

iSPY

iSPY: Image Surveillance and Protection Yielder

Group 5 Authors:

Alejandro Lopez Zelaya
Armani Garcia
Gabryella Baldaci
Luke Reyes
Robert Josephson

Photonic Science and Engineering
Electrical Engineering
Photonic Science and Engineering
Computer Engineering
Computer Engineering

Reviewer Committee:

Dr. Guifang Li
CREOL
Dr. Suboh Suboh
Dr. Nazanin Rahnavard
Dr. Stephen Eikenberry

ECE
ECE
CREOL

Mentors/Instructors:

Dr. Aravinda Kar
Dr. Lei Wei

Chapter 2 – Project Description

2.1 Project Background and Motivation

2.1.1 Background

The history of espionage has always been intertwined to optics and photonics. From the research on lightweight cameras for the U2 surveillance aircraft in the 1940s to the recent creation of unmanned aerial vehicles (UAV) [1], [2]. The need for such technologies guides the direction of present research.

Joe Houston, a SPIE Fellow and past SPIE president highlighted the importance of the innovation in already known systems, “The most covert systems that I know of are those systems – either image communications or electronic communications – that are hidden in the skirts of strong signals in a location that you know, but no one else knows” [1].

Our system, iSPY is then a new chapter of this history. The idea is not to hide the cameras but make the transmission of their captured video something untraceable and uncrackable.

Just like many other technologies that were once created having the military in mind, we hope that iSPY can make its way to the everyday market and bring safety to civilian users of surveillance systems since security is a priority in our daily life. Whether you are a homeowner, a soldier, or a first responder, having eyes on the place you want to protect is a necessity. To prevent crimes from happening, save lives during emergency situations, or to reduce the reaction time if an incident does take place, surveillance systems are widely used.

These surveillance systems usually consist of several cameras that capture live footage of their surroundings. The footage is then transmitted to a device(s), via cable or radio, to display a real-time video of the observed location or objects.

2.1.2 Motivation

Due to the rise of crime, the need for surveillance systems has increased. And unfortunately, these technologies, when poorly used, can lead to more safety issues. Several news outlets have reported the hacking of camera systems [3], [4]. The issue is not only in home security systems, but it is a subject of international security as reported by Scott Ikeda from CPO Magazine. He wrote on Russian military attaining crucial information related to landscape of cities by hacking into cameras Ukrainian citizens used in their homes and businesses [5].

Most security systems are connected through Wi-Fi, which unfortunately, is the most common path for hackers. One of the shared tips on how to secure these devices is to disconnect them from the internet [6]. However, another problem arises with this measure, data transfer cannot take place without Wi-Fi or other means of radio communication.

In iSPY, the need for Wi-Fi or radio signal is to be replaced by a free-space optical (FSO) communication system, eliminating the security risk associated with using Wi-Fi or other radio connection. Since data will be transmitted through light, it will have some features that will make it more secure than the alternative version. One of them is proximity. Light diffracts much less than radio signals, making it hard for an outside entity to tap into the signal being transmitted. Without access to the area where the light is shone it is impossible to access the information. Another attribute of the system is that the receiver's detector is designed and tuned for that specific transmitted signal, the detector is unique to that system, decreasing and almost eliminating the chances of someone matching the specifications needed to decode the data.

2.2 Commercial Technologies and Existing Projects

Market research was done to better understand the specifications and parameters we should keep in mind when building our project. We searched for different products and/or projects, we took notice of how they work, and we tried to understand their strengths and flaws. In this section, we will discuss those products and/or projects and relate them to our own iSPY.

2.2.1 Koruza

Koruza is a wireless optical internet-access system that advertises the secure transmission of data and easy installation process. A product on the market that matches some of the goals we want to achieve with iSPY. They are also the only open-source free space optical system available.

The system was designed to facilitate data transfer in environments where other options (wireless and fiber optics) are not feasible solutions. Koruza is also easily upgraded, as they say, it is “future proof” [7], [8].

2.2.2 MOSTCOM

MOSTCOM is a well-established Russian company that specializes in free-space optics research and development on earth and space. They have a product line that includes stationary optical systems (OGS), ground-to-space terminals, and mobile terminals. Their current research is on underwater communication terminals [9].

2.2.3 Drone Mounted Object Detection & Recognition using ML and IR Sensors

This was a previous senior design project from a group based out of the University of Central Florida in Spring 2023. This effort combined the ideas of artificial intelligence (AI) and unmanned aerial vehicles to create an autonomous object detection drone. The primary goal for this project was to create a drone-mountable object detection system enabling artificial intelligent visual navigation. The secondary goal was to develop and integrate autonomous peripherals for the drone. The need for this sort of drone technology is

widespread for the defense, emergency response, and agriculture industries. The idea, combined goals, and industries of interest of this project are very relevant for iSPY and must be considered closely.

Without knowledge of this project, iSPY intends to build on and design around the successes of this previous collaboration. For example, one of iSPY's stretch goals is the integration of an object detection and free-space optical communication systems on to a drone. Since iSPY is a surveillance system, object detection through computer vision increases the amount of information made available for the end user's discretion. In defense situations, drone-mounted, AI-assisted object detection systems can be employed to detect guns, explosives, or enemy targets in a region of interest. In emergency situations like wildfire, a version of the drone-mounted, AI assisted object detection system could be used to detect trapped people through the foliage. In agriculture, another version of the AI-assisted system could be used to detect pests amongst crops like groundhogs or certain weeds. It's important to note that iSPY doesn't plan to use a drone to demonstrate AI-assisted object detection; however, it's a stretch goal the team is considering if time permits.

The main difference between iSPY and this previous collaboration is the means of communication with the end user. One of the main motivations for iSPY is to implement free-space optical communication as a means of limiting outsider interference to the signal. If surveillance is to be achieved in secret, then a wire or radio signal cannot be used. Instead, an optically confined light pulse is employed. To interfere, or tap, the signal, the optical must be physically intercepted, making the intrusion known.

2.2.4 S.T.E.A.L.T.H. Secure Transmission via Electronic and Laser Technology Hardware

The project was a previous senior design project from a group at the University of Central in Spring 2022. The project focused on a bidirectional Optical link communication system which utilizes the transmission of data and receiving of said data. In this project the goal was to make the optical wireless communication system more modular and mobile as well as focusing on low size, weight plus power budget within reason.

S.T.E.A.L.T.H. used two different PCB's. One of the PCB's was for serializing and the other was for de-serializing. The design of iSPY intends to utilize a similar transmission of data except without serialization and de-serialization PCB's. Instead iSPY will be using FSO transmission which sends data via digital signal through free space and not an optical link or through PCB's that encode/decode the data.

2.3 Goals and Objectives

The main idea of this project is to build a secure communication, computer-vision-capable object detection surveillance system. A working iSPY device will relay a real-time surveillance video feed to a display by means of FSO. A product of this caliber has the capacity to support defense, emergency response, and agriculture industries, by yielding

the end user valuable information for their discretion and mission. The basic, advanced, and stretch goals and objectives for this project are outlined below.

2.3.1 Goals

To get the project running, the team devised three tier-lists of goals – basic, advanced, and stretch goals. The goals are organized in increasing difficulty, respectively. This give the team a map on which to make the project a success.

Basic Goals

- Capture faint images of the environment surrounding the sensor.
- Transmit and receive a laser pulse.
- Designed PCB must effectively house a processing unit, electrical components, power systems, and the necessary connection pins.
- Able to detect objects with the use of computer vision algorithm in video feed.
- Achieve stable power distribution across all components.
- Establish a stable method that can efficiently compress data for transmission and decompress data for viewing on display.

Advanced Goals

- Capture a clear image of the surrounding environment with 80 degrees FOV.
- Transmit and receive usable optical signals using integrated circuits at 1Gbps.
- The power systems must power the imaging acquisition system and an external processing board.
- Capture body recognition as well as count the number of individuals recognized from video feed.
- Minimize signal interference and ensure data integrity on high-speed transmission lines.
- Achieve a high-quality video feed of 720p after compression and decompression.

Stretch Goals

- The iSPY system is modular – will be spread across a drone and a tower.
 - Imaging system and transmitter optics found on the drone.
 - Reception optics found on the tower.
- Aberration-corrected, high-resolution, 120° FOV images.
- Ultra-fast transmission speeds using high-speed FSO transceivers (+\$2000 ea.).
- Tracking body movement of individuals in video feed and enhancing recognition of people to determine if male or female, young or old and adult or child.
- Optimize thermal management for extended operation.
- Implement compression that is capable of dynamically adapting to video complexity which would allow much higher quality of video feed.

Regardless of the tier level achieved, the project will carry out the same functions and look similarly to the below figure.

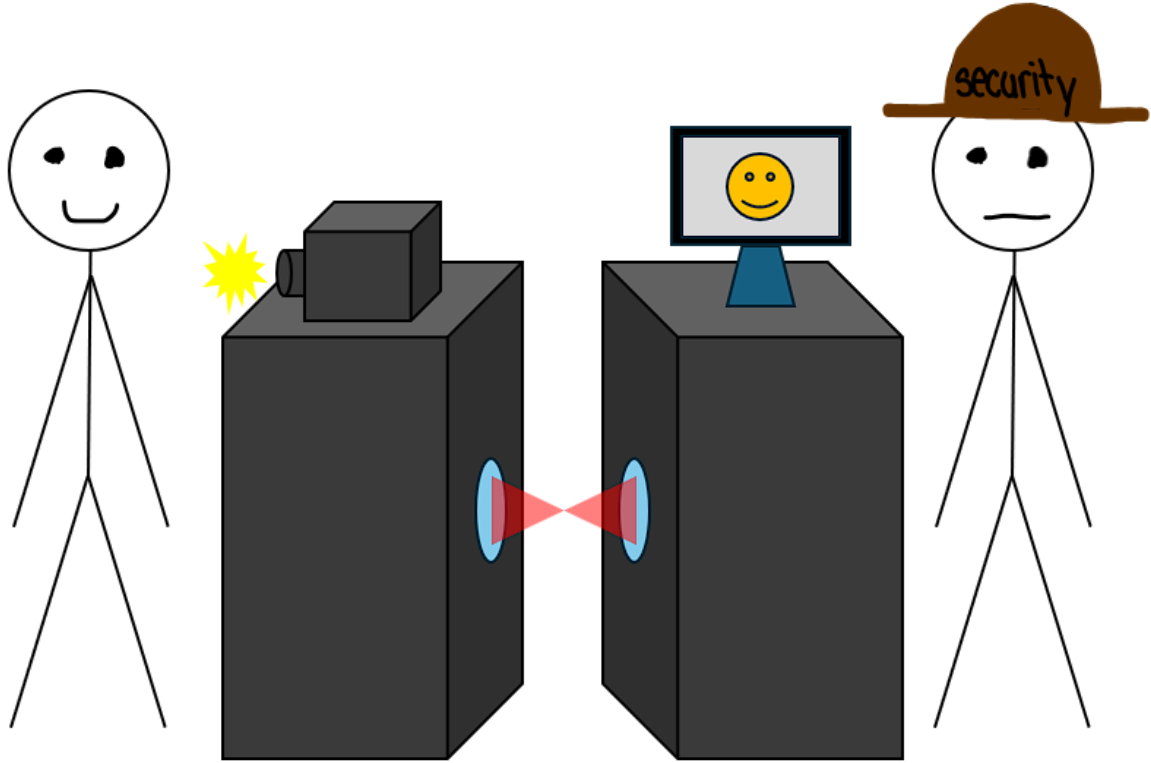


Figure 2.1. General project diagram by Authors.

2.3.2 Objectives

To effectively achieve the listed goals for the project, objectives must be established and understood for each. The objectives, divided according to goal type, are discussed below.

Objectives for Basic Goals

- OOK laser diode and measure electrical signal on photodetector.
- Create a two-lens system to capture an image.
- Design a PCB layout that will work for the envisioned system.
- Utilize a computer vision algorithm to detect objects.
- Design a robust power management circuit that can regulate voltage and current for the envisioned system.
- Implement a compression algorithm in our system that can reduce the size of data to reliably transfer data via the FSO system.

Objectives for Advanced Goals

- Integrate optoelectronic transceivers to transmit and receive usable optical signal.
- Create a wide field of view ($\geq 80^\circ$), continuous image feed.
- Test PCB with all the components and confirm that power system is giving desired output ($\leq 10V$).
- Program code for body recognition with an accuracy level of ($\geq 75\%$) and implement a counter for recognized bodies.

- Implement shielding and controlled impedance techniques in the layout of the PCB.
- Implement advanced codecs for real-time compression to achieve low latency and quality display.

Objectives for Stretch Goals

- Design a lens system that captures a high-resolution video and correct aberrations.
- Create a transmission/receiver system with a transceiver, evaluator board, and MCUs.
- Integrate iSPY into an aircraft to survey larger areas and increase applicability.
- Enhance the code to track the bodies and break recognition of person into sub-categories as well as increase accuracy of recognition to ($\geq 90\%$).
- Integrate thermal management solutions (heat sinks, thermal vias, etc).
- Develop an AI-driven algorithm to compress and decompress video with higher quality and faster processing.

From a very high level and broad perspective, after achieving the listed objectives, the project will have a layout like below.

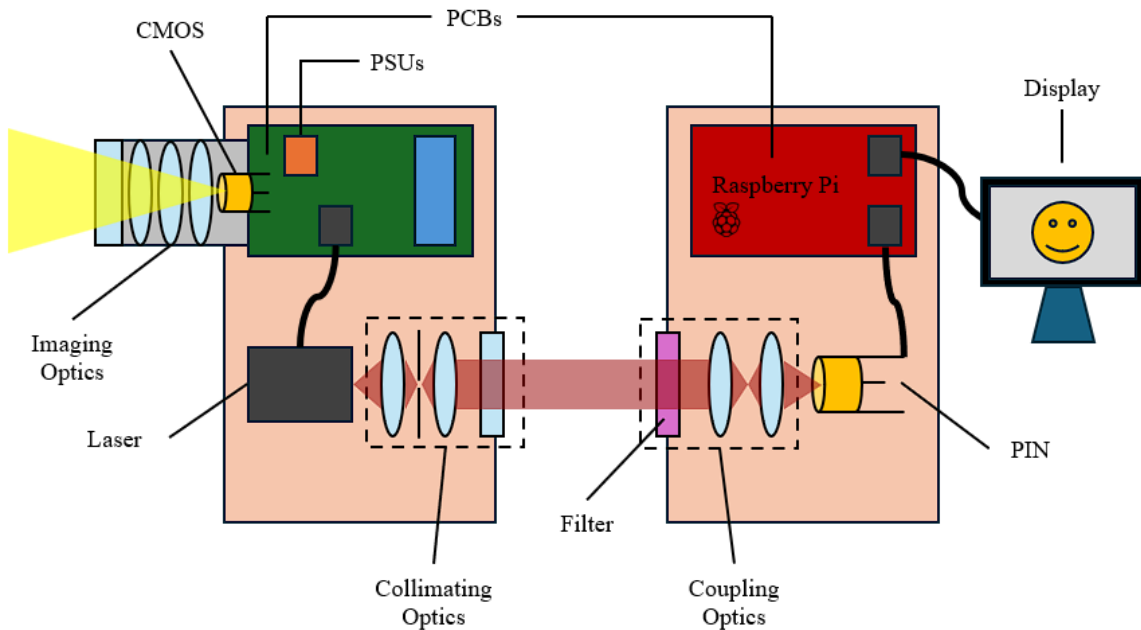


Figure 2.2. General program schematic by Authors.

2.4 Required Specifications

Considering the proposed goals and objectives, the required product specifications are:

Table 2.1. Engineering requirements specifications for the performance and physical features of the envisioned system.

Product requirement	Requirement specification	Unit
The system will capture a wide field of view (horizontal)	≥ 80	Degrees
The system will capture a wide field of view (vertical)	≥ 60	Degrees
The system will produce a high image resolution	720	Progressive scan
The system will be designed for low power consumption	≤ 15	Watts
The system will have a high data transmission rate	$\geq 1\text{G}$	Bits per second
The system will be capable of simulating a long-haul comms link.	1	Kilometer
The system will have some latency for processing time	1000	milliseconds
The system will have a large optical link range	≥ 1	Meter
The system will be reliable when detecting people (accuracy)	≥ 75	Percent
The system will be able to detect multiple individuals	3	People

Table 2.2 Engineering requirements specifications for the components that will be used in this project to build the envisioned system.

Component(s)	Parameter	Specifications
Image acquisition system (Camera)	Field of view	80 degrees (horizontally) 60 degrees (vertically)
Computer vision algorithm (MCU)	Accuracy of detection	75%
Photodetector	Incident optical power	$< -10\text{ dBm}$
MCU	Clock Speed	16 MHz
Camera sensor	Framerate	30 frames per second
Power Supply (External)	Voltage Output	5V DC regulated
Display	Framerate	30 frames per second

Table 2.1 and 2.2 show the engineering requirements specifications for this project. Our goal is to showcase these specifications, especially the ones in the highlight rows.

Table 2.3 – Optical components used in the envisioned system.

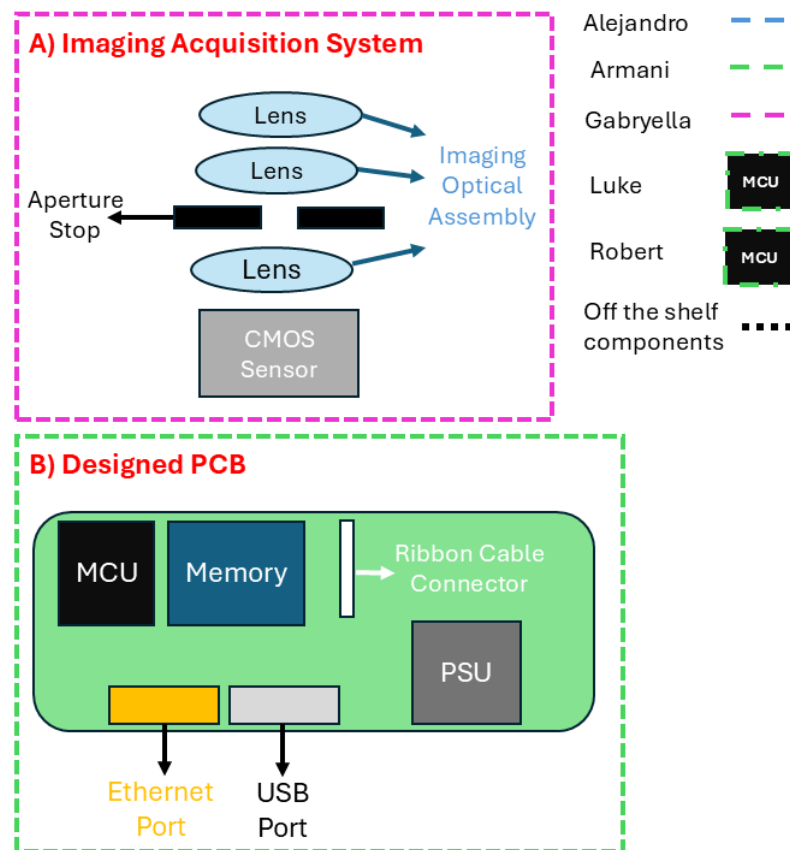
Component	Quantity
Lenses	7
Transceiver	2
Optical Fibers	2
ND Filters	2

2.5 System Diagrams

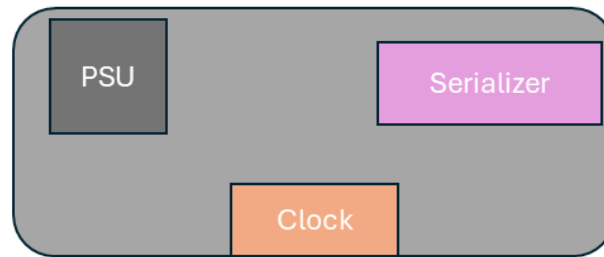
The engineering scope of this project is divided amongst hardware and software systems, each working in tandem to make iSPY work. This section discusses and presents the hardware and software diagrams necessary to make this project a success.

2.5.1 Hardware

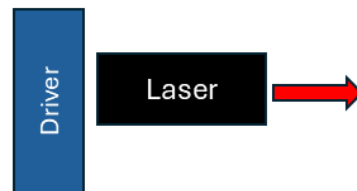
The hardware is divided into three separate sub-systems: imaging acquisition, FSO communications, and electrical hardware. The hardware layout and separation of work is demonstrated below.



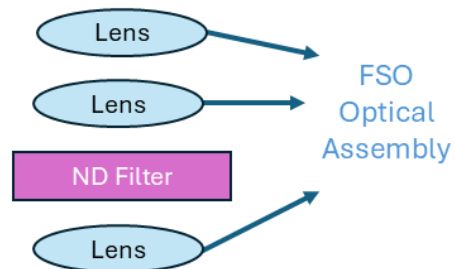
C) Evaluation Board (1 out of 2)



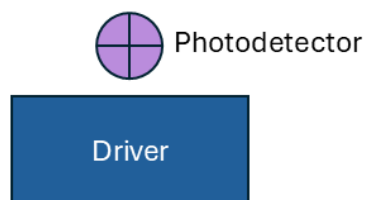
D) Transmitter



E) FSO System



F) Receiver



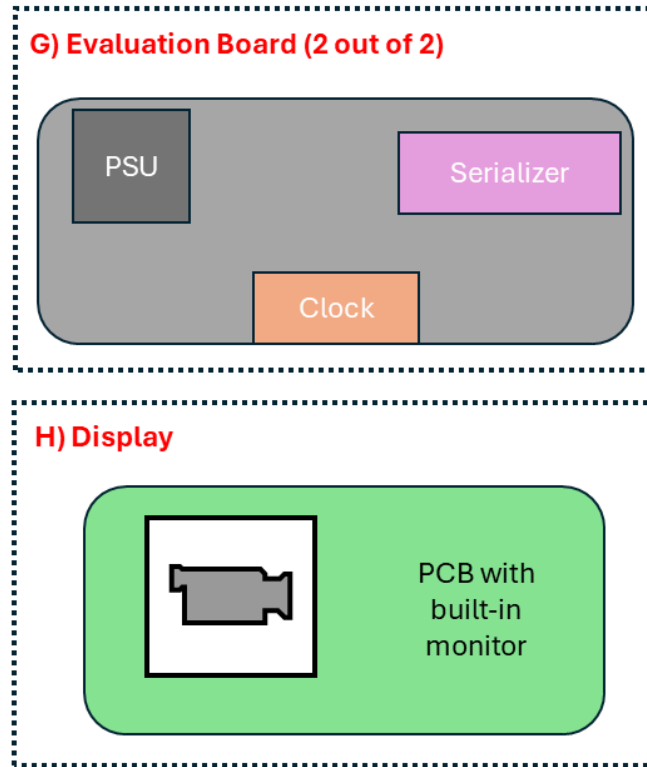


Figure 2.3. Hardware diagram by Authors

2.5.1.1 Image Acquisition Sub-System

One of the most crucial parameters for surveillance systems is the camera's field of view. Field of view "is the open, observable area a person can see through their eyes or via an optical device, such as a camera" [10].

The goal with iSPY is to have a wide field of view without sacrificing too much of the resolution. A wide field of view is needed to capture a large portion of the scene and decrease blind spots. At the same time, high resolution is necessary since the computer vision system will require as much detail as possible to determine what is taking place. Additionally, high resolutions enable the end user to best discern between two or more objects that are close to each other.

The system will be designed to be placed in spaces where observing a larger portion of the surroundings is more important than seeing "crisp" details. This system will consist of an assembly of lenses designed exclusively for this project. The lenses will focus the light, create an image on the sensor, and correct optical aberrations affecting the imaged scenario. These optical components will be mounted before a CMOS sensor – forming iSPY's own camera and our image acquisition system.

At the moment, we are leaning towards creating a fisheye lens system since it gives us the most field of view when compared to other systems. When dealing with fisheye lenses, distortions will take place, and through researching and testing we will attempt to minimize

the effect of distortions and aberrations, optimize the system, and change accordingly. Some important points that will play a role when developing this system are the desired field of view, which dictates the focal length and the sensor size; the projection type we want to obtain, since we will be mounting it to a sensor; the lenses used for aberration correction; overall lens alignment; and sensor compatibility with the rest of the system.

Choosing the different parts of this system, understanding their interaction, and simulating outcomes on the ray optics simulation software Zemax will be the key to obtaining a successful prototype.

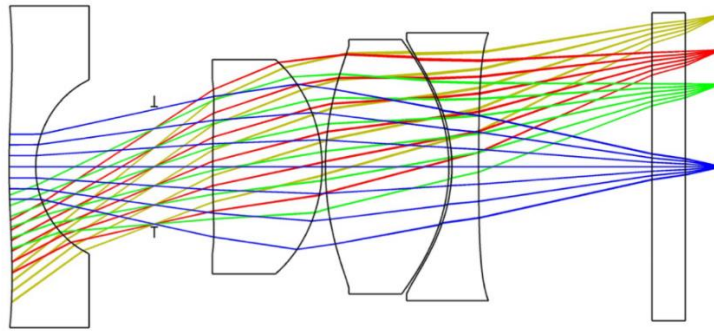


Figure 2.4. Layout of a wide-lens system from Reference 11. [Permission Requested]

Figure 2.3 shows the design of a wide-lens system [11]. Wide-lens and fisheye lens systems usually follow a similar structure, and this is what we currently have in mind. Our system will follow industry's frameworks and expand to match the requirements of our design.

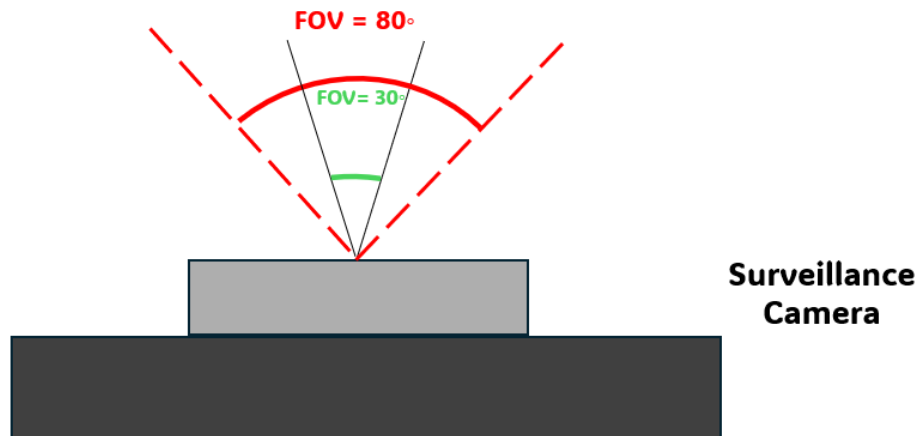


Figure 2.5. Field of view captured by Authors.

Figure 2.4 shows a ray optics model of the field of view captured by the detector. Some cameras can cover the larger region (180 degrees). However, the optics introduce many optical aberrations making high resolution, minimally distorted images difficult to

produce for this project. Therefore, we will focus on a smaller portion of the observable area and focus on the quality of the image.

Having the scope of this class in mind and facing some uncertainty due to the cost and size of equipment, our visual monitoring system will consist of only one image detector and the focus will be on capturing live videos of a well-lit environment, such as, inside a building or outdoors during daytime. One of our future goals (stretch goals) is to implement the system outside, mount iSPY onto drones, and have it running regardless of the time of the day.

2.5.1.2 Free-Space Optical Communications Sub-System

For iSPY to achieve secure communications link, FSO communications is employed. An FSO communications system consists of a transceiver (Tx/Rx) for optical signal transmission (Tx) and reception (Rx), optical components, and a signal evaluation board. Though these components are distinct, they must work as a unit to achieve a common goal: to relay the captured live-video feed to the end user. Together, the FSO communications sub-system looks as follows:

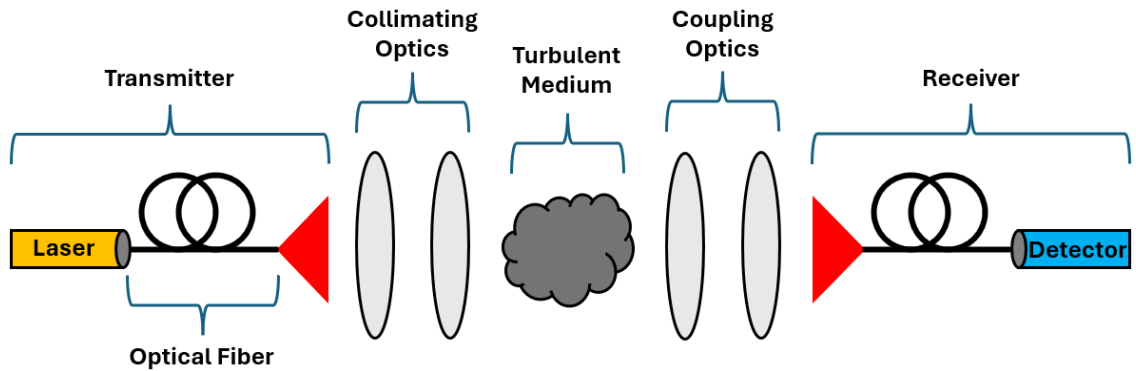


Figure 2.6. FSO transmission/receiver system diagram by Authors.

Optical communications systems, like fiber optic communication or FSO communication systems, typically reach transmission speeds in the order of a couple MHz to tens of GHz. In telecommunications, especially fiber optic communications, bidirectional communication is common and done frequently using an optical transceiver. An optical transceiver is an optoelectronic integrated circuit consisting of a laser diode and photodetector. In practice, a transceiver takes in two optical fibers: one for Tx and one for Rx. The Tx optical fiber is connected directly to the laser diode output, whereas the Rx optical fiber is connected directly to the photodetector input. Transceiver technology has improved tremendously over the last 40 years, making it an enabling technology for the success of this project.

In iSPY, a transceiver will be used as the optoelectronic relay to accomplish Tx and Rx. On the Tx side, the transceiver will be responsible for generating an intensity modulated laser pulses that correspond to the live video digital signal bitstream. The laser diode is directly responsible for this process. The generated laser pulses will first be transmitted via

optical fiber. The optical fiber will transport the laser pulses to the free-space channel. To have adequate control of the laser pulses in the free-space channel, the pulses must be Gaussian in shape. This can be achieved using an optical fiber or a spatial filter within the FSO transmission optics. On the Rx side, the transceiver will be responsible for generating an electrical signal that corresponds to the detected incident laser pulses. The photodetector is directly responsible for this process. The incident free-space laser pulses will first be coupled into an optical fiber to guide the pulses to the transceiver photodetector, terminating the Tx/Rx link.

Unlike other optical communications systems, like optical fiber communications where the means of data transmission is confined in a fiber optic cable, FSO communications systems transmit data in free-space, open-air channels. This type of system will not be confined by any conventional means like a copper or fiber optic cable. Thus, FSO systems must adequately control the beam size and divergence of its free-space laser beam.

The free-space input and output laser pulses will be guided from and to the transceiver by an optical fiber. Optical fibers are low-loss, light waveguides with unique optical properties. Using an optical fiber as an input to the free-space channel comes with some considerations. Single-mode or multi-mode optical fibers may be used for this project; single-mode fibers only guide one spatial light mode whereas multi-mode fibers can guide hundreds. Multi-mode fibers have high light-capturing capabilities because of their high numerical aperture compared to single-mode fibers. However, the bitrate must be closely considered for multi-mode fibers because of the limited modal bandwidth associated to all multi-mode fibers. Single-mode fibers enable very high bitrates over long distances but are hard to use for free-space applications due to their low light-capturing capabilities. Besides, all fibers suffer from material attenuation and dispersion. Material attenuation is how much light is absorbed by the optical fiber, decreasing the strength of the guided signals. Dispersion is the spreading of the guided signals by means of color, mode, or polarization. Managing both attenuation and dispersion must be done under close consideration of each other. Thus, choosing an optical fiber that enables the bitrates of the system, has good light-capturing capabilities, is cheap, and can deliver signals over short distances is critical.

When light exits from an optical fiber, it diverges according to the guided optical mode present at the output – Gaussian for single-mode fibers and some higher-order mode for multi-mode fibers. Controlling the divergence and shape of the free-space pulse is important as it will determine how it interacts with the turbulent channel and how well it can launch into the receiver optical fiber. To control the divergence to the desired amount, two optical lenses will be used at the input. Using different focal length lenses will change the shape of the spatial dimensions of the signal pulse and, depending on the alignment of these optical components, the divergence can be increased or mitigated, giving us total control over the beam. Additionally, if the input lenses are aligned accordingly, a 4F lens system can be configured to make use of Fourier optics, enabling spatial filtering of the input optical signal pulse.

When light is to enter the receiver optical fiber, an optical design that mimics the input will be used – two lenses with different focal lengths. The process of launching free-space light

into an optical fiber is known as coupling. When coupling light into an optical fiber, the fiber core size, the fiber numerical aperture, and the lenses' numerical apertures must be considered. When focusing the free-space light signal, the spot size must closely match the fiber core size. The numerical aperture will determine how much of the focused light enters the fiber; the higher the numerical aperture, the lighter the fiber will accept. For the fiber to accept most of the focused light, the numerical of the lens system must be less than numerical aperture of the fiber.

An external evaluator board is the signal processing center for the iSPY system. Housed on a PCB, the evaluator board housing a processing unit and transceiver, seamlessly integrating critical signal processing hardware. Per advisor suggestions, the evaluator board is an aftermarket component. Which evaluator board is used depends on the transceiver of choice and cost. The trade-offs associated with this choice will be explored in the next chapter.

The evaluator board will receive data from PCB #1 and convert it to a functional optical signal for the transceiver; this optical signal modulates the laser diode on the transceiver, generating the necessary laser pulses for the FSO component of the project. For incoming laser pulses on the photodetector of the transceiver, the electrical signals generated by the photodetector will be interpreted by the evaluator board and transmitted to the next electrical component of the project – the display. This means that the signal will be looped back into the evaluator board using two fiber optic cables acting as a sort of antennae, sending the data over the air between themselves back into the evaluator board to be processed. Essentially, the evaluator board acts as a bridge that connects the display to the Image Acquisition System by facilitating the necessary processing for FSO communications.

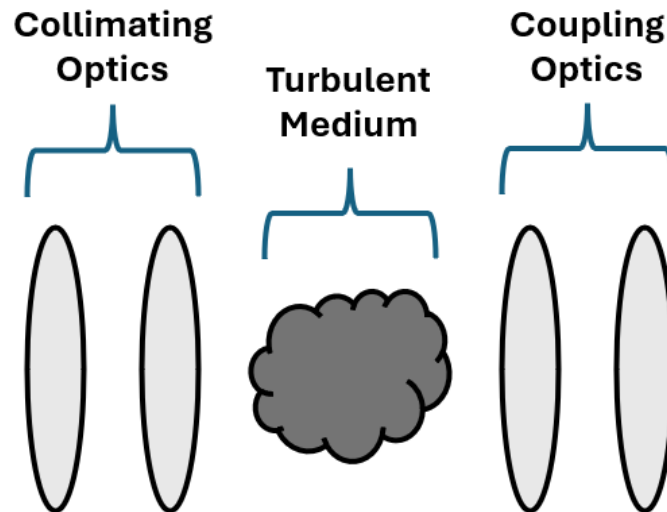


Figure 2.7. Optical schematic for FSO by Authors.

Due to this PCB being connected to PCB #1, we can also use the same power supply to distribute enough power to both PCBs. We would have to take extra care in the construction of the circuits connecting the components as we want to make sure that everything can receive enough power to operate but not consume excess power. The distribution of power will have to be carefully managed across the board to make sure the components are safe,

and the important parts of the system won't get burnt out. Signal integrity is a major issue that would make or break the process since we want clear data that we can work with and get an output from.

2.5.1.3 Electronics Housing PCB

The goal of PCB #1 is to be the main brain for the system. This is where the bulk of our software will be as we work with computer vision and video data. This PCB will be designed and constructed with the intention to take in live video without audio, process that data using computer vision, counting the number of people in the camera's field of view, and prepare the transmission of data to the Evaluator Board, that is PCB #2. It should perform these tasks efficiently, balancing processing power, data integrity, and power consumption to ensure real-time operation. We would also like to be able to store some of the video feed into a memory cache to be able to have some playback if necessary.

The MCU will receive an uninterrupted continuous video feed in a digital format from the Image Acquisition system. With this uninterrupted digital signal, the MCU can break it down into individual frames where it can be analyzed further, including preprocessing if necessary. This could include adjusting the brightness or contrast of the image, if need be, so that the image can be cleaner for better computer vision performance.

Before processing any data and cleaning up the video, we will train a computer vision model to detect and count how many people are in the field of view of the external camera. This program will analyze the video feed and keep a record of the number of present individuals, ensuring that it is identifying each unique person individually. Then the video will be cleaned up more and the resulting data will be compressed into a binary bitstream as well as a suitable format that will enable efficient transmission. The Transmitter Board will process the video feed using a MCU into a binary sequence that can be broken up into 4 or 8 bits at a time creating this bitstream. A bitstream is a sequence of bits that are transferred from one device to another. Remember that all types of data are made up of bits, the most basic unit of information. As such, a bitstream applies to all data, be it an audio, image, or any other type of file [12].

Once the data has been processed and compressed properly, PCB #1 will send this data over to PCB #2 through a copper cable connection. The two PCBs will be physically connected to ensure proper transmission of data, while the FSO transmission will occur later into PCB #2. However, before the data is sent over, we can store the data into a memory cache to allow easy access to playback any part of our video files.

This PCB will be connected directly to a power supply as well as power management circuitry that can produce enough power to the system while also ensuring that the computer vision and data transmission can take place while not using up too much power. Also, since the system will be used in a way that will need to be operational for long periods of time, we will utilize low-power operations with the MCU so that we can reduce the amount of power that is utilized to complete its functionality.

2.5.1.4 Display

An external monitor will be used to display the captured and relayed real-time video feed. For the monitor to generate an image, it will be connected to an external power source and a video-out port. The acceptable signal format for the monitor of choice will be determined from the beginning processing unit discussed in the next section. The signal format will vary between monitor models and video resolution.

2.5.2 Software

The software is divided into two separate sub-systems: object detection processing and digital signal processing. The software flowchart and separation of work is demonstrated below.

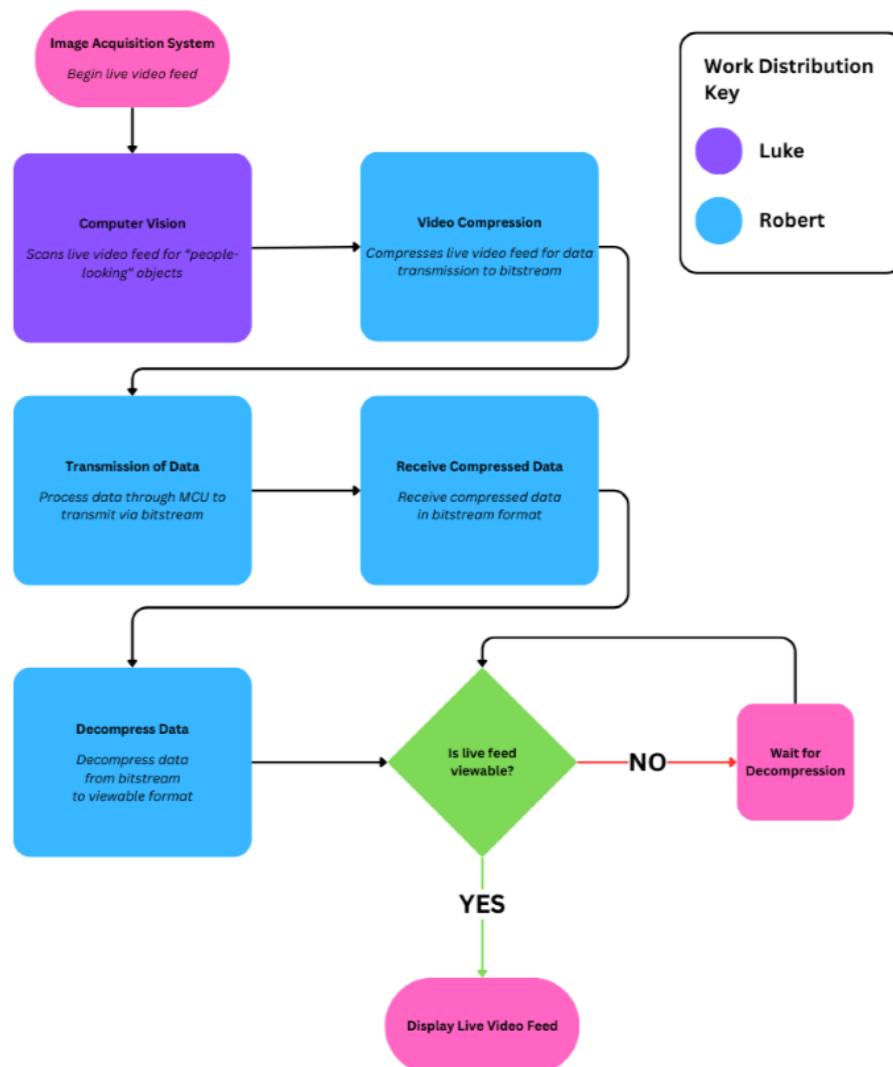


Figure 2.8. Software diagram by Authors

2.5.2.1 Object Detection Processing

The goal of the computer vision side of this project is to detect the number of people on the live video transmission at any time. This can be achieved with real-time processing of the live video feed to acquire an accuracy of 75% or more. A suitable approach to attain these kinds of results could be utilizing deep learning. Neural networking with many layers can give our computer vision the ability to understand more complex detection. Examples of this complex detection could include poor lighting, night vision, different angles of the body, and even clothing covering crucial parts of the body for detection. The program would detect the total number of people in the image and display this number on the output screen so that the viewer of the surveillance system knows exactly how many people are currently on the feed.

Neural networking is a process of machine learning that utilizes a “neural net” of many nodes. These nodes receive data and assign a “weight” to the data received. Each node will pass down this data with the assigned weight to the next node and so on. There is a desired weight set for each node. If the node is under the desired weight, it will not transfer the data. If it is above the desired weight, it will pass on the data. The weight of the data determines how important the information gathered is. Once the data has gone through the entire neural network, it will arrive at the output layer. The output layer is where the result will be displayed. [13]

In this case, the program is attempting to recognize human bodies in a live video feed. The program will almost certainly not produce an accurate result on the first attempt. That is where the process of training comes in. Algorithms in OpenCV using neural networking will be used and the best attempts from previous runs will be remembered. The program will remember the previous data and use it as a comparison when going through the neural network in future runs. This process will repeat until the program can accurately detect a human body 75% of the time.

2.5.2.2 Digital Signal Processing

The image acquisition system will relay a live video feed processed by the computer vision algorithm which recognizes an object as a body. Once the process of recognition is done then another algorithm will compress the data into a bitstream which is sent to laser diode as a digital signal. The photodiode will receive this digital signal as an optical signal and convert it to a digital signal. The MCU will decode the digital signal into a working bitstream for the display.

2.6 House of Quality

Efforts to carefully plan the project and closely consider customer needs and engineering requirements must be evaluated. A house of quality (HOQ) is a conceptual map that integrates planning, customer needs, and engineering requirements onto one chart. Additionally, it enables customer needs and engineering requirements to be directly related

to each other in an organized manner. In this section, the team's HOQ is shown in the figure below.

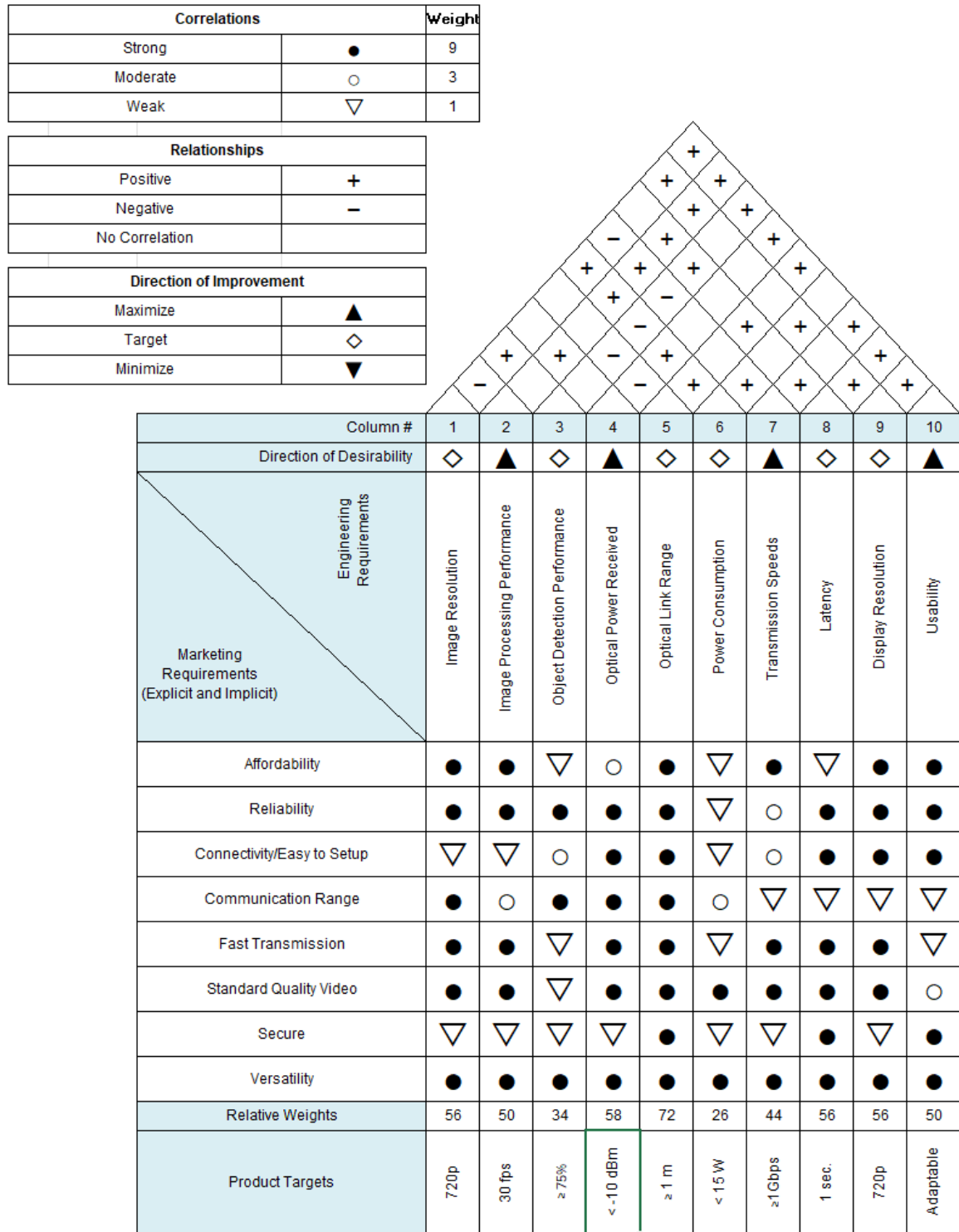


Figure 2.9. House of Quality

Chapter 3 – Research

The success of this project depends on the preparedness of each author and a knowledge-based decision-making process. This section presents the research sought by each author and decision-making process for various components embedded in project.

3.1 Imaging Sub-system

iSPY begins with the image acquisition system and in this section, we will explore the research that took place in order to design it. We separate this section into several sub-sections, starting with the sensor and ending with the lens choice, in later chapters we will talk about the integration of all of these parts.

3.1.1 Sensor Parameters

The first step on building iSPY was choosing the camera sensor. The camera sensor would dictate what kind of signal and video format we would be provided to the rest of the system.

Choosing the sensor is the first step in many optical design projects because there is a standard to be followed. They are usually not custom-made, and especially for a project like ours it would be impossible to design and fabricate one. However, there is freedom for lens choices. Therefore, we followed the steps of other optical engineers and decided to work around a sensor. [14]

When choosing a camera sensor one of the most important parameters is its size. There are so many options, and each choice comes with a tradeoff. Understanding what is important for our system is key. To make sure all parameters were considered ChatGPT was used to point out the main factors to look out for when choosing a sensor. The parameters were: sensor size, resolution, frame rate, dynamic range, shutter type, data interface compatible with the system. I also integrated other parameters into the discussion like depth of field, F-number, and image circle.

Sensor format size is the starting point when choosing a sensor. It dictates nearly all operational metrics of the imaging system [15]. An equation that shows this strong relationship between sensor size and the parameters is:

$$\text{FOV (in degrees)} = 2 \tan^{-1}\left(\frac{w/2}{f}\right) \text{ (Eq. 3.1)}$$

Where f is the focal length and w is the sensor width [16].

For our system (and surveillance systems in general), a wide field of view is crucial. Field of view (FOV) is an angle measurement of how much of a scene is visible to a device or individual. It is the parameter that lets the viewer know how much they can see. For this

system, it will be a parameter based solely in one frame from a single device (only one camera).

With equation 3.1 in mind, we can see that a larger sensor will increase the field of view. Other benefits of a larger sensor are that it decreases diffraction, gathers more light, in other words, it acquires more information increasing the quality of the electrical signal that is converted by the pixels and having better performance in lowlight environments [17] [18].

The downside of a large sensor is its size, with a larger sensor we also need larger and consequently heavier lenses. The price also increases dramatically [17] [19]. However, an important feature is that for sensors designated for video recording/streaming increasing the size of the sensor does not add too much to the system, it is not a clear linear relationship like it is for photography. A larger sensor should only be considered in specific cases like extremely shallow or deep depth of field and shooting low-light performance [17].

If the aperture (opening) is to remain the same, something we have to take into consideration because in this project that is the case, the depth of field (range of distances where the images are clear) will decrease as the size of the sensor increases. The way to counterattack this effect would be to get closer to the subject or to increase the focal length to fill the frame. However, both options are nonviable in our system. The camera will not move closer to a subject and the focal length will not change [19]. So even though larger sensors will provide greater detail, depth of field is not something a surveillance system can tradeoff.

A stop is a hard physical aperture that limits the amount of light into an optical system. An example of a stop is an iris; by opening and closing it one can determine how much light passes onto the system. The aperture stop changes light levels and it is also used to help control optical aberrations [20].

Another crucial parameter equation 3.1 displays are the relationship between focal length and field of view. We will go into more detail about focal length and effective focal length of the system in the second part of this second where we explore optical parameters, but even now we see that a short focal length will increase our field of view.

Resolution can be defined as “the dimension of the finest spatial detail in the object that can be discerned in the acquired image” [21]. In other words, we decrease the distance between two points until we cannot resolve them separately anymore. Resolution is the ability to tell points/objects apart, to resolve them.

Many elements play a role in the resolution of a system. We will separate these differences in two, the resolution of the sensor and the resolution of the lens system will be explored later in this section.

It is common knowledge that the pixel size and quantity define the resolution of a sensor. Megapixels is a measurement of how many millions of pixels a sensor has. The sensor

sizes are standard, but not all sensor sizes have the same number of pixels. A large number of megapixels increases the resolution.

High quality surveillance systems have 4K sensors (four times as many pixels as a 1080p sensor). Cameras on a budget will oscillate between 1080p and 2K. 1080p sensors have a resolution of 1920x1080 pixels. 2K sensors have a resolution of 2560x1440 pixels, and 4K sensors have a resolution of 3840x2160 pixels [22] [23].

But when exploring resolution there is much more to it. The highest frequency a sensor can resolve is called the Nyquist frequency. It can be defined as two pixels or one line pair (2 pixels=1 cycle=1 line pair, or lp). In the industry, a common way to approach the sensor resolution (ξ_{Sensor}), also referred to as the system's image space resolution ($\xi_{\text{Image Space}}$), is “by multiplying the pixel size (s), in microns, by 2 (to create a pair), and dividing that into 1000 to convert to mm.” [24]

$$\xi_{\text{Sensor}} = \xi_{\text{Image Space}} = \left(\frac{1 \text{ lp}}{2 \times s}\right) \times \left(\frac{1000 \mu\text{m}}{1 \text{ mm}}\right)$$

We can also determine the minimum resolvable spot in a scene by taking the ratio between the sensor size (horizontal, vertical, or diagonal) and field of view (FOV). This relationship is also called system magnification [24].

$$m = \frac{H}{FOV}$$

By determining the imaging space resolution and the system magnification we can make the lens choice more convenient.

Choosing between the sensor resolution is also a matter of bandwidth. Higher resolution means more data transfer, and it requires faster speed. Another detail to consider is the frame rate the sensor enables. The rate of frames per second (FPS) indicates how many images are taken in one second. More frames per second results in a smoother and higher quality video. Usually, the frame rate found in security cameras varies from 30 to 60 FPS, where the high end of this range is used for traffic monitoring [25].

The tradeoff between a higher number of frames per second is that it also increases the bandwidth and storage requirements. Since the goal with this surveillance system is not to track fast moving objects but people inside a room, a 60 FPS is not necessary. Just as important as the sensor size, the type of imaging sensor dictates the system: charge-coupled device (CCD) or active-pixel sensor (CMOS). When choosing between the two through the lenses of this project, a CMOS sensor aligns more effectively with the goals we plan to achieve. A CMOS sensor offers a higher resolution, low power consumption and the sensors are more cost effective than the CCD ones [26].

The shutter of a sensor controls how long it is exposed to light to record a frame. This period of exposure is called the shutter speed [27]. When this speed is not as fast as we would like blur will occur, therefore, even in well-lit places a shutter speed of at least 1/30s

is recommended. With the basic and advanced goals in mind, we will make sure this recommendation is met. However, for the stretch goals this speed would need to increase dramatically. In lowlight environments, shutter speed should be around 1/500s [28].

There are two modes for shutters, global and rolling. In global shutters the whole sensor array is exposed simultaneously, a good way to picture it is to image a flash. Rolling shutters on the other hand, have different array lines being exposed at different times, line-by-line, like a ‘wave sweep’.

Rolling shutters have a disadvantage, they are known for spatial distortion. Because the array is exposed line by line, objects that move extremely fast will not be captured accurately if the speed at which the lines (of the sensor) are exposed is lower than the speed of the object moving across the sensor. However, rolling shutters are commonly used in industry and the majority of CMOS cameras have them. Since we are not tracking fast-moving objects, they are enough for our purpose [29] [30].

There were two choices between the sensor for the image acquisition system, the sensor module and the camera module. They have similar functions; however, the sensor module on its own has a limited image processing capacity, making the host side (the PCB with the MCU it is connected to) in charge of both pre and post processing. Also, if the sensor were to change the MCU would have to be reset since the program is unique for each sensor.

On the other hand, the camera module already comes with pre-processing capacity and the host side can take the already clean image and focus on other aspects such as object recognition. It is also more flexible since it does not solely rely on the host side for the processing of images [31].

For this project we will use computer vision to detect how many people are at a given place, therefore, our choice is of a camera module instead of a sensor module. The camera module will enable us to focus our efforts on the usage of the images since it will obtain and process it for us. We will also be able to have higher quality video because we can have two parts of the system (the module and the host) working on polishing the frames.

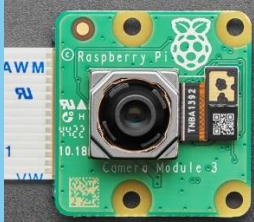
A method by which data is transferred from one device to another is called data interface. Interfaces are standardized to ensure that electronic components can smoothly work in conjunction. When talking about cameras, the data interface refers to how the frames (image or video) are obtained and transferred to a processing unit [32].

Since the interface is the connection between the camera sensor and the host it dictates the overall system performance and speed. The interface supports a two-way communication where the sensor sends images and videos (frames) to the host and the host controls the camera sensor. Choosing a suitable interface is crucial to meet bandwidth and resolution requirements [33]. The most common interfaces in the industry for machine vision are FireWire, Gig-E Vision, USB3 Vision, Camera Link and MIPI.

3.1.2 Comparison of Camera Modules

After analyzing all the parameters from the previous sections, three camera modules were selected for the final cut. They can be seen on the table above. I went ahead and named them camera module “A”, “B” and “C” to simplify the comparison when referring to each one of them.

Table 3.1 Comparison table of different camera modules.

				
Camera Module A				
Model	Frame rate (fps)	Sensor size	Interface	Shutter Type
daA3840-45uc	45	1/1.8”	USB 3.0	Rolling
Brand	Pixel size (H x V)	Resolution	Weight	Price
Basler	2.0 x 2.0 μm	8.3 MP	< 15g	\$199
				
Camera Module B				
Model	Frame rate (fps)	Sensor size	Interface	Shutter Type
Pi Camera Module 3	30	1/2.43”	MIPI CSI-2	Rolling
Brand	Pixel size (H x V)	Resolution	Weight	Price
Raspberry Pi	1.4 x 1.4 μm	11.9 MP	4g	\$25
				
Camera Module C				
Model	Frame rate (fps)	Sensor size	Interface	Shutter Type
IMX519	30	1/2.53”	-	Rolling
Brand	Pixel size (H x V)	Resolution	Weight	Price
ArduCam	1.22 x 1.22 μm	16 MP	8.5g	\$37.99

As noted above, camera module A is the most complete one. It already comes mounted (the plastic surface that surrounds the sensor); it has the highest frame rate and sensor size, and it is equipped with the newest interface. However, it is also the most expensive sensor. One of our goals is to keep the project low cost offering the same, or better quality, therefore, sensor A was not chosen.

The decision then was between camera module B and C. They are similar to each other, so when choosing between the two we went for the one with the largest sensor and of course, the cheapest price. And that is why for iSPY we will be using camera module B, Raspberry Pi's Camera Module 3.

The brand, Raspberry Pi, offers great documentation for their components which has made understanding the sensor a lot better. The specs table for this camera will be attached to the appendix. The camera module comes with a lens; however, it does not fit our requirements so we will replace it with a different lens assembly that will improve our field of view.

The camera module chosen uses the interface MIPI CSI 2, one of the most popular interfaces for computer vision. It is a reliable option because of its bandwidth capacities and price [34].

3.1.3 The Optical Parameters

One of the most critical optical parameters is the focal length. The simplest definition of focal length is the distance between a lens and its focal point. This relationship can be described by the equation:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}$$

Where f is the focal length, d_o is the object distance and d_i is the image distance.

However, for a lens system as complicated as ours this relationship is not so straightforward. Our focal length will actually be called effective focal length, since it will be a combination of different focal lengths in our system. Fortunately, the software that will be used for the simulation of our system, Zemax, will help us to keep our desired effective focal length in check as we add new surfaces and optimize the design.

Equation 3.1 shows that a short focal length is needed for a wide field of view. It is no surprise then that surveillance cameras have focal lengths of just a couple of millimeters. The optical system for iSPY will have a fixed focal length and fixed focus. In other words, the camera will not move around; it will always be "observing" that same part of the room/of the scene. Even if objects/people come close to it the focus will not adjust as our main interest is to count how many people we can see and not facial recognition. The use of fixed focal length lenses is common when working with machine vision, another aspect of iSPY [35]. Our surveillance system, at least to meet the basic and advanced goals, will be designed to work on environments where there is not lighting variations (i.e. inside

buildings such as lobbies and hallways), and the fixed focal length design is advantageous because it gives a stable image quality [36].

The three ways that a focal length can be changed are by moving the lens system, replacing the lens system and/or by changing the sensor size. All these options are not feasible for iSPY [10]. We will use these constraints in focal length and focus to simplify our system. Our lens design will already be complex and adding features like autofocus will increase the cost and time needed to complete the project.

The F-number ($f/\#$) or f-stop indicates how much light passes through an optical system. It is the ratio between the focal length and aperture diameter. Low F-numbers are related to large apertures, more light enters through a lens and the system is great in low light conditions. Lenses with a low F-number are called fast lenses. High F-numbers, on the other hand, describe small apertures, where less light comes into the system and are best used in well-lit places. They also have a greater depth of field than low F-numbers. Lenses with a high F-number are called slow lenses.

Since the security we want to build will be used during the daytime and inside well-lit places, lenses with low F-number are not needed. And this also decreases the cost of our system since moderate to high F-number lenses are cheaper.

Moderate F-number lenses were previously mentioned because it is a trade-off that needs to take place to decrease diffraction. A high F-number is not ideal. When talking about F-number, another parameter that comes in question is the numerical aperture, NA. The two are inversely related for infinite conjugate systems by equation 3.2.

$$NA = \frac{1}{2 \times f/\#} \quad [\text{Eq. 3.2}]$$

NA is used when we refer to the cone of acceptance of a system. It is the range of angles that are allowed, “a measure of the collection efficiency of an optical system” defined as:

$$NA = n \sin \theta$$

Where θ is the half angle of acceptance and n is the refractive index of the lens [37].

Diffraction is the bend of light (waves) when passing through narrow apertures or around objects. When diffraction results from a finite circular aperture, an airy disc pattern is the outcome [38]. This airy disc is known as the diffraction limit of the system, and it is always present, no matter how perfect the lenses are, “the ultimate limit of resolution of an optical system is set by diffraction” [39] [40].

To create a successful system, we should always take into account the diffraction limited spot. When designing our lens system using the optical design software, great care is taken when comparing the current design to the airy disc. We always want to make sure our resolution is not too different from it.

Lens compatibility is crucial when matching a lens system to a sensor. The size of the lens needs to be greater than or equal to the size of the sensor to totally cover the image circle.

When the sensor is larger than the lens, part of the image is cut-off making dark circles around the edge of the image visible, the optical vignetting [41] [42]. For surveillance systems, smaller sensor format sizes are the most popular. They have a smaller lens image circle making the lens and the system more affordable [42].

Optical vignetting is one of the several aberrations that can take place in an optical system. One of the strategies to reduce its effects is to use aspheric lenses; however, this method requires a larger budget since these lenses are expensive. Another procedure to decrease optical vignetting is to purposely limit the field of view, and this is what we will do in our design. The field of view of our camera will be deliberately limited due to cost, size and weight of the final product [16].

3.1.4 Image Acquisition System: The Optical Design

The optical design is an extensive one, from choosing a sensor, to finding the lenses with the correct diameter and focal length. It was a process that required a lot of time, trials, and tradeoffs. Wide field of view cameras are complex. When one wants to receive light from larger angles aberrations greatly increase and to minimize their effects several techniques need to be used.

The first step to design a lens system for our camera was to research what is done in the industry, what others with more experience and time have put together. This research showed us that even standard designs are complex and require aspheric lenses. Image 3.X displays a lens system for a camera of 100 degrees field of view.

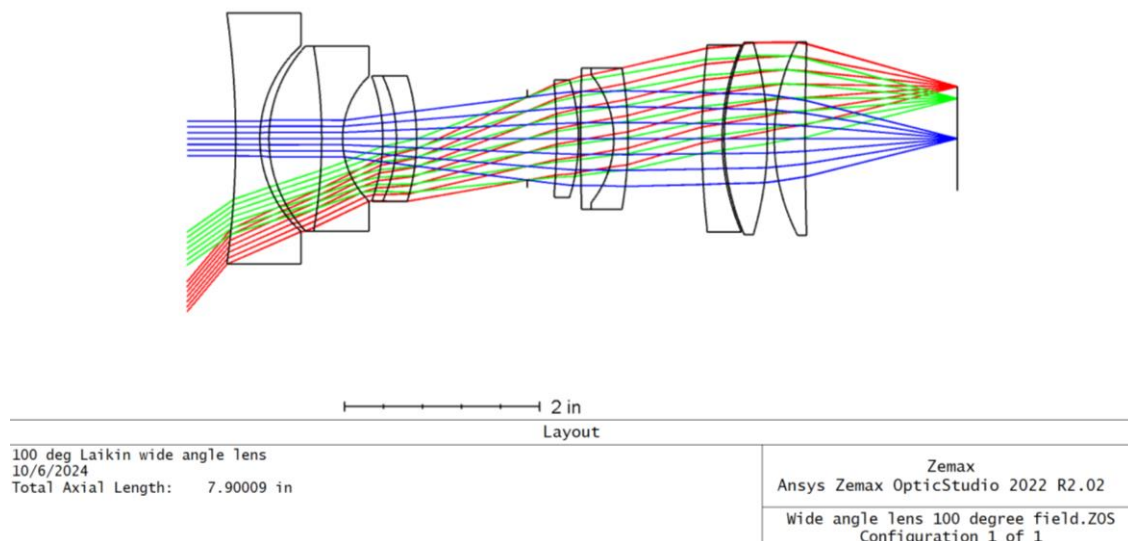


Figure 3.1. Zemax layout of a 100-degree field of view system. This layout can be found under ‘Samples’ inside the Zemax folder when the software is downloaded.

After an extensive investigation on wide-field-of-view cameras used in the medical field, and open patents for phone cameras [11] [43], and seeing the complexity of such systems, I was oriented by supervisors to go for a simpler design, the ones previously mentioned are above the scope of a senior design project, they require years of dedicated work – something like a Master’s thesis. The designs fall under the constraints that will be explored in Chapter 4.

The idea was to find a framework for our design and not start from zero, a common practice in industry. And that is when I stumbled upon the new SPIE’s book, “Designing Optics Using Zemax” by Donald C. O’Shea and Julie L. Bentley. In this book, the writers explore the lens design software, Zemax, with the readers and give them a chance to design a system at the end. This system, that can be found in chapter 12, is the skeleton for the optical design of iSPY.

Following the framework laid by the authors of the book I was able to achieve a similar system with the desired focal length and consequently, the required field of view. However, just like any other simulation some of its approaches involve custom-made lenses, and/or are made for a “perfect scenario” and ignore the feasibility of the design. Therefore, it is up to the optical designer to translate this optimum/perfect design to something tangible that can be recreated.

Both designs, the original one and the one I worked on following the steps from the book [Appendix G] ran into an issue in its early states. The optimal thickness of the lenses was negative, which of course, cannot happen. This issue was fixed by configuring the software and letting it know the limits – setting up boundaries.

Since the book’s intent was to design a lens system for the sole purpose of proving a concept the writers had the freedom to use any type of surface combinations, basically creating their own lens. When following their steps, I also use the surface combinations necessary to make my system work; however, since my plan is to physically build this lens system, to bring it to life, I knew the design would have to be modified.

Having the scope of this project in mind, off-the-shelf lenses are the cheapest and less risky option. Fortunately, the main websites for purchasing lenses from Edmund Optics and Thorlabs have Zemax files for most of their lenses, which facilitates the transition from simulation to a feasible physical system.

3.1.5 Lens Background

The most popular lenses are the standard convex and concave ones. Convex lenses are positive, they can be biconvex (the two surfaces have a positive radius of curvature) or plano-convex (one surface is flat while the other has a positive curvature). Convex lenses are used to focus/converge light onto a specific spot, they also are used to decrease focal lengths. Concave lenses are negative lenses, they can be biconcave (the two surfaces have a negative radius of curvature) or plano-concave (one surface is flat while the other has a

negative curvature). Concave lenses are used to diverge light, they also increase focal lengths.

The physics of lens refraction is the key to understanding any optical system. These standards lenses have been around for centuries, they are the bread and butter of the world of optics. However, the need for more complex systems pushed scientists to come up with ideas such as new surfaces configurations, and the use of different materials to make and coat the lenses.

The standard lenses can also be called singlets. They are the simplest form of a lens. By combining singlets, we can get doublets (cementing two singlets together) or triplets (cementing three singlets together). The reason for such new configurations is to gain more control over the system and refine the imaging process.

An example of a doublet is the achromatic doublet, which is formed by combining a positive and negative lens. The positive lens is called the crown, and for this doublet to work as intended, the crown needs to be made of a material with low refractive index compared to the one material from the flint, the negative lens [44].

The Abbe number is a measure of dispersion when passing through a lens. A high Abbe number is an indicator that different wavelengths have similar index of refraction in a medium, in other words, lower dispersion. In achromatic doublets the crown has a high abbe number while the flint has a low one, and by combining the two lenses chromatic aberration is decreased [45]. We can also use a triplet achromatic or aspherized (more on aspheric lenses in following paragraphs) achromatic for color correction in our system [46].

Convex lenses bend outwardly, concave lenses bend inwardly; however, sometimes a combination of both bends is necessary and that is why meniscus lenses were introduced. Meniscus lenses have curvatures for both of their surfaces, one outwardly and the other inwardly, they can also be called convex-concave lenses. Positive meniscuses are used to reduce the beam's focal spot, while negative meniscuses are used on applications that require beam expansion [47], [48], [49], [50].

A special type of lens is the aspheric lens since it varies its curvature unlike regular lenses. Aspheric lenses are used to reduce aberrations (especially spherical aberrations) and decrease the size of the system since a few aspheric surfaces work just as well if not better than several spheric surfaces when correcting aberrations [51]. Another unique lens is the Fresnel one. Fresnel lenses are formed by etching a series of concentric grooves into plastic. They are used in diverse applications, especially the ones that involve light gathering [52].

The lens design for a wide field of view system is complex, it usually requires several surfaces to decrease aberrations and focal length. While researching the designs used in industry, I found that aspheric lenses are the most popular lens type for such a system, which makes sense since they are extremely efficient in decreasing aberrations. They are custom-made for each application.

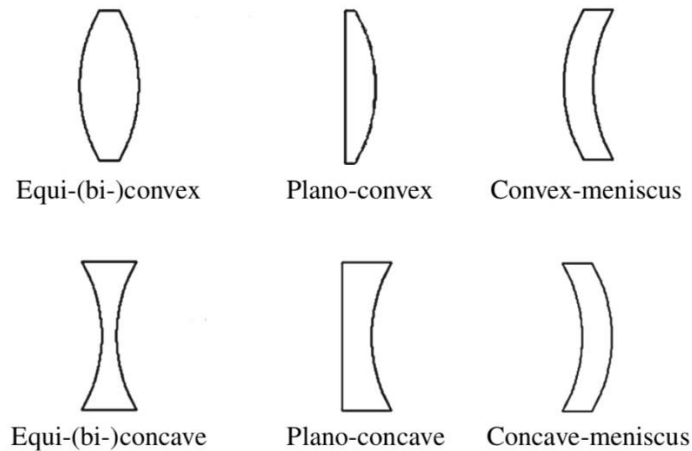


Figure 3.2. A comparison between the standard lenses and the meniscuses found in reference 49.

A widely known method to increase the field of view relies solely on standard lenses, the inverted telephoto lens system. The regular telephoto lens system is made of a “group” of positive lenses followed by another “group” of negative lenses. This configuration enables the focal length to be longer than the physical length of the system. In the inverted version the negative lens or lenses come first than the positive ones, and it is used to widen the field of view [53], [54].

3.1.6 Lens Material

The most common glass types to make lenses are N-BK7, N-SF5 and N-SF11. N-BK7 glass is known for its exceptional optical properties, purity, low-cost, and constant refractive index throughout its surface [55].

N-SF5 is popular due to its high refractive index, and low Abbe number. Due to its high dispersion, it is usually paired with another material with a higher Abbe number to avoid chromatic aberrations, an example of that is the achromatic doublet. N-SF5 is used in applications where high precision is required [56].

N-SF11 is popular due to its high resistance in harsh environments. Lenses with this type of glass are used in systems where the bending of light needs to be significant such as high magnification microscopes [57], [56].

3.1.7 Lens Coating

Another important characteristic of a lens is its coating. Optical coating is the action of depositing a thin layer of materials onto optical elements such as mirrors, lenses, and prisms to change the way light is reflected and/or transmitted. A commonly used optical coating is the antireflection coating in eyeglasses [58], [59].

When choosing a lens, it is important to take into consideration if the lens is coated or uncoated, and what type of coating is being used if any. Depending on the wavelength of the light source different types of coating are used. Another detail is that coated lenses tend to be more expensive than uncoated ones.

For our project, we decided to use lenses with an AR coating. According to a lens manufacturer, uncoated lenses lose 4% of light at each optical surface due to reflections, our system will have several optical surfaces and such loss is undesirable [60]. Since we are dealing with the visible part of the spectrum, the wavelength range of our lenses will be from 400-700nm. And the material used for the coating is MgF_2 . Magnesium fluoride is the most common compound used for anti-reflection coating. It is known for its durability and cost-effectiveness [61].

In Figure 3.3 we can see the comparison between a coated and uncoated lens. The uncoated lens (Graph A) transmission band is broad, enabling the transmission of a large range of wavelengths, while the coated lens (Graph B) reflects an average of 0.5% along the wavelength range of 400-700nm, in other words, the transmission in this range is higher and wavelengths outside of the visible spectrum are blocked.

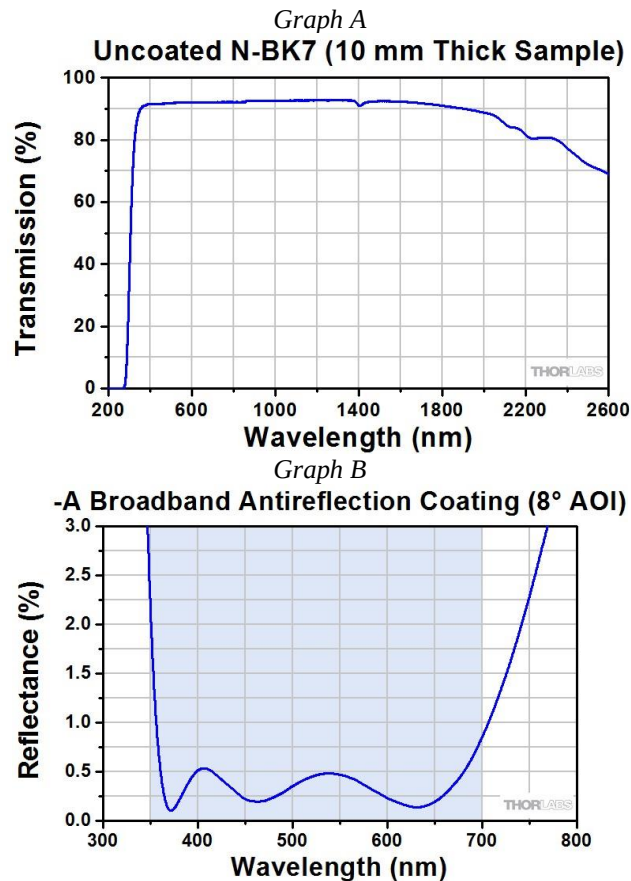


Figure 3.3 A comparison between uncoated [Graph A] and coated [Graph B] lenses from reference 60.

3.1.8 Optical Aberrations

In theory, a lens will always perfectly image an object. However, that is not the case. As the rays move away from the axis, we see a failure to image it, and this distinction between the ideal scenario and the result of actual lenses are a measure of optical errors, also called aberrations. As defined on the Edmund Optics website, “optical aberrations are deviations from a perfect, mathematical model” [62].

When talking about aberrations, we usually discuss first and third order aberrations. Higher order aberrations exist but they are too complex and trying to decrease them improves the image quality only slightly [62].

The first order aberrations are tilt and defocus. Tilt as the name suggests, deals with a shift from the reference wavefront. The actual image is formed at an angle. Defocus happens when the image is formed in the wrong plane [62].

The third order aberrations are spherical aberration, coma, astigmatism, Petzval curvature, and distortion. Spherical aberration happens when light does not focus on the same plane, a mismatch when converging light. This type of aberration blurs the final image, and to decrease we can use aspheric lenses, reduce the entrance pupil of the system so the rays at the edge are cut off, add defocus, use lenses with a higher refractive index or use lens splitting – increase the number of surfaces [62] [14].

Imaging depends on ray heights, and when the rays are coming from far away these heights change – they do not hit the spot for perfect imaging. This change in ray height and magnification with respect to the location of the image causes coma [63], [62]. A way to decrease coma is to move the aperture spot to the center of the system.

Astigmatism refers to the aberration caused when the lens does not perform the same way in the sagittal (horizontal) and tangential (vertical) direction, two focus points are created. This aberration can be decreased by defocusing the system and moving the aperture stop.

Field curvature is an aberration that occurs when imaging onto flat planes (screens, sensors) rather than an ideal curved one, the paraboloidal surface known as the Petzval surface. Its effects can be minimized using spaced doublets [62].

Distortion aberration does not cause blur, but it changes the magnification of the image, it produces images that looked stretched or compressed. Positive distortion is called a barrel distortion, and a negative distortion is called a pincushion distortion. In lens systems designed for wide field of view distortion is the main aberration. The effects of distortion can be decreased by changing the location of the aperture stop [14], [64].

Another aberration that we should take into account especially for this system is chromatic aberration. Chromatic aberration happens when an optical system fails to converge different wavelengths on the same axis, it causes color inconsistency and distortion. To reduce this type of aberration a combination of positive and negative lenses in an

achromatic doublet is used [14]. Achromatic doublets are used to eliminate chromatic aberrations; these types of lenses work by focusing a parallel input beam to a perfect “point” [44].

3.1.9 Lens Selection

The selection of the lenses is one of the most crucial steps in this project, and for this reason it also becomes one of the most time-consuming tasks. Several iterations of this lens system have taken place, and adjustments will continue to be made until the end of the semester to obtain the highest resolution for our image acquisition system.

On top of considering the details previously mentioned such as lens material, coating, its effect on the overall quality of the image (aberration), we also had to take into account price, and the availability of the lenses. Unfortunately, not all the lenses are in stock and since time is a huge constraint in this project, we had to work around that.

Another limitation was that using off-the-shelf lenses required us to work with already set parameters. We do not have a say on what focal length and diameter we want, we must work with what is available to us.

In the latest iteration of the lens system for iSpy biconvex and convex plano lenses are used. The biconvex is made of N-SF5 and the convex plano is made of N-SF11. Both lenses are coated with MgF_2 so only wavelengths in the visible spectrum can be transmitted in our system. This iteration of the lens design gives us the highest field of view (shortest focal length) and the best image.

3.2 Computer Vision

The computer vision aspect of iSPY is the most important part of our system. Computer vision uses mainly deep learning algorithms such as convolutional neural networks (CNN), recurrent neural networks (RNN) and natural language processing (NLP) just to mention a few.[65], [66] Convolutional neural network is used for images whereas recurrent neural networks are for video.[65] Both handle their own things to process in a very similar way.

When it comes to CNN, they use a method that breaks down the image into pixels and from the pixelated form of the image then gives a label and tag.[65] CNN also uses a process that utilizes a kernel over the image creating a sort mapping of features or edges.[66] RNN uses this same process though it focuses on frames of a video and the progression of those frames.

The use of both algorithms set up the overall functionality and process of computer vision. There is one other thing which is Deep Neural Network (DNN). If deep learning and computer vision combined into a hybrid process that could send the detected object to a DNN then the number of resources when computing is decreased.[66] The process mentioned is more common today. Even though there are other things out there like 2DCNN and 3D CNN that are out of scope for our project.

The use of neural networks has advanced over time to become more compact into simpler coding commands for computer vision purposes like image classification, object detection, human body detection and even tracking people or objects.[65] The coding libraries that utilize the simple coding commands that can handle computer vision vary from simple to more complex. The coding libraries were created by different programming languages. All these programming languages have their own library to handle computer vision tasks. These programming languages include C++, Java and Python.

3.2.1 C++ Computer Vision Library

The C++ library for computer vision is Convolution Architecture for Fast Feature Embedding (CAFFE) which was created by a research group from University of California, Berkley.[67], [68] The library is an open-source framework that can handle multimedia, vision and speech like tasks. Some features that CAFFE has are object detection, object classification and learning semantic features.[67] CAFFE uses something known as CUDA which was developed by NVIDIA for GPU computations.[67] CAFFE is modular and can be used with other programming languages such as Python, MATLAB and NumPy.[67]

The library is very useful for certain things in computer vision however the library does have its own pros and cons. Somethings that are useful about CAFFE are its high performance for fast feature embedding and it supports the use of many deep learning architectures.[68] However there are a few downfalls which are more manual coding for certain tasks and limited to only a few features that computer vision can do.[68] CAFFE is only one out of many libraries used for handling computer vision like tasks.

3.2.2 Java Computer Vision Library

The Java based library is known as BoofCV which is an open-source library as well.[68], [69] BoofCV has various features ranging from image processing to geometric vision. It can render features in 2D and 3D spaces.[68], [69] BoofCV was made from scratch for real time applications such as image association, mosaics, visual odometry and 3D structure estimation.[67] The library can also stabilize images. Even though this library has more in the way of computer vision aspects and can handle more tasks it has its flaws. The only flaw that can be determined and seen is slower performance in certain low-level operations.[68] Java based library BoofCV is faster than most other libraries when it comes to certain operation handling.

3.2.3 Python Computer Vision Libraries

Python has various computer vision libraries that are used when it comes to detection, recognition and image segmentation. The libraries utilize pre-built algorithms to achieve different performance factors whether it be accuracy to efficiency or both at once. In this section I will examine three different Python computer vision libraries and discuss briefly the choice made for the library I chose to use for iSPY.

3.2.3.1 TensorFlow

TensorFlow is a library that was developed by a research group at Google.[70] It is one of the many libraries in Python that can do computer vision duties. TensorFlow supports many other programming languages and platforms, not just Python or PyTorch. It supports developing models on mobile and embedded devices opening up to more systems.[70] The TensorFlow library for computer vision is simple to use for anyone that needs to use it. It has a few features which are object identification, picture categorization and facial recognition to name a few.[68] TensorFlow has regular updates with new features and enhancements offering more to the user.[68] It has strong features with effective operation though it consumes a lot of resources.[68] There are other components and versions of TensorFlow which to name one is TensorFlow Lite. Although it relies on a bunch of resources it is a very great toolkit for computer vision aspects.

3.2.3.2 SimpleCV

SimpleCV is a collection of libraries and software used for vision applications.[71] It allows you to work with images and video streams from different sources. Although this library can only be used on other platforms like Mac, Windows and Linux, which means it can't be used with other programming languages.[71] SimpleCV has similar features as the other libraries mentioned though can-do object detection, image segmentation and image arithmetic.[70] It can also deal with detecting and recognition in a video. It is more concise and simpler to use based on its functionality. SimpleCV includes libraries like OpenCV without the knowledge of image processing. It is very useful for a beginner in computer vision and understanding how to detect items in an image or video depending on the system. SimpleCV is limited when it comes to features than OpenCV. It also is more used for prototyping.[70]

3.2.3.3 OpenCV

OpenCV is more popular among those that work with computer vision applications. It is very versatile in the number of functionalities it can do and handle. Unlike its predecessor it can deal with more complex computer vision imaging and video aspects. OpenCV has been used by beginners and more experienced individuals for many projects like detection of objects in each image input into the program. One of the only libraries that have over 47 thousand people that use it and exceed 18 million downloads.[68], [72] It was also written and developed in C++ programming language.[68], [72] Although it is commonly used in Python by the users which is OpenCV-Python.[73] OpenCV-Python uses the Python API and combines Python with OpenCV C++ API.[73]

OpenCV was created by Intel in 1999 and was first released in 2000.[73] The library system has over 2500 optimized algorithms which is more than any other library system when it comes to computer vision.[68], [70], [72] Stands the reason as to why OpenCV is not limited in features as other libraries. OpenCV also has many versions that have added more

features and functionalities over time. Some of the features that OpenCV can do are human action classification in videos, object motion tracking, eye movement tracking and 3D model extraction just to name a few of the features.[68], [72] There are many more features as denoted before.

However, this library is not as easy to use and does take some time to learn.[68] Although the benefits outweigh these issues since this library is so expensive. Earlier I mentioned OpenCV-Python and how it uses a combination of certain things. OpenCV-Python is a library with Python bindings meaning we can utilize Python wrappers with the original OpenCV C++ implementation for computer vision applications.[73] I have decided to use OpenCV-Python for our project. OpenCV uses many algorithms for body recognition which is what we are focusing on. One of them is Viola-Jones Haar Cascade and another is Histogram of Oriented Gradients (HOG).[74]

3.3 Haar Cascade vs. Histogram of Oriented Gradients (HOG)

Algorithms are specifically useful in handling computer vision tasks. The Haar Cascade and Histogram of Oriented Gradients (HOG) are two of the basic and classical algorithms used for handling such tasks like detection and image segmentation to name a few. The two algorithms are the building blocks for more advanced algorithms that are discussed later on in this chapter.

3.3.1 Haar Cascade

The Haar Cascade is an algorithm used to detect objects, bodies and faces in video as well as images. It uses CNN for processing objects, bodies and faces in a video or image. Haar Cascade was introduced by two individuals Viola (Paul Viola) and Jones (Michael Jones) in 2001.[75], [76] They wrote a paper known as “Rapid Detection using a Boosted Cascade of Simple Features” that introduced this technique that has been used for a long time ever since it was introduced.[76] When this technique was developed it utilized the combination of features. These features known as Haar features were able to detect patterns for edges, lines and center-surrounded features.[75] When the features are detected they are recognized and processed so that when scanning an image or real time video those features are not repetitive. The reason for this is all about time and efficiency. This is ideal for detecting objects in images however there is cascade which breaks down the process of detection[75].

The technique is very popular and is used among the users of OpenCV as well as OpenCV-Python. Haar Cascade is very useful due to its ability to exclude irrelevant regions when scanning an image.[75] Haar Cascade is an algorithm that is used in various ways. Some of these include recognition and detection in real time video or images. The way Haar Cascade works is with pre-built packages known as classifiers and with these pre-trained classifiers we can use them to our advantage in our computer vision program.[75] The Haar Cascade pre-trained classifiers are stored as XML files.[75] The only way for this to work

in the computer vision program is to reference to the XML file. Some of these XML files are trained better than others and can be modified to be more accurate.

The detection of objects can have errors due to the use of these classifiers which affect the accuracy of detection or recognition. However, these classifiers can handle almost detecting everything you can think of in an image or real time video such as face, body (lower/upper/full) and even color detection.[75] Haar Cascade is only one option for detecting bodies in a real time video there is another which is HOG.

3.3.2 Histogram of Oriented Gradients (HOG)

Another algorithm used for body detection in real time is HOG. HOG is more of an accurate and fast choice when it comes to body detection. HOG was introduced by Navneet Dalal and Bill Triggs in a paper they wrote which was “Histogram of Oriented Gradients for Human Detection”.[77] The algorithm extracts out features based on the gradients from the pixel intensities which it analyzes to determine the edges of the body.[78] When this is done those edges of the body shall be highlighted to make it clear and known that these are the edges of the person that is being detected. HOG doesn’t allow appearance change due to lighting or position in fact it can handle the different appearances that can be caused by lighting or position.[78]

It recognizes patterns that capture an individual’s body shape. The creation of these patterns helps even if the person is partially shown in an image or video which means it can still identify the visible parts of the body. The process is known as descriptor patterns.[78] HOG features are then input into a classifier which is a Support Vector Machine (SVM) allowing accurate predictions.[78] The HOG feature descriptor can be calculated by certain steps which are preprocessing, calculate the gradient images, calculate histogram of gradient, normalization and calculate the Histogram of Oriented Gradients feature vector.[77], [79], [80], [81]

The algorithm has a process that it utilizes for the best result in detecting objects, people and hand gestures. HOG is a better choice than Haar Cascade due to its efficiency in real-time applications and its ability to handle variations such as different lighting. Behind the HOG algorithm comes an extensive process of how the histogram is created. The first thing that is done is image or frame preprocessing which is adjustment of size on image or frame creating a patch (smaller part from larger image or frame).[80], [81] Next is calculating the histogram of gradient images or frames by using kernels for filtering the image or frame.[81] If OpenCV is used there is an operator known as Sobel that takes in kernel size which should be set to 1.[81]

Then the magnitude and direction of the gradient is determined by using a function in OpenCV known as `cartToPolar`. [81] Once every pixel’s gradient’s magnitude and direction is calculated then it is broken into cells which can vary in size the larger the cell the faster the process runs. The calculation of histogram of gradients happens for every one of the cells.[81] After this the histogram is formed and normalization begins.

The last part of the process brings all the vectors that were generated in the last two steps together to create a giant vector this is known as calculating the Histogram of Oriented Gradients feature vector.[81] All of these steps are used in the HOG algorithm and shows that this algorithm is more structured for real time human detection. Although it still has

errors that occur like false positives (non-people) and false negatives (people not detected) when detecting either in an image or video even if it is more accurate than Haar Cascade.[79] Also OpenCV only has detectors for people.[79] There are newer OpenCV algorithm's that have recently come out.

3.4 Newer OpenCV Algorithm's

Recently newer more advanced algorithms have been developed and designed to be used in OpenCV. The newer algorithms use certain aspects of older algorithms previously mentioned. These newer algorithms can do more advanced tasks such as real-time detection, recognition and tracking movement which improves the accuracy as well as the efficiency of those tasks.

3.4.1 YOLO (You Only Look Once)

In OpenCV there is another set of algorithms known for their accuracy, speed and efficiency which is YOLO. YOLO has many versions that all vary in different performance rates and abilities in detection. YOLO was developed by Joseph Redmon and Ali Farhadi in 2015.[82] The real time object detection algorithm both classifies and detects the object at the same time. The process makes this algorithm very versatile in its ability to process the detection of objects faster and more efficiently, which is useful in detection in videos. The algorithm is a single stage object detector which means it can classify and detect in one single pass through a CNN.[82], [83]

YOLO is used in self-driving cars, surveillance systems and robotic systems. Also, it is very useful in real time detection as well as recognition such as real time video analytics and surveillance. Due to the algorithm being useful in various applications it makes other algorithmic models non-existent. It has a pipeline architecture meaning it combines specific regions together.[83] YOLO has its benefits such as the ability to detect objects instantly, resource use is optimized, performance, accuracy, flexibility and efficiency.[83]

Even though YOLO has great benefits making it one of the top algorithms it does have its downfalls. The downfalls for YOLO are that it only focuses on object detection not image segmentation, less accurate than some other methods like two-shot object detection methods, struggles with detecting very small objects and no tracking capability for tracking objects over time. However, there are many versions of YOLO meaning that the limitations might not be the same which is the reason I will focus on the most recent three which are YOLOv9, YOLOv10 and YOLOv11.

3.4.1.1 YOLOv9

YOLOv9 is the 9th version of the YOLO algorithm. It combines models together to create a more accurate and efficient algorithm. YOLOv9 was created by Chien-Yao Wang, I-Hau Yeh and Hong-Yuan Mark Liao on February 29th, 2024.[84] Version 9 was mentioned and described in detail in a paper written by Chien-Yao Wang with his team known as

“YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information”.[84] The paper went into detail on the design and the structure of YOLOv9.

One of the issues when it comes to deep neural networks (DNN) is that it loses input data information which is known as information bottleneck.[84], [85], [86] There are currently three different methods used to combat this issue which can reduce the loss of information. The three methods used are reversible architecture, masked modeling and deep supervision.[84], [86] Although they reduce information loss they are not well developed in training and inference processes.[86] YOLOv9 introduced two very useful techniques which happen to be Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN).[84], [85], [86] Both making YOLOv9’s learning capacity increase and able to retain crucial information when detecting objects.[85] PGI and GELAN also help with the performance of version 9. Version 9 tackles the Information Bottleneck Principle and the use of Reversible Functions.[85]

The Information Bottleneck Principle is a challenge in DNN which is as data passes through multiple layers of a network there is potentially an increase in loss of information.[85] YOLOv9 implements PGI to handle this challenge by preserving data throughout the network which ensures reliable gradient generation, better model convergence and performance.[85] YOLOv9 can use reversible functions which are reversible if the function can be inverted without any loss of information and with the use of reversible functions it can mitigate risk of information degradation.[85]

PGI is a combination of a few things that deal with DNN issues. It has a main branch for inference, an auxiliary reverse branch for reliable gradient calculation and multi-level auxiliary information to deal with deep supervision issues effectively without extra inference costs which is shown in Figure 1.[86]

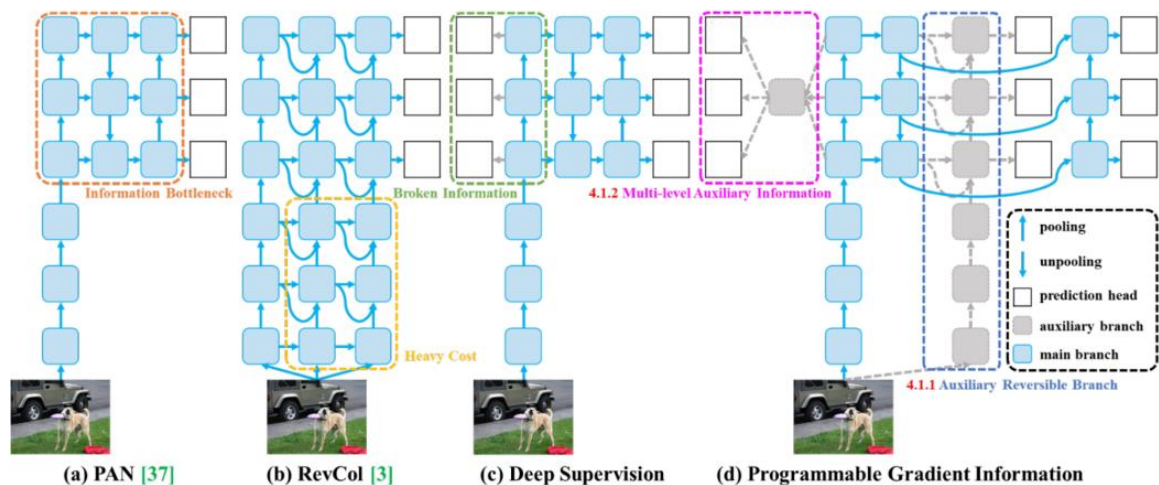


Figure 3.4 PGI Architecture from Reference [86]

PGI also includes an auxiliary supervision node that helps with backpropagation of gradients.[86] The main branch maintains high performance during inference process and can bypass auxiliary components as needed to keep this performance the same.[86] The auxiliary reversible branch utilizes the reversible architecture to maintain information from

getting lost in the layers of the DNN and retaining the information for learning.[86] It helps by not allowing anything to affect the inference speed. The last part of PGI is the multi-level auxiliary information which helps with making better predictions of objects no matter the size.[86]

GELAN is more of a supporter for the PGI in that it provides a flexible and efficient design that supports various computational blocks shown in Figure 5.

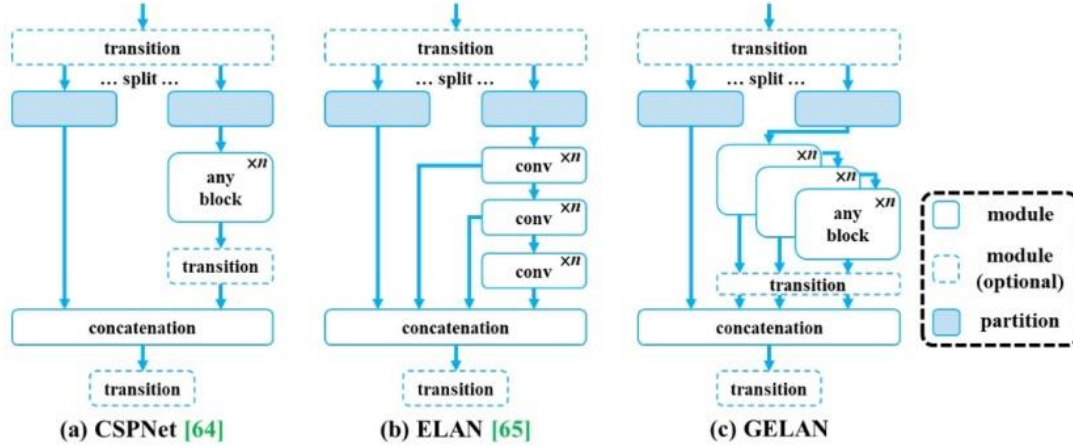


Figure 3.5 GELAN Architecture from Reference [86]

It combines features from CSPNet and ELAN. GELAN takes CSPNet’s gradient path plan with ELAN inference speed optimization to make it a versatile architecture.[86] It has quick inference times with retaining accuracy. GELAN has two critical aggregation techniques which happen to be SPPELAN and RepNCSPPELAN4 models.[86] SPPELAN is the implementation of Spatial Pyramid Pooling (SPP) within ELAN architecture. SPPELAN deals with layer aggregation in such a way that it is able to extract from spatial hierarchies.[86] RepNCSPPELAN4 is a component that advances the CSP-ELAN to improve the feature extraction process.[86] GELAN shows that it can keep and retain information of features after many layers in a DNN.

YOLOv9 has many models which are broken down into lightweight, medium and large models. The models vary in performance ratings for real time object detection which were tested with the MS COCO dataset shown in Table 1.

Detection (COCO)		Segmentation (COCO)			
Model	size (pixels)	mAP ^{val} ₅₀₋₉₅	mAP ^{val} ₅₀	params (M)	FLOPs (B)
YOLOv9t	640	38.3	53.1	2.0	7.7
YOLOv9s	640	46.8	63.4	7.2	26.7
YOLOv9m	640	51.4	68.1	20.1	76.8
YOLOv9c	640	53.0	70.2	25.5	102.8
YOLOv9e	640	55.6	72.8	58.1	192.5

Figure 3.6 Comparison of State-of-the-Art Real-Time Object Detectors from Reference [85]

In the table above there is the number of parameters and computational needs (FLOPs) for each model of YOLOv9.[85]

3.4.1.2 YOLOv10

YOLOv10 was introduced by Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Jungong Han and Guiguang Ding in a paper known as “YOLOv10: Real-Time End-to-End Object Detection” which was released May 23rd, 2024.[87] YOLOv10 is a built up from the previous version of YOLO though includes techniques to handle even more accurate detection in real-time. It has two major components, which are Consistent Dual Assignments for NMS-free Training and Efficiency-Accuracy Driven Model Design.[87], [88] The Consistent Dual Assignments for NMS-free Training consists of two components which are dual label assignments and consistent matching metrics.[87], [88]

The process deals with Non-Maximum Suppression (NMS) post-processing with help from Detection Transformer (DETR) Architecture due to redundancies of bounding boxes in detection which is usually caused by NMS post-processing.[87], [88] Dual label assignments assist with duplication and redundancy from NMS post-processing by avoiding the use of NMS post-processing all together. It combines one-to-one and one-to-many matching techniques to make a more sufficient prediction during the inference process.[87], [88] When it goes through the inference process it will disregard the one-to-many head and use the one-to-one head to make the prediction however during training both heads are jointly optimized for the model.[87]

The joint optimization allows greater supervision with use of the one-to-many assignment which improves the supervision from previous versions of YOLO.[87] Dual label assignments reduce inference cost and adopts the one selection which performs the same as the Hungarian matching reducing the training time.[87] Consistent matching metric is used for making sure that one-to-one head aligns with one-to-many head during training.[88] The metric for both heads allow for the model to use the best samples instead of using samples that are not best for both heads.

Next major part to YOLOv10 is Efficiency-Accuracy Driven Model Design. The Efficiency-Accuracy Driven Model Design is for balancing the efficiency and accuracy of the algorithm.[88] Unlike previous YOLO versions that do have improvements in their efficiency and accuracy, they still have redundancies as well as limitations. The Efficiency-Accuracy Driven Model Design is split into two models, one is for efficiency and the other is for accuracy. Efficiency driven model design reduces computational cost without reducing its overall performance.[88] There are three ways that the Efficiency driven model does this which are lightweight classification head, spatial channel downsampling and rank-guided block design.[88]

YOLOv10 adopted a lightweight classification head which uses two depthwise separable convolutions.[88] Spatial channel downsampling is common in typical YOLO versions and models, however, cause computational cost. YOLOv10 decouples the spatial reduction and channel increase operations allowing efficient downsampling.[87] Besides doing both depthwise convolution for reduction and increasing number of channels in a single

convolution layer like most algorithms. YOLOv10 can do both by pointwise convolution which can retain more information during downsampling.[88]

The last way Efficiency driven model design reduces the computational cost is by rank-guided block design which assigns ranks to stages in the convolution layers. The rank-guided block design calculates a rank of each stage, identifies redundant stages and replaces them with compact inverted blocks (CIB) structure.[88]

The Accuracy driven model design has two ways to increase the accuracy of the algorithm. One of the ways is large-kernel convolution and the other is partial self-attention (PSA).[87], [88] The large-kernel convolution resizes convolution area to be larger than its original size. When it comes to YOLOv10 the use of large-kernel convolution is used in CIB within deeper stages to increase the size of the second depthwise convolution which is 3x3 to 7x7.[87], [88] Also, adds structural reparameterization techniques which includes the addition of another 3x3 depthwise convolution branch.[88] The addition of this branch helps optimization issues with no additional inference cost.[88]

Version 10 can detect smaller objects by the increase in their performance without any cost to either inference or computation. PSA is introduced in version 10 to use global representation learning with reduction in computational cost.[88] PSA splits features into two parts after a 1x1 convolution which is merged later by 1x1 convolution after being concatenated.[88] The PSA is only applied to low resolution stages to reduce excessive overhead from complexity of self-attention and improves overall capabilities of YOLOv10 algorithm.[88]

Version 10 has many features that are included to improve detection of individuals or objects in real-time. The features are advanced backbone architecture, enhanced feature pyramid network, adaptive anchor-free detection, advanced data augmentation and hybrid precision training.[89] Most of these features are built off from previous versions of YOLO and mainly from version 9. Although they are improved by YOLOv10's structure and design. There are also different variations of version 10.

Model	Test Size	#Params	FLOPs	AP ^{val}	Latency
YOLOv10-N	640	2.3M	6.7G	38.5%	1.84ms
YOLOv10-S	640	7.2M	21.6G	46.3%	2.49ms
YOLOv10-M	640	15.4M	59.1G	51.1%	4.74ms
YOLOv10-B	640	19.1M	92.0G	52.5%	5.74ms
YOLOv10-L	640	24.4M	120.3G	53.2%	7.28ms
YOLOv10-X	640	29.5M	160.4G	54.4%	10.70ms

Figure 3.7 YOLOv10 Model Variations from Reference [90]

YOLOv10 has many different models like YOLOv9. It is separated the same way as well with lightweight, medium and large models. The six different models depend on efficiency and size which are shown in Figure 3.

3.4.1.3 YOLOv11

The newest version of YOLO is YOLOv11. It is currently the most recent version of YOLO to date and improves based off version 8 however it is more accurate and efficient over version 10. YOLOv11 was developed by Ultralytics on September 27th, 2024.[91] It was announced at Ultralytics YOLOVision event making its official release to the public.[91], [92], [93] YOLOv11 is an improvement from the team at Ultralytics from version 8 which was the latest version from Ultralytics that was released. YOLOv11 architecture has new integrations and parameter tuning.[92]

YOLOv11 uses a C3k2 block which is a custom implementation of the CSP Bottleneck.[92] The C3k2 block is faster and efficient in computation without loss of performance.[92] It uses two convolutions and not one large convolution.[92] When it comes to Cross Stage Partial (CSP) Bottleneck it is a network that splits the feature map and takes one part through a bottleneck layer then merges it to the other part with the output of the bottleneck.[92] In other words the CSP Bottleneck module divides the feature map into two parts which one part goes through bottleneck layers and the other goes to the output.[92] YOLOv11 also utilizes Cross Stage Partial with Spatial Attention (C2PSA) which improves spatial attention in the feature maps giving the ability to focus on a given region of interest more effectively.[92] If a region is focused on then features can be filtered out spatially.

The algorithm adds C3k2 and C2PSA blocks to the backbone which deals with extraction of features from an input image at multiple scales and resolutions.[92] The additions to the backbone increase the capability of spatial attention and efficiency in computations making it faster and perform a bit better than version 10. The C3k2 block is also in the neck which is the aggregating of features from multiple resolutions and is passed to the head for prediction.[92] If the C3k2 block is included in the neck it once again improves speed and performance of feature aggregation process.[92] The way that YOLOv11 uses C2PSA to focus on key regions for better detection makes it able to detect obscured and small objects.[92]

The head part of the architecture deals with making the final predictions meaning bounding boxes and classifying objects.[92] In the head part of the architecture for YOLOv11 inference is faster and parameters are more efficiently defined by the addition of C3k2 block making this possible.[92] Very similar to previous versions of YOLO YOLOv11 has five different variations.

Model	size (pixels)	mAP ^{val} 50-95	Speed CPU ONNX (ms)	Speed T4 TensorRT10 (ms)	params (M)	FLOPs (B)
YOLO11n	640	39.5	56.1 ± 0.8	1.5 ± 0.0	2.6	6.5
YOLO11s	640	47.0	90.0 ± 1.2	2.5 ± 0.0	9.4	21.5
YOLO11m	640	51.5	183.2 ± 2.0	4.7 ± 0.1	20.1	68.0
YOLO11l	640	53.4	238.6 ± 1.4	6.2 ± 0.1	25.3	86.9
YOLO11x	640	54.7	462.8 ± 6.7	11.3 ± 0.2	56.9	194.9

Figure 3.8 YOLOv11 Model Variations by Reference [90]

YOLOv11 has lightweight, medium and large models just like the other prior versions. Not only does it have the models, but it also has features like enhanced feature extraction, optimized for efficiency and speed, greater accuracy with less parameters, adaptable to be accessed from other environments and a range of tasks supported by version 11.[91] The tasks that YOLOv11 can handle are object detection, image segmentation, image classification, pose estimation and oriented bounding boxes (OBB).[92], [93]

3.4.1.4 YOLO-NAS

Another version of YOLO is YOLO-NAS which is different than the others by not only structure but its functionality. YOLO-NAS was developed by Deci AI and was designed using Deci's AutoNAC NAS technology.[94], [95], [96] The way the algorithm model works is by Neural Architecture Search (NAS) technique meaning it can search for the best network architecture for a certain task without human intervention.[94], [96], [97] NAS searches for an architecture that has the best trade-off between multiple limitations of prior YOLO versions like accuracy, complexity and model size.[95], [96] Although NAS is applied by AutoNAC (Automated Neural Architecture Construction) which is an engine that allows task input, data characteristics, inference setup and precise performance tasks.[94], [97]

After AutoNAC takes all the required information it then finds and identifies the most optimal architecture. The architecture that is found has a balance between precision as well as inference speed for the certain task provided by the user to AutoNAC. In inference there is a stack known as inference stack that has quantization and compilers. AutoNAC accounts for the quantization and compilers that are included in the inference stack.[97]

YOLO-NAS introduces a brand-new block which is a quantization aware block and selectivity of quantization.[94], [97] The selectivity of quantization is mainly for performance optimization making YOLO-NAS perform at its best. Another newly introduced feature is these quantized versions of YOLO-NAS which are INT8 versions. The INT8 versions use quantization techniques to convert from floating-point models to integer models, allowing more efficient inference.[94], [97] YOLO-NAS still uses inference reparameterization and spatial attention like prior YOLO versions. When the quantized versions are used there is a slight precision drop which is a bit different than other versions of YOLO.

Just like the other versions of YOLO that have features so do YOLO-NAS. The main features of YOLO-NAS are quantization-friendly basic block, sophisticated training and quantization as well as AutoNAC optimization which includes pre-training.[95] Also, YOLO-NAS has its own models like those that have been mentioned before in other YOLO versions. It has lightweight, medium and large models. However, it only has three models YOLO-NAS S for small, YOLO-NAS M for medium and YOLO-NAS L for large unlike other versions that have more variations. Although it does have the quantized variation making the total number of models six which is shown in Figure 5.

Model	mAP	Latency (ms)
YOLO-NAS S	47.5	3.21
YOLO-NAS M	51.55	5.85
YOLO-NAS L	52.22	7.87
YOLO-NAS S INT-8	47.03	2.36
YOLO-NAS M INT-8	51.0	3.78
YOLO-NAS L INT-8	52.1	4.78

Figure 3.9 YOLO-NAS Model Variations by Reference [97]

The algorithm model is very flexible, scalable and useful for a variety of things including security systems. It has more performability for speed, accuracy and efficiency for many areas of limitations in other versions.

3.5 Comparison of Computer Vision Algorithm Models

After collecting and gathering the data from the other sections, three of the top computer vision models were chosen to compare. The models can be seen in the above tables with their own variations included. I decided to label them “Model 1”, “Model 2” and “Model 3” for simplicity purposes when referring to them.

Table 3.2 Comparison table of computer vision algorithms.

Model 1		
Model	mAP	Latency (ms)
YOLOv10 N	38.5	1.84
YOLOv10 S	46.3	2.49
YOLOv10 M	51.1	4.74
YOLOv10 L	53.2	7.28
YOLOv10 X	54.4	10.70

Model 2		
Model	mAP	Latency (ms)
YOLOv11 N	39.5	1.55
YOLOv11 S	47.0	2.46
YOLOv11 M	51.5	4.70
YOLOv11 L	53.4	6.16
YOLOv11 X	54.7	11.31

Model 3		
Model	mAP	Latency (ms)
YOLO-NAS S	47.5	3.21
YOLO-NAS S INT8	47.03	2.36
YOLO-NAS M	51.55	5.85
YOLO-NAS M INT8	51.0	3.78
YOLO-NAS L	52.22	7.87
YOLO-NAS L INT8	52.1	4.78

In the tables above Model 1 seems to be everything we look for in a computer vision algorithm model with high precision and low latency levels. Model 1 is referencing YOLOv10 and is broken into comparable variations with Model 2. It has a high mAP rate which is the mean average precision rate. The mean average precision is the precision of detection for objects in either image or video and the accuracy of bounding boxes. Model 1 has low latency levels meaning that the speed of detection is at a faster rate. YOLOv10 N has a 38.5 mAP and latency is at 1.84 milliseconds as shown above. Although this seems like the route to go, one of our goals is to have high precision as well as accuracy without the cost of speed. This meant that Model 1 was not chosen to be the algorithm model for iSPY.

The decision then was focused on Model 2 and 3. Both models had their own pros and cons, especially when it came to the large model variations. Model 2, which references YOLOv11 has varying rates dependent on model. YOLOv11 N has a 39.5 mAP and latency is at 1.55 milliseconds which is better than YOLOv10 N. It has higher precision with lower latency which is needed for iSPY. Although Model 3, which refers to YOLO-NAS, also has varied rates. YOLO-NAS S is at 47.5 mAP with a 3.21 millisecond latency were as YOLOv11 S has 47.0 mAP with 2.46 millisecond latency. If we move to YOLO-NAS S INT8 then the mAP is 47.03 with 2.36 millisecond latency.

However, if we use YOLO-NAS M then mAP is 51.55 with 5.85 millisecond latency and YOLOv11 M has 51.5 mAP with 4.70 millisecond latency. If YOLO-NAS M INT8 is considered, then mAP is 51.0 with 3.78 millisecond latency. So, Model 3 which is YOLO-NAS will have lower latency at the cost of lower precision on higher models specifically INT8 models for YOLO-NAS than Model 2 which is YOLOv11. Even though YOLOv11 higher end models have higher precision with higher latencies that are not affecting the

precision meaning that precision is not at the cost of lower latency though latency is at the cost of higher precision which is why for iSPY we will use Model 2 YOLOv11 as our algorithm model for computer vision.

If we are going for high precision and accuracy with lower latency, then either YOLOv11 N or YOLOv11 S are the models to utilize. Even though YOLOv11 M is in the middle of the range it has a much higher latency than the lower two models which makes it a possible choice if YOLOv11 N or S is not selected. Overall YOLOv11 models outperform in speed and accuracy among the other versions.

3.6 Classifiers and Object Detectors

A classifier is used in most computer vision algorithms to classify input data such as input image and video. The classifier can be a pre-built package which is already pre-trained. If the classifier is already pre-trained, then the only thing needed is to reference the file that contains this data. Some of the classifiers used in computer vision algorithms have been mentioned like cascade and SVM classifiers. Many classifiers have switched to object detectors to handle more rigorous tasks such as tracking, segmenting and pose detection. The most recently developed algorithm models for computer vision can do all or at least most of these tasks meaning that the object detector is more advanced than the older classifiers like the ones Haar Cascade and HOG utilize.

Newer algorithm models like YOLO use an object detector. YOLOv9, YOLOv10, YOLOv11 and YOLO-NAS have an object detector which is not a classifier like cascade or SVM, instead it is a single-stage object detector.[82] Also, just like classifiers object detectors can be pre-built packages and pre-trained meaning they can be used by referencing to their file in the computer vision program. The newer models rely on these object detectors to do the detection and classification together for input image and video data. However, the object detectors still end up utilizing the classifiers to do the classification part as the detector mainly handles the detection and not the classification of the object.

Object detectors and classifiers have their own important way of handling tasks. Object detectors can handle two things at once which helps the algorithms precision, accuracy and efficiency without the cost of loss data such as double bound boxes or bound boxing an item that is mislabeled. The detector can determine objects in real-time which is the reason that improved YOLO versions can focus on pose detection and object tracking. Classifiers can only handle one thing which is classify objects in an image or video usually not used in real-time applications although HOG can handle real-time applications with the SVM classifier.

Even though object detectors and classifiers differ slightly they have commonalities. Both can handle segmentation, captioning and recognition. The only difference when it comes to the commonalities is the way classifiers and detectors handle each one of them. When it comes to segmentation classifiers can give a class label to each pixel in the image whereas detectors can identify then segment the object out of the image. The use of either classifier

or detector is dependent on some factors. These factors are computer vision algorithm model used, computer vision task and performance for given task. All these factors will determine which one is utilized in the computer vision program.

3.7 Body Recognition vs Body Detection

There is a fine line between recognition and detection which is usually why people mix the two of them up. Security systems can either do one or the other and most current security systems can do both. Body detection differs from body recognition by a couple factors such as labeling, feature extraction and confidence leveling. All these factors have a role in the way detection and recognition work. Although there is a difference it doesn't mean that they can't co-exist together in one system like mentioned earlier in this section.

3.7.1 Body Recognition

Body recognition is like facial recognition and the process is the same with minor additions. Body recognition or facial recognition uses an input image or video frame and scans it using convolution. Breaking the image or video frame up into regions also known as cells which can be 3x3 or larger. After the image is broken up, features are extracted although these features are more detailed. The features that are extracted for facial recognition can be eyebrows, lips and facial hair if applicable. If the facial recognition program is trained with images of different eyes, eyebrows and lips with varying colors then it can determine what eye, eyebrow or lip color the person has by recognition. Also, a facial mapping can be generated using facial recognition.

Body recognition has the same training setup except for size of the individual, movement and pose being added. It differentiates between object classes such as humans, animals and any other object classes that can be created from the image or frame. Body recognition can break down a real-time video and determine the people in each frame by labeling. Labels are used to identify individuals in the frame or image rather than other objects. Mainly body recognition is used for classification purposes to classify an individual based off its features and movement.

3.7.2 Body Detection

Body detection is simply distinguishing an individual in each frame or image from the background. It is trained with features though not to the same extent as body recognition. The features are more rough and rigid than those from body recognition. Body detection will box out an individual from the image although it will not label them as a person like body recognition would.

When it comes to real-time detection the use of body detection can be used to only box out the person or people in the real-time video frame or image. It utilizes bounding boxes to show the detection of the object whether it is a person, animal or any other object. However, it will not label the object it detects, which means that it is not detailed enough to do the

classification. This is why for iSPY we decided to use body recognition instead of body detection for our system.

3.8 Live Video Compression and Decompression

The first thing to discuss is the means of compressing the video to bitstream so that we can properly transfer the data over via the laser. We need to make sure that the live video feed is transmitted efficiently and retains a level of quality that is acceptable to our standards. There are a number of ways of doing this. The most promising option is using a codec to compress the video.

3.8.1 Codec

A codec stands for both encoder and decoder. A codec can be used to compress a large amount of data. They are usually used to compress media to a smaller format for either storage or more efficient transfer of the file. A codec will allow us to compress our data so that we are not wasting a large amount of time and resources transferring the original data. Even further, we could use a predictive codec which uses an algorithm to turn data into bytes. This will help give us a much more efficient way of transmitting the data via bit [98].

3.8.2 Lossy vs Lossless

There are two kinds of compression: lossy and lossless. Lossy compression is able to compress the data by removing some of the original data. Lossless compression compresses data by “removing unnecessary metadata” [99]. Lossy compression is used when a file doesn’t need all the data or if there is a need to have storage freed up. It would basically get rid of things that are not important to the average user viewing the feed. Lossless compression keeps the original quality of the image but doesn’t reduce the size as much.

3.8.2.1 Lossy Compression

Let’s compare the two types of compression. Lossy compression reduces the size of the files greatly at the cost of reducing quality. The bandwidth of data transfer would be lower due to the reduced size of the file. However, because it reduces the size of the files by getting rid of unimportant looking areas, the data lost from the compression is permanently lost and cannot be salvaged. Lossy compression is typically used when it comes to web-based compression but that won’t be an issue for our project. Lossy compression also might cause blurring or pixelation on our feed [100].

3.8.2.2 Lossless Compression

Lossless compression might seem like the best way to move forward but there are a few aspects to consider. Lossless compression will maintain the original quality of the image. This results in no data loss. We can compress the data and return it to its original format without any data loss. Since we are building a surveillance system, quality is a big aspect

to consider when choosing our codec. We wouldn't want to not be able to detect a person in the live feed due to quality loss from compression [100].

3.8.3 Video Compression Techniques

There are many different ways of compressing a video feed. It's important to consider all of them when planning out a project. We'll be looking at a few methods on how to compress a video and what kind of process it takes. We'll be looking at spatial compression, temporal compression, and entropy coding.

3.8.3.1 Spatial Compression

Spatial compression (intraframe compression) is a method of compressing data by reducing "...the size of each individual frame independently by reducing redundant information within the frame..." [101]. This allows us to reduce the size of the data while maintaining a level of quality to our image. Let's say there is an image of a field of flowers of all different kinds of colors and a clear blue sky. This technique will look at the pixels of the blue sky and determine if they can merge some of the pixels into a larger area. This will require less information to store as it's just a large blue area. However, a field of flowers of all different colors will have a lot of variation. In this scenario, compression would hurt the quality of the image by merging colors together. An example of how spatial compression works can be seen below in Figure 3.10.

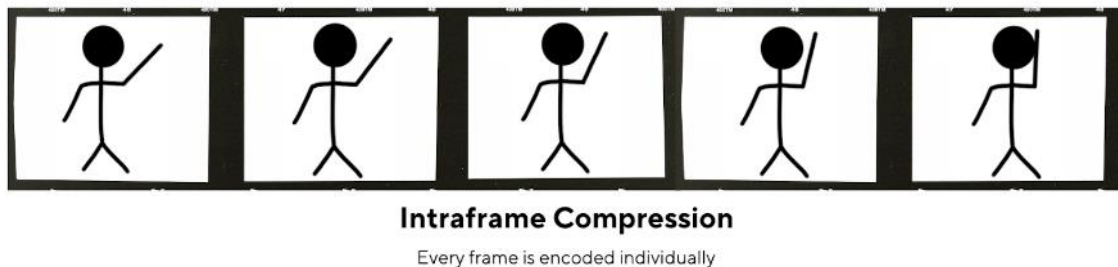


Figure 3.10 – Image Depicting Spatial Compression (Intraframe Compression) [101]

3.8.3.2 Temporal Compression

Temporal compression (interframe compression) is a method of compression that starts with a single image. This image acts as a starting point of reference for the coming frames. It saves the image and only saves changes that occur after the starting point. This starting image is called the "I-frame". The next frames are classified as P-frames (Predictive frames) and B-frames (Bi-directional frames). P-frames store data based on changes in previous frames. B-frames store data based on changes in both previous and coming frames. This helps in determining what information is important or not to make the least number of changes possible [101]. This method compresses data into a much smaller size than spatial compression but it has potential to lose the quality of the original video. An example of how temporal compression works can be seen below in Figure 3.11.



Interframe Compression

Only the differences between frames are encoded

Figure 3.11 Image Depicting Temporal Compression (Interframe Compression) [101]

3.8.3.3 Entropy Coding

Entropy coding is the process of data compression based on statistics. One example of entropy coding is Huffman coding. Huffman coding basically takes frequently occurring symbols (symbols could letters, numbers, words, or in our case pixels) and assigns a small bit value to the symbols. In contrast, the symbols that occur less are assigned longer bit lengths [GPTA]. A good example to explain Huffman coding is an example from Prateek Joshi who wrote an article about entropy coding [102]. Let's say we have a sequence of "SSCAAACCKKKKAKKKKKKKKCCCAKKKK". It takes 5 bits to represent every letter of the alphabet and the length of the sequence is 30. It would take $30 \times 5 = 150$ bits to display the sequence given. Huffman coding allows us to shorten the sequence by quite a margin. S only occurs 2 times which makes it the least occurring letter in the sequence [102]. We can do this for every letter in the sequence and determine its frequency. The assigned bits are shown in the table below (Figure 3.12).

S = 000 (3 bits)

A = 001 (3 bits)

C = 01 (2 bits)

K = 1 (1 bit)

Figure 3.12 Assigned Bit Lengths of Letters in Sequence [102]

We assign the longest bit length to the symbol that occurs least, and we assign the shortest bit length to the symbol that occurs most. Now we can show the sequence using only 51 bits ($3 \times 2 + 3 \times 5 + 2 \times 7 + 1 \times 16 = 51$) [102]. That is nearly a third of the size of what we were initially going to transmit. We also kept the original data intact by doing this method. Huffman coding is also compressed into the form of bitstream which allows us to transmit the data via bitstream with our FSO system.

Arithmetic coding is another example of entropy coding. Arithmetic coding can be more efficient than Huffman coding because it encodes "entire sequences of symbols as a single number (usually a fraction)" [GPTA]. Arithmetic coding looks at the individual frames of

the video and predicts the probability of different symbols. Huffman coding assigns bit lengths to each symbol based on how frequently it shows up. However, "...arithmetic coding represents the entire sequence of symbols as a fractional number within a range based on their probabilities" [GPTA]. This will result in a better compression to bitstream. Arithmetic coding requires more computational power which could result in issues later on.

3.8.4 Live Transmission Limitations

There are two main factors to consider when talking about the live transmission of data. The first factor is bandwidth. In our case, bandwidth is the maximum amount of data we can transfer over a set amount of time [103]. Lower bandwidth will give us less room to work and may require us to compress the video feed further, which would also reduce the quality of the original feed. Higher bandwidth gives us the freedom to transmit higher resolution imaging and avoiding delays. We will have to choose between methods of compression to best suit our needs according to the bandwidth.

The second factor is latency. The entire process of compression and decompression will take time and is something that should be considered when dealing with a live transmission. Compression of more data will decrease the bandwidth required but will increase latency due to the amount of time it will take to compress more data. Inversely, less compression will result in a higher bandwidth requirement but lower the latency [GPTB]. Propagation delay is also a factor in latency. The laser will take a certain amount of time to reach the receiver.

3.8.5 Codec Comparison

We've gone over what a codec is and the different kinds of methods when it comes to video compression in codecs. Now we will discuss the different kinds of codecs developed by different companies. There are a wide range of codecs to choose from when it comes to video compression. We'll be going over the most popular and practical ones that we can use in our system.

3.8.5.1 H.264

One of the most popular codecs is H.264. It allows for streaming at higher qualities while maintaining a lower bitrate. It is a lossy codec which could result in permanent data loss of the original image. While the data loss could be mostly unnoticeable, it is something to consider when comparing it to other codecs. However, there is a lossless version of H.264. Due to H.264 and its popularity, it is widely available and compatible with many devices [104].

3.8.5.2 H.265

H.265 is the successor to H.264 (there is also a lossless version of H.265). It compresses data at about half the bitrate of H.264. The stream should support much higher resolutions

due to the decreased bandwidth and has better quality over H.264. However, it is a newer model and is not as widely used as H.264. This could cause run ins with compatibility issues. H.265 also requires much more computational power because of the higher complexity of the process [104].

3.8.5.3 VP8/VP9

VP8 is a lossy codec designed by On2 Technologies and now Google owns it. Google later improved upon it and developed VP9 (also a lossy codec). VP9 is good for web streaming and is fairly good at compressing and transmitting data with lower bitrates. The codec is also open source which allows for easy access [104].

3.8.5.4 AV1

AV1 is a lossy codec developed by Alliance of Open Media (AOMedia). It's main purpose is to stream quality video over an internet connection. This codec is open source and supports developments in technology like higher resolution or even virtual reality. However, AV1 is not as widely used as other codecs. It also requires more computational power compared to other codecs [104].

3.8.5.5 FFV1

FFV1 is a lossless codec developed by Michael Niedermayer. It is popular for both its speed and size. The codec is mainly used for archival or editing purposes and was acknowledged to be a “viable archival encoding” by the U.S. Library of Congress. The codec also utilizes entropy coding and is compatible with a wide variety of devices and software [105].

Table 3.3 Comparison table of different codecs.

Parameter	Codecs				
Brand	H.264	H.265	VP8/9	AV1	FFV1
Type	Lossy/Lossless	Lossy/Lossless	Lossy	Lossy	Lossless
Efficiency	High	Very High	High	Very High	High
Pros	Widely used and compatible with most devices	Better version of H.264	Open source and efficient on web	Open source and very efficient	No data loss during compression, open source, efficient
Cons	Not the best for higher resolutions	More computational power and not as widely used	Not as widely used	Requires more computational power	Not typically used for live transmission

3.8.6 Transmission of Data

The codec transforms the live video feed into a compact bitstream format. This bitstream format is exactly what we need to transmit the data via the FSO system. There are a few ways we could approach the issue of transmission. Pulse code modulation (PCM) is when an analog signal is converted into a binary sequence. In the process of demodulation, we get the original analog signal [106]. However, since we have already converted the original data into bitstream via a codec, we can transmit the data directly using PCM.

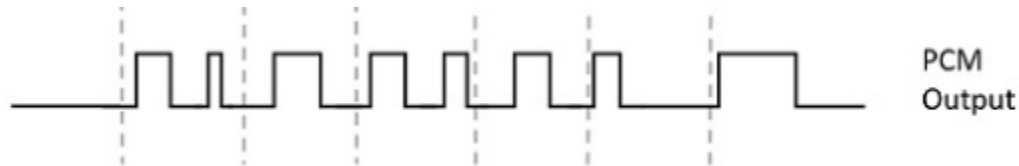


Figure 3.13 Demonstration of PCM Output [106]

The laser can pulse in intervals to indicate different binary digits to properly decompress the transmission. As seen in Figure 3.13 above, there are varying lengths of pulses of which could be different values being sent based on the length of the pulse [107]. When the laser is off, the binary value will be set to 0. When the laser is on, the binary value will be set to 1.

Amplitude modulation (AM) can act in a similar way to PCM. We can amplify the laser to different levels of intensity to indicate a change in data. We could also accomplish this with frequency modulation (FM). While not necessarily the best option, we could use phase modulation (PM) as well [108] [GPTE].

3.8.7 Loss of Data

What do we do in the scenario that some data is lost during transmission? There are a few ways we can solve that. Lost data is bound to happen from time to time due to interference with the laser. This could be from a bird flying through the laser or dust blocking it for a moment. The first way to solve this problem is to simply store records of the recorded footage so that, if needed, the footage can be viewed at a later time if something was missed.

Another way to solve this problem is with error detection. Reed-Solomon codes are one of the ways to solve this problem. Reed-Solomon codes were developed by Irving S. Reed and Gustave Solomon. How it works is that it takes a data block and adds extra parity bits (redundant bits). When the transmission has been received, the decoder checks to see if any bits have been corrupted and replaces them with the parity bits [109].

Hamming code is another code that is used for bit errors and correcting these errors. It was developed by R.W. Hamming. It is like Reed-Solomon code in that it inserts parity bits (redundant bits) before transmission. It then performs recalculations to detect errors and find out the position of said errors [110].

3.8.8 Video Compression Libraries

Libraries are very helpful when it comes to any kind of code. There are some things that are extremely complex when it comes to individual coding. For example, writing a codec on your own is very complex and time consuming. This is why we'll be looking at the different libraries when it comes to video compression in three potential languages.

3.8.8.1 C++

C++ has a few different libraries that allow us to utilize codecs. The first library is FFmpeg (libavcodec). This library supports all kinds of codecs. It includes H.264, VP8, VP9, AAC, MP3, and a lot more. It is widely used and documented well which will allow for easy understanding. It is also very optimized to include all of these different codecs [111] [GPTC].

The second library is GStreamer. GStreamer is much more complex and allows different kinds of codecs to act as plugins. While it is complex, it allows for customization that other libraries do not offer [112] [GPTC].

The third library is OpenH264. This codec only focuses on H.264 but since it only focuses on H.264, it is quite efficient. This library is useful for real-time video compression. The library is also open source which allows anyone to use H.264 for free without any licensing costs. The only drawback to this library is that it only uses H.264 and does not support other formats [113] [GPTC].

Table 3.4 Comparison table of different libraries for C++

Parameter	Libraries		
Library	FFmpeg (libavcodec).	GStreamer	OpenH264
Performance	High	High	High
Supports Multiple Codecs?	Yes	Yes	No
Well documented?	Yes	Yes	Yes
Optimized For Live Feed?	Yes	Yes	Yes

3.8.8.2 Java

Java has its own libraries as well. The first library is called JCodec. It is purely written in Java and supports codecs like H.264 and VP8. Unfortunately, it has limited codec support and lack optimization for real-time applications [GPTC]. The developers of the program have acknowledged that since the library is all in java, it will be “an order of magnitude slower than native implementations (such as FFMpeg)” [114].

Xuggle (Xuggler) is a library in java that allows FFmpeg to be implemented into java. Because it is based on FFmpeg, it supports most codecs and is able to handle encoding and decoding well [GPTC]. Xuggler is open source and has cross platform support. It allows for flexibility and community support due to the project being open source [115].

The third library for video compression in java is JMF (Java Media Framework) [116]. It is a much older library that has support for a few codecs but is no longer being worked on [GPTC]. Due to the outdated nature of the library, I think it's safe to say there are better options.

Table 3.5 Comparison table of different libraries for java

Parameter	Libraries		
Library	JCodec	Xuggle (Xuggler)	JMF (Java Media Framework)
High Performance	Medium	High	Low
Supports Multiple Codecs?	Yes (but limited)	Yes	No
Well documented?	Yes	Yes	No
Optimized For Live Feed?	No	Yes	No

3.8.8.3 Python

PyAV is the first library we're going to discuss that is used for video compression [GPTC]. According to the official website of PyAV, "PyAV is a Pythonic binding for FFmpeg. We aim to provide all the power and control of the underlying library but manage the gritty details as much as possible" [117]. PyAV can be used to manipulate data via specific processes like codecs, containers, or streams. The developers do warn that PyAV can be a hindrance at times, they also claim that it is a "critical tool" when it must be used [117].

The second library that we're going to compare is MoviePy. MoviePy is different in that it allows many different functions when it comes to video editing. Some of these features include "...cutting, concatenations, title insertions, video compositing (a.k.a. non-linear editing), video processing, and creation of custom effects" [118]. While it can handle codec related tasks, the primary function of this library is to edit videos [GPTC].

The third library used in Python for video compression is FFmpeg-Python. This is essentially a "Python wrapper for FFmpeg" [GPTC]. This essentially allows the user to fully use FFmpeg commands and associated codecs while using python. It is well documented as the library comes from the same FFmpeg website as the C++ FFmpeg library [111].

The common denominator between all three of these coding languages and the libraries they use seems to be that they use FFmpeg. After researching several different libraries across several languages, FFmpeg keeps popping up as a reliable and efficient library to use for our video compression needs.

Table 3.6 Comparison table of different libraries for python.

Parameter	Libraries		
Library	PyAV	MoviePy	FFmpeg-Python
Performance	High	High	High
Supports Multiple Codecs?	Yes	No (mainly a video editing tool)	Yes
Well documented?	Yes	Yes	Yes
Optimized For Live Feed?	Yes	No (mainly a video editing tool)	Yes

3.8.9 MCU Selection

We now know that certain codecs require more computing power than others. We also know that computer vision requires processing power as well. One of the most important decisions for this project is the MCU selection. It is what allows us to run our code efficiently and quickly. ChatGPT was asked about three potential MCUs that could work well with our needs [GPTD]. The first MCU provided was an ESP32-S3. The second MCU provided was an STM32H7 Series. The final MCU listed was a Raspberry Pi Pico W (RP2040 with Wireless). I did have the initial idea to use the MSP boards from previous classes at UCF but after research, it was found that they are not powerful enough for our needs.

The first MCU to consider is the ESP32-S3 from Espressif. It is an “Xtensa 32-bit LX7 dual-core processor that operates at up to 240 MHz” [119]. It contains 512 KB of SRAM. It does well with neural network computing which will be helpful with the computer vision aspect of the project. The chip also handles signal processing well. The storage temperature of this chip is between -40 °C and 150 °C [120]. It costs around \$1.85 at the current time of writing this [121].

The second MCU to consider is the STM32H7 Series. It is a “...32-bit Arm Cortex®-M7 core, running at up to 600 MHz” [122]. It can come in both single and dual core (Cortex®-M7 + Cortex®-M4). It offers flash memory between 64 KB and 2 MB. The chip comes with between 564 KB to 1.4 MB of SRAM. The storage temperature is between -40 °C and 125 °C [122]. I found chips from this series to be ranging from 7 to 10 dollars [123].

The third MCU is a Raspberry Pi Pico W (RP2040 with Wireless). It is a “Dual-core Arm Cortex-M0+ processor, flexible clock running up to 133 MHz” [124]. There are 264 KB of SRAM on board. It’s storage temperature is between -40 °C and 85 °C. On CanaKit, this cost around \$1.00 [125]. One thing to note about raspberry pi is that it known for its modularity when it comes to programming. We were also told by our advisors that raspberry pi boards are powerful.

After careful consideration, we decided to go with the brand raspberry pi. Raspberry pi was recommended by our advisors. It is known for its flexibility with programming. In addition, it’s very easy to find components that fit well with a Raspberry Pi. This includes things like the sensor for our project.

Table 3.7 Comparison table for different brands of MCUs

Parameter	MCUs		
Brand	ESP32-S3	STM32H7 Series	Raspberry Pi Pico W
Clock	240 MHz	600 MHz	133 MHz
SRAM	512 KB	564 KB to 1.4 MB	264 KB
Flash Memory	N/A	64 KB to 2 MB	N/A
Storage Temp	-40 °C to 150 °C	-40 °C to 125 °C	-40 °C to 85 °C
Cost	\$1.85	\$7 to \$10	\$1

We know that we want to use Raspberry Pi as the MCU for our project. Now we need to decide which raspberry pi to use that best suit our needs. Once again, ChatGPT was asked to provide potential raspberry pi boards that could fit our needs [GPTD]. The first raspberry pi board was a Raspberry Pi 4 Model B. The second raspberry pi board was a Raspberry Pi Zero 2 W. The third board suggested was a Raspberry Pi 3 Model B+.

The Raspberry Pi 4 Model B is a “Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.8GHz” [126]. It has 1 GB, 2 GB, 4 GB, or 8 GB SDRAM depending on which model we choose. This board comes with H.265 decoding (at 4k 60fps) and H.264 encoding (1080p at 60fps) and decoding (1080p at 30fps) which is perfect for our needs. It comes with a variety of ports like USB ports, a USB-C connector, micro-HDMI ports and more. It has gigabit ethernet. The operating temperature is 0 °C to 50 °C [127]. The cost varies depending on the amount of on-board RAM you decide to get. It ranges from \$35 to \$75 on CanaKit [128].

The Raspberry Pi Zero 2 W is a “1GHz quad-core 64-bit Arm Cortex-A53 CPU” [129]. It comes with 512 MB of SDRAM. It comes with similar ports as the Raspberry Pi 4 Model B but since it is smaller, it has less ports. It comes with both H.264 encoding and decoding (both 1080p 30fps). The operating temperature is -20 °C to 70 °C [130]. The board cost on CanaKit is \$15 [131].

The Raspberry Pi 3 Model B+ is a “Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit SoC @ 1.4GHz” [132]. It comes with 1 GB of LPDDR2 SDRAM. It comes with H.264 decoding and encoding (both 1080p 30fps). It has gigabit ethernet but a maximum throughput of 300 Mbps. The operating temperature is 0 °C to 50 °C [133]. It was unavailable at CanKit but is \$35 dollars on Chicago Electronic Distributors [134].

All of these boards have camera connectors and H.264 encoding and decoding capabilities. However, The Raspberry Pi 4 Model B is a bit above the rest in terms of performance. It has gigabit ethernet and has more RAM options. It is the same price as the Raspberry Pi 3 Model B+ while having higher specs.

Table 3.8 Comparison table for different models of Raspberry Pi boards

Parameter	MCUs		
Model	Raspberry Pi 4 Model B	Raspberry Pi Zero 2 W	Raspberry Pi 3 Model B+
Clock	1.8 GHz	1 GHz	1.4 GHz
SDRAM	1 GB, 2 GB, 4 GB, or 8 GB	512 MB	1 GB of LPDDR2
Codec	H.265 decoding (at 4k 60fps) and H.264 encoding (1080p at 60fps) and decoding (1080p at 30fps)	H.264 encoding and decoding (both 1080p 30fps)	H.264 decoding and encoding (both 1080p 30fps).
Ethernet	Gigabit	N/A	Gigabit (300 Mbps throughput)
Operating Temp	0 °C to 50 °C	-20 °C to 70 °C	0 °C to 50 °C
Cost	\$35 to \$75 (depending on RAM)	\$15	\$35

3.9 Communication Sub-System

Signal communication is a critical component for the project. Without clear signals communication, the recorded data will not be displayed. Understanding what communication is, the different types of communication (i.e. interpersonal, electrical, and optical), and how to achieve communication is necessary. The following sections will explore these questions and highlight the decision-making for the communication sub-system for the project.

3.9.1 Communication

Communication is the form of relaying messages to people or groups. The relayed messages are the ideas conveyed by a sender, the communicator, to a receiver, the participant. The process for sending messages is known as transmission. This unique exchange of information is transmitted through different channels. A human channel uses verbal and non-verbal communication – verbal involves the exchange of spoken and written information while non-verbal happens without the exchange of spoken or written information. Technological advancements in analog and digital electrical systems have enabled new and distinct channels for communication, like copper cable communications, radio communications, fiber optic communications, and free-space optic communications. These avenues will be discussed in succeeding sections. Regardless of the channel of choice, the agreed roles between the sender and receiver dictates the style of communication to be one-way or two-way.

3.9.1.1 One-way Communications

One-way communication, also known as linear or non-transactional communication, is the form of transmission in which information flows in a single, linear direction from sender to receiver [135]. An example is shown in fig. 3.14

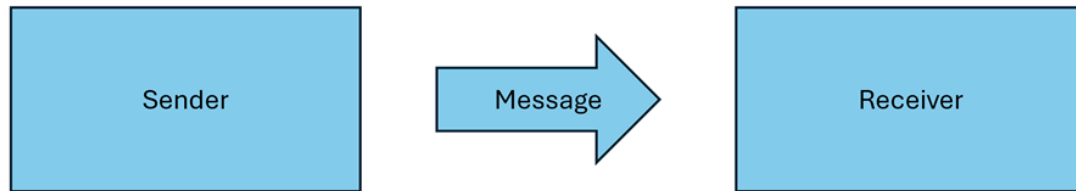


Figure 3.14 Linearity of one-way communication. Fig. created by Authors.

As seen on fig. 1, the sender controls the message by determining what information is shared, how it is presented, and when it is shared; the receiver's role is passive and has little influence over the communication process [136]. The sender transmits information without any feedback or response from the receiver, making the event non-transactional [137]. The limited interaction between sender and receiver severely restricts engagements, understanding, and trust; additionally, information is transmitted without consideration of the receivers' feedback [137]. One-way communication is critical for information dissemination, instructional purposes, persuasive influence, and entertainment like radio and motion pictures [135]. One-way communication is widely applicable to technologies like one-way video surveillance systems and notifications. Perfect examples include baby monitors and natural disaster alerts. In the case of a baby monitor, the sender is the baby while the receiver are the parents. The baby's actions dictate what the parents observe or hear giving them the necessary information to intervene on the baby or not. For natural disaster warnings, one-way communication enables the transmission of important information to the public with the goal of rapid dissemination. In this case, the sender is law enforcement, and the receiver is the publics.

3.9.1.2 Two-way Communications

Two-way communication is the form of transmission in which information is exchanged by both sender and receiver directly or indirectly [138]. An example is shown in fig. 3.15.

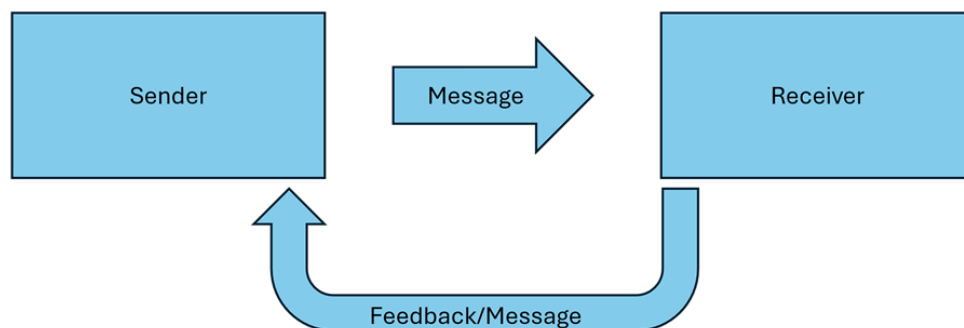


Figure 3.15 Direct (message) and non-direct (feedback) two-way communication. Fig. created by Authors.

As shown in fig. 3.15, in this style of communication, information flows first from sender to receiver then receiver to sender, creating a feedback loop between the sender and receiver; unlike one-way communication, two-way communication encourages interaction and transactions amongst sender and receiver [139]. The highly interactive nature of two-way communication, engagements, understanding, and trust is well established amongst sender and receiver. Two-way communication is the way humans communicate. Naturally, two-way communication is widely applicable to technologies that facilitate interpersonal interactions amongst people like two-way radio systems, telephones, and social media platforms. Two-way radios enable law enforcement officers and security teams to immediate back-and-forth communication; in this case, the communication channel is a low-frequency electromagnetic wave in air. Telephones operate under the same principle as two-way radio. Social media platforms enable users to communicate and interact through posts, comments, and likes; this style of two-way communication takes place different servers (one for the sender and one for the receiver) connected via fiber optic cables transmitting information using light pulses.

3.9.1.3 One-way Communications over Two-way Communications

This project intends to build a one-way communications surveillance system because receiver feedback is not necessary. The intended functionality of a surveillance system is to transmit information from the sender, the immediate environment, to the receiver, the end user, for their discretion. Two-way surveillance systems would complicate the scope of the project because it would involve the receiver providing the sender with some information. When analyzing the stretch goals and applications for the project, two-way communication is not necessary for military, first response, or agricultural surveillance. For example, a two-way channel for military surveillance could compromise the surveillance effort.

3.9.2 Electrical Communications

In communications, a channel is any way information is transmitted. Verbal channels transmit information using voice. Non-verbal channels transmit information using facial expressions, body language, or writing. Technological channels transmit information using analog or digital electrical signals via copper cables. Over the last 150 years, copper has proven to be a reliable conductor for transmitting electrical signals. Copper cables like coaxial cables, twisted pair cables, and ethernet cables have demonstrated copper can be used for short-haul and long-haul electrical communication but with limitations.

Coaxial cables are a type of shielded transmission line designed to conduct high frequency electromagnetic signals [140]. The structure of a coaxial cable consists of a central copper core, a metallic shield, a dielectric insulator, and a plastic jacket [140]. Invented in the 1800s by Oliver Heaviside, coaxial cables offered exceptional low loss and low inter-cable induction interference [140], [141]. The invention of the coaxial cable revolutionized the communications landscape of the past and continues excelling communications

capabilities of today. Coaxial cables have been so reliable over the years their design hasn't changed much [142]. Some ways coaxial cables have enabled communication include television broadcasting, long-distance phone service, and internet infrastructure.

Data transmitted through coaxial cables are typically in the form of an electromagnetic radio frequency (RF) wave. RF signals are analog or digital electrical signals that oscillate at frequencies from 300 GHz to as low as 9 kHz, the frequencies associated with radio electromagnetic radiation [143]. Radio signals are typically associated with free-space propagation. However, radio electromagnetic disturbances radiate through shielded copper conductors, like a coaxial cable. The shield coating of coaxial cables is what maintains the RF signals inside the cable [144]. To create a distinction, telecommunication services refer to free-space RF signals as wireless RF and cabled RF signals as cabled RF [144]. RF signals originate at a transmitter and are picked up by a receiver.

3.9.3 Optical Communications

Electrical communication offered a good solution for many avenues of computer communication. Research and engineering efforts during the mid-1900s demonstrated the limitations of electrical communication and introduced and proliferated optical communication into the telecommunication industry. Optical communication has enabled today's world internet connectivity by means of long-haul optical fiber networks and data center inter-server connectivity. In this section, optical communication methods relevant to the scope of the project are discussed.

3.9.3.1 Fiber vs. Free-Space Communications

Over the last 30 years, the proliferation of internet traffic and the rapid development of the communication industry has optical communication a cornerstone strategic industry to meet the world's connectivity needs [145].

3.9.3.2 Optical Transceivers

Optical modules, also known as optical transceivers or fiber optic transceivers, are the building blocks for all network connectivity inside and outside of data centers [145]. Optical transceivers have revolutionized the communication industry with high bandwidth communication capabilities. Some optical transceivers offer optical connectivity bandwidths, also known as data rates, up to 100 gigabits per second (Gbps) over link distances (transmission distances) of 10 km and beyond [146].

A transceiver is a modular hot-pluggable interconnect device that facilitates bi-directional optical communications using various wavelengths over different types of physical media [145], [147]. In general, optical transceivers are electrical-to-optical and optical-to-electrical converters – they have an electrical interface on the side that connects to the electronic system, like a PCB plug-in connection, and an optical interface that connects to

the optical system, like a single mode or multimode optical fiber [145], [146]. Transceivers look like illustrated below in fig. 3.16A.

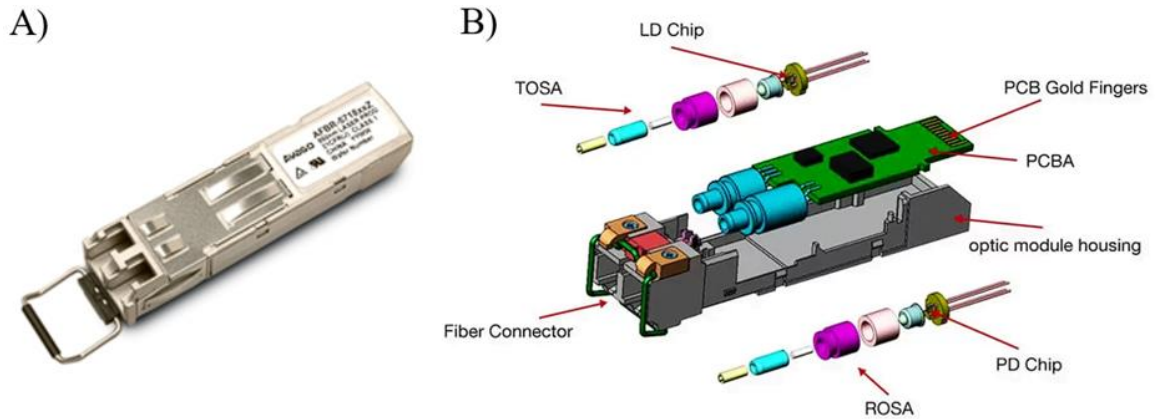


Figure 3.16 An optical transceiver and its breakdown. A) Image of optical module AFBR-5718PZ sold by Broadcom Inc. Fig. taken from Ref. [148]. B) General makeup of an optical transceiver. Fig. taken from Ref. [149].

Optical transceivers have the functionality of both a transmitter and receiver, hence the name transceiver – data is transmitted via optical pulses from a laser diode or light-emitting diode and received using a photodiode [149], [150]. Laser diodes are more commonly used across the telecommunications industry and the reasoning will be explained in a later section. Other components that make up a transceiver include integrated functional circuits, main control circuit board, housing, and optical and electrical interfaces [149]. A breakdown of all the components together is shown in fig. 3.16B.

The transmission side of the transceiver is usually housed in what is called a transmitter optical sub-assembly, or TOSA; the TOSA is responsible for converting electrical signals into optical signals [149]. When the TOSA's internal driver chip receives an input electrical signal, it drives the laser diode to emit an equally modulated optical signal equivalent to the input electrical signal [149]. Optical fiber modulation formats will be discussed in a later section. The receiver side of the transceiver is usually housed in what is a receiving optical sub-assembly, or ROSA; the ROSA is the essential component of any optical communication system because it's responsible for converting optical signals transmitted by the TOSA into usable electrical signals [149]. The work of the ROSA is mostly carried out by a photodiode – an optoelectronic device that is photosensitive and generates a current or voltage signal. When a photodiode is used in tandem with a trans-impedance amplifier and post-amplifier, the incident optical pulses then become a usable digital signal [149]. Both TOSA and ROSA systems integrate optical components, like lenses, necessary to couple light into an optical fiber (TOSA) and focus an optical fiber output to the photodiode (ROSA).

Today, telecommunication connectivity is facilitated by an optical fiber network and a pair of transceivers [151]. Consider a one-way communication computer connection, like shown in fig. 3.17.

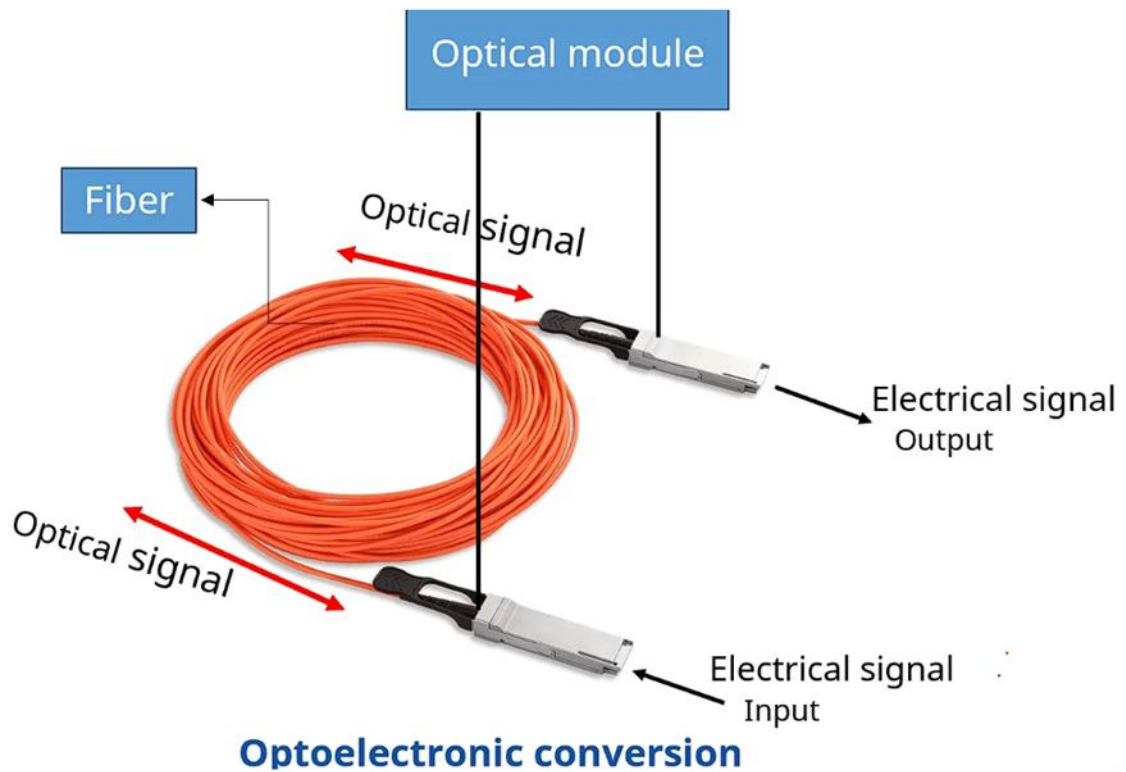


Figure 3.17 An optical fiber interfaces with an optical transceiver. Fig. taken from Ref. [149].

To simplify this picture, assume you have a pair of computers – computer-1 is transmitting information via a fiber optic channel to computer-2. Each computer has a transceiver to transmit and receive information but only computer-1 is transmitting and only computer-2 is receiving. When Computer-1 is ready to transmit a signal, it generates an electrical signal on the electrical interface of its transceiver to drive the laser diode and generate optical pulses that represent the electrical signal – this is electro-to-optical conversion [149], [151]. This optical signal travels down the length of the optical fiber onto the transceiver on computer-2. The received optical signal induces an electrical response on the photodiode generating an electrical signal for computer-2 to interpret – this is opto-to-electrical conversion [149], [151]. This happens millions of times between data center server modules. Data centers are famous for their extreme number of connections and cable links between server modules. Server modules have tens of transceivers working independently of one another and are in bi-directional communication with another server module since transceivers offer that capability.

Transceivers are differentiated across a plethora of different classification methods. These classification methods often describe a specific working parameter of interest. The most common classification methods include form factor, data (transmission) rate, wavelength, light source, fiber type, transmission distance, modulation format, communication mode, receiver type, connector type and pluggability [149].

Classification by form factor, also known as package type, is by far the most popular because it determines the additional equipment to be used; for example, Broadcom Inc. lists optical transceiver products by form factor on their website [146]. This method of classification pertains to the size, shape, and overall dimensions of the optical module [149]. As is discussed in a later section, standards and regulations, the dimensions of all optical transceivers are closely monitored by a standards committee and a multisource agreement [149], [152]. Some package type examples include small form-factor pluggables (SFPs), enhanced-SFP (eSFP), SFP-plus (SFP+), SFP28, X form-factor pluggable (XFP), quad small form-factor pluggable (QSFP), C form-factor pluggable (CFP), QSFP28, and QSFP-DualDensity (QSFP-DD) [152]. A few of these are shown below on fig. 3.18.



Figure 3.18 Examples of different form-factor modules. A) SFP+ module. B) QSFP+ module. C) QSFP-DD module. Fig. created by Authors of images collected from Ref. [149].

Classification by data rate is almost equally as popular as form factor. Transceiver transmission rates have only increased since the release of the gigabit interface converter (GBIC), the first small (relatively large compared to SFP) form-factor transceiver capable of reaching speeds of 1 Gbps for ethernet and up to 4 Gbps for an optical fiber channel, in 1995 [149], [152]. The optical modules that flood today's market provide optical connectivity at rates up to 100 Gbps and beyond [146], [152]. The first big jump in transceiver transmission rates occurred in 2006 with the release of SFP+, the still dominant industry format; this form factor enabled minimal rates of 8 Gbps and up to 16 Gbps [152]. In 2014, QSFP28 became the standard interface of choice for high-speed, large-volume data transmission; this form factor had four transmission lanes operating at 25 Gbps with slightly larger physical dimensions than an SFP [152]. The fastest achievable data rates per transceiver now reach 800 Gbps with a maximum of 400 Gbps per transmitter-receiver pair; this became available in 2019 with the release of QSP-DD transceiver [146], [152].

The next two classifications go hand in hand – wavelength and light source. As discussed earlier, the light source housed within an optical transceiver is critical for a transceiver to transmit data and accounts for a large sum of the cost to buy a transceiver [153]. In general, the light sources used for optical fiber communication are either laser diodes or light-emitting diodes (LEDs), both made from semiconductor material. A laser diode is any direct bandgap semiconductor material that amplifies generated or input light by means of stimulated emission of radiation (photons). An LED is any direct bandgap semiconductor material that generates light by means of spontaneous emission of radiation. Both processes generate light from the recombination of electron-hole pairs in the semiconductor material. Regardless, what also differentiates a laser diode from an LED is the concept of optical feedback and optical coherence. Optical feedback is a laser process that utilizes generated

light to generate more light. All lasers, whether semiconductor or solid-state, achieve optical feedback by means of a reflective optical cavity; these cavities have a type of reflective material that feeds some optical energy back into the gain material while letting the rest leave the system. Optical coherence refers to the phase difference between the electromagnetic waves that make up the generated light. LEDs generate incoherent light; this type of light spreads away (diverge) from a straight path making LED light non-directional. Lasers generate highly coherent light; this type of light is tightly confined and directional enabling photons to travel together without spread over long distances. Besides, laser diodes and LEDs emit narrow and relatively broad, respectively, spectrums of light centered about a wavelength. The central emission wavelength is determined by the semiconductor material bandgap energy since the emitted photon energy is equivalent to the bandgap energy following an electron-hole recombination event. The light sources that make up the transmitted of optical modules include Fabry-Perot lasers (FPL), distributed feedback lasers (DFB), vertical-cavity surface-emitting lasers (VCSEL), LEDs, and electro-absorption modulated lasers (EAM) [149]. Which light source is used depends on the link communication link length, transmission rate, and optical fiber being used (single mode vs. multimode optical fibers) [149], [153]. Unlike VCSELs, like shown in fig. 3.19, FPLs and DFBs are edge-emitting semiconductor lasers, also known as in-plane lasers [153].

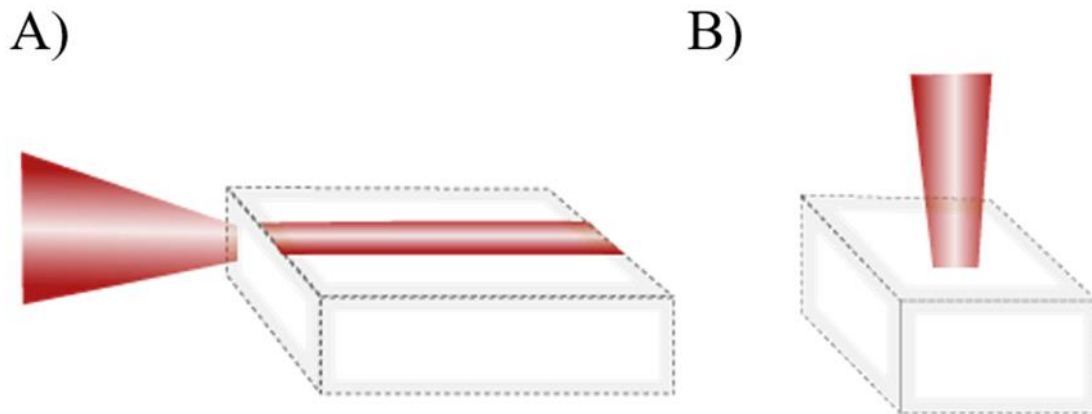


Figure 3.19 A) Edge-emitting semiconductor laser. B) Vertical-cavity surface-emitting laser. Fig. created by Authors of images collected from Ref. [153].

EAMs are a different type of semiconductor laser diode in which it has a built-in modulator [153]. The most common transceiver central wavelengths are 850 nm, 1310 nm, and 1550 nm. These wavelengths are fundamental wavelengths in the optical communications industry because of the low material loss associated with these wavelengths on silica glass, the material that makes up optical fibers. From these three wavelengths, only VCSELs are used to generate 850 nm light [153]. Additionally, LEDs were used in the early years of optical communications; however, the invention and proliferation of VCSELs has taken over the 850 nm industry that LEDs once dominated [154]. Today, FPLs and DFBs are found in most optical transceivers because of their narrow spectrums, low power demands, and highly coherent nature [149], [153]. The coherence of lasers and the central wavelength

spread of laser diodes enable transceivers to push transmission capacity and optical fiber limits using modulation methods discussed later.

Transceiver classification by fiber type and transmission distance go hand in hand. Transceivers made to work with SMFs are known as single-mode transceivers and transceivers made to work with MMFs are known as multi-mode transceivers [155]. As discussed earlier, SMFs enable long-haul communication over tens of kilometers whereas MMFs are bandwidth limited for long-haul communications and are, thus, better suited for short-haul communication spanning several meters. The main forces affecting the transmission distance are dispersion and attenuation; dispersion refers to the spread of colors (intra-modal) or the spread between light modes (inter-mode) while attenuation refers to material absorption and scattering [149]. Transceivers are hence optical fiber limited in this classification duality – the transmission distance of a transceiver depends on the fiber of choice. Additionally, the optical fiber of choice also limits the light source of the transceiver. SMFs typically operate around 1310 nm and 1550 nm central wavelengths, commonly known as the O-band (original band) and C-band (conventional band) in telecommunications, and MMFs operate around the 850 nm central wavelength, called 850 nm band for multimode fibers [149], [156]. Since SMFs are manufactured to operate at 1310 nm and 1550 nm, the light sources of choice are FPLs and DFBs while the light sources of choice for a MMF are LEDs and VCSELs [153]. When choosing between single-mode and multi-mode transceivers, the cost of multi-mode transceivers will always be cheaper [155]. The other classification types, except modulation formats, are not discussed further since they're not as popular and are not relevant to the scope of the project.

For transceivers to reach ultra-high transmission rates, like the 400 Gbps transmission speeds of a QSFP-DD transceiver, the transmitted optical signal must be modulated, also known as encoded, with respect to the input electrical signal. There are two modulation methods: intensity modulation, also known as pulse-amplitude modulation, and coherent modulation, also known as quadrature amplitude modulation [157], [158].

Intensity modulation can be achieved using one of two ways: non-return-to-zero modulation (NRZ), also known as pulse-amplitude modulation 2-level, and pulse-amplitude modulation 4-level, also known as PAM4 [158], [159]. NRZ is a two-level modulation technique that encodes digital signals in binary logic, either a one or a zero, as two different intensity levels; put simply, NRZ establishes logic 0 to be a lower intensity than logic 1 [158]. PAM4 is a four-level modulation technique that encodes digital signals in 2-bit logic; put simply, the intensity waveform has four different levels with each level corresponding to one of either logic 00, 01, 10, or 11 [158]. An example of the intensity difference is shown in figure 3.20.

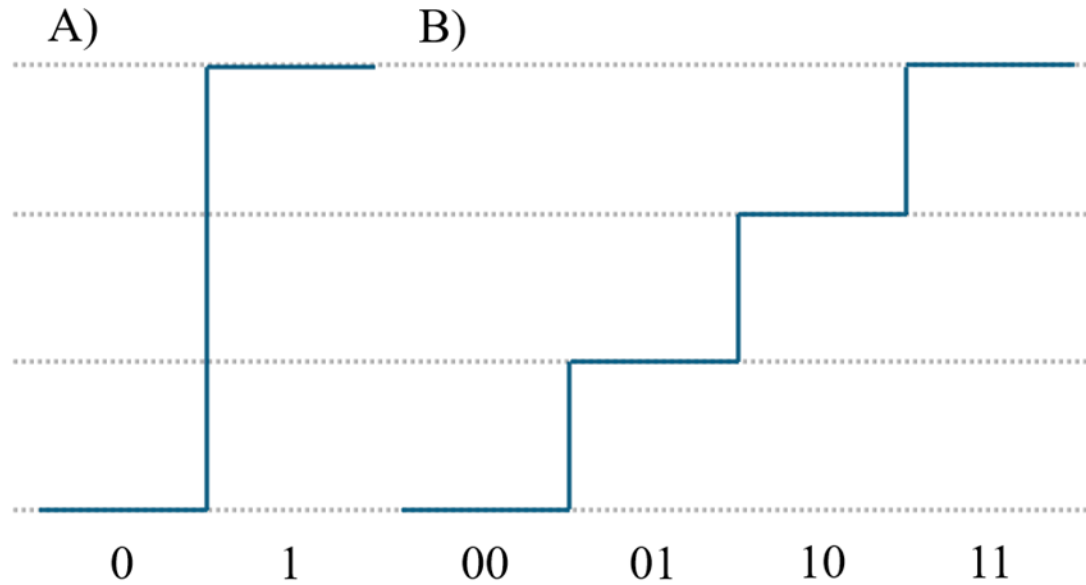


Figure 3.20 Logic difference between NRZ and PAM4. A) Intensity logic of NRZ. B) Intensity logic of PAM4. Fig. created by authors.

Figure 3.20 does a great job at highlighting one of the main differences between NRZ and PAM4: bit rate. NRZ only enables 1 bit per clock cycle; for this reason, the bit rate will always be the same as the baud rate [159]. Whereas PAM4 enables 2 bits per clock cycle, doubling the bit rate for a given baud rate [159]. Without having to change many components, PAM4 doubles the amount of data that is transmitted in each channel by using four levels of intensity. Conveniently, signals loss is tremendously reduced in PAM4 because of the doubled bit rate at the same baud rate [159]. However, PAM4 comes with a signal-to-noise ratio and bit error rate cost. By increasing the encoding levels, the intensity difference between levels decreases; this increases the link budget penalty making PAM4 sensitive to noise resulting in higher bit error rates [158], [159]. In general, a transceiver modulating signals with PAM4 will always be more complex and consume higher power than an NRZ transceiver [158].

Coherent modulation can only be achieved through quadrature amplitude modulation (QAM), a technique that modulates the amplitude and phase of light [160]. QAM amplitude modulation is the same as NRZ or PAM4 – the intensity of the optical pulse is varied [160]. The phase modulation aspect of QAMs is more complicated – each optical pulse waveform has a different phase from each other [160]. Like intensity modulation, multi-level phase modulation can be achieved to create modulated signals like PAM4, but that differ in phase rather than amplitude. An NRZ equivalent modulation in phase is called binary phase shift keying (BPSK) and a PAM4 equivalent modulation in phase is called quadrature phase shift keying (QPSK) – either keying technique falls under the umbrella of phase shift keying digital modulation formats [160]. QAM nomenclature is determined by the evaluation of M , the number of different waveforms possible, where n is the number of bits that make up the transmission symbol (i.e. 01, 0101, 101010, etc.); for example, 16-QAM tells us there are 16 different waveforms for a transmission symbol made up of 4 bits (i.e.

1010) [160]. The modulation levels of QAMs are observed using polar coordinates in a constellation diagram, like shown in figure 3.21 [160].

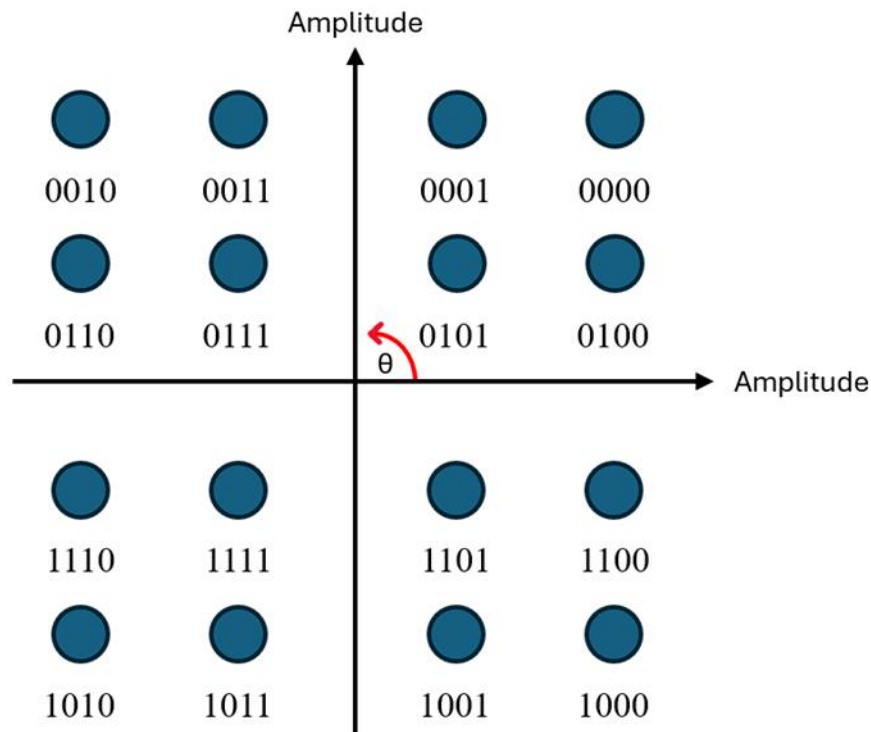


Figure 3.21 Constellation diagram of a 16-QAM modulated signals. Fig. created by authors.

QAMs have become ubiquitous in radio communications; for example, WiFi 5 uses 256-QAM whereas WiFi 6 uses 1024-QAM [160]. In optical communications, QAM has become the method of choice for long-haul data links, like trans-ocean optical fiber networks [161].

Other methods of increasing the data capacity per channel exist (i.e. wavelength-division multiplexing techniques and its derivatives) but are not relevant to the scope of the project.

3.9.3.3 Optical Transceiver Parameters of Interest

To effectively transmit a high-resolution video feed, knowing the transmission rate of a transceiver is important and must be near the order of GHz. The maximum transmitted power is necessary to calculate the associated optical losses across all the mediums in the communication channel. Receiver sensitivity establishes the sensitivity of the project's communication channel; the lower the sensitivity, the lesser the constraints are. Wavelength will determine the optical components used throughout the communication channel. Fiber transmission distance is a parameter found in the data sheets of all transceivers; this parameter serves as a reality check for the system since it gives insight into maximum transmitted power, sensitivity, and type of fiber to use. Fiber type is an important parameter: it gives insight into the type of fiber used during the testing process

of the devices. Form factor constrains the evaluation board used. Cost is necessary to account for the budget.

For the scope of the project, the transceiver parameters of interest, in order of most to least important, include (1) transmission rate, (2) maximum transmitted power, (3) receiver sensitivity, (4) wavelength, (5) fiber transmission distance, (6) fiber type, (7) form factor, and (8) cost. The trade-offs between these parameters determine the cost. The most valued option will be closely considered, and potentially used, for the success of the project.

3.9.3.4 Optical Transceiver Comparisons

In this section, different transceivers from the same manufacturer, across manufacturers, and across fiber types will be compared. First, the differences between form factors of the same brand, Cisco, are compared below.

From the comparison on table 3.9, optical transceivers across form factors generally have similar values for many of their parameters, except transmission rate and cost. As transmission rate increases, the cost increases tremendously. For example, a 10 Gbps increase from 1 Gbps transmission rates doubles the cost; but a 40 Gbps rise in transmission rate increases the cost by a couple hundred dollars. Since the project does not benefit from ultra-fast transmission rates (10 Gbps and above), then SFP form factor transceivers with transmission speeds of 1 Gbps will be looked at more closely in table 3.10.

Table 3.9 Differences across different module form factors of the same brand.

Parameter	Optical Modules			
Mfg. P/N	GLC-LX-SM-RGD	SFP-10G-LR	QSFP-ELR4-40G	QDD-PLR8-800G
Standard	1000Base-LX/LH	10GBase-LR	40GBase-ELR4	800GBase-PLR8
Form Factor	SFP	SFP+	QSFP+	QSFP-DD
Transmission Rate (Gbps)	1.25	10.3125	41.2 (4x 10.3)	850 (8x 106.25)
Tx Power (dBm)	-9.0 - -3.0	-8.2 – 0.5	-4.5 - -2.3	-1.9 – 4.8
Rx Sensitivity (dBm)	< -21	< 14.4	< -12.5	-8.2
Wavelength (nm)	1310	1310	1310	1310
Transmission Distance	10 km	10 km	20	10 km
Fiber Type	SMF	SMF	SMF	SMF
Cost (\$)	10.00	27.00	319	1899.00

Table 3.10 Differences across SMF and MMF SFP modules of different brands.

Parameter	Optical Modules				
Brand	Cisco	Juniper	Arista	Cisco	Juniper
Mfg. P/N	GLC-LX-SM-RGD	QFX-SFP-1GE-LX	SFP-1G-LX	GLC-SX-MM-RGD	QFX-SFP-1GE-SX
Standard	1000Base-LX/LH	1000Base-LX	1000Base-LX	1000BASE-SX	1000BASE-SX
Transmission Rate (Gbps)	1.25	1.25	1.25	1.25	1
Tx Power (dBm)	-9.0 - -3.0	-9.5 - -3.0	-9.5 - -3.0	-9.5 - -3	-9.5 - -3
Rx Sensitivity (dBm)	< -21	< -20	< -20	< -17	< -17
Wavelength (nm)	1310	1310	1310	850	850
Transmission Distance	10 km	10 km	10 km	0.55	0.55
Fiber Type	SMF	SMF	SMF	MMF	MMF
Form Factor	SFP	SFP	SFP	SFP	SFP
Cost (\$)	10.00	8.00	8.00	9.00	9.00

From the comparison on table 2, it's very apparent that 10 km, SMF, 1310 nm and 550 m, MMF, 850 nm SFP transceiver modules have relatively standardized parameters. Transmission rates are at least 1 Gbps, transmission optical power ranges between -9.0 to -3.0 dBm, and photodiode optical power sensitivity is below -20 and -17 dBm for SMF and MMF modules, respectively. The cost of an SFP module is very cheap too. Since which SFP module is chosen for the success of the project is, then, not a trivial question.

Between SMF and MMF modules, certain SMF modules can be adapted to fit MMFs [162]. This makes SMF transceivers more versatile than SMF modules for the scope of the project. This further narrows the decision down to only considering SMF SFP transceivers. Due to the limited difference and very low-price difference between Cisco, Juniper, and Arista transceivers, either will suffice the scope of the project. However, the Cisco branded transceivers are the most available, making it a good choice for the project.

3.10 Different Types of Power Supplies

3.10.1 Mean Well

Mean Well is a reputable brand founded in 1982 that can be recognized for its power supplies that are high-quality and reliable. With a strong standing in the industry from a solid reputation, Mean Well's products are used commonly across a range of applications, allowing us to become one of the leading manufacturers of standard power supply products. Industrial automation, LED lighting, and communication systems are just some of those applications. Mean Well is known for producing highly efficient and durable products for electrical needs.

Mean Well has a broad collection of products that are known for high efficiency, boasting a robust design and extensive protection features. The efficiency of some of their models boasts up to 96% while having a wide input voltage range. The power supplies can protect against short circuits, overloads, and over-voltages. Mean Well also includes long warranties for their products, where some can range up to 7 years. On top of this, Mean Well products comply with different international safety and quality standards like the Underwriters Laboratories (UL), Technischer Überwachungsverein (TUV), Conformité Européenne (CE), and more, ensuring that their products are reliable and safe for the multiple uses in the industry.

3.10.1.1 Mean Well LRS-50-5

The LRS series of products under Mean Well's stock offer compact and cost-effective power supplies that boast high efficiency and reliability. As an enclosed style power supply, it holds the AC-DC circuitry with a metal mesh frame, having the connections made with screw-in terminals. Compared to other power sources within Mean Well's repertoire, this is the easiest to work with outside of the GST series. The LRS delivers the best price-to-performance solution on their site while having a full range of protection functions [163].

The Mean Well LRS-50-5 is a 5V power supply unit with a 10A output current that boasts a high reliability and efficiency (up to 90%) while reducing energy consumption and heat generation, which is very helpful for a project like this with a lot of sensitive components that can be damaged easily. In addition to a 90% high efficiency, the metallic mesh case enhances heat dissipation while operating in a -30°C to +70°C temperature range under air convection without a fan [164]. Without the need for a fan, the noise generated by the unit is reduced. Mentioned previously, the LRS-50-5 boasts high reliability, but it has a mean time between failures (MTBF) of 645K hours minimum [165], making it a suitable long-term power source for a security system. With one of the applications listed for this unit as security systems, therefore, it could be a good choice for our product. Also, it is designed to operate in humid conditions, which is important since we are based in a humid area.

3.10.1.2 Mean Well RS-15-5

The Mean Well RS-15-5 comes from the RS-15 series that offers a highly reliable, compact AC/DC power supply. Owning a small footprint with high power density is a suitable choice to combat space constraints depending on the application. Like the LRS, it is an enclosed style power supply, with that similar metal mesh frame. The RS-15-5 can also operate at a range of temperatures from -20°C to +70°C, while also utilizing open air convection for cooling, making it suitable for a variety of environments. Mean Well products boast the inclusion of various protections against electrical complications and this unit is no exception.

The RS-15-5 offers a high efficiency which is around 77%, reducing energy consumption and heat generation. On top of that, with a MTBF of 108.8K hours minimum [166], this unit offers long-term reliability that is essential for a product focusing on security, allowing it to run for long periods at a time while ensuring electrical protection.

3.10.2 RS Pro

RS Pro is a distributor and manufacturer of electronic components founded in 1937 and has grown to be one of the largest distributors in the industry. It is known for its extensive range of products as well as its reliable service, garnishing it as a very reputable brand. These products include electronic, electrical, and industrial components as well as tools and measurement equipment. Having a global reach and appearing as a one-stop shop for an engineer or technician's needs, its online platform allows customers to easily browse and find any products that suit their applications.

In addition to their extensive product range, RS Pro vaunts comprehensive customer service and support. Providing technical support and assistance, customers can find what they need. Based on some reviews, customers have noted that order fulfillment and delivery times had some issues, but with customer support aiding, those issues are resolved in a timely manner.

3.10.2.1 RS Pro 5V DC Power Supply

The RS Pro 5V DC Power Supply or Embedded Switch Mode Power Supply is a 5V output voltage and 7A output current PSU that has an output power of 35W. It is an enclosed configuration that has been designed to offer high efficiency, long life and high reliability for various applications [167]. Although this PSU has a smaller operating temperature range compared to other PSUs, from 0°C to +45°C, it is still a wide enough range where it can operate in humid environments, allowing a variety of environments. There are also built-in protections against complications and burnouts while still meeting various safety standards to ensure it is safe to use in household and industrial applications.

Based on this product's datasheet, this PSU boasts an efficiency of up to 82%, putting it on the lower end of the PSUs being compared but it still allows for a lower energy

consumption and heat generation. Like both Mean Well PSUs, the RS Pro is an enclosed configuration which allows it to use open air convection to cool itself, promoting a long-lasting life.

3.10.3 CUI Inc.

CUI Inc. is a well-known manufacturer of electronic components and power supplies founded in 1989, establishing itself as a leading supplier for the Original Equipment Manufacturer (OEM) market. Focusing on creating ease of access to power supply products and information, they aim to improve the way power products are integrated into designs. Like Mean Well and RS Pro, CUI Inc. possesses a wide variety of power supply products as well as other electrical components. These products are designed to meet a range of applications, whether that be industrial automation, medical devices, or consumer electronics.

CUI Inc. products are designed to meet high efficiency standards as well as comply with international quality standards, including DoE Level VI as well as UL, Puget Sound Energy (PSE), and Federal Communications Commission (FCC) standards. This company is looking to innovate, meaning that they want to offer compact and versatile solutions to meet the needs of their customers. Providing great customer service and technical support, engineers gain a lot of support in their questions regarding integrating their components.

3.10.3.1 CUI Inc. SWI5-5-N-P5

The SWI5-N-P5 is the only power supply that is of a different configuration. The other three are that of an enclosed metal mesh casing but this is an efficient AC-DC wall plug power supply. With an output voltage, current, and power of 5V DC, 1A, and 5W respectively, the SWI5-5-N-P5 is a suitable option for a wide range of applications. These applications could include consumer electronics, industrial equipment, communication devices, as well as security systems, which our product is described as. As a wall plug-in adaptor, we would have to take into mind the distance that we can reach from the point of the outlet to where our system would sit since the cord that comes with the unit is about 1.5 meters in length. The SWI5 has an operating temperature range of 0°C to 40°C, making it the one unit to have the smallest range to be compared. While having a smaller temperature range, this unit also has the smallest MTBF, which is 50K hours as per Telcordia SR-332 [168].

Along with this unit having a range of applications, it makes sense that it would have a range of protections as well as safety approvals. With Class II construction, the SWI5 provides double insulation which enhances safety, while having protections against over-voltage, over-current, and short circuits, further ensuring the longevity of the delicate system. The SWI5 also has UL, PSE, and FCC safety approvals, reinforcing the reliability of the product. On top of that, The SWI5 meets the Department of Energy (DoE) Level VI efficiency standards, meaning that as an AC-DC external power supply with an output between 1W and 49W, we have a minimum average efficiency calculated by:

$$0.071 \times \ln(P_{out}) - 0.0014 \times P_{out} + 0.67$$

While having a maximum power in no-load mode of less than or equal to 0.100, being at less than 0.075W.

3.10.4 Product Comparison

Based on all four of these different PSUs, there are very clear differences while also having a good number of similarities. Looking at the output power of each PSU, we can see that the Mean Well LRS-50-5 gives off the most wattage at 50W, while also has the longest minimum MTBF at 645K hours with an impressive 90% efficiency rate. However, due to these key values, the cost of this unit averages from \$15-20 which makes it the most expensive option. As university students without a sponsor, the total cost of the project must be taken into consideration as we are paying out of pocket. Based on all the comparisons, we have chosen to use the Mean Well LRS-50-5, which has been highlighted in the table below.

Table 3.11 Power Supply Comparison

Parameters	Power Supply Units			
Brand	Mean Well	RS Pro	CUI Inc.	Mean Well
MFG. P/N	LRS-50-5	777-6171	SWI5-5-N-P5	RS-15-5
Output Voltage	5V DC	5V DC	5V DC	5V DC
Output Current	10A	7A	1A	3A
Output Power	50W	35W	5W	15W
Efficiency	Up to 90%	Up to 82%	Up to 85%	~77%
Operating Temperature	-30°C to +70°C	0°C to +45°C	0°C to 40°C	-20°C to +70°C
Form Factor	Enclosed	Enclosed	Wall Plug Mount Adaptor	Enclosed
Size	99x82x30mm	99x82x30mm	22.6x36x47.7mm 1.5m Cord	62.5x51x28mm
MTBF Min. (Hrs)	645K	N/A	50K	108.8K
Price	~\$15-\$20	~\$10-15	~\$10	~\$10-15

Chapter 4 – Related Standards and Design Constraints

4.1 Industrial Standards

Industrial standards are parameters and guidelines specified by governments, companies, or non-profit scientific or engineering organizations to create products and/or systems. To produce a safe, universally understood, and high-quality product, this section discusses the industrial standards relevant for the success of the project.

4.1.1 Related Electrical Standards

Electrical standards include all standards relevant to electrical systems. This section discusses standards with PCB design and electrical power standards.

4.1.1.1 PCB Design Standards

IPC-2221 establishes standards for PCB design such as schematic, material selection, thermal management, DFM, DFA, DFT, and quality assurance [169]. The IPC-2221 standard provides comprehensive guidelines and recommendations for PCB design, which covers numerous aspects that must be adhered to ensure that reliability, manufacturability, and performance are emphasized.

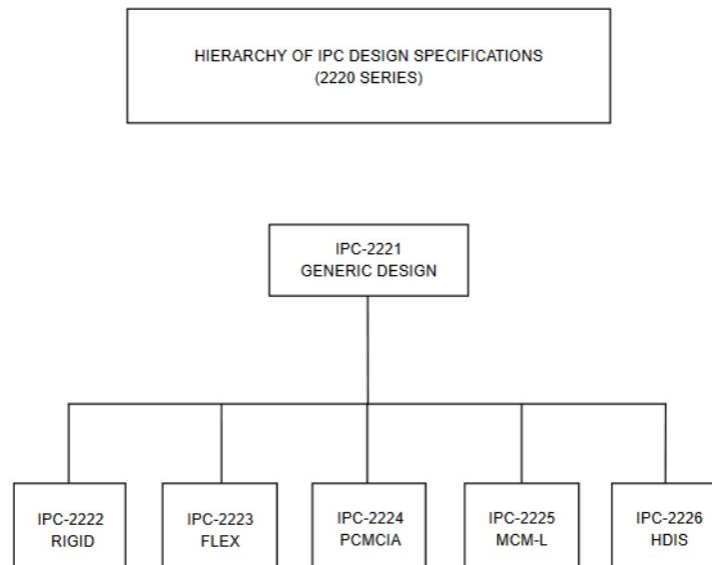


Figure 4.1 Hierarchy of IPC Design Standards

The key areas that IPC-2221 addresses are:

Design Guidelines: Guidelines are offered to provide guidance on the parameters that are critical to PCB design. These parameters include trace width, spacing, and layering. They are essential to maintain proper power distribution as well as maintaining the integrity of the signal being sent.

Material Selection: Recommendations are given for materials to be chosen based on specific properties that should be kept in mind when designing PCBs. These properties include those of dielectric strength, mechanical durability, as well as thermal conductivity. To meet performance specifications, it is important to carefully choose the right materials.

Component Placement: The placement of the components chosen on the PCB is incredibly important and IPC-2221 outlines practices that can be followed to enhance the functionality and reliability of the PCB. Placement can minimize signal interference or even improve thermal management.

Trace Routing: The guidelines provided for trace routing emphasize the best practices to minimize signal interference with high-speed signal routing. These practices detail impedance control, signal integrity, and electromagnetic compatibility (EMC). Having efficient routing makes the system more efficient and stable.

Thermal Considerations: Using strategies such as the use of thermal vias and the placement of heat-generating components are addressed by the IPC-2221 to handle thermal management. Keeping effective thermal management in mind will prevent overheating and improve the longevity of the device as well as the components inside.

Quality and Reliability: To ensure that the designed PCBs meet quality standards, criteria are included in IPC-2221. Reliability and compliance with industry regulations are ensured when the criteria are followed.

Design for Manufacturability (DFM): With manufacturability kept in importance, IPC-2221 reduces issues and streamlines production with the DFM guidelines that are included. These guidelines cover aspects like ease of fabrication, assembly, and testing.

Overall, using IPC-2221 as an encompassing guide for designing our PCB, we can create a reliable system that can adhere to industry standards. Utilizing these standards will ensure that our designed PCB will be manufacturable and function properly at a minimum cost.

4.1.1.2 IEC 62368-1 and UL 62368-1

IEC 62368-1 and UL 62368-1 are standards that apply to audio/video, information, and communication equipment that cover safety requirements to protect against electrical, thermal, mechanical, and radiation hazards. UL 62368-1 is the North American equivalent to IEC 62368-1 while ensuring that products meet both international and North American safety requirements. These standards focus on ensuring that products follow the requirements that provide safeguards against likely and consequential faults, foreseeable misuse and external influences such as temperature, altitude, pollution, moisture, and

overvoltage [170]. Products and manufacturers, like Mean Well, must undergo testing and certification by UL to ensure compliance while also requiring proper documentation to inform users of safety precautions and compliance.

4.1.1.3 DoE Level VI

The U.S. Department of Energy (DoE) established a set of regulations labeled as the DoE Level VI efficiency standard to improve energy efficiency within external power supplies. Based on the graph provided below, the requirements are listed.

AC-DC External. Power Supplies (except low voltage and multiple voltage output)		
Nameplate Output Power (P_{out})	Minimum Average Efficiency in Active Mode (expressed as a decimal)	Maximum Power in No-Load Mode (W)
$P_{out} \leq 1W$	$\geq 0.5 \times P_{out} + 0.16$	≤ 0.100
$1W < P_{out} \leq 49W$	$\geq 0.071 \times \ln(P_{out}) - 0.0014 \times P_{out} + 0.67$	≤ 0.100
$49W < P_{out} \leq 250W$	≥ 0.880	≤ 0.210
$P_{out} > 250W$	≥ 0.875	≤ 0.500

Table 4.2 DoE Level VI Requirements [171]

The implementation of DoE Level VI is expected to increase energy savings significantly while contributing to lower electricity bills for customers. Although this standard is mainly for external power supplies, it is still beneficial to explain and understand the benefits and requirements of implementing an efficient power supply to power our system.

4.1.1.4 IEC 61000-4-5 and Other IEC 61000 Standards

For manufacturers to produce safe and reliable electrical and electronic equipment, the products would have to be tested to comply with standards given out like IEC 61000-4-5, which describes the immunity requirements, test methods, and range of recommended test levels for equipment subjected to surges caused by over-voltages. While putting equipment under testing, the standard aims to make sure that this equipment can withstand without suffering from malfunctions. The standard has several performance criteria to pass, with those being operating as intended during and after the test, recovering automatically after temporary loss of function, or requiring manual intervention to restore operation after temporary loss of function.

It is important for companies to comply with the standards presented as they enhance the performance and reliability of the product they are trying to push. Having safe products and those that work as intended gives the company the reputations that they gain. Without compliance, the purchased products are deemed unsafe to sell and would risk the systems that they are integrated into. As mentioned before, one of the companies that our power supply will come from, Mean Well, follows IEC 61000-4-5 as well as other IEC 61000

standards [172]. What this means is that they comply with the standard as well as test their products so that they manufacture reliable and safe products.

4.1.2 Related Optics Standards

Optical standards are an important part of the project. These standards range from optical design standards to telecommunication standards. The following section only outlines the optical standards deemed relevant to this project.

4.1.2.1 ISO 9001, ISO 10110, and ISO 9345

The lenses used in this project were purchased by Edmund Optics and Thorlabs. These two manufacturers follow strict regulations to ensure that their products are on par with the standards.

Edmund Optics is ISO 9001 certified in several of its facilities, including the company's headquarters. Their products "undergo rigorous and thorough testing as part of the company's quality program and in compliance with Edmund Optics' global quality procedure" [173]. ISO 9001 is the most used standard for quality management globally [174]. [ISO 9001:2015] Edmund Optics has also implemented ISO 14001 standards in their company to improve resource efficiency and reduce waste [173].

Edmund Optics is also ISO 10110 certified. ISO 10110 is the standard for optical drawings. "ISO 10110, Optics and optical Instruments – Preparation of optical drawings for optical elements and systems, is the primary reference for preparation of drawings for optical elements and systems" [175]. [175]. This standard is divided into 13 parts, and they each highlight details that should be taken into consideration when drawing and describing the drawings for optical elements. All of the optical components fabricated by the company have a drawing in which we can see its schematic and description (Appendix D). [175].

Thorlabs is ISO 9001, ISO 10110, and ISO 9345 certified. ISO 9345 is the standard for optics and photonics instruments, it refers to the imaging distances related to the mechanical reference planes [175], [176].

4.1.2.2 Optical Transceiver Multisource Agreement

Optical transceiver modules across companies do not share the same form factor and basic design by mistake – these companies share an MSA. A multi-source agreement, also known as an MSA, is a manufacturing agreement that binds manufacturers to make a product that has the same functionality across different vendors [177]. This agreement was created to address the form, fit, and function, like electrical and optical interfaces, of optical transceiver modules before IEEE created its standards [177].

At a high level, MSA standards make all MSA-compliant transceivers the same while keeping costs low – third-party transceivers, like HiFiber, will offer the same functionality

as a name-brand transceiver, like Cisco, for the same price [177]. With an MSA, customers have more options at their disposal without compromising performance or costs. Additionally, this effort enables customers to efficiently exchange transceivers across switches and routers [177]. All optical transceivers listed in section 3.3.3.3.1 are MSA compliant.

4.1.3 AI Standards

AI Standards are very important when it comes to our project. They are focused on the safety and security for individuals. When it comes to the use of AI functionality it is key especially when used globally. Although that is not the only thing the AI Standards are for when it comes to government and business, AI must be relevant to the work or job depending on the circumstances. Threats from the use of AI must be considered and maintenance of systems that require AI. The work that AI does in designing scripts, computer vision techniques and web searching is a huge area of concern. The standard mentions the transparency of data characteristics and the system that is using AI. AI can also cause risks to happen in process or design which there is a focus on risk management for AI as well. Then there are the smaller things about AI use such as terminology, measurements and metrics in development. AI Standards are for prevention as well. Human interaction with AI can cause harm and invasion of privacy. AI Standards cover privacy, security, safety and prevention of harm towards humans.

4.1.4 Surveillance Systems Standards

For surveillance systems the suggested standards are ISO 9001, ISO 27001 and, ISO 22301. The first one was discussed in the previous section, it is a standard about quality management. ISO 27001 refers to information security, and ISO 22301 is about business continuity [178].

ISO 27001 includes requirements for information backup, information deletion, access to source code, and clock synchronization to name a few [179], [180]. According to HighTable, a company that specializes in consulting and implementation of this standard, ISO 27001 was designed to be used “for implementing information security controls based on internationally recognized best practices; for developing organization specific information security management guidelines.”

ISO 22301 provides a “framework for organizations to plan, establish, implement, operate, monitor, review, maintain, and continually improve a documented management system to protect against, reduce the likelihood of, and ensure recovery from disruptive incidents.” [181].

Another standard for surveillance systems is IEC 62676-4:2014. This standard gives recommendations and sets requirements for all the stages, from product selection to installation and testing video surveillance (VSS) [182].

4.1.5 Privacy Regarding Surveillance Systems

Surveillance systems that include audio transmission and recording have many benefits as they can alert the user if someone is entering their property or wake up the parents if their baby starts to cry.

However, the recording and storing of audio is a legal issue that deals with the privacy of the individuals being recorded and we had to consider when planning this for project.

The federal Wiretap Act prohibits individuals from secretly recording conversations, and on top of that Florida is one of the states that requires two-party consent for a conversation to be recorded. In other words, for iSPY to transmit, store and record audio we would need the consent of everyone in the room [183], [184]. Therefore, we decided to not add audio recording to our project. Even if installed in a public place any recording done could not be used in court and/or as part of an investigation. [183], [184].

4.2 Practical Constraints

Constraints can take the shape of any obstacle that a team faces during a project and are an important aspect of the design process that need to be taken into consideration. This is to ensure that the project meets parameters set by various variables that the project is encompassed by. In our case, as student engineers in senior design, we are faced with a multitude of constraints that can set us back if not managed.

4.2.1 Time

For this project, the primary constraint is time. Since this is a senior design project, we have very strict regulations and deadlines that must be met. This only gives our group a limited amount of time to conceptualize, research, design, and fabricate our system and idea. Due to this restraint, we have had to come up with different solutions and follow different paths to focus on meeting our deadlines rather than fully fleshing out a fully optimized final product.

One of the key challenges we faced is that we only had a single academic calendar year to construct and demonstrate a functional system. We are given no extensions and have firm deadlines that must be met to successfully complete the course and pass. Fortunately, we were able to gain a bit of a head start as half of our group was formed in the previous summer semester and started workshopping the idea that we wanted to do. This allowed us to mitigate some pressure as we progressed because we had the general idea, and we can jump straight into research and development immediately.

On top of designing the system, we also must have ample time set aside to conduct extensive testing and debugging, all within the timeframe we were given. In any sort of

design project, we assume that we will run into setbacks and errors, so we must allocate a sufficient timeframe that can encapsulate this threat if need be. Unforeseen issues that weren't accounted for are a hindrance to the success of the project plan and the passing of the course.

Another related issue would be the delay that we could see when ordering the parts for the project, as a numerous number of variables can affect whether the components, we ordered arrive in time to complete the project. Projects like this, with such a limited time span, often face delays at this stage because there could be potential supply shortages or issues with shipping that could deter the design's completion. Without the components for the project, we cannot move forward with the critical portion of constructing, testing, and debugging the system, eating up valuable time.

Having such strict time constraints might make us want to take shortcuts to achieve our final goal. However, having that mindset will only lead to a rushed and sloppy project that compromises the quality of the overall function. Adjustments are bound to be made to help solve issues but if those adjustments aren't drastic at such a late time, we can avoid jeopardizing the structure of the project.

One of the ways the deadline has constrained our project is seen in the optical design. When designing a wide-field-of-view system a lot of optical designers can count on parts being custom-made for their prototype to work. However, after talking to some people in the industry who actively need custom-made lenses, we were told that the waiting time is at least three months. Unfortunately, that would not be enough time for us to receive and test our system. Therefore, the system will have to adapt lenses/surfaces that are already available in the market.

Our team has a strong value that we must hold everyone accountable and withhold our responsibilities that we are given or take on. We follow the principle that was explained to us in our class where we will "Do What You Say You Will Do" or DWYSYWD. Having a set goal and working towards that goal consistently and efficiently will lead to a better product. Communication and proper planning help mitigate these potential issues that could pop up and allow us to complete our project on time and to a quality standard.

4.2.2 Economic

One of the main constraints that must be taken note of was the economic constraint. Our project has no sponsors, so the whole cost will be coming out of pocket. Due to this, we must take extra care in constructing a budget so that we have enough funds to cover the entirety of the project or not spend more than we need to. We had to carefully investigate

different components and comparisons so that we can find the products that have the capability to do what we want to accomplish.

We hope to keep the entire cost below \$700 while still having extra funds if something breaks or burns out during testing. By having a buffer of funds, we can ensure that in an unfortunate event, we allow ourselves to have less worries when working with such delicate components. Although we can replace anything that breaks, we'd like to avoid that possibility since it will only rack up a higher cost in the long run of the project.

Our research in part selection is essential when carrying out the project, as we don't want to spend an absurd amount of money on a component that has no overall functionality for the purpose of the project. Therefore, we must consider the price of components and the functionality in our research, making sure that we can use the component, or it would be a waste of time and money, both of which are very limited as Senior Design students.

The economic constraints that we place upon ourselves can be a detriment to the success of the project because of the need to buy the best products on the market that can do what we would like to achieve while also not being absurdly expensive. This requires us to do exceptional research so that we can compare different products and make sure that the cost ratio is sufficient. We can also avoid going over the budget if we give ourselves a cost buffer. This cost buffer, as explained previously, will give the team a safety net when we're in the delicate fabrication process. If something were to break, we would already have a time issue since we would need to wait for a component to come in, but instead of being over budget and over time, we can allow some leeway in terms of cost.

As long as we keep researching and making sufficient comparisons, we can choose the best products that can handle the scope of the project while also maintaining our budget and keeping that safety buffer intact.

4.2.3 Environmental

The environmental constraint is one of the biggest design constraints that we will have to factor in for this project since we rely heavily on our FSO system. Within FSO, we must work with open air acting as our conductive material for light. This could lead to numerous variables that could potentially affect the functionality of the system since these types of systems are very sensitive to a range of factors.

One of the biggest aspects that can affect our system is light interference. Since a portion of our system is in the open air, sunlight and ambient indoor light can distort and overwhelm the system and the transmitted bitstream, resulting in noisy data and output.

Just like light, we can also be wary of electromagnetic interference from any other electronic devices in the vicinity of our system.

A significant challenge that our FSO system can face would be interference from the atmosphere around the system, specifically any air impurities that may be present. For example, in an indoor environment, dust or airborne particulates can be a source of interference as they can diminish the integrity of the signal. Or, in an outdoor environment, foggy and rainy areas can absorb and disrupt the light being transmitted. If our signal gets disrupted or absorbed by other variables then we wouldn't get the right data back, if at all, compromising the range that the signal can be received.

Of course, since we are working with electronics, we would have to think about how temperature and humidity can affect our components. Living in Florida, we get hot temperatures as well as humid or moist weather. Humidity is one of the bigger threats as our components can be short circuited or corroded from the condensation that is gathered. And with temperature, we would have to think about the overheating that can occur from the operation of our system on top of the environment's temperature.

Furthermore, having optical paths makes the system complicated since we must be precise and accurate in our construction. Therefore, any external movement such as vibrations and mechanical shocks can disrupt the system and null out the output we expect to gain. A vibration could knock our pathing out of alignment and disrupt the system so having a robust housing along with secure installation can mitigate the risk that this variable imposes on our system.

For any system relying on sensitive components, it is essential to consider the effects of these constraints so that we can make the proper choices to construct a strong design. We would have to design a robust housing structure for the system that can withstand environmental variabilities just in case any of these constraints prove to be an issue. Maintenance and system redundancy can also improve the longevity of the system.

4.2.4 Remaining Constraints

Constraints are encountered throughout all engineering projects. Some constraints can be accounted for while others cannot. In this section, constraints like sustainability, manufacturability, and ethics are discussed.

4.2.4.1 Sustainability

Since this system has the sole purpose of security, sustainability is a major constraint since we want a system that will last a long time without needing much to maintain it. Having a system that can last a good amount of time energy-wise, while also being able to withstand some weather conditions if used outside, is what a good security system can do.

We would have to make the system with the idea that natural weather or indoor conditions vary, so having a strong casing will have to be a priority in the overall construction of the system. On top of that, having a strong power supply and efficient power management can have a huge benefit to the overall functionality of the project as it will be able to utilize power for a long time without components burning out or not having enough power.

4.2.4.2 Manufacturability

Another constraint that must remain considered for this project is that it needs to be manufacturable, which is important for a group of undergraduate students working on this project. We are limited to resources provided by the university; therefore, we are limited in the scope of the project. This results in our group having to keep the design of our project simple enough so that we can utilize resources provided to us.

However, as we are not required to create things from scratch, we can purchase our components from online resources and modify anything as needed, which reduces the need to go overboard with extensive designing. By allowing us the ability to outsource aftermarket parts, we can assemble our system using techniques that we have gathered, thereby minimizing the workload of constructing a functioning system.

Since our project deals with optics, we have opted to purchase some pre-made parts that will lessen the burden the fabrication of our system will have on the group. Of course, we will design a lot of how the components come together, but we wouldn't be able to construct the major components from scratch.

4.2.4.3 Ethical

As a group, we have decided that we would like to factor in an ethical constraint into our project, ensuring that we choose companies that are ethical in nature. We would like to avoid purchasing parts and components from companies that outsource from countries that use any type of slavery or labor of an unethical manner. Third-world countries are a prime

target for an exploited workforce. Being aware that trade-off is deemed important to our group as we all believe in ethical practices in the engineering industry.

Since our group has a strong belief in the ethical acquirement of our parts, we want to make sure that this constraint is factored properly in our planning. Therefore, we want to only use parts from companies that are deemed trusted based on track records and extra research. To the best of the group's knowledge, we will avoid dealing with companies that use unfair labor practices and only source parts from fair businesses that hold a high standard. Limiting unfair business practices will be ensured with the extra research our group will conduct into the choices we make in order for this product to take shape.

Chapter 5 – Large-Language Model Engine

In the recent past, large-language models have enabled tools for engineers across disciplines. As this tool continues to proliferate throughout the industry, it's important to consider it for the success of the project. In this section, large-language model outputs are discussed in detail with regards to details of the project.

5.1 Case Study 1: Asking ChatGPT and Copilot for Camera Module Specifications

Large Language Models (LLM) are a great tool when used correctly. For this project, it was used as a search engine, like the way we use Google, it was also used to help with brainstorming and reviewing our grammar.

In prompt 1, we asked ChatGPT what specifications/parameters we should consider when searching for a camera module. The outcome was helpful, it gave us a list of details to look for and it explained the role of each parameter and why it mattered. We did not mention that we were on a budget, but the LLM went ahead and told us to watch out for the financial aspect of it.

The same prompt was given to Microsoft's Copilot and the outcome was similar. Copilot went a step further and gave us references (links) for each parameter it mentioned, and it also highlighted the importance of the sensor's availability.

For this case study the results obtained through LLM were helpful, they summarized a big portion of the key aspects we should research before purchasing our camera module, and one of them even gave us websites to check the information.

5.2 Case Study 2: ChatGPT vs. Zemax

The lens design software Zemax is very a unique tool, and it is mostly used by optical engineers. The software is not that popular outside of the field of optics and photonics. Therefore, there is a limited amount of information that can be found making it harder for LLMs like ChatGPT to have accurate data on it. ChatGPT does not know how to design a lens system, the outcome 2 shows that it has ideas on how things should be but when its inputs are brought to Zemax the design does not work.

Also, whenever ChatGPT tells the user to do something in Zemax the instructions are not clear. "Optimize your design iteratively, using the Merit Function Editor in Zemax, focusing on minimizing chromatic aberration, spherical aberration, and distortion." Such direction is not so straightforward as the LLM implies. To use the Merit Function Editor the user needs to set up boundaries, add specifications, check parameters and make sure they fall under the design that is being implemented.

When dealing with peculiar softwares, using LLMs is not beneficial and in fact I believe it only hurts the learning and designing process. Since it does not have accurate data but only a faint plan/idea on how it should be, following their steps leads to issues.

5.3 Case Study 3: Asking Multiple AI Models for Power Supply Recommendations

Using LLM as a secondary search engine is useful for quick research while providing additional resources to the prompts that are inputted. With the help of multiple AI resources, namely ChatGPT, Copilot, and Google AI, I was able to amass a good list of different PSUs that could be integrated into our design to supply power. Some are more beneficial than others as one could provide more information in one pass, whereas another could only provide one general response. As seen in ChatGPT's response, it gave me very broad types of PSUs that could be used rather than actual products currently on the market. It could be asked for more information but that would require a more in-depth inquiry. Copilot and Google AI both provide different products that are currently in the market and available to purchase, with the exception of some products being out of stock. Not only did they provide different products from different brands, but also provided links and references to investigate each product more individually.

One thing that ChatGPT and Copilot have in common, whilst differing from Google AI, is that we can add files and pictures into our prompt to give further information to our question. This allows the model to learn more about the topics and provide a response that is more in line with what we are looking for. Google AI doesn't have that feature since we are searching directly on Google, but it was still able to provide suitable answers.

5.4 Case Study 4: Camera Module Options Using ChatGPT

We also used ChatGPT when looking for camera modules [Appendix F – prompt 4]. Since the LLM can search its database quickly and give us options it was a good way to narrow down our choices, it was a lot easier to ask ChatGPT what the options were then to visit several websites and manually check every single camera module. The LLM also gave us a range of values for each module.

The following question was if it knew what the shutter type of one of the modules was and if there was a way we could replace the lenses from the modules to add the lens design we wish to have. The first question was answered with some uncertainty. It did give the sensor type, however it also explained both types of sensors (rolling and global) as if it was justifying its answer, and then at the end it told us to check the specification tables of the module.

Even though it narrowed our options down ChatGPT could not offer certainty about the specifications, the LLM still has limitations, and it is upon every user to go and check the information.

ChatGPT told us that yes, we can replace the lens for one of the modules and even gave us steps in how to do so. The steps were not so specific but it came up with it without us even asking for instructions. The LLM also warned us of the issues we might face when removing the original lens and placing something else on its spot.

5.5 Takeaways

When used as a search engine, a tool to brainstorm ideas, a place to obtain an overall picture before getting into the details, LLMs are extremely beneficial. Since websites and applications like ChatGPT and Copilot speed up the proofreading process and answer our questions almost instantly and with great accuracy, they enable us to direct our energy into understanding how all our areas will come together in this project.

We have an idea, each one of us understands their part and how we got there (the mathematics, physics and engineering behind), we know what to look for when purchasing the parts and now is the time to figure out how to place this puzzle together.

I think this is the best part about using LLMs. Of course we must understand what is taking place, but after that they can save us a lot of time and energy. The way LLMs can harm our learning is if we completely replace books, lecturers, and research papers for their output. As good as it is in acquiring and delivering information, it does not obtain the answer for everything, especially if the subject is unique.

It's important to look at ChatGPT as a tool in your arsenal instead of it being your go to for everything related to research. ChatGPT has greatly benefited our learning experience in Senior Design because we used it for guidance instead of using it as our sole means of research. In the beginning of writing sections, it can be difficult to know where to start. Using ChatGPT to give a guideline for what to write can be incredibly helpful to jump start ideas and get rid of brain fog.

For example, the prompt was given to ChatGPT asking what someone should write about when researching codecs for live video compression to transfer data across a laser using bitstream. ChatGPT gave a comprehensive list of what should be covered and what is important to note when researching. ChatGPT's response worked well as a guide rather than a sole source of information. ChatGPT listed the types of codecs that should be covered. Instead of just relying on what ChatGPT said, research was done to get more information that can be used as quality research. Numerous articles were visited trying to research the topic to their fullest so that an accurate decision can be made on what to use.

ChatGPT was always supplemented with proper research done by researching topics from various other sources. ChatGPT can be wrong. As helpful as it can be, it is not perfect and is prone to mistakes. That's why it's important to not rely on it entirely. In that way, ChatGPT can be harmful. If a person blindly follows what ChatGPT says without verifying it for themselves, then they are doomed to make mistakes.

Chapter 10 – Administrative Content

The success of the project equally depends on the administration of its many parts and engineers. A detailed discussion on the group's financial budget and distribution of work is necessary to understand the decision-making process. This section discusses the project's budget, bill of materials, distribution of work, and overall milestones.

10.1 Budget and Bill of Materials

The iSPY project is to be entirely self-funded by its group members. Sponsors capable of supplying critical higher-cost items and components, like lenses, processing units, or electrical hardware, are currently being analyzed. It's the Authors goal to minimize costs and limit expenditures to less than \$1000.

The table below highlights the distribution of costs across the different major sub-systems and potential components that make up iSPY. Historically, optical and photonic components are priced higher than the electronics components; allocating a larger sum of the budget for this is acceptable.

Table 10.1. Itemized bill of materials.

Sub-System	Components	Budget	Unit Cost	Quantity	Total Cost
Imaging Acquisition	Camera Module	\$ 200.00	\$ 75.59	1	\$ 75.59
	Imaging Lenses		\$ 43.91	4	\$ 175.64
Transmission/ Receiver System	Near-IR Photodiode	\$ 300.00	\$ 58.91	1	\$ 58.91
	Near-IR Laser Diode		\$ 54.16	1	\$ 54.16
	Tx/Rx Lenses		\$ 43.91	3	\$ 131.73
	Miscellaneous Optical Components		\$ 50.00	1	\$ 50.00
Computer Hardware	Microcontroller Unit	\$ 50.00	\$ 10.00	2	\$ 20.00
Electrical Housing and Power Systems	Transmission PCB	\$ 50.00	\$ 20.00	1	\$ 20.00
	Receiver PCB		\$ 20.00	1	\$ 20.00
	Power Supply Unit		\$ 12.42	2	\$ 24.84

	Miscellaneous Electrical Components		\$ 20.00	1	\$ 20.00
Display	Monitor	\$ 25.00	\$ 25.00	1	\$ 25.00

10.2 Distribution Worktable

The distribution worktable below details the primary and secondary members assigned to each responsibility. Each group member is assigned as the primary member for a specific sub-system that makes up the overall system. However, it's important to note each sub-system strongly depends on the adjacent sub-systems of another member, making secondary member assignments equally important. Each group member will also be assigned as secondary member for two or more responsibilities across sub-systems to meet this need, encouraging collaboration and discussion members.

Table 10.2 Distribution worktable.

Sub-System	Responsibilities	Primary Member	Secondary Member
FSO	FSO Lens Design and Implementation	Alejandro	Gabryella
	Laser and Detector Selection and Implementation		Gabryella
	Optoelectronics integrations with PCB Housing		Robert
Imaging	Design And Align Lenses for Image Acquisition System	Gabryella	Alejandro
	Selection of Appropriate Image Sensor		Alejandro
	Optimization of Image Quality		Luke
	Integration of Image Acquisition System		Luke
Object Detection	Computer-Vision Software Design and Implementation	Robert	Luke
	Appropriate MCU Selection		Armani
	Machine Learning Algorithm Development		Luke
Image Processing	Image Correctness	Luke	Armani
	Output Video Processing and Display		Robert
	Compression Software Design and Implementation		Robert
Electrical Hardware	PCB Design and Fabrication	Armani	Luke
	PSU Design and Implementation		Robert

	Interface Design for Data Transmission		Robert
	Signal Conditioning and Processing		Robert
Website Management	Organize and update group's website with assignments before deadline	Luke	Robert

10.4 Milestones

The success of the project depends on individual and collective achievements. To create a good dynamic amongst group members and to maintain pace, the collective milestones encountered in Senior Design 1 and Senior Design 2 are highlighted below.

Table 10.3. Senior Design 1 Milestones - Project design phase.

Senior Design I Milestones		
Task	Status	Deadline
First group meeting	Completed	August 15, 2024
Project selection/Brainstorming	Completed	August 25, 2024
Group meeting to discuss D&C document and committee formation	Completed	August 27, 2024
Idea presentation to Dr. Kar	Completed	August 29, 2024
Submit Divide and Conquer 10 pages report	Completed	September 6, 2024
Committee Formation	In progress	September 6, 2024
Divide and Conquer Meeting with Advisors	Completed	September 9, 2024
Fix Divide and Conquer report	Completed	September 27, 2024
Upload Divide and Conquer Report on website	Completed	September 27, 2024
Midterm Demo (CREOL)	Completed	October 17, 2024
Upload 75 Page Report	Completed	October 25, 2024
75 Pages Report Meeting with Advisors	Completed	October 29, 2024
Fix 75 Pages Report	Standby	November 8, 2024
Upload 75 Pages report on website	Standby	November 8, 2024
Final Demo (CREOL)	Standby	November 21, 2024
SD1 Final Report	In progress	November 26, 2024
Upload Mini Demo Video on website	Standby	November 26, 2024
Project Research	In progress	-
Order Parts	Standby	Before Winter Break

Table 10.4. *Senior Design 2 Milestones - Project creation phase.*

Senior Design II Milestones		
Task	Status	Deadline
Project Research	In progress	-
Prototype Development	Standby	TBD
Testing	Standby	TBD
Troubleshooting	Standby	TBD
Optimizing	Standby	TBD
CREOL Showcase	Standby	April 25, 2025

10.5 Acknowledgements and Declaration

Declaration: We hereby declare that we have not copied more than 7 pages from the Large Language Model (LLM). We have utilized LLM for drafting, outlining, comparing, summarizing, and proofreading purposes.

Appendix A – References

- [1] W. G. Schulz, “Eye Spy: Optics, Photonics, and Espionage,” *Photonics Focus*, vol. 5, no. 2, pp. 19–21, Mar. 2024.
- [2] G. Chen, “Advancing Mobile Sensing with Resilient Distributed Computing,” *Photonics Focus*, vol. 5, no. 2, pp. 19–21, Mar. 2024.
- [3] S. M. Kelly, “That security camera and smart doorbell you’re using may have some major security flaws | CNN Business,” CNN. Accessed: Aug. 30, 2024. [Online]. Available: <https://www.cnn.com/2024/03/09/tech/smart-home-cameras-hackers-security/index.html>
- [4] N. Vigdor, “Somebody’s Watching: Hackers Breach Ring Home Security Cameras,” *The New York Times*, Dec. 15, 2019. Accessed: Aug. 30, 2024. [Online]. Available: <https://www.nytimes.com/2019/12/15/us/Hacked-ring-home-security-cameras.html>
- [5] S. Ikeda, “Russian Agents Hacking Residential Surveillance Cameras to Gather Intel in Ukraine,” CPO Magazine. Accessed: Aug. 30, 2024. [Online]. Available: <https://www.cpomagazine.com/cyber-security/russian-agents-hacking-residential-surveillance-cameras-to-gather-intel-in-ukraine/>
- [6] J. Pattison Tuohy and R. Henderson, “How To Keep Your Home Security Cameras Safe,” US News & World Report. Accessed: Aug. 30, 2024. [Online]. Available: [//www.usnews.com/360-reviews/services/security-cameras/how-to-keep-your-security-cameras-safe](https://www.usnews.com/360-reviews/services/security-cameras/how-to-keep-your-security-cameras-safe)
- [7] “Features – Koruza.” Accessed: Sep. 05, 2024. [Online]. Available: <http://www.koruza.net/features/>
- [8] “Wireless Optical Data Transmission Technology KORUZA: Upgrade Announcement! - IRNAS.” Accessed: Sep. 05, 2024. [Online]. Available: <https://www.irnas.eu/wireless-optical-data-transmission-technology-koruza-upgrade-announcement/>
- [9] “MOSTCOM – PathFinder Digital Wiki.” Accessed: Sep. 05, 2024. [Online]. Available: <https://wiki.pathfinderdigital.com/wiki/mostcom/>
- [10] R. Awati, “What is field of view (FOV)?,” TechTarget. Accessed: Sep. 01, 2024. [Online]. Available: <https://www.techtarget.com/whatis/definition/field-of-view-FOV>
- [11] D. Kim, S. Chang, and H.-S. Kwon, “Wide Field-of-View, High-Resolution Endoscopic Lens Design with Low F-Number for Disposable Endoscopy,” *Photonics*, vol. 8, no. 4, Art. no. 4, Apr. 2021, doi: 10.3390/photonics8040089.
- [12] Threat.Media, “What is a Bitstream?,” [Online]. Available: <https://threat.media/definition/what-is-a-bitstream/#:~:text=A%20bitstream%20is%20a%20sequence,any%20other%20type%20of%20file>
- [13] L. Hardesty, “Explained: Neural networks,” MIT News | Massachusetts Institute of Technology. Accessed: Sep. 05, 2024. [Online]. Available: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>
- [14] D. C. O’Shea and J. L. Bentley, *Designing Optics Using Zemax OpticStudio*, 1st ed., vol. PM367. SPIE, 2024.
- [15] “CMOS Sensor Size Comparison Guide and Lens Calculator,” Commonlands Optics. Accessed: Sep. 14, 2024. [Online]. Available: <https://commonlands.com/blogs/technical/cmos-sensor-size>

- [16] R. Paschotta, "Field of view." Accessed: Sep. 03, 2024. [Online]. Available: https://www.rp-photonics.com/field_of_view.html
- [17] T. Don, *Camera Sensor Size Explained*, (Jan. 25, 2017). [Online Video]. Available: https://www.youtube.com/watch?v=cBV8wpNmnpY&ab_channel=Channel8
- [18] S. d'Entremont, *Does sensor size affect aperture? The TRUTH about CROP FACTOR.*, (Apr. 11, 2024). [Online Video]. Available: https://www.youtube.com/watch?v=Apum3Ezqo0U&ab_channel=Simond%27Entremont
- [19] "Digital Camera Sensor Sizes: How it Influences Your Photography." Accessed: Sep. 14, 2024. [Online]. Available: <https://www.cambridgeincolour.com/tutorials/digital-camera-sensor-size.htm>
- [20] J. M. Geary, *Introduction to Lens Design: with practical ZEMAX examples*, 1st ed. Richmond, Virginia: Willmann-Bell, Inc, 2022.
- [21] B. Saleh, Ed., *Introduction to Subsurface Imaging*. Cambridge: Cambridge University Press, 2011. doi: 10.1017/CBO9780511732577.
- [22] S. d'Entremont, "HOW MANY MEGAPIXELS DO YOU NEED? YOU MAY BE SURPRISED! - YouTube." Accessed: Sep. 16, 2024. [Online]. Available: https://www.youtube.com/watch?v=ThpQWhOfKO4&ab_channel=Simond%27Entremont
- [23] "1080p vs. 2K vs. 4K Security Cameras: Understanding the Differences | Mammoth Security Inc." Accessed: Sep. 16, 2024. [Online]. Available: <https://mammothsecurity.com/blog/1080p-vs-2k-vs-4k-security-cameras>
- [24] N. James and G. Hollows, "Resolution," Edmund Optics. Accessed: Sep. 16, 2024. [Online]. Available: <https://www.edmundoptics.com/knowledge-center/application-notes/imaging/resolution/>
- [25] "What are Frames Per Second (FPS)?" Accessed: Sep. 18, 2024. [Online]. Available: <https://www.isarsoft.com/knowledge-hub/fps>
- [26] Elvia, "Complete Guide to Security Camera CMOS VS CCD Image Sensors," Reolink Blog. Accessed: Sep. 20, 2024. [Online]. Available: <https://reolink.com/blog/security-camera-cmos-vs-ccd-image-sensors/>
- [27] "Electronic shutter vs mechanical shutter," Canon Europe. Accessed: Sep. 18, 2024. [Online]. Available: <https://www.canon-europe.com/pro/infobank/electronic-vs-mechanical-shutter/>
- [28] IPVM Team, "Shutter Speed / Exposure Tutorial," IPVM. Accessed: Sep. 20, 2024. [Online]. Available: <https://ipvm.com/reports/exposure-setting-tutorial>
- [29] "Rolling Shutter vs Global Shutter sCMOS Camera Mode," Oxford Instruments. Accessed: Sep. 20, 2024. [Online]. Available: <https://andor.oxinst.com/learning/view/article/rolling-and-global-shutter>
- [30] C. Yang, "Global Shutter vs. Rolling Shutter in Embedded Vision," TechNexion. Accessed: Sep. 20, 2024. [Online]. Available: <https://www.technexion.com/resources/global-shutter-vs-rolling-shutter-in-embedded-vision/>
- [31] AlliedVisionTV, "Sensor module vs. camera module." Accessed: Sep. 21, 2024. [Online]. Available: https://www.youtube.com/watch?v=Oi4S7SwumGs&ab_channel=AlliedVisionTV

- [32] B. McCurrach, “A Quick Overview Of Industrial Camera Interfaces,” *Quality Magazine*. Accessed: Sep. 21, 2024. [Online]. Available: <https://www.qualitymag.com/articles/97108-a-quick-overview-of-industrial-camera-interfaces>
- [33] AlliedVisionTV, “Which camera interface is the right one for my application?” Accessed: Sep. 21, 2024. [Online]. Available: https://www.youtube.com/watch?v=Co2PR1wDRLk&ab_channel=AlliedVisionTV
- [34] P. Kumar, “What is a MIPI Interface? How does MIPI Camera Work?,” *e-con Systems*. Accessed: Sep. 29, 2024. [Online]. Available: <https://www.e-consystems.com/blog/camera/technology/what-is-a-mipi-camera-how-does-mipi-camera-work/>
- [35] “Fixed focal length lenses for machine vision applications,” *Opto Engineering Basics*. Accessed: Sep. 28, 2024. [Online]. Available: <https://www.opto-e.com/en/basics/fixed-focal-length-lenses>
- [36] P. Kumar, “Fixed-focus lens vs. autofocus lens: How to choose the best lens for your application,” *e-con Systems*. Accessed: Sep. 29, 2024. [Online]. Available: <https://www.e-consystems.com/blog/camera/technology/fixed-focus-lens-vs-autofocus-lens-how-to-choose-the-best-lens-for-your-application/>
- [37] “Numerical Aperture (NA),” *MEETOPTICS*. Accessed: Sep. 28, 2024. [Online]. Available: <https://www.meetoptics.com/academy/numerical-aperture?srsltid=AfmBOorbSLQJmJttjanMOndBKgcGwBhzVMk5bEIVihX9W5EDwPb9ngw3>
- [38] C. Olson and J. Bentley, *Field Guide to Lens Design*, vol. FG27. SPIE, 2012. [Online]. Available: <https://doi.org/10.1117/3.934997>
- [39] M. J. Riedl, “Diffraction Limit,” *SPIE*. Accessed: Sep. 29, 2024. [Online]. Available: https://spie.org/publications/spie-publication-resources/optipedia-free-optics-information/tt48_38_diffraction_limit
- [40] Ansys Learning, *The Airy Disk and Diffraction Limit — Lesson 4*, (2023). Accessed: Sep. 29, 2024. [Online Video]. Available: <https://www.youtube.com/watch?v=EbN7rNe98k4>
- [41] J. Chouinard, “What Is a Lens Optical Format? | Sensor Size vs Sensor Format,” *1stVision Inc. - Machine Vision Articles*. Accessed: Sep. 15, 2024. [Online]. Available: <https://www.1stvision.com/machine-vision-solutions/2017/08/how-does-a-lens-optical-format-relate-to-machine-vision-cameras.html>
- [42] “Image Circle,” *VS Technology*. Accessed: Sep. 15, 2024. [Online]. Available: <https://vst.co.jp/en/glossary/image-circle/>
- [43] F. C. Wippermann, E. Beckert, P. Dannberg, B. Messerschmidt, and G. Seyffert, “Disposable low-cost video endoscopes for straight and oblique viewing direction with simplified integration,” in *Design and Quality for Biomedical Technologies III*, SPIE, Feb. 2010, pp. 25–34. doi: 10.1117/12.840692.
- [44] S. Sparrold, *Optics Tutorial - 10 - Achromatic Doublets*, (Jul. 29, 2013). Accessed: Oct. 18, 2024. [Online Video]. Available: <https://www.youtube.com/watch?v=DspKhNiygDU>
- [45] “Abbe-Number,” *Edmund Optics*. Accessed: Oct. 18, 2024. [Online]. Available: <https://www.edmundoptics.com/knowledge-center/glossary/abbe-number/>

- [46] “Why Use an Achromatic Lens?” Accessed: Oct. 18, 2024. [Online]. Available: <https://www.edmundoptics.com/knowledge-center/application-notes/optics/why-use-an-achromatic-lens/>
- [47] “Meniscus lens,” VY Optoelectronics. Accessed: Oct. 18, 2024. [Online]. Available: <https://www.vyoptics.com/optical-meniscus-lens.html>
- [48] “Meniscus Lenses,” Altechna - Custom Laser Optics & Technologies. Accessed: Oct. 18, 2024. [Online]. Available: <https://www.altechna.com/products/lenses/meniscus-lenses/>
- [49] “Meniscus Lenses Selection Guide: Types, Features, Applications,” GlobalSpec. Accessed: Oct. 18, 2024. [Online]. Available: https://www.globalspec.com/learnmore/optics_optical_components/optical_components/meniscus_lenses
- [50] E. L. Dereniak and T. D. Dereniak, *Geometrical and Trigonometric Optics*, vol. 1. Cambridge University Press, 2008.
- [51] “All About Aspheric Lenses.” Accessed: Oct. 06, 2024. [Online]. Available: <https://www.edmundoptics.com/knowledge-center/application-notes/optics/all-about-aspheric-lenses/>
- [52] “Advantages of Fresnel Lenses,” Edmund Optics. Accessed: Oct. 18, 2024. [Online]. Available: <https://www.edmundoptics.com/knowledge-center/application-notes/optics/advantages-of-fresnel-lenses/>
- [53] H. Grigonis, “What is a Telephoto Lens? (Why and When you Need One),” ExpertPhotography. Accessed: Oct. 18, 2024. [Online]. Available: <https://cms.expertphotography.com/telephoto-lens-camera-equipment/>
- [54] J. E. Greivenkamp, “Elementary Optical Systems - Section 12 - Objectives.” College of Optical Sciences - The University of Arizona, 2019. [Online]. Available: <https://wp.optics.arizona.edu/jgreivenkamp/wp-content/uploads/sites/11/2019/08/502-12-Objectives.pdf>
- [55] “N-BK7 and BK7.” Accessed: Oct. 18, 2024. [Online]. Available: <https://www.mindrum.com/glass/n-bk7-and-bk7/>
- [56] M. N. Polyanskiy, “Refractiveindex.info database of optical constants,” *Sci. Data*, vol. 11, no. 1, p. 94, Jan. 2024, doi: 10.1038/s41597-023-02898-2.
- [57] “N-SF11 Plano-Convex Lenses,” Thorlabs. Accessed: Oct. 18, 2024. [Online]. Available: <https://www.thorlabs.com>
- [58] P. Sahoo, S. K. Das, and J. Paulo Davim, “3.3 Surface Finish Coatings,” in *Comprehensive Materials Finishing*, M. Hashmi, Ed., Oxford: Elsevier, 2017, pp. 38–55. doi: 10.1016/B978-0-12-803581-8.09167-0.
- [59] “An Introduction to Optical Coatings,” Edmund Optics. Accessed: Oct. 19, 2024. [Online]. Available: <https://www.edmundoptics.com/knowledge-center/application-notes/lasers/an-introduction-to-optical-coatings/>
- [60] D. Ristau, “OPTICAL COATINGS | Thin Film Optical Coatings,” in *Encyclopedia of Modern Optics*, R. D. Guenther, Ed., Oxford: Elsevier, 2005, pp. 360–369. doi: 10.1016/B0-12-369395-0/00880-0.
- [61] “Coatings, Anti-Reflective, Single Layer MgF2,” Esco Optics, Inc. Accessed: Oct. 19, 2024. [Online]. Available: <https://escooptics.com/products/coatings-anti-reflective-single-layer-mgf2>

- [62] “Comparison of Optical Aberrations.” Accessed: Oct. 19, 2024. [Online]. Available: <https://www.edmundoptics.com/knowledge-center/application-notes/optics/comparison-of-optical-aberrations/>
- [63] “Coma.” Accessed: Oct. 19, 2024. [Online]. Available: <https://site.physics.georgetown.edu/~vankeu/webtext/Aberrations/Coma.html>
- [64] N. James and G. Hollows, “Distortion,” Edmund Optics. Accessed: Oct. 19, 2024. [Online]. Available: <https://www.edmundoptics.com/knowledge-center/application-notes/imaging/distortion/>
- [65] Junyi Chai, H. Zeng, A. Li, and E. W. T. Ngai, “Deep learning in computer vision: A critical review of emerging techniques and application scenarios,” *Mach. Learn. Appl.*, vol. 6, p. 100134, Dec. 2021, doi: 10.1016/j.mlwa.2021.100134.
- [66] N. O’Mahony *et al.*, “Deep Learning vs. Traditional Computer Vision,” in *Advances in Computer Vision*, vol. 943, K. Arai and S. Kapoor, Eds., in *Advances in Intelligent Systems and Computing*, vol. 943. , Cham: Springer International Publishing, 2020, pp. 128–144. doi: 10.1007/978-3-030-17795-9_10.
- [67] Y. Jia *et al.*, “Caffe: Convolutional Architecture for Fast Feature Embedding,” in *Proceedings of the 22nd ACM international conference on Multimedia*, Orlando Florida USA: ACM, Nov. 2014, pp. 675–678. doi: 10.1145/2647868.2654889.
- [68] “7 Best Computer Vision Development Libraries in 2024,” Labellerr. Accessed: Sep. 24, 2024. [Online]. Available: <https://www.labellerr.com/blog/top-7-computer-vision-development-tools-2024/>
- [69] “BoofCV.” Accessed: Sep. 24, 2024. [Online]. Available: https://boofcv.org/index.php?title=Main_Page
- [70] “7 Best Computer Vision Libraries in Python,” ProjectPro. Accessed: Sep. 24, 2024. [Online]. Available: <https://www.projectpro.io/article/computer-vision-libraries/772>
- [71] “SimpleCV Tutorial — Tutorial.” Accessed: Sep. 24, 2024. [Online]. Available: <https://tutorial.simplecv.org/en/latest/>
- [72] “About,” OpenCV. Accessed: Sep. 25, 2024. [Online]. Available: <https://opencv.org/about/>
- [73] “OpenCV: Introduction to OpenCV-Python Tutorials.” Accessed: Sep. 25, 2024. [Online]. Available: https://docs.opencv.org/4.x/d0/de3/tutorial_py_intro.html
- [74] “Body Detection with Computer Vision | Latest | Instrument.” Accessed: Sep. 26, 2024. [Online]. Available: <https://www.instrument.com/latest/body-detection-with-computer-vision>
- [75] A. Tam, “Using Haar Cascade for Object Detection,” *MachineLearningMastery.com*. Accessed: Sep. 26, 2024. [Online]. Available: <https://machinelearningmastery.com/using-haar-cascade-for-object-detection/>
- [76] A. Rosebrock, “OpenCV Haar Cascades,” PyImageSearch. Accessed: Sep. 26, 2024. [Online]. Available: <https://pyimagesearch.com/2021/04/12/opencv-haar-cascades/>
- [77] 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)*, San Diego, CA, USA: IEEE, 2005, pp. 886–893. doi: 10.1109/CVPR.2005.177.

- [78] A. Drake, "Unveiling the Potential of Histogram of Oriented Gradients (HOG) in Computer Vision," Comet. Accessed: Oct. 06, 2024. [Online]. Available: <https://www.comet.com/site/blog/unveiling-the-potential-of-histogram-of-oriented-gradients-hog-in-computer-vision>
- [79] A. Tam, "Extracting Histogram of Gradients with OpenCV," MachineLearningMastery.com. Accessed: Oct. 06, 2024. [Online]. Available: https://machinelearningmastery.com/opencv_hog/
- [80] S. AI, "Histogram of Oriented Gradients (HOG) — Simplest Intuition," Medium. Accessed: Oct. 06, 2024. [Online]. Available: <https://medium.com/@skillcate/histogram-of-oriented-gradients-hog-simplest-intuition-2392995f8010>
- [81] "Histogram of Oriented Gradients explained using OpenCV." Accessed: Oct. 06, 2024. [Online]. Available: <https://learnopencv.com/histogram-of-oriented-gradients/>
- [82] A. Mirkhan, "YOLO Algorithm: Real-Time Object Detection from A to Z," Kili Technology. Accessed: Oct. 13, 2024. [Online]. Available: <https://kili-technology.com/data-labeling/machine-learning/yolo-algorithm-real-time-object-detection-from-a-to-z>
- [83] Melanie, "You Only Look Once (YOLO): What is it?," Data Science Courses | DataScientest. Accessed: Oct. 13, 2024. [Online]. Available: <https://datascientest.com/en/you-only-look-once-yolo-what-is-it>
- [84] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," Feb. 29, 2024, *arXiv*: arXiv:2402.13616. Accessed: Oct. 15, 2024. [Online]. Available: <http://arxiv.org/abs/2402.13616>
- [85] Ultralytics, "YOLOv9." Accessed: Oct. 15, 2024. [Online]. Available: <https://docs.ultralytics.com/models/yolov9>
- [86] "YOLOv9: Advancing the YOLO Legacy." Accessed: Oct. 15, 2024. [Online]. Available: <https://learnopencv.com/yolov9-advancing-the-yolo-legacy/>
- [87] "YOLOv10: Real-Time End-to-End Object Detection." Accessed: Oct. 15, 2024. [Online]. Available: <https://arxiv.org/html/2405.14458v1>
- [88] "YOLOv10: Paper Explanation and Inference Results." Accessed: Oct. 15, 2024. [Online]. Available: <https://learnopencv.com/yolov10/>
- [89] Saiwa, "YOLOv10 | explanation, features, and implementation," Medium. Accessed: Oct. 15, 2024. [Online]. Available: <https://medium.com/@saiwadotai/yolov10-explanation-features-and-implementation-c67b49e44120>
- [90] S. Arya, "YOLOv10: Revolutionizing Real-Time Object Detection," Analytics Vidhya. Accessed: Nov. 03, 2024. [Online]. Available: <https://www.analyticsvidhya.com/blog/2024/07/yolov10-for-realtime-object-detection/>
- [91] Ultralytics, "YOLO11 🚀 NEW." Accessed: Oct. 17, 2024. [Online]. Available: <https://docs.ultralytics.com/models/yolo11>
- [92] "YOLO11: Faster Than You Can Imagine!" Accessed: Oct. 17, 2024. [Online]. Available: <https://learnopencv.com/yolo11/>
- [93] Z. Shariff, "YOLOv11 Is Officially Out! What You Need To Know!," Medium. Accessed: Oct. 17, 2024. [Online]. Available:

- <https://medium.com/@zainshariff6506/yolov11-is-officially-out-what-you-need-to-know-6738c5d25be1>
- [94] N. Singh, "Meet YOLO-NAS: An Open-Sourced YOLO-based Architecture Redefining State-of-the-Art in Object Detection," MarkTechPost. Accessed: Oct. 18, 2024. [Online]. Available: <https://www.marktechpost.com/2023/05/08/meet-yolo-nas-an-open-sourced-yolo-based-architecture-redefining-state-of-the-art-in-object-detection/>
 - [95] Ultralytics, "YOLO-NAS (Neural Architecture Search)." Accessed: Oct. 18, 2024. [Online]. Available: <https://docs.ultralytics.com/models/yolo-nas>
 - [96] "YOLO-NAS: New Object Detection Model Beats YOLOv6 & YOLOv8." Accessed: Oct. 18, 2024. [Online]. Available: <https://learnopencv.com/yolo-nas/>
 - [97] R. Vaidya, "Deci's YOLO-NAS: Next-Generation Model for Object Detection.," CodeX. Accessed: Oct. 18, 2024. [Online]. Available: <https://medium.com/codex/decis-yolo-nas-next-generation-model-for-object-detection-8ccb7f2013a7>
 - [98] A. S. Gillis, "What is a codec?," Unified Communications. Accessed: Oct. 22, 2024. [Online]. Available: <https://www.techtarget.com/searchunifiedcommunications/definition/codec>
 - [99] "Lossy vs Lossless Compression: Differences & Advantages | Adobe." Accessed: Oct. 22, 2024. [Online]. Available: <https://www.adobe.com/uk/creativecloud/photography/discover/lossy-vs-lossless.html>
 - [100] Z. Powell, "Lossy vs Lossless Compression: Comprehensive Analysis - ShortPixel Blog." Accessed: Oct. 23, 2024. [Online]. Available: <https://shortpixel.com/blog/lossy-vs-lossless/>
 - [101] D. Molenhouse, "Understanding the Difference between Interframe & Intraframe Compression." Accessed: Oct. 23, 2024. [Online]. Available: <https://info.verkada.com/Interframe-vs-Intraframe-Compression/>
 - [102] ~ Prateek Joshi, "What Is Entropy Coding?," Perpetual Enigma. Accessed: Oct. 23, 2024. [Online]. Available: <https://prateekvjoshi.com/2014/12/06/what-is-entropy-coding/>
 - [103] "Network Bandwidth: What Is Bandwidth in Networking? - IT Glossary | SolarWinds." Accessed: Oct. 23, 2024. [Online]. Available: <https://www.solarwinds.com/resources/it-glossary/network-bandwidth>
 - [104] "Video Codecs - List of the best codecs and how they work." Accessed: Oct. 23, 2024. [Online]. Available: <https://getstream.io/glossary/video-codecs/>
 - [105] "FFV1," *Wikipedia*. Sep. 06, 2024. Accessed: Oct. 23, 2024. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=FFV1&oldid=1244343241>
 - [106] "Pulse Code Modulation - Modulation, Types, Advantages and Disadvantages, Applications," BYJUS. Accessed: Oct. 23, 2024. [Online]. Available: <https://byjus.com/physics/pulse-code-modulation/>
 - [107] "Pulse Code Modulation - an overview | ScienceDirect Topics." Accessed: Oct. 23, 2024. [Online]. Available: <https://www.sciencedirect.com/topics/engineering/pulse-code-modulation>
 - [108] "Modulation Basics – Wavelength Electronics." Accessed: Oct. 23, 2024. [Online]. Available: <https://www.teamwavelength.com/modulation-basics/>

- [109] “Error Correcting Codes - Reed-Solomon codes.” Accessed: Oct. 23, 2024. [Online]. Available: <https://www.tutorialspoint.com/error-correcting-codes-reed-solomon-codes>
- [110] “Error Correcting Codes - Hamming codes.” Accessed: Oct. 23, 2024. [Online]. Available: <https://www.tutorialspoint.com/error-correcting-codes-hamming-codes>
- [111] “About FFmpeg.” Accessed: Oct. 23, 2024. [Online]. Available: <https://ffmpeg.org/about.html>
- [112] “GStreamer: features.” Accessed: Oct. 23, 2024. [Online]. Available: <https://gstreamer.freedesktop.org/features/>
- [113] “OpenH264.” Accessed: Oct. 23, 2024. [Online]. Available: <https://www.openh264.org/>
- [114] “JCodec -.” Accessed: Oct. 23, 2024. [Online]. Available: <http://jcodec.org/>
- [115] B. Louis, “Xuggler: The Easy Way to Work with Media Files in Java,” Xu 2Gle. Accessed: Oct. 23, 2024. [Online]. Available: <https://www.xuggle.com/xuggler/>
- [116] “Java Media Framework.” Accessed: Oct. 23, 2024. [Online]. Available: <https://www.oracle.com/java/technologies/javase/java-media-framework.html>
- [117] “PyAV Documentation — PyAV 9.0.2 documentation.” Accessed: Oct. 23, 2024. [Online]. Available: <https://pyav.org/docs/stable/>
- [118] *moviepy: Video editing with Python*. Python. Accessed: Oct. 23, 2024. [Online]. Available: <https://zulko.github.io/moviepy/>
- [119] “ESP SoCs | Espressif Systems.” Accessed: Oct. 23, 2024. [Online]. Available: <https://www.espressif.com/en/products/socs>
- [120] “esp32-s3_datasheet_en.” Espressif. [Online]. Available: https://www.espressif.com/sites/default/files/documentation/esp32-s3_datasheet_en.pdf
- [121] “ESP32-S3,” DigiKey Electronics. Accessed: Oct. 23, 2024. [Online]. Available: <https://www.digikey.com/en/products/detail/espressif-systems/ESP32-S3/15822445>
- [122] “STM32H7 - Arm Cortex-M7 and Cortex-M4 MCUs (480 MHz) - STMicroelectronics.” Accessed: Oct. 23, 2024. [Online]. Available: <https://www.st.com/en/microcontrollers-microprocessors/stm32h7-series.html>
- [123] “Buy STM32H7S3L8 - eStore - STMicroelectronics.” Accessed: Oct. 23, 2024. [Online]. Available: <https://estore.st.com/en/products/microcontrollers-microprocessors/stm32-32-bit-arm-cortex-mcus/stm32-high-performance-mcus/stm32h7-series/stm32h7r3-7s3/stm32h7s3l8.html>
- [124] R. P. Ltd, “RP2040 specifications,” Raspberry Pi. Accessed: Oct. 23, 2024. [Online]. Available: <https://www.raspberrypi.com/products/rp2040/>
- [125] “Raspberry Pi RP2040 Microcontroller,” CanaKit. Accessed: Oct. 23, 2024. [Online]. Available: <https://www.canakit.com/raspberry-pi-rp2040.html>
- [126] R. P. Ltd, “Raspberry Pi 4 Model B specifications,” Raspberry Pi. Accessed: Oct. 23, 2024. [Online]. Available: <https://www.raspberrypi.com/products/raspberry-pi-4-model-b/specifications/>
- [127] “raspberrypi-4-product-brief.” Raspberry Pi. [Online]. Available: <https://datasheets.raspberrypi.com/rpi4/raspberry-pi-4-product-brief.pdf>
- [128] “Raspberry Pi 4 1GB Model B,” CanaKit. Accessed: Oct. 23, 2024. [Online]. Available: <https://www.canakit.com/raspberry-pi-4.html>

- [129] R. P. Ltd, "Buy a Raspberry Pi Zero 2 W," Raspberry Pi. Accessed: Oct. 23, 2024. [Online]. Available: <https://www.raspberrypi.com/products/raspberry-pi-zero-2-w/>
- [130] "raspberrypi-zero-2-w-product-brief." Raspberry Pi. [Online]. Available: <https://datasheets.raspberrypi.com/rpizero2/raspberry-pi-zero-2-w-product-brief.pdf>
- [131] "Raspberry Pi Zero 2 W," CanaKit. Accessed: Oct. 23, 2024. [Online]. Available: <https://www.canakit.com/raspberry-pi-zero-2-w.html>
- [132] R. P. Ltd, "Buy a Raspberry Pi 3 Model B+," Raspberry Pi. Accessed: Oct. 23, 2024. [Online]. Available: <https://www.raspberrypi.com/products/raspberry-pi-3-model-b-plus/>
- [133] "raspberrypi-3-b-plus-product-brief." Raspberry Pi. [Online]. Available: <https://datasheets.raspberrypi.com/rpi3/raspberry-pi-3-b-plus-product-brief.pdf>
- [134] "Raspberry Pi 3 Model B+ 1.4 GHz," Chicago Electronic Distributors. Accessed: Oct. 23, 2024. [Online]. Available: <https://chicagodist.com/products/raspberry-pi-model-3-b-1-4-ghz>
- [135] "One-way communication: Definition and applications." Accessed: Oct. 24, 2024. [Online]. Available: <https://www.oneteam.io/en/blog/what-is-one-way-communication>
- [136] "Communication; principles for a lifetime," *Ref. Res. Book News*, vol. 16, no. 3, 2001.
- [137] "One-way Communication: Definition, Examples, Benefits & Process," Clearinfo. Accessed: Oct. 24, 2024. [Online]. Available: <https://clearinfo.in/blog/what-is-one-way-communication/>
- [138] "What is two-way communication? Importance of two-way communication." Accessed: Oct. 24, 2024. [Online]. Available: <https://thebusinesscommunication.com/two-way-communication-and-importance/>
- [139] A. removed at request of original Publisher, "1.2 The Communication Process," Sep. 2016, Accessed: Oct. 24, 2024. [Online]. Available: <https://open.lib.umn.edu/communication/chapter/1-2-the-communication-process/>
- [140] "Everything You Need to Know About Coaxial Cable | Lenovo US." Accessed: Oct. 24, 2024. [Online]. Available: <https://www.lenovo.com/us/en/glossary/coaxial-cable/>
- [141] "Oliver Heaviside and the coaxial cable," Waveform. Accessed: Oct. 24, 2024. [Online]. Available: <https://www.waveform.com/blogs/main/146807175-oliver-heaviside-and-the-coaxial-cable>
- [142] "HeavisideUKPatent1407.pdf." Accessed: Oct. 24, 2024. [Online]. Available: <https://cdn.shopify.com/s/files/1/0358/5537/files/HeavisideUKPatent1407.pdf?3837882249986977540>
- [143] "What is radio frequency (RF, rf)? - Definition from WhatIs.com," Networking. Accessed: Oct. 24, 2024. [Online]. Available: <https://www.techtarget.com/searchnetworking/definition/radio-frequency>
- [144] "RF cable and RF wireless transmission: how does it work?" Accessed: Oct. 24, 2024. [Online]. Available: <https://www.microcontrollertips.com/rf-wireless-rf-cable-transmission/>

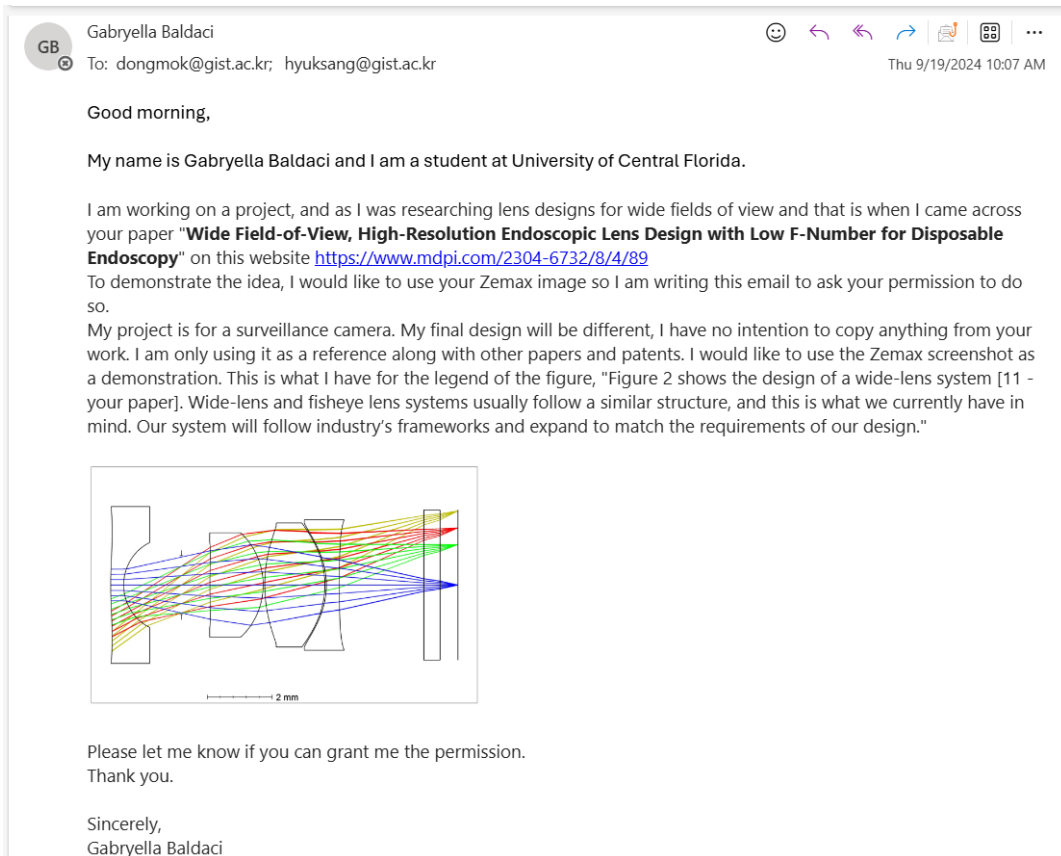
- [145] “What is an optical module? Optical module wiki | FS Community,” Knowledge. Accessed: Oct. 24, 2024. [Online]. Available: <https://community.fs.com/article/what-is-an-optical-module.html>
- [146] “Optical Transceivers.” Accessed: Oct. 24, 2024. [Online]. Available: <https://www.broadcom.com/products/fiber-optic-modules-components/networking/optical-transceivers>
- [147] “What is an SFP? All things optical transceivers - YouTube.” Accessed: Oct. 24, 2024. [Online]. Available: <https://www.youtube.com/watch?v=WqTNdUatOho>
- [148] “1.25 GBd MMF Transceiver with DMI for Gigabit Ethernet, SFP, MSA, AFBR-5718PZ.” Accessed: Oct. 24, 2024. [Online]. Available: <https://www.broadcom.com/products/fiber-optic-modules-components/networking/optical-transceivers/sfp/afbr-5718pz>
- [149] “The Most Comprehensive Guide Of Optical Modules - Asterfusion Data Technologies.” Accessed: Oct. 24, 2024. [Online]. Available: <https://cloudswit.ch/blogs/the-most-comprehensive-guide-of-optical-modules/>
- [150] “Optical Transceivers.” Accessed: Oct. 24, 2024. [Online]. Available: <https://www.amphenol-cs.com/optics/transceivers.html>
- [151] “Optical transceivers - turning data into light,” Smartoptics. Accessed: Oct. 24, 2024. [Online]. Available: <https://smartoptics.com/knowledgebank-post/optical-transceivers-turning-data-into-light/>
- [152] C. Bunce, “Transceiver Form Factors: The History,” Pro Optix. Accessed: Oct. 24, 2024. [Online]. Available: <https://www.prooptix.com/news/transceiver-form-factors/>
- [153] C. Bunce, “Fiber Optic Lasers: Understanding Lasers in Optical Transceivers,” Pro Optix. Accessed: Oct. 24, 2024. [Online]. Available: <https://www.prooptix.com/news/fiber-optic-lasers/>
- [154] “101 Series: LEDs vs. VCSELs,” Fluke Networks. Accessed: Oct. 24, 2024. [Online]. Available: <https://www.flukenetworks.com/blog/cabling-chronicles/101-series-leds-vs-vcsls>
- [155] “Deep Analysis on Optical Transceiver Module | FS Community,” Knowledge. Accessed: Oct. 24, 2024. [Online]. Available: <https://community.fs.com/article/deep-analysis-on-optical-transceiver-module.html>
- [156] “Optical Communication Band - FiberLabs Inc.,” FiberLabs Inc. - Specialty Fiber Amplifiers and Fluoride Fibers. Accessed: Oct. 24, 2024. [Online]. Available: <https://www.fiberlabs.com/glossary/optical-communication-band/>
- [157] “Coherent Modulation vs. PAM4 in 800G Optical Transmission | FS Community,” Knowledge. Accessed: Oct. 24, 2024. [Online]. Available: <https://community.fs.com/article/coherent-modulation-vs-pam4-in-800g-optical-transmission.html>
- [158] “1.1. NRZ Fundamentals,” Intel. Accessed: Oct. 24, 2024. [Online]. Available: <https://www.intel.com/content/www/us/en/docs/programmable/683852/current/nrz-fundamentals.html>
- [159] “NRZ Vs. PAM4 Modulation Techniques | FS Community,” Knowledge. Accessed: Oct. 24, 2024. [Online]. Available: <https://community.fs.com/article/nrz-vs-pam4-modulation-techniques.html>

- [160] Z. Xia, “What Is QAM? How Does QAM Work?,” Huawei. Accessed: Oct. 24, 2024. [Online]. Available: <https://info.support.huawei.com/info-finder/encyclopedia/en/QAM.html>
- [161] “Coherent Modulation vs. PAM4 in 800G Optical Transmission | FS Community,” Knowledge. Accessed: Oct. 24, 2024. [Online]. Available: <https://community.fs.com/article/coherent-modulation-vs-pam4-in-800g-optical-transmission.html>
- [162] “Cisco Transceiver Modules - Use of Mode Conditioning Patch Cables in Gigabit Ethernet and 10 Gigabit Ethernet Laser-Based Transmissions,” Cisco. Accessed: Oct. 24, 2024. [Online]. Available: https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/transceiver-modules/product_bulletin_c25-530836.html
- [163] Taylor Scully, “How to Choose a Mean Well Power Supply,” How to Choose a Mean Well Power Supply: LED Driver and Power Supply Guide. [Online]. Available: <https://www.ledsupply.com/blog/meanwell-power-supply-guide/>
- [164] Mouser Electronics, “LRS-50-5.” [Online]. Available: <https://www.mouser.com/ProductDetail/MEAN-WELL/LRS-50-5?qs=vDxCgdWo2h9hVrbBDwCVQA%3D%3D>
- [165] Mean Well, “LRS-50-5 Datasheet.” [Online]. Available: https://www.farnell.com/datasheets/2547982.pdf?_gl=1*1qk8ied*_gcl_au*MjM2NzY4MDI3LjE3Mjk1NjkwMjA
- [166] Mean Well, “RS-15-5.” [Online]. Available: <https://us.rs-online.com/product/mean-well/rs-15-5/70069559/>
- [167] RS Pro, “RS Pro 5V DC Power Supply.” [Online]. Available: <https://us.rs-online.com/product/rs-pro/1618272/71291688/>
- [168] CUI Inc., “CUI SWI5-N Datasheet.” Sep. 27, 2023. [Online]. Available: <https://www.cui.com/product/resource/swi5-n.pdf>
- [169] Poulomi Ghosh, “Applying IPC-2221 Standards in Circuit Board Design.” [Online]. Available: <https://www.protoexpress.com/blog/ipc-2221-circuit-board-design/>
- [170] “EN IEC 62368 Testing,” D.L.S. Electronic Systems, Inc. Accessed: Oct. 24, 2024. [Online]. Available: <https://www.dlsemc.com/en-62368-testing/>
- [171] CUI Inc., “Efficiency Standards for External Power Supplies,” Standards. [Online]. Available: <https://www.cui.com/efficiency-standards>
- [172] Alex Deng, “Technical Compliance Statement CE EMC Test Report,” Mean Well, Test Report C1M2106134. [Online]. Available: <https://www.meanwell.com/Upload/PDF/UHP-1500/UHP-1500-EMC-cer.pdf>
- [173] “Quality | Edmund Optics.” Accessed: Oct. 19, 2024. [Online]. Available: <https://www.edmundoptics.com/capabilities/quality/>
- [174] “ISO 9001:2015,” ISO. Accessed: Oct. 19, 2024. [Online]. Available: <https://www.iso.org/standard/62085.html>
- [175] J. Lane, “Tutorial on ISO 10110 Optical Drawing Standard OPTI 521 – Intro to Opto-Mechanical Engineering.” The University of Arizona, Dec. 13, 2009. [Online]. Available: <https://wp.optics.arizona.edu/optomech/wp-content/uploads/sites/53/2016/10/Tutorial-on-ISO-10110-Optical-Drawing-Standards.pdf>

- [176] “Custom Optics Manufacturing and OEM Ordering.” Accessed: Oct. 19, 2024. [Online]. Available: <https://www.thorlabs.com>
- [177] “Introduction to Optical Transceiver MSA Standards | FS Community,” Knowledge. Accessed: Oct. 24, 2024. [Online]. Available: <https://community.fs.com/article/how-much-do-you-know-about-msa-standard.html>
- [178] “Security & Surveillance,” Assent Risk Management. Accessed: Oct. 20, 2024. [Online]. Available: <https://www.assentriskmanagement.co.uk/industries/iso-certification-for-security-surveillance-organisations/>
- [179] “ISO 27001 Annex A Controls Archives,” High Table. Accessed: Oct. 20, 2024. [Online]. Available: <https://hightable.io/category/iso27001-annex-a-controls/>
- [180] S. Barker, “The complete guide to ISO/IEC 27002:2022,” High Table. Accessed: Oct. 20, 2024. [Online]. Available: <https://hightable.io/the-ultimate-guide-to-iso-27002-changes-2022/>
- [181] “ISO 22301:2019,” ISO. Accessed: Oct. 20, 2024. [Online]. Available: <https://www.iso.org/standard/75106.html>
- [182] “IEC 62676-4 Ed. 1.0 b:2014 - Video surveillance systems for use in security applications - Part 4: Application guidelines.” Accessed: Oct. 20, 2024. [Online]. Available: <https://webstore.ansi.org/standards/iec/iec62676ed2014>
- [183] “Is It Legal for Security Cameras to Record Audio?,” Brinks HomeTM. Accessed: Oct. 20, 2024. [Online]. Available: <https://brinkshome.com/smartcenter/are-security-cameras-allowed-to-record-audio>
- [184] R. Edwards, “Security Camera Laws, Rights, and Rules,” SafeWise. Accessed: Oct. 20, 2024. [Online]. Available: <https://www.safewise.com/security-camera-laws/>

Appendix B – Copyright Permission

Most of the figures and tables in this document were made by the authors themselves. Permission was requested to use Figure 2.3



Appendix C – Data Sheet

1. Camera Module: Raspberry Pi Camera Module 3

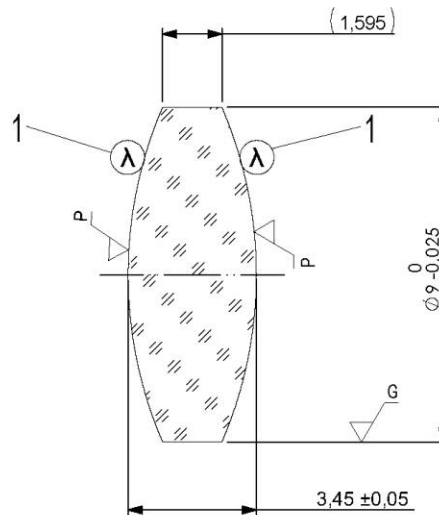
Specification

Sensor:	Sony IMX708
Resolution:	11.9 megapixels
Sensor size:	7.4mm sensor diagonal
Pixel size:	1.4µm × 1.4µm
Horizontal/vertical:	4608 × 2592 pixels
Common video modes:	1080p50, 720p100, 480p120
Output:	RAW10
IR cut filter:	Integrated in standard variants; not present in NoIR variants
Autofocus system:	Phase Detection Autofocus
Dimensions:	25 × 24 × 11.5mm (12.4mm height for Wide variants)
Ribbon cable length:	200mm
Cable connector:	15 × 1mm FPC
Operating temperature:	0°C to 50°C
Compliance:	FCC 47 CFR Part 15, Subpart B, Class B Digital Device Electromagnetic Compatibility Directive (EMC) 2014/30/EU Restriction of Hazardous Substances (RoHS) Directive 2011/65/EU
Production lifetime:	Raspberry Pi Camera Module 3 will remain in production until at least January 2030

Variants

	Camera Module 3	Camera Module 3 NoIR	Camera Module 3 Wide	Camera Module 3 Wide NoIR
Focus range	10cm–∞	10cm–∞	5cm–∞	5cm–∞
Focal length	4.74mm	4.74mm	2.75mm	2.75mm
Diagonal field of view	75 degrees	75 degrees	120 degrees	120 degrees
Horizontal field of view	66 degrees	66 degrees	102 degrees	102 degrees
Vertical field of view	41 degrees	41 degrees	67 degrees	67 degrees
Focal ratio (F-stop)	F1.8	F1.8	F2.2	F2.2
Infrared-sensitive	No	Yes	No	Yes

Appendix D – ISO 10110 Drawing



FOCAL LENGTH (EFL): $9 \pm 0,09$
 BACK FOCAL LENGTH (BFL): 7,9

1 AR COATING:

$\frac{1}{4}$ WAVE MgF2 @ 550nm
 $R(\text{AVG}) < 1,75\%$ FROM 400-700nm (N-BK7)

LEFT SURFACE		MATERIAL	RIGHT SURFACE	
R	11,38±0,011 CX	GLASS N-SF5 673/322	R	11,38±0,011 CX
Øe	8,1		Øe	8,1
3/	3 (0,5) $\lambda=632,8$ nm		3/	3 (0,5) $\lambda=632,8$ nm
4/	-		4/	4,46'
5/	40 - 20 (MIL-PRF-13830B)		5/	40 - 20 (MIL-PRF-13830B)
6/	-		6/	-

PROTECTIVE CHAMFERS AS NEEDED			FIRST ANGLE PROJECTION		 Edmund Optics®			
ROHS: COMPLIANT	SCALE: NONE	$\lambda = 587,6\text{nm}$						
SPECIFICATIONS SUBJECT TO CHANGE WITHOUT NOTICE. DIMENSIONS ARE FOR REFERENCE ONLY					TITLE 9mm Dia, x 9mm FL, MgF2 Coated, Double-Convex Lens			
FOR INFORMATION ONLY: DO NOT MANUFACTURE PARTS TO THIS DRAWING								
INDICATIONS IN ACCORDANCE WITH ISO 10110			ALL DIMS IN:	mm	DWG NO 32019		SHEET SIZE A4	SHEET 1 OF 1

Appendix E – Software Code

Appendix F – ChatGPT prompts and outcomes

Case Study 1:

Prompt 1

I am designing a surveillance camera with a wide field of view. This camera will be placed inside a well-lit place or used for daytime surveillance. I will design a lens system/assembly for it. However, before doing so I need to find a camera sensor that I will eventually mount the lens on it. What are some aspects I should take into account when choosing a sensor for a camera that will transmit live video to my surveillance system?

ChatGPT Outcome 1

When choosing a camera sensor for a surveillance camera that will transmit live video with a wide field of view in a well-lit or daytime environment, here are some key aspects to consider:

1. Sensor Size:

- The sensor size will directly influence the field of view and the depth of field. A larger sensor will capture more of the lens's image circle, providing a wider field of view for a given focal length.
- Common sensor sizes to consider:
 - 1/2.7-inch or larger (e.g., 1/1.8-inch) for high-quality imaging.
 - Full-frame sensors (if budget and space allow) can provide excellent image quality and flexibility but are often more expensive and larger in physical size.
 - APS-C sensors can offer a balance between size and cost, with a moderate crop factor (~1.5x).

2. Resolution:

- Choose a resolution that meets your surveillance needs while ensuring smooth video transmission.
- HD (1080p) or 4K resolution is common for modern surveillance cameras.
- 4K resolution offers more detail and is ideal for wide-area monitoring, but it requires more bandwidth for live transmission.

- 1080p may be sufficient if high detail isn't critical, and it requires less bandwidth.

3. Frame Rate:

- For smooth live video, aim for a frame rate of 30 fps or higher.
- If you're dealing with fast-moving objects, 60 fps or more can provide smoother motion capture.
- A higher frame rate increases the bandwidth requirements but results in better video quality for fast-moving subjects.

4. Dynamic Range:

- A sensor with a good dynamic range is important, especially if the camera will face varying lighting conditions (e.g., shadows, bright spots in the scene). Wide dynamic range (WDR) capabilities can handle such lighting variations, ensuring details in both bright and dark areas are captured well.
- In a well-lit environment, WDR can still be useful to prevent overexposure or underexposure in certain parts of the image.

5. Low-Light Performance:

- While your design is intended for daytime or well-lit environments, you should still consider a sensor with reasonable low-light performance in case lighting conditions vary or for occasional use in dimmer settings.
- Low-light sensors with large pixels (i.e., lower resolution for a given sensor size) capture more light, resulting in better image quality in low-light situations.

6. Shutter Type:

- Consider using a global shutter if your camera will be capturing fast-moving objects, as it eliminates the rolling-shutter effect (which can cause motion distortion).
- If motion artifacts are not a major concern, a rolling shutter may suffice, and sensors with rolling shutters tend to be more affordable.

7. Data Interface:

- Ensure the sensor supports a data interface that is compatible with your transmission system (e.g., MIPI CSI, USB, or Ethernet).
- Since this is for live video transmission, consider the bandwidth requirements. Higher resolutions and frame rates demand faster data interfaces.

8. Power Consumption:

- Low power consumption is critical if the camera is running continuously or if you have specific power constraints. Sensors with lower power consumption will generate less heat and are ideal for a system that operates 24/7.

9. Lens Compatibility:

- The sensor must be compatible with the lens system you're designing, particularly in terms of mount type (e.g., C-mount, M12), image circle, and field of view.

- Sensor size impacts how the lens will project the image, so consider both the focal length of your lens and the sensor size to achieve the required FOV.

10. Latency:

- For live video, low latency is crucial to ensure real-time video streaming. The sensor, along with the camera's processing pipeline, should support low-latency video output.

- Hardware video encoding (e.g., H.264 or H.265) directly on the sensor or module can reduce the delay in live transmission.

11. Compression and Bandwidth:

- Consider a sensor that supports onboard compression (such as H.264, H.265) to reduce the data bandwidth needed for live transmission. This helps with smoother video streams without sacrificing too much image quality.

12. Environmental Durability:

- Ensure the sensor is built to handle environmental factors, especially if the camera may be exposed to temperature fluctuations, dust, or vibrations in an indoor environment.

- A sensor that can function in a range of temperatures may be beneficial, especially for areas that can experience heat from lighting or equipment.

13. Cost:

- Sensor costs can vary significantly depending on their size, resolution, and additional features like WDR or low-light capabilities.

- Balancing cost and performance is key—ensure the sensor meets your performance requirements for field of view, resolution, and frame rate while staying within budget.

By considering these factors: size, resolution, frame rate, dynamic range, data interface, and power consumption—you can choose a sensor that not only fits your design needs but also ensures optimal performance for live video transmission in your surveillance system.

Copilot Outcome 1

When choosing a sensor for your surveillance camera, especially for a well-lit, daytime environment, there are several key aspects to consider:

Sensor Size and Resolution: Larger sensors generally capture more light, resulting in better image quality.

Field of View (FoV): Since you need a wide field of view, ensure the sensor can support lenses that offer a wide-angle perspective. [This is crucial for covering large areas with fewer cameras².](#)

Frame Rate: For live video transmission, a higher frame rate (e.g., 30 fps or higher) ensures smooth video playback, which is important for real-time monitoring¹.

Dynamic Range: A sensor with a good dynamic range can handle varying lighting conditions within the scene, ensuring that both bright and shadowed areas are captured clearly¹.

Lens Compatibility: Make sure the sensor is compatible with the lens system you plan to design. [The lens and sensor should be well-matched to maximize image quality¹.](#)

Connectivity and Compression: Consider how the sensor handles video compression and transmission. [Efficient compression \(e.g., H.264 or H.265\) reduces bandwidth usage without significantly compromising image quality³.](#)

Power Consumption: For continuous operation, choose a sensor with low power consumption to ensure reliability and reduce operational costs¹.

Cost and Availability: Finally, balance your budget with the sensor's features and availability. [Sometimes, the latest sensors might offer advanced features but at a higher cost¹.](#)

By considering these factors, you can select a sensor that meets your surveillance needs and integrates well with your lens design. If you have any specific requirements or need further assistance, feel free to ask!

Case Study 2:

Prompt 2

I want to build a lens system that will be mounted on a CMOS sensor. I want my camera to have at least 80 degree field of view, I want to minimize aberrations. Can you give me ideas for the design? How many lenses should I use? Which type of lenses should I use?

ChatGPT Outcome 2

To design a lens system with at least 80-degree field of view (FOV) for a camera mounted on a CMOS sensor, while minimizing aberrations, you need a multi-lens design. Here's a detailed guide on how to approach it:

Key Design Considerations:

1. **Wide Field of View (80 degrees):** Achieving a wide FOV requires short focal lengths relative to the sensor size, and the lens design must manage distortion and field curvature.
2. **Minimizing Aberrations:** To reduce spherical, chromatic, and coma aberrations, a multi-lens approach is necessary. You'll likely need at least 3-5 lens elements.
3. **CMOS Sensor Dimensions:** The sensor size determines the required focal length for the desired FOV. If your sensor size is 6.45mm x 3.43mm (diagonal 7.4mm), the focal length needs to be short, around 3-4mm for a 80-degree FOV.

Recommended Design Outline:

1. Use a Retrofocus Design (Reverse Telephoto)

- For wide-angle lenses, a retrofocus design works well. This is a design where the back focal length is longer than the effective focal length, which is essential for maintaining a wide FOV while allowing space for the sensor and camera body.
- This approach helps control aberrations, especially off-axis aberrations like coma and astigmatism.

2. Number of Lenses: 3 to 5 Elements

- Three-element system: This would be a minimal design where you can use two positive lenses and one negative lens to correct for spherical aberrations and chromatic aberrations. However, further optimization might be required for edge-to-edge sharpness.
- Four to five-element system: This will give you more control over aberrations, especially field curvature, distortion, and chromatic aberrations.

3. Types of Lenses:

- **Achromatic Doublets:** Use at least one or two achromatic doublets to correct chromatic aberrations. These are pairs of lenses made from different glass types to reduce the focus shift between different wavelengths of light (reduce color fringing).
- **Meniscus Lenses:** Positive or negative meniscus lenses can help in shaping the field curvature while controlling distortion. For a wide-angle system, positive meniscus lenses are often placed at the front.
- **Plano-Convex/Concave Lenses:** These lenses can be used to manage spherical aberrations, especially in the central region of the lens.

- Symmetry: Symmetric or near-symmetric lens designs (using two positive lenses with a negative lens in between) help minimize distortion and reduce coma.

4. Lens Layout Suggestion:

A 5-element design could look like this:

- First element: A positive meniscus lens near the object side to control distortion.
- Second element: A negative lens (concave) to help manage field curvature and reduce the system's effective focal length.
- Third and fourth elements: A pair of achromatic doublets to handle chromatic aberration and further reduce spherical aberrations.
- Fifth element: Another positive element to refine focus and manage edge sharpness.

5. Coatings and Glass Types:

- Use anti-reflective coatings to reduce ghosting and flare, especially since wide-angle systems are prone to internal reflections.
- Choose glass materials that balance cost and performance, such as N-BK7 for positive lenses and N-F2 or SF5 for the negative lens.

6. Lens Diameter:

- Given the field of view and focal length, your lens elements need to have diameters large enough to capture the full field. Depending on the final design, the front elements could be between 7mm and 9mm in diameter.

Optimizing the Design:

- Minimize Distortion: Since wide-angle lenses are prone to barrel distortion, incorporate lens elements that correct this, like the meniscus lenses in the early part of the design.
- Minimize Chromatic Aberration: Use achromatic doublets, particularly near the middle of the system, to ensure that the light of different wavelengths focuses at the same point.
- Minimize Field Curvature: Use lenses with appropriately designed curvatures to correct field curvature, ensuring the image remains sharp from the center to the edge.

Example Configuration (for Guidance):

1. First Element: Positive meniscus lens (convex toward the object).
2. Second Element: Negative plano-concave lens (controls focal length and field curvature).
3. Third Element: Achromatic doublet (for chromatic aberration correction).

4. Fourth Element: Another achromatic doublet (to refine aberration corrections and image quality).

5. Fifth Element: Plano-convex lens (final focusing element).

Final Notes:

- Start with ray tracing in Zemax to adjust the radii of curvatures, thicknesses, and spacings between lenses to achieve the desired FOV and minimize aberrations.

- Optimize your design iteratively, using the Merit Function Editor in Zemax, focusing on minimizing chromatic aberration, spherical aberration, and distortion.

This design approach should give you a solid starting point for a high-performance lens system for your CMOS sensor with a > 80 -degree field of view.

Case Study 3:

Prompt 3

Based on this project, what are some recommended power supplies that are currently in the market?

ChatGPT Outcome:

For a project like the iSPY optical security system, which combines advanced imaging, processing, and free-space optical (FSO) communication technologies, selecting the right power supplies is crucial to ensure system reliability, efficiency, and safety. Here are some recommended types of power supplies that would be appropriate:

1. Switch Mode Power Supply (SMPS)

- Description: SMPS units convert AC to DC power efficiently, providing stable DC output while handling varying load conditions and power factors. They are preferred for their efficiency and compact size.

- Application: Useful for powering the MCU, camera modules, and other digital components on PCB #1 and #2 where efficiency and space are critical.

2. Battery Packs

- Description: Rechargeable battery packs, such as lithium-ion or lithium-polymer, offer portability and can provide a high-energy density, which is vital for systems that may need to operate independently of a fixed power source.
- Application: Batteries can power the entire system during mobile operations or when installed in locations without direct access to mains electricity. They're particularly crucial for backup power to ensure the system remains operational during power outages.

3. Linear Power Supply

- Description: Known for providing extremely low noise and ripple compared to SMPS, which is beneficial for sensitive electronic components.
- Application: Ideal for powering noise-sensitive components such as the analog circuits in transceivers and sensors, especially where signal integrity is paramount.

4. Power over Ethernet (PoE)

- Description: PoE can transmit electrical power, along with data, to remote devices over standard twisted-pair cable in an Ethernet network.
- Application: Suitable for networked components within the system, such as cameras or any part of the system that communicates data over Ethernet. This simplifies wiring and reduces installation costs and complexity.

5. Uninterruptible Power Supply (UPS)

- Description: UPS systems provide power backup using batteries and an inverter to supply AC or DC power when main power is unavailable, and also condition incoming power to protect against surges, spikes, and brownouts.
- Application: Ensures continuous operation of the entire security system during power failures, protecting against data loss and hardware damage during critical operations.

6. Solar Power Solutions

- Description: For outdoor installations, incorporating solar panels can offer a renewable energy solution, reducing reliance on external power sources and enhancing the system's sustainability.

- Application: Powers components during daylight with battery storage to provide energy during nighttime or cloudy conditions, ideal for remote or environmentally sensitive locations.

7. Regulated DC Power Supplies

- Description: These provide a constant voltage or current despite variations in load current or input voltage, which is crucial for maintaining the stability and performance of the electronic components.

- Application: Feeds stable DC power to sensitive circuits and components within the iSPY system, ensuring reliable operation and preventing damage due to voltage fluctuations.

Each of these power supply types brings specific benefits that can enhance the performance and reliability of your iSPY system, making it crucial to choose based on the component requirements, operational environment, and intended deployment scenario.

Copilot Outcome:

For your iSPY project, which requires a 5V DC regulated output and low power consumption, here are some recommended power supplies currently available on the market:

1. Mean Well LRS-50-5:

- a. **Output:** 5V DC, 10A (50W)
- b. **Efficiency:** Up to 88%
- c. **Features:** Compact size, high reliability, low cost, and built-in protections (short circuit, overload, over voltage).
- d. **Price:** [Around \\$15-\\$20¹](#).

2. Adafruit 5V 4A Switching Power Supply:

- a. **Output:** 5V DC, 4A (20W)
- b. **Features:** High efficiency, compact design, and includes a 2.1mm barrel jack.
- c. **Price:** [Around \\$15²](#).

3. CUI Inc. SWI5-5-N-P5:

- a. **Output:** 5V DC, 1A (5W)
- b. **Features:** Wall plug-in power supply, high efficiency, and includes multiple plug options for international use.
- c. **Price:** [Around \\$10³](#).

4. **RS Pro 5V DC Power Supply:**

- a. **Output:** 5V DC, 3A (15W)
- b. **Features:** Compact, reliable, and includes over voltage, over current, and short circuit protection.
- c. **Price:** Around \$20.

5. **Mean Well RS-15-5:**

- a. **Output:** 5V DC, 3A (15W)
- b. **Efficiency:** Up to 78%
- c. **Features:** Compact size, high reliability, and built-in protections (short circuit, overload, over voltage).
- d. **Price:** Around \$10-\$15.

These power supplies are well-suited for low-power applications like your iSPY project, providing reliable and efficient power with necessary protections. If you need more specific recommendations or have any other questions, feel free to ask!

Google AI Outcome:

When looking for a power supply for a designed project, some well-regarded options include Meanwell power supplies, Cincon products, RECOM power supplies, and depending on your specific needs, the ZKETECH EBD-A20H variable DC power supply could be a good choice as well; all offering a variety of options for different voltage and current requirements, with good reliability and industry standard certifications.

Key points to consider when choosing a power supply:

- **Voltage and Current Requirements:**

Match the output voltage and current of the power supply to your project's needs, ensuring sufficient capacity to handle peak loads.

- **Form Factor:**

Consider the physical size and mounting options needed for your project.

- **Efficiency:**

Look for power supplies with high efficiency ratings to minimize heat generation.

- **Safety Certifications:**

Ensure the power supply meets relevant safety standards like UL, CE, and TUV.

Some recommended brands and series:

- **Meanwell:** Considered a top choice for a wide range of power supplies, known for reliability and quality.
- **Cincon:** Another reliable option with a focus on efficiency and long lifespan
- **RECOM:** Offers a broad selection of AC-DC and DC-DC converters, often suitable for compact applications
- **WAGO ECO 2:** Compact industrial power supply with good space-saving design

Carlo Gavazzi SPMA Series: Compact DIN-mount power supply suitable for industrial and building automation

Case Study 4:

Prompt 4

Can you recommend me some camera modules for a surveillance camera I am trying to build? I will work on the lens design later and choose the best lens or lenses but I want to have a camera module in mind. I am looking for something that can give me a wide field of view but not too expensive.

ChatGPT Outcome 4

Here are some camera modules that may suit your surveillance camera design, offering a wide field of view while remaining cost-effective:

1. **Arducam IMX219 Camera Module**

- **Sensor**: Sony IMX219
- **Resolution**: 8 MP (3280x2464)
- **Field of View**: Options available with lenses that provide up to 160° wide-angle.
- **Interface**: MIPI CSI-2
- **Price**: Around \$25-40 USD
- **Features**: This is a popular camera module for Raspberry Pi, offering decent image quality for surveillance and other general-purpose applications. You can replace the lens to achieve your desired field of view.

2. **Raspberry Pi Camera Module 3

- **Sensor**: Sony IMX708
- **Resolution**: 12 MP (4608x2592)
- **Field of View**: 120° (Wide-angle version)
- **Interface**: MIPI CSI-2
- **Price**: Around \$30-40 USD
- **Features**: It offers a wide field of view (120°) in the wide-angle version, which fits your requirement for wide surveillance. It supports HDR (High Dynamic Range), which is useful for well-lit environments.

3. **E-Con Systems See3CAM_CU55

- **Sensor**: Sony IMX547
- **Resolution**: 5 MP (2592x1944)
- **Field of View**: 100° with the standard lens; other lenses available for wider angles.
- **Interface**: USB 3.0

- **Price**: Around \$60-80 USD
- **Features**: This module provides high-quality imaging and supports a range of lens options. Its USB interface simplifies integration, and it's well-suited for real-time applications.

4. **OV5647 Camera Module**

- **Sensor**: OmniVision OV5647
- **Resolution**: 5 MP (2592x1944)
- **Field of View**: Available with lenses that can offer up to 120°.
- **Interface**: MIPI CSI-2
- **Price**: Around \$15-25 USD
- **Features**: A budget-friendly option, this module is used in many low-cost surveillance and robotics systems. It supports interchangeable lenses for a customizable field of view.

5. **AI-Thinker ESP32-CAM**

- **Sensor**: OV2640
- **Resolution**: 2 MP (1600x1200)
- **Field of View**: Around 100° with the stock lens.
- **Interface**: Wi-Fi/Bluetooth (ESP32-based)
- **Price**: Around \$10-20 USD
- **Features**: This module is extremely affordable and includes an ESP32 microcontroller for wireless communication, making it suitable for low-cost wireless surveillance solutions. While the resolution is lower, it could work for certain low-budget surveillance applications.

Considerations for Choosing a Camera Module:

- **Field of View**: Since you're focusing on a wide field of view, you can choose modules that either come with wide-angle lenses or support lens swapping, allowing you to design a custom lens later.
- **Interface**: Depending on your system setup, decide between MIPI CSI-2 (common in embedded systems) or USB (easier for plug-and-play integration).
- **Resolution**: Higher resolutions are better for capturing more detail, especially if you need to zoom in on portions of the image.
- **Price**: All the listed options are relatively affordable and range from \$10 to \$80 depending on the features and resolution.

These camera modules should give you flexibility for your lens design while being budget-friendly for your surveillance system.

Prompt 4.1

Can I replace the lenses of the 2 option? 0.3 MegaPixels USB Camera for Raspberry Pi / NVIDIA Jetson Nano / UNIHAKER what is the shutter type?

ChatGPT Outcome 4.1

Yes, you can replace the lenses on the **Raspberry Pi Camera Module 3**, but it requires some modifications. By default, the camera module comes with a fixed lens, but third-party accessories, like lens adapters, can allow you to swap lenses to customize the field of view or optical performance.

Steps to Replace the Lens:

1. **Remove the Existing Lens**: You would need to carefully remove the fixed lens, which usually involves unscrewing or detaching the plastic housing.
2. **Use a Lens Adapter**: Third-party vendors offer adapters that can be attached to the camera module, allowing you to use different lenses (like M12 or CS-mount lenses).
3. **Choose Compatible Lenses**: Make sure to choose a lens that matches the sensor size (1/2.43" for the IMX708 sensor in this case) to ensure proper image coverage and minimize vignetting.

Since you are planning to design a custom lens system later on, replacing the lens on this module could give you more flexibility in achieving your desired field of view. Just keep in mind that modification requires precision, and any changes to the lens could affect calibration and focus.

For a **0.3 MegaPixel USB camera** designed for platforms like the Raspberry Pi, NVIDIA Jetson Nano, or UNIHAKER, the shutter type is most likely **rolling shutter**. This is because:

- **Rolling shutters** are more common in lower-cost and lower-resolution camera modules due to their simpler design and lower power requirements.
- **Global shutters** are typically found in more expensive, high-performance cameras that require accurate image capture of fast-moving objects, which isn't typical for a 0.3 MP camera.

To confirm the exact shutter type, it's best to check the datasheet or specifications provided by the manufacturer, as some models may vary. Would you like help finding those details for a specific model?

GPTA

Prompt 1

Can you describe entropy coding (e.g. huffman and arithmetic) in terms of live video feed compression?

Outcome 1

Entropy coding, such as Huffman and arithmetic coding, plays a key role in the compression of live video feeds by efficiently encoding data based on its statistical properties. Here's a breakdown of how these methods work in the context of video compression:

1. Huffman Coding in Video Compression:

Huffman coding is a lossless entropy coding technique used to compress data by encoding frequently occurring symbols with shorter bit lengths and less frequent symbols with longer bit lengths. In a live video feed, frames consist of a large amount of data that can be processed symbolically (e.g., pixel values, differences between pixel values in adjacent frames, or quantized coefficients from a transform like DCT used in formats like H.264).

- **Process:**

- **Frame Analysis:** The video is broken down into frames, and each frame is analyzed to determine the frequency of certain patterns (e.g., pixel intensity values or motion vectors).
- **Symbol Mapping:** Huffman coding then maps more frequently occurring patterns to shorter codes and less frequent ones to longer codes.
- **Transmission:** The compressed frame data is transmitted in the form of a bitstream, where shorter codes dominate, leading to reduced overall data size.

Example: Suppose a video has many repeated pixel values for background color in a stationary scene. These frequent pixel values can be assigned shorter codes, reducing the number of bits needed for the static portion of the video.

- **Advantages for Live Video:**

- Huffman coding can be used to encode residual data in video compression algorithms like H.264, which first applies motion compensation to remove temporal redundancy between frames.
- The compression speed of Huffman coding makes it suitable for real-time video feed compression.

- **Limitations:**

- Huffman coding is relatively simple and may not achieve the highest compression rates for highly complex video data, where more adaptive techniques like arithmetic coding might outperform it.

2. Arithmetic Coding in Video Compression:

Arithmetic coding is a more sophisticated entropy coding technique that often outperforms Huffman coding by encoding entire sequences of symbols as a single number (usually a fraction). This provides greater flexibility and efficiency, especially for symbols with probabilities that aren't powers of two.

- **Process:**

- **Frame Analysis:** Like Huffman coding, arithmetic coding starts by analyzing video frames and predicting the probability of different symbols (e.g., pixel values or residuals from motion compensation).
- **Range Assignment:** Instead of assigning fixed-length codes, arithmetic coding represents the entire sequence of symbols as a fractional number within a range based on their probabilities.
- **Bitstream Generation:** The resulting bitstream is more compact, as arithmetic coding can efficiently represent fractional probabilities with high precision.

Example: Suppose a region of the video has a predictable color gradient or motion pattern. Arithmetic coding can represent this predictability in a compact form, minimizing the data required for transmission.

- **Advantages for Live Video:**

- Arithmetic coding is more adaptive than Huffman coding, meaning it can handle video data with complex or varying statistical properties (e.g., scenes with sudden movement or high texture).
- It achieves near-optimal compression rates, which can be critical for bandwidth-limited live video transmission.

- **Limitations:**

- Arithmetic coding is more computationally intensive, which could make it challenging for very low-latency live video compression systems unless hardware-accelerated or highly optimized implementations are used.

Integration in Live Video Compression Standards (e.g., H.264/HEVC):

Both Huffman and arithmetic coding are used in popular video compression standards. For instance:

- **H.264/AVC:** Uses a combination of Huffman coding (for variable-length coding) and Context-Adaptive Binary Arithmetic Coding (CABAC) for more efficient

encoding. CABAC plays a crucial role in compressing the residual data after motion estimation.

- **HEVC (H.265):** Leverages arithmetic coding (specifically CABAC) more heavily to reduce the bit rate even further, making it particularly suitable for higher resolution live streams like 4K or 8K.

Summary:

- **Huffman Coding:** Quick, simple, and effective for coding frequent symbols in predictable scenes, but less optimal for complex or highly dynamic video data.
- **Arithmetic Coding:** Offers better compression efficiency for complex video content but at the cost of higher computational complexity, which may impact real-time encoding unless optimized.

In a live video feed compression system, both Huffman and arithmetic coding work to minimize the bitstream size by efficiently encoding video data based on its statistical properties, which is essential for transmitting high-quality video over bandwidth-constrained channels.

GPTB

Prompt 1

Could you explain bandwidth and latency in terms of my project?

Outcome 1

In the context of your video compression and transmission system, **bandwidth** and **latency** are two critical factors to consider for efficient data transmission over free space using a laser.

Bandwidth:

Bandwidth refers to the maximum amount of data that can be transmitted over the laser communication channel in a given period of time. It is usually measured in bits per second (bps). For your project, the **bandwidth** defines how much compressed video data (in bitstream form) can be sent per second through the laser.

- **Higher bandwidth** means you can transmit more video data in less time, which is especially important if you're dealing with high-resolution or high-frame-rate video.
- **Lower bandwidth** can result in slower transmission, potentially causing delays or the need to use more aggressive video compression, which can reduce quality.

Since you are working with video compression, it's essential to optimize the amount of data sent per second. Efficient compression algorithms will help minimize the bitstream size, allowing it to fit within the available bandwidth.

Latency:

Latency is the time delay between when the video data is sent from the source (via laser) and when it is received and decoded at the destination. In your case, latency is affected by several factors:

- **Propagation delay:** the time it takes for the laser signal to travel through free space from the transmitter to the receiver.
- **Processing delay:** the time spent compressing the video, encoding it into a bitstream, and decoding it on the receiving end.
- **Buffering delay:** the time spent waiting for enough data to be buffered to avoid interruptions in the video stream.

Low latency is crucial for real-time applications, such as live video transmission, where any noticeable delay can disrupt the user experience.

In summary:

- **Bandwidth** determines how much video data can be sent per second, affecting video quality and frame rate.
- **Latency** affects the delay between video capture, transmission, and playback, critical for real-time streaming.

Balancing compression efficiency (to minimize bandwidth use) and ensuring low latency will be key to the success of your laser-based video transmission system.

Prompt 2

Would compression also play a factor in latency?

Outcome 2

Yes, **compression** plays a significant role in latency as well, especially in your video compression and transmission system.

Here's how compression affects latency:

1. Compression Time (Encoding Latency):

Compression involves encoding the video data into a smaller bitstream, which can introduce latency due to the time it takes to process the video frames. The more complex the compression algorithm (e.g., higher compression ratios or advanced encoding techniques), the longer it may take to compress the video. This contributes to **encoding latency**.

- **Higher compression** (more processing): can reduce bandwidth needs but takes more time to encode, thus increasing the latency.
- **Lower compression** (less processing): reduces the encoding time, thus lowering latency but may require more bandwidth for transmission.

2. Decompression Time (Decoding Latency):

On the receiving end, the compressed video bitstream needs to be **decompressed** (decoded) before it can be displayed. Like compression, decompression also requires processing time, which introduces additional latency, especially if the compression method is complex.

- **Complex codecs:** like H.264 or H.265 (often used in video compression) might take more time to decode, increasing latency.
- **Simpler codecs:** may have faster decompression times, reducing latency but might not compress as effectively.

3. Trade-off Between Compression Ratio and Latency:

There's often a trade-off between how much compression you apply and the latency you introduce:

- **Aggressive compression:** reduces the amount of data to be transmitted (saving bandwidth), but adds time for encoding and decoding, which increases latency.
- **Light compression:** transmits larger amounts of data (requiring more bandwidth), but reduces the time spent compressing and decompressing, thus lowering latency.

4. Keyframe Interval and Buffering:

Most video compression schemes use keyframes (complete frames) and intermediate frames (which store only the differences between frames). The **keyframe interval** (how often a full frame is encoded) can impact latency. Longer intervals between keyframes might save bandwidth but introduce latency as the receiver needs to wait for the next keyframe to start decoding a portion of the stream.

- **Short keyframe intervals** reduce latency but increase bandwidth consumption.

5. Compression Mode:

Some modern video codecs, like H.265, support **low-latency modes**, where the compression is optimized to minimize latency at the expense of a slightly larger bitstream.

In summary, **compression does affect latency** due to the time required for both encoding and decoding. For your project, you'll need to balance the level of compression with the system's latency requirements. If real-time transmission is a priority, you might opt for faster compression algorithms (low-latency codecs) even if they don't achieve the highest compression ratios, to ensure minimal delay.

GPTC

Prompt 1

Can you give me libraries for codecs in c++, java, and python? Please provide an explanation of each library as well.

Outcome 1

Here are some libraries for working with codecs in C++, Java, and Python:

C++ Codec Libraries

1. FFmpeg (libavcodec)

Description: FFmpeg is a complete, cross-platform solution to record, convert and stream audio and video. Its codec library, libavcodec, is widely used for encoding and decoding audio and video. It supports a vast range of codecs including H.264, VP8, VP9, AAC, MP3, and many others.

- **Use case:** Video and audio encoding, decoding, and transcoding.
- **Pros:** Supports many formats, well-documented, highly optimized.
- **Cons:** Steeper learning curve; LGPL license may impose restrictions.
- **Website:** [FFmpeg](https://ffmpeg.org/)

2. GStreamer

Description: GStreamer is a multimedia framework that allows building media-handling components like codecs and transcoders. It supports a variety of codecs via plugins, which can be combined to create pipelines for media processing.

- **Use case:** Real-time streaming, media playback, and processing.
- **Pros:** Highly modular, great for building complex media pipelines.
- **Cons:** More complex than FFmpeg, especially for beginners.
- **Website:** [GStreamer](https://gstreamer.net/)

3. OpenH264

Description: OpenH264 is a codec library that provides real-time H.264 encoding and decoding. It's developed by Cisco and is widely used for video streaming and teleconferencing solutions.

- **Use case:** Real-time video compression (H.264).
- **Pros:** Lightweight, focused on H.264.
- **Cons:** Limited to H.264; doesn't support other formats.
- **Website:** [OpenH264](https://openh264.org/)

Java Codec Libraries

1. JCodec

Description: JCodec is a pure Java library for video and audio codec processing. It supports basic operations like encoding and decoding H.264 and VP8 formats, as well as media container operations such as MP4.

- **Use case:** Basic encoding and decoding of common formats in Java.
- **Pros:** Written in pure Java, no native dependencies.

- **Cons:** Limited codec support, lacks real-time performance optimizations.
 - **Website:** [JCodec](#)
2. **Xuggle** **(Xuggler)**
Description: Xuggle is a Java wrapper for FFmpeg, providing an easy way to work with audio and video codecs in Java. It supports a wide variety of formats and offers access to FFmpeg's vast functionality.
- **Use case:** Handling multimedia streams and encoding/decoding them in Java.
 - **Pros:** Supports most codecs, built on FFmpeg.
 - **Cons:** Project maintenance has been inconsistent in recent years.
 - **Website:** [Xuggle](#)
3. **JMF** **(Java Media Framework)**
Description: JMF was an older framework for capturing, playing, streaming, and transcoding media. It includes support for some codecs but is no longer actively developed or maintained.
- **Use case:** Legacy systems and simple multimedia tasks.
 - **Pros:** Provides basic multimedia support.
 - **Cons:** Outdated, limited codec support.
 - **Website:** [JMF](#) (deprecated)
-

Python Codec Libraries

1. **PyAV**
Description: PyAV is a Pythonic binding for FFmpeg, allowing easy access to FFmpeg's audio and video codec capabilities. It can be used to encode, decode, and manipulate multimedia files using FFmpeg's libavcodec.
- **Use case:** Working with audio and video files for tasks like encoding/decoding, file manipulation, and transcoding in Python.
 - **Pros:** Simple FFmpeg binding; easy-to-use API.
 - **Cons:** Limited to FFmpeg's capabilities; must install FFmpeg separately.
 - **Website:** [PyAV](#)
2. **MoviePy**
Description: MoviePy is a high-level library for editing video files in Python. It supports reading and writing video files, as well as applying transformations such as cutting, concatenation, effects, and basic audio/video codec handling.
- **Use case:** Simple video editing and codec manipulation.

- **Pros:** User-friendly, high-level API; integrates with FFmpeg.
- **Cons:** Not suitable for complex, low-level codec operations.
- **Website:** MoviePy

3. FFmpeg-Python

Description: FFmpeg-Python is a convenient Python wrapper for FFmpeg, enabling users to build and execute FFmpeg commands programmatically. It supports all FFmpeg codecs, containers, and filters.

- **Use case:** Complex multimedia processing pipelines using FFmpeg.
- **Pros:** Provides access to all of FFmpeg's functionality; flexible.
- **Cons:** Steeper learning curve for FFmpeg command syntax.
- **Website:** [FFmpeg-Python](#)

These libraries span a wide range of codec support and are often built on FFmpeg or similar core technologies, providing different levels of abstraction depending on your needs.

Prompt 2

Does ffmpeg support H.265 as well?

Outcome 2

Yes, **FFmpeg** supports **H.265 (HEVC)** encoding and decoding. The H.265 codec offers better compression efficiency than H.264, which is especially useful for high-definition and 4K video content.

Key Details About H.265 Support in FFmpeg:

1. Encoding:

- FFmpeg can encode videos to H.265 using the **libx265** encoder.
- libx265 is an open-source encoder for H.265/HEVC, offering high-efficiency video compression.
- You can specify libx265 during encoding with the `-c:v libx265` option.

2. Decoding:

- FFmpeg can also decode H.265 videos using the **libx265** decoder.
- You can process and view H.265-encoded content by simply specifying the input file.

Example FFmpeg Command to Encode Video Using H.265:

bash

Copy code

```
ffmpeg -i input.mp4 -c:v libx265 -crf 28 -preset slow output.mp4
```

- **-c:v libx265:** Specifies the H.265 encoder.
- **-crf 28:** Controls the quality (Constant Rate Factor). Lower values give better quality but larger files.
- **-preset slow:** Optimizes the encoding process (slow gives better compression at the cost of speed).

Additional Notes:

- **Licensing:** H.265 (HEVC) is patented, and although libx265 is open-source, you may need to consider licensing fees if you plan to use it in a commercial product.
- **Hardware Acceleration:** FFmpeg also supports hardware-accelerated encoding and decoding for H.265 if your GPU supports it (e.g., using nvenc for NVIDIA GPUs).

FFmpeg provides comprehensive support for H.265, making it an excellent choice for projects requiring high-efficiency video compression.

GPTD

Prompt 1

What are three potential MCUs that could work well with my project of having computer vision and video compression?

Outcome 1

For a project involving computer vision and video compression, you need a microcontroller (MCU) with sufficient processing power, memory, and support for hardware acceleration for video tasks. Here are three potential MCUs that could work well:

1. ESP32-S3

- **Key Features:**
 - **Dual-core 32-bit Xtensa LX7 CPU**, up to 240 MHz.
 - Integrated **AI accelerator** and **vector instructions** for efficient handling of image and signal processing tasks.
 - Supports **8/16-bit parallel cameras**, making it suitable for basic computer vision tasks.
 - **Wi-Fi and Bluetooth** integrated for wireless video transmission.
- **Advantages:**
 - Low cost, relatively low power consumption, and built-in wireless support.
 - Suitable for lightweight video compression tasks like MJPEG.

- **Limitations:**

- Limited for heavy video compression tasks like H.264 without external hardware.

2. STM32H7 Series

- **Key Features:**

- **Cortex-M7** core running at up to 480 MHz with **2MB flash** and **1MB RAM**.
- **Hardware JPEG encoding/decoding** support, which can significantly help with video compression.
- **FMC and DCMI interfaces** for connecting cameras and high-speed data transfers.

- **Advantages:**

- High processing power, making it suitable for more advanced video compression algorithms.
- The hardware JPEG codec offloads the compression workload from the CPU.

- **Limitations:**

- More complex to program and higher power consumption than ESP32.

3. Raspberry Pi Pico W (RP2040 with Wireless)

- **Key Features:**

- Dual-core **Cortex-M0+** running at 133 MHz, with **264KB of RAM**.
- **Pico W** version includes **Wi-Fi** for wireless video transmission.
- Support for **Pio** (Programmable I/O) that allows handling of parallel interfaces like camera input.

- **Advantages:**

- Affordable, with a large developer community and libraries for interfacing with cameras.
- While it doesn't have built-in video compression, it can interface with external compression modules (e.g., using an external H.264 encoder).

- **Limitations:**

- Limited onboard memory and processing power compared to the STM32H7 and ESP32-S3, so external hardware would be needed for high-end compression.

The STM32H7 or ESP32-S3 would be ideal for handling lightweight computer vision tasks and basic video compression natively, while the Raspberry Pi Pico W could be suitable if external components are added for more complex tasks.

Prompt 2

What about the msp 430fr6989

Outcome 2

The **MSP430FR6989** is a low-power microcontroller from Texas Instruments, part of the MSP430 series, known for ultra-low power consumption and high energy efficiency. However, it may not be the best choice for a project involving computer vision and video compression due to the following reasons:

Limitations of the MSP430FR6989:

1. Processing Power:

- The MSP430FR6989 is based on a **16-bit RISC architecture** with a maximum clock speed of **16 MHz**, which is much lower compared to the MCUs typically used in computer vision or video compression tasks.
- Video compression and computer vision tasks are computationally intensive, requiring much higher processing power, such as 32-bit architectures with higher clock speeds and additional computational capabilities (like hardware accelerators).

2. Memory:

- It comes with **128 KB FRAM** and **2 KB SRAM**, which is quite limited for handling video data or even image buffers.
- Computer vision tasks, especially video processing, require a significant amount of memory for storing frames, applying transformations, and buffering data streams, which would quickly exceed this capacity.

3. No Camera or Graphics Interfaces:

- Unlike some other MCUs, the MSP430FR6989 does not have dedicated camera or display interfaces (like DCMI or parallel camera interfaces), which are crucial for real-time image acquisition and video processing.

4. No Dedicated Video/Compression Hardware:

- The MSP430 lacks any hardware acceleration for video encoding or decoding (like JPEG or H.264 hardware codecs) that are common in other microcontrollers like the STM32H7 series.
- Without hardware support, video compression would need to be done entirely in software, which would be far too slow on this chip.

When the MSP430FR6989 Could Be Useful:

- If your project required ultra-low power consumption and only basic sensing tasks (like reading from low-power sensors or basic IoT applications), the MSP430FR6989 would be an excellent choice.
- It could be used in a peripheral role, such as controlling sensors or handling communication tasks in a more complex system where another device does the heavy lifting for video and image processing.

Conclusion:

The MSP430FR6989 is not a good fit for video compression or computer vision applications due to its limited processing power, memory, and lack of video-related hardware. You would be better off with a more powerful MCU like the **ESP32-S3**, **STM32H7**, or even a small single-board computer like the **Raspberry Pi** if your project demands more computational power and memory.

Prompt 3

What are three potential raspberry pi boards that I could use for this project?

Outcome 3

For a project involving computer vision and video compression, Raspberry Pi boards offer a good balance of computational power, flexibility, and available software libraries. Here are three Raspberry Pi boards that could be a great fit:

1. Raspberry Pi 4 Model B

- **Key Features:**
 - **Quad-core Cortex-A72 (ARM v8)** 64-bit SoC @ 1.5GHz.
 - Available with **2GB, 4GB, or 8GB RAM**, which is more than enough for video processing and compression tasks.
 - **Dual-display support** with 4K output, which is useful if you need to monitor your video feed or test video quality.
 - Includes **VideoCore VI GPU** with hardware acceleration for **H.264** and **H.265** video encoding/decoding.
- **Advantages:**
 - Great for computer vision tasks using frameworks like OpenCV and TensorFlow Lite.
 - High processing power and memory make it capable of handling real-time video compression and transmission.
 - A large community and a wide range of libraries for interfacing with cameras, such as the **Raspberry Pi Camera Module**.
- **Ideal For:**

- More complex video compression tasks, live streaming, and image processing applications.

2. Raspberry Pi Zero 2 W

- **Key Features:**

- **Quad-core Cortex-A53 CPU @ 1GHz.**
- **512MB RAM**, which is lower than the Pi 4 but can still handle lightweight video compression and basic computer vision tasks.
- Built-in **Wi-Fi and Bluetooth**, making it suitable for wireless video transmission projects.
- Compatible with the **Raspberry Pi Camera Module**.

- **Advantages:**

- Extremely compact and power-efficient, making it a great option for portable or battery-powered projects.
- Cheaper than the Raspberry Pi 4, which makes it appealing for simpler projects or cost-sensitive applications.
- Supports hardware-accelerated H.264 encoding/decoding for efficient video compression.

- **Ideal For:**

- Lightweight video processing tasks, basic computer vision, and low-cost implementations of video streaming.

3. Raspberry Pi 3 Model B+

- **Key Features:**

- **Quad-core Cortex-A53 CPU @ 1.4GHz.**
- **1GB RAM**, which is enough for mid-range video compression tasks and basic computer vision.
- **Dual-band 2.4GHz and 5GHz Wi-Fi** with Gigabit Ethernet, allowing faster video transmission over networks.
- Hardware support for **H.264 encoding/decoding**.

- **Advantages:**

- More affordable than the Pi 4 while still providing decent processing power for video compression.
- Supports most of the same peripherals as the Pi 4, including the camera modules, making it a versatile choice for video projects.

- **Ideal For:**

- Medium complexity video compression, remote video monitoring, and projects where high-end performance isn't a strict requirement.

Comparison Overview:

- **Raspberry Pi 4 Model B:** Best choice for handling complex video compression (H.264/H.265), real-time image processing, and more resource-demanding computer vision tasks.
- **Raspberry Pi Zero 2 W:** Great for compact, portable solutions with lighter video processing tasks. Its small size and low power draw make it suitable for projects where space or energy efficiency is key.
- **Raspberry Pi 3 Model B+:** Balanced option for mid-range video tasks, where cost is a factor but performance is still important for moderate video processing.

These Raspberry Pi boards provide a variety of options depending on the complexity of your video compression and computer vision needs. The **Pi 4 Model B** would be ideal for more advanced tasks, while the **Zero 2 W** could be a solid choice for a simpler or more space-constrained solution.

GPTE

Prompt 1

what should I write about when writing research on codecs and how to compress a live video stream to transfer across a laser using bitstream.

Outcome 1

When writing research on codecs and how to compress a live video stream for transmission via a laser using a bitstream, you should cover both the theoretical and practical aspects of video compression and transmission. Below is a structured approach to what you can include in your research:

1. Introduction to Codecs and Video Compression

- **Definition of Codecs:** Explain what a codec is—combination of *encoder* and *decoder*—and its role in video compression.
- **Types of Codecs:** Discuss the major categories:
 - **Lossless Codecs** (e.g., Huffman coding, Run-Length Encoding)
 - **Lossy Codecs** (e.g., H.264, H.265, VP9)
- **Importance of Compression:** Emphasize the need for compression in reducing bandwidth and storage requirements, especially for real-time applications like live streaming.

2. Understanding Video Compression Techniques

- **Spatial Compression:** Compression within individual frames (Intra-frame compression) by removing redundant data within a single image (e.g., JPEG-based).

- **Temporal Compression:** Compression across frames (Inter-frame compression) by encoding the differences between consecutive frames, such as using motion estimation and compensation (e.g., P-frames, B-frames).
- **Entropy Coding:** Methods to further reduce the size of encoded frames by efficiently coding the data (e.g., Huffman coding, Arithmetic coding).
- **Challenges in Live Video Compression:** Discuss the complexity of achieving low-latency and high-efficiency compression without significant quality loss.

3. Real-Time Streaming Requirements

- **Latency and Bandwidth Considerations:** Detail how latency is critical for live streaming, as is minimizing the bit rate for efficient transmission over limited bandwidth (such as a laser).
- **Quality vs. Compression Trade-Off:** Explain how compression reduces quality but allows for higher transmission efficiency. Highlight tools like bitrate control, adaptive compression (CRF in H.264), and variable vs constant bit rates.
- **Codec Comparison:** Compare commonly used codecs for live video streaming (H.264/AVC, H.265/HEVC, VP8/VP9, AV1), discussing factors like compression efficiency, computational complexity, and compatibility.

4. Bitstream Transmission Using Laser

- **Overview of Bitstream Transmission:** Explain how video streams are converted into a bitstream, including formatting into packets for transmission over different media, in this case, laser communication.
- **Laser Communication Overview:**
 - **Free-Space Optical Communication (FSO):** How lasers are used to transmit data wirelessly, including an overview of modulation schemes (e.g., amplitude modulation, pulse-width modulation) to encode the bitstream.
 - **Advantages:** High bandwidth, low interference, and long-range communication potential.
 - **Challenges:** Susceptibility to environmental factors like weather (fog, rain) and line-of-sight requirements.

5. Modulation Techniques for Laser Transmission

- **Pulse Code Modulation (PCM):** Explain how data from a compressed video bitstream can be encoded using PCM.
- **Amplitude Modulation:** Technique where the strength (amplitude) of the laser beam is modulated to represent binary data.
- **Frequency/Phase Modulation:** Alternative methods to improve the signal quality or robustness.

- **Error Correction and Recovery:** Discuss error detection/correction techniques (like Reed-Solomon codes, Hamming codes) to ensure data integrity, especially for long-distance transmission or over noisy environments.

6. Compression Techniques in Relation to Transmission Medium

- **Bitrate Control:** How to manage bitrate and compression settings to match the available bandwidth of the laser transmission.
- **Adaptive Streaming:** Investigate how adaptive bitrate streaming (ABR) techniques could be employed to adjust the compression dynamically based on transmission conditions.
- **Fragmentation of Video Stream:** Methods for dividing the compressed video into smaller packets suitable for transmission over a bitstream and how to synchronize the frames during reception.

7. Implementation and Performance Evaluation

- **Experimentation:** Explain how you might implement such a system, including software (like FFmpeg) and hardware components (laser transmitter/receiver, camera, FPGA, microcontroller).
- **Simulations:** Use simulation tools to model the performance of the transmission under different conditions, such as varying distances, environmental factors, and data rates.
- **Metrics for Evaluation:** Quality of experience (QoE) metrics such as peak signal-to-noise ratio (PSNR), Structural Similarity Index (SSIM), bitrate, frame rate, and latency.

8. Security Considerations

- **Encryption for Transmission:** Discuss techniques for securing the bitstream during transmission, such as AES encryption applied to the video stream before modulation.
- **Authentication and Tamper Detection:** Ways to ensure that the transmitted data has not been altered during transmission.

9. Conclusion and Future Directions

- **Current Limitations:** Summarize the current challenges and limitations in using laser communication for real-time video transmission, including environmental susceptibility and codec computational complexity.
- **Future Research:** Discuss potential areas for further research, such as improving error correction techniques, enhancing codec efficiency, or developing new modulation schemes for laser-based communication.

By covering these areas, your research will provide a comprehensive overview of the topic, integrating both the theoretical underpinnings of video compression and practical applications related to laser-based transmission.

Appendix G – Zemax Images

Image 1: Design from book “Designing Optics Using Zemax”

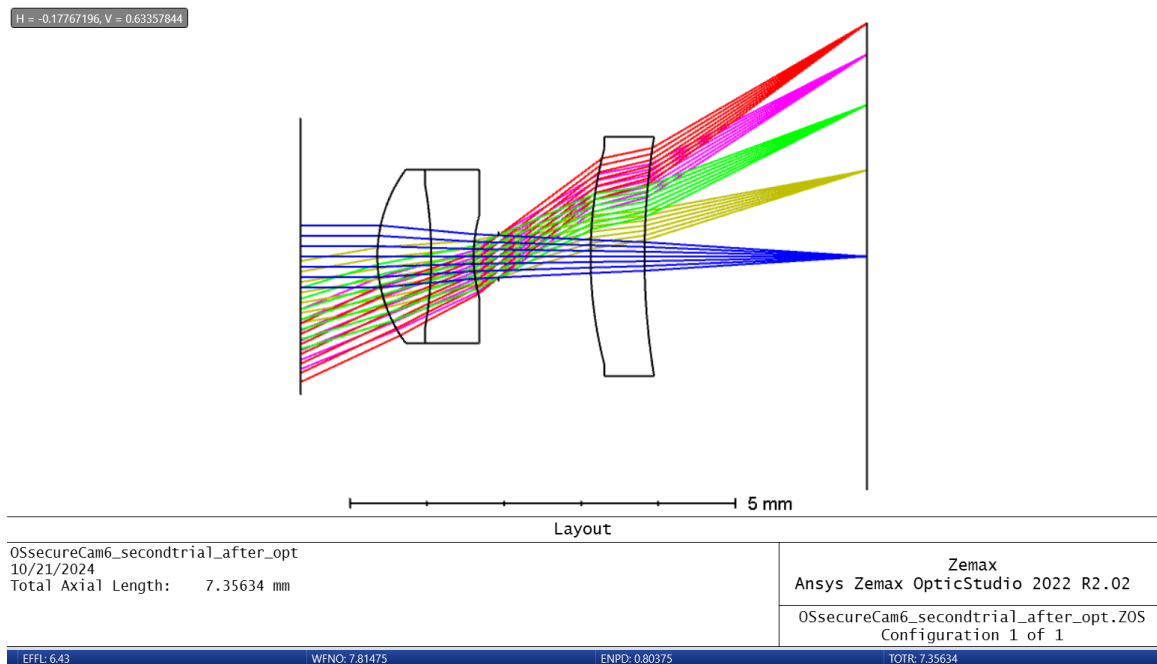
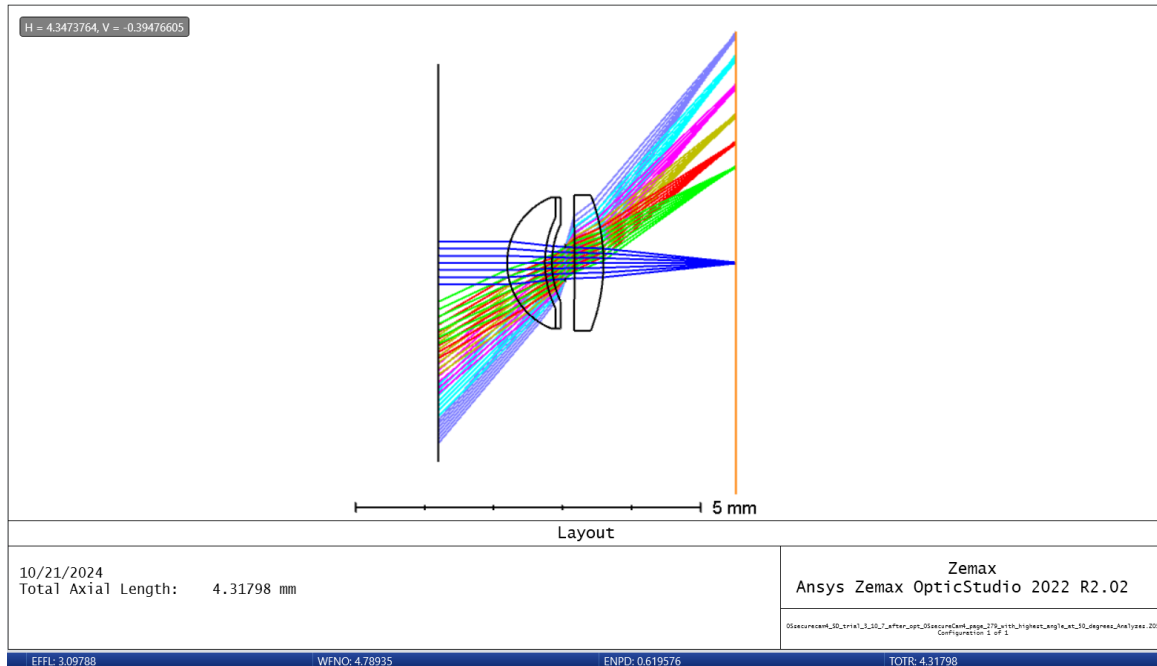


Image 2: Author’s own design following framework from “Designing Optics Using Zemax”



Appendix H – Computer Vision

Image 1: PGI Architecture from “YOLOv9: Advancing the YOLO Legacy.”

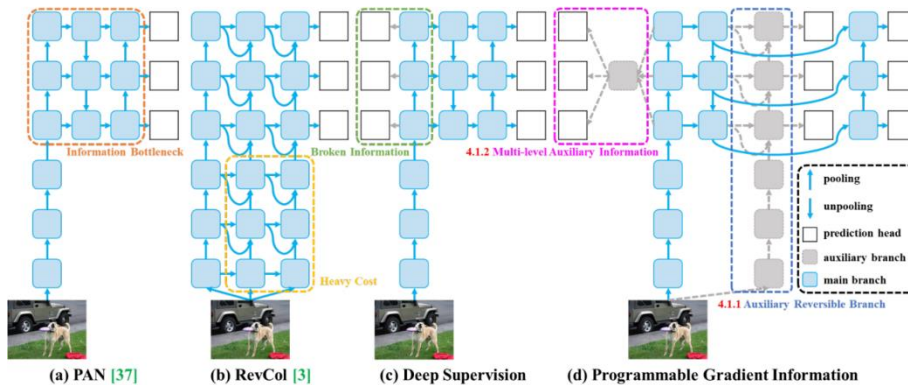


Image 2: GELAN Architecture from “YOLOv9: Advancing the YOLO Legacy.”

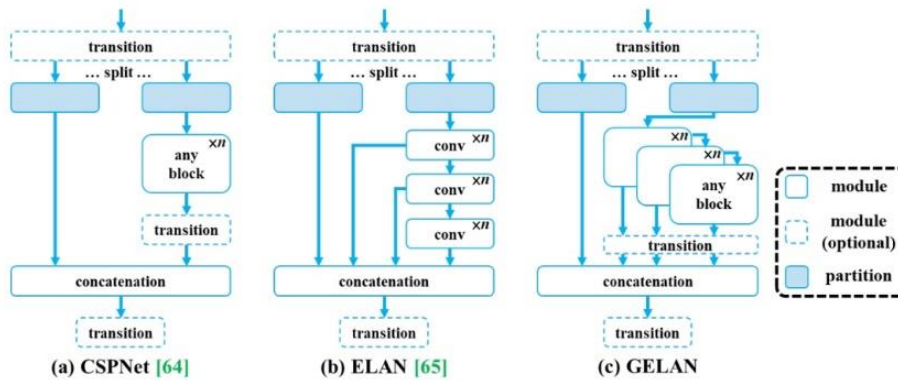


Image 3: Comparison of State-of-the-Art Real-Time Object Detectors from “YOLOv9”

Detection (COCO)		Segmentation (COCO)			
Model	size (pixels)	mAP ^{val} ₅₀₋₉₅	mAP ^{val} ₅₀	params (M)	FLOPs (B)
YOLOv9t	640	38.3	53.1	2.0	7.7
YOLOv9s	640	46.8	63.4	7.2	26.7
YOLOv9m	640	51.4	68.1	20.1	76.8
YOLOv9c	640	53.0	70.2	25.5	102.8
YOLOv9e	640	55.6	72.8	58.1	192.5

Image 4: YOLOv10 Model Variations from “YOLOv10: Paper Explanation and Inference Results.”

Model	Input Size	AP ^{Val}	FLOPS (G)	Latency (ms)
YOLOv10-N	640	38.5	6.7	1.84
YOLOv10-S	640	46.3	21.6	2.49
YOLOv10-M	640	51.1	59.1	4.74
YOLOv10-B	640	52.5	92.0	5.74
YOLOv10-L	640	53.2	120.3	7.28
YOLOv10-X	640	54.4	160.4	10.70

Image 5: YOLOv11 Model Variations from “YOLO11 NEW.”

Model	size (pixels)	mAP ^{val} ₅₀₋₉₅	Speed CPU ONNX (ms)	Speed T4 TensorRT10 (ms)	params (M)	FLOPs (B)
YOLO11n	640	39.5	56.1 ± 0.8	1.5 ± 0.0	2.6	6.5
YOLO11s	640	47.0	90.0 ± 1.2	2.5 ± 0.0	9.4	21.5
YOLO11m	640	51.5	183.2 ± 2.0	4.7 ± 0.1	20.1	68.0
YOLO11l	640	53.4	238.6 ± 1.4	6.2 ± 0.1	25.3	86.9
YOLO11x	640	54.7	462.8 ± 6.7	11.3 ± 0.2	56.9	194.9

Image 6: YOLO-NAS Model Variations from “Deci’s YOLO-NAS: Next-Generation for Object Detection”

Model	mAP	Latency (ms)
YOLO-NAS S	47.5	3.21
YOLO-NAS M	51.55	5.85
YOLO-NAS L	52.22	7.87
YOLO-NAS S INT-8	47.03	2.36
YOLO-NAS M INT-8	51.0	3.78
YOLO-NAS L INT-8	52.1	4.78

Appendix I – Video Compression Images

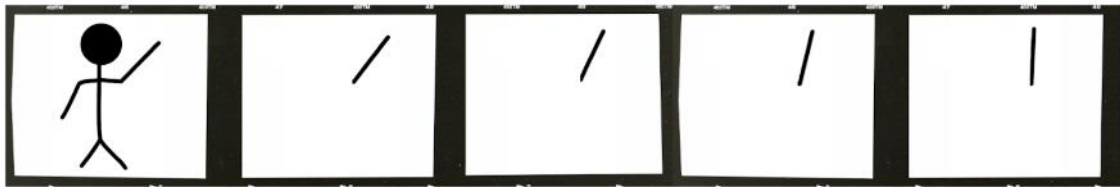
Image 1: Image Depicting Spatial Compression (Intraframe Compression)



Intraframe Compression

Every frame is encoded individually

Image 2: Image Depicting Temporal Compression (Interframe Compression)



Interframe Compression

Only the differences between frames are encoded

Image 3: Assigned Bit Lengths of Letters in Sequence

S = 000 (3 bits)

A = 001 (3 bits)

C = 01 (2 bits)

K = 1 (1 bit)

Image 4: Demonstration of PCM Output

