required optical resolution for deep learning-based long-range UAV detection," In Pattern Recognition and Tracking XXXV, vol. 13040, pp. 78-84. SPIE, 2024

D. Ojdanić, C. Naverschnigg, A. Sinn, and G. Schitter, "Algorithm evaluation for parallel detection and tracking of UAVs," In Optics, Photonics, and Digital Technologies for Imaging Applications VIII, vol. 12998, pp. 327-333. SPIE, 2024.