Acknowledgement

We first would like to thank god for giving us the opportunity to be here and make this project. Secondly, we thank our families for their unwavering support and great sacrifices in order to help us reach this moment. We would also like to thank the faculty of Benghazi university for their work and efforts to give us the best education possible in order to reach our full potential, especially our supervisor Dr.Mohammed Musbah for his guidance and assistance in not only this book but throughout our entire journey. We would also like to extend our thanks to our friend Taha Fannoush for his help and advice in the writing of this project.

.

Abbreviations and Notations

AI	Artificial Intelligence
ASCII	American Standard Code for Information Interchange
Bit	Binary Digit (0 or 1)
Blob	Binary Large Object
Cat5/Cat6	Category 5/6 (Ethernet Cable)
CCTV	Closed Circuit Television
CIF	Common Intermediate Format
CIF	Common Intermediate Format
CNN	Convolutional Neural Network
CRT	Cathode Ray Tube
CV	Computer Vision
DL	Deep Learning
DPI	Dots per Inch
DVD	Digital Versatile DISC
DVR	Digital Video Recorder
EMI	Electromagnetic Interference
FCC	Federal Communication Commission
FPS	Frame Rate Per Second
HDD	Hard Disk Drive
HDTV	High Definition Television
IP	Internet Protocol
IT	Information Technology
LAN	Local Area Network
LCD	Liquid Crystal Display

LED	Light Emitting Diodes
LOS	Line Of Sight
Mbps	Mega its per Second
MJPEG	Motion Joint Photographic Experts Group
ML	Machine Learning
Mm	Millimeter
MP	Megapixel
MPEG	Moving Picture Experts Group
NLP	Natural Language Processing
NTSC	National Television System Committee
NVR	Network Video Recorder
OUI	Organizational Unique Identifier
PAL	Phase Alternate Line
PC	Personal Computer
POE	Power Over Ethernet
PTZ	Pan/Tilt/Zoom
RF	Radio-Frequency
RFI	Radio-Frequency Interference
RG	Radio Guide
RGB	Red/Green/Blue
ROI	Region of Interest
RPA	Robotic Process Automation
SD	Standard Definition
SD-card	Secure Digital Card
SECAM	Sequential Color with Memory

SQCIF/QCIF	Sub-Quarter/Quarter CIF
S-VHS	Super VHS
TB	Tera Byte
TCP	Transfer control Protocol
TV	Television
UDP	User Datagram Protocol
UPS	Uninterruptable Power Supply
UTP	Unshielded Twisted Pair
VCR	Video Cassette Recorder
VGA	Video Graphic Array
VHS	Video Home System

Contents

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The evolution of the camera has led to the introduction of CCTV systems and their important role in increasing the security and integrity of institutions, businesses and homes.

And with fast paced technological improvement, CCTV systems are now capable of integrating with IP networks, which means that the CCTV system will have access to a much wider range of features.

However, consumption of resources done by CCTV systems is highly impractical in today's standards, because CCTV cameras capture, record and store the footage almost all the time and with the same resolution, which is inefficient because sometimes cameras record and store unimportant footage that does not contain any events of interest.

Therefore, implementing intelligent techniques such as motion detection and object recognition helps in determining whether the frames of the captured footage are useful if not, moreover, reducing the resolution of these frames will reduce the size of the captured footage, which in turn will result in less network resource consumption.

In this project, we are going to implement a system consisting of multiple algorithms on a CCTV footage in order to reduce its resource consumption through applying a series of basic image processing techniques and machine learning algorithms.

CHAPTER ONE: INTRODUCTION

1.1 General Background:

Video surveillance was first introduced in the late 19th century as one of the methods that can be used to detect escaping techniques, by providing a visual oversight of the entire prison grounds. Due to its astronomic price; the video surveillance system was confined to big cash-flow institutions and organizations such as banks, casinos and high-profile government buildings. It was not until the mid-20th century that video surveillance was widely adopted by smaller entities, such as businesses and homes, due to the reduced cost of manufacturing of video cameras, which resulted in a decline in prices [1].

CCTV (Closed-circuit television) also known as Video Surveillance can be categorized into two general types which are: Analog CCTV Systems and Digital CCTV Systems, the method of transmission can vary between wired or wireless; however, the essential components of the systems are basically the same but they may differ in terms of architecture: a lens, a camera, a monitor and cables (for wired systems) and video recorders.

The camera captures the area with its own eyes (the lenses) and then transmits the recorded signal to a monitor (it can be a simple television or a computer) via a wired or wireless medium. The monitor is often accompanied by recorders. A video recorder can be a VCR (Video Cassette Recorder), DVR (Digital Video Recorder), or an NVR (Network Video Recorder) which is used on IP-based CCTV system [2].

IP based CCTV Systems are a result of the technological changes that have tremendously affected the industry in a positive way. In recent years, the low cost of IP bandwidth and digital storage utilities has paved the way into a less complex and faster footage transmission and storage over the network, also made the process of manufacturing and acquiring all the components required to set up IP-CCTV Systems a whole lot cheaper.

Recently, Video Surveillance systems are being integrated with emerging technologies based on machine intelligence (ex: Machine Learning and Artificial Intelligence) that aim to optimize the overall performance of capturing, processing and storage of the footage.

1.2 Problem Statement:

While IP-based CCTV Systems have greatly improved the working mechanism of surveillance, it does come with its own set of drawbacks, which are:

- Bandwidth consuming: the higher the quality and quantity of the footage, the higher the bandwidth consumption.
- Waste of storage and unnecessary processing: IP-based CCTV cameras work 24/7 and store all of the footage, which is highly impractical.
- The tedious process of monitoring: it is difficult for a human to efficiently monitor and detect unusual activities, monitoring also can be a very boring activity, thus they always sleep in the movies.

1.3 The Aim and Objectives of the Project:

The aim of the project is to tackle all the previously mentioned problems of IP-Based CCTV Cameras by utilizing machine learning and computer vision.

This will result in significant improvements in surveillance systems, such as:

- Studying CCTV Technology, Machine Learning and Computer Vision.
- Better bandwidth consumption by controlling the resolution of the footage.
- Less storage consumption by storing only relevant footage.
- Smart resource allocation.
- More efficient processing.

1.4 Project Scope:

This project will focus on the area of video surveillance from the perspective of networking, computer vision and machine learning using python as a programming language. It is going to deal with integrating computer vision with the utilization of machine learning to improve the overall performance of the surveillance system. In addition, it is going to cover the objective of reducing the bandwidth consumption on the connected network.

1.5 Project Methodology:

Resource utilization by IP cameras is noticeably inefficient, especially in network and footage storage. Therefore, we decided to implement a certain methodology, where we first study CCTV systems and its components in a theoretical manner, keeping in mind the networking aspect of the inner-workings of CCTV. Next, to improve the bandwidth and storage consumption taken by the CCTV systems, we are going to incorporate these systems with machine learning and computer vision. Finally, we will compare the results of our

research with previous ones to assess the improvement of resource utilization by CCTV systems.

1.6 Requirements and Resources:

The followings are a number of requirements and resources for the implementation of this project:

- A server (virtual) with high processing power, large storage capacity and at least 8 GB of RAM.
- An IP camera.
- A Local Area Network.
- OpenCV Programming Library.
- Iftop network bandwidth montering tool.

1.7 The Targeted CCTV System:

A proper CCTV system is one that fulfills all the security requirements without undermining the integrity of the system at a reasonable cost; however, these requirements are difficult to achieve in a realistic scenario without the utilization of machine learning and the implementation of smart computer vision algorithms.

Ultimately, with the rapid evolution of AI, who knows where it is going to take us in the field of Surveillance and technology?

1.8 Project Structure:

This project dives deep into the workings of IP based CCTV systems and proceeds to mention all the drawbacks of the typical systems used today. Furthermore, it is going to introduce contemporary smart technologies that can be used to solve these issues. Subsequently, the final chapters will include a practical example of an implementation of one of those techniques.

Accordingly, this project has been structured in the following way:

- ***Chapter One (INTRODUCTION):*** *introduces a general background with an overview of the problem, scope, methodology and required tools.*
- ***Chapter Two (SURVEILLANCE SYSTEMS AND ITS COMPONENTS):*** *sheds a light on what is a CCTV System, how it works and what it comprised of.*

- ***Chapter Three (IP NETWORKING AND VIDEO SURVEILLANCE):*** *explains how networks are affected by IP-Based CCTV Systems and how real-time technology has been implemented in its systems.*
- ***Chapter Four (COMPUTER VISION AND MACHINE LEARNING****): this chapter sheds a light on the different components, types and applications of computer vision and machine learning, as well as the relationship between the two fields.*
- ***Chapter Five (DESIGN AND IMPLEMENTATION):*** *the design and build-up to the actual simulation and result analysis will be done in this chapter.*
- ***Chapter Six (Conclusion and Future Work):*** *this chapter will include the conclusion and the future plans for the algorithm that will be introduced in this project****.***

CHAPTER TWO: SURVEILLANCE SYSTEMS AND ITS COMPONENTS

This chapter gives a brief explanation of the digital video technology and its forms, and goes into the inner-workings of surveillance systems with an increased focus on its components while also comparing the different types of CCTV systems.

2.1 Overview of CCTV

With the fast-paced technological development and the expansion of networking technologies in particular; securing important and valuable infrastructures has become an essential part, from homes and small shops to huge organizations and institutions. Therefore, the video surveillance industry has swiftly implemented those technologies.

The availability of the gracious internet infrastructure has given an absolute leverage for the video surveillance migration to a much easier and cost-effective deployment that the previous traditional hardwired video surveillance (CCTV) systems (Figure 2-1) lack.

CCTV systems are composed of three main elements:

- Video capture units.
- Network transmission devices.
- Central control module.

The video capture units are a set of either analog or digital cameras that are supported by a video encoder device that has the ability to perform analog to digital conversion. This encoder captures and then compresses raw data into one of the popularly used standard formats (MPEG, H261, H263, etc.).

The second element in the components of a CCTV system is used for the transmission of the encoded video stream over an IP network, whether it is a local network or the internet.

The last element of the CCTV system is the central control module that can record and display each video channel separately and also controls the camera's behavior by sending control commands through the network to the designated camera. The control command has some unique features such as the relatively small packet size that only consists of a few ASCII characters, and that the control commands are not sent unless the user starts operating the system or a certain event occurs. It is also worth mentioning that the control commands security is crucial, as they might be carrying confidential data such as usernames, passwords, system admin configurations and other key information concerning the network.

There are two types of data flow in between the three composing elements of a CCTV system. The first type is the control flow that consists of the command issued by the control command transmitter, while the second type is the video stream of which the source can be an IP camera, IP video server, or a DVR.

The recipients of the digital video can be either a computer or a digital video decoder. The video data also has some distinct features, one of which is its huge consumption of bandwidth due to the fact that digital video stream requires up to 4 Mbps network bandwidth in most cases. Data in CCTV systems is live and the footage must be reachable instantly, and to fulfill the requirement of real-time video streaming and to cater to the vital property of time sensitivity, the system must have the ability to handle large network throughput and high processing power [2].

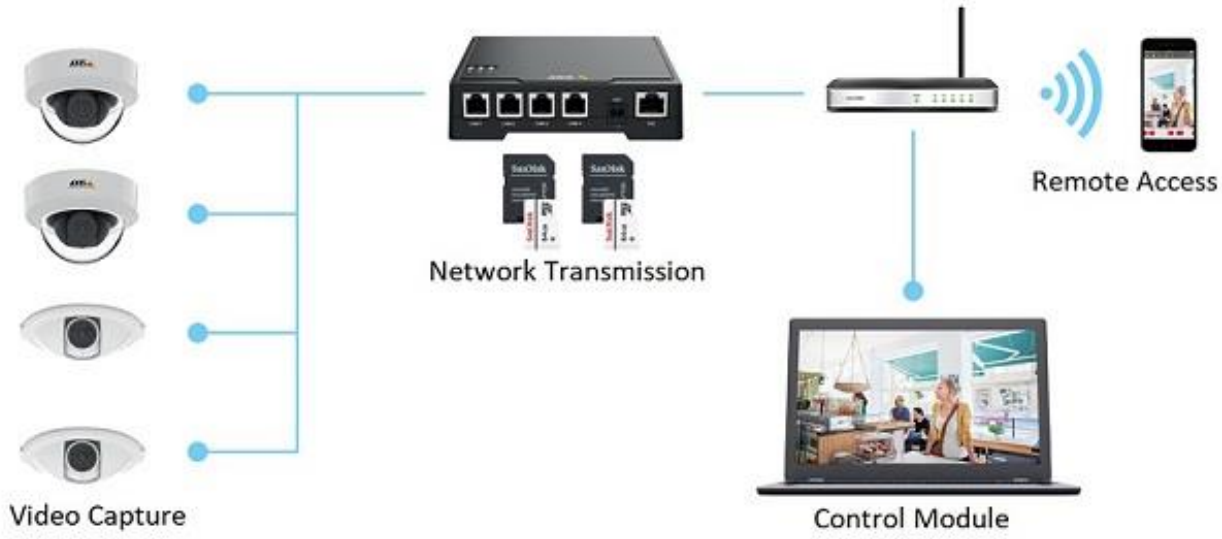

Figure (2-1): Video Surveillance System

2.2 Digital video:

CCTV security systems are ever-evolving. Businesses and institutions are finding themselves facing an important question: if they have an analog system, do they upgrade? And if they are installing a new one, do they select analog or digital? This section aims to answer and to elucidate those questions.

2.2.1 Analog, Digital and Megapixel:

The main difference between the analog CCTV system and the digital IP-based CCTV system is the method or the process of how videos are recorded and delivered.

In the analog CCTV system, the recorded footage is sent through coaxial cables to a DVR. Afterwards, the DVR converts the video from an analog signal to a digital one, and stores

the file on a storage device (hard drive, SD card, etc.) after compressing it. To view the video stream, either all monitors must be connected to the DVR, or the DVR is connected to a modem or a router to broadcast the footage over a network [2].

Digital cameras on IP-based CCTV systems, do not need a DVR to convert the signal, because it records in digital format from the beginning of the process. Then they proceed by sending or receiving data over a computer network instead of needing to go through a DVR. The two systems are very dissimilar in some characteristics, and each one of them comes with its own advantages and disadvantages [3].

Advantages of analog CCTV:

- Analog systems tend to be cheaper than digital systems, due to the less complex structure of its cameras.
- DVRs installation is easier than configuring a digital CCTV network, because it has just one device. Therefore, it costs less, and the installation is a bit more straightforward.
- Analog CCTV uses less bandwidth, because the recorded video files are smaller, and are transmitted to the DVR via coaxial cable instead of the commonly used LAN cables so the transmission does not take much bandwidth. Moreover, DVRs only use the bandwidth when the video is played, not on a constant basis.
- Analog cameras come with a lot more variety of options, therefore, it may be easier to find a camera model that suits your required properties.

Disadvantages of analog CCTV:

- There are a lot more cables in the analog CCTV system, due to the fact that each camera must be connected to both a power supply and a DVR. Therefore, the working area is untidier than its digital counterpart, moreover, coaxial cables are more expensive than cat5 or cat6, which are used for digital systems.
- Image quality on analog cameras is very low, which makes it difficult to identify suspects with a high level of confidence. In addition, it is not possible to zoom in on analog video, because the image will only get blurrier and more distorted when zoomed in.
- Analog cameras have a much narrower field of view, due to its inability to record high resolutions.
- The analog cameras must be positioned relatively close to the DVR to ensure the reliability of the connection, accordingly, it becomes limited to a specific range.

- DVRs have a constrained number of ports; therefore, you have to acquire a second DVR when you exceed the camera limit on your first device.
- Analog wireless systems don't work very well due to governmental regulations regarding the analog frequencies and signal strength. As a result, other wireless devices may interfere with the video signal.
- Encryption, or rather the lack of it. Due to the fact analog signals cannot be encrypted, they are a lot easier to spy on which is considered a security concern.

Advantages of digital CCTV:

- The image quality is noticeably better than that of an analog system, with many cameras that have the ability to record and transmit high definition video footage. Also, the digital video has the ability to zoom in on objects several meters away (may even reach 30m) while keeping up the quality of the video to a certain extent.
- Digital cameras have a coverage area of four or maybe five analog cameras. Therefore, fewer number of cameras is required to be able to achieve proper surveillance over the required area.
- Less cabling is needed, because, instead of wiring each camera to a power supply and to a DVR, all cameras can be connected to a switch. This switch can be connected to a NVR using a single cable.
- Digital cameras only need to be connected to the LAN network, therefore, there is no limitation concerning the distance between the cameras.
- By utilizing PoE (Power over Ethernet), switches enable your cables to provide power to the digital camera. Thus, reducing the need for extra cabling.
- The wireless abilities of digital cameras are much better than that of the analog cameras, due to the fact that it is not susceptible to the same interference that the analog cameras are affected by.
- Most digital cameras come with the ability to encrypt data. As a result, the footage is, to a certain extent, safe from malicious attacks.

Disadvantages of digital CCTV:

- The setup of the digital CCTV might be complicated, if the network is not already set up (e.g. the switches are not in place). This may increase the price of setting up the digital CCTV system. However, fewer parts are needed in setting up the network of the digital CCTV system, therefore, it sorts of stabilizes itself.
- Bandwidth consumption is much higher than in an analog system, because of the higher resolution and frame rate. Even with compression, we can expect around

720Kbps without even considering the newer cameras that have a megapixel resolution.

- Both higher resolution and higher frame rate lead to working with larger files. Consequently, a hard drive that has a significant storage space should be used to accommodate them.

2.2.2 Videos and formats:

All video formats are a result of a video compression algorithm. Video compression is the operation of using a codec to filter through the video file in order to reduce unnecessary files, thus, the video file size also reduces.

The two main types of codecs are H.264 and MJPEG. However, MPEG4 is considered as a relatively old version. Nearly all IP cameras come with a video compression codec, some of which are mentioned in detail below [1].

- **H.264 compression:** it is the most recent of the three codecs, it captures small groups of frames and eliminates any duplicate content that appears in each frame.
- **MJPEG compression:** also known as motion JPEG, it evaluates each frame of the video, and then proceeds to compress and send them as an individual JPEG images.
- **MPEG4 compression**: it is the oldest of the bunch. In case the audio is recorded with the footage, it is compressed separately. It has been mostly replaced by H.264.

2.2.3 Resolution:

The images in any screen or monitor, whether it is an analog CRT (Cathode Ray Tube) television or a computer monitor, is the grouping of the rectangular dots known as pixels, which refer to picture elements. All visual media is measured in pixels. In the analog world, each pixel is a combination of three colors: Red, Green, and Blue (RGB). Analog TV screens have a color depth of about 256 levels for each of the three colored layers, therefore, each pixel has around 16,8 million colors [1].

There are two common types of scanning techniques used by monitors to display images and videos, both of these techniques are discussed below:

Interlaced lines:

Analog videos are presented using something called Interlaced method of scanning. It operates on low-resolution analog TVs and monitors, interlaced scan-based images were developed for CRT televisions, it splits the visible horizontal lines on a TV screen into two

even and odd fields that alternately refresh 29.94 FPS (Frame Rate per Second). Any sort of delay in the rate of refresh will cause a jagged and distorted pattern between the two groups, such delay might be caused by slow shutter speed cameras because it they are simply not fast enough to capture the motion.

Progressive scanning:

a de-interlacing filter is given by the digital encoders that is used in the prevention of the jaggedness from an interlaced video, it is also worth mentioning that progressive scans require more processing power especially at high resolutions.

Progressive scan scans the entire picture line by line every sixteenth of a second and in sequential order without separating the scene into two separate groups. Digital technologies do not require the use of interlaced scanning for video presentation. Digital video does a very good job of eliminating the "flickering" effect as long as the monitor is also of suitable quality.

The analog TV format uses rectangular pixels; however, the HDTV format uses the smaller square pixels of the digital world. One NTSC (National Television System Committee) analog pixel holds four and a half times digital pixels (8-bit to 32-bit). therefore, each digital pixel holds four and a half more detail which results in a sharper and a clearer image.

All digital imagery and resolutions including printed material are pixel-based. The majority of digital monitors whether they are CRT or LCD have either 72 or 96 DPI (Dots per Inch), on analog CRT televisions the resolution is fixed, that is why the DPI will decrease depending on the screen size, the width of the pixel divided by the size of the screen will result in the screen's DPI depth, so you might have an excellent video signal, but the display might not be able to reach that level of quality.

Table (2-1): Digital Color Depth is The Number of Bits Per Pixel

Number of Bits	Number of Colors	Formula
8 bits	256 colors	2^8 ($2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$)
16 bits	65,536 colors	2^{16}
24 bits	16,777,215 colors	2^{27}

2.2.4 The World of Pixels:

In the mathematical world of digital video pixels are measured in bits and bytes rather than light emitting phosphor. In the digital world, 16 or 24 bits per pixel include the same 256 levels of color usually found in traditional televisions. The color resolution, also referred to as color depth, is how many bits of data that need to be used to store the information regarding the color of each pixel. Naturally, if more detail and colors are required in the image, the resolution must be increased (which in turn will increase the file size). The most commonly used color resolutions are 8-bit, 16-bit and 32-bit. [1]

Since everything digital is measured in pixels, there must be a balance between the display and the displayed resolution. for instance, a 50-inch LCD HDTV monitor has the capability of 1920 x 1080p, a regular low-resolution camera has the ability to generate about 352 x 240 CIF (Common Intermediate Format) image, and DVD players are usually 720 x 480. [1]

The CCTV software has to control how the footage is displayed on the security monitor, where DVD players automatically upscale the footage to full screen. However, since the monitor's resolution is much higher that of the DVD player; it must upconvert the 480p to fit the screen of the monitor. To simulate HDTV (1080p), an operation known as upsampling is needed. Upsampling is the operation of interpolating between neighboring pixels to estimate a reasonable value at the new pixel location. In other words, whatever digital display device is used, it most probably has the capability not to just stretch the 720 x 480 DVD video to full screen, but to also calculate the pixel valleys and fill up blank spaces to create a 1080i resolution.

There is a certain correlation between the pixels printed for hard copy, the pixels generated to be displayed and the pixels displayed. Most of the time we see on television shows, the law enforcement agencies capturing an image of the bad guy on the CCTV camera, and then they proceed to print it out as an extremely glossy high-resolution photo for everyone to see. This is not a result of a very capable CCTV system it is just the magic of Hollywood. In order to print an image of similar quality to the ones we see on television shows, nearly three times the number of pixels is required to produce such high-quality printed image. [1]

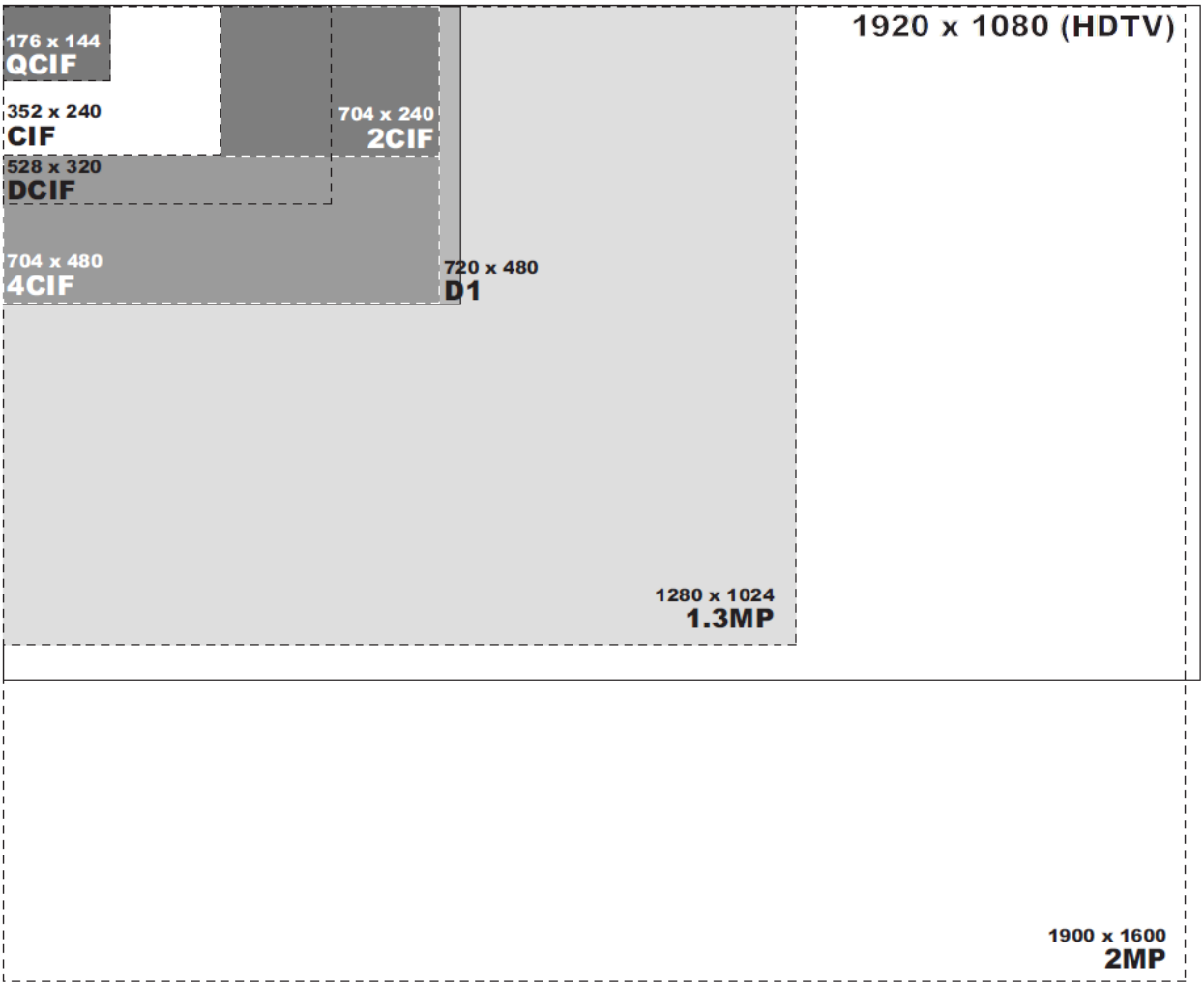

Figure (2-2): Digital Video Resolution and HDTV Monitor

Table (2-2): Display Standards and Their Attributes

Display Standard	Pixel Settings	Aspect Ratio	Total Pixels per Frame	Megapixels Identification
SQCIF	128 × 96	4:3	12,228	
QCIF	176 × 144	4:3	25,344	
VHS (NTSC)	320 × 480	4:3	115,200	
CIF	352 × 240	4:3	76,800	
½ D1*	352 × 480	4:3	168.960	
2/3 D1*	464 × 480	4:3	222,720	
DCIF	528 × 320	16:9	168.960	
S-VHS (NTSC)	530 × 480	4:3	192,000	
VGA	640 × 480	4:3	307,200	
Standard NTSC TV 480i	640 × 480	4:3	153,600 × 2 (Interlaced)	
2CIF*	704 × 240	4:3	168.960	

4CIF	704 × 480	4:3	337, 920	
Standard NTSC DVD 480p	720 × 480	16:9	345,600	
D1	720 × 480	16:9	345,600	
SVGA	800 × 600	4:3	480,000	
XVGA	1024 × 768	4:3	786,432	
XVGA+	1152 × 864	4:3	995,328	
	1280 × 1024	4:3	1,310,720	1.3 MP
SVGA+	1400 × 1050	4:3	1,470,000	1.5 MP
	1600 × 1200	4:3	1.920,000	1.9 MP
WSVGA	1680 × 1050	16:10	1,764,000	1.8 MP
WUXGA	1920 × 1200	16:10	2,304,000	2.3 MP
QXGA	2048 × 1536	4:3	3,145,728	3 MP
HDTV 1080p	1920 × 1080	16:9	2,074,000	2 MP
HDTV 1080i	1920 × 1080	16:9	1,037,000	
HDTV 720p	1280 × 720	16:9	922,000	
HDTV 720i	1280 × 720	16:9	461,000 × 2 (Interlaced)	

2.3 CCTV Components:

This section dives deep into the main components of a CCTV system, with an increased focus into cameras and their working mechanism as well as storage techniques for their active role in the creation of this project. [2]

2.3.1 The Eyes of Surveillance (the cameras):

The word “camera” is originated from the Latin word "camara obscura" which translates to "dark chamber". Cameras are considered the starting point of the video signal in CCTV systems. Therefore, it is a crucial component in the surveillance system.

2.3.1.1 Types of cameras from a technological point of view:

This section delves into the diverse types of security cameras according to the different technologies they use. Some of these types are going to be discussed in the following [2].

High Definition (HD) Digital/IP Camera:

Digital IP camera was created long after the invention of the analog camera. It was first introduced as a method for simplifying the process of linking CCTV cameras with an internal

network. It also came with a huge new abundance of features that the traditional analog camera lacked, such as:

- Higher possible resolution, given that it could easily reach 720 or 1080 rows of pixels.
- Better jitter handling, due to the use of progressive scanning.
- More flexibility in terms of physical capabilities, such as movement, not to mention the camera's ability to zoom in on objects without significantly distorting the image.
- It has a 16:9 format, which works better with wide screens.

Mega-Pixel Digital/IP Camera:

Megapixel IP cameras came as an advanced option to the HD Digital cameras, and its main feature is its ability to capture very high-resolution footage. However, this may result in an increased bandwidth usage, which in turn might strain the network. Therefore, megapixel cameras are another reason for developing an architectural approach to surveillance system design, due to the high network bandwidth consumption of the captured video. Some of its main features are:

- Capturing footage with 720 or 1080 rows of pixels easily.
- Has a 4:3 format.
- Has the ability to zoom in on small detailed objects while keeping the image’s integrity.

Standard Definition (SD) Analog camera:

Standard definition (analog camera) is the oldest of the three camera types, and as previously mentioned, analog security cameras have their foundation measured in television lines. And it is worth mentioning that the resolutions do not go a lot higher than what an analog television can display. However, analog cameras follow two international standards in the world of television, which are NTSC (used in North America and Japan) and PAL (used in other countries, mostly Europe). Some of its main features are:

- 704 x 576 resolution (PAL).
- 704 x 408 resolution (NTSC).
- Sufficient to meet many surveillance applications.

2.3.1.2 Types of cameras from physical design point of view:

This section mentions the different types of security cameras according to their physical design rather than what type of technology they use. The following are the five popular camera types that are worth mentioning [4].

Dome CCTV Camera:

Dome CCTV cameras get their name form their dome-shaped housing. This particular design's purpose is not to make the camera hidden or covert, but to make it unobtrusive. Thus, it lets the "bad guys" know that the facility is being watched, and it also makes the customers or employees feel safe.

Bullet CCTV Camera:

Bullet CCTV cameras take their name from the cylindrical and tapered shape of a "rifle bullet". This camera is not specifically designed to pan, tile, move and zoom, however, it is meant to capture footage from a fixed area. Many bullet cameras are water proof and are installed in a protective casing.

C-Mount CCTV Camera:

The special feature about C-Mount CCTV cameras is that it is possible to detach their lenses to fit different situations.

Day/night CCTV Camera:

Day/night CCTV cameras have the ability to operate in both normally-lit and poorly-lit areas, they are mostly used in outdoor environments.

Infrared/Night vision CCTV Camera:

Infrared/Night vision CCTV cameras have the distinct advantage of seeing in pitch black conditions using infrared LEDs.

2.3.2 Lenses:

Using the lenses with the right properties is crucial for the CCTV system to operate in the required way. There are certain attributes that must be taken into account to ensure the selection of the right lens, some of which are the following [5].

- **Manual Iris:** usually used indoors where light levels are fixed, due to its inability to adjust to any change in lighting.
- **Auto-Iris:** can be used in different environments due to the aperture ability to automatically adjust as the light levels change.
- **Focal Length:** the size of the lens (2.8 –60mm), or the distance between the center of the camera's lens and its focus.

Today's cameras have a better ability to adjust to the various changes in the light levels without the need of an auto-iris.

2.3.3 Transmission Medium:

There is a number of different methods that can be used to transfer the footage between the elements of CCTV systems, the most common mediums will be discussed below [1].

RG-59/U Coax Cable (traditional method):

An electrical device that converts balanced and unbalanced signals is called a balun. It can be used to convert the video signal to TCP/IP, so that the older existing cable plants (RG6/RG59) can be used in today's ethernet environment.

Category 5/5E/6 Unshielded Twisted Pair (UTP):

These are the cables that are usually used to connect network devices in a LAN network. Thus, using it to connect CCTV cameras to the network might be very beneficial, because it will have the ability to utilize existing LAN network cables, as it also reduces the space used in conduit trays.

Fiber Optic Cable:

Fiber optic cables are the fastest option; however, it is also the most expensive one due to its immunity to strong EMI/RFI signals. It also comes with a very large amount of bandwidth.

RF Wireless Systems:

This option requires fewer cables, which results in reducing the tedious work of having to set up routes between each camera and the network device. However, it does come with its own set of drawbacks that include the need for a LOS (line of sight) between the sender and the receiver, not to mention that a government license may be required.

2.3.4 Monitors:

In the end of the 20th century a computer bug caused a worldwide problem that costed nearly 300 billion dollars to fix. This bug's name is Y2K, and it happened because the software at the time would only store two digits of the four digits in a year, i.e. 1970 would be stored as 70. Unfortunately, when the calendar rolled to Jan 1 ,2000, most running software at the time thought the date was Jan 1, 1900 (or in some rare cases Jan 1,19100).

This bug caused multiple issues such as alarms going off at the wrong time, and bills shown to people as a century overdue (with billions of dollars in late charges), but it is also regarded as a reference point in the development of monitors and recording devices, due to the very different characteristics of the devices that came after and before Y2K [5]:

Pre-Y2K:

- Monitor sizes were limited to 9" and 12" black and white tube monitors.
- Burn-in images were a concern. Burn-in is the discoloration of certain areas on an electronic display, which is caused by cumulative non-uniform use of pixels.
- Very high maintenance cost (replacing tubes).

Post-Y2k:

- LCD was introduced and followed by newer technologies of LED & plasma.
- Very wide range of sizes 4" to 52".
- Longer product lifespan

2.3.5 Recorders:

Monitors were not the only elements that evolved in that period. Video recording also took huge strides in its evolution [5].

Pre-Y2K:

- Video footage was recorded on time lapse video recorders (tapes).
- It had a very limited recording capacity.
- Difficult to find specific time/date of images.

Post-Y2k:

- Images can be easily stored on electronic storage devices (Hard Drive, SD card, etc.…).
- DVRs (Digital Video Recorders) were introduced as a standalone option.
- NVRs (Network Video Recorders) were introduced to assist in the linking of CCTV cameras to a local network (LAN) or the internet.
- Higher level of search and analytics capabilities with increased control options.
- Transferring images became much easier due to the fact that it is able to store them on electronic storage.

2.3.6 Controlling Techniques:

The control mechanisms of CCTV systems were replaced by more advanced techniques, some of which are listed below [5].

Baluns: convert between a balanced and an unbalanced signal to run over an Unshielded Twisted Pair (UTP).

Encoder: converts the video signal from analog to digital and compresses it.

Enclosures: containers to host the sensitive equipment of the CCTV system, they may be basic outdoor compartments to bullet/explosion proof rooms.

Pan/Tilt/Zoom (PTZ): controls used by the system admin or user to operate the surveillance camera.

Transmitter/Receiver: used in a wireless environment to transfer the signal, or in Fiber Optic to convert the electrical signal to light and vice versa.

Video Switchers: used in the routing process of either analog or IP video signal to multiple monitors.

Cooling Devices: some high performing CCTV systems require cooling devices to keep an ideal ambient temperature of 85 degrees.

Mounts: (i.e. wall, ceiling, parapet, pole, etc.).

Power Supplies: the devices or outlets used to power the various components of the system (e.g. 24Vac, POE, Integrated UTP).

Racks: enclosures ranging from wall mount to uprights.

Uninterruptable Power Supplies (UPS): protection against power sags, surges and blackouts.

2.3.7 The Storage (how data is stored in CCTV systems):

Storage in CCTV systems has become so intertwined with the network that the functionality should go beyond just the camera capacity or the software capabilities and factor in the IT infrastructure of the entity that is implementing the system in order to be able to service the required professional equipment. [6]

Normally, CCTV camera footage is stored on the NVR's HDD, DVR's HDD, computer, SD card or even through a cloud storage service. NVR, DVR and computers support 24/7 recording and motion detection recording, while security cameras with SD cards only supports motion detection recording.

The DVR and NVR are equipped with an HDD that starts from 8 TB up to 32 TB in storage, thus, there won't be any issues in storing footage of long durations.

Data storage in video surveillance could require multi-thread sequential operations. When a surveillance system consists of multiple cameras, which usually happens, all writing

operations go to the same storage, and this could cause considerable multi-thread write workload.

It is important to balance out between the storage performance, density and cost of the storage method that the surveillance system will be using. Therefore, it is necessary to include a high-speed software defined technology that complies the HDD to deal with sequential workload requirements.

The storage performs the worst when there are multiple workstations from various operators running playback. However, this is highly impractical given the fact that the vast majority of the workload is for writing, with some insignificant reading workload, which usually comes from viewing archived footage [1] [6].

2.4 Types of CCTVs:

The section below outlines the evolution of CCTV surveillance systems [7]:

2.4.1 VCR-based analog CCTV system:

This is one of the first used CCTV systems, which was used to utilize analog cameras that were connected to a VCR for recording video. The entire system (Figure 2-3) was analog, the VCRs used the same cassettes that were used in home VCRs. A coaxial cable had to be stretched out from the VCR to every analog camera. The captured video was not compressed in any way, and when the cameras were recording at full frame rate, one tape would only last about eight hours.

After a period of time, the time lapse mode was incorporated into the VCRs to make the tape last much longer. The time lapse mode enabled the industry to come up with some specifications, such as 15 FPS, 3.75 FPS, and 1.875 FPS, as these were the possible recording frame rates made by an analog system.

Thenceforward, the quad mode was presented, which was basically the method used to deal with CCTV systems that had multiple cameras. Quads simply took the input from four cameras and displayed them in one, which reduced the need for extra television screens.

As time went on, a set of drawbacks of the analog system started to appear. These drawbacks included limitations in scalability and the tedious manual job of changing the tapes, not to mention that the quality slowly deteriorated and became obsolete over time, and the cameras only recorded in black and white which quickly became unacceptable, hence impractical.

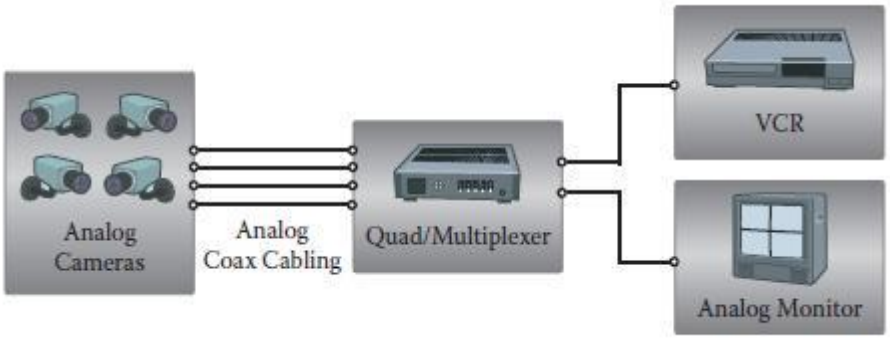

Figure (2-3): VCR-based Analog CCTV System

2.4.2 DVR based analog CCTV system:

In the late 20th century, the video surveillance industry witnessed its first ever digital resolution, with the introduction of DVRs. DVRs were better than VCRs in many aspects, because the recorded video was digitized and then compressed to store as many days' worth of footage as possible. It also replaced the tapes that were previously used by VCRs with HDDs.

In the early years of DVRs, the hard drive storage capacity was limited, so a lower frame rate had to be used or a limit to the recording duration had to be specified. Therefore, manufacturers started developing proprietary compression algorithms in order to reduce the storage space that was occupied by the recorded footage. Although with the drastic decrease in the prices of hard disks over time, most manufacturers gave up their development of compression algorithms in the favor of standards such as MPEG-4.

DVR also replaced the multiplexer as well as the VCR, and thereby reduced the number of components in the CCTV system. Also, because the DVR made the digital video available, it became possible to transmit digital video over longer distances. DVR based CCTV can be seen in (Figure 2-4).

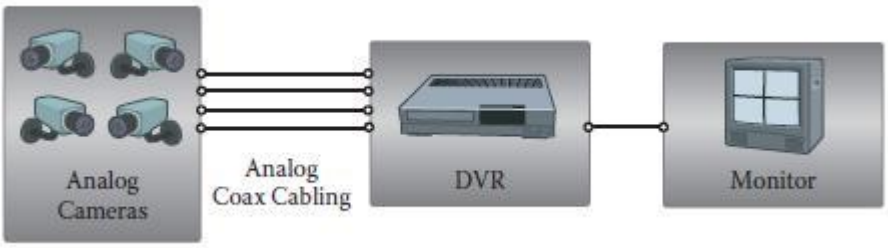

Figure (2-4): DVR Based Analog CCTV system

2.4.3 Network DVR based CCTV system:

After a decent period of time, DVRs were equipped with an ethernet port to provide internet connectivity. This presented a new type of DVRs called the network DVR (Figure 2-5) that gave the admin or the user the ability to remotely monitor the video stream by using PCs. Some DVRs requires special software to monitor the video, whereas others use a regular web browser to do so, which is a lot more flexible.

Even though the DVR came with many advantages over the VCR, it did come with a few drawbacks. For instance, the device was burdened with a large number of tasks, such as video compression, recording, digitization, and networking. Virus protection was also a challenge, even though the DVR was mostly run on a Windows environment. Its proprietary interface did not allow virus protection, moreover, it was not very cost effective, as most DVRs came with 16 or 32 inputs, which is highly counterproductive if you have a system that only needs 10 inputs for instance, and it also makes scalability a bit difficult.

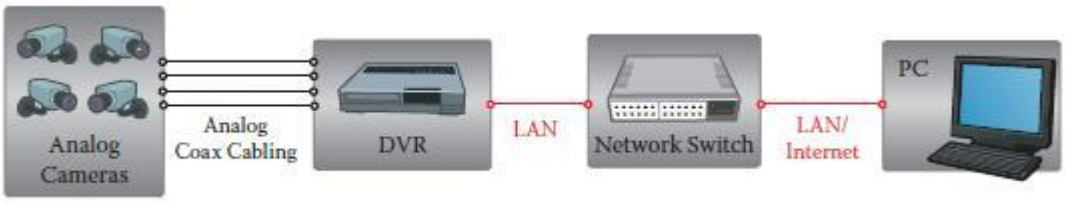

Figure (2-5): Network DVR based CCTV System

2.4.4 Video encoder-based network video system:

To tackle the scalability issues of the network DVR based CCTV system, a new networked video system was introduced (Figure 2-6), which separated the DVR into two components: the video encoder and the PC server. The video encoder connects to the analog cameras, digitizes and compresses the captured footage, then this footage is sent through the network to the PC server that runs a video management software that is used for monitoring and recording of the footage. Therefore, both the digitization and compression are handled by the video encoder, and the recording and storage are handled by the PC server which ultimately simplified the process of scaling up this network.

Even better alternatives are the NVRs and Hybrid DVRs, they are a proprietary hardware box with reinstalled video management software for the management of video encoders and cameras. Hybrid DVRs handle both analog inputs and network videos, whereas NVRs only handle network inputs. The main benefit of using NVRs and Hybrid DVRs is the easy installation, because all the management and recording is done in one place.

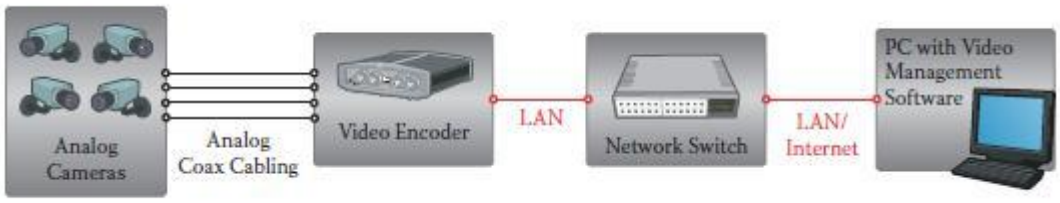

Figure (2-6): Video Encoder based Network Video System

2.4.5 Network camera-based video system:

The network camera-based video system (Figure 2-7), commonly known as the IP camera, is a camera that has the ability to establish an IP network connection. Since the video is transported across an IP network, the system is fully digitized starting from the camera and passing through switches and other IP network components all the way to a PC server that is running a video management software.

One of the main advantages of this system is that the captured footage is recorded in digital format from the beginning, which ensures a consistent image quality across the system. It is important to note that for every conversion between analog and digital, a certain level of image quality is lost.

Some of the other benefits of network camera-based CCTV systems is that it provides a means for IP cameras to share the same physical cable. Also, it gives the ability to carry power to the cameras through an ethernet cable (PoE), it also has the ability to carry two-way audio. Not to mention the ability to remotely configure the network.

Therefore, network-based CCTV is considered as one of the best options for a surveillance system, because it provides a flexible and cost-effective transmission medium, which in turn gives it the scalability of network systems, thus, the video surveillance system can be scaled up to hundreds or even thousands of cameras.

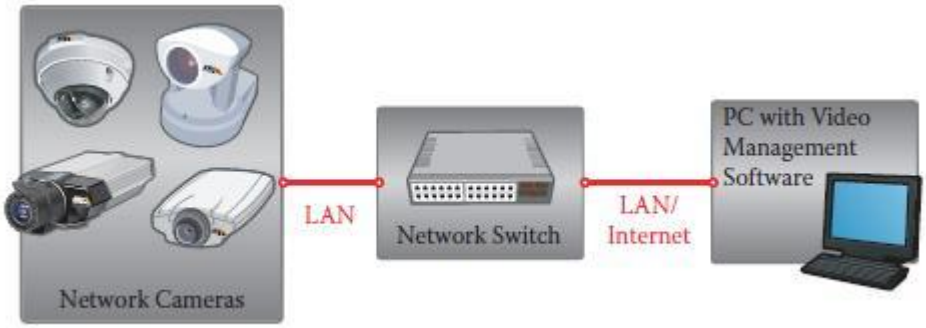

Figure (2-7): Network Camera-based Video System

2.5 Wired vs Wireless:

There are two types of transmission mediums that can be used for CCTV systems, this section focuses on dissecting and comparing them.

2.5.1 Wired:

CCTV systems with wired transmission medium have been around the longest. They are hardwired into landlines and home electrical systems [2].

Advantages of wired security systems:

- Less likability to fail than wireless systems, because as long as the wiring is intact, the signal will reach its destination. Unlike wireless systems, which rely on sensors.
- Suitable for large areas, because the cameras only need to be set up and connected to the wiring system.
- More secure than wireless systems, because the hacker needs physical access to the system to tamper with it.
- Preferred by professionals, for the reason that hardwired systems are usually more reliable, secure and consistent.

Disadvantages of wired security systems:

- Costs more to install, because the installation of wired systems involves connecting sensors with low-voltage wires that are inside the walls. Accordingly, professionals are required to drill holes and set it up.
- One significant vulnerability is that if the malicious attacker figures out that the security system is wired, he may be able to cut the phone lines outside the facility.
- Very difficult to uninstall.
- Usually controlled from one central control panel, and for buildings with multiple stories, each floor is equipped with its own control panel.
- It may take a long time to install in complicated environments.

2.5.2 Wireless:

Wireless systems aim to address the most significant downsides of wired CCTV systems. Its main concept is to connect all your network devices through WI-FI. However, the devices still require wired power [2].

Advantages of wireless security systems:

- Simple and fast to install, all you need to do is to set up the cameras and connect them to your WI-FI network, without the need to run a cable from your cameras to

the recorder. The only limitation in installing wireless cameras is probably the need to attach the camera to a power source.

- Easy to uninstall and move, which is why wireless security systems are ideal for renters or businesses in temporary locations.
- Easily modifiable, because its components are easy to upgrade and modify since it is not hardwired.
- Can be accessed remotely and through multiple devices, unlike wired systems where it is confined to a single control panel.

Disadvantages of wireless security systems:

- Even though it is very uncommon, wireless security systems are susceptible to interference. However, there are many factors that can affect the signal, whether the interference came from electromagnetic waves, power lines, or through metal filing cabinets.
- Battery life, if the installed devices run on battery they must be periodically checked.
- Distance limitations, where the WI-FI signal cannot travel long distances. Therefore, wireless security systems are best suited for small to medium environments.
- Poor wireless security systems can be hacked if the feed is not encrypted properly.

Summary:

In this chapter, we briefly explained CCTV systems and their working mechanism, as well as concisely described its components. Regularly, we began by an overview on what a CCTV system is, and talked about how a video is constructed, and how computers and monitors interpret visual media. After that, we dissected the CCTV components while focusing on two main components: the camera and the storage. Afterwards, we proceeded to list the different types of CCTV systems. Finally, we concluded by comparing different types of transmission mediums while listing their pros and cons.

CHAPTER THREE: IP NETWORKING AND VIDEO SURVEILLANCE:

3.1 The Effect of Networks on Video Surveillance:

In the world of technology, everything is evolving rapidly, every day a new technology is introduced that renders another one obsolete, which applies to the video surveillance world as well. The video surveillance world had been using analog as a starting technology. Afterwards, digital technology was introduced, and it took the surveillance world by storm due to its clear advantages over its analog counterpart. However, it did not contribute in simplifying the tedious installation, management and configuration that the analog system had. Therefore, network-based surveillance systems were introduced, which utilized IP cameras to link the cameras to a LAN network, which therefore was a huge step for the video surveillance world, because integrating the CCTV system with the network meant that it could benefit from all the features the network had, such as reliability, easy management, flexible camera deployment and easier data integrity preservation.

3.2 IP-Based CCTV Cameras:

IP camera or (Internet Protocol Camera) is a digital video camera that receives and sends data over a LAN network or through the internet. Its most popular application is in video surveillance, due to its wide adaptation as a new solution in the industry. Some IP cameras require an NVR to deal with the captured footage, while others act in a decentralized manner with no NVR needed, because it stores the footage directly into a local or a remote storage media.

The early generation of IP cameras used common television broadcasting formats, such as CIF (Common Intermediate Format), NTSC, PAL and SECAM. However, since the start of the 21st century, there has been a significant shift towards higher resolutions (1080p, 4K) and widescreen formats (16:9) [8].

3.3 The Advantages of IP Networks for CCTV:

IP cameras are dominating the video surveillance market, and the rise of the network-based security systems has elevated user expectations and changed the way surveillance systems are conceived, managed and implemented. On the whole, IP cameras are very different from

analog cameras [8].

This section mentions the main advantages of IP based CCTV systems [8]:

- IP cameras can have multiple sensors, which means an IP camera can contain three or four cameras, which therefore gives one IP camera the ability to cover an area that normally requires four analog cameras.
- Technology products prices are forever going down, and IP cameras are no exception. Even though when they were first conceived, only strong and well-funded institutions could afford an IP based CCTV system. Nowadays, the prices of IP cameras have gone down drastically that the average person can now afford to install a system in his home.
- Simplicity in installation. While in analog systems two wires are needed for each camera, one for power, and another for it to be attached to a device (usually a DVR). However, only one wire is needed for both data and power in IP cameras.
- The resolution, where the quality of footage in video cameras just keeps getting better and better, and IP cameras have a clear superiority over their analog counterparts.
- Intelligent cameras, where IP cameras have a very distinct feature that the analog cameras lack, which is its ability to perform data analytics within the camera, because IP cameras are basically small computers that compress and send the captured footage. They can also detect movement, identify shapes and track certain colors, keeping in mind that an analog CCTV system needs to separate devices for capturing and compression.
- One other clear and very important advantage is the IP camera’s ability to encrypt the footage, which results in securing the transmission.
- IP cameras do not need encoders and decoders; therefore, it uses less equipment than analog cameras.
- IP cameras have the ability to use wireless networks or WI-FI.
- Remote access.

3.4 The Drawbacks of IP Networks for CCTV:

This section lists the disadvantages that come with IP based surveillance systems [8]:

- The initial cost of setting up the system may be high if you are migrating from an old analog surveillance system. However, when it is already set up, it is much easier to tailor it according to your needs.

- IP cameras record videos at a much higher resolution, therefore, a large storage capacity is needed.
- Setting up IP cameras for broadcasting over the internet might be a bit difficult.

3.5 IP Video Architectures:

There are three different options in terms of IP video architectures (Figure 3-1). For the systems that are still using analog cameras, network DVRs and analog encoders make it possible to digitally encode the analog feed and link it to an IP network. These systems are usually confined to a low-resolution video feed.

IP cameras allow the capture and recording of much higher resolutions, however, these IP cameras require an entirely different infrastructure when compared to DVRs and encoders, and they may even require more network infrastructure work when upgrading from an analog system. Figure (3-1) shows a block diagram of the usual IP based video surveillance network [2].

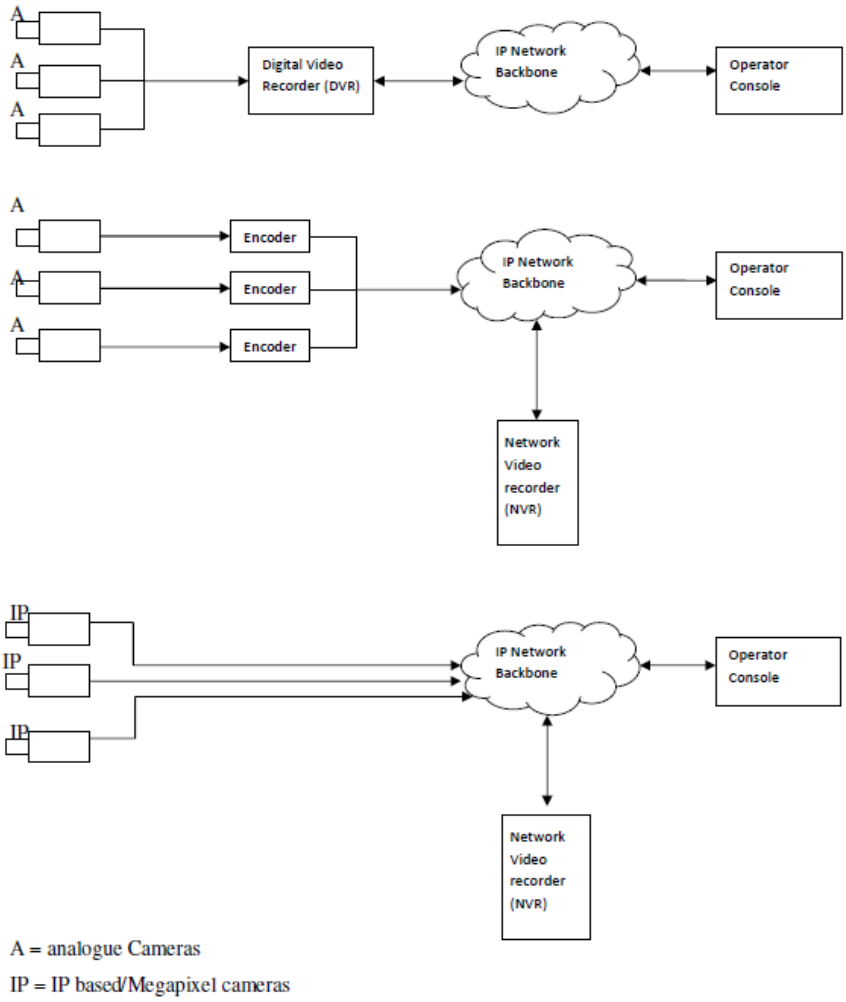

Figure (3-1): CCTV Systems Architectures

3.6 Digital Data Compression and Decompression:

When footage is captured by a CCTV camera, it is necessary to compress the captured footage to reduce the used bandwidth and storage space taken by the video. Compression is done by what is called a CCTV codec algorithm, such as MJPEG, MPEG-4, H.264 or H.265. A CCTV codec algorithm works the same way as most of the compression algorithms work. It looks for and eliminates redundant information to reduce the size of the captured footage. The part in charge of capturing and transmitting the footage across the network is also responsible for the compression of the video, while the receiving side (computer, mobile phone, tablet, etc.) is in charge of decompressing it.

In analog CCTV systems (Figure 3-2), a DVR is needed to convert the received signal from analog to digital, and it also uses a compression algorithm to compress the video before storing it in the HDD or transmitting it live to a remote host.

Figure (3-2): Analog CCTV System

In the IP based CCTV (Figure 3-3), the IP camera has the ability to digitize and compress the video, so that it is sent directly to the NVR to be stored in the HDD.

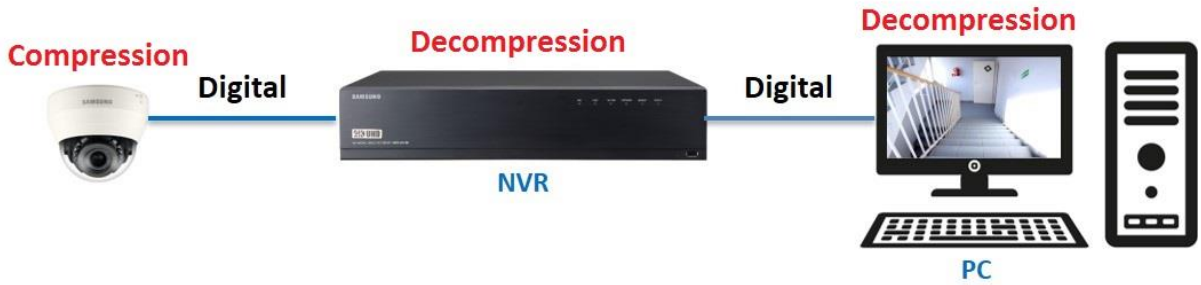

Figure (3-3): IP-Based CCTV System

However, video compression must be a carefully thought out process, because if some information is discarded from the digitized video; the video quality decreases. And this is

why the compression algorithm must discard just as much information as possible with minimal impact on the quality of the video, and every codec algorithm has its own way of accomplishing this task [9].

The most common CCTV codecs are [9]**:**

- MJPEG: this codec follows a very simple concept, where it sends a set of complete images in rapid succession to give the impression of movement. It is usually limited to 15 frames per second which is enough to show the video in good quality without taking too much bandwidth and storage (Figure 3-4).

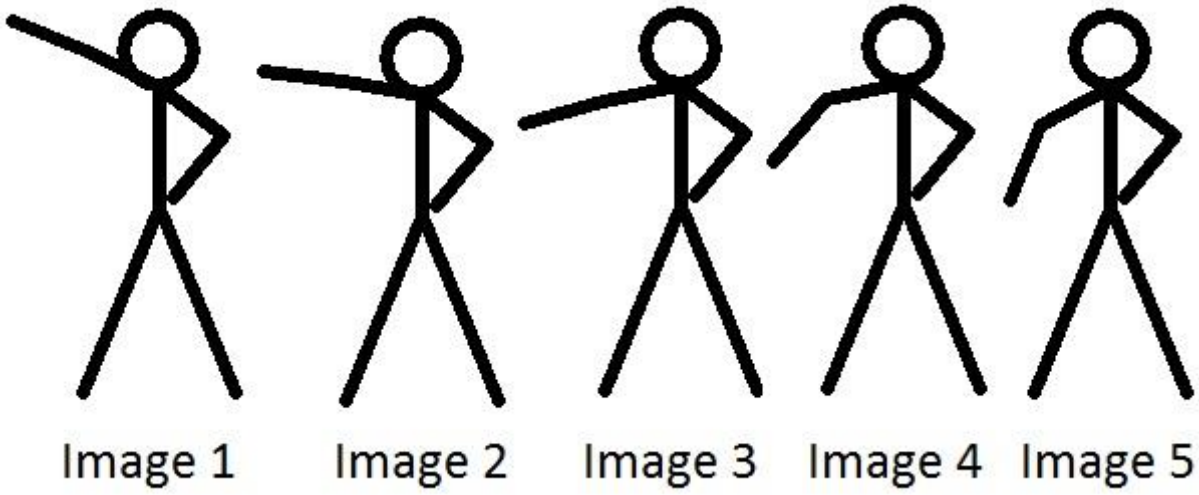

Figure (3-4): ADdemonstration of MJPEG Compression Algorithm.

- MPEG-4: while the MJPEG sends full images to be put in a sequence, MPEG-4 sends a combination of full and partial images. The concept is that the receiving device (NVR, computer, etc.) is in charge of organizing and assembling the video that was constructed from the sent images (Figure 3-5).

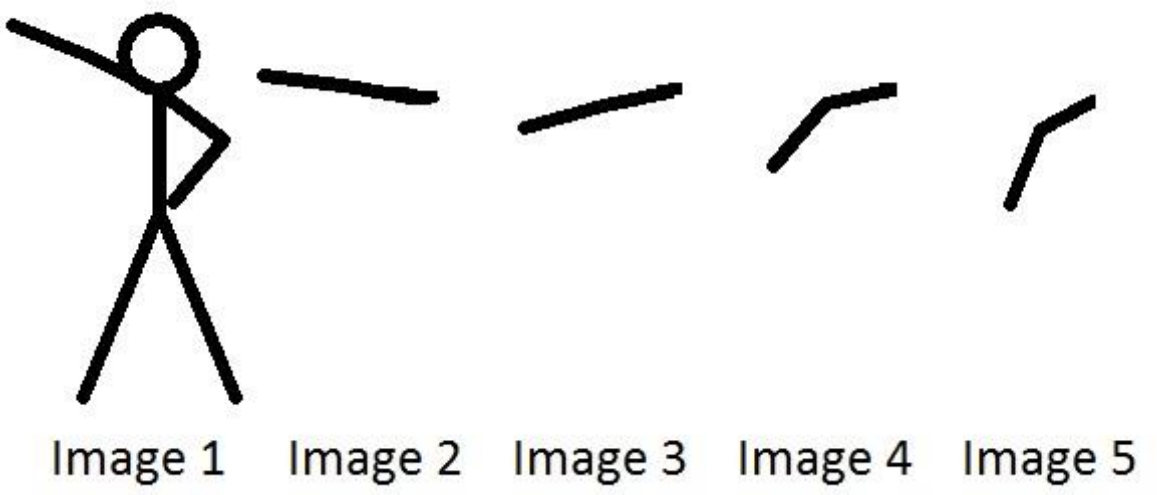

Figure (3-5): A Demonstration of MPEG-4 Compression Algorithm.

- H.264: this codec is basically an improved MPEG-4. It uses the exact same concept but with less bandwidth and storage consumption.

3.7 Network-Based Devices: Servers and Workstations:

With the introduction of IP cameras into the surveillance world, a lot has changed in terms of required equipment and knowledge needed to operate the system. While network admins used to run software products on commercial servers in the past, these days they must have a decent amount of knowledge on how the server's hardware and software works, both to implement the system and to maintain it.

Computer security now comes into play as well, because it is necessary to authenticate the users and authorize them to access a certain stream from the network. Therefore, it has become crucial to implement proper security precautions and protocols. And servers must also maintain a carefully audited access control list to prevent unauthorized access.

The software that the servers will be using must be selected according to the conditions and requirements of the site. For instance, stable server platforms such as Linux are an option. It is highly recommended to lock these workstations to a point where minimal services are running, which prevents the allocation of resources to non-security operations because downtime or rebooting is unacceptable in any security system [2].

3.8 Network Bandwidth Consumption of IP Cameras:

This section will address exactly how to measure the bandwidth that an IP camera will consume, in order to have the ability to make an accurate estimation of the overall bandwidth required to operate a certain network.

There are four elements that directly affect the bandwidth used by IP cameras:

- Resolution: the higher the resolution, the more bandwidth is required to carry the footage.
- FPS: the higher the FPS, the more bandwidth will be required to handle it.
- The Codec: basically, what compresses the video footage, where the size of the footage affects the bandwidth.
- The number of IP security cameras.

In order to determine the bandwidth that the IP cameras will consume, a bandwidth calculation formula must be used:

$$\text{Bandwidth (Mbps)} = \text{Bitrate (Main)} * n + \text{Bitrate (sub)} * M \quad (3\text{-}1)$$

N & M represent the number of IP cameras for mainstream and sub-stream. As for the bitrate, its value can be found in the specifications of the IP camera [10].

3.9 Network Delivery Methods and Protocols:

There are a number of protocols used to transfer data across the network. In this section, we will bring up a set of protocols that are used in video streaming in CCTV systems [1]:

3.9.1 Transfer Control Protocol (TCP):

The TCP is not the ideal choice for any sort of digital stream, especially live streams since it needs a connection-based transmission channel. If TCP is used, the demand on network resources will increase, which will cause the bandwidth to close up and acknowledgment packets to be lost, which therefore will cause network latency issues. TCP is considered a synchronous transmission method, due to the fact that if it does not detect a connection, it will keep sending data until it is synchronized.

3.9.2 User Datagram Protocol (UDP):

UDP is an asynchronous protocol that does not contain any dialog between the sender and the receiver, which makes it an ideal choice for streaming live video footage. However, it is worth mentioning that most firewalls block UDP packets by default for security reasons,

which makes it difficult for a video stream to reach its destination. UDP is an unreliable and a connectionless protocol with no acknowledgment whatsoever on whether the UDP packet has reached its destination or not. Nevertheless, dropping a couple of frames every second seems like a worthwhile risk to ensure the continuity of live stream.

There are three main methods of delivery that are used by CCTV systems, which are [1]**:**

- Unicast: this method of delivery is used when there are only one sender and one receiver on a one on one connection. Unicast (Figure 3-6) packets are encapsulated and delivered using TCP. The majority of network and internet traffic is unicast, where sent data is only addressed to a single recipient. However, in some cases, unicast can be counterproductive, because it streams a single stream of footage per host. As in each host will have its own broadcast of the streaming media. i.e. if we have a stream with an average of 2 Mbps and five clients decide to connect concurrently, that is 10 Mbps streaming through the network.

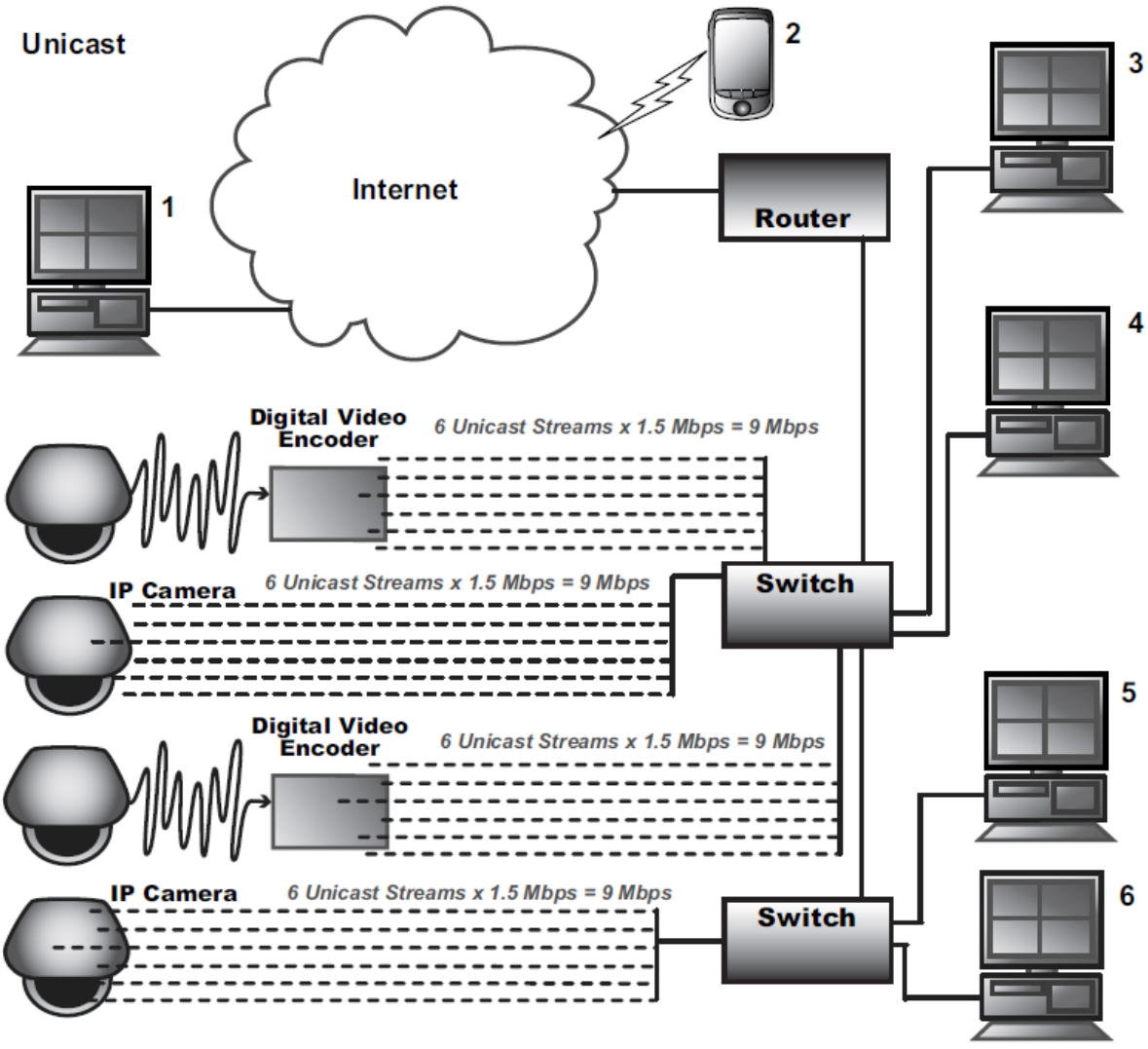

Figure (3-6): Unicast Delivery Method.

- Anycast: basically broadcasting, it is similar in a way to television broadcasting where one sender transmits a signal and it reaches all the connected devices on the network, whether they want to receive it or not. However, broadcasting is not recommended for CCTV systems as they might cause broadcast storms.

- Multicast: multicast (Figure 3-7) is the ideal delivery method for CCTV networks that has multiple recipients for the captured footage, because it reduces the network traffic drastically by sending only one video stream to multiple users. This delivery method cannot use TCP, due to the fact that it is only one way, therefore, multicasting is only sent using UDP.

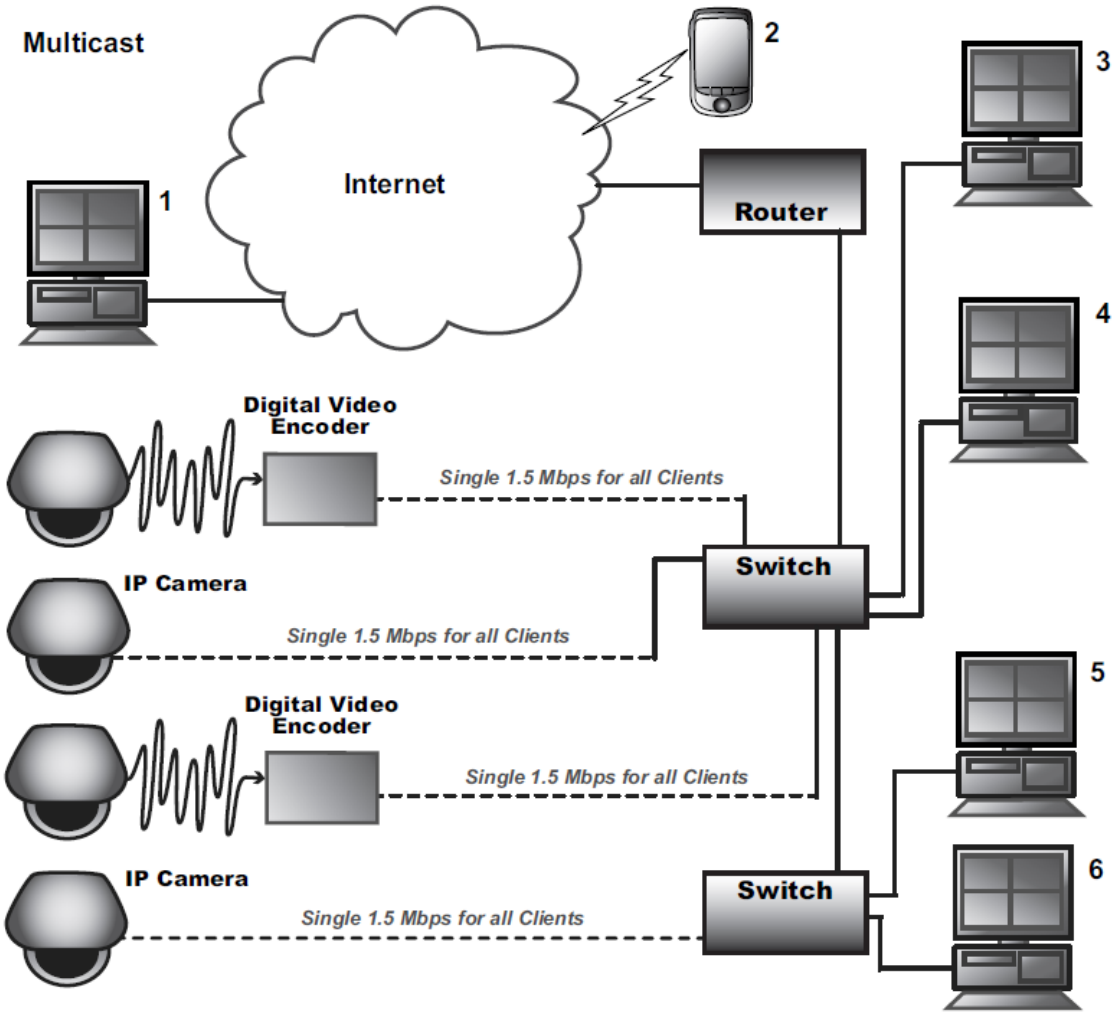

Figure (3-7): Multicast Delivery Method.

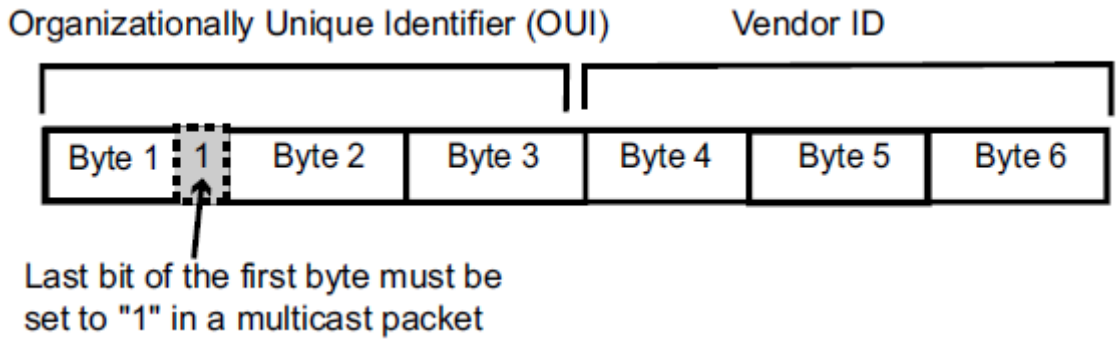

Figure (3-8): OUI of a Multicast Packet.

3.9.3 Real-time Transport Protocol:

Real-Time Transport Protocol or RTP is a protocol that is typically used via UDP. It does not actually guarantee real-time, but it does optimize the synchronization and control of the streamed media. The MPEG-1 and MPEG-2 formats come with their own synchronization and video conferencing, but they usually take too much bandwidth for just the streaming of a few digital video cameras. MPEG-4 does not have a built-in video/audio synchronization. Therefore, the RTP provides the necessary timestamps for video/audio synchronization for MPEG-4. Most IP cameras and encoders provide interoperability with other devices to support RTP for transport.

3.9.4 Real-Time Streaming Protocol:

Real-Time Streaming Protocol or RTSP is the control protocol for RTP. What RTSP does is that it calls the unicast stream into the VMS and/or video player interface.

3.9.5 HyperText Transfer Protocol:

HyperText Transfer Protocol or HTTP is the protocol used by web browsers to access the web through web browsers. And it is used to access the interface of the web server for the digital encoders and IP cameras using a plug-in for the browser to view the cameras.

3.10 Remote Access:

One of the main features that became available to the system admin or users once they integrate their CCTV system with their network is Remote Access. As the system needs the ability to be available on any device at any time as long as there is an internet connection. Because once the system and IP cameras are configured properly with internet connection, it becomes possible to access the video stream from any device that is connected to the

internet (assuming the owner of the device has the required credentials to access the video stream) [1].

Summary:

In this chapter, we delved into how exactly networking affected the video surveillance world, and how the introduction of IP cameras presented a new set of features to video surveillance. Furthermore, we demonstrated how the network handles these new components in terms of bandwidth and protocols to be used.

CHAPTER FOUR: MACHINE LEARNING AND COMPUTER VISION:

This chapter gives a simple introductory to machine learning and the computer vision technology, which is an emerging interdisciplinary scientific field.

4.1 What is Computer Vision?

Computer vision (also known as CV) is often defined as a field of study that has the aim of developing techniques in order to help computers see and extract useful information from digital images, such as photos and videos. However, this has proven to be a very challenging task, where one of the problems of computer vision is that it appears as if it is unsophisticated, which is simply due to the fact that the problem is pettily solved by people. A huge part of this problem is the complexity of visual data, for example: imagine a photo which contains cars, high buildings, and a sky. As you may have imagined, in this particular photo there are hundreds of objects and almost all of them are partly occluded. For humans, identification of objects and the differentiation between them is much simpler than for a computer vision algorithm. For instance, a human can simply determine where one object ends and another one begins, whereas a computer vision algorithm will find that very challenging, especially if not trained well [11].

4.2 Artificial Intelligence, Machine Learning and Deep Learning:

Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) (Figure 4-1) are all considered as computer science fields. Even though the three terminologies are regularly used interchangeably and are strongly related to each other, they do not quite refer to the same thing and one might not distinguish between them. Nevertheless, to fully understand these terminologies, we are going to go through each term individually.

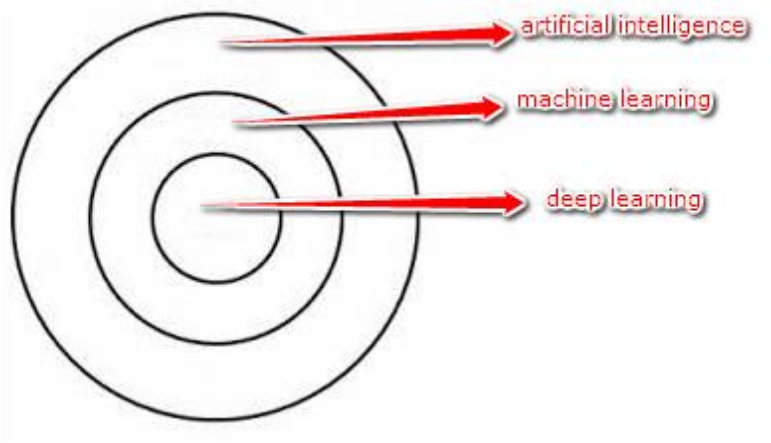

Figure (4-1): The Relationship of Artificial Intelligence, Machine Learning and Deep Learning

4.2.1 Artificial Intelligence:

There are arguably many definitions for the term "artificial intelligence". The term itself awakens emotions and raise questions in the human mind. Questions such as "What is Intelligence?", "How does our brain exactly work?" and "is intelligence measurable?". Artificial intelligence has always been an area of debate from the many different perspectives and points of view that are sometimes scientific and other times are philosophical. All these questions are significative when trying to understand this interesting field. However, questioning the intelligence of machines that behave like a person, and how machines are showing intelligent behavior, are the type of questions that engineers in general, computer scientists in particular, find most interesting [12].

In the area of computer science, artificial Intelligence can be defined as intelligence demonstrated by machines in distinction to the natural intelligence exhibited by humans. To put it bluntly, AI is often used to describe machines that are developed to imitate cognitive functions that humans relate with the human mind, for example, learning and problem-solving [13].

Due to the fact AI is integrated into different types of technologies; there are numerous examples of AI technology. Here is a list of the most common applications:

- Automation: is the technology by which a process is executed with minimal human intervention. Keep in mind that Automation can be achieved with or without artificial intelligence, but integrating both will undeniably achieve much more. For example, Robotic Process Automation (RPA) is the use of software with AI to perform and execute large repeatable tasks that are normally performed by humans. These tasks could be queries, calculations, maintenance of records and transactions.

- Natural Language Processing (NLP): The processing of human language by a computer program. One of the older and known examples of NLP is spam detection, which looks at the subject and the content of an email and decides if it is spam or not. Recently, NLP approaches are based on machine learning. NLP tasks include text translation, sentiment analysis and speech recognition.
- Self-Driving Cars: are vehicles that are capable of detecting their environment while moving safely with minimal human input, and by using computer vision and machine learning.
- Machine Learning: It is simply the science of getting a computer to act without programming.

4.2.2 Machine Learning:

Machine learning is a subfield of computer science that deals with getting computers to perform certain actions without being explicitly programmed. It is mainly concerned with building algorithms. In order for these algorithms to be useful, they depend on a collection of examples. These examples can come from nature, made by humans or produced by another algorithm.

Machine learning can also be defined as the process of solving a particular problem by following two main steps:

1- Gathering a data set. This part is concerned with collecting data.
2- Building a model that is based on the previously gathered dataset by using an algorithm or a set of algorithms.

Eventually, the resulting model is used to solve the problem that was introduced in the first place.

To elaborate further on the subject. A simple program of adding two numbers would probably work with addition only since that is its hard-coded rule. However, trying to multiply two numbers with the exact same program would not work, due to the fact that it is hard coded to do addition only. Therefore, a separate program needs to be programmed for each logic or operation.

The essence of machine learning part comes when a system can do both operations without being explicitly programmed, where there is no need to write the programs with hard-coded rules, instead, you let the system understands the logic and produce the required results. Figure 4-2 illustrates the difference between traditional programming and machine learning.

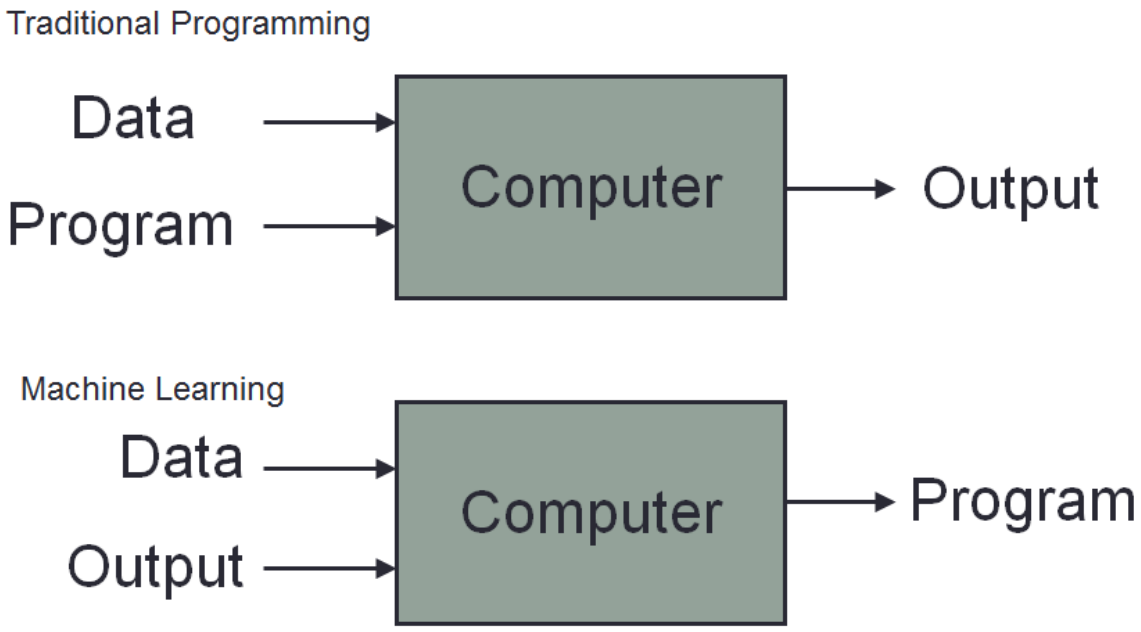

Figure (4-2): Traditional Programming vs Machine Learning

As you can see, we are now giving the **output** as an input to the program instead of the **logic**. For example, in traditional programming, we give the **data** and the **logic** for the addition of two numbers (A = 60, B = 9) then we get the output '69'.

On the other hand, in machine learning, we give the **data** (A = 60, B = 9) and the **output** '69', and then the system understands how the process (60 + 9 = 69) took place. Thenceforth, we ask the program to add another two numbers to test it, for example, we ask the program what is '3' plus '5', however, the answer might be '7.988' or '8.01', which is understandable due to the fact that the system is still in the process of **learning.** It is also worth mentioning that the more data provided, the better predictions and results the system will output.

Training Data, also called **Labelled Data**, is data composed of a collection of training examples, where each instance is either a couple of inputs and required labels (targets) or an input data only.

Machine learning is about learning from examples (data), building a logic and predicting the output for a given input. It learns from past experiences to improve the system of intelligent programs.

4.2.3 Types of Machine Learning:

The types of machine learning (Figure 4-3) are commonly divided into the following three main categories [14]:

- Supervised Learning

- Unsupervised Learning
- Reinforcement Learning

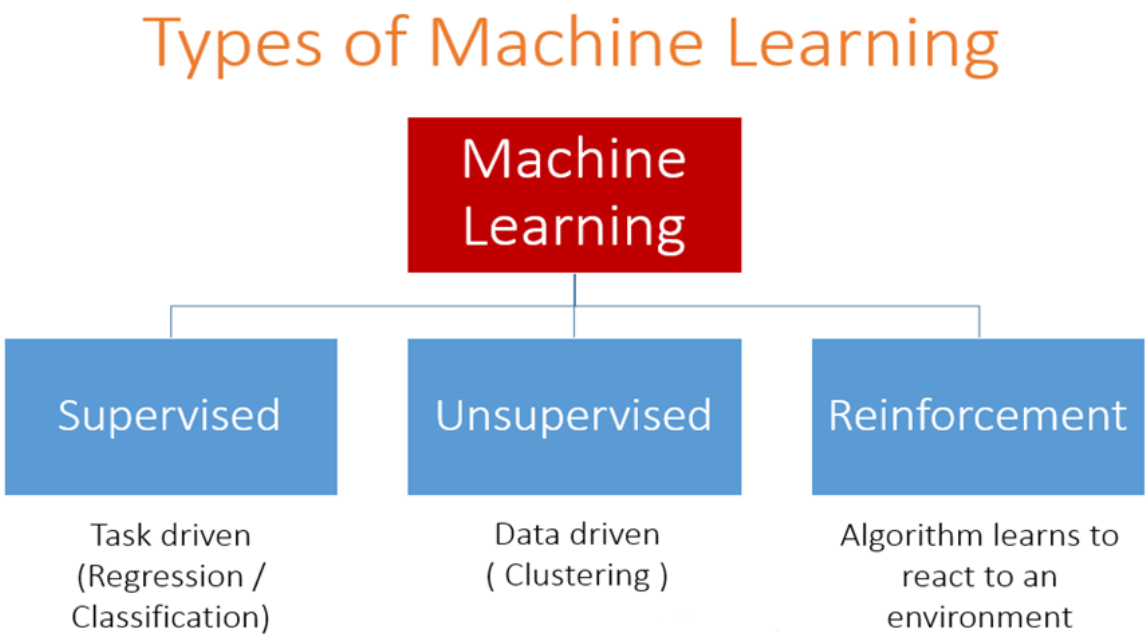

Figure (4-3): Types of Machine Learning

4.2.3.1 Supervised Learning:

Supervised learning is machine learning's most popular paradigm. Due to various data in the form of examples with labels, we can supply these example-label combinations into a learning algorithm one by one, allowing the algorithm to predict the label for each instance and to provide feedback as to whether or not the correct response was anticipated. The algorithm will learn about the exact nature of the relationship between the examples and their labels over time. When fully trained, the supervised learning algorithm will be prepared to follow and predict a good label for a fresh, never-before-seen example.

To further explain the form of training data regarding supervised learning algorithms, we will take the example of adding the two numbers ‘5’ and ‘6’, which results in ‘11’, by giving ‘5’ and ’6’ as inputs and giving ‘11’ as a target or label.

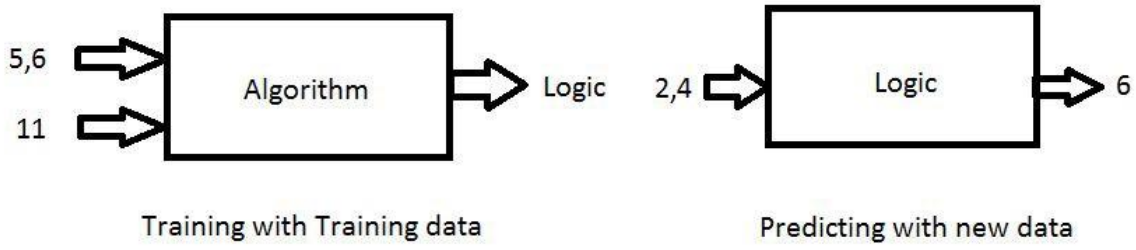

Figure (4-4): An Example of How a Supervised Algorithm is Trained

As shown in figure (4-4), we first trained the model with the training data, thenceforward, with the logic that the algorithm has learned about before, it predicts the output of the new entered data.

This is often why supervised learning is described as task-oriented. It is extremely concentrated on a particular job, which is feeding the algorithm with more and more examples, until it can perform correctly. This is the type of learning you will most probably encounter.

There are two general types of supervised learning (Figure 4-5):

1- Classification.
2- Regression.

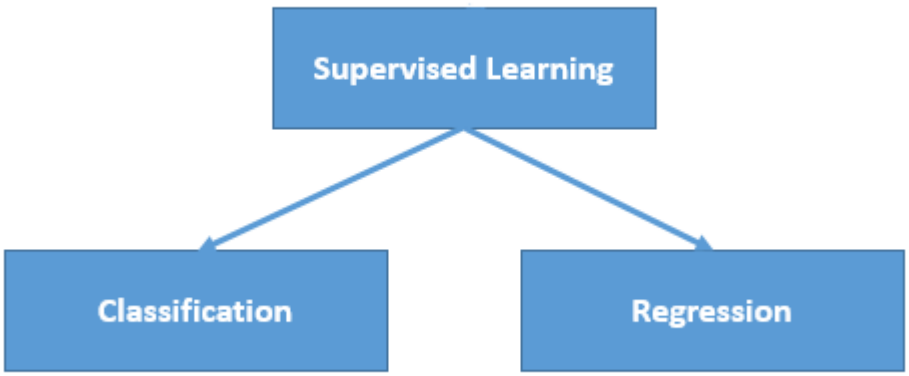

Figure (4-5): Types of Supervised Learning

Classification: this is the type of problem in which the categorized response value is estimated, where data is divided into certain "classes" (i.e. one of the values is predicted in the set of values), for example, {spam, not spam}, {true, false}.

Some of the popular applications in supervised learning classification are the followings:

- **Spam Classification:** if you're using a modern email system, you have probably come across a spam filter, which is a supervised learning system. These systems have been given examples of emails and labels (spam / not spam) and taught to filter malicious emails pre-emptively in order not to harass their users. Moreover, many of these systems also function in way so that a user can give the system new labels and subsequently learn the preference of users.
- **Facial Recognition:** if you are using Facebook, most likely your face was used to train a supervised learning algorithm in order to recognize your face in pictures. It is a supervised process to have a system that photographs, identifies faces and guesses who is in the picture (proposing a tag). It has several layers to it, finding and identifying faces, but is nevertheless still supervised.

Regression: this is a type of problem where the continuous-response value is to be predicted (e.g. we predict a number that can vary from -infinity to + infinity).

Some of the popular applications in unsupervised learning classification are the followings:

- **House Price Prediction:** predicting a house's price given the characteristics of a house, such as size, price etc.
- **Stock Price Prediction.**

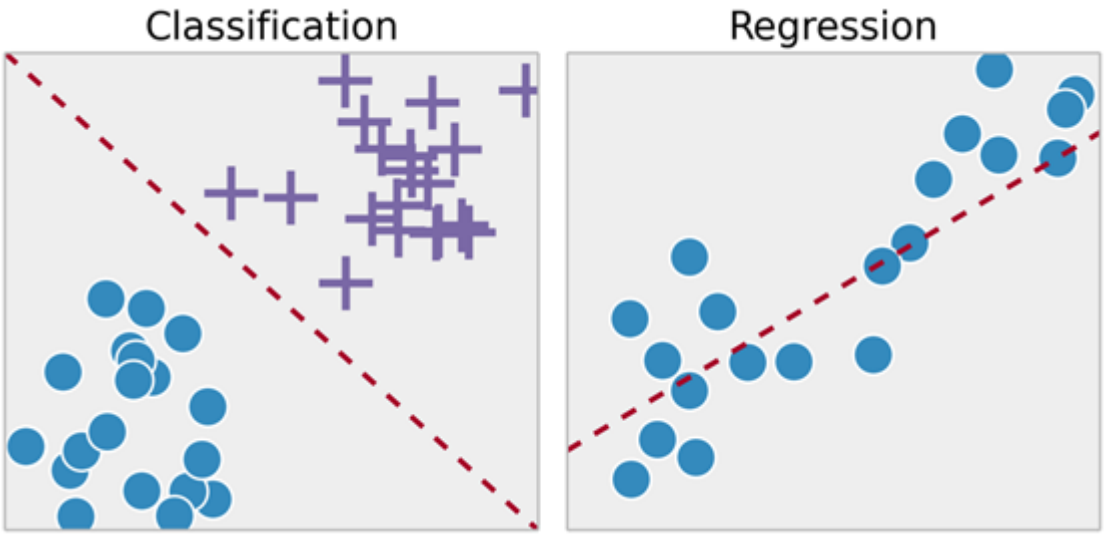

Figure (4-6): Classification and Regression

Classification splits the data into classes. Regression fits the data (both seen in Figure 4-6).

List of Common Algorithms:

- Nearest Neighbor.

- Neural Networks.
- Linear Regression.
- Decision Tress.
- Support Vector Machines (SVM).

4.2.3.2 Unsupervised Learning:

Unsupervised learning (Figure 4-7) is quite the opposite of supervised learning. It does not contain any labels. Instead, a lot of data and techniques would be fed to the algorithm to understand the data properties. From there, it can learn how to group, cluster, and/or organize the data in such a manner that a person (or another smart algorithm) can enter and make sense of the freshly structured data.

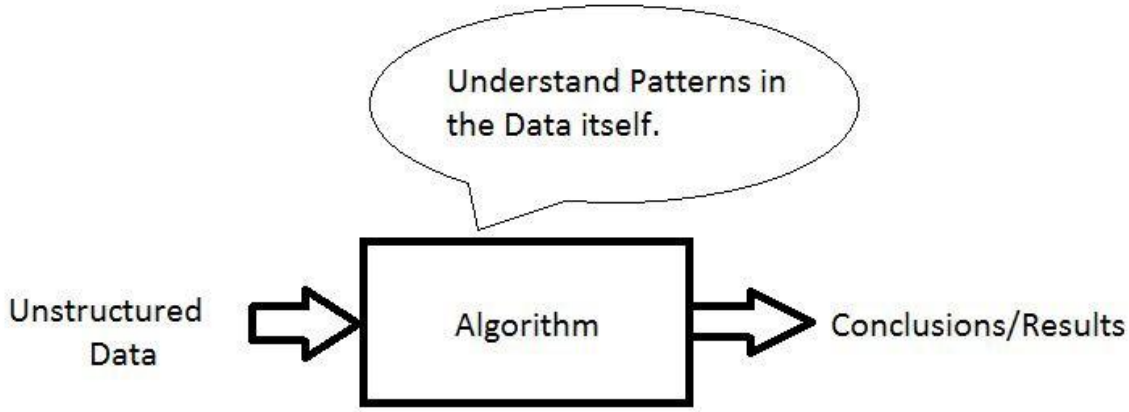

Figure (4-7): An Example of how an Unsupervised Algorithm is Trained

What makes unsupervised learning such an interesting area is that there is an overwhelming majority of unlabeled data in this world. However, having smart algorithms that can handle multiple terabytes of unlabeled data and make sense of it is a tremendous potential source of profit for many sectors. That alone would help to boost productivity in a variety of fields.
There are different types of unsupervised learning such as: Clustering, Association Rule and Anomaly Detection. However, clustering algorithms are the most popular unsupervised learning algorithms.

Clustering: This is a kind of problem where similar things are grouped together. This might be similar to classification, but in clustering (Figure 4-8), labels are not provided, hence, the system learns from the data itself and then groups it into clusters.

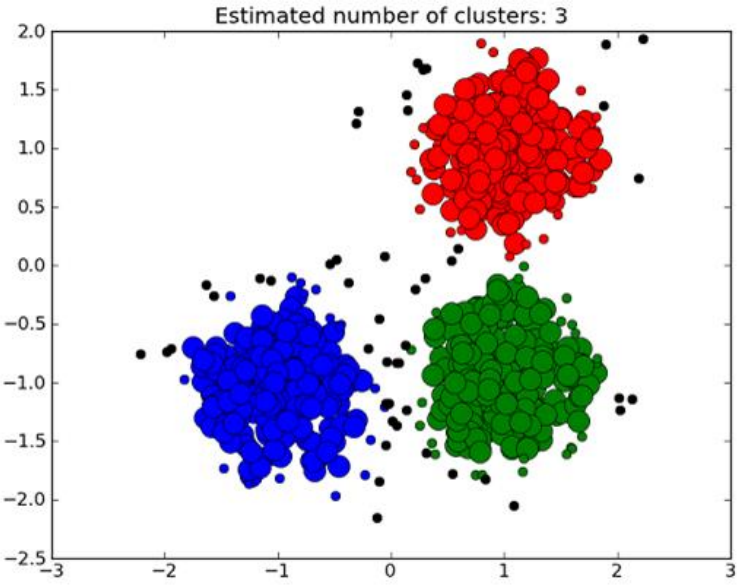

Figure (4-8): An Example of a Clustering Process with 3 Clusters

Due to the fact that unsupervised learning is based on the data and its characteristics, we can conclude that unsupervised learning is data-driven. The results of an unsupervised learning test are controlled by the data and how it is formatted. Here are some areas where unsupervised learning can be seen:

- **Recommender Systems**: you most probably have experienced a video recommendation system, if you have ever used YouTube or Netflix. These systems are often put in the unsupervised domain. Since we know videos properties, perhaps their duration, genre, etc., as we are also aware of many users' watch history. Considering customers who have watched similar videos, and have also liked videos that you still have not seen yet, a recommender system will see this connection in the data, and will offer the suggested videos to you.
- **Purchasing Habits:** your purchasing habits are probably somewhere in a database, and the information is currently being bought and sold at the moment. These purchasing habits can be used to group customers into similar purchasing segments in unsupervised learning algorithms, which enables businesses to target these grouped segments and look like recommendation systems.

List of common algorithms:

- K-means Clustering.
- Association Rules.

4.2.3.3 Reinforcement Learning:

In comparison with supervised and unsupervised learning, reinforcement learning (Figure 4-9) is quite different, where the relationship between supervised and unsupervised can be easily seen (label presence or absence), however, the relationship to reinforcement learning is a bit more confusing.

Reinforcement learning can be simply described as learning from errors. If you place a reinforcement learning algorithm in any environment, it will make many mistakes in the beginning. However, As soon as we provide the algorithm with some kind of signals that link good behavior to positive signals and poor behavior to negative ones, we can reinforce our algorithm to prefer a good behavior over a bad one. After that, our learning algorithm will learn to make fewer errors over time than it used to.

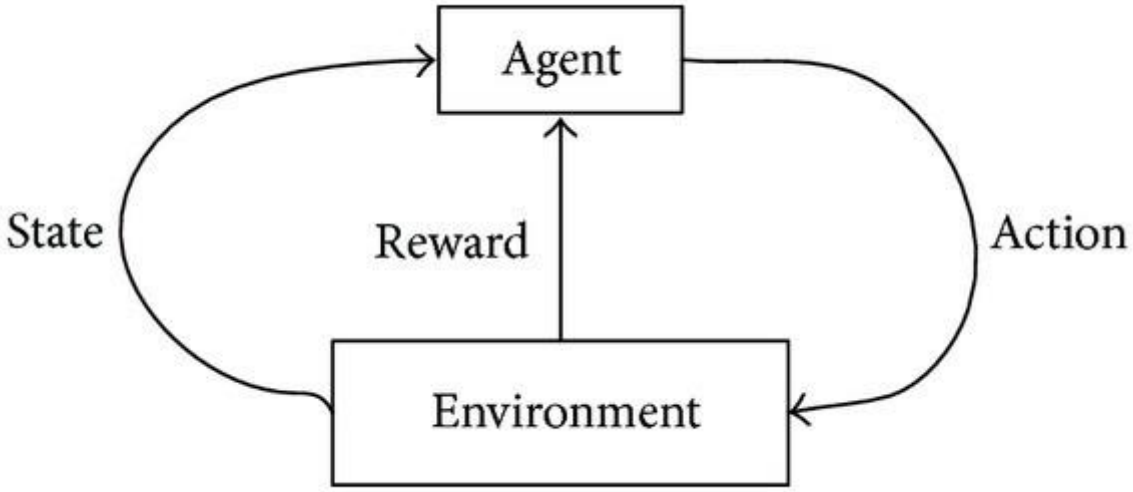

Figure (4-9): How Reinforcement Learning Works

In order to produce smart programs (also known as agents), reinforcement learning goes through the following steps:

1- The agent observes the input status.
2- In order to make the agent conduct an action, the decision-making function is used.
3- The agent receives either a reward or a reinforcement from the environment after the action is performed.
4- The reward information regarding the state-action pair is stored.

The following are some applications that utilize reinforcement learning algorithms:

- **Video Games:** one of the most common scenarios for reinforcement learning algorithms is to learn how to play video games, for instance, Chess and Go.
- **Self-driving cars.**

List of Common Algorithms:

- Q-Learning.
- Deep Adversarial Networks.
- Temporal Difference

4.2.3.4 Deep learning:

Deep learning is a subset of machine learning based on the concept of artificial neural networks. Deep learning architectures, such as deep neural networks and convolutional neural networks have been applied to multiple fields including computer vision, speech recognition and machine translation, where they have been able to produce results comparable to and in some cases superior to human experts.

4.3 The Utilization of Machine Learning in Computer Vision:

As mentioned before, computer vision is arguably considered as a multidisciplinary field of study and can be generally called as a subfield of artificial intelligence and machine learning (Figure 4-10),

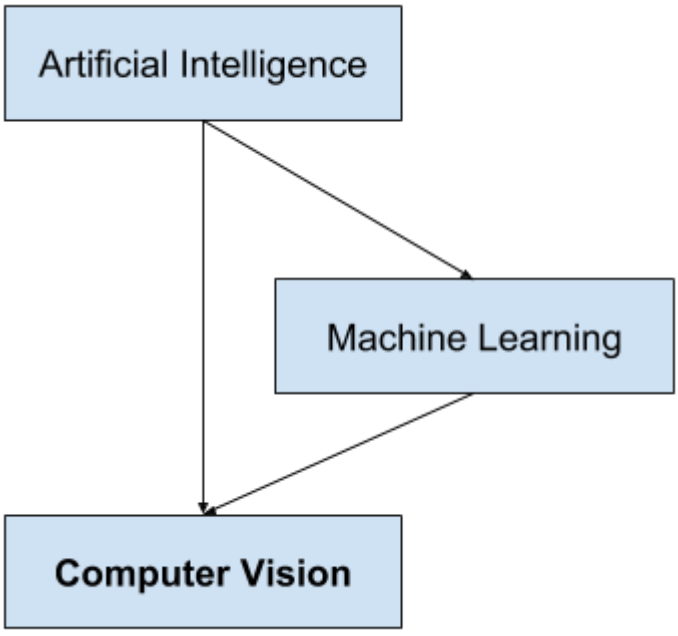

Figure (4-10): The Relationship of Artificial Intelligence, Machine Learning and Computer Vision

Computer Vision is the procedure of utilizing machines to comprehend and analyze both videos and photos. The main goal is to understand the content of the imagery, and to extract useful information from them, which may involve the use of particular algorithms and methods that try to imitate the ability of human vision.

Computer vision has a wide variety of applications, some of the most popular ones are motion detection and object recognition. Each one of these applications has its own characteristics and components. However, every one of these applications must go through an essential step, and that step is known by its common name: digital image processing.

4.3.1 Digital Image Processing:

Digital image processing is a sub-field of digital signal processing that has many significant advantages over the relatively old analog image processing. However, digital signal processing allows a very wide range of algorithms that are applied to the input data in order to avoid built-up noise and signal distortion during processing.

There are many different techniques used by different digital signal processing algorithms, such algorithms may include a combination of different approaches to the processing of images, such as, grey-scaling, image enhancement, color processing and wavelets etc [15].

4.3.2 Motion Detection:

Motion detection may be defined as the detection of a change in an object's position relative to its surroundings, or the change in the surroundings relative to the object's position. Such a task can only be achieved by successfully executing a combination of sequential image processing techniques and concepts. Some of the most important techniques are going to be discussed below [16].

Image Scaling: image scaling refers to the resizing of digital images. Image scaling has many applications and is used differently according to the context. Where in some cases the image needs to be upscaled (increased in size), other cases require the image to be downscaled (reduced in size).

Grey Scaling: image grey scaling may be defined as the representation of an image in shades of grey, regardless of what colors the image is composed of. Each pixel in an image has a luminance value, which can be described as the brightness or intensity of a pixel, where the

luminance of a pixel is measured by a scale from black to white. However, grey scaling an image may be done by measuring the luminance value of each pixel in an image [17].

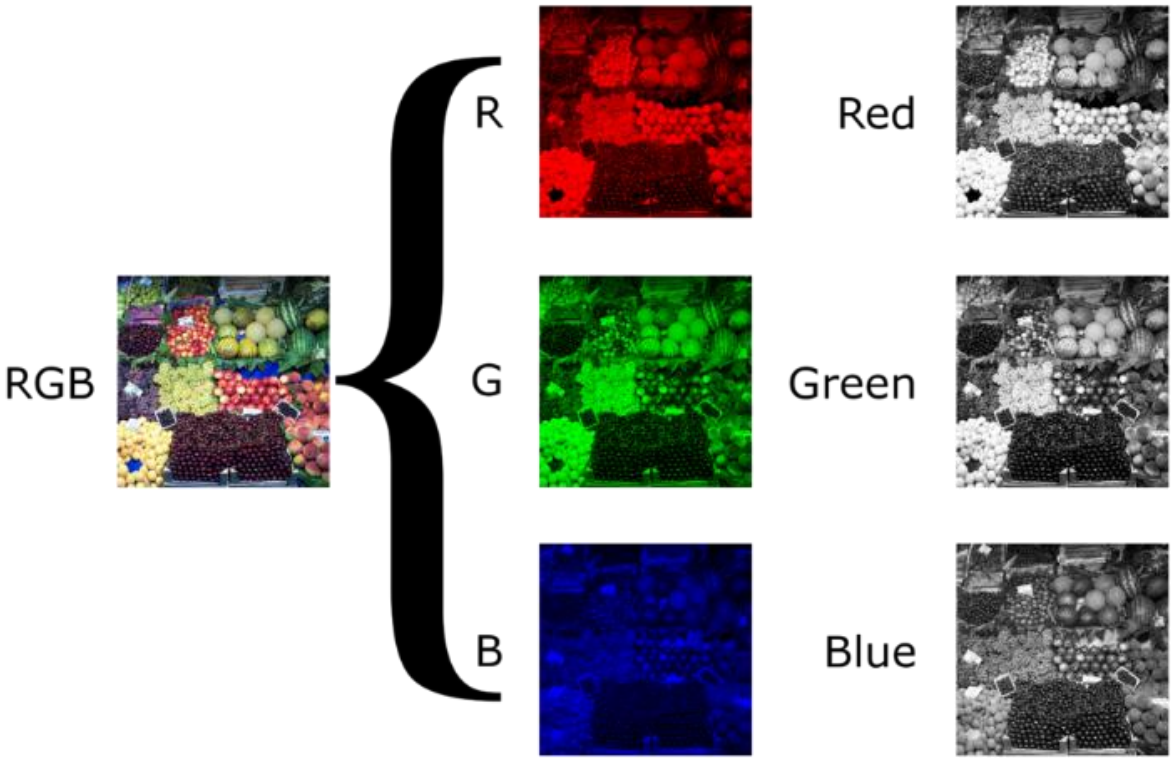

Figure (4-11): Composition of RGB from 3 Grayscale Images

Contours: a contour is a closed curve of points or line segments that represents the boundaries of an object in an image. In other words, contours represent the shapes of objects found in an image [18].

Blobs: a blob (Binary Large Object) usually represents some media format (images, videos, audio etc.), which is composed of a collection of binary data as a single entity. In the context of image processing and computer vision, a blob is a collection of connected pixels, where each pair of connected pixels are neighbors in an image. However, a method known as **blob detection** is used to detect regions in an image that differ in properties [19].

Blurring: blurring images is a technique often used to reduce the level of noise within an image, which prepares the image in a way that makes identifying features much simpler. It is done by creating a weighted combination of each pixel with its neighboring pixels. One example of blurring an image is the **gaussian smoothing**, which will apply a very soft blur on an image.

4.3.3 Object Recognition:

In the previous section, it was pointed out that there are techniques that may be applied to a video footage to detect moving objects through analyzing the changes in pixels and their properties. However, the detection of objects does not indicate what the object is, and here is where object recognition comes in.

Object recognition is the process of detecting objects (whether it is a person, an animal or an inanimate object) in an image of a video frame and identifying what that object is. In other words, it is the process of putting a label on the detected object. Object recognition, much like motion detection, has multiple different algorithms, and selecting the optimal algorithm depends on the context that the object recognition is applied to. Some of the object recognition concepts used in this project are listed below.

4.3.3.1 CNN (Convolutional Neural Network):

CNN (Figure 4-12) is a network that consists of multiple layers, such as the input layer, at least one hidden layer and the output layer. It works in a way that an input is introduced to the CNN through the input layer, afterwards, the input is put through a series of hidden layers, each of which is a convolutional layer that transforms the input using a specific feature/pattern, and sends the input to the next layer. With each layer the input passes it is transformed in a different way [20].

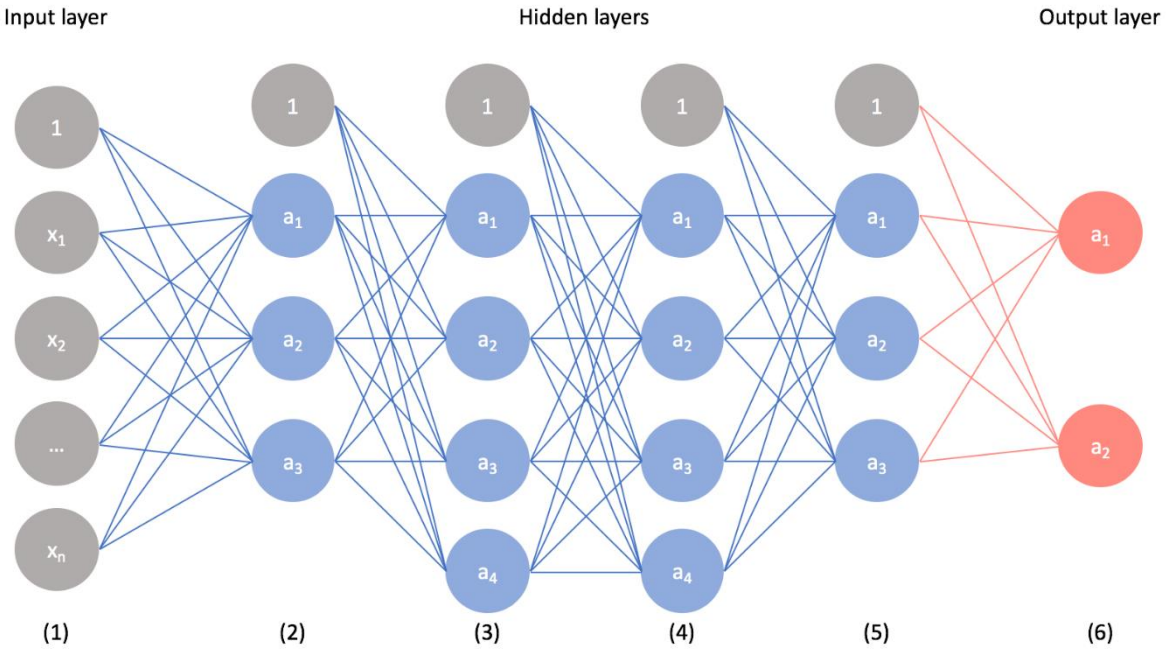

Figure (4-12): an Example of a Convolutional Neural Network

Embeddings: embeddings in neural networks serve a simple task, which is to reduce the dimensionality of categorical variables and meaningfully represent the categories in a transformed space. The embeddings of neural networks have three main purposes:

1- Serve as an input to a supervised machine learning algorithm.
2- Create a visualization of different concepts and relations between categories.
3- Detect and locate the nearest neighbor in an embedding space.

The main usage of embedding in this project is as an input to the object detection algorithm that will be introduced in the next chapter [21].

Region of Interest: the region of interest (often abbreviated as ROI) is the groups of samples identified for a particular reason. It is mostly used to identify and mark groups of samples that have specific characteristics, which match the patterns that a certain detection or recognition algorithm is looking for. For example, when a traffic camera records footage of passing cars, an algorithm analyzes the frames and extracts the ROI, which are in this case the number plates and the face of the driver [22].

Computer Vision Model: computer vision model is a processing block that takes images or videos as input, and extracts certain pre-learned attributes and labels from them. The model is trained to detect and extract certain features, by giving it multiple datasets of examples and their labels. The more datasets the model is given, the higher is the possibility of it making a correct prediction.

Summary:

In this chapter, we explored the field of computer vision and what is the nature of its relationship with the fields of machine learning and artificial intelligence. We also talked about the various types of machine learning in detail. Afterwards, we mentioned popular computer vision applications that utilize machine learning in its working mechanism.
In the end, artificial intelligence is definitely going to shape the future more than any other innovation in the century.

CHAPTER FIVE: DESIGN AND IMPLEMENTATION:

Before we delve into the design and the implementation of this project, it is important to note that the nature of video surveillance systems depends on the environment that the surveillance system is required to operate in. Also, an in-depth site study is needed before reaching the optimal structure, the equipment and the algorithm that suits a certain site. Therefore, the following simulation goes into a certain scenario that has its own specific circumstances, as it goes into how implementing machine learning and computer vision can positively impact the network resources consumed by an IP-based surveillance system.

5.1 Design:

The purpose of this project is to utilize machine learning and computer vision with an IP-based CCTV system in order to reduce or control the CCTV system's overall consumption of its resources. The simulation process will follow a series of steps, each of which implements a technique related to machine learning or to computer vision. In this section each step will be mentioned in the same sequence it was executed in the simulation.

Recording: the first step in the simulation involves recording video footage using a camera, which is the footage that the algorithm will be applied to. Afterwards the footage will be compared to a non-modified version of the footage. The length of the footage is of relative significance, because the comparison will be based on a list of metrics that has little to no relation to the length of the footage.

Processing: a series of image processing techniques will be applied to every frame in the footage in order to prepare the footage for the next steps. This step will include resizing, grey scaling, dilating and transforming the frames into a collection of blobs. The processed frame will be analyzed and scanned in order to extract the contours of the objects, which in turn will contribute in the detection of ROIs.

Detection and Recognition: the processed frames will be passed into two separate algorithms, one of which is for motion detection and the other is for object recognition. The object detection algorithm will detect motion by monitoring if the number of contours that changed position has crossed a predefined threshold.

The object recognition algorithm will grab the generated blobs and will pass them through a pre-trained recognizer model with the purpose of achieving facial recognition. Once the resulted detections pass a certain level of confidence, the detections will be classified with the purpose of labeling each detection.

The Algorithm: after all the frames of the footage has been processed, analyzed and put through the motion detection and object detection. An algorithm will detect the frame that have been identified to have motion or faces, and proceed to lower the resolution of the frames that do not contain any significant activity, by applying a certain level of Gaussian Blur that will simulate the process of reducing resolution. Which in turn will theoretically result in a reduction in the size of the footage, which will make the routing of the footage through the network consume less resources.

5.2 Implementation:

This section includes the execution of all the steps the simulation applied in this project will consist of, and in each step, we will reflect on how the video footage has been affected.

Recording the Footage: the footage used in this simulation will be recorded using an AHD-BL 180 HD security camera as in figure (5-1). Due to the similar qualities this camera has with the average CCTV camera in terms of resolution and framerate, the footage will be recorded and sent through an IP network to a virtual server resembling the workstation that the footage is usually sent to in a standard IP-based CCTV system.

Figure (5-1): AHD-BL 180 HD Security Camera

Processing the Footage: the footage captured from the AHD-BL 180 HD security camera will be put through multiple image processing techniques. Here is a raw frame (Figure 5-1) from the footage that this simulation will be applied to.

Figure (5-2): Raw Frame From the CCTV Footage

Every frame in the footage will be put through two algorithms, one of which is for motion detection and the other is for facial recognition. However, both of the algorithms will be executed using a well renowned library in the field of computer vision known as OpenCV. In terms of motion detection, the first stage is grey-scaling the frames. In order to calculate the luminance levels of every pixel in the frame (Figure 5-3).

Figure (5-3): Gray-scaled Frame

Motion Detection: after the frames are grey-scaled, the process of motion detection can now begin, and the first step is to calculate the absolute difference between each two consecutive frames (Figure 5-4) through a simple subtraction equation.

$$delta = |background\_frame - current\_frame|$$

Where the background frame is the first of the pair.

Figure (5-4): an Example of the Frame Delta (right), The Difference Between The Background Frame and Current Frame

Notice how the majority of the delta frame is black, while areas that contain movement are much lighter, which means that large frame deltas indicate movement. Consequently, after the delta is calculated for the frame, the delta frame is then put through a threshold to only reveal the areas that contain a significant change in pixel intensity (Figure 5-5). However, if the delta is greater than 20, the area will be set to white, and if the delta is less than 20 the area will be set to black.

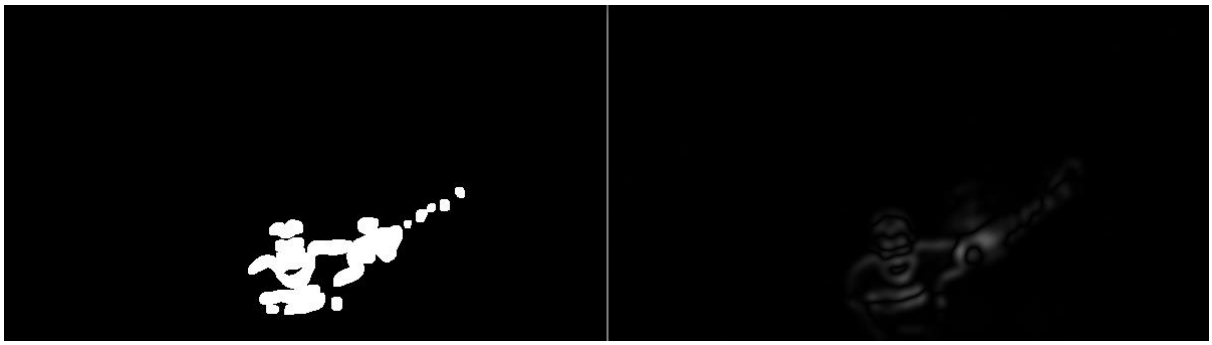

Figure (5-5): The threshold crossed Frame (left)

After the frame has crossed the threshold, it is easier to apply contour detection to find the outlines of the white regions. The frame then is put through a filter to discard any small and irrelevant contours. However, if the contour passes though the filter, a bounding box is drawn around the foreground and motion region (Figure 5-6).

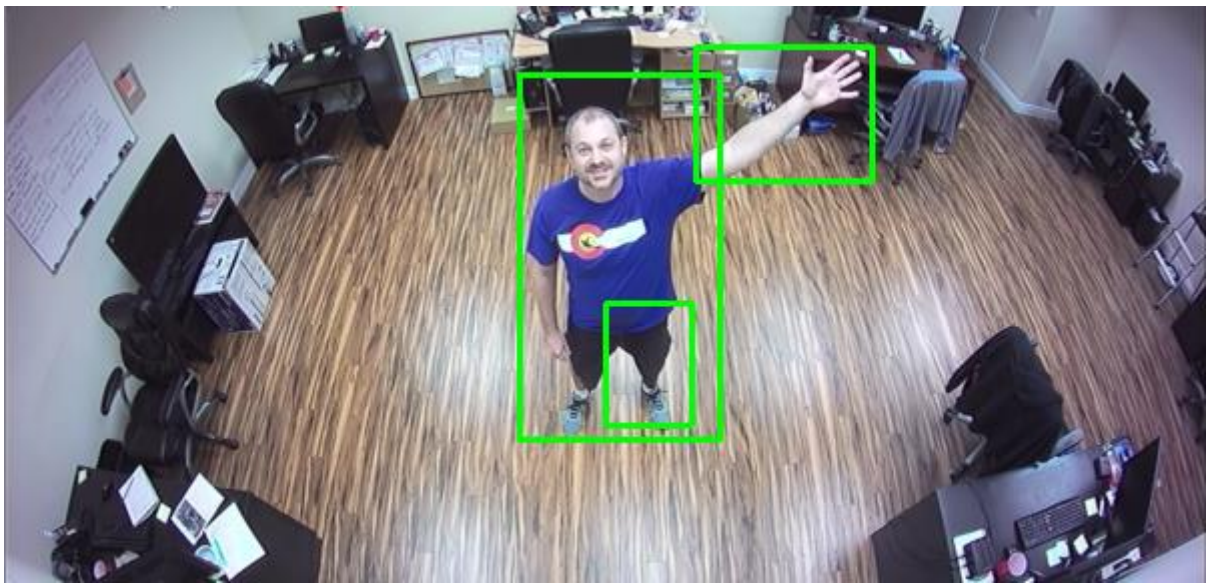

Figure (5-6): Surrounding the Contours with Bounding Blocks

As a result, any movement in the frames will be detected and marked with a green bounding box.

Facial Recognition:

After detecting the movement in the frames of the footage, the facial recognition phase will take place in a series of steps. These steps are going to be explained below.

- **Extracting embeddings from a dataset of faces:** the second algorithm is concerned with the recognition of faces in the video footage. Such a feature is achieved by executing a series of modifications to the frames. The first step is the analysis of a dataset of faces to be recognized in the footage. This is done first by resizing the images from the dataset, where every image will be resized to 600px in width, while keeping the same aspect ratio. Afterwards, a blob is constructed from the resized image, which is then sent to a deep-learning face detector network. The resulted detections contain possibilities and coordinates to localize faces in a given image. However, with the assumption that there is at least one face in the image. The detections with the highest value for the "confidence" indicator will be passed through a confidence threshold to filter out weak detections. Assuming that one or more detections has crossed the threshold, the face ROI is then extracted and converted to another blob. However, this time it is constructed from the face ROI. Subsequently, the face blob is passed through the embedder CNN, which will result in a 128-D vector that describes the face. And then every vector will be labeled with a name taken from the dataset.

  This process of looping through frames will continue throughout the entire list of images in the dataset, where every image is transformed and scanned for any face detections.
- **Training the face recognition model:** now that the 128-D embedding has been extracted, a machine learning model must be trained on top of these embeddings. In this simulation, a supervised learning algorithm known as Linear Support Vector Machine (SVM) is going to be used for that particular task.
- **Recognizing faces in video footage:** after the 128-D vectors are passed to the SVM model, the resulted predictions of who is in the ROI are taken and queried through the label encoder in order to find the most probable name for the face in the ROI.

Afterwards, a string is constructed to display the name and the probability, and rectangle is drawn over the face (Figure 5-7).

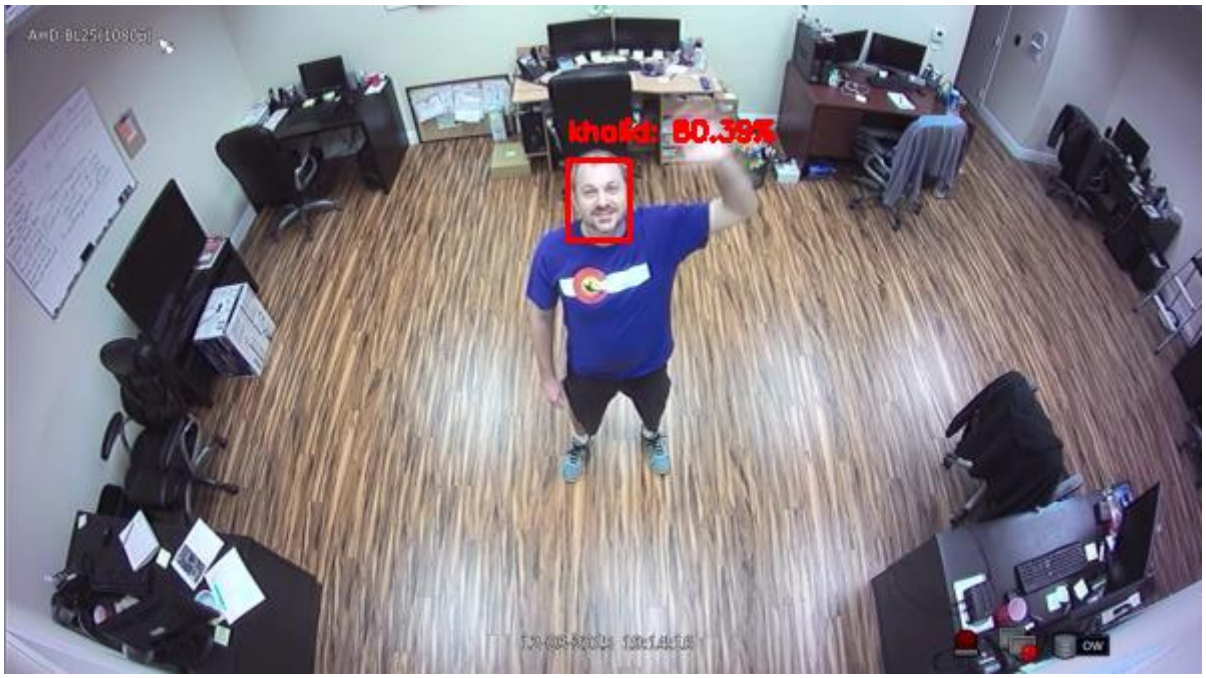

Figure (5-7): Facial Recognition

All the previous algorithms are then combined and executed in sequence in order to produce a system that has both motion detection and facial recognition abilities (Figure 5-8).

Figure (5-8): Motion Detection and Image Recognition Running Simultaneously

Optimizing Network Resource Consumption: all of the previous steps only contribute to motion detection and facial recognition, however, none of them help optimize or improve on the resource consumption that an IP-based CCTV system makes.

And here comes the integral part of the simulation that this project is based on, which is the optimization of network resources by implementing computer vision and machine learning to the captured footage.

The additional step that will achieve this task is the control of footage resolution depending on its content. Therefore, the algorithm will issue a command to the camera to reduce the resolution, however, if the camera does not support resolution control the algorithm will apply a certain level of gaussian blur to the frames of the footage that do not contain any abnormal activity, which in turn will reduce the size of the recorded footage. Accordingly, the reduction in the size of the recorded footage will result in additional saved storage space and less bandwidth consumption.

In the following section, a comparison will be made between the resource consumption of a raw CCTV footage, this footage has been taken from an office space, and it includes two periods, the working hours where there is a lot of acrivity and breakfast time where the office is empty and there is little to no activity whatsoever. This comparison will be measured according to three main metrics, which are storage, performance and bandwidth consumption.

For the storage, the capacity taken by the video is relative to the compression codec it uses more than the video's length. For instance, the video sample used in this simulation has a total size of 240MB when it is an AVI file, whereas it is almost half that size if it was an MP4 file.

And now it is time to put the video footage into our algorithm and observe exactly how using this algorithm contributed positively to the size of the footage.

Table (5-1): Comparing File Sizes

name	Size(before)	Duration(Hours)	effect	Size(after)
Sample.avi	240,701 KB	Approx. 1H	none	240,701 KB
Sample.avi	240,701 KB	Approx. 1H	Gaussian Blur	190,206 KB

As we can see from the table in the previous page, applying the algorithm has impacted greatly on the size of the recorded footage. With nearly 20% decrease in size, lowering the resolution to save storage space is definitely a valid approach to achieve better resource management.

For the performance, this section indicates approximately how much CPU and RAM usage is required to service the algorithm without undermining the rest of the system components. The following table contain the comparison in performance between the two versions of footage.

Table (5-2): Comparing Processing Power Usage

Name	Effect on footage	CPU	RAM usage
Sample.mp4	none	0.57 GHz	Approx. 30 MB
Sample.mp4	Multiple algorithms	1 GHz	Approx. 70 MB

From the previous table, a tradeoff can be noticed. Because while the algorithm significantly reduces the size of the stored footage, it does consume nearly double the CPU power. Therefore, more modifications may be applied to the algorithm to find the perfect balance between the consumed processing power and the preserved storage capacity.

Last but certainly not least, is the network bandwidth consumption, where a comparison will be made between the bandwidth consumption of the regular video stream and the modified one.

The AHD-BL 180 HD captures footage with a resolution of approximately 1280*720, and with a frame rate of 30 FPS. Below is a comparison between the bandwidth consumption of a regular footage and a stream of the same footage modified by the algorithm. Keeping in mind that the following measurements have been taken using a bandwidth monitoring tool called iftop.

Table (5-3): Comparing The Bandwidth

Stream	Resolution	Duration(Hours)	Bandwidth
Regular	1280*720	Apprx. 1H	Approx. 1.9 Mb/s

Modified	1280*720 (approx.)	Approx. 1H	Approx. 1.3 Mb/s

As you may have noticed, another tradeoff took place. Due to the fact that the footage has taken more processing power, it allowed more bandwidth and storage space to be preserved for other operations.

Summary:

In this chapter we started the design and implementation of the simulation this project is based on, and continued with the execution of the algorithm by following a series of steps. And in each step we described in detail what the purpose of it was and how it was executed. Afterwards we ran a comparison between a raw CCTV footage and a one modified by our algorithm, and extracted the differences by taking into account three main metrics which are storage, processing power and bandwidth consumption. And we noticed that the storage capacity taken by the modified footage decreased as well as its bandwidth consumption, however, an increase in processing power has been detected. Therefore, a balance must be found in the tradeoff of resources in order to find the optimal levels of network resource consumption for the given system.

CHAPTER SIX: CONCLUSION AND FUTURE WORK:

In conclusion, the algorithm does do its job in terms of reducing the network resources used by the IP-based CCTV system, however, that optimization does come at the cost of the processing power of the network.

Therefore, more work must be done to ensure that the system operates at the minimum processing power, without undermining the work it does in reducing consumed network bandwidth and used storage capacity.

However, it is important to note that the efficiency of the algorithms is relative to the environment it is required to operate in. for instance, an environment where there is always frequent movement like airports or train stations will reduce the difference in size between the resulted recordings. Unlike an environment where any activity only happens at specific hours of the day such as office security cameras, and where there is a specific time in the day when there is no reason for the camera to be recording at full resolution.

For now, the algorithm only reduces the resolution when there is a detection of motion or a recognition of a face. However, our future plans will be concentrated for developing the algorithm in order to give it the ability to spot and learn the common and repetitive events that occur on a daily basis, so that it will not register any movement or face as an anomaly. But instead it will detect any abnormal movement or new faces and improve the resolution of the camera to its maximum capabilities

Bibliography

[1] A. C. Caputo, Digital Video Surveillance, 2nd ed., Oxford: Elsevier Inc., 2014.

[2] M. MURUNGI, "VIDEO SURVEILLANCE SYSTEM DESIGN," University of Nairobi, NAIROBI, 2009.

[3] "c1c.net," Customer 1st. Communications, 2017. [Online]. Available: https://www.c1c.net/blog/analog-vs-digital-security-cameras-cctv/. [Accessed 2017].

[4] J. Wilson, "sonitrolwesterncanada.com," Sonitrol Western Canada Inc, 18 June 2015. [Online]. Available: https://www.sonitrolwesterncanada.com/blog/what-type-of-cctv-camera-should-i-buy.

[5] E. P. M. T. Kevin Loresch, "Advanced CCTV and what it means to your operation," in *EFC Conference*, Leavenworth, 2011.

[6] "www.raidix.com," RAIDIX, 14 March 2016. [Online]. Available: https://www.raidix.com/blog/2018/03/data-storage-system-for-high-end-cctv-infrastructures-5000-hd-cameras-and-above/.

[7] A. C. Fredrik Nilsson, Intelligent Network Video: Understanding Modern Video Surveillance Systems, New York: Auerbach Publications, 2009.

[8] "netgear.com," 2012. [Online]. Available: https://www.netgear.com/images/pdf/IP-Networking_Impact-On-Video-Surveillance.pdf.

[9] C. Martins, "learncctv.com," Socrates, 2018. [Online]. Available: https://learncctv.com/how-cctv-codecs-work/.

[10] Elvia, "reolink.com," Reolink Innovation Limited, 28 April 2019. [Online]. Available: https://reolink.com/ip-camera-bandwidth-calculation/.

[11] S. J. D. Prince, Computer Vision: Models, Learning, and Inference, 2012.

[12] W. Ertel, Introduction to Artificial Intelligence, Springer International Publishing AG, 2017.

[13] Wikipedia, "Artificial intelligence," [Online]. Available: https://en.wikipedia.org/wiki/Artificial_intelligence#cite_note-FOOTNOTERussellNorvig20092-1.

[14] D. Patel, "Different Types Of Machine Learning," 23 August 2018 . [Online]. Available: https://www.digitalvidya.com/blog/types-of-machine-learning/.

[15] "Digital image processing," [Online]. Available: https://en.wikipedia.org/wiki/Digital_image_processing.

[16] "Motion Detection," [Online]. Available: https://en.wikipedia.org/wiki/Motion_detection.

[17] "grayscale," [Online]. Available: https://techterms.com/definition/grayscale.

[18] "contours," Data Carpentry, [Online]. Available: https://datacarpentry.org/image-processing/09-contours/.

[19] "Binary Large Object," [Online]. Available: ttps://en.wikipedia.org/wiki/Binary_large_object.

[20] P. Bavsar, "Object Detection with Convolutional Neural Networks," 18 January 2019. [Online]. Available: https://medium.com/datadriveninvestor/object-detection-with-convolutional-neural-networks-dde190eb7180.

[21] W. Koehrsen, "Neural Network Embeddings Explained," 2 October 2019. [Online]. Available: https://towardsdatascience.com/neural-network-embeddings-explained-4d028e6f0526.

[22] "Region of interest," 26 May 2019. [Online]. Available: https://en.wikipedia.org/wiki/Region_of_interest.

[23] S. Ludwig, "Pro's and Cons for IP vs. Analog Video Surveillance," *Security Magazine*, 2019.

[24] "wikipedia.org," the Wikimedia Foundation, Inc., 3 May 2019. [Online]. Available: https://en.wikipedia.org/wiki/IP_camera#Surveillance_cameras.

[25] A. Young, "safewise.com," Safewise, 9 May 2019. [Online]. Available: https://www.safewise.com/home-security-faq/wired-vs-wireless/.