A System for Ocular Surface Temperature Measurement Using Infrared Thermography

by

Ehsan Zare Bidaki

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Vision Science and System Design Engineering

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Austin Roorda Professor
Supervisor(s)	Dr. Paul J Murphy Professor
	Dr. Alexander Wong Associate professor
Internal Member	Dr. Lyndon Jones Professor
	Dr. Natalie Hutchings Associate professor
Internal-external Member	Dr. Clark Dickerson Professor

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Background: Ocular surface temperature (OST) is affected by changes in the physiology of the eye caused by normal homeostasis, environmental changes, or systemic and local disease. OST can help a physician to diagnose eye disease with improved accuracy and provide useful information for eye research. OST is assessed non-invasively using a thermal (infrared radiation (IR)) camera. Current methods of OST measurement are restricted in their ability to analyze individual pixel data across the area of measurement due to being unable to localize and track the cornea accurately during a period of measurement. They are also unable to detect eye blinks and are dependent on manual management of the data collection.

Purpose: This thesis presents a novel hardware design, as well as several novel algorithms, for control of the hardware and for image processing of the captured data stream as part of a novel system to measure and track OST from the cornea automatically over any period of time.

Methods/Results: The system uses an IR camera and a visible light camera to capture thermal and visible videos, respectively, from the eye surface. The videos are captured synchronously using designed hardware and an implemented algorithm (data acquisition). The frames for the two video sequences are then registered together (video registration) using two sets of control points. The points are manually selected on the first pair of timestamped thermal and visible frames, and then tracked over the subsequent frames using the Lucas–Kanade optical flow algorithm (point tracking). A mean square error of 5.43±2.01 pixels (equal to 5.43 * 0.09 mm) was reported for salient point tracking for the thermal video and 6.81±2.32 pixels (equal to 6.81 * 0.09 mm) for the visible video. The mean square error for the registration was 5.03 ±1.82, which is approximately 0.45 mm. The corneal area was segmented in the visible images and localized on the images using semantic segmentation method (corneal segmentation). A mean Intersection over Union (IoU) of 94.6% was found, representing the accuracy in identifying corneal pixels in the tracked corneal segmentation, was achieved. Using video registration, the corneal segmentation in the visible image was mapped to the thermal image. OST data extraction from the segmented corneal area in the thermal image was then possible.

Conclusion: A system for measuring and tracking eye surface temperature over time was developed. The system captures thermal and visible image sequences synchronously from the eye

surface of corresponding thermal and visible images taken at the same time. The system is able to localise the cornea on both visible and thermal images. The system is able to report temperature profiles of the cornea over the period of measurement. Experimental results shows that the whole system can work as a tool for measuring and tracking OST over time.

Acknowledgement

I would like to thank you my supervisors Dr. Paul Murphy and Dr. Alexander Wong for their guidance, patience, and moral support during my PhD study. This project and the results would not have been possible without the tremendous contribution of your advices and supports.

Thank you to my committee members Dr. Lyndon Jones and Dr. Natalie Hutchings for their advices during my work. It helps me a lot to find my path.
Thank you to my examining committee, Dr. Roorda and Dr. Dickerson for reviewing my thesis and helpful comments.

I deeply thanks my parents for their unconditional trust, encouragement, and endless support. I am grateful for my wife whose constant love and support keep me motivated and confident. I am forever thankful for her patience, unconditional love and support throughout the entire study process.

Special thanks to my friends Javad (Dr. Shafiee), Mahmoud (Dr. Famouri), and Amir Nazemi for their kind support and helps during my study.

Many thanks to members of the MLEO lab for their support during my study.

Thank you to all of the members of the optometry school for their kind support during my PhD study.

Finally, thank you to all participants and volunteers who took part in my experiments and bringing me an opportunity to finish my work during the pandemic.

Dedication

To my parents, my wife, and our children. I cannot thank you enough for all that you have done for me.

Table of Contents

List of tables

List of figures

Chapter 1: Introduction

Body temperature reflects physiological information about human health. Several methods have been used to assess body temperature. Of these methods, thermal (infrared radiation (IR)) imaging has drawn much attention because of its many advantages, including that it is non-invasive, can be obtained in real-time, and is very accurate. Clinically, IR thermography can be used to observe areas of inflammation in the body, and it has been used to assist clinicians in diagnosing several ocular diseases. A precise IR thermogram can help physicians to diagnose eye diseases with much improved accuracy and opens new avenues in research. IR thermography has been used to monitor temperature changes over the surface of the eye for over 50 years. It has been used to look at tear-film dynamics, temperature changes during contact lens wear, and for disease diagnosis.

Current methods employ either a single camera or dual camera arrangement. A single camera uses an IR camera, but localisation of the cornea as the area of interest is limited due to the lack of thermal distinction across the ocular surface. Dual camera systems attempt to use a visual camera in combination with the thermal camera to assist in locating specific areas of interest on the ocular surface. However, these systems do not synchronise their video streams from each camera and are thus unable to accurately locate and track the corneal area of interest in the thermal image during a measurement sequence. This limits the available options for tracking changes in temperature over time, as well as limiting the possible analysis of OST across the cornea during the measurement period. Previous systems are also unable to detect and remove artefacts in the OST data due to eye movement and eyelid closure. Lastly, they are not fully automatic. Hence, a customised system for imaging the eye and gathering useful OST data across the full eye surface that incorporates a reliable and precise method to detect the corneal boundary on the thermal images, and can track the temperature across all parts of the cornea over time, is needed.

1.1 Thesis objective

To develop:

1- A novel dual camera system design for image/video recording of the eye surface

2- A novel system for measuring and tracking OST over a period of time that incorporates:

a. A novel algorithm for managing the cameras and record synchronous thermal and visible videos from the eye surface.
b. A novel algorithm that is able to register the video output from each camera
c. A novel algorithm that is able to segment the cornea in the thermal camera video image sequence
d. A novel algorithm that is able to remove the artifacts from the video files
e. A novel algorithm that is able to extract the temperature profile within the segmented area from all of the thermal video frames and calculates statistics of the temperature profile

1.2 Thesis overview

The main purpose of this thesis was to develop a system that could image the eye and adnexa across two electromagnetic radiation spectrums: visible light and infrared radiation, using two different cameras, both aimed and focused on the same region of interest; to use the overlapping regions of interest to isolate the corneal region as the primary area of interest in the infrared camera image; to use the infrared radiation data detected by the thermal camera to measure the ocular surface temperature of the cornea over a period of time; and to use removed artefacts produced by eye movement and eyelid blinking during the period of measurement.

The designed system included hardware and software components capable of addressing all of the associated problems. A general system diagram of the development steps for the dual camera system is given in Figure 1-1.

Background information and literature review are given in Chapter 2. The first step of the system development was physical system design and hardware selection. In this step, two cameras, one thermal (IR) and one visible, were selected for image sequence recording. These cameras were installed on a purpose-designed camera mount. Next, the initial steps were made in developing data acquisition and control software for the two cameras. A dual camera management algorithm was developed to access and control both cameras (Chapter 3).

The images produced by each camera were normalized to remove any distortions in the images produced either by each camera's sensor (intrinsic) or due to the relative camera positions (extrinsic) aberrations. A customised algorithm was developed to complete this task (Chapter 4).

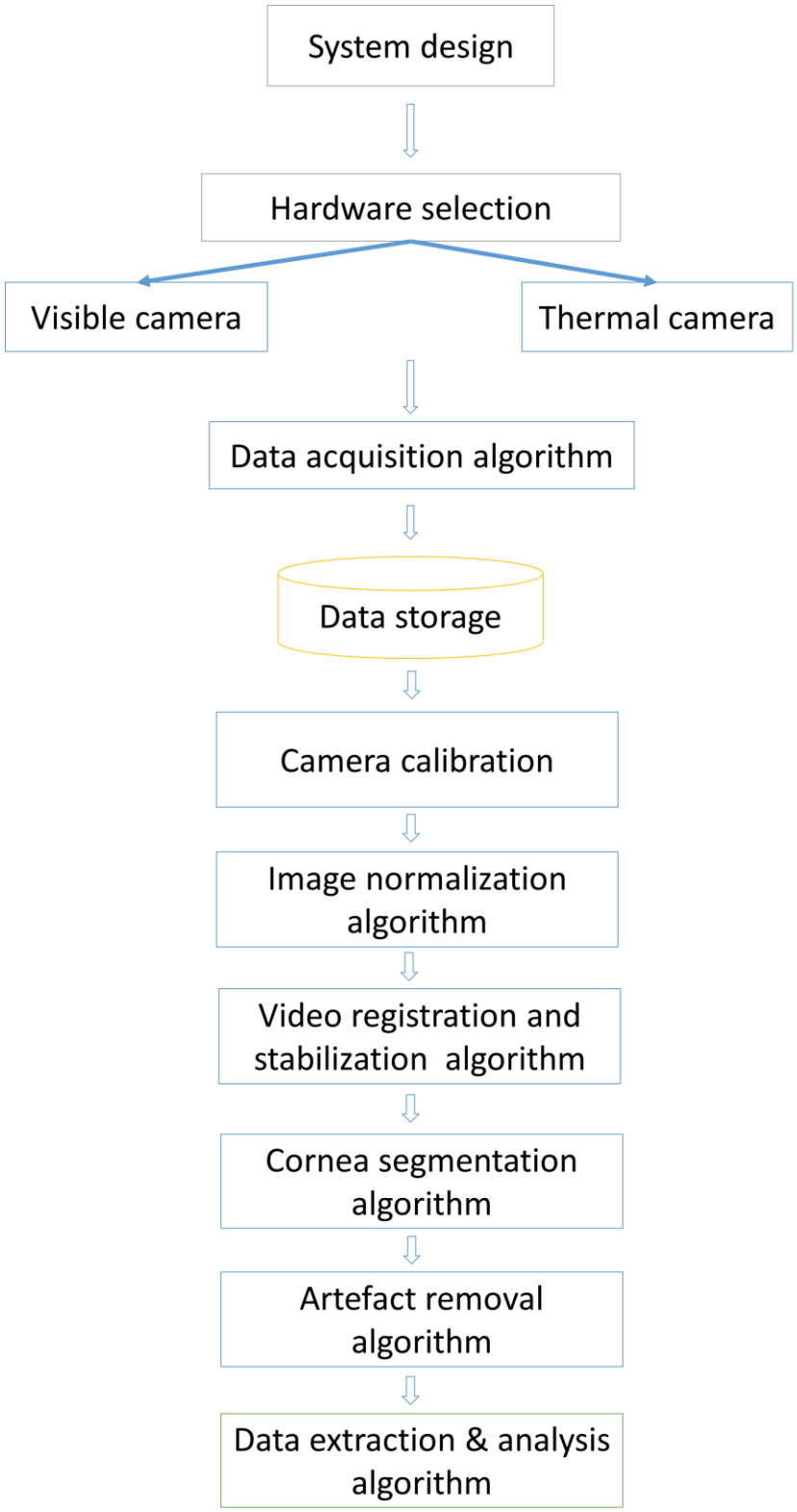


Figure 1-1: The system diagram showing the different steps required for development of the ocular surface temperature measurement system.

Further algorithms were then developed for video registration and salient point selection of the eye and surrounding surface to map the timestamped video files from both cameras (Chapter 5).

The salient points were selected on the first frames and then tracked and localized on all of the subsequent frames using optical flow algorithm for corresponding frame registration and video stabilization. Then, the video files were registered in this step using an image processing algorithm to overlay (map) each frame with its matching pair (Chapter 6).

With the two video files mapped, the cornea was then localized on the visible video sequence. This enabled localization of the cornea on the thermal video sequences using a further corneal segmentation algorithm. By locating the corneal location in the thermal video sequence, the temperature data could then be extracted for the corneal area of interest (Chapter 7).

To remove the eyelid blink artifact from the video sequences, an algorithm was developed to detect the blink frames in the video sequences and remove them from the sequences. To detect the blink, machine learning algorithms were used to extract blink frames from the video files (Chapter 7). A final discussion and conclusion are reported in Chapter 8.

Chapter 2: Overview of image formation and OST measurement methods

This chapter provides background information in the area of thesis research. Section 2.1 provides background information on electromagnetic radiation (EMR) and image formation. Image capturing devices are also described in this section. In Section 2.2, human body temperature, human eye structure, and tear film structure are reviewed. Section 2.3 describes previous and current methods of the OST measurements. Finally, in Section 2.4, the specific issues of eye imaging and OST measurement and tracking are discussed.

2.1 Background

2.1.1 Electromagnetic radiation

Electromagnetic radiation (EMR) refers to the waves or photons radiating from any object in nature and carrying electromagnetic energy. The basic concept of EMR is that a wave is produced when an electrically charged particle is being accelerated. The wave possesses both magnetic and electrical characteristics that cannot be separated – in the same way that time and space cannot be separated. It is thus a fundamental principle of physics: a moving electric field will have an associated magnetic field that changes as the electric field changes; a changing magnetic field will also have an associated electric field that varies as the magnetic field varies.[1] This concept is used practically to produce electricity using a dynamo (generator). EMR is classified by the wavelength or frequency of the wave. Figure 2-1 shows a representation of the electromagnetic spectrum.

EMR can be described as both a wave and as a particle, and experiments using visible light and IR (two types of EMR) have revealed these two basic characteristics. Although simplistic as an explanation, considering how visible light interacts with the light-sensitive proteins in a retinal photoreceptor, can be helpful in understanding how these two EMR characteristics interact. The particle aspect describes the energy carried by the wave which is transferred when the wave/particle interacts with the light-sensitive protein. However, the protein is ‘tuned’ to respond to a limited range of wavelengths, so the wave aspect, as described by the wavelength or frequency, is also important.

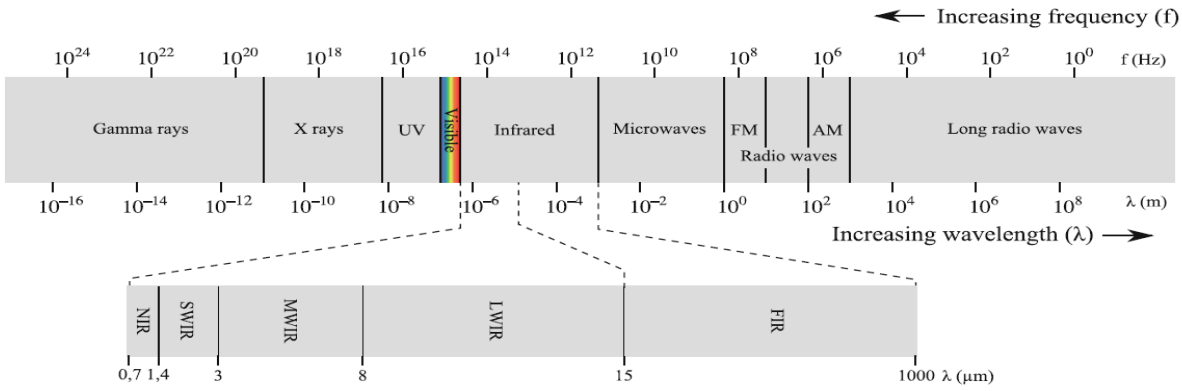


Figure 2-1: The electromagnetic spectrum, showing visible light in color. EMR to the left of visible light on the spectrum has a shorter wavelength, while EMR to the right of visible light has a longer wavelength. The sub-divisions for the infrared spectrum are also shown.[2] *(Reproduced from: Gade R, Moeslund TB. Thermal cameras and applications: A survey. Mach Vision Appl. 2014;25(1):245-262)*

For the light-sensitive protein, specific differences in the molecular structure of the protein make it more, or less, sensitive to EMR of a particular wavelength, depending on how close to the optimal wavelength the incident wave is. A wave closer to the optimal wavelength is more likely to produce the structural change needed to produce a neural signal that registers the detection of a light photon. The photoreceptors use three difference protein structures centered around three wavelength bands of 564–575 nm, 534–545 nm, and 420–440 nm (Figure 2-2).[3,4] The range of response for any specific wavelength of light will vary for each light-sensitive protein, and this enables the retinal photoreceptors to be sensitive to the full spectrum of visible light. This process is called phototransduction.[4]

EMR can interact with sub-atomic particles, atoms and molecules, with differing interactions and effects dependent on the EMR wavelength, the concentration of EMR particles, and the atomic structure and density of the target material. In a vacuum, no interactions occur since there is an absence of particles or molecules. In a gaseous atmosphere, some interactions will occur with the molecules in the gas, to the extent of the gas concentration and the molecular size of the gas molecules. In a liquid or solid, the concentration of available particles, atoms and molecules is much greater, greatly increasing the likelihood of an interaction.

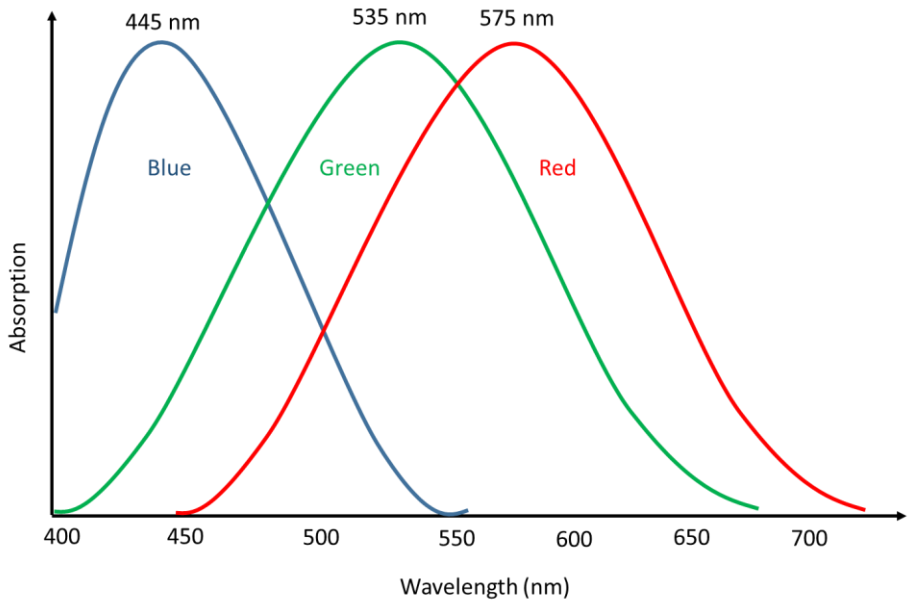


Figure 2-2: Absorption spectra for each type of color sensitive photoreceptor, showing their peak wavelength for maximum sensitivity.

There are three main forms of interaction: reflection, absorption, and transmission. The likelihood of any particular interaction depends on the EMR wavelength and the size of the atoms or molecules in the target material. Reflection occurs when the electrical charge of the atom or molecule of the material interacts with the EMR to alter its trajectory, but not its wavelength or frequency (e.g., visible light reflecting from a mirror). Absorption occurs when the radiation is absorbed by the atomic particles of the atom or of the molecule producing a destruction of the incident wave. The absorbed energy of the EMR can change the structure of the atom or particle, e.g., freeing an electron from the atom (photoablation), breaking the atomic bonds within the atom itself (nuclear fission), or altering the chemical bonds between atoms in a molecule to alter its structure (phototransduction). With this change in atomic or molecular structure, a new form of EMR may also be released and emitted by the target molecule (e.g., fluorescence, infrared radiation), which can act as a marker for the interaction by the EMR on the target material. Transmission occurs when radiation passes through the target material without any interaction. Whether a material is transparent for a particular EMR wavelength also depends on the atomic and molecular structure of the target material. For example, infrared radiation cannot transmit through crown glass, but it can through germanium, and x-rays can pass through most materials, but not a lead screen.

It is possible for all three interactions to occur simultaneously, with the proportion of each form of interaction depending on the EMR wavelength and the atomic structure of the material. For example, when observing a red flower under full spectrum visible light illumination, the longer (red) visible wavelengths are reflected towards the observer, while the shorter wavelengths are absorbed by the atoms/molecules of the flower.

2.1.1.1 Types of electromagnetic radiation

The concept of EMR was first proposed by James Clerk Maxwell in 1865, and since then each type of EMR has been described by many previous researchers. The variation in wavelength and frequency that characterizes each type of radiation and how that interacts with other particles or objects has allowed each type of EMR to be applied in a practical way, e.g., radio waves for communication, x-rays for imaging, microwaves for communication and cooking.

Visible light is a small portion of the spectrum and ranges from 400-700 nm. It is described as visible light since the human eye contains specialized receptor cells that are able to detect these wavelengths and produce a neural signal that is processed by the brain to create a perception of the surrounding world.

In Figure 2-1, the EMR to the right of the visible spectrum has longer wavelengths that are of lower frequencies. These longer wavelengths have lower energy and are less able to cause damage by directly interacting with other particles or materials and are not usually harmful to biological material. They consist of three broad radiation types: infrared (2.5-25 µm), microwaves (25 µm-1 mm), and radio waves (>1 mm).

The EMR to the left of the visible light has shorter wavelengths and higher frequencies. These wavelengths have higher energies that are more able to interact with particles and materials to cause damage, and they can be harmful to biological material. There are three main types of radiation: ultraviolet (400-1 nm), x-rays (1 nm-1 pm), and gamma rays (<1 pm).

2.1.1.2 Wavelength/frequency and energy of a photon

One key aspect of this spectrum is that as the EMR wavelength increases, the wavefront frequency decreases. This relationship is important when considering the energy possessed by a photon of EMR of any particular wavelength. The relationship is described using Planck's equation:

$$E = hf$$

where E is the energy per photon, f is the frequency of the photon, and h is Planck's constant.

Using the equation, it can be shown that a single photon of gamma radiation may possess ~100,000 times more energy than a photon of visible light. The frequency of the wave is thus integral to the energy the wave possesses, and it explains why shorter wavelengths have more energy and are thus more dangerous for biological tissue.

2.1.1.3 Infrared radiation

Infrared radiation (IR) is EMR generated by the movement of particles in matter. The speed of movement reflects the kinetic energy of the particles, thus particles that possess more energy will have greater kinetic movement, and will produce higher amounts of IR. The object can also be said to have a higher temperature. Any object with a temperature above absolute zero (0 K) will therefore emit IR.[5] It also follows that the temperature (kinetic energy of the particles) of the object influences the amount and wavelength of IR produced. An object of temperature between 190-1000 K will emit IR in the mid- or long-wavelength range.

IR describes those wavelengths that are immediately beyond visible red light (infrared comes from the Latin *infra* meaning 'below', and from *red* for the longer wavelength visible light that is perceived as being red in color and which marks the boundary wavelength for IR). IR is therefore not detectable by the photoreceptors of the human eye. IR lies between visible light and microwave radiation with a wavelength range of 0.78-1000 µm (Figure 2.1) and is divided into four sub-divisions: near infrared (NIR) 0.78-1.4 µm, small-wave infrared (SWIR) 1.4-3 µm, mid-wave infrared (MWIR) 3-8 µm, long-wave infrared (LWIR) 8-15 µm, and far infrared (FIR) 15-1000 µm.[6,7] IR in the mid-wave and long-wave range is also referred to as thermal radiation.

Since IR has longer wavelengths within the EMR spectrum, it has lower frequencies, meaning that the energy contained in each photon is not large. When a photon of IR energy interacts with a surface, the energy is insufficient to produce changes in the atomic or molecular structure of the surface, but the energy is still absorbed by the atoms and molecules, which increases the kinetic energy of the object, producing heat. Heat is the term used to describe the increase in temperature (kinetic energy) that is occurring in the object. Depending on the molecular structure of the object, the heat may produce destructive effects on the molecular structure of the object by changing the physical state of the object from a solid to a liquid, or from a liquid to a gas. In biological tissues, heat can be very destructive to the molecular components of the tissues, leading to burning of the tissue. The increase in temperature also causes the emission of IR from the material at a wavelength that is dependent on the new temperature level produced by the incident radiation. An observer that can detect the IR and measure the wavelength would then be able to determine the temperature of the material.

2.1.2 Black Body Physics and Emissivity

As described above, EMR is produced by emission of radiation from a material. The ability of a material to emit EMR is called its emissivity, which is described as the ratio of the energy radiated from the surface of the material compared to the energy radiated from a black body (perfect emitter) at the same temperature and wavelength and under the same viewing conditions. Emissivity is described as a number between 0 and 1, where 0 is total reflection (and thus no absorption and no re-emission, e.g., polished silver surface has an emissivity of 0.02) and 1 represents perfect absorption of EMR radiant on the surface and perfect emission of radiation from the surface. The emissivity depends on the surface and its nature. For example, a clean polished metal surface will reflect most of the radiation and has low emissivity. The polishing has the effect of purifying the molecular characteristics of the surface layer, thereby improving the regularity of the structure at a molecular level, with the effect that visible light is reflected from the surface. The reflection from the surface also means that there is less EMR absorption at the surface, and so less re-emission of EMR from the surface. On the other hand, a rough, oxidised metal surface has high emissivity. The irregular surface encourages EMR absorption at the surface, and subsequently increased re-emission from the surface. A surface covered by lampblack paint will absorb about

97% of the incident radiation and can be considered as a black body. On the other hand, a polished metal surface will only absorb 6% of the radiation and reflect the rest.

A black body is a surface that absorbs all of the radiation falling on it. It is named a black body because the visible light incident on the surface is absorbed and none is reflected, and so the surface appears black. The best example of a black body is a blackened spherical cavity, supported within a box, with a small hole in the wall of the cavity through which radiation can enter the cavity (Figure 2-3). The inside of the cavity radiator is rough and blackened, and this surface absorbs the radiation. For a perfect black body surface, any radiation entering the cavity will be trapped inside – none of the radiation entering the cavity can escape. It will be absorbed, emitted, and reabsorbed by the surface in a continuing cycle within the cavity.

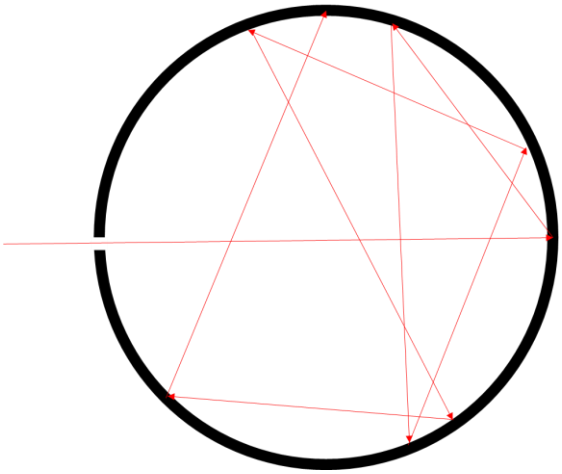


Figure 2-3: A cavity radiator black body. The red line represents the ray-path for EMR that enters the cavity, and which is repeatedly reflected from the surface.

In a thermodynamic equilibrium, the cavity absorbs radiation as much as it emits. For a non-perfect black body surface, some of radiation entering the cavity will not be absorbed and, along with some of the emitted radiation from the surface, will be able to escape form the cavity through the hole. In this way, the emission spectrum of the surface can be obtained using an analysis of the light radiating from the hole.

The EMR from a black body is called blackbody radiation. The intensity of the blackbody radiation depends on the incident EMR wavelength and surface temperature. The Stephan-Boltzmann law describes the relationship between EMR radiation and temperature.[8,9] Based on the law, the temperature of an object can be calculated if the radiation and emissivity are known. The total radiation emitted by an object is proportional to the surface area, emissivity, and the fourth power of its absolute temperature:

$$E = \sigma T^4$$

where E is radiant energy emitted by surface, σ is the Stefan-Boltzman constant, and T is temperature.

Emissivity is very important in non-contact temperature measurement and for heat transfer calculations. Radiation thermometers detect the radiated thermal energy emitted from a surface. The thermometers are generally calibrated using a black body reference surface, with an emissivity close to 1.

2.1.3 Charge-Coupled Devices

A charge-coupled device (CCD) is an integrated circuit of linked (coupled) capacitors (photodiodes). The photodiode (PD) has a dielectric component that can convert an incident photon of radiation into an electrical signal. The size of the electrical signal from the pixel describes the light intensity falling on the PD. Each PD represents one pixel in the CCD array and by combining the CCD with the optical components of a camera, the image of an object can be captured. The sampling resolution of the camera is directly-related to the pixel count on the camera sensor – the higher pixel count, the higher the resolution possible.

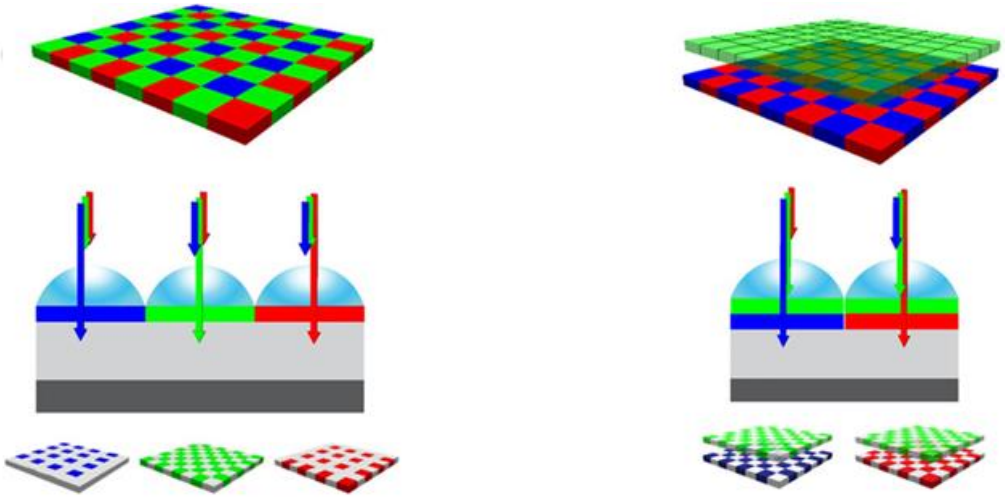
An alternative to the CCD is the complementary metal-oxide-semiconductor or CMOS. Both sensors use a PD to capture light and convert it into an electric charge. The main difference between them comes from how they process these electronic signals. The CCD sensor groups together the electronic signals from the different PDs and then amplifies them all together. The CMOS sensor has one amplifier for each PD. This allows the output from each pixel to be accessed individually and at a much higher speed than the CCD. This also means that CMOS sensors use less power than the CCD. It also means that there is less noise in the signal since each signal is considered individually. However, the need to have an amplifier for each PD means that there are

less PDs per unit area on a CMOS sensor, compared to a CCD sensor, reducing the sensitivity of the CMOS sensor. To overcome this limitation, micro-lenses are placed in front of each PD to gather any light that might miss the sensor, thereby increasing the incident light falling on the PD. Lastly, since CMOS sensors are based on semiconductors used in computer memory, they are cheaper to produce.

As a result of their advantages, CMOS sensors are more commonly used in still photography cameras and consumer level cameras, whereas the better sensitivity in a CCD sensor means that they are used for high-end, broadcast-quality, video cameras. The PDs in the CCD or CMOS can also be designed to be sensitive to different EMR wavelengths, making it possible to have cameras for visible light or IR.

2.1.3.1 Visible Camera

A visible camera sensor is designed to gather visible EMR (400-700nm) and convert it to an electric signal. The CMOS typically used is designed with PDs that are sensitive to visible light in the red, green, and blue wavelengths. This is made possible by two main methods. The first is to place a colored filter over each PD that permits the transmission of only one wavelength, e.g., red, and absorbs the other wavelengths. The electronic signal from each pixel is thus color-coded. The second is to use an array of layered sensors, each of which is sensitive to one wavelength. Every pixel location is now able to detect all three wavelengths. The signal is gathered by the CMOS for accurate color representation (Figure 2-4).



a *b*

Figure 2-4: Schematic diagram of the photodiode arrangement in CMOS image sensors: a) Two-dimensional, b) Three-dimensional [10] (Reproduced from: Lim S, Leem D, Park K, et al. Organic-on-silicon complementary metal–oxide–semiconductor colour image sensors. Scientific reports. 2015;5(1):1-7)

Figure 2-4 a) shows a two-dimensional CMOS sensor where each PD measures the intensity of incident visible light for blue (B), green (G) and red (R) wavelengths after passing through B, G, and R color filters, respectively. In this side-by-side PD array, twice as many photodiodes are assigned to detect G light than for B and R light. Figure 2.4 b) shows a three-dimensional, multi-stacked, organic-on-Si hybrid CMOS sensor. The G array, which is placed on top of the B and R colored filters, detects the amount of G light, after which the B and R light are selectively detected through the B and R colored filters.

2.1.3.2 Detecting IR

Although IR was discovered by William Herschel in 1800, the practical use of non-contact thermography did not begin until the 1950's for military applications and infrastructure inspection. The delay occurred because specialized detector technology was required to capture IR from an object and to measure its intensity. This also delayed the development of the first commercial products, which were introduced in 1983. These initial devices were dependent on liquid nitrogen cooling to cool the detector and improve sensitivity making them impractical for everyday use and very expensive. The development of the CCD removed the need for nitrogen cooling and simplified camera design. In combination with commercialization of the technology, access to imaging cameras has become cheaper and their use more widespread.

There are two types of detectors used to image IR in thermal cameras: photon detectors and thermal detectors. Photon detectors convert the absorbed EMR into an electrical signal in a semiconductor, with the size of the change in energy distribution reflective of the intensity of the EMR. Photon detectors work in the MWIR band, where the temperature contrast is high. The detectors are very sensitive to small temperature changes and have a higher refresh rate compared to thermal detectors. However, they need to be cooled to a temperature below 77K to reduce the effect of thermal noise. Liquid nitrogen was typically used for cooling, but a cryocooler is now used for that purpose, which has high maintenance costs. Therefore, the overall price for a photon detector is higher than a thermal detector.[2]

Thermal detectors also work by absorbing the EMR and in so doing convert the radiation into thermal energy. This increases the temperature of the detector, which alters the resistance of the

material and, in turn, the electrical current that can pass through the detector. Thermal detectors do not require external cooling and work in the LWIR band. There are two different types of uncooled thermal detectors: ferroelectric and microbolometers.[11]

Ferroelectric detectors use the ferroelectric phase transition in certain dielectric materials. A dielectric material is one which functions as an electrical insulator below a specific temperature. Small fluctuations in temperature at or near the phase transition cause large changes in electrical polarization of the material. When the material receives EMR, it absorbs the IR, and the temperature of the material increases. In other words, the kinetic energy of the material at the atomic level increases, allowing electrons to move more easily. In effect, the electrical resistance of the material is changed by the IR and an electrical current can pass through the material, with a flow depending on the amount of incident IR. Barium strontium titanate (BST) is the material commonly used in ferroelectric detectors.

A microbolometer, usually made of vanadium oxide and amorphous silicon, is a specific type of resistor. The electrical resistance of the material is changed by the IR when the radiation is absorbed by the material producing a change in temperature. This alters the electrical resistance, which is then measured. Microbolometers have a higher sensitivity, a smaller pixel size on the detector, and higher spatial resolution than ferroelectric sensors, and have gained a larger market share.[2]

2.1.3.3 IR detector instruments/cameras

The measurement instruments available today can be grouped into three categories: point sensors, line scanners, and focal plane arrays. A point sensor captures radiation emitting from a single point, and if the sensor is scanned across a visual scene, a line scan of the scene can be produced. With a 2-D focal plane array, all of the image elements are captured at the same time. The 2-D array is the dominant method as it has no moving parts, is faster, and has a higher spatial resolution than scanning devices.

After choosing the detector for the thermal camera (2-D focal array), a lens should be selected. The lens serves to protect the sensor, as well as to focus the incident radiation on the sensor. Crown

glass has a very low transmittance for thermal radiation and is not suitable. Instead, germanium, a metalloid, grey-white material, is most often used in thermal camera lenses as it is transparent to IR light and reflects visible light. Germanium is, however, very expensive.[2]

In appearance, a thermal camera is similar in shape and design to those used for visible light. A light-proof box contains the sensor array, with a transparent lens that permits entry of the IR to the sensor. The lens also acts to focus the IR on the sensor. Data transfer from the camera to a computer is by USB, Ethernet, Rs232, or FireWire. The data collection is managed by a customised software, which can display images in a greyscale format on an LCD screen. The images can also be displayed as a color map, with different temperatures corresponding to different colors.

The thermal sensitivity range for most commercial instruments is 20 - 40 mK for uncooled and cooled thermal cameras. The spatial resolution of the cameras varies from 160 x 120 pixels to 1280 x 1024 pixels, and the field of view from 1 to 58 degrees.

There are two main factors that should be considered when choosing a thermal camera: sensitivity and resolution. Sensitivity can be described in two ways: the desired temperature range, and the ability to detect a difference in temperature. Spectral range describes the wavelength range that the camera can detect, and thus the temperature range. Thermal sensitivity describes the ability of the camera sensor to detect a change in temperature. It is the minimum difference in measurement value between two consecutive temperatures, which is equivalent to the thermal resolution.

The sampling resolution of the thermal camera sensor or detector describes the number of pixels in the detector. The higher number of pixels, the higher the resolution. A higher resolution provides a sharper image, a more accurate measurement of the temperature at each individual point in the image, and a more reliable measurement. A higher resolution also provides the ability to measure smaller differences in temperature at a greater distance.

The precision with which the measurement for each pixel is reported is relates to whether 8-bit or 16-bit encoding is used. 8-bit provides a smaller number of data points available to describe the pixel output signal (256) versus 16-bit, which has a much greater number of data points (65536).

A greater number of data points will allow a more refined measurement. However, for most situations, 8-bit encoding is sufficient.

For temperature measurement of the eye, the detector should be able to detect wavelengths between 8-12 microns, which is within the range of IR emitted from the human body. With this range, the camera can obtain a passive image of the eye at room temperature.

2.2 Human body temperature

The temperature of the human body is produced as a result of respiratory activity in the cells and tissues of the body. The cells and tissues of the body also function best within a narrow temperature range. Maintaining body temperature is therefore very important and is part of the autonomic processes of the body. Humans can maintain body temperature within a narrow range even as the ambient environmental temperature varies. This is important since any abnormal change in body temperature is detrimental to health.

A change in body temperature can occur due to systemic or local infection, or malfunction in a body system part. Hence, temperature data of the human body could be a good indicator of health status. Medical thermography has been used to detect different human body conditions, with most of the reported thermography studies being used to detect angiogenesis in tumours and abnormality of blood flow or inflammation. Inflammation is the body's response to an injury or infection and results in a higher blood flow to the site of infection. This increased blood flow raises the local temperature of the area, and the increase in temperature is one of the four signs of inflammation, along with pain, swelling and redness. If an infection becomes more widespread in the body, a general increase in body temperature occurs, which is described as a fever. Inflammation is part of the body's immune system. Both inflammation and fever can be harmful to human body if the immune response is too severe.

Measurement of body temperature is a key part of any medical assessment for a disease that is causing, or can cause, a fever. A medical thermometer is used to take this measurement. Local measurement of temperature at the site of an infection may also be helpful in the diagnosis of

inflammation within a tissue. Thermography has become a useful method in many medical fields, including neurology, oncology, dentistry, and dermatology.[12]

2.2.1 Structure of the human eye

The eye is the most delicate and important sensory organ of the human body. It collects sunlight or reflected light (visible EMR) from the surrounding environment and converts it into a nerve impulse. The impulses are transferred to the visual cortex in the brain, via the optic nerves, where a perception of the surrounding visual scene is formed, and so providing the sense of sight.

The human eyeball is in the shape of a globe and is covered by fatty and fibrous tissues to protect the eye. For additional protection, each eye rests within an orbit in the skull. The bones of the orbit also provide connection points for the six extra-orbital muscles that are attached to the outside of the eyeball. These muscles enable movement of the eyeball. The interior of the globe is formed by the posterior chamber that contains vitreous humor, the lens, and the anterior chamber that contains aqueous humor. Figure 2.5 shows a cross-sectional view of the structure of the eye.

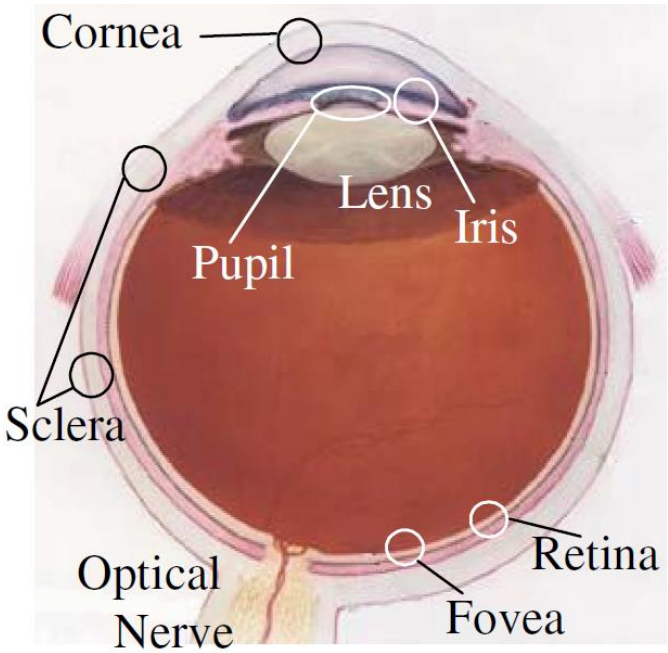


Figure 2-5: Cross-sectional view of the structure of the human eye[13] (Reproduced from: Nishino K, Nayar SK. The world in an eye. 2004;1:I-I)

The enclosing wall of the eyeball consists of three coating layers: the outer, middle, and inner coats. The coats consist of different tissues and have different functions. The outer layer, also known as the fibrous tunic, is made of dense connective tissue that helps the eye to maintain its shape and to protect it from injury. The fibrous tunic parts comprise the cornea and sclera. The sclera covers nearly the entire eye surface and is white in color. The sclera provides attachments for the muscles that control eye movement. The cornea is a transparent tissue located in the front center of the external tunic. The cornea refracts the light passing into the eye and provides most of the eye's focusing power.

The middle layer is named the vascular tunic and includes the choroid, ciliary body, and iris. The choroid is the primary component of this layer and mainly consists of blood vessels. The major function of the layer is oxygen and nutrition supply for the retinal layer. In order to control unwanted scattering of light within the eye, the choroid contains a dark pigment. The anterior part of the choroid is connected to the ciliary body, to which the lens is attached. The ciliary muscles within the ciliary body are able to change the lens shape to adjust the focus of the eye between far and near objects. This process is called accommodation or control of the refractive power of the eye. The ciliary body also produces aqueous humor fluid that circulates in the anterior chamber of the eye and is a source of oxygen and nutrition to the cornea. The fluid fills the anterior and posterior chambers of the eye. The iris is connected to the anterior part of the ciliary body and is positioned in front of the lens. The iris controls the amount of light entering the eye through the pupil. The iris forms a circular, thin structure within the eyeball that controls the size of the pupil. The iris contains pigments to absorb light and so prevent it from entering the eye, so acting as an aperture stop. The density of the pigment determines eye color – less pigment for blue and more pigment for brown.

The inner coat of the eye is the retina, which is responsible for detecting the light entering the eye and converting it into a neural signal. The retina consists of specialized photoreceptor cells and connecting nerves that work together to process the neural signals and to transfer the signal to the visual cortex. There are two types of photoreceptor cells: rods and cones, that function best at different light levels. The rods operate during low light levels and are linked together to improve the overall sensitivity of the eye to light. The rods thus provide a perception of black and white

only. The cones operate at normal light levels, and provide the perception of color, as described in Section 2.1.

The conjunctiva forms the superficial surface of the eye over the anterior surface of the sclera. The conjunctiva is a mucous membrane that is important in the formation and structure of the tear film. It also acts as a physical barrier to prevent a foreign body entering the orbit.

The tears, while not part of the eyeball, form the outermost component of the eye system. The tears help in lubrication of the eyelids during blinking, they assist in forming an in-focus image of an object on the retina by reducing optical aberrations, they provide nutrition to the cornea, they help to protect the eye by trapping dust and foreign objects, and they contain proteins important in protecting the eye from infection. The tears are formed of three main components: lipids, produced by the meibomian glands in the eyelids; aqueous, produced by the lacrimal glands; and mucus primarily produced by goblet cells in the conjunctiva.

Surrounding the eye are the eyelids and eyelashes that have a protective function for the eye. Closing the eyelids presents a physical barrier for the eye, and the eye lashes act to prevent dust from entering the eye.

2.2.2 Tear film

The tear film is a moist, superficial layer between the ocular surface and the ambient environment. Classically described, the tear film is composed of three layers consisting of a superficial lipid layer, an intermediate aqueous layer, and underlying mucus layer.[14] Figure 2.6 shows an illustration of the tear film layers.

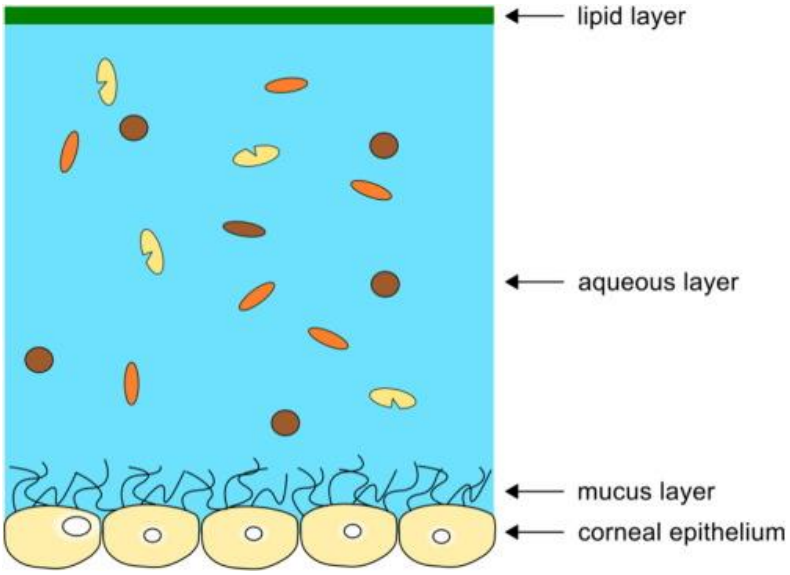


Figure 2-6: An illustration of the tear film showing the tri-laminar structure.[15] *(Reproduced from: Cwiklik L. Tear film lipid layer: A molecular level view. Biochimica Et Biophysica Acta (BBA)-Biomembranes. 2016;1858(10):2421-2430)*

2.2.2.1 Lipid layer

The lipid layer is a thin (around 0.1 µm thick), oily layer that is produced by the meibomian glands in the upper and lower eyelids.[16] The layer contains a mixture of esters, tri-acylglycerols, fatty acids and free sterols that form into two phases: a thick, non-polar, outer layer, and a thin, polar, inner layer.[17] The polar layer spreads over the aqueous layer to counteract the aqueous layer surface tension and provide a surface for spreading of the non-polar layer. This enables the lipid layer to provide a stable layer over the other layers of the tear film. The main function of the lipid layer is to reduce aqueous layer evaporation. Eye blinking provides the opportunity for the lipid layer to distribute over the eye surface. With each blink, new lipid is expressed from the meibomian gland and added to the lipid layer. Between blinks, the lipid layer structure becomes less stable and thins over time, which leads to the phenomenon of tear film break-up. A break-up of the tear film enables more evaporation from the tear film and a consequent increase in tear osmolarity.

2.2.2.2 Aqueous layer

The aqueous layer is the major part (between 6-10 µm thick) of the tear film and is formed of aqueous that is secreted from the main lacrimal gland, the accessory glands of Wolfring, and the accessory glands of Krause.[18] It is primarily composed of water, but also includes numerous electrolytes, proteins, vitamins, peptide growth factors, hormones, anti-microbial factors, immunoglobulins, and cytokines. These components work to protect the ocular surface and

maintain tear film structure by lowering the tension of the ocular surface, allowing the tear film to spread smoothly, giving ocular defense against infection agents, moderating the osmolality of tears, and working as a buffer to maintain pH.[17]

2.2.2.3 Mucus layer

The mucus layer is the lowest layer of the tear film and is adjacent to superficial layers of the cornea. It is mainly secreted by the conjunctival goblet cells, with a small proportion secreted by the stratified epithelial cells of the ocular surface. The mucus layer mainly consists of glycoprotein molecules, which are high molecular weight proteins with a high carbohydrate-to-protein ratio and inorganic salts suspended in water.[18,19] The mucus layer has different functions, including eyelid lubrication that helps the eyelid margins and palpebral conjunctiva to slide smoothly during blinking and ocular movements, and protection of the cornea from abrasion and foreign bodies.

2.2.2.4 Tear break-up time (TBUT)

The tear film loses its stability gradually after a blink. This happens because of insufficient quantity or quality in one or more components in the ordered structure.[20] Normally, the tear film breaks up within 15-40 seconds after a blink, and dry spots begin to appear over the cornea.[21] The time between a blink and the first appearance of the dry spots on the tear film is named the tear break-up time (TBUT). The current standard method is to apply a moistened sodium fluorescein strip to the tear film and allow the fluorescein to spread in the tears after two or three blinks. The subject is asked to blink naturally and to then hold their eye open without blinking. The tear film is observed using a slit-lamp microscope under cobalt blue light illumination, and any fluorescein pattern break-up observed through a Wratten 12 yellow filter. Any break-up in the tear film is observed as a dark area and the TBUT is recorded as the time elapsed since the last blink. A TBUT of less than 10 seconds suggests an abnormal tear film.[19] TBUT is one of the most commonly-used tests of tear film stability, and a reduced tear break-up time or limited ocular surface wetting is one of the main signs of tear film abnormality.[19]

2.2.2.5 Tear evaporation

Evaporation occurs when molecules in a liquid have sufficient energy to change from a liquid state to a gas state, and so escape from the liquid into the ambient environment immediately superficial to the liquid surface. In this way, heat is transferred from the liquid into the atmosphere, and the

liquid cools slightly. The process of sweating takes advantage of this phenomenon to reduce surface body heat.

In the eye, evaporation can occur from the aqueous component of the tear film. The primary role of the lipid layer is to form a superficial layer over the tear film that is resistant to evaporation.[22,23] During tear break-up, the organized structure of the lipid layer breaks down and increased evaporation from the tear film becomes possible. The ocular surface cools as a result of the tear film evaporation.[24] Observation of the cooling can be made using thermographic methods.

The process of evaporation and evaporation rate is very important in ocular surface disease diagnosis. The tear evaporation rate has been shown to increase in pathological dry eye disease (DED). This increased evaporation rate may be responsible, at least in part, for the increased rate of cooling of the tear film observed in DED.[25]

2.2.2.6 Tear production and drainage

Tear production forms part of a sensory loop that includes the ocular surface sensory nerves and the main lacrimal glands.[26] The relationship is most easily demonstrated in the presence of a foreign body on the ocular surface that stimulates the ocular surface sensory nerves. The neural response is processed centrally, leading to activation of the lacrimal glands to release a large volume of tears that flushes the foreign body from the eye surface. The sensory loop is also thought to be part of normal aqueous tear production stimulating a basal production rate.[27] The stimulus for production is thought to be the cooling of the tear film caused by evaporation. The evaporation causes a reduction in tear volume, and basal production is necessary to replenish the tears.[27]

Tear volume loss also occurs through normal tear drainage. Tears drain from the superior and inferior tear menisci, along the upper and lower eyelid margins, through the lacrimal puncta that are situated at the nasal side of each eyelid. The tears pass along the canaliculi or tear canals into the lacrimal sac and are absorbed through the nasal mucosa (Figure 2-7).[28,29]

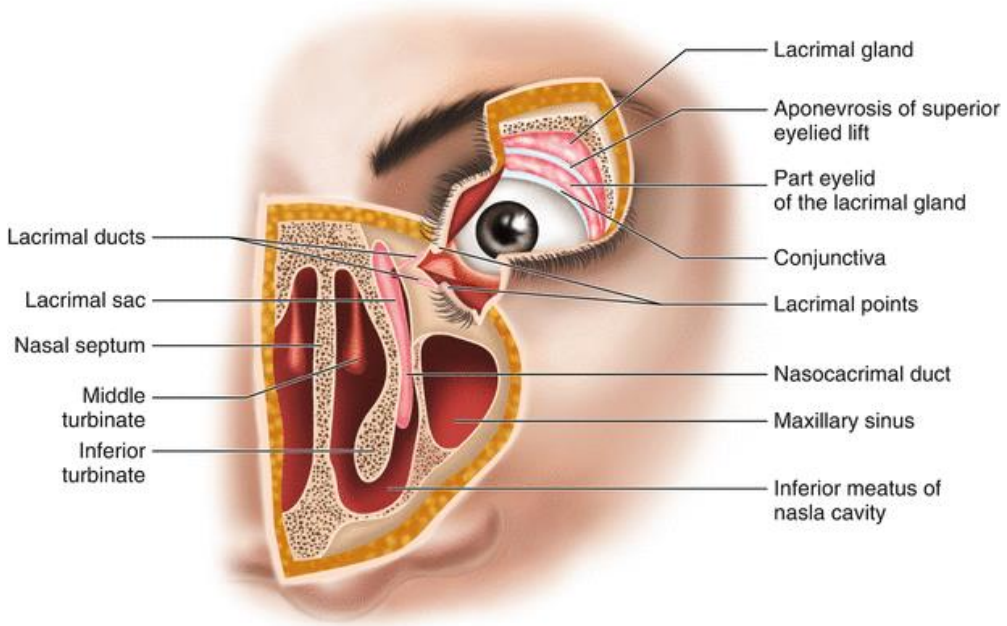


Figure 2-7: Anatomy of tear production and drainage [30](Reproduced from:Marchioni D, Bettini M, Soloperto D. Anatomy of the lacrimal drainage system. In: Endoscopic surgery of the lacrimal drainage system. Springer; 2016:1-12)

2.2.3 Ocular surface temperature (OST)

Various methods are described in the literature for measuring ocular surface temperature (OST) either invasively or non-invasively. Since the eye is a very delicate organ, a non-invasive method for temperature measurement is preferable. Hence, thermography is a key solution to this problem. Data on OST changes are useful in the diagnosis of some eye diseases and for analysing tear-film stability. For example, OST measurement has been used in the diagnosis of detecting vascular, neoplastic and inflammatory pathologies[31], lacrimal drainage system inflammation[32], DED[33], glaucoma disease[34], unilateral proptosis[35], and ophthalmic post-herpetic neuralgia[36]. It has also been used to diagnose retinoblastoma in children[37] and vascular neuritis[38] of the optic nerve. Hence, temperature measurement can be used in the diagnosis of a wide variety of ocular diseases. In addition, temperature measurement is useful in the study of tear-film stability[39].

Tear film instability will affect OST[12,40] by increasing the level of evaporation from the surface, and is one of the key factors in DED[41]. Typically in DED patients, OST changes more rapidly, the size of the change is greater,[33,42,43] and there is greater variation across the ocular surface.[44,45] These effects can be observed using IR thermography[40,46,47]. Under normal conditions, the tear-film layer undergoes a repeated cycle of formation, destabilisation, break-up, and reformation. Since the tear

film is inherently unstable, variations in evaporation across the surface is a natural phenomenon, which may have a role in triggering a blink or in detecting changes in local ambient environmental conditions. Reformation occurs by the action of the eyelids during a blink cycle. TBUT is monitored clinically to provide a measure of the quality of the tear-film[48]. Using thermal imaging, the tear-film changes could be tracked over time over the corneal area. However, current methods for reporting the variation in OST across the cornea and within the tear film break-up area are very limited and frequently rely on manual assignment of OST recording at single locations.

Previous studies have measured OST across the cornea either at single pixel locations, at multiple single locations, as the mean within a selected area, or as the mean across the whole cornea. For example, Efron et al. showed that the corneal temperature ranges between 32.8 to 35.4°C. However, there is considerable temperature variation across the eye surface. For instance, the temperature of the corneal center is different to that of the limbal edge in a healthy eye (due to the presence of the limbal blood supply). Also, during tear break-up, there is a rapid temperature decrease in the break-up area due to evaporation. Ideally, OST should be analyzed locally and globally to be useful in clinical application.

2.2.4 Emissivity of human body and eye

As noted, the strength of EMR emission from a body is described as the emissivity, with an emissivity of 1 for a perfect black-body emitter. Neither the human body nor the eye can act as perfect black body, but in certain EMR wavelengths, they may perform very close to a black body. By assuming the human body temperature to be 37°C, the emitted wavelength of 9.35 µm falls within the long-wave IR region. Previous studies have shown that the absorption band of the ocular surface is similar to water (emissivity of 1) and behaves like a black body radiator in wavelengths above 3 µm. The emitted spectrum of the ocular tissues is between 1 to 30 µm under normal circumstances (32°C). However, this is limited to a maximum distance of 9µm. As a result, the spectrum radiated by the ocular tissues are fully absorbed by the ocular tissues anterior to it. For example, the spectrum emitted by the vitreous is absorbed by the lens.

This relationship also occurs with the cornea and tear film, but the effect varies depending on the thickness of the tear film. When the tear film thickness is about 10µm, the IR transmittance is

about 30%, and when the thickness is above 20µm, the transmittance is zero. Since tear film thickness ranges from 2-5.5 µm, the tear film will absorb 20-60% of the IR from deeper tissues.[49]

For the eye, Mapstone[50] determined that the ocular surface closely approximates a perfect black body with an emissivity of 0.97. This occurs since water behaves in the same way as a black body for IR radiation of wavelengths above 3µm.[7] Thus, the emitted IR from the eye surface can be measured to determine OST. Reported ocular surface temperatures range between 32.8°C to 35.4°C. Efron et al.[51]

2.3 Overview of OST measurement methods

2.3.1 Earlier methods of OST measurements

Earlier methods for corneal temperature measurement can be categorized into two main categories: invasive and non-invasive methods. Several previously reported methods for OST measurement were invasive in nature.[44] The most commonly-reported methods used a thermistor in a probe or sandwiched between two contact lens layers. These methods lacked consistency in reporting normal corneal temperature. This was due to the type of sensor used, the placement of the probe on the eye, the location of measurement, the need for anesthesia, a single point of measurement, the influence of the probe temperature, and the ambient environment temperature effect.

Mapstone was the first person to use a non-contact method to measure OST.[50] A bolometer (thermal detector) was used to measure the radiation from the surface of the eye. In the 1970s, the emergence of new cameras with better sensitivity and magnification provided an improved facility to capture temperature data from the ocular surface. These early cameras incorporated some form of detector and required liquid nitrogen for cooling of the detector to enable detection of IR radiation. With the subsequent development of semi-conductor CCDs, the detectors have been revolutionized in their level of sensitivity, speed of operation, and they can work at room temperature.

2.3.2 Current methods of OST measurement

For IR measurement of the eye, studies suggest that radiated heat detected by a thermal camera derives principally from the tear film. The tear film is both absorbing and emitting radiation – it absorbs IR from deeper components of the eye, such as the cornea, and re-emits this absorbed

radiation, which is then detected. Therefore, the thickness of the tear film layer is a key factor in how much radiation is absorbed from deeper ocular layers. If the tear film was absent (due to thinning to irrelevant levels), the temperature detected would be that of the cornea itself. When the tear film thickness is between 4-10 µm, the tear film absorbs between 55-80% of IR radiation from deeper ocular layers, respectively. If the tear film thickness is more than 20µm, it will absorb 100% of the radiated infrared from deeper layers, and the emitted radiation detected will arise solely from the tear film.[52,53] It is therefore important to note that the majority of IR detected by the thermal camera is emitted from the tear film, albeit with some absorption and re-emission of radiation from deeper layers. This concentration of information from the tear film, however, makes thermography a very useful tool for observing changes in tear film stability.

With the invention of thermal cameras, thermography has become a valuable method to measure OST. Thermography has two advantages over the previous methods. First, it is non-invasive. Second, it reflects temperature variations over the total eye surface.[2] IR thermography provides an opportunity to measure the temperature of the whole cornea, or a specific area, such as a circle around the center, or limbus. Many studies have been undertaken on the use of IR thermography and OST. Published studies reporting on the measurement of OST describe varying methods for analysis of the captured IR thermography image. The methods for estimation of the OST can be grouped into two categories: single camera and dual camera system.

2.3.2.1 Single camera

In these methods, a single thermal camera was used to measure the OST. Generally, these methods are divided into three categories: manual, semi-automatic, and automatic methods.

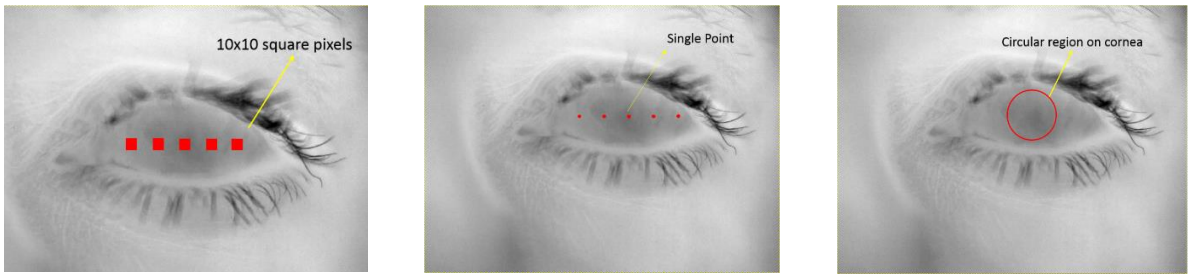
2.3.2.1.1 Manual methods

Originating with the earliest thermal cameras, the speed of response was slow and only a few frames could be captured in series, perhaps with a frame rate of 4Hz. Each image produced was analysed by the user manually selecting a single point or multiple points in the area of interest. In most cases, OST was then studied by comparison of several points or areas on the corneal surface. For example, Efron et al.[51] (1989) measured OST at 11 points on the surface of the cornea to observe changes in OST after a blink. The points were placed along a horizontal line that crossed the geometric center of the cornea, and the points were separated by 0.5mm.

In another study, Morgan et al.[54] (1995) used five 10x10 pixel boxes (Figure 2-8a), each of approximately 1mm^2 area, located on the horizontal line across the center of the cornea. The mean OST of each area was used for statistical analysis. They found that DED patients have a higher OST in comparison to a normal eye group with the same age and conditions. Also, that tear film evaporation in a dry eye was approximately three times faster than from a normal eye.

Galassi et al.[55] (2007) used five different points in their study instead of five boxes (Figure 2-8b). The points were positioned along the horizontal line of the cornea, one point at the center, one on the internal and external canthi, and one half-way between the internal canthus and nasal limbus, and one half-way between the temporal limbus and external canthus. They applied the method to subjects with primary open-angle glaucoma (POAG) and found that OST was significantly lower in POAG patients compared to healthy eyes.

A similar study by Sodi et al.[56] (2007) assigned five different points on the horizontal line of the cornea. The points were placed equidistant from each other, with one placed on the estimated center of the cornea. They used IR thermography to measure OST in central retinal vein occlusion (CRVO) patients compared with a healthy control group.



a *b* *c*

Figure 2-8: Manual selection, a. Drawing multiple squares (10x10 pixels), Morgan et al., b. Single-point measurements (5 pixels), Galassi et al., c. Drawing a circle over the cornea region, Chiang et al

Murphy et al.[57] (1999) used five boxes placed along a horizontal line centered on the cornea for OST measurement. They found that during non-contact cornea aesthesiometer (NCCA) air pulse usage, the rate of OST change was the principal mode of corneal nerve stimulation. Mori et al.[43] (1997) used OST measurement to evaluate the tear film layer by using a 20x20 pixel box (3.3mm^2)

located at the center of cornea. They found that the rate of decrease in OST post-blink for DED patients was considerably less than for normal subjects. Chiang et al.[58,59] (2006) placed an encircled area (4.4mm diameter) on the center of the cornea (Figure 2.8c). They used the extracted data to diagnose DED. Ng et al.[59] (2008) also placed a circular area on the center of the cornea. They used the data to find a relationship between age and OST. In their experiment, OST was found to decrease with age at a rate of 0.0383°C per year. Purslow et al.[12] (2005) recorded temperature at 23 points across the cornea, grouped in five regions: central, superior, inferior, nasal, and temporal. They used thermography to measure OST change that occurred while a person wore different types of contact lenses.

Matteoli et al.[60] (2016) used thermography to investigate OST change in the eyes affected by age-related macular degeneration (AMD). They attempted to control for tear evaporation by only using the first frame of the opened eye for analysis to set the baseline. The images were acquired in RGB digital format and converted to greyscale for implementation of an image processing algorithm. Then, using a MATLAB software program, they selected five areas (small, oval shape) on the surface of the eye in which the OST was analysed in selected frames of a recorded movie file. They found that the OST of the AMD patients was significantly lower in comparison to a control group.

Tan et al.[61] (2016) used thermography to investigate the repeatability of thermal measurement in DED patients. OST was recorded and the region of interest (ROI) was selected manually using the diamond method, and the acquired data was recorded at 0, 5, and 10 secs. Inter-image, inter-occasion and inter-examiner repeatability of the OST was evaluated by calculating coefficients of repeatability (COR). The study showed that most of the coefficients used in their experiment were highly repeatable.

Sudarshan et al.[62] (2017) also worked on DED using thermal imaging and extracted the OST ROI from the original image manually. They attempted further image processing by using discrete wavelet transforms and Gabor transforms for feature extraction. By analysing the features, they found that the OST changes in the lower part of the cornea were more accurate, using their algorithm, for diagnosis of DED.

Ding et al.[63] (2021) investigated the effect of ocular surface cooling on the maximum inter-blink period. Using the Teledyne FLIR A655sc camera to record OST, Matlab programming was used to segment the cornea using the first full frame after a blink. The subject was asked to maintain a steady fixation during the recording, and the position of the cornea was considered to be the same for all subsequent frames after the first detected frame. Like other single camera methods, the OST measurement was considered on the estimated corneal area and the method was unable to track the cornea movement and precisely localize corneal area on the frames.

2.3.2.1.2 Semi-automatic methods

The manual methods, while simple in their performance, have deficiencies arising either from the camera technology itself – low thermal and temporal accuracy, varying degrees of invasiveness – or from the limitations of data analysis – they are non-automatic and require much post-processing. By improving both the thermal camera capabilities and quality, researchers have tried to improve their measurement results from the eye surface. Consequently, some image processing algorithms have been implemented on the thermal images to reduce user manual input.

Acharya et al.[45] (2009) proposed a semi-automated method to acquire OST. The original image of the eye was converted to grayscale for image processing purposes. Then, the image was cropped manually to localise the ROI of the image and resized to 400x200 pixels. The circular center of the cornea was detected by their algorithm by assuming that the corneal radius was one-fourth of the entire anterior ocular surface. Then, the profile line was plotted in the middle of the cornea for temperature measurements. Their main finding was that the average OST and temperature distribution reduced with age.

In another study, Acharya et al.[64] (2014) used thermography and a feature extraction method to find the treatment results in DED patients. The thermographs (thermal images) were captured immediately after a blink. Then the ocular region was segmented from the image. Feature extraction was used to extract some texture information of the thermal images. Using the extracted information, a decision could be made on the patient's response to treatment.

Matteoli et al.[65] (2018) proposed a new method based on image processing techniques for thermal image analysis. A fixed eye shape, bounded by two arches, symmetrical both horizontally and vertically, and narrower in the area representing both canthi, represented an ideal eye. Then, the thermal images of the eyes were adapted to this fixed eye shape by reversible geometric transformation. Their method did not depend on ocular geometry and could be applied to compare OST between the left and right eyes, or even between different groups of people.

2.3.2.1.3 Fully automatic methods

Although an improvement in both hardware and software processing, the semi-automatic methods still required user input. In response, some groups tried to produce a fully automatic method. The key technical challenge here is to locate the corneal boundary in the ocular thermogram.

Tan et al.[66] (2009) proposed a fully automated method to measure OST. Using their method, which included a snake algorithm and target tracing function, the eye was localised on the image, and the eyelash effects removed from the image. Finally, a circle was drawn in the center of the eye automatically to achieve cornea temperature data. However, their algorithm could not omit the influence of the eyelash completely and the accuracy of segmentation was not satisfactory.

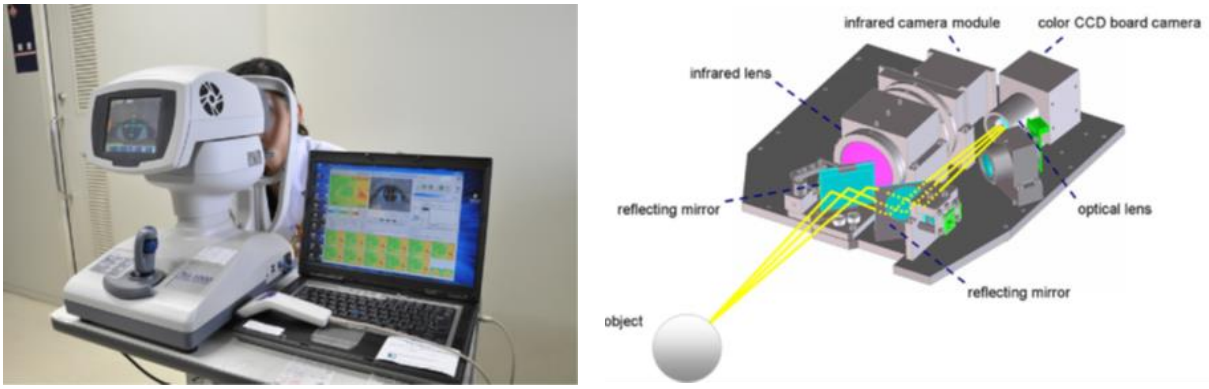
Shuang et al.[67] (2016) used active contour modelling to extract the eye from other parts of the image. Their algorithm was able to deal with low resolution and blurred images. The algorithm was compared with the Tan et al. algorithm and, although the method had more processing time, it was improved in terms of accuracy compared with the other method.

2.3.2.2 Dual camera systems

The manual, semi-automatic and fully automatic methods described thus far depend on a single camera system. They are not suitable for OST measurement either because they are manual or have low accuracy in locating the corneal ROI. All of the algorithms only worked on the thermal images, but since the pixel intensity of a thermal image is based on temperature distribution, there is no clear edge between the corneal and surrounding area in the images. Hence, it is impossible to accurately segment (locate) the corneal area from the surface of the eye in the thermal images. To solve this problem, some groups have used a dual camera system to improve the accuracy of

the segmentation. They took the advantage of the visible image to find the corneal boundary in the thermal image.

Kamao et al.[68] (2011) used a visible (EMR) camera embedded in a thermal camera system to measure OST in DED patients in a semi-automatic method. Figure 2-9a shows the dual camera system. In the pre-processing level, the pupil of the subject's eye was first located in the center of the visible camera image by touching the computer control screen. By means of camera alignment, the thermal camera was therefore also centered on the pupil of the subject's eye. To provide semi-simultaneous viewing, two mirrors (one fixed and one moving) were used to reflect visible light from the subject eye into the visible camera detector. Figure 2-9 b shows the mirror installation in the system. The moving mirror rotated 45° every 0.25 secs to reflect the visible light to a fixed mirror and finally into the visible camera. During visible light transmission to the camera, the IR radiation was blocked from entering the thermal camera by the moving mirror. After recording the data, a customised software was used for further processing. The ROI was segmented from the images manually and the extracted data analysed. The images were compared for TBUT in DED patients by visual comparison of the coincident images.



a *b*

Figure 2-9: Dual camera system introduced by Kamao et al. (2011) a. camera setup, b. internal camera architecture and mirror installation[68] (Reproduced from: Kamao T, Yamaguchi M, Kawasaki S, Mizoue S, Shiraishi A, Ohashi Y. Screening for dry eye with newly developed ocular surface thermographer. Am J Ophthalmol. 2011;151(5):782-791)

Introducing this method was a big achievement at that time (2011). The corneal boundary was detected with much improved accuracy, and the location of tear film break-up on the surface of the eye could be estimated by an experienced user in a more precise way. However, the method was not synchronous and required a lot of user input to extract the data.

Su et al.[69] (2014) used a different dual camera system (thermal and visible camera) to analyse tear film break-up patterns. To permit simultaneous imaging of the thermal and visible cameras, a germanium beam-splitter was used (Figure 2-10). The beam-splitter transmitted IR light and reflected visible light. The thermal camera and beam-splitter were installed in front of the patient's eye and the visible camera was located alongside. The system was an improvement in terms of synchronisation over the previous method. However, user input was still needed to analyse the tear-film patterns.

Figure 2-10: Dual camera system using a semi-reflecting germanium beam-splitter to permit simultaneous visible and IR imaging of the eye surface, Su et al.[69] (Reproduced from: Su T, Chang S, Yang C, Chiang HK. Direct observation and validation of fluorescein tear film break-up patterns by using a dual thermal-fluorescent imaging system. Biomedical optics express. 2014;5(8):2614-2619)

In 2015, Li et al.[70] used a different camera installation to record two video files from the surface of the eye synchronously. The thermal and visible cameras were installed in front of the cornea within -15° and +15° off-axis from geometric center of the cornea. Then, two movies of the eye, one from each camera, were recorded for a specified amount of the time. After that, the movies were synchronised using FINAL CUT PRO X software (Apple Inc., Cupertino, CA, USA). Then, the tear film break-up area was estimated by an experienced user looking at the thermal and visible videos simultaneously.

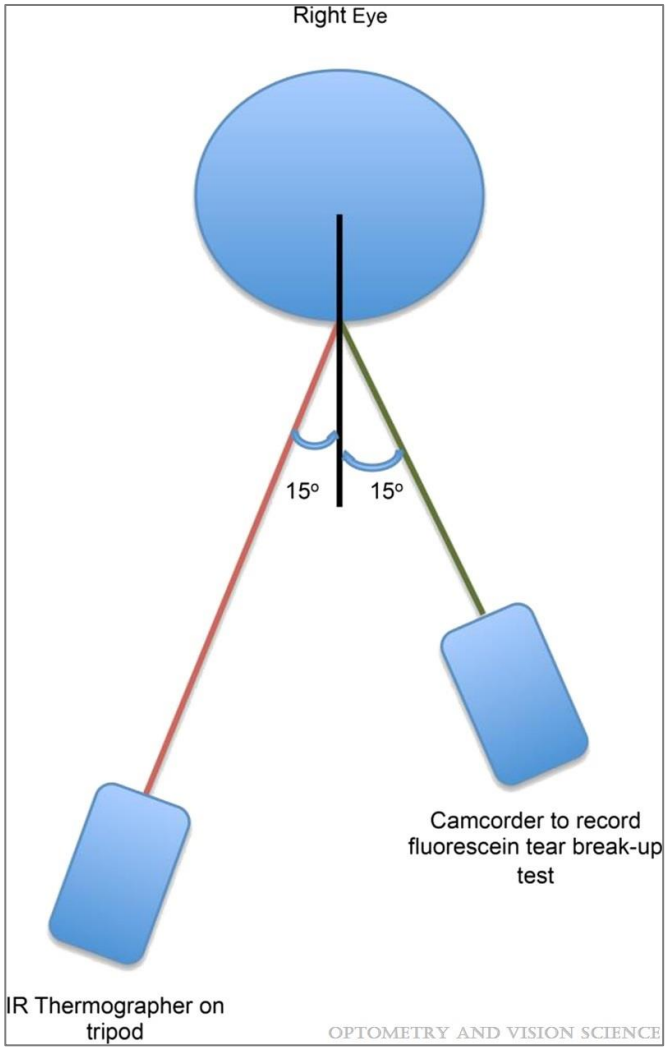


Figure 2-11: Topology of the camera installation introduced by Li et al.[70](Reproduced from: Li W, Graham AD, Selvin S, Lin MC. Ocular surface cooling corresponds to tear film thinning and breakup. Optom Vis Sci. 2015;92(9):e248-56)

Kricancic et al.[71] (2017) demonstrated another camera installation method (Figure 2-12) to monitor OST change on the surface of the eye. They used a germanium filter in a fixed position in front of the patient's eye. The system was used to monitor OST change in the contact lens wearer. However, their method still required manual input of the user.

Figure 2-12: New camera installation using semi-reflecting Germanium beam-splitter introduced by Kricancic et al.[71](Reproduced from: Kricancic H, McNeill H, Titze M, Alonso-Caneiro D, Collins MJ. Instrument for simultaneous assessment of fluorescein and thermal dynamics of the tear film. 2017)

A dual camera system is a simple method that permits location of the corneal and eye contour area in the thermal image. All of the reported dual camera system methods tried to have the best camera installation and adjustment in order to have the same image from the eye surface. However, adjusting two cameras to have the same field of view from a scene is difficult, and the resultant images from the two different sources are not always identical. Images from different sensors can be affected by many factors, such as sensor pixel size, image resolution, lens distortion, environment light, distance, angle of photography, and sensor type. Therefore, working only on camera installation alignment will not yield data extraction with a high spatial accuracy. To improve the accuracy of cornea localisation, the images should be processed after finding the best camera installation.

2.4 Specific issues for imaging of eye

To conclude, most of the described methods estimate the corneal location based on eyelid position and cannot track the corneal area during eye movement, blinking, or eccentric gaze. This occurs because the eye thermograph lacks a corneal boundary, which is clearly visible in color images. Also, other methods, which may have more accuracy, need user input to extract the data. Table 2-1 summarize the previous and current methods of the OST measurement.

Table 2-1: Summary of the previous and current methods of the OST measurement

Device	Author	Manual/Auto	Method
Single camera	Efron et al. (1989)[51]	Manual	Selecting 11points on the surface of the cornea and extracting temperature
	Morgan et al. (1995)[54]	Manual	Drawing five 10x10 pixel boxes on the cornea surface
	Mori et al. (1997)[43]	Manual	Extracting OST using a 20x20 pixel box (3.3mm2) located at the center of cornea
	Murphy et al. (1999)[57]	Manual	Five boxes placed along a horizontal line centered on the cornea
	Purslow et al. (2005)[12]	Manual	Locating 23 points across the cornea, grouped in five regions: central, superior, inferior, nasal, and temporal to measure OST
	Chiang et al. (2006)[72]	Manual	Drawing an encircled area (4.4mm diameter) on the center of the cornea
	Galassi et al. (2007)[55]	Manual	Locating five different points on the eye surface to measure the temperature
	Sodi et al. (2007)[56]	Manual	Locating five different points on the horizontal line of the cornea
	Ng et al. (2008) [59]	Manual	Drawing a circular area on the center of the cornea
	Matteoli et al. (2016)[60]	Manual	Drawing five areas (small, oval shape) on the surface of the eye
	Tan et al. (2016)[61]	Manual	The region of interest (ROI) was selected manually using the diamond method
	Sudarshan et al. (2017)[62]	Manual	Drawing a circle on the original image to extract the temperature manually
	Ding et al. (2021)[63]	Manual	Best fitted circle on the center of the eye was used to extract the temperature
	Acharya et al. (2014)[64]	Semi-automatic	Cropped the eye part of the image and fitted a circle in the center of the image
	Matteoli et al. (2018)[65]	Semi-automatic	A fixed eye shape, bounded by two arches, symmetrical both horizontally and vertically, and narrower in the area representing both canthi, represented an ideal eye
	Tan et al. (2009)[66]	Automatic	Using snake algorithm and target tracing function, the eye was localised on the image automatically
	Shuang et al. (2016)[67]	Automatic	Active contour algorithm was used to localize the cornea and extract the temperature
Dual camera	Kamao et al. (2011)[68]	Semi-automatic	Using a moving mirror to send visible light to visible camera every 0.25 sec-manual OST extraction by looking at visible and thermal images side by side
	Su et al. (2014)[69]	Semi-automatic	A germanium filter was used as a beam splitter to capture synchronous images- manual OST extraction by looking at visible and thermal images side by side
	Li et al. (2015)[70]	Semi-automatic	The cameras were installed in front of the cornea within -15° and +15° off-axis from geometric center of the cornea to capture images- manual OST extraction by looking at visible and thermal images side by side
	Kricancic et al. (2017)[71]	Semi-automatic	Using germanium beam splitter with different installation- manual OST extraction by looking at visible and thermal images side by side

All of the above methods made a good contribution to OST measurement, but there are some weaknesses with each method. For example, the manual methods need an experienced user input, and the semi-automatic and automatic methods (single thermal camera) using images in which the corneal boundary is indistinct and therefore only an approximation of the boundary can be provided. The dual camera systems provided better localisation accuracy, but the method still requires user input. None of the described methods are able to localize the cornea with high accuracy and track it over time. Also, the methods cannot detect and remove blink artefacts from the image sequences.

Improvements can be made by overlaying high-resolution visible and thermal images, captured using a dual camera system, to maintain tracking of OST during eye movement and blinking, and detection of the elliptical palpebral aperture. Such a system provides an opportunity to analyse the ocular surface with much improved accuracy. As a result, the system could be applied in the diagnosis and management of DED, in monitoring changes in contact lenses during lens wear, and in understanding the role of temperature change as a stimulus for blinking. The main purpose of this thesis is to report on the design of an automatic system for recording and analysing OST of the human eye.

Chapter 3: System Design

The best system for measuring and tracking OST over time is to use a dual camera system consisting of a thermal and a visible camera. The first step of the system design was to consider system requirements, and this is reported in Section 3.1. Hardware design and selected hardware are explained in Section 3.2. Section 3.3 describes the data acquisition algorithm and data gathering method.

3.1 System Requirements

To develop a system for measuring and tracking OST over time, both hardware infrastructure and algorithm developments are required.

The hardware consists of:

- Thermal camera
- Visible camera
- Camera lenses
- Connection cables
- Camera mounts
- Patient chin/head rest for adjustment

To design the hardware infrastructure for imaging the eye surface, two cameras are required: a thermal (IR) camera and a visible (machine) camera. In selecting each camera, the following parameters should be considered: thermal resolution, thermal sensitivity, image resolution, frame rate (frames per second (fps)), field of view (FOV), and availability of digital I/O connection.

- Thermal and image resolution describe the same characteristic: the ability to detect and record detailed information from the eye surface. A resolution of 640x480 pixels or higher provides enough information from the eye surface for the analysis step.
- Thermal sensitivity describes the ability of the thermal camera to detect small changes in OST across the ocular surface. The temperature range across the eye is small and the precision of the measurement is very important. A detected temperature variation of at least two decimal places provides enough information for decision making.

- The frame rate describes the number of frames per second (fps). It is important to provide enough data for processing, but if the fps is too high, an excess of data is generated which would need additional processing time and a more powerful processor. If the fps is too low, then insufficient information is generated to track OST over time. A framerate of 12-25 fps provides enough information to track OST between two blinks.
- The FOV is the coverage area across the ocular surface and nearby eye and face (adnexa) that can be seen by the camera. The FOV can be changed by lenses, with a close-up lens reducing the FOV, but enabling more detailed information of the eye surface. The lens should be selected based on the distance of photography and the required FOV.
- For temperature measurement, the thermal camera should be close to the eye to capture the temperature with the highest accuracy. A close-up lens must therefore be fitted to enable a clear focus. To match the thermal camera FOV, the visible camera should also have a close-up lens with a similar FOV. The visible camera should also be able to image the eye surface from the same viewing distance, or from a little further or closer for best mounting flexibility.
- Another requirement for camera selection is a digital I/O port. The port helps to connect each camera via cable and enable synchronous imaging.
- A further requirement for the system is a suitable camera mount. Each camera should be installed on the same mounting plate to ensure that alignment adjustments for each camera are made in conjunction with the other camera
- The camera mount must also be attached to an adjustable mechanism that allows changes to the height and angle of the camera alignment with the head and eye of a subject.
- Lastly, a head and chin holder are required to stabilise the subject's head during image capture.

A general schematic of the system is shown in Figure 3-1. The thermal (1) and visible (2) cameras were installed on the camera mount (3), which is flexible and can rotate to provide different angles of view and provide different relative positions of the cameras to the eye. The camera mount was installed on an adjustable arm (4) of a modified slit-lamp biomicroscope that could rotate the whole camera system in front of the subject eye. The adjustable arm was an integral component of the biomicroscope and could be rotated in front of the eye. The biomicroscope could also be moved

(5)(6) up/down, left/right, and forward/backward. The biomicroscope incorporated a chin/head rest (7) that could be adjusted for eye alignment and patient comfort (8). The camera imaging system was controllable by a computer system (9).

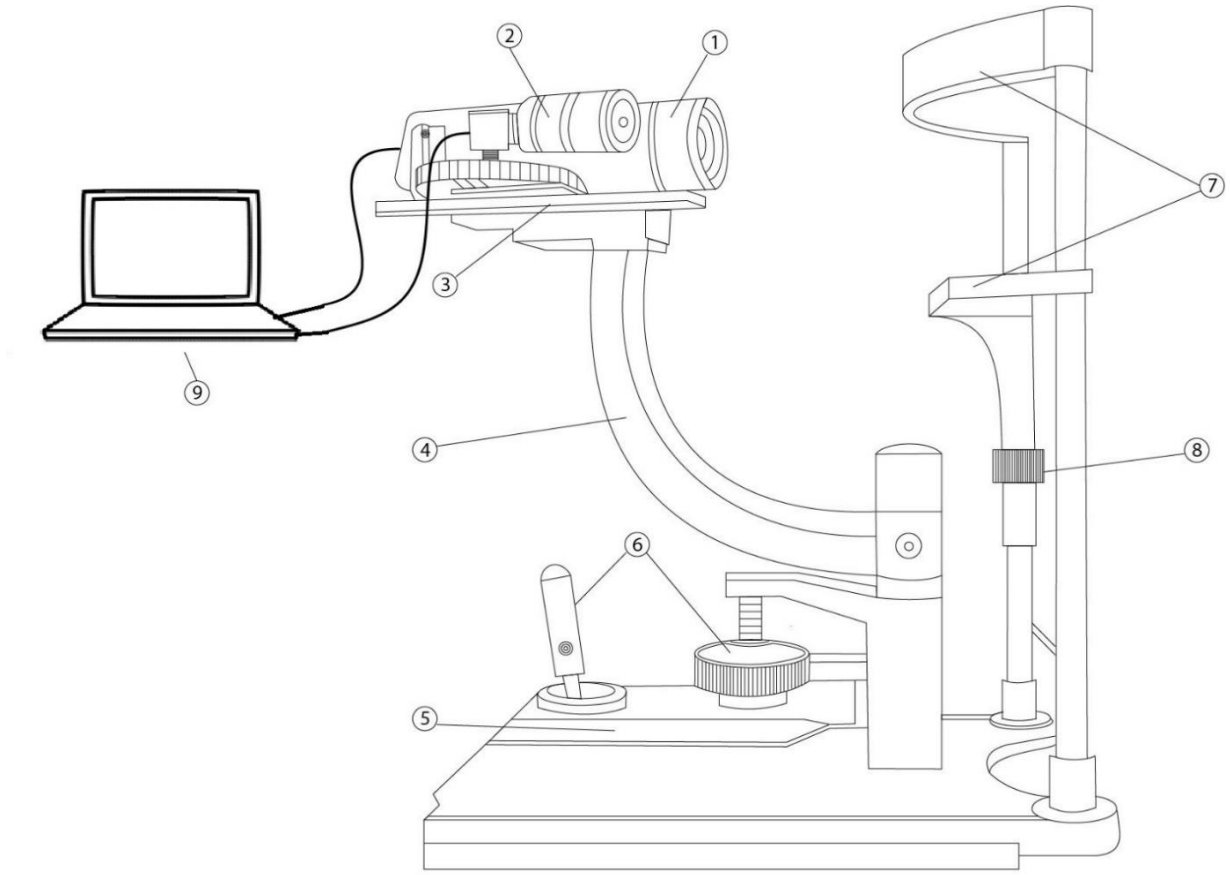


Figure 3-1: General schematic of the instrument set-up for the system.

3.2 Hardware selection

3.2.1 Thermal camera

A Teledyne FLIR IR A655sc thermal camera (Teledyne FLIR LLC, Wilsonville, OR, USA) was chosen for this project. This camera uses an un-cooled CCD detector (microbolometer) that is sensitive for IR wavelengths from 7.5 to 14µm and is able to distinguish temperature differences down to 30mK. This wavelength range falls within the IR-C or Far IR range. The thermal camera can capture an image in temperature or radiometric (raw) format. For this project, the radiometric format was selected as it provided more sensitivity in the data, and the data could be converted into any desired format, i.e., Celsius, Fahrenheit, radiometric, or grey value.

The camera was able to capture full frame size (640x480) images at 50Hz, or a smaller window size at speeds up to 200Hz. For this project, the full frame size and 25Hz frame rate was chosen. Individual pixel size reported by the camera manufacturer on the CCD is 17µm, which provided detailed information of the ocular surface at the short focal lengths used, since the pixel size projected onto the ocular surface is 0.09 mm). For more detailed imaging of the eye and adnexa, a polished germanium close-up lens was attached to the objective lens of the camera. Germanium lenses have a high refractive index (about 4.0 for a 2-14 µm incident wavelength) and are transparent for IR radiation. This lens is designed to provide sampling resolution of extremely small targets (640 x 480 pixels with an on-eye pixel size of 0.09 mm) and has a field of view (FOV) of 64 x 48 mm with a 172 mm working distance, and a magnifying factor of 5.8x. By using the close-up lens, the image of the eye could be adjusted to fill the full frame size. Camera focus was controlled manually. Figure 3-2 shows the thermal camera and the germanium close-up lens.

a *b*

Figure 3-2: a) Teledyne FLIR IR A655sc thermal camera with close-up lens attached used for temperature recording; b) Germanium close-up lens (Image from: Teledyne FLIR LLC, Wilsonville, OR, USA)

For operating control, the thermal camera used the Spinnaker Software Development Kit (SDK) (Teledyne FLIR LLC, Wilsonville, OR, USA).

3.2.2 Visible camera

Prior to selecting a suitable visible light camera, a prototype dual camera system was developed using a webcam in combination with the thermal camera. By analysing the system, it was found that a machine vision camera that could be controlled by a customized software was the best fit for the project. Therefore, the FLIR BFS 51S5C-C camera (Teledyne FLIR LLC, Wilsonville, OR, USA) was selected. This machine vision camera has a high frame-rate sampling speed, high spatial

resolution, and a distortion free lens. It uses a CMOS Sony IMX250 sensor (SONY Group Corp, Tokyo, Japan) with a pixel size of 3.45 µm which provides detailed information of the scene. The maximum resolution (progressive scan) of the camera is 2464x2056 pixels, which provides very good quality images of 5 mega pixels at a frame rate of 75 frames per second (fps). It has a dynamic range of 71 dB and saturation capacity of 10330 which provides an excellent dynamic range, permitting good quality images in poor lighting environments. The red/green/blue (RGB) pixel format was selected for image output as this provided the full range of color information for each image.

The size of the camera is very small (3x3x4cm), which allowed it to be mounted beside the thermal camera on a single mount and with a minimal separation distance between the two cameras. A Fujinon HF12.5SA-1 close-up lens (Fujifilm Corp, Tokyo, Japan), which is designed for macro photography, was fitted to the camera. This lens has a focal length of 12.5 mm, which gives a field of view (FOV) of 12 x 9 cm at 17 cm. Figure 3-3 shows the selected machine vision camera and the close-up lens used for this project. Camera focus was controlled manually

The visible camera could also be controlled using the Spinnaker Software Development Kit (SDK) (Teledyne FLIR LLC, Wilsonville, OR, USA), which enabled the development of a single software program that could control both cameras together.

a *b*

Figure 3-3: a) Teledyne FLIR BFS 51S5C-C machine vision camera selected for visible image recording, b) Fujinon HF12.5SA-1 close-up lens. (Image from: Teledyne FLIR LLC, Wilsonville, OR, USA)

3.2.3 System integration

Having selected the two cameras for the project, the next step was to design the dual camera system. Since the intention was to have both cameras focused on the same area of the eye and its adnexa, both the thermal and visible cameras needed to be positioned on a mount in front of the

subject's eye. Ideally, both cameras should be aligned directly in front of, and perpendicular to, the eye, but with two separate cameras, that was not possible. A zero relative angle between the cameras is optimal as this would cause less skewing between the images.

A decision was made to install the thermal camera perpendicularly in front of the patient's eye (zero angle) on a horizontal mount positioned on top of an adjustable arm, and to have the visible camera installed on one side of the thermal camera on the same mount. The visible camera was small in size and could be installed very close to the thermal camera. Thus, for simplicity in hardware set-up, it was assumed that there was a zero angle between the cameras.

The adjustable arm on which the camera mounting was positioned formed part of a slit-lamp biomicroscope that was mounted on a table, and this arrangement allowed the cameras to be moved up/down, left/right, forwards/backwards, and to rotate around the subject's eye. This flexible arrangement allowed the capture of thermal and visible images at different angles and distances. The camera mount also allowed the visible camera to be adjusted to a range of different angles relative to the thermal camera, if needed (Figure 3-4 b). Lastly, the slit-lamp instrument included a headrest on which the subject could position their head.

The cameras were installed in such a way that each camera imaged the same area of the face and eye during photography. To get the best performance in the image processing steps, it was better to have the same image size for both thermal and visible images. The maximum image size of the thermal camera is 640x480 pixels, therefore the same image size was chosen for the visible camera. The FOV for the visible camera was also adjusted to match the FOV of the thermal camera. By designing the system to have similar image sizes from both cameras, this gave the best images in terms of similar eye size in the images across the same pixel dimensions and size.

a *b*

Figure 3-4: a) Mounting arrangement for the two cameras showing the movable arm mounted on the slit-lamp biomicroscope, the camera mounting and the subject headrest, b) different views of the visible camera mount (top, middle) that was attached to the thermal camera (bottom).

3.2.4 Camera synchronization

When two or more cameras are used to capture an image of the same scene or event at the same time, it is possible to synchronize when image capture occurs. Synchronization permits each camera to capture an image at the same time point or with any desired time interval. In this system, two cameras were used, and camera synchronization was an essential component of the image capture system. Camera synchronization can be arranged using either software or a hardware trigger.

3.2.4.1 Software trigger

In this arrangement, the different cameras are controlled by a single software program that acts as the trigger to control image capture by the cameras. When the software is running, a signal is sent to each camera commanding it to capture an image. However, although the initiation of the command for image capture can be synchronized, the exposure time of each camera might be different, since that can be affected by factors such as system processing power, memory, or type

of coding. Usually, thread programming is necessary for the camera management software to manage these problems, which is costly.

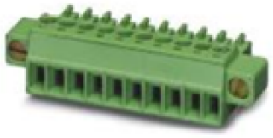
3.2.4.2 Hardware trigger

Another way of camera synchronization is hardware triggering, in which each camera is physically connected to the other camera by a cable. To synchronize the cameras, one camera is considered as the primary and the other as the secondary. The primary camera strobe is used to trigger the secondary to start image capture. The strobe is an electronic signal that occurs when a camera starts to capture an image. Since the two cameras are linked by a cable, the primary camera strobe triggers the secondary camera. In effect, the secondary camera is synchronized with the strobe, and so the frame rate of an image sequence for the secondary is the same as primary camera. If the system memory for the synchronized cameras is not able to store the images in real-time, the cameras must use their buffer to record the frames to slow the data stream and allow the system to read it by its speed. In this way, there is no loss of frames because of the system speed, and the system cannot affect the camera synchronized capturing. Since the secondary camera is controlled by the primary, the frames are recorded within a millisecond delay.

3.2.5 System camera connection and synchronization

The hardware trigger mechanism was chosen for this project, with the thermal camera used as the primary and the visible camera as the secondary. To connect the cameras, the data ports of each camera were connected in such a way that the thermal camera could send the trigger signal to the visible camera. Figure 3-5 and Figure 3-6 shows the port map for the thermal and visible cameras, respectively.

Digital I/O connector



Pin	Function	Data
1	IN 1	opto-isolated, 0 - 1.5V = low, 3 - 25 V = high
2	IN 2	opto-isolated, 0 - 1.5V = low, 3 - 25 V = high
3	OUT 1	opto-isolated, ON = supply (max. 100 mA), OFF = open
4	OUT 2	opto-isolated, ON = supply (max. 100 mA), OFF = open
5	Supply VCC	6 - 24 VDC, max 200 mA
6	Supply Gnd	Gnd

Figure 3-5: Teledyne FLIR A655sc digital in/out data connectors (Image from: Teledyne FLIR LLC, Wilsonville, OR, USA)

For the thermal camera s primary, either Pin 3 or 4 could be used as an output. Pins 5 and 6 were used for the 12V power supply.

Diagram	Color	Pin	Line	Function	Description
	Green	1	3	V_{AUX}	Auxiliary Input Voltage (DC)
				GPI	Non-isolated Input
	Black	2	0	OPTOIN	Opto-isolated Input
	Red	3	2	VOUT	Camera Power Output
				GPIO	Non-isolated Input/Ouput
	White	4	1	OPTOOUT	Opto-isolated Output
	Blue	5	N/A	Opto GND	Opto-isolated Ground
	Brown	6	N/A	GND	Camera Power Ground

Figure 3-6: Teledyne FLIR BFS GPIO digital in/out port connectors. (Image from: Teledyne FLIR LLC, Wilsonville, OR, USA)

For the visible camera, Pins 2 and 5 were used for the trigger input from the thermal camera. To connect the cameras, a GPIO cable with a 6 pin Hirose circular connector was used (Figure 3-7).

Figure 3-7: 1m GPIO Cable with a 6 pin Hirose HR10 circular connector. (Image from: Teledyne FLIR LLC, Wilsonville, OR, USA)

The cable is color-coded. Therefore, the GPIO connector was connected to the GPIO port of the visible camera and the trailing black wire from the cable was connected to pin 3 of the thermal camera. The blue trailing wire from the cable was connected to pin 6 of the thermal camera.
Both cameras were connected to a PC computer. The thermal camera was connected via the GigE vision port to provide the maximum speed and minimum delay between frames in the video sequence. The visible camera was connected to the computer using a USB 3.1 cable.

3.2.6 Frame rate

There were three options for camera frame rate: 12.5, 25, or 50 fps, based on the available frame rates of the thermal camera at full resolution. A frame rate of 25 fps has a 40ms gap between frames. This was felt to be the optimum frame rate available since it would be fast enough to detect eyelid blinking during any video sequence, while not producing lots of data, and slow enough to provide detailed information of any temperature changes while the eye was open. In particular, during a blink there will always be a few frames captured that contain all or part of the eyelid closure. A 12.5 fps would reduce the overall temperature data collected while also providing less information about the blink – eyelid movement may only be observed in one or two frames. A 50 fps frame rate would provide more detailed information, but processing of the data would be computationally costly and need more time. Therefore, the frame rate of 25fps was selected for the project.

3.3 Camera Management and Data Acquisition

To acquire data from the eye surface using the dual camera system, a software was required to manage the cameras and record the data on a disk. Before developing the software, an algorithm should be developed for the camera management. The algorithm should be designed and developed based on the system requirements. For the camera management and data acquisition software, the algorithm should be developed in a way to capture synchronous frames from both cameras and save it to the computer memory. Also, the algorithm should be developed in a way to write a timestamp on each image file for synchronization and to record time after blink. After developing the algorithm, a programming language should be selected to implement the algorithm. In the next section the programming method is described.

3.3.1 Programming language

Having selected the two primary cameras (IR and Visible) for the system, the next step was to develop a camera management software that could interact with the proprietary camera driver software of each camera. The software needed to be able to recognize the camera drivers and have access to each camera registry to be able to control each camera's hardware. Each camera has a dynamic link library (dll) file and software development kit (SDK). Specifically, it was the SDK that was used to access the camera hardware and control it. Since the two cameras selected for the project were from Teledyne FLIR, both cameras were able to support the same SDK called Spinnaker SDK. Spinnaker SDK is a Teledyne FLIR next generation GenICam3 API library designed for machine vision software developers. Using the Spinnaker SDK, a customized software was developed to control the cameras. The SDK could also be imported as a library to a C++, C#, Matlab, or Python programming environment.

After setting up the camera system installation and connecting the two cameras together using the synchronization cable for hardware triggering, the data acquisition and image processing steps were controlled by a customised software program. For this system, the Python programming language (Centrum Wiskunde & Informatica, Amsterdam, Netherlands) was selected to develop the camera management software, data acquisition software and image processing software as it was found to offer more flexibility and performance for these future steps of the project, and especially when using deep learning algorithms.

3.3.2 Management software interface

A graphical user interface (GUI) was designed for the software to enable easier control of the system. Figure 3-8 shows the GUI of the dual camera management software. The graphical user interface (GUI) was designed from scratch. Before the design, all of the requirement for the program were considered. Then, a form was developed and two placements for the video were designed on the form. Six control buttons, each with a different function, were designed with a suitable background to provide the user-friendly user interface. The buttons were defined to work with a keyboard shortcut or mouse click. After designing the GUI, code development was started. The code was written for each part of the software. For example, each button needed several lines of codes to be functional, and the code written was based on the designed algorithm in the previous steps.

Using the GUI interface, a preview of the video stream from each camera was displayed side-by-side, and synchronous image sequences could be recorded using a single button click. This enabled the user to make a visual check that both cameras were operating and focused on the same area of the subject's eye, and that the subject was not making excessive eye movements or blinking (Figure 3-10). The software also had the ability to replay recorded video files beside each other to allow observation of any changes on the eye surface during the recorded video sequence or to see the final video output after image processing.

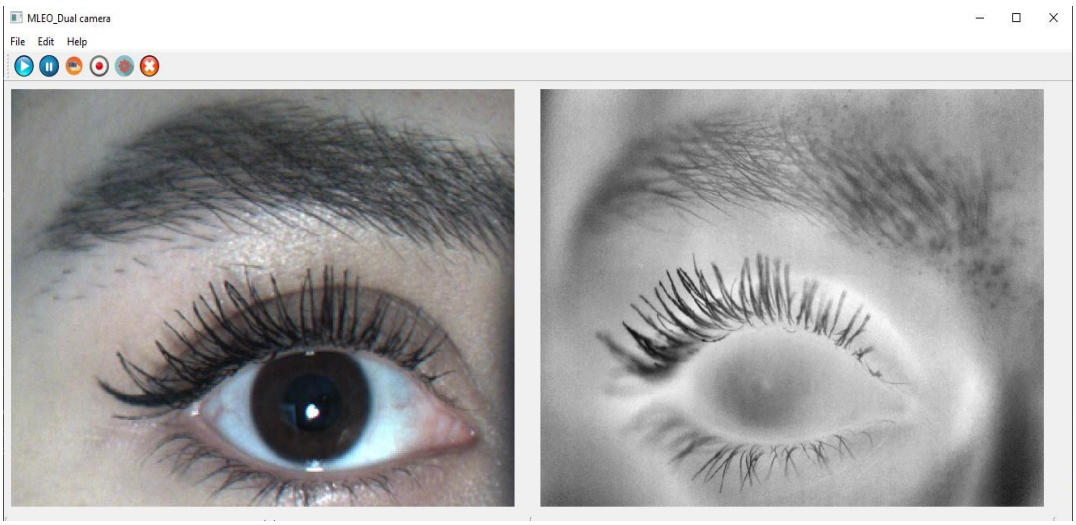


Figure 3-8: Graphical user interface (GUI) for the dual camera management software showing the visible camera feed (left) and thermal camera feed (right), and software control buttons (top left).

The specific software management controls are shown in Figure 3-9.

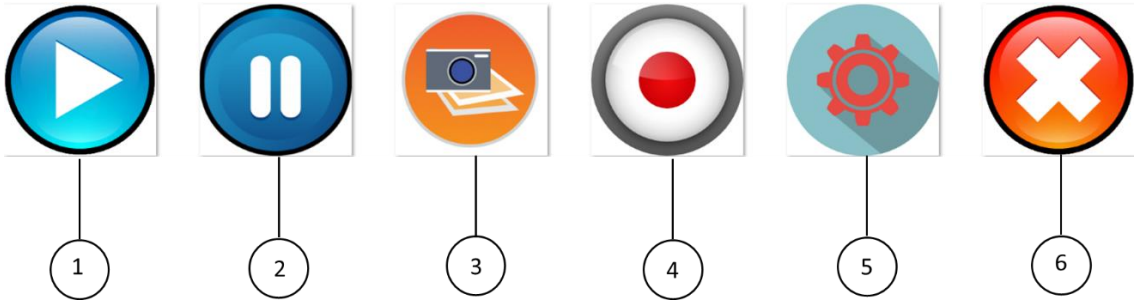


Figure 3-9: GUI control buttons for the dual camera management software. Button function: 1: Start/Resume camera preview; 2: Pause camera preview; 3: Capture image from the current view; 4: Record button; 5: Camera and record settings, 6: Stop cameras and Close program.

3.3.3 Data acquisition

Data acquisition from the two cameras was one of the primary requirements for the system, and a reliable method of collecting and recording the image sequences from each camera was essential. The customized software had the ability to record individual image frames from each camera in a synchronized sequence that started from a single reference frame in the thermal camera.

To ensure synchronous data acquisition, the software relied on the video image timestamp for each image frame. When a camera captures an image it notes the exact time of the capture, which can be the system time or the camera time, and the time is attached (stamped) to the image frame. This is called the timestamp and it is a permanent feature of each image frame. The timestamp was used to check the synchronization between the cameras by looking for any time difference between two frames. In this system, although both cameras were manufactured by Teledyne FLIR, each camera used a different timestamp format. Thus, to create a single timestamp for both cameras, the timestamp format from the thermal camera was converted to the format of the visible camera. The timestamps for the two recorded frames of the target object (one from each camera) were compared to check that the cameras were capturing simultaneous images. Since both cameras were set to capture images at 25 fps, the timestamp was able to reveal a time difference between frames for both cameras of 40 milliseconds.

To obtain a set of paired thermal and visible video sequences recorded from the eye surface, the subject was asked to maintain fixation in primary gaze (straight-ahead) and to blink normally during the recording. Image synchronization of the two video streams could be checked visually by reviewing the video files from each camera that were presented side-by-side in the GUI.
For subsequent image processing, a sequence of individual and timestamped frames was extracted from each video file and numbered in order from an initial reference frame. Figure 3-10 shows an example of the extracted frames from the visible and thermal videos. By looking at the frames, it is obvious that the cameras captured the frames synchronously. A blink can be observed to start at frame 32 and end at frame 40 in both video sequences.

Figure 3-10: Visible (a) and thermal (b) extracted video frames used for side-by-side comparison, numbered from an initial reference frame.

The algorithm was designed to record the video files and store them on a memory disk with a recorded timestamp for each frame. The frames were then extracted from the video files using a developed software. The frames were then sent to the image processing software for further processing.

Chapter 4: Image normalization

After developing the dual camera operating software, the image sequences recorded by each camera in the system must be processed to remove the inherent aberrations in the image using a customised image processing algorithm. Each image captured by each camera is affected by the quality of the camera's optics and by the sensor characteristics. These features can be calculated for each camera and used to calibrate the image by removing the associated artefacts.

4.1 Image normalization

By having knowledge of the effect of the camera's optical and sensor characteristics, the images can be pre-processed to remove these effects and be prepared for future processing steps. This process of identifying and correcting the errors is called image normalisation. Image normalisation can include a variety of pre-processing steps, such as contrast adjustment, size change, color change, and distortion removal. For this project, only image distortion will be corrected.

Image normalization ultimately improves the output from the image processing algorithms used for object recognition, pattern recognition, corner detection, and image segmentation. In medical image processing (such as this project), it is important to have precise and accurate output images after image processing algorithm implementation, and so ensuring that the input images for that process are as unaffected by artefacts as possible is essential to that task. Hence, image pre-processing is an essential part of the project.

The first step of the pre-processing is the camera matrix calculation which involves determining and extracting the intrinsic and extrinsic camera parameters. The intrinsic parameters represent the optical characteristics of the camera, such as focal length, principal points, distortion, and skew. The extrinsic parameters represent the location and position of the camera in relation to the object which can produce rotational and translational artefacts. The calculated intrinsic parameters can be used to calculate the lens distortion coefficient, by which the images taken by the cameras can be undistorted.

Based on the manufacturer's camera specification (Teledyne FLIR BFS), there should be no intrinsic image distortion with the images produced by the visible camera. However, the manufacturer's camera specifications are measured under ideal conditions that do not match the operating parameters of this system, and some intrinsic effects are expected. Similarly, the thermal camera will be affected by a small number of intrinsic effects. Both cameras will have some extrinsic distortion, due to the side-by-side camera mounting needed to align the two cameras that must be removed from the images.

4.2 Intrinsic optical characteristics

The optical performance of the camera lens (or lenses) has a direct impact on the quality of the image produced – in other words, the inherent optical aberrations of the optical system distort the image produced by that system of the object. Aberrations occur when the rays of light from a point source object that pass through a lens or optical system do not converge to form the image at a single point. These aberrations are an inevitable effect of how the EMR waves (either visible or IR) interact with the refracting lens material, and how the lens material is distributed across and through the body of the lens.

There are two types of lens aberrations: monochromatic and chromatic. Monochromatic aberrations produce effects in the focus, magnification and distortion of the image, and chromatic aberrations produce dispersion of the EMR waves (according to their wavelength) to produce colored fringes around the locus of image formation.

4.2.1 Monochromatic aberrations

These types of aberrations are best described using a single wavelength of EMR and a point source as an object. The effect produced is that the image formed by the lens is not focused at a single point (locus). There are five different sub-types of monochromatic aberrations with a 3rd order approximation to Snell's law.

4.2.1.1 Focus (or defocus) error

This is the simplest type of aberration and occurs when the intended image plane of the optical system (e.g., the camera's sensor or film) is not coincident with the actual image plane of the lens. It is a familiar error for any photographer and is corrected by adjusting either the focal power or

position of the focusing lens to move the image plane of the lens forward or backwards in optical space to be coincident with the image plane of the optical system. Defocus produces errors in the sharpness and contrast of the image.

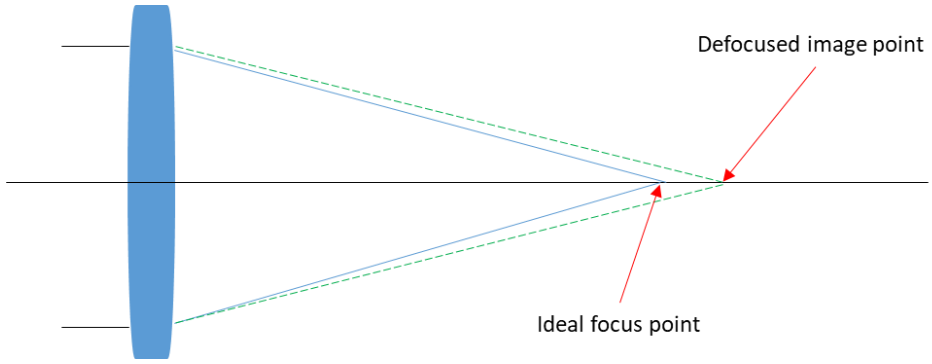


Figure 4-1: Ray diagram illustrating defocus aberration

4.2.1.2 Spherical aberration

This error is similar to focus error, but it is a consequence of how the EMR waves interact with the focusing lens. Even when defocus error has been removed by adjusting the lens image plane position, a small amount of focus error remains. This is caused by spherical aberration.

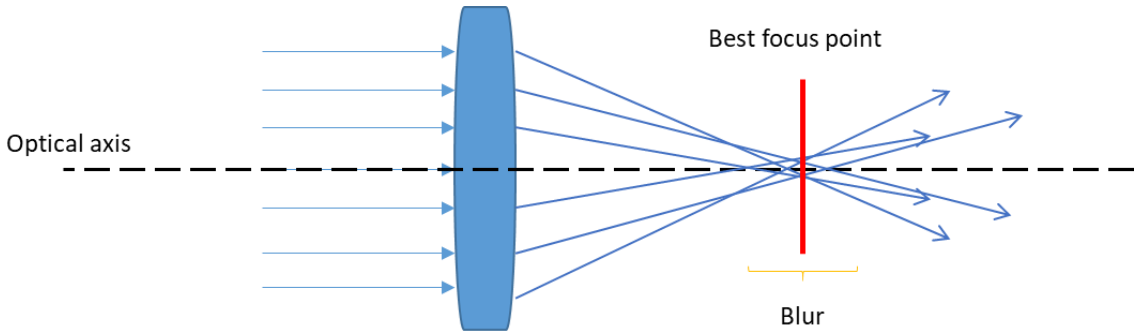


Figure 4-2: Ray diagram illustrating spherical aberration for a distant object

EMR waves (visible light or IR) passing through an optically transparent structure (e.g., a lens, prism, or glass block) are affected by the material in a process called refraction. The lens has an optical density that describes the strength of the interaction – the refractive index. A material with a higher refractive index will have a stronger effect. One benefit of this is that a thinner lens of higher refractive index will produce the same refractive effect as a thicker lens of lower refractive index. This thickness effect is also important for spherical aberration.

The thickness of a lens changes across its surface – a positive (convex) lens is thicker in the center of the lens compared to the periphery, while a negative (concave) lens is thicker in the periphery compared to the center. A ray of light passing through a thinner or thicker part of the lens will be affected (refracted) by the lens by different amounts – in this case, less and more, respectively – since it encounters differing amounts of the optically dense material. Light passing through air is unaffected since air considered to have a negligible optical density.

There is a second feature of refraction that relates to the angle that the ray of light from an object makes with the surface of the transparent, refracting material. The central ray from the object that passes along the optical axis of the system will enter the material perpendicular to the front surface and will appear to be unaffected by the structure – the light will pass through the material un-deviated. In comparison, a ray of light that leaves the object at an angle to the optical axis will meet the surface of the structure at an angle and, as a consequence, will be deviated as it passes through the material, with the size of deviation dependent on the angle. When the ray then exits the material, it experiences a further deviation. It is the cumulative effects of these two deviations that enables the lens to focus the light. This relationship is described by Snell's Law (Figure 4-3: Ray diagram illustrating Snell's Law. The blue line represents the incident ray path of light passing through air (refractive index n1=1). At the interface between the two refractive indices (refractive index n2>1), the light is refracted. The new ray path is represented by the green line.).

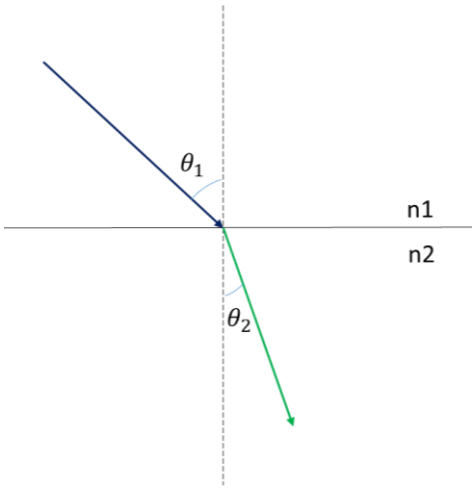


Figure 4-3: Ray diagram illustrating Snell's Law. The blue line represents the incident ray path of light passing through air (refractive index n_1=1). At the interface between the two refractive indices (refractive index n_2>1), the light is refracted. The new ray path is represented by the green line.

Spherical aberration is a consequence of these two effects. A ray of light travelling through the periphery of the convex camera lens has a greater incident angle than light passing through the center of the lens, and so experiences a greater level of refraction. This is also modified by the distance the light must travel within the optically denser material. The optical effect is that light from a point source of light passing through the lens is not all focused exactly at the same image locus. Rather, the image is spread out along the optical axis of the system (Figure 4-2). On the image plane, the visual effect for a point source of light is to see a central focused spot surrounded by a halo of light. The size of the halo represents the magnitude of the spherical aberration, and this is controlled by the refractive index of the lens, the dioptric power of the lens, the wavelength of light, and whether an aperture stop is used. By using an aperture stop in the optical system, peripheral light rays can be prevented from contributing to the image, but at the cost of reducing the image intensity.

4.2.1.3 Coma or comatic aberration

When considering a normal visual scene, rather than a point source, the effect of spherical aberration is repeated for each object point across the image that is imaged onto the camera's film or sensor, but it is also further affected by the relative position of each object point in relation to the optical axis.

The central ray of light from an off-axis object will not pass along the optical axis (by definition, since the object is off-axis) and so will not pass perpendicularly through the lens – indeed, no ray of any angle emitted from an off-axis object will be able to pass through the lens unaffected. Each off-axis object point is thus affected by spherical aberration that is distorted by the off-axis location of the object point - the further off-axis the object is, the greater will be the effect. The visual effect for an off-axis point source is that the central focus and circular halo of the object is distorted, and the image appears to have a tail (coma), like a comet (Figure 4-4). The magnitude of the comatic aberration is also controlled by the refractive index of the lens, the dioptric power of the lens, the wavelength of light, and whether an aperture stop is used.

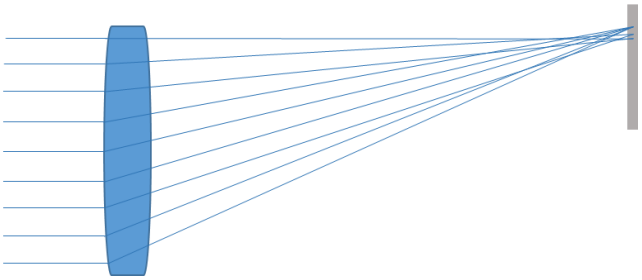


Figure 4-4: Ray diagram illustrating comatic aberration for a distant target

4.2.1.4 Astigmatism or astigmatic aberration

Although a lens may be designed to have a spherical surface, errors in the manufacturing process may produce a lens surface that is slightly distorted. This effect is called astigmatism and all lenses have some degree of astigmatism unless they are very carefully manufactured. The aberration may be regular, in that it can be defined as following a pattern, or irregular, in having no pattern or predictability. The optical effect of this distortion is to produce multiple focal points for the lens that represent the multiple surface curvatures of the lens.

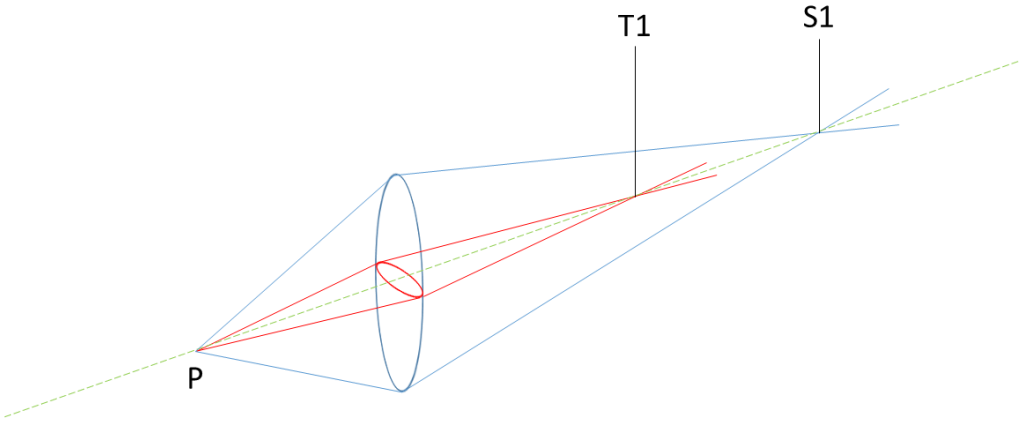


Figure 4-5: Ray diagram illustrating astigmatic aberration showing location of the two line images produced from a single point source

The simplest example of astigmatism is to consider the situation of regular astigmatism where the surface of the lens has been distorted to produce two axes of dioptric power that are at 90° to each other. (This is the familiar situation for an astigmatic lens used for correction of an ocular refractive error). The consequence of the two dioptric powers is that rays of light from a point source on the optical axis will be refracted by differing amounts depending on where they pass through the lens surface, and which results in two principal focal points, one for each dioptric surface. The image

produced at each focal point will take the form of a line of light, rather than a single spot (Figure 4-5).

The optical effect on the image depends on where the image plane of the system is placed in relation to each of the focal lines. If it is placed at one of the focal lines, the image will be in focus in one meridian and out of focus in the other. The best overall focus (least blur) is obtained halfway between the two focal lines, where a blurred image of the object will be produced. This point is often called the circle of least confusion.

When the astigmatism is irregular, multiple focal lines are produced from the differing lens curvatures distributed over the surface of the lens. Identifying individual focal lines becomes impossible, and the effect is to blur the image.

For an off-axis object, the effect of astigmatism is modified in a similar way that spherical aberration is affected to produce coma. In this situation, the off-axis effect produces oblique astigmatism.

4.2.1.5 Field curvature

Field curvature is an effect that relates to the shape of the image plane produced by the lens. The image plane has a curvature and is not flat (Figure 4-6). It is due to the different refractive effect on rays of light that pass through the periphery of a lens – an effect noted previously in spherical aberration and coma. It affects the whole image such that the center of the image is in focus, while the periphery is not. Modern lens design can control the majority of this aberration, but it cannot remove it entirely. Using an aperture stop to remove the peripheral rays will help to reduce the effect, but at the cost of image intensity.

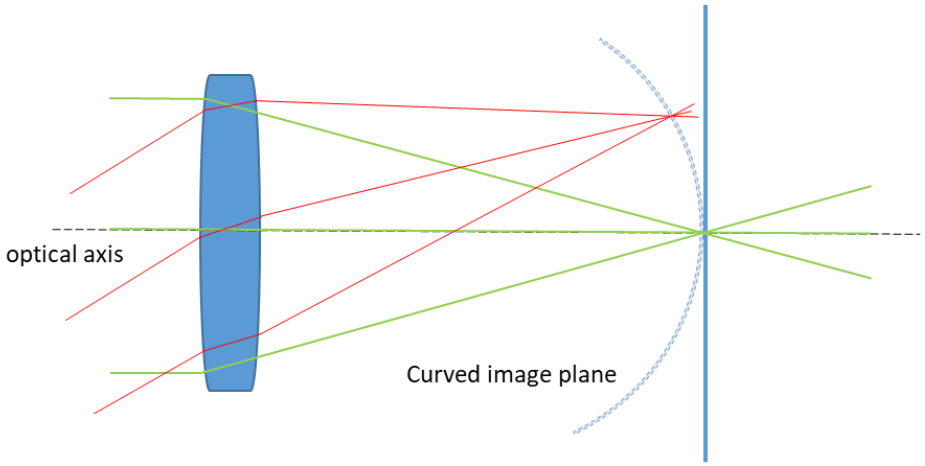
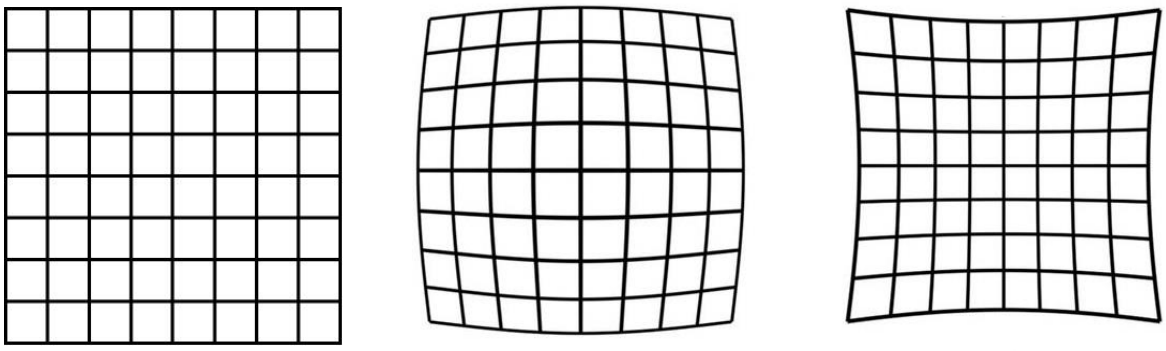


Figure 4-6: Ray diagram illustrating field curvature aberration

4.2.1.6 Distortion

Distortion is a further example of the effect of light rays passing through the periphery of the lens. The effect is more obvious for a wide object scene, and the image produced is distorted to produce a barrel distortion, a pin-cushion distortion, or a moustache distortion (which is a combination of barrel and pin-cushion). The distortion can occur in any meridian. The visual effect is to cause a bending in the image, such that straight lines become bowed. Distortion is a greater problem when using short focal lengths.



a *b* *c*

Figure 4-7: Distortion in image produced for a square object (a) barrel (b) pincushion

4.2.2 Chromatic aberration

Visual and IR light can be presented in a single (monochromatic) wavelength or in a broad spectrum of wavelengths (chromatic). As discussed in Section 2.1, EMR wavelength has a controlling effect on how the EMR interacts with a material. This effect can be demonstrated using

white visible light, which contains all of the wavelengths of visible light. When white light is passed through a prism, the individual wavelengths of light are separated out, and the effect can be visibly observed as the colored light spectrum. The differing wavelengths of light have interacted with the prism material in slightly different ways, with the consequence that the speed of each wavelength as it travels through the material is altered and the waves become separated and ordered according to their wavelength. When the light leaves the prism, the effect on the wavelengths is retained and is seen visually as a colored spectrum. Since the effect is related to the optical density of the material, the magnitude of the chromatic aberration is controlled by the refractive index of the lens, the dioptric power of the lens, and the range of wavelengths of light used.

The chromatic aberrations describe the effect that this EMR wavelength differentiation has on an image from a light source possessing multiple wavelengths. There are two sub-types of chromatic aberration.

4.2.2.1 Longitudinal chromatic aberration

This occurs when multi-wavelength light passes through a lens – each wavelength is focused at a separate focal point along the optical axis and not at a single locus. This type of aberration produces a colored fringe around the subject that spreads from the center of the object to the edge of the frame (Figure 4-8).

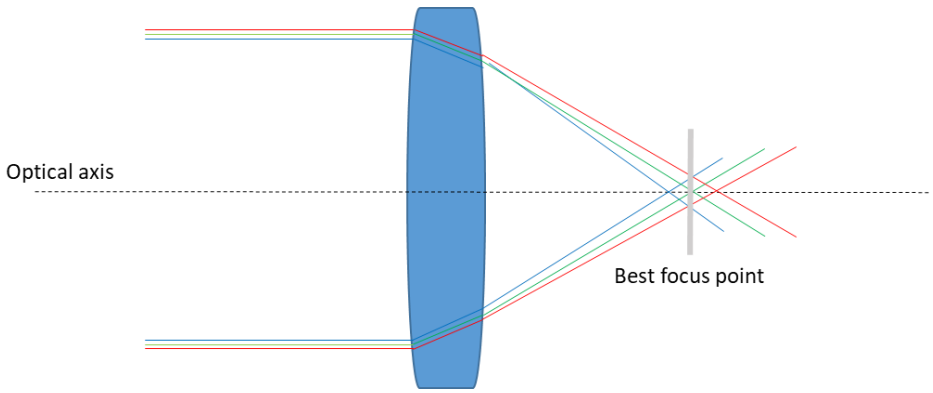


Figure 4-8: Ray diagram illustrating longitudinal chromatic aberration

4.2.2.2 *Lateral chromatic aberration*

Lateral chromatic aberration arises from an off-axis object. The different wavelengths are focused in the same focal plane, but are displaced laterally from each other. The produces a colored fringe that is visible at the edges of the frame spreading towards the center of the object (Figure 4-9).

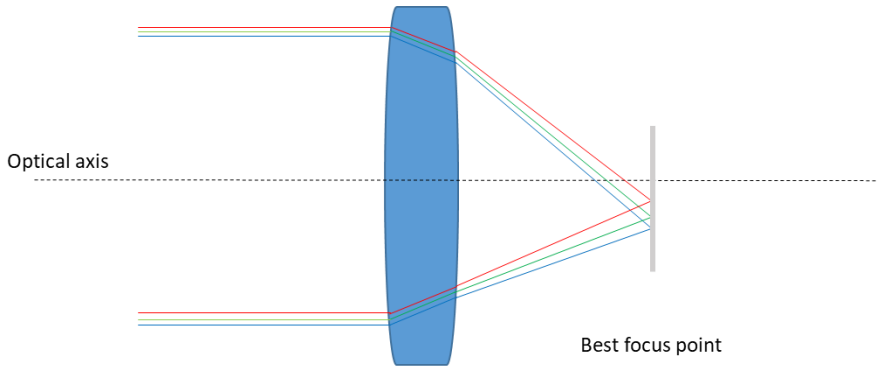


Figure 4-9: Ray diagram illustrating lateral chromatic aberration

Chromatic aberration can be corrected by using an aperture stop to block the peripheral rays passing through the lens or with achromatic doublets, triplets, etc. With the technological advances in camera lens manufacture, most modern cameras are not significantly affected by chromatic aberration.

4.2.3 Control of aberrations

Aberrations are a significant issue for visible light and IR cameras, especially for cameras that use small objective lens size and/or short focal lengths. The ultimate effect is that they cause light to converge at different places on the camera sensor. The optical quality of the optical components, the efficiency of the lens design, the manufacturing quality, and the relative positioning of the optical components within the system all affect the size and effect of each aberration. However, with recent technological advances, visible cameras have very low amounts of chromatic aberration. IR cameras are affected by mono-chromatic aberrations. IR cameras with close-up lenses are more sensitive to focus error aberration as they have a small depth of view. The effect of other aberrations can be removed or minimized from the cameras.

The main way of controlling aberrations for both visible and IR cameras is to use an aperture stop. The aperture stop is effective by blocking (stopping) the light rays from the object that pass through the periphery of the lens contributing to the image formation. Distortion can be modified by changing the position of the lens elements.

However, while these methods can help to control the effect from aberrations on an image, it cannot remove them all. It is therefore important to identify the remaining optical effects from these aberrations in an optical system so that they can be removed from the image, and thus produce a true image of the object.

4.2.4 Intrinsic camera sensor characteristics

Imaging electronics and optics play important roles in the quality of imaging system output. All of the camera components, including camera sensor, capture board, cables, and software, have an effect on the overall system performance. The heart of each camera is the sensor. Modern visible camera sensors contain millions of photodetector sites called pixels located on a charge-coupled device (CCD), and so an image is divided into small, discrete pixels each of which represents one small part of the object. The pixel might be a photodiode or a photo-capacitor, which generates an electrical signal with the charge varying according to the amount of light that shines on the pixel (see Section 2.4). The information from the pixels is collected and organized by the CCD for onward transmission and is then transferred to a monitor to be displayed or transferred to a computer for further processing.

Individual pixels are usually square in shape in digital cameras, with a size of 3-10 μm. The pixels can be arranged on the sensor in any formation, but are most commonly organised into a rectangle. The individual pixel size, and the density of pixel distribution across the sensor, plays an important role in the sensitivity and resolution of the sensor. More pixels/unit area will produce a greater image resolution of the object since smaller details of the object can be detected by individual pixels. Alternatively, sensitivity can be increased by electronically combining the output from several pixels to enable less bright objects to be detected. However, this occurs at the price of reduced resolution.

Pixel size will also influence sensitivity. A larger pixel will gather light from a larger part of the object, increasing sensitivity, but reducing resolution. A large pixel has a high charge saturation capacity and a high signal-to-noise ratio. On the other hand, a smaller pixel will improve resolution, but reduce sensitivity. Small pixels are a fairly easy way to achieve high resolution (more details) for a fixed sensor size and magnification. A simple measure of sensor resolution is the number of pixels per millimetre.

The size of the sensor active area (pixel array on the sensor) determines the camera's field of view (FOV). By assuming a fixed focal length (which depends on the optical components of the camera), a greater FOV will be produced by a larger sensor. There are several standard area-scan sensor sizes: ¼", 1/3", ½", 1/1.8", 2/3", 1" and 1.2", with larger available.

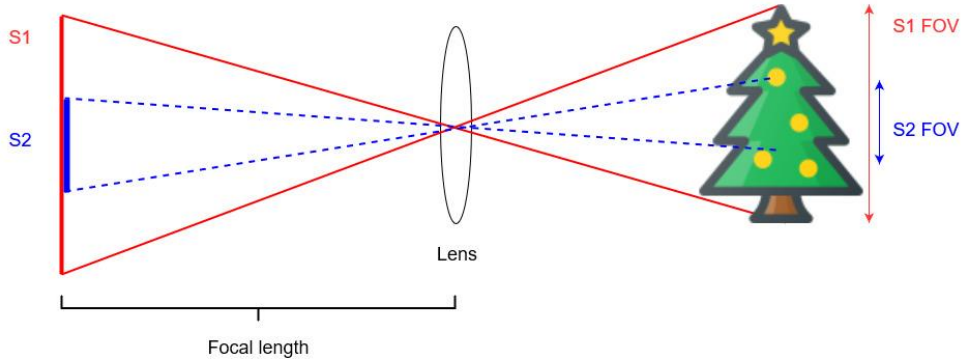


Figure 4-10: Effect of sensor size and focal point on camera field of view

Although manufactured to high quality standards, the CCD sensors used in visible and thermal cameras can have some irregularity in their pixel arrangement, which may produce some localised distortion in the image. It might also be expected that two cameras manufactured with the same sensor by the same company should function in a similar manner, but small differences in the sensor manufacture and interface design can have an impact on sensor, and thus camera, performance. Any effects from these manufacturing defects must be identified and removed from the images.

4.3 Pinhole camera

Image formation is the process of turning a three-dimensional (3D) object into a two-dimensional (2D) image projection. The simplest method to produce this effect is to use a pinhole camera

(Figure 4-11). In this model, a small, single aperture is used to produce an image without the action of a focusing lens. By removing the lens, there is no lens-related distortion in the image.

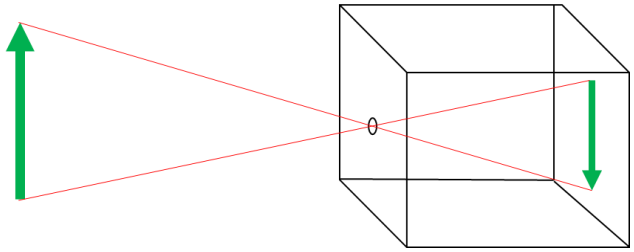


Figure 4-11: Pinhole camera model

In the pinhole camera model (Figure 4-12), the center of projection is placed at the center of the pinhole and is known as the *camera center (C)*. The line passing perpendicularly through the camera center to the image plane is the *principal axis (Z)*, and the point where the axis intersects with the image plane is the *principal point (p)*. The distance between the camera center and the principal point is the *focal length (f)*. *X* represents the location of the object (the *world point*) with coordinates $X = (X, Y, Z)^T$. Light from the object passes through *C*. This projects X onto the image plane to form the *image point* with coordinates $x=(x,y)^T$.

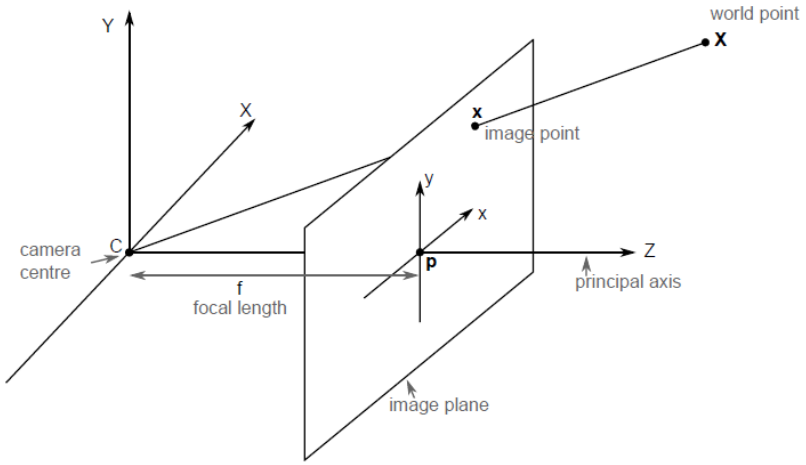


Figure 4-12: Geometric description of the pinhole camera model[73]*(Reproduced from: Hartley R, Zisserman A. Multiple view geometry in computer vision 2nd ed., 4th print. 2006)*

The mathematical relationship between the world point coordinates and image point coordinates is described in Equation 4-1, where *f* is the focal length.

$$(X, Y, Z)^T \rightarrow (fX/Z, fY/Z)^T$$ *Equation 4-1*

By using homogeneous coordinates, Equation 4-1can be written as Equation 4-2:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$ *Equation 4-2*

Equation 4-2 describes the pinhole camera model when the world points are expressed in the pinhole camera coordinate system. However, points are often described in a different coordinate system in the real world and are affected by rotation and translation, which are artefacts produced by the location and position of the camera in relation to the object (Figure 4-13: Demonstrating camera and world coordinate system relationship).

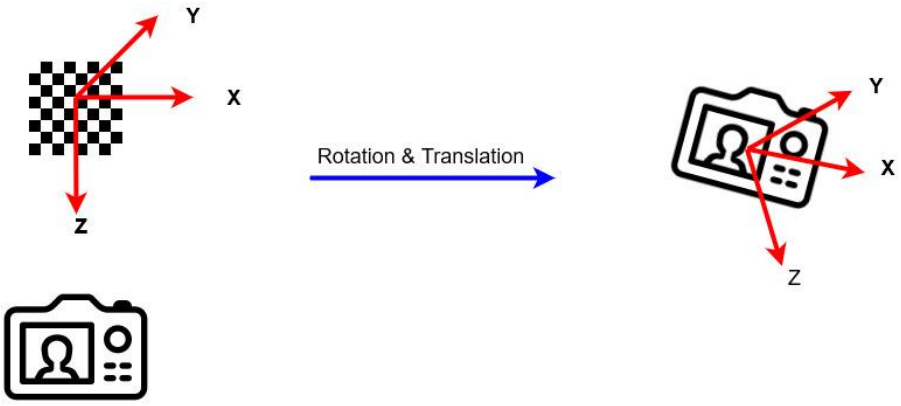


Figure 4-13: Demonstrating camera and world coordinate system relationship

Rotation controls how the camera, is aligned to the x, y, and z axes. Translation refers to where an object is in space and whether the camera and object are in alignment along an axis. Both the rotation and translation parameters are using to transform 3D world coordinates to 3D camera coordinates. Rotation can be considered as a 3x3 matrix, with each column describing the rotation along an axis. Since the whole coordinate system rotates around the x, y, and z axes, this gives three columns of 3-vectors that results in a 3x3 matrix that is used to rotate the whole system based on another coordinate system. The translation vector maps the center of camera onto the center of

the world coordinate system. The translation is a 3-vector number that describes the difference between two coordinates along the x, y, and z axes. Translation and rotation change during photography with each new object or camera set-up and must be calculated for each image.

Consequently, the coordinates from a 3D world point need to be rotated and translated into the camera coordinate system. The coordinate systems are rotated by a 3×3 matrix and translated by a 3-vector.

Therefore, Equation 4-2 can be written as Equation 4-3:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix} \begin{bmatrix} r1 & r2 & r3 & t1 \\ r4 & r5 & r6 & t2 \\ r7 & r8 & r9 & t3 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad \textit{Equation 4-3}$$

where $r1..r9$ represents the rotation matrix and $t1..t3$ represents the translation vector.

4.4 Camera manufacturer parameters

Equation 4-3 describes an idealized projection of the points to the image plane. However, in reality some physical imperfection or manufacturer inaccuracy of some parameters of the camera system will affect the image projection.

4.4.1 6.3.2.1 Principal point

In the pinhole camera model, the principal point is located at the image plane. However, in a camera the principal point often does not lie exactly at the image plane. Therefore, the parameters p_x and p_y are used to express the coordinates of the principal point on the image. These parameters, in pixel dimension, are $x_0 = m_x p_x$ and $y_0 = m_y p_y$ in the x and y directions, respectively, where m_x and m_y = the number of pixels per unit.

4.4.2 Non-square sensor

For the pinhole camera, it is assumed that the number of the pixels in the image plane in both the x and y directions is equal in scale. However, for most CCD cameras, the pixels are arranged in a

rectangle. If it is assumed that a rectangular sensor is square, there will be a scaling error, producing distortion in the image. If the pixel coordinates are measured in pixels, the focal length should be multiplied by scaling factors m_x and m_y, to be presented in pixel dimension at x and y direction. The focal length is generally expressed as: $\alpha_x = fm_x$, $\alpha_y = fm_y$ in pixel dimension.

4.4.3 Skew

The arrangement of the pixels on a CCD sensor may not be perfectly square, resulting in a small distortion in the X or Y directions. There is an expected number of pixels per unit length in each direction on the CCD sensor. Where the distribution of pixels is not regular, the error can be used to produce the skew coefficient *s*. Usually, where there is no error, *s* is zero.

Equation 4-3 can be modified to include *s* as:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} \begin{bmatrix} r1 & r2 & r3 & t1 \\ r4 & r5 & r6 & t2 \\ r7 & r8 & r9 & t3 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad \textit{Equation 4-4}$$

In Equation 4-4, the 3×3 matrix describes the intrinsic parameters of the camera that remain unchanged during photography and so can be calculated.

To take these into account, Equation 4-4 can be simplified as Equation 4-5:

$$P = K[R|t] \qquad \textit{Equation 4-5}$$

where *P* is the camera matrix, *K* is the intrinsic parameters, *R* is the rotation matrix, and *t* is the translation vector.

4.5 Inherent camera aberrations

A general description of optical aberrations that can affect an optical system was given in Section 4.2. In considering the development of the camera matrix, only non-optical factors have been

considered by assuming that there is no lens attached to the camera. For this system, two kinds of distortion will occur: radial and tangential.

4.5.1 Radial distortion

Optical or radial distortion is the most common type of optical artefact that appears in an image. It is produced by the variation in magnification from the center to the edge of the lens. The amount of distortion is higher when the lens has poor quality of design and material or has a short focal length. Radial distortion is noticeable when the object distance to principal point is large and results in straight lines in the object appearing bent in the image. The radial distortion can be calculated using Equation 4-6:

$$x_d = x(1 + k_1 * r^2 + k_2 * r^4 + k_3 * r^6)$$
$$y_d = y(1 + k_1 * r^2 + k_2 * r^4 + k_3 * r^6)$$

Equation 4-6

where x_d and y_d are the coordinates of the distorted points, x and y are the coordinates of the undistorted pixel location in a normalized image.

The normalized image coordinates are calculated from the pixel coordinates by translating their location with respect to the optical center and dividing by the focal length in pixels. Thus, x and y are dimensionless. k_1..k_3 are lens radial distortion coefficients, and $r^2=x^2+y^2$.

4.5.2 Tangential distortion

Tangential distortion occurs when the lens and the image plane are not parallel. It occurs during lens assembly. Equation 4-7 describes the distorted pixel coordinates.

$$x_d = x + (2 * p_1 * x * y + p_2 * (r^2 + 2 * x^2))$$
$$y_d = y + (p_1 * (r^2 + 2 * y^2) + 2 * p_2 * x * y)$$

Equation 4-7

where x_d and y_d are distorted point coordinates, x and y are undistorted pixel location on a normalized image.

The normalized image coordinates are calculated from the pixel coordinates by translating their location with respect to the optical center and dividing by the focal length in pixels. Thus, x and y are dimensionless, p_1 and p_2 are tangential coefficients of the lens, and $r^2=x^2+y^2$.

4.6 Camera matrix estimation

Camera calibration, also known as camera re-sectioning, is the process of estimating the camera matrix. The camera matrix is the equation that defines the intrinsic and extrinsic artefacts for a camera system that incorporates an optical component. Equation 4-4 has 11 degrees of freedom and so to find the parameters, 11 equations should be solved. To solve the matrix, the corresponding points of a 3D object should be located in a 2D image. After point correspondence, each of them gives us two degrees of freedom. Hence, at least 6 correspondences are required to solve the camera matrix. To estimate the camera matrix parameters, the image produced by the camera of a known object should be analysed.

4.6.1 Calibration checkerboard

To determine the camera matrix parameters and so calibrate a camera system, several images produced by the system of an object of known dimensions are needed. In the photography literature, a variety of different test object patterns used to measure the camera matrix parameters have been reported. The most common object pattern is a checkerboard. A series of photos must be taken that capture the object at different depths and viewpoint angles from the camera objective lens to reveal the effect of rotation and translation on the image. Usually, 10 to 20 different views provide sufficient information for camera calibration. The full image of the checkerboard should be in focus to reduce the noise rate in the image and so reduce the potential size of errors in the camera matrix estimation. Since the data cannot be perfect and without noise, a large number of data points must be identifiable on the object to improve the signal/noise ratio on the image, and thus achieve a robust solution. A checkerboard is an ideal object since it is composed of a series of equally sized squares of alternating black and white color (Figure 4-14: Checkerboard pattern).

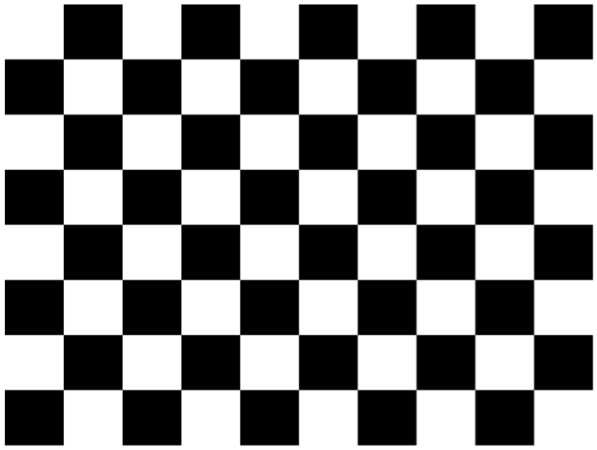


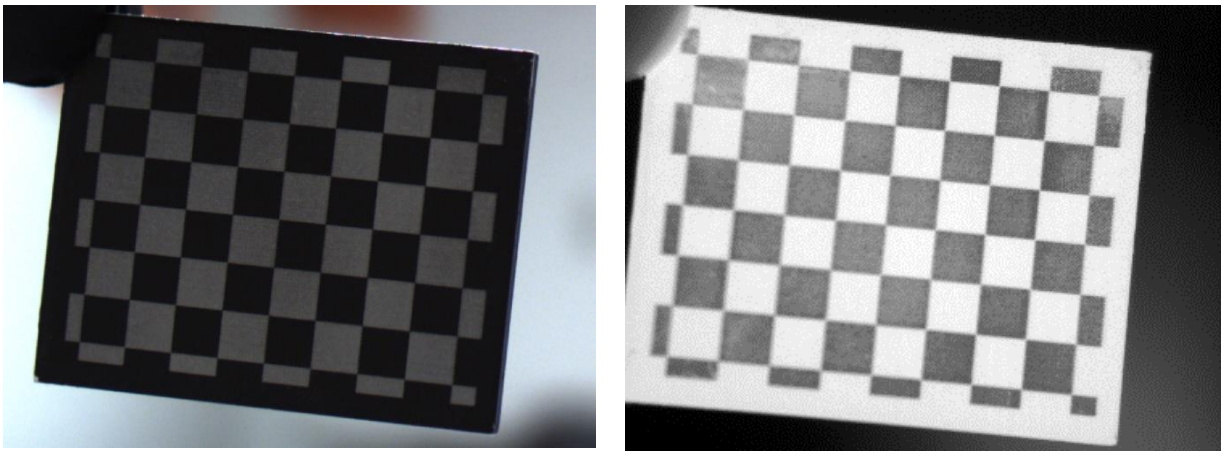
Figure 4-14: Checkerboard pattern

For this project, finding a checkerboard that could be detectable by both the visible and IR cameras was a challenge. To be easily detectable to the thermal camera, the white areas of the checkerboard need to have a low emissivity, while the black areas need a high emissivity. However, while a black and white checkerboard printed on card will be suitable for the visible camera, the difference in emissivity between the black and white areas of the card will not produce a strong thermal image.

In reviewing the literature, different methods for thermal camera calibration have been described. Saponaro et al.[74] (2015) took a flat, glazed, ceramic tile, which has the capacity to maintain an even temperature across its surface, and attached a thin sheet of paper in front surface with a 10x7 square, black and white checkerboard pattern printed on it. The front surface of the ceramic tile was heated using a lamp to produce a thermal source that could heat the paper. The checkerboard pattern on the surface of the paper created a strong visible image, but the emissivity difference between the black and white squares was insufficient to produce a strong thermal image. An aggressive corner detection algorithm was needed to find the corners with minimum error. Ursine et al.[75] (2012) used a copper plate painted with black squares to produce a checkerboard. This solution worked well in an outdoor environment, but did not work well indoors because of specular reflection. Peric et al.[76] (2014) and Campo et al.[77] (2012) used aluminium foils to form the checkerboard, but they also had problems with specular reflection. Harguess et al.[78] (2014) used a Dibond® board (a sandwich material covered by two sheets of aluminium) on which was painted

black and white circles. The circle targets provided a good contrast on sunny days, but had poor contrast in other conditions. St-Laurent et al.[79] (2017) demonstrated that aluminium is the best foundation material since it possesses a low emissivity. They combined painting of the aluminium surface a black matte color with sand-blasted areas on the plain aluminium sheet to produce thermal images with very good visual and thermal contrast.

Based on this previous literature, a checkerboard design was produced that could be printed onto an aluminium sheet. The aluminium sheet was sand-blasted before printing to increase the emissivity. Figure 4-15 shows the checkerboard target as (a) a visible image and (b) a thermal image.



a *b*

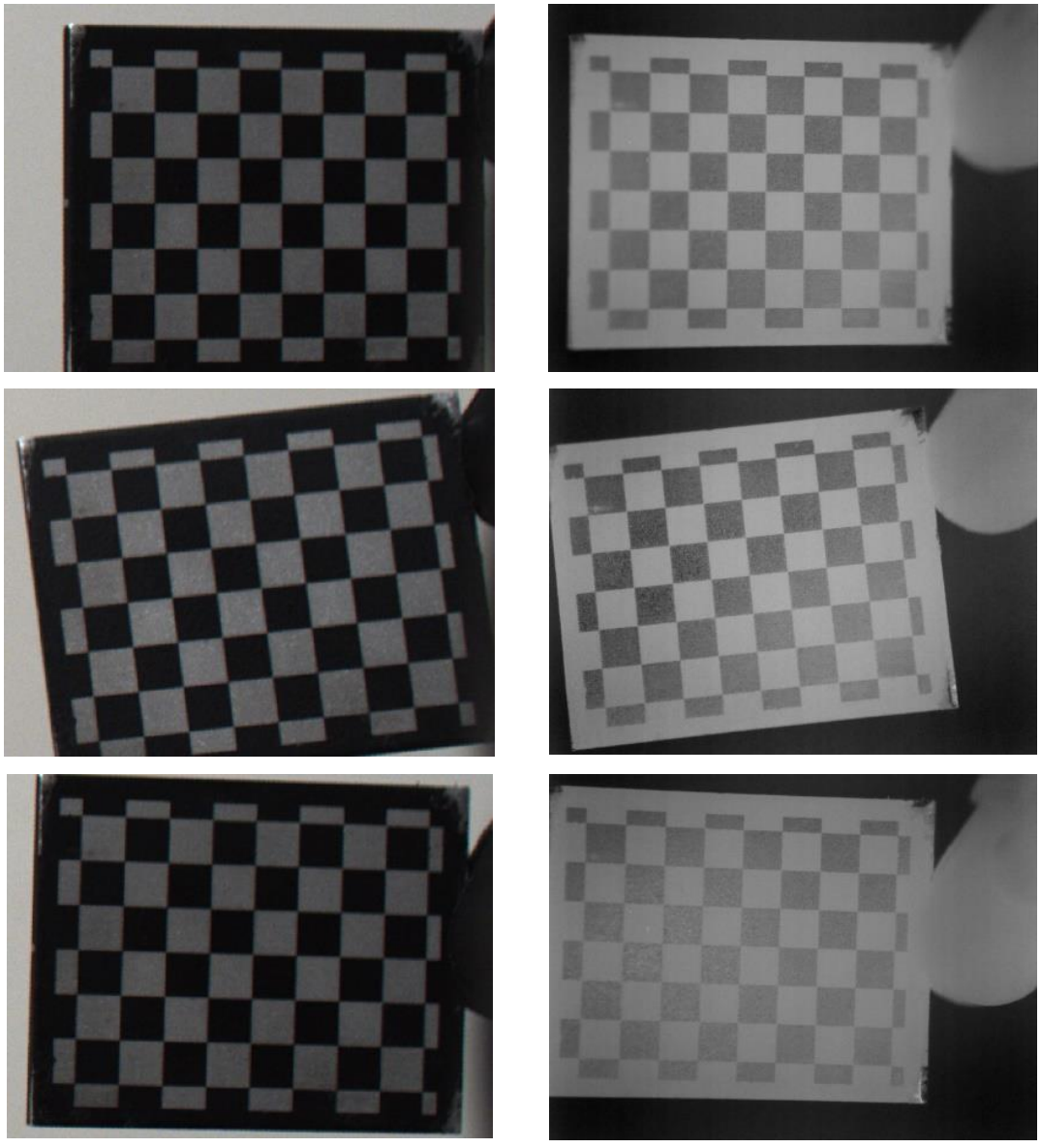
Figure 4-15: Checkerboard printed on an aluminium sheet showing its appearance as: a) visible image, b) thermal image.

Comparing the two images, the light grey squares on the thermal image represent the black squares of the board, thus the thermal camera produces a negative image of the object. For this project, since a close-up lens was attached to the thermal camera with a very small field of view (around 4 × 6 cm), a 5 × 4 cm checkerboard plate, with individual square size of 5 mm, was used for the camera calibration.

4.6.2 Image Gathering Method

After preparing the checkerboard calibration object, a series of images were captured using the dual camera system. The cameras were locked in position and the checkerboard was manually held directly in front of the two cameras in a position that allowed a full image of the object to be produced by both cameras. (It was not necessary to position the checkerboard in a holder at known offsets/angles since the angle and distance of the object are not required input parameters). Each

camera was separately focused on the checkerboard. A series of 30 images was then captured with the checkerboard located in different on-axis and off-axis positions, and object depth from the cameras. During photography, both the focal length and relative position of both cameras were not changed. As part of the standard camera set-up for the system, the resolution of the images was adjusted to a pixel array of 640 x 480 pixels and illumination was kept consistent. Figure 4-16 shows a series of paired sample images produced by the two cameras from the checkerboard pattern in three different object positions.

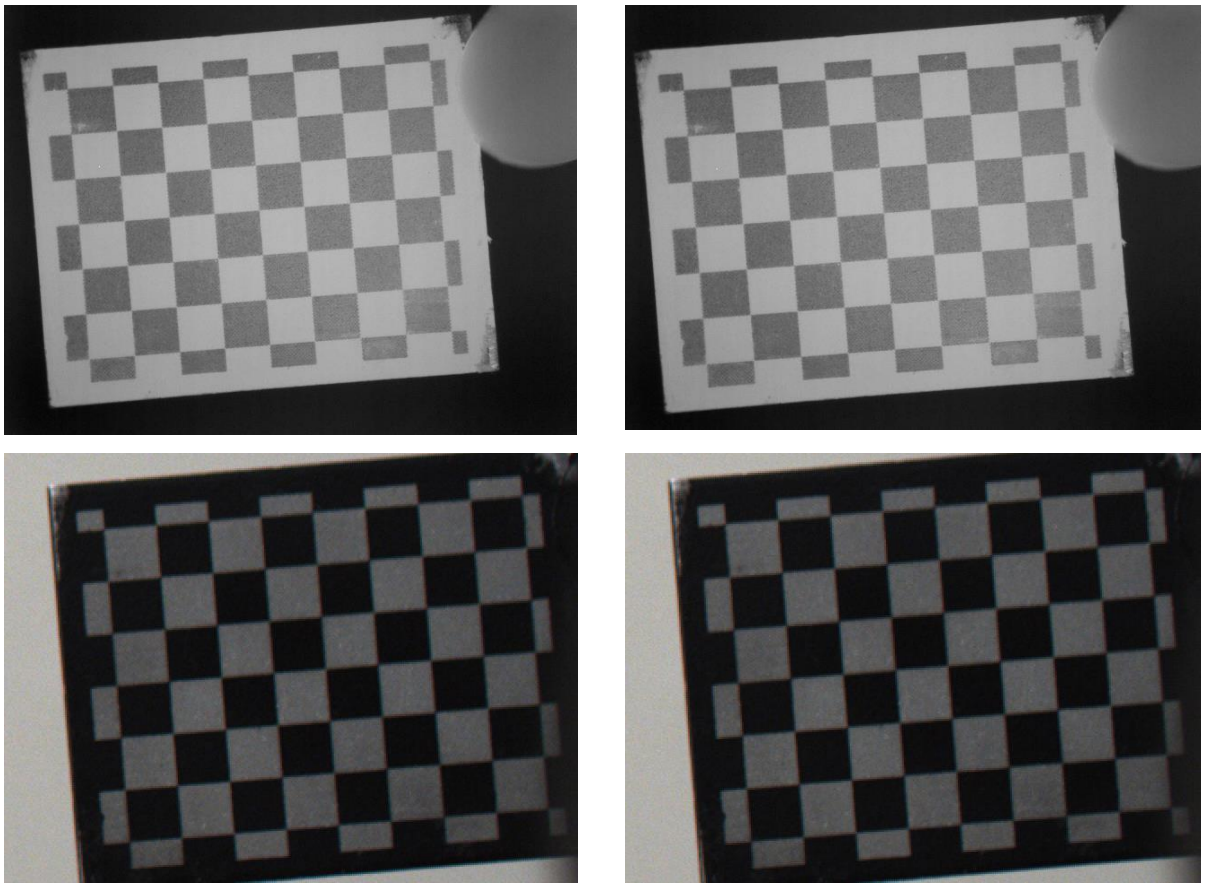


a *b*

Figure 4-16: Corresponding calibration images of the checkerboard test object in three object positions: a) visible camera, b) thermal camera.

4.6.3 Checkerboard image normalization

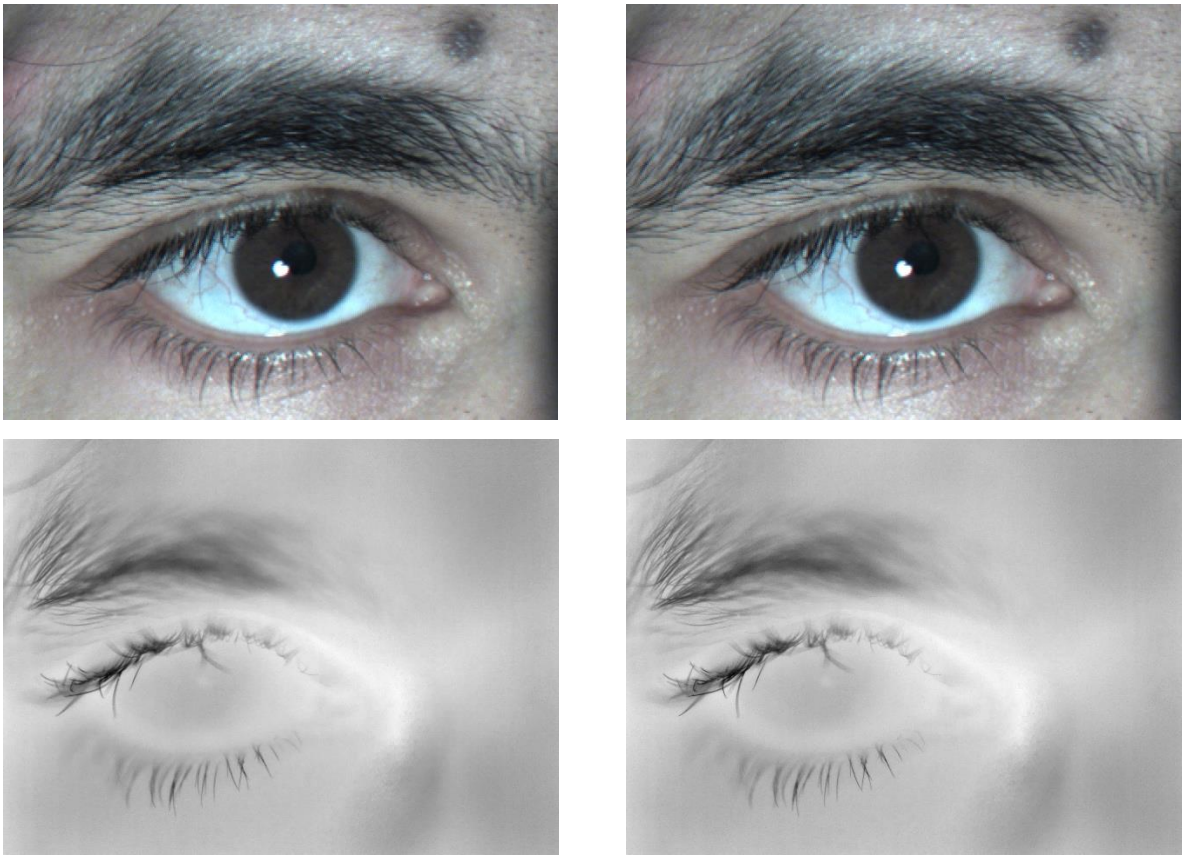
After camera parameter matrix calculation, the image distortion can be removed from the images. When the distortion is removed from the images, each pixel is in its correct size and location (Figure 4-17). As previously mentioned, the intrinsic parameters do not change during the photography and should be calculated only once. However, the extrinsic parameters of rotation and translation change when the camera position changes. To address this problem, the relative position of the cameras was fixed during camera installation and did not change for the remainder of the project. Hence, the calculated translation and rotation parameters can be used in all future steps.



a *b*

Figure 4-17: Demonstration of the effect of distortion removal on the checkerboard image from the visible camera (upper) and thermal camera (lower): a) distorted images, b) undistorted images.

Figure 4-18 shows a sample of two video frames, one for the visible camera and one for the thermal camera, before and after normalization. Since the amount of lens distortion is very small, the image appearance does not change sufficiently that it can be detected by the human eye.



a *b*

Figure 4-18: Demonstration of the effect of distortion removal on visible (upper) and thermal (lower) frames taken using the dual camera system: a) before normalization, b) after normalization.

4.7 Implementation details for image normalization

The camera calibration algorithm was implemented using the same programming language used in the calibration process. A stereo camera calibration algorithm was used in for this system.

Camera calibration was completed using the Python V 3.7 programming language (Guido van Rossum, Centrum Wiskunde & Informatica (CWI) in the Netherlands, 1980) and the OpenCV library (Gary Bradsky, Open Computer Vision, Intel, 1999). Python is an open source, high-level, object-oriented programming language with built-in high-level objects. The built-in objects make it easy to use and allows rapid software development. It is also easy to debug. Python programming enables the programmers to express their ideas with fewer lines of codes with high readability. OpenCV-python is a python library that can be used to solve computer vision problems.

Using Python and the OpenCV library, code was developed that used the checkerboard images as an input to calculate the camera parameters and matrix. Once the camera parameters were collected and stored, the OpenCV function for image distortion removal (cv.undistort) was used to undistort images. The function uses the checkerboard image and camera parameters as an input and calculates the distortion parameters using equations Equation 4-6 and Equation 4-7. Then, the function removes the distortion from the input image and gives an undistorted image as an output. Since the only parameter considered for normalization is distortion removal from the images, image normalization is completed after this step and the resultant image is available for the next image processing step.

4.8 Testing for error

After camera calibration using the checkerboard patterns, the re-projection error was calculated. Re-projection error provides a qualitative measure of accuracy. It is the distance between detected key points and corresponding points in the world coordinate in the same image. The re-projection error for the visible image was calculated as 0.38 pixels which is equal to 0.38 * 0.09 mm, and the error for the thermal camera was calculated as 0.52 pixels, which is equal to 0.52 * 0.09 mm.

4.9 Next steps

After normalizing each frame in each video sequence, the next step was to register each paired visible and thermal video frame. In the next chapter, the image registration methods are discussed.

Chapter 5: Image registration

One key objective of the image calibration process was to produce an image from each camera that was unaffected by the respective intrinsic and extrinsic errors of each camera. Image calibration has resulted in two images, one from each camera, taken of the same object at the same timestamp that are 480 x 640 pixels in dimension. Each pixel on the visible camera image now has a corresponding pixel in the thermal camera image, but because of the misalignment of the two cameras, these matching pixels do not represent corresponding points in each image. There are translational and rotational errors between the two images. In effect, the two images do not overlap precisely.

Processing the two images so that the corresponding pixels in each timestamped set of images also represent corresponding image points is crucial for this system. In later steps of system development, the visible camera image will be used to localize the cornea, and then the corneal localization from the visible camera will be used to identify the corneal area in the thermal image. To do that, the coordinates of the iris will be identified in the visible image and then those coordinates will be used as the location of the cornea in the thermal image. For this to work, the images from each camera must be mapped on top of each other.

In this chapter the image registration techniques are described. Section 5.1 describes the camera alignment problem. Section 5.2 explains the concept of image registration, image registration techniques, and the evaluation method for image registration. Thermal-visible image registration is described in Section 5.3. Video registration and stabilization algorithm are described in Section 5.4.

5.1 Camera alignment

When the cameras were installed on the lens mounting (Section 3.2), the cameras were positioned in such a way that each camera was focused on the same general area of the eye and adnexa, so that they both produced an image of the same area. However, adjusting each camera's physical position to ensure that the image precisely overlap from the two cameras is a difficult task. Even if such alignment could be obtained easily, the physical separation of the two cameras produces a

relative distortion between the two images. The result of these optical and physical errors is that a common location on the eye is not located at corresponding image points in the two images. The key factors involved in producing these errors are: pixel size, lens characteristics, field of view, distance, and angle of photography. The resulting effect is that the same location on the eye or adnexa does not have the same geometric coordinates in both images.

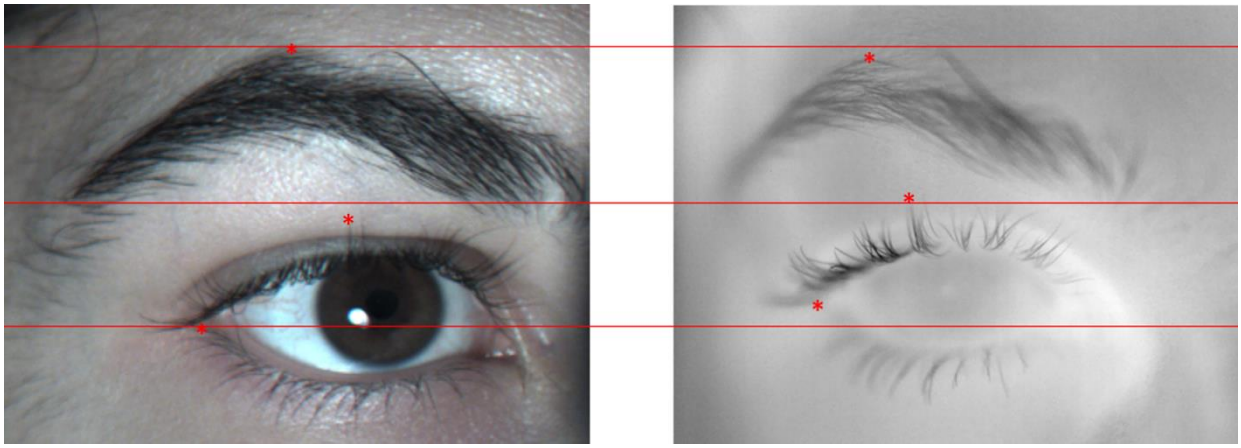


Figure 5-1: An example of the misalignment of sample eye locations on two corresponding images taken by the dual camera system (left: visible camera; right: thermal camera).

Figure 5-1 shows an example of two corresponding images taken by the visible and thermal cameras for the same area of the eye and adnexa at the same time point. Examining the images reveals that the eye size is different in the images and that three test eye locations (red dots) are not located at corresponding points on the two images. Three horizontal lines have been added to the images to help with this comparison. In the visible image (left), each line has been drawn through a physical feature of the eye and adnexa. By following the lines across the two images, it can be observed that the lines do not intersect at the same physical feature in both images, i.e., they are not mapped on top of each other.

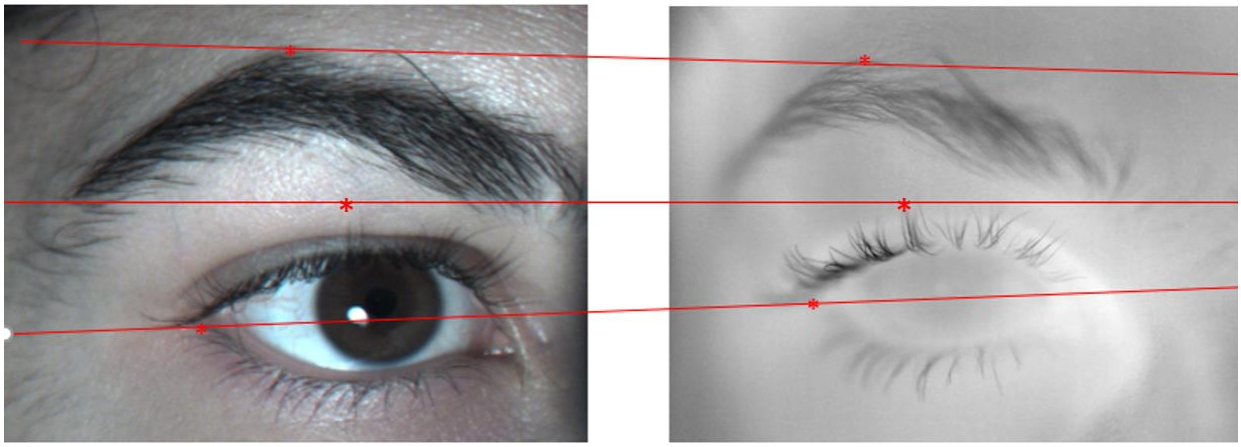


Figure 5-2: An example of the error produced by point correspondence of the images taken by the visible (left) and thermal (right) cameras.

Figure 5-2 gives an example of the relative image distortion that must be corrected. The lines have been drawn to intersect with the same physical feature of the eye, and it can be seen that the lines are no longer parallel to each other. The process of image registration will correct this distortion.

5.2 Image registration

Image registration is a technique for overlaying two images taken at different times or from different viewpoints with the same camera, or at different times or different viewpoints with different sensors. When the images are taken from different sensors, the aim of the registration is to overlap the corresponding image pixels from both images taken by each camera for the same scene. The processing of the image re-aligns the images in geometric space using a transformation matrix. The transformation matrix is a mathematical formula that describes the error between the two images, and which can be used for image transformation. The pixel coordinates on one image are transformed using the matrix to form a transformed image in which the common location on the object is now imaged at corresponding pixel points on the two images. This process does not change the pixel values, but the geometry of the pixels. For this transformation, one image is assigned as the reference image – *the fixed image* – to which the other image will be aligned – *the moving image*. For this system, the visible camera image was assigned as the fixed image and the thermal camera image was assigned as the moving image.

Generally, image registration methods consist of four steps:

1- Feature detection

Salient image features are identified and selected. Each feature should have a distinctive characteristic, such as an edge, corner, or line intersection, that can be easily identified in both images. Figure 5-3 shows an example of feature selection in this system.

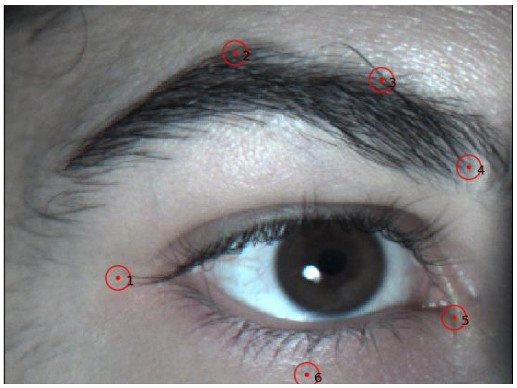


Figure 5-3: Selected feature on the visible image

2- Feature matching

The next step is to find correspondence between the selected features in both images. The image features can be detected either automatically using a feature detection algorithm, or manually selected by the user. For automatic feature selection, a feature matching algorithm is used to identify and locate corresponding features in the two images. For manual point selection, the corresponding points are selected in each image by the user sequentially clicking on each item – the feature is marked in the first image and then the corresponding point marked in the second image. Therefore, the feature matching algorithm is not required. Figure 5-4 shows an example of the selected feature in the thermal and visible images. The matched features are the same, and there is the same number of matched features, in the thermal and visible images.

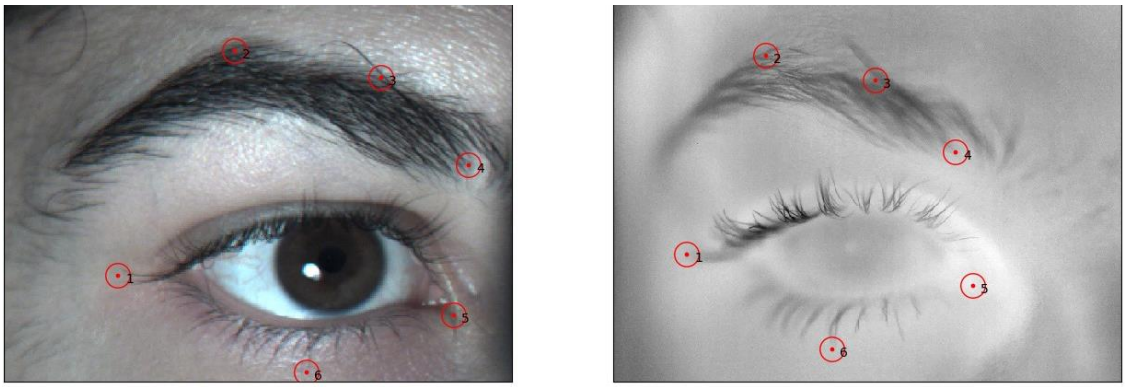
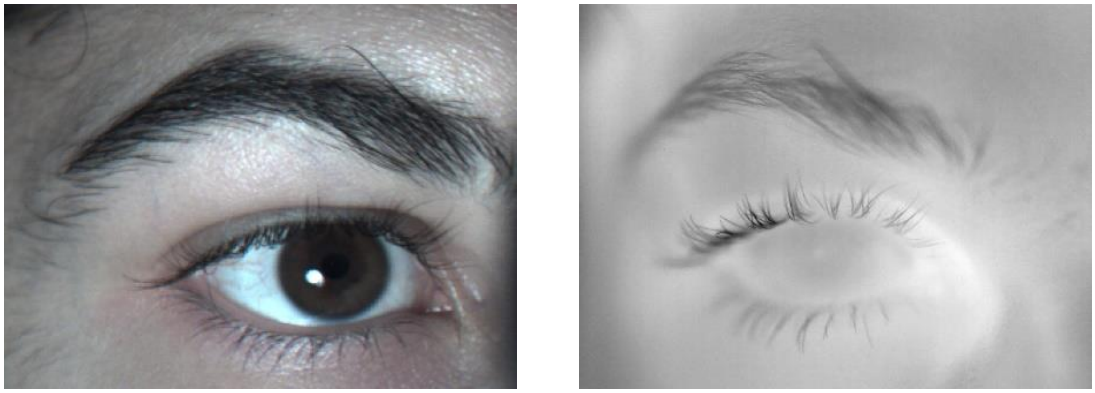


Figure 5-4: Matched features on the corresponding frames

3- Transformation estimation

Now that the two sets of corresponding points have been identified on each image, the geometric transformation between the images can be performed. Since only a small number of points have been selected, the resulting transformation is estimated for all of the other image locations. Figure 5-5 shows an example of the original images before transformation.



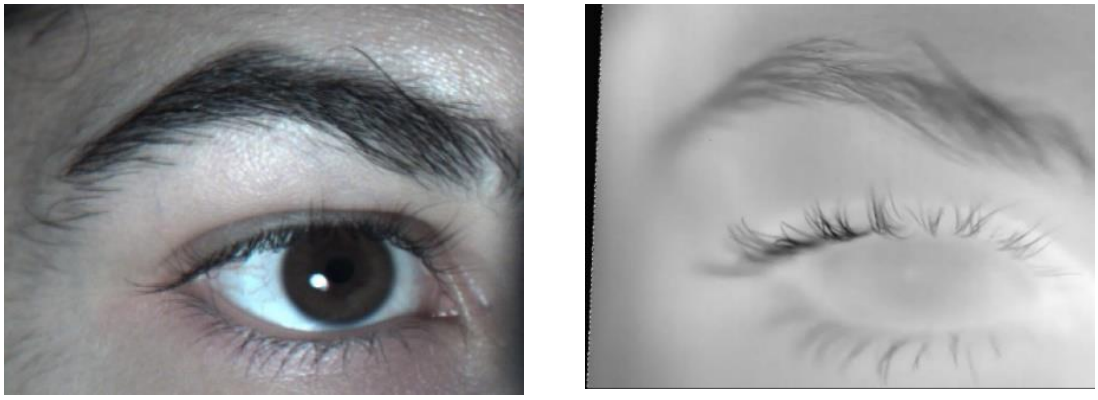
a *b*

Figure 5-5: Original visible(a) and thermal(b) images before transformation

4- Image wrapping

The last step is image resampling or transformation by applying the estimated transformation matrix. In this step, the transformation matrix gained from the previous step is applied to map the moving (thermal) image onto the fixed (visible) image. Figure 5-6 shows an example of transformation applied on the sample image. The estimated

transformation function applied on the thermal image (Figure 5-6 b) to be registered on the visible image (Figure 5-6 a).



a *b*

Figure 5-6: Wrapped thermal image(b) to be mapped on the visible image(a)

The final quality and accuracy of the image registration method is related to a careful application of these separate steps.

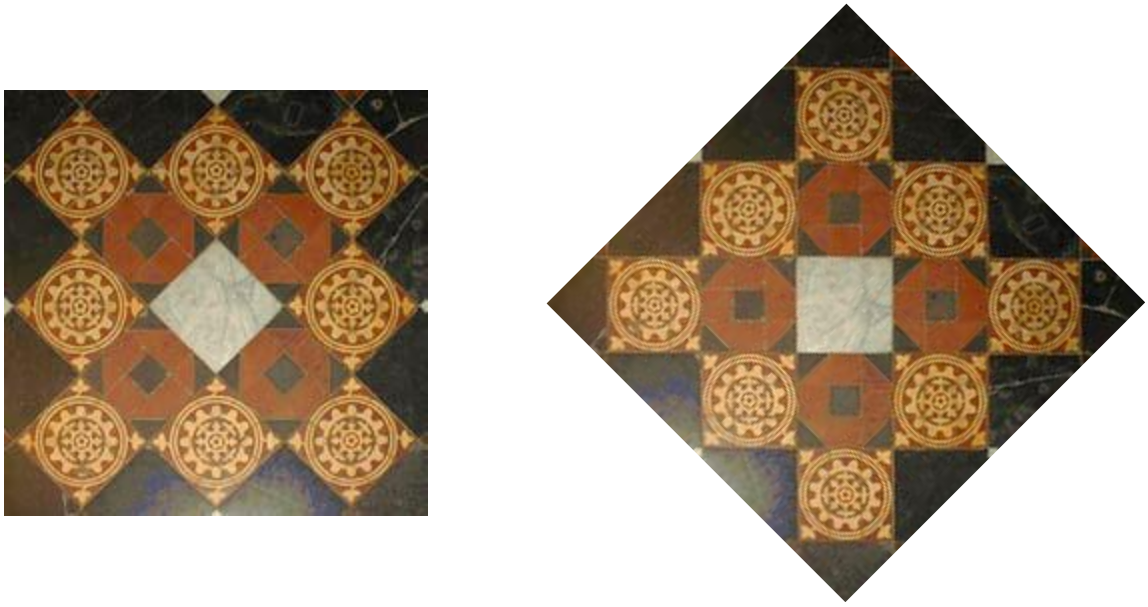
5.2.1 Transformation matrix

The fundamental step in any image registration algorithm is to choose a proper transformation method that is able to properly map the images. The most common transformation methods are: rigid body (Euclidean), similarity, affine, and projective transformation. [80]

5.2.1.1 Euclidean transformation

This method is also known as rigid transformation and the algorithm acts to preserve the Euclidean distance between each pair of points. By doing so, the rigid transformation preserves the size of each feature and the angles between them. The only parameters which are affected by this transformation are rotation and translation. Translation means an adjustment that moves corresponding image points along the x-axis, y-axis, or both. Rotation means rotating corresponding points around a common point of rotation (usually the center of the image) by a value between 0° to 360°. If an image is only affected by these two parameters, the Euclidean transformation can be applied to overlay the images. To do so, a Euclidean transformation needs at least two pairs of corresponding control points in the matching images. An example of Euclidean transformation is illustrated in Figure 5.7 in the image taken of a floor tile. The transformed image

has been rotated and translated, but the size of the squares or circles and the distances between the points are unchanged between.



a *b*

Figure 5-7: Euclidean transformation: a) original image of the floor tile, b) transformed image[73] (Reproduced from: Hartley R, Zisserman A. Multiple view geometry in computer vision 2nd ed., 4th print. 2006)

The Euclidean transformation can be calculated using Equation 5-1.

$$p' = Rp + t$$

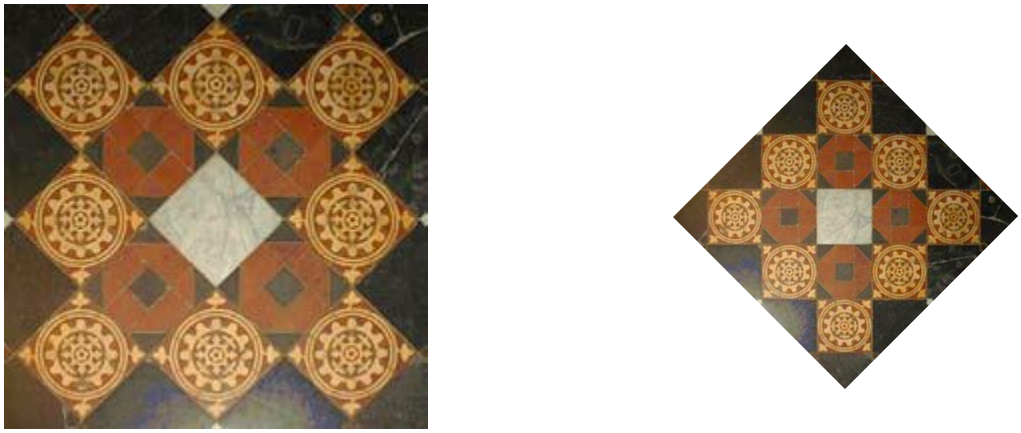
where $R = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$ *Equation 5-1*

where R is the rotation matrix, t is the translation vector, and p and p' are the original and transformed points, respectively.

5.2.1.2 Similarity transformation

The similarity transformation method preserves the angles between each feature point, as well as the ratio of length and ratio of areas. In addition to translation and rotation, the magnification of an object in the image is considered. The key point in this transformation is that the ratio of the length and areas are preserved. To register an image to a reference image that is affected by

translation, rotation, and scaling (resizing), the similarity transformation is the best method. Usually, two pairs of corresponding points are enough to calculate the similarity transformation between a pair of images. Figure 5-8 shows an example of a similarity transformation. In the transformed image, the circular patterns are imaged as circles and the square patterns are imaged as squares, which means the ratios are preserved. The parallel and perpendicular lines in the image preserve their relation after transformation. However, the size of the image is changed. The rotation is 30° and there is no translation in this example. Therefore, the only parameter needed to register the images is the scaling factor.



a *b*

Figure 5-8: Similarity transformation: a) original image, b) transformed image[73](Reproduced from: Hartley R, Zisserman A. Multiple view geometry in computer vision 2nd ed., 4th print. 2006)

The similarity transformation can be calculated using Equation 5-2.

$$p' = sRp + t$$
$$\text{where } R = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \qquad \textit{Equation 5-2}$$

where R is the rotation matrix, t is the translation vector, p and p' are the original and transformed points, respectively, and s is the scaling factor.

5.2.1.3 Affine transformation

An affine transformation is the most commonly used method. It preserves collinearity (i.e., all points initially lying on a line still lie on the line after transformation) and the ratios of distances

(e.g., the midpoint of a line segment remains the midpoint after transformation). All parallel lines remain parallel after an affine transformation, but not the angles between the lines. The affine transformation is composed of, translation, rotation, scaling, and shearing. Shearing produces distortion (or perhaps corrects distortion) in the image by moving one part of the image in one direction, and another part of the image in another direction. Affine transformation needs at least 3 pairs of corresponding points to be calculated. Figure 5-9 shows an example of affine transformation. The sides of the square tile that are parallel in the untransformed image remained parallel in the transformed image, but the orthogonal lines in the untransformed image do not remain orthogonal in the transformed image. The ratio of the lines (distances) in the images are changed and there is some shearing in the image - the circles in the transformed image have been modified to appear as an ellipse.

a *b*

Figure 5-9: Affine transformation: a) original image, b) transformed image[73](Reproduced from: Hartley R, Zisserman A. Multiple view geometry in computer vision 2nd ed., 4th print. 2006)

The affine transformation can be calculated using Equation 5-3.

$$p' = Ap + t$$
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

Equation 5-3

where p and p' are the original and transformed points, respectively, t is translation, and A is the matrix for scaling, rotation, and shearing. The matrix is calculated from multiplication of the matrices for each parameter.

5.2.1.4 Projective transformation

A transformation that maps lines to lines (but does not necessarily preserve parallelism) is a projective transformation. A projective transformation is the general case of a linear transformation on points in homogeneous coordinates. Therefore, the set of projective transformations in three-dimensional (3D) space is the set of all four-by-four matrices operating on the homogeneous coordinate representation of 3D space.

Figure5-10 shows an example of projective transformation. The circles and squares do not preserve their shape, and parallel world lines converge in the image. As a result, the circles closer to the camera are larger in size on the transformed image.

a *b*

Figure5-10: Projective transformation: a) original image, b) transformed image [73](Reproduced from: Hartley R, Zisserman A. Multiple view geometry in computer vision 2nd ed., 4th print. 2006)

The projective transformation can be calculated using Equation 5-4.

$$p' = Hp$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Equation 5-4

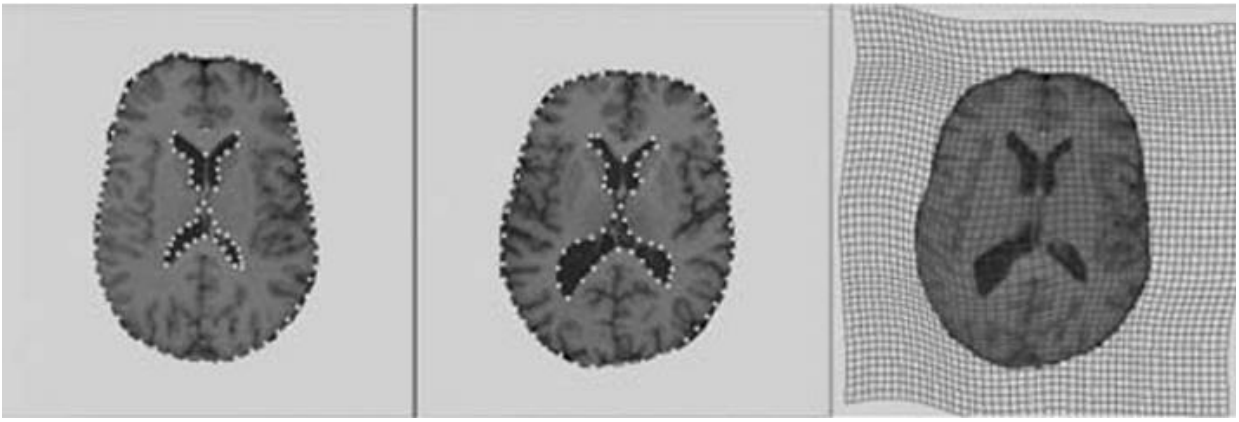
where p and p' are the original and transformed points, respectively, and H is the homogeneous matrix to transform the points. $h11$, $h12$, $h21$, and $h22$ are the rotation matrix elements (this matrix defines the rotation, scaling, and shearing), $h13$ and $h23$ are the translation vectors, and $h31$ and $h32$ are the projection vectors.

5.2.2 Image registration techniques

There are several different methods reported in the literature for image registration. These can be divided into three main categories: point-based, surface-based, and intensity-based methods.

5.2.2.1 Point-based methods

Point-based methods use corresponding points on each image, called control or fiducial points, to perform the image registration task. Using a transformation matrix, the points can be mapped on each other. Ideally, every possible corresponding point in the two images should be identified and mapped together, but this is a processor-heavy image processing task. Instead, only some of the corresponding points are selected in the pair of images, and from that image registration, the transformation is applied to the whole of the image. This is called an estimated transformation matrix. Figure 5-11 shows an example of point-based image registration: a) moving image with control points, b) fixed image with control points, c) after finding the transformation matrix, the moving image is transformed and mapped onto the fixed image.



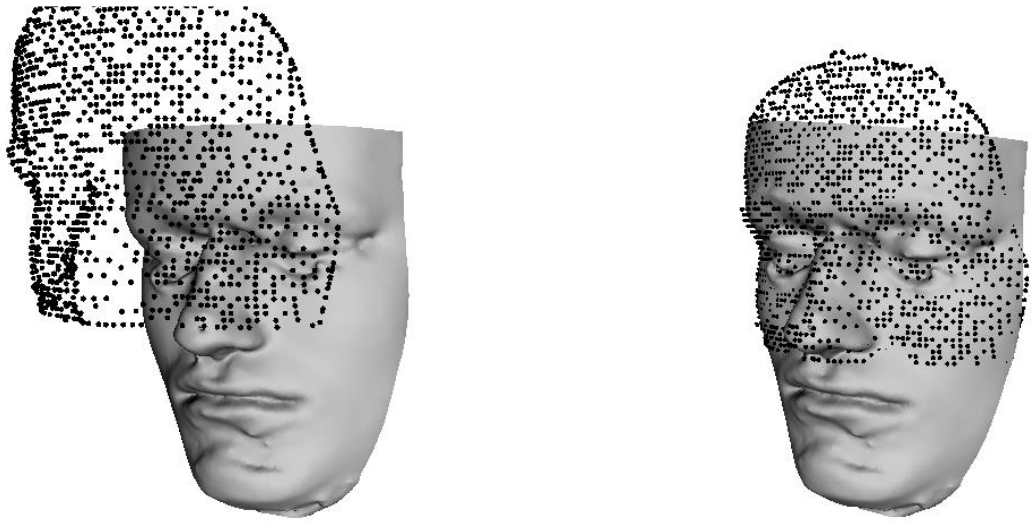
a *b* *c*

Figure 5-11:An example of point-based image registration; a) Moving image, b) Fixed image, c) Registered image showing the effect of the transformation on an image grid. [81]*(Reproduced from: Xie Z, Farin GE. Image registration using hierarchical B-splines. IEEE Trans Visual Comput Graphics. 2004;10(1):85-94)*

5.2.2.2 Surface-based methods

Surface contours or boundaries are usually distinct in medical images. A surface-based image registration looks for corresponding surfaces in different images and calculates the best transformation that aligns these surfaces. For example, the skin boundary surface and the outer cranial surface are two surface examples that could be used for image-to-image and image to physical registration in head images.

The surface representation can be a point set, a faceted surface, an implicit surface, or a parametric surface. The surface that covers the largest area on the image is usually considered as complete and is used for surface model generation. After that, iterative closest point and feature matching algorithms are used for surface registration.



a *b*

Figure 5-12: An example of surface-based image registration: a) before registration, b) after registration[82](Reproduced from: Fitzpatrick JM, Hill DL, Maurer CR. Image registration. Handbook of medical imaging. 2000;2:447-513.)

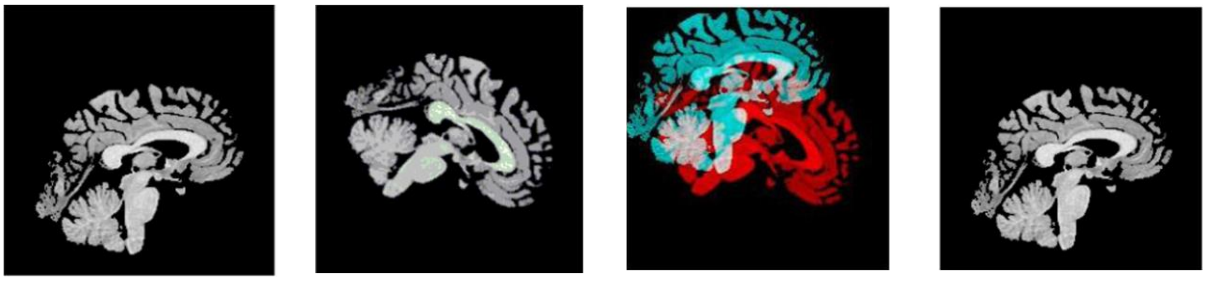
Figure 5-12 shows an example of surface-based registration. In this example, the head is registered using the facial skin surface. The dots represent skin surface points acquired with a four-camera photogrammetry system. The surface rendering represents a triangle set model of the skin surface extracted from a magnetic resonance image. The left panel (a) shows the initial position of the data sets, and the right panel (b) shows the data sets after registration. The registration was performed using an independent implementation of the iterative closest point algorithm [68]. The surfaces

overlap only partially, which is a common situation in surface-based registration. The problem of partial overlap is dealt with by setting the weights of any outliers to zero after the first search, before running a second search.

5.2.2.3 Intensity-based methods

For this method, intensity refers to the scalar values in each image pixel. Intensity-based registration involves calculating a transformation between two images using the pixel values alone. The intensity-based registration works by iteratively optimizing some similarity measures and calculating the transformation between all pixel values. Intensity-based methods can also be used for 3D image registration, where the voxel similarity measure is used for each data point (voxel). Most intensity-based algorithms used a sub-set of pixels or voxels and require some sort of pre-processing. The sub-set can be chosen using a regular grid or by randomly choosing points, or an algorithm might work in a defined region of interest in an image. In this kind of registration some sort of pre-segmentation of the image is required, which depends on the modalities being registered and the part of the image being studied. In some other intensity-based algorithms, the similarity measures work on derived image parameters, such as image gradients, rather than the original voxel values. The advantage of the intensity-based method over the point-based and surface-based methods is lower user interaction and pre-processing. This makes the method easier to automate.

Figure 5-13 shows an example of intensity-based image registration: a) and b) show the fixed and moving images, respectively; c) shows the combined, but unregistered images; d) is the final registered image.



a *b* *c* *d*

Figure 5-13: Intensity-based image registration: a) fixed image, b) moving image, c) combined unregistered images, d) final registered image. [83]*(Reproduced from: Abdel-Basset M, Fakhry AE, El-Henawy I, Qiu T, Sangaiah AK. Feature and intensity based medical image registration using particle swarm optimization. J Med Syst. 2017;41(12):1-15)*

In this project, since two different sensors were used for image capture, the sensor-dependent image formation for each camera is completely different and intensity-based methods cannot be used. Surface-based methods were also not suitable because they use 3D surface matching to determine the transformation between images. Therefore, the best method for thermal and visible image registration in this system is the point-based method.

5.2.3 Image registration for a sequence of images

The system is designed to record a series of images from the visible and thermal cameras in order to track OST over time. After recording the image sequence from each camera, each pair of corresponding timestamped image frames should be registered for the whole video sequence. To do this, a series of control points for each pair of frames should be selected. However, to identify salient points for each pair of images is impractical and inefficient. Instead, corresponding salient features are identified in the first image of each image sequence for each camera and then an algorithm continues to track these features in each subsequent image frame for each sequence. The image registration algorithm estimates the transformation between the features and then processes each image pair to produce a sequence of overlapped (registered) images.

5.2.4 Feature matching

The feature matching algorithm in Step 2 of the image registration process (Section 5.2) should be robust and efficient, but it can be affected by the pixel characteristics of the selected features. When selecting a suitable feature to mark for image registration, it is best to choose a feature that is distinctive (an edge, corner, or line intersection) that occurs in both sets of images. When a series of images is to be registered, this requirement is even more important. The features must consistently (or at least frequently) appear in each image frame from each image sequence for both cameras. This ensures that there is a consistent point of reference for each pair of images throughout the image sequence. For multimodal (i.e., two cameras, as used in this system) image registration, this step plays an important role in the result of the registration. The physical interpretability of the features is very important in this step. If automatic feature selection is used, the algorithm should also be able to detect the same features from different views of the projection.

As well as needing to appear in each image sequence, the selected feature should have similar pixel characteristics for the best matching results. (Pixel characteristics can be defined according

to RBG/greyscale). However, in a multimodal system, the pixel characteristics for corresponding features can be dissimilar in the images from each camera due to different sensor sensitivities or capturing conditions that are more suitable for one camera than the other. This can make it more difficult for the matching algorithm to consistently identify the corresponding features that are used in image registration.

Finally, the mapping function should be selected based on the image capturing device and acquisition process information. Based on this information, the effect of image degradation can be considered, and a suitable transformation function used. To sum up, the three steps are: accuracy of feature selection, reliability of feature correspondence detection, and accuracy of transformation estimation all play important roles in the process of image registration.

5.2.5 Fiducial Registration Error

A measure of the accuracy of the image registration produced by a transformation matrix can be determined by using the location of the marked features on the fixed image and comparing the alignment (registration) of the features in the transformed moving image. Different transformation matrices will produce different amounts of error in the registration – an ideal matrix would produce no error in the alignment. The size of the error can then be used as a measure of the accuracy of the transformation matrix. Since the marked features used for registration lie in discernible features (which makes them consistently visible and therefore reliable), they are called fiducial features. (Fiducial comes from the Latin *fiducialis* meaning reliable; and *fiducia* meaning trust). Fiducial localization is the process of determining these reference points and can be done using automatic or manual selection of the points.

The fiducial registration error (FRE) can be calculated after transformation calculation. After selecting some points on the images (fixed and moving), the transformation matrix is calculated. To calculate the FRE, some salient points (not the first set of point used for transformation estimation) should be selected on the moving images. Then, the points should be transformed using the estimated transformation matrix and the actual position of the salient points on the transformed image selected manually as a reference. Figure 5-14 shows an example of FRE calculation, where the black filled circles are the manually selected reference points, and the unfilled circles are the location of the salient points as located by the transformation function.

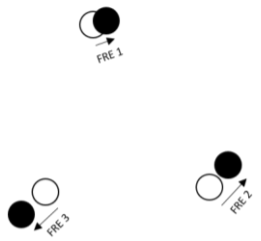


Figure 5-14: FRE demonstration

5.3 Thermal-visible image registration

To meet the main goal of the project, which was to extract temperature data from the thermal image by selecting the corresponding region of interest (ROI) in the visible image, and then mapping that ROI onto the thermal image, accurate image registration between the visible and thermal images is a critical requirement. As discussed previously, it is necessary to find some common features that can be consistently identified in both images. However, the sensors used in the project are designed to detect different wavelength spectrums for image formation. This makes it hard to match the features using an automatic feature-matching algorithm, as the pixel characteristics produced by the two sensors are dissimilar. Hence, all of the control points on the images (thermal and visible) used for transformation matrix estimation were selected manually by the user.

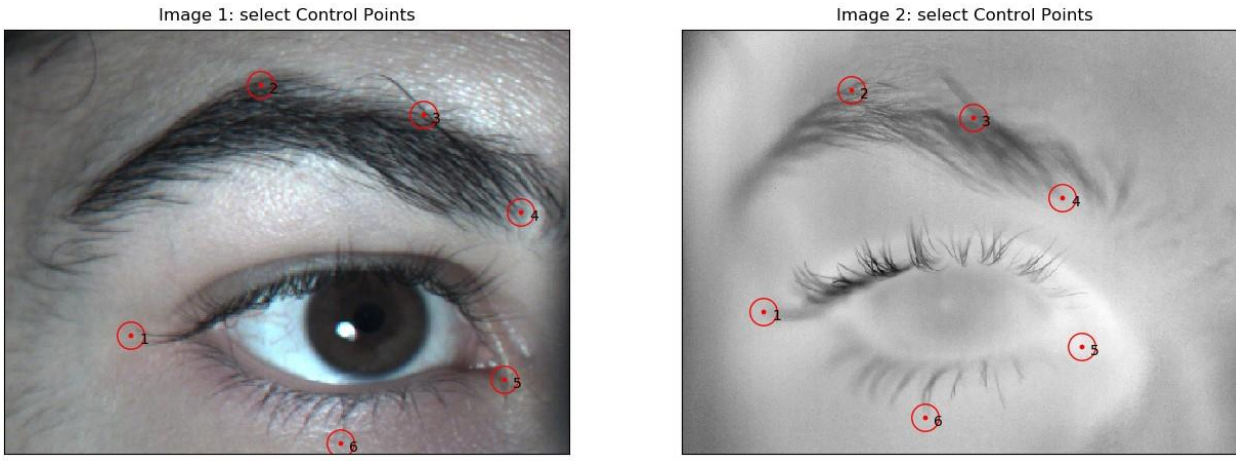


Figure 5-15: Control point selection showing the corresponding salient points in a pair of visible (left and thermal (right) images.

Figure 5-15 shows an example of five corresponding points selected on the visible and thermal images, each marked by a red dot surrounded by a red circle. The points selected were judged to give a high degree of distinctiveness that would help in point identification in subsequent images. The points were also selected to be on the face rather than the eyelid or eye itself, as these facial areas will have less movement during blinking and eye movement, and not be obscured by eyelid closure.

After manually selecting the two sets of points, the transformation between the points was calculated. In this project, since the visible image had a smaller pixel size and better quality compared to thermal image, it was used as the fixed image. Therefore, the calculated transformation matrix was applied to the thermal image to transform its shape and allow it to be overlayed onto the visible image. As any lens distortion was removed from the images using image normalization (Chapter 4), an affine transformation was used for the registration step. An affine transformation needs at least three corresponding points on the images to be calculated. However, more than three points were selected to improve the accuracy of the registration.

a *b* *c*

Figure 5-16: Image registration results showing a) the reference fixed visible image, b) the unregistered thermal image, and c) the transformed thermal image.

Figure 5-16 shows the registration result with the thermal image mapped to the co-ordinates of the visible image. By looking at the images side-by-side, the change in eye size of the thermal image can be observed – the eye size is now the same in both images. The small black wedge on the left side of the transformed image reveals how the thermal image (Figure 5-16c) has been rotated as part of the transformation. Figure 5-17 demonstrates how the two images are now registered

together. The marked corresponding locations on each image are linked together by parallel horizontal lines.

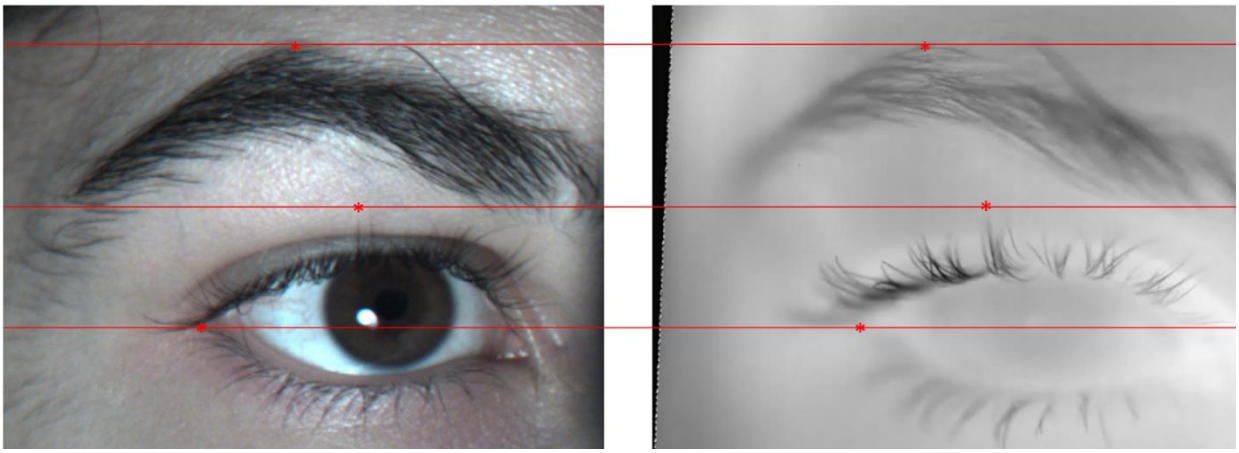


Figure 5-17: Alignment of the registered images is shown by the parallel lines connecting the corresponding points on the two images.

Manual selection of the points results a good fiducial localization. The algorithm uses two methods to compute robust transformation, RANSAC and Least-Median robust method (LMEDS). The methods will find the best three matching points using the RANSAC or LMEDS at the first step, and then using the selected points on the images for affine transformation calculation.

5.4 Video registration and stabilization

When considering two paired images, selection of the salient control points can be made manually or automatically using an algorithm. For this project, the first set of corresponding control points were manually selected on the images. However, this project uses the video output from each camera to record two separate video sequences. This raises the question of how to maintain image registration between paired images in two video sequences. Manual selection of control points on each paired frame set in the video sequences is not viable, and so for each subsequent frame pair after the first frame pair, the selected points must be tracked automatically.

The output of the video registration should be two aligned video files that have the same coordinates on the corresponding frames points for the same location on the face. The most important part of video registration is control point selection. For a video file with a duration of 20 seconds recorded at 25 fps, there are 500 frames for each video. It is not possible to select the control points on all of the corresponding frames. Instead, the control points must be selected on

the first frames pair and automatically followed on the remaining frames using an optical flow algorithm. There is also a further issue arises from movement by the subject during the video recording. This will mean that the position of the control points in the fixed images are not stable. This movement will have an effect on registration accuracy. However, by stabilizing the video files, the effect of subject movement can be removed from the video files and improved accuracy can be achieved.

Stabilization can be done by registering the video frames on a reference frame. The reference frame should be selected manually in this algorithm. For this system, the first frame of the visible video sequence is taken as the reference frame. To stabilize the visible video, all of the visible frames are then registered with the reference frame (Figure 5-18).

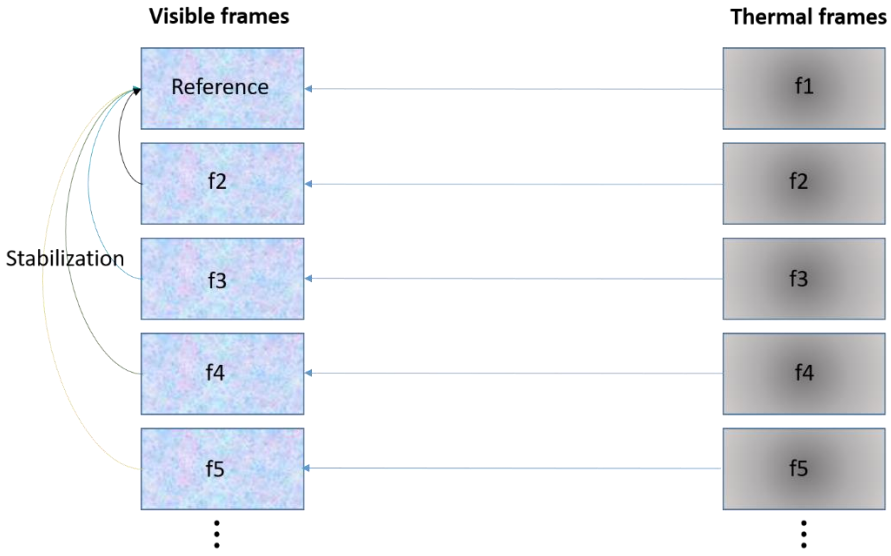


Figure 5-18: Video registration showing stabilization of the visible frame sequence

5.5 Next steps

As previously mentioned, to register two video files, a series of control points must be identified on each pair of corresponding images. To do this, control points should be selected for all of the frames, but it is impossible to locate the control points on all of the frames manually. Instead, an optical flow algorithm was used to automatically locate the control points on all of the frames. In the next chapter, the details of the optical flow algorithm are described.

Chapter 6: Point tracking/Video registration

The previous chapters have described the methods for image and video registration. However, one of the major problems for video registration is the effect of camera or object movement. In this system, the location of the camera is fixed and there is no camera movement. However, the subject might move their head or eyes during the recording. To remove the effect of subject movement, the video files must be stabilized. Video stabilization refers to algorithms that improve video quality by removing unwanted camera/subject movement. As discussed in the previous chapter, to register and stabilize two video files, corresponding control points are required on each pair of frames. However, it is intensely time-demanding and inefficient to select the control points on the frames manually. Therefore, in this project, the control points were selected on the first pair of frames manually, and the location of the salient points automatically tracked and localized on all subsequent frames using an optical flow algorithm. The concept of optical flow is described in Section 6.1. The optical flow algorithm and calculation method is described in Section 6.2. The application of the optical flow algorithm in this system is demonstrated in Section 6.3, and performance evaluation of the method is reported in Section 6.4.

6.1 Optical flow

Optical flow is the process for following the actual or apparent motion of an object between two consecutive frames of a video sequence caused by object or camera movement, respectively. Optical flow algorithms provide a 2D vector in which the elements show the displacement of the control points from their position in the first frame to their position in the second frame. Salient, or control, points are located in the first frame of a video sequence either manually or automatically, and then their positions tracked in the subsequent image sequences using an optical flow algorithm. For this system, an optical flow algorithm was used to track the displacement of the manually selected control points in the video frames for both the visible and thermal video sequences.

There are two main assumptions in the application of an optical flow algorithm. First, that the pixel intensities of the selected points do not change in two consecutive frames – a variation in pixel intensity between frames will make it more difficult to correctly identify the control point in each

subsequent frame. Second, that neighboring pixels move together – it is the contrasting pixel characteristics between neighboring pixels that enables the algorithm to locate the control point pixel.

If we consider a control point in the first image frame of the video sequence as I (x,y,t) where the x,y is the coordinate of the point and t is the time, then the point can be displayed in the next image frame by I(x+dx, y+dy, t+dt), where the dx and dy represent displacement and dt shows the time difference. By assuming that the pixel intensity of the point has not changed between the two consecutive frames, then:

$$I(x, y, t) = I(x+dx, y+dy, t+dt) \qquad \textit{Equation 6-1}$$

Then the Taylor series approximation can be used to removing common terms, to produce:

$$I(x+\delta x, y+\delta y, t+\delta t) = I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + \ldots \qquad \textit{Equation 6-2}$$

$$\Rightarrow \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = 0$$

Equation 6-2 can be divided by dt to gain the optical flow equation:

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \qquad \textit{Equation 6-3}$$

where u=dx/dt, v=dy/dt, dI/dx is the image gradient of the horizontal axis, dI/dy is the image gradient of the vertical axis, and dI/dt is the image gradient of time.

By solving u(dx/dt) and v(dy/dt), the movement of a control point between consecutive frames can be determined.

6.2 Optical flow algorithms

Optical flow algorithms are divided into two main categories: sparse and dense optical flow. In sparse optical flow, there are some salient features, such as a corner or edge, that are tracked in the consecutive frames to produce a vector for the selected points. Dense optical flow, on the other hand, tracks the flow between frames for all of the pixels in the frame and produces a vector for each pixel. Dense optical flow is more accurate, but also needs more computation time. The main goal for the optical flow algorithm in this system was to follow the salient points on the thermal and visible images used for point-based image registration, hence, the sparse optical flow method was used.

6.2.1 Lucas-Kanade (LK) Optical Flow

An effective sparse optical flow method to estimate motion between two consecutive frames was introduced by Lucas and Kanade (LK)[84]. There are three main assumptions in the LK algorithm. First, that there is a very short time difference between the two consecutive frames. Second, that the pixel intensity of the selected features changes very smoothly in the two consecutive frames. Third, a window of 3x3 pixels is considered instead of a single pixel, and it is assumed that the neighbor pixels are moving together. Taking the optical flow equation (Equation 6-3), then for a window of n pixels the equation will be:

$$\begin{aligned} I_x(q1)V_x + I_y(q1)V_y &= -I_t(q1) \\ I_x(q2)V_x + I_y(q2)V_y &= -I_t(q2) \\ &\vdots \\ I_x(qn)V_x + I_y(qn)V_y &= -I_t(qn) \end{aligned} \qquad \textit{Equation 6-4}$$

where q1, q2,…, qn denote the pixels inside the sample window (n = 9 for a 3x3 window) and $I_x(qi)$, $I_y(qi)$, and $I_t(qi)$ denote the partial derivatives of image I with respect to position (x,y) and time t, for pixel qi at the current time.

Equation 6-4 can then be written in a matrix form of Av=b as Equation 6-5:

$$A = \begin{bmatrix} I_x(q1) & I_y(q1) \\ I_x(q2) & I_y(q2) \\ \vdots & \vdots \\ I_x(qn) & I_y(qn) \end{bmatrix} \qquad v = \begin{bmatrix} V_x \\ V_y \end{bmatrix} \qquad b = \begin{bmatrix} -I_t(q1) \\ -I_t(q2) \\ \vdots \\ -I_t(qn) \end{bmatrix}$$

Equation 6-5

where A is the pixel matrix, v is the movement vector, and b is vector of the partial derivitives of pixel qi and time t. This produces n x n equations.

By applying least square fitting, a two-equation/two-unknowns equation is gained as follows:

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum_i I_x(qi)^2 & \sum_i I_x(qi)I_y(qi) \\ \sum_i I_y(qi)I_x(qi) & \sum_i I_y(qi)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_x(qi)I_t(qi) \\ -\sum_i I_y(qi)I_t(qi) \end{bmatrix}$$

Equation 6-6

where V_x=dx/dt and V_y=dy/dt demonstrate movement in x and y directions over time, respectively. Solving for these two variables completes the optical flow problem.

6.2.2 Pyramid Model for Optical Flow

The LK algorithm only works for small motions. If the object has a large movement in the frame, the algorithm will fail. To solve this problem, the pyramid algorithm was introduced. The pyramid model is used with the LK algorithm to detect large movements. The solution is simple: the image is re-sized to a smaller size while the sample window size remains unchanged. The window can now cover a larger area on the image and detect the larger movement. The algorithm is called a pyramid model because the optical flow is calculated for each pair of images in a series of resizing steps to produce a range of optical flows from coarse to fine. It starts with the smallest image size and the optical flow calculated, then the image is warped (re-sized) and the optical flow re-calculated, and so on until the final image size is reached.

6.3 Applying Optical Flow to the system

As stated, optical flow is needed in this system to remove the artefacts produced by head and eye movement. The LK algorithm incorporating a 3-level pyramid was used for this purpose, and it can complete its' calculations using the localized control points previously marked on the first

thermal and visible frames of the video files for image registration. The salient points selected were, for example, the corner of the eye or some points along the eyebrows. Any head or eye movement can then be tracked and localized on all of the subsequent frames on each video file independently.

To implement the LK algorithm, the first step was to select some corresponding control points on the first thermal and visible frames. Since we are using two different sensors for recording, the pixel features on the corresponding frames have different values, and, as noted earlier, it is not possible to use automatic feature detection algorithms to find corresponding features. Since very accurate point tracking is required, the control points were selected manually. Figure 6-1 shows an example of six selected and corresponding control points on time-stamped visible and thermal frames.

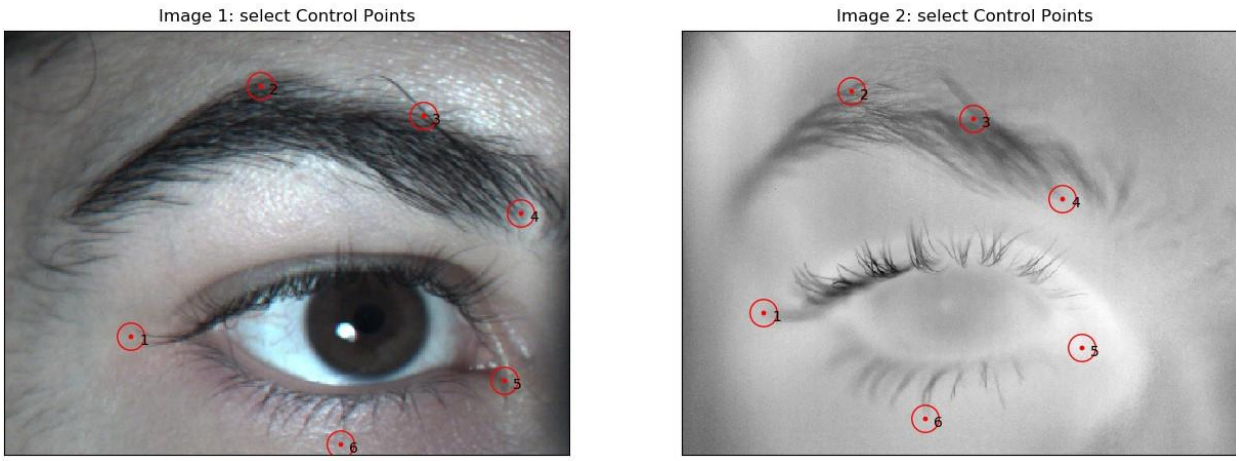


Figure 6-1: Control point selection on the first visible and thermal frames for optical flow algorithm input

As noted in Section 5.4.3 three points is usually enough for an affine image transformation, but using more points bring higher accuracy for the image registration process. Consequently, between 5 to 8 corresponding points were manually selected on the initial reference frames of each video sequence. The location of these points was then used as the input for the LK algorithm to track optical flow on the subsequent frames in each video file. For this system, the subject sits in front of the cameras on a headrest and so only very small head movements are expected. However, since both the head and the eye can move, and the eye can move independently from the face, some points were selected on the face as well as around the eye. Points were not selected on the surface

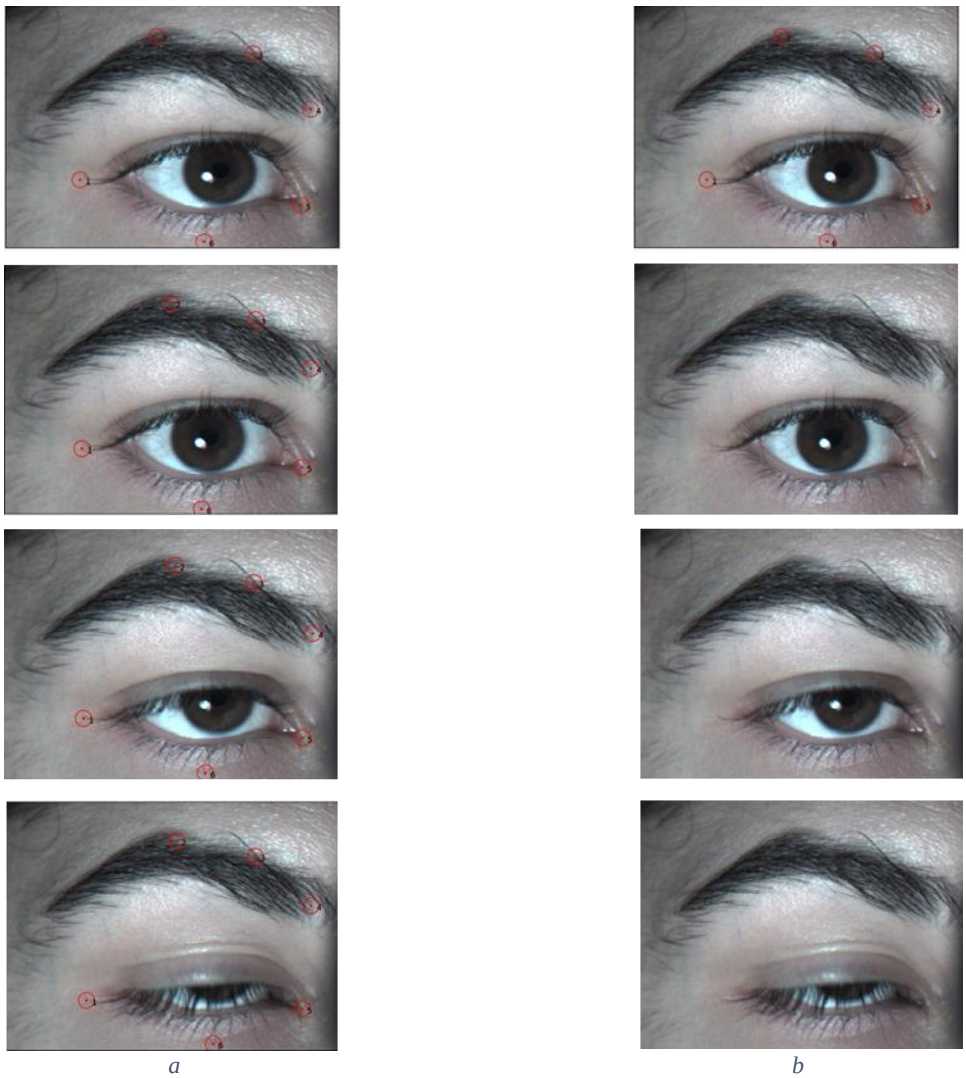
of the eye itself, but movement of the eye will produce movement in the eyelids, and this can be tracked. Consequently, points on the eye corners and eyebrows were selected.

It is very important to have a minimum point selection error as this can have an impact on the registration accuracy. Minimum point selection error describes the error in location between two corresponding points in the two different images. A relatively small window size (5 x 5) was used in the optical flow algorithm.

6.4 Performance evaluation

For any point-tracking method, the performance of the algorithm applied should be calculated. The optical flow algorithm error can be calculated by measuring the distance between the control point in the reference frame and the localized point produced by the optical flow algorithm. A timestamped and matched pair of thermal and visible video files of a subject's face and eye, containing over 500 frames, was selected for performance evaluation. The L-K algorithm was applied to each video sequence separately. Figure 6-2a shows a sequence of reference images from the visible camera with manually-selected control points marked by red dots. Figure 6-2b shows the same sequence of video frames, this time unmarked there were used as the input for the optical flow algorithm.

A software program was developed that could detect the location coordinates of each manually selected control point on a video frame and record the coordinates in a file (Figure 6-2a). Using this software, the control points were selected on the first frame of the video file (Figure 6-2b). After that, the video sequence was processed through the optical flow algorithm, and each point was tracked and localized using the software, with the location of the points in each subsequent frame stored in a file. The Euclidean distance between the locations for each matching point between two subsequent frames was calculated as the optical flow error. Then, the average of the distances was used to calculate the optical flow error for the whole video. The same process was repeated for the thermal video sequences. Table 6-1 shows the calculated error.



a b

Figure 6-2: Performance evaluation of the point tracking method a) manually selected reference frames b) video frames produced as an input of the algorithm.

Table 6-1: Calculated optical flow algorithm error

Visible video	MSE ± SD = 5.43 ±2.01 pixels=5.43 * 0.09 mm
Thermal video	MSE ± SD = 6.81 ±2.32 pixels= 6.81 * 0.09 mm

After localizing the control points in each frame, the video files were registered using the point coordinates of the control locations on the visible and thermal frames \ using the image registration Affine transformation algorithm.

Any misplacement of the points in the registration process is known as the registration error. A common measure of overall point misalignment is the root-mean square (RMS) error, which is

also called the fiducial registration error (FRE). To calculate the FRE, the Euclidean distance between the fixed and transformed points should be calculated, with the average of the distances as the FRE. In this project, the FRE calculation was done as follows: about 10 points were selected in both the fixed (visible) and moving (thermal) images. Then, the transformation matrix was calculated using 3 corresponding pairs of the points. After that, the remaining points were transformed by the transformation matrix. At the end, the transformed points distance with selected points was calculated and averaged as the FRE. This process was repeated for all of the frames and FRE of all frames were averaged as video registration error. Mean FRE ± SD= 5.03 ±1.82 pixels, gained over 500 frames. With a pixel size of 0.09mm, the error is approximately 0.45mm.

6.5 Next steps

The stabilized video files are aligned and prepared for the next step. In the next step, the cornea should be located on the visible frames. Then, the coordinates of the cornea should be used for temperature extraction. In the following section the methods for corneal segmentation and localization are described.

Chapter 7: Cornea localization

To extract the OST values for the cornea from each thermal video frame, the corneal area must be localized on the thermal image of the eye. However, thermal image formation is based on the temperature distribution across the ocular surface and there is no clear boundary that separates the cornea from the conjunctiva in the thermal images.

The previously reported steps of image and video registration for the dual camera system have produced an aligned visible and thermal image, with the effect that the same pixel coordinates in each corresponding visible and thermal frame represent the same location on the eye. Using this arrangement, the cornea can be localizing in the thermal image by first using the visible image to identify and segment the cornea, and then using this segmented area as a mask over the thermal image to locate the cornea in the thermal image. The limits of the cornea on the visible image are best found by identifying the underlying iris, since the iris provides a strong contrasting signal to the visual contrast of the neighboring sclera. The more precise the localization of the cornea in the thermal images, the more accurate and reliable the data will be for the data analysis steps. The concepts of image segmentation using deep learning are discussed in Sections 7.1 – 7.4. Previous methods of iris segmentation are reviewed in Section 7.5. The network used in this project and its performance evaluation are described in Section 7.6. The blink detection algorithm is described in Section 7.7, and the final outcomes of the system are reported in Section 7.8.

7.1 Image segmentation

Image segmentation is the process of partitioning an image into a multiple number of segments that have similar features or characteristics. The objective is to divide an image into more meaningful segments, each of which could then be analysed separately. The effect is to simplify the analysis step of the project. Figure 7-1 shows two examples of image segmentation. Each image pair shows the original image and its segmentation mask, which shows the segmented areas in pixels of different colors.

Figure 7-1: An example of image segmentation with the original image on the left side and the segmentation mask on the right, with segmented areas shown in different colors.

Image segmentation is one of the most important components of image processing. Indeed, it has been considered as the first step of image processing. An efficient image segmentation brings more accurate and reliable results in the later analysis of an image. It can be used as part of the analysis of a single frame or image, as well as in the dynamic situation of a moving scene in computer vision. It is an essential part of image analysis in many systems, enabling the automatic identification of areas of interest. It has many applications in different fields, such as medical image analysis, autonomous vehicle control, video surveillance, and robotic surgery.

A wide variety of methods have been reported in the literature for image segmentation tasks, such as thresholding, region growing, clustering, morphological, edge detection, active contour, graph cuts, and neural network (learning-based). All of the methods have their own advantages and

disadvantages. Therefore, the selected method should be based on the needs of the project and perspective. Learning-based methods have shown a remarkable performance in the field of image segmentation and generally out-perform all previous methods in terms of accuracy and performance. To understand learning-based methods, it is first necessary to know about machine learning.

7.2 Machine learning

Machine learning is an artificial intelligence (AI) application that provides the opportunity for a computer software system to learn from sample data on what should be correctly identified within an image, and by so learning to improve the output from more complex data without explicit programming. The developed algorithm can then be used to improve the efficiency and/or speed of identifying areas of interest in images for large data sets. The process of learning includes looking for patterns in the images that are consistent in each image and to use these patterns to improve decision making. Machine learning algorithms are categorized into three categories: supervised, unsupervised, and reinforcement learning.

7.2.1 Supervised learning

Supervised learning is a basic method that trains the algorithm on labeled images. The images need to be manually labeled accurately, but this method has very high accuracy when it is chosen for the right project. In this method, the original images and labeled images are given to the algorithm for training purposes. The algorithm finds the relationship between the parameters given by the labeled data and establishes the cause-and-effect relationship between the variables in the dataset. Supervised machine learning algorithms learn from the available labeled data and produce a function to predict output values. At the end of training, the model has the ability to compare the output data with ground truth and find the error, and then use the error to improve the model.

7.2.2 Unsupervised learning

An algorithm designed for unsupervised learning has the ability to learn from unlabeled data. Since the dataset is not required to be labeled by the user, much larger datasets can be used in the program. In supervised learning, the labels allow the algorithm to learn the exact nature of the relationship between the data points. On the other hand, in unsupervised learning there is no label to learn from. Instead, the relationship between the data points is calculated by the algorithm with

no outside input. The creation of these hidden structures is what makes unsupervised learning algorithms versatile. Instead of a defined problem statement, unsupervised learning algorithms can adapt to the data by dynamically identifying hidden structures. This offers more post-deployment development than supervised learning algorithms.

7.2.3 Reinforcement learning

Reinforcement learning imitates human learning behavior. The algorithm learns using a trial-and-error method and gradually improves the accuracy of its output. Favorable outputs are reinforced (encouraged) and non-favorable outputs are punished (discouraged). The algorithm is put in a working environment accompanied by an interpreter and reward system. In every iteration of the algorithm, the output result goes to the interpreter to decide whether the outcome is favorable or not. The algorithm gets a reward when the outcome is favorable, and when the outcome is not favorable, the algorithm is forced to reiterate until a better result is found. The reward system is most-often defined on the effectiveness of the output.

7.3 Deep learning

Deep learning is a sub-field of machine learning that tries to imitate human brain activity in learning. It is representation learning or, to put it another way, learning from examples. Deep learning is inspired by human brain structure and function to produce an artificial neural network (ANN), sometimes also called a deep neural network (DNN) when multiple layers are used. A DNN represents the structure of a human brain modeled in a computer and consists of neurons and synapses organized into layers. A DNN has millions of neurons connected together that enable the system to analyze large data sets or to memorize them. A DNN consists of: neurons, synapses, weights, biases, and functions. It automates the formation of data representation (machine learning) that could be useful in decision making.

Neural networks are used to solve complex problems that need lots of computation and analytical calculations, like those of the human brain. The most common tasks for a neural network are classification, prediction, and recognition.

7.3.1 Neurons

A neuron is the basic unit of a DNN that receives information, performs some simple calculations, and passes it on to another unit. Each neuron is composed of a cluster of pixels from one part of the image. Within that cluster of pixels there may be a variation in the textural information. Calculations are performed on this variation to determine whether the variation is sufficiently large (or small, depending on the purpose of the calculation) for the neuron to be activated. In this way, every neuron in the DNN contributes input data to the algorithm which is learning to extract a feature.

Neurons are classified as input, hidden, or output. In a large DNN containing a large number of neurons and connections between them, neurons are organized in layers. Generally, there is an input layer to receive information, some hidden layers that contribute to information processing, and an output layer providing the results.

7.3.2 Synapses and weights

A synapse connects one neuron to another neuron. To add variation to the output from each neuron and to help the algorithm learn, every synapse has a weighting. This weighting is added to the output information, when the information is passed to another neuron that reflects the importance of the neuron in the network (Figure 7-2: An example of input weighting).

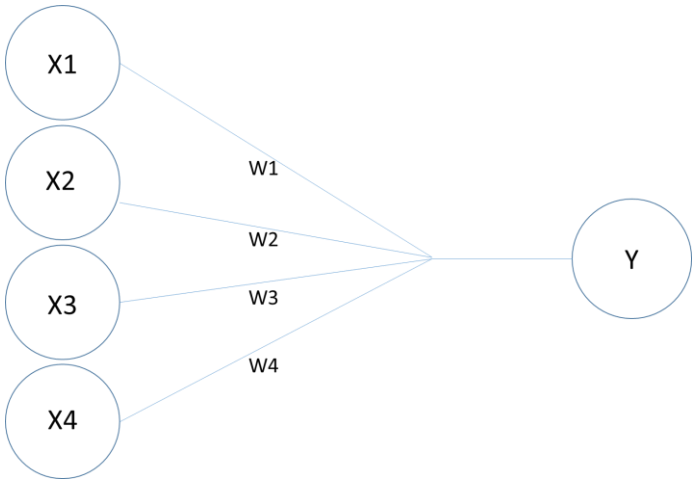


Figure 7-2: An example of input weighting

The amount of weighting guides the receiving neuron on whether to pass the information on to another neuron or not. Neurons with greater weights are dominant in how the receiving neuron

processes the information. During network initialization, weights are randomly assigned to each neuron, but as the system learns, the weights are updated by optimization.

Some neurons are designated as bias neurons. A bias neuron (b1 & b2) allows for more variation in the weightings stored in a network (Figure 7-3).

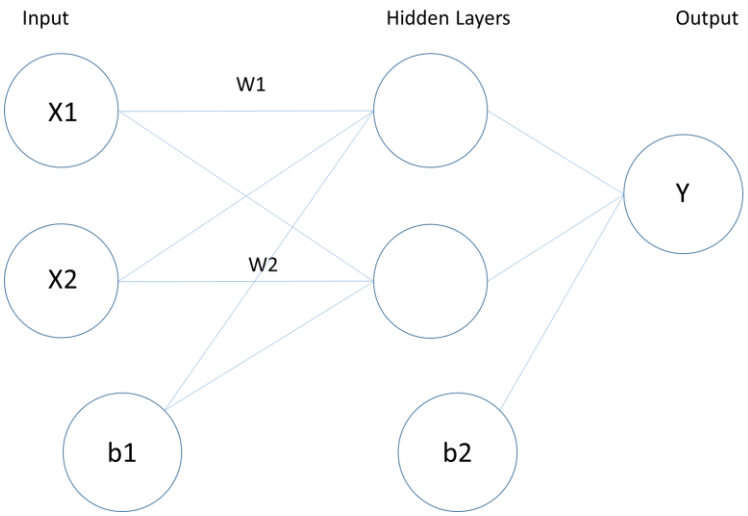


Figure 7-3:Demonstration of bias neurons in a neural network

Although a DNN can work without bias neurons, they are an indispensable part of a neural network model. They add a richer representation of the input to the model weights and produce some variation in the output. When a bias neuron is added to a layer, it modifies the weightings to make it more or less likely that the receiving neuron receives sufficient input to be activated to transmit its signal on to another neuron, or to be in another way, to make it possible for the activation function of a feature to be promoted or demoted (Figure 7-4). This will affect the algorithm output in either a positive or negative way, thereby helping the algorithm to learn. In the same way as the weightings for processing neurons can be changed, the weightings from bias neurons can also be changed. It is this iterative process that allows the network algorithm to learn and improve the accuracy of the final output to produce a high-performance model.

7.3.3 Activation functions

As noted, each neuron is completing a series of calculations based on the input it receives. This input may be from the image pixels or the weighted input from a neuron in another layer of the DNN. These calculations are called activation functions. With each calculation, there is an

activation threshold that allows the continued transmission of the input data along the processing pathway of the model. Promotion of the activation function accentuates the feature in the next layer, thus producing reinforcement of the feature for subsequent analysis. Combining the functions togethers produces an image transformation that describes the formula behind the DNN function.

There are a lot of activation functions, but the most common are linear, sigmoid, and hyperbolic tangent. Each function is looking for a different variation in textural pattern across the pixel cluster for each neuron. The main differences between them are the profile of change they respond to and the range of values they work with.

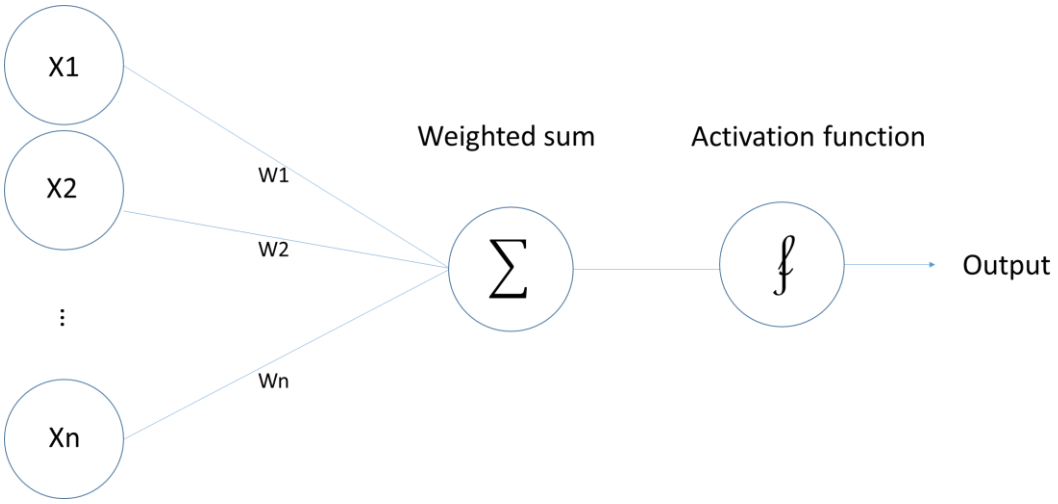


Figure 7-4: Demonstration of the weight assignment to each input and producing final function

7.3.4 Neural network training

Using the output from the neurons, the DNN calculates the difference or error (delta) between the input data and the output data. It does so by using two masks: an input mask (ground truth) and a output mask (predicted). The input mask may be manually labelled or the output from a previous iteration of the DNN. The predicted mask is the output from the DNN. There are two main ways to calculate the error: Arctan and Mean Square Error.

The amount of error reflects the deviation between the desired output (ground truth) and the calculated output, and it can be used to guide weighting changes for the next model iteration. In

this way, by repeatedly running the model, the prediction accuracy in the model is improved. The amount of error should decrease after each epoch (iteration). (An epoch is a term used to describe each training run through the entire set of training sets). The network layers optimize the weightings of the network through each epoch to reach a delta that is equal or close to zero, indicating that the network was able to predict the examples accurately.

Training based on only one image will produce an algorithm that is excellent at identifying the areas of interest in that image, but will be of limited value for other images. The intent is therefore to produce an algorithm that is generalizable to other similar images. To do so, a large number (batch) of images is used, with the output from the images all contributing to DNN learning. Batch size describes the total number of training examples used. While having a larger batch size will improve the generalisability of the model, batch size should also be considered according to the available memory size. The higher the batch size, the more memory that will be required.

7.3.5 Neural network types

DNN design is based around two main methods: feed-forward and recurrent. Feed-forward neural networks are the simplest neural network which doesn't need any memory. There is no backward movement in these types of networks, which makes it impractical for many tasks. Feed-forward networks can be used in supervised learning when the data are not sequential or time-dependent. The main advantage of feed-forward networks is that they are very fast and easy to build. Recurrent neural networks remember the results of the previous iteration and use them to make better decisions. These kinds of networks are widely used in natural language processing and speech recognition.

7.3.5.1 Convolutional neural network

Convolutional neural networks (CNN) are the standard method of deep machine learning (DNN) and are used to solve problems in many different fields (Figure 7-5). These networks can be either feed-forward or recurrent neural networks models. As described earlier, in CNN network architecture each neuron is connected to a small group of pixels in an image, each of which has the same weight. The network is given an image as an input, and each neuron in each layer performs a series of calculations on the pixel set to which it is linked. In this way, the pixel information in the image is transformed to produce a layer of neurons, each of which possesses a

weighted activation function. If the required activation function level is reached, the neuron is identified with this activation function. This process is called image convolution, and each neuron layer is called a convolutional layer. After each convolutional layer, the output from each neuron is assessed for a chosen activation function.

A CNN has the ability to learn from a large number of filters running in parallel in each convolutional layer that are specific to a training dataset under the constraints of a specific problem, like classification or segmentation. This produces an algorithm for a very specific feature that can be detected anywhere on the input image.

A key aspect of learning is the ability to summarize the neuron outputs after each convolutional layer to identify similar areas across the image. To do so, the output from all of the neurons in the layer is gathered together (pooled), ready to be passed through another convolutional layer. The output from this second layer is again assessed for activation function and the neuron output pooled. Pooling layers down-sample the output feature maps from each convolutional layer by summarizing the presence of features in each neuron. The two common pooling methods are average pooling and max pooling. Average pooling summarizes the average presence of a feature and max pooling summarizes the most activated presence of a feature.

A third component, included in the pooling stage, is the rectified linear unit (ReLU). This is a linear function that assesses the received activation function input from a neuron. If the ReLU is positive, then the input is directly transferred as neuron output, but if the ReLU is negative, then input value is reduced to zero. In this way, those activation functions that are identifying features of interest are accentuated, while those functions that are not identifying features of interest are removed from the algorithm. ReLu is a default feature of many networks because it makes the network easier to train and gives better performance.

After as many convolutional/pooling cycles as are necessary, those neurons that are linked together – they have similar activation functions – are identified. By the end of the processing, the network makes a prediction to find the output image class. In this example, the network has been asked to distinguish dog in another image.

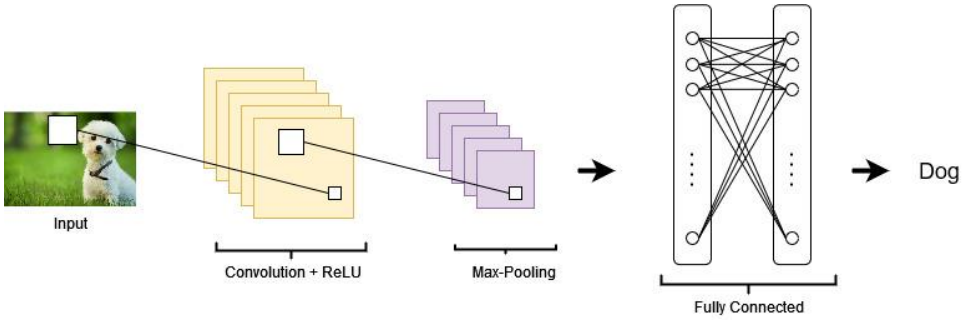


Figure 7-5: Demonstration of a DNN in a classification task

7.3.5.2 Encoder-decoder network

Higher level neural networks are composed of encoder, decoder, or encoder-decoder networks. An encoder network finds patterns in the raw data to form a useful representation, while a decoder network generates high resolution data from the representation. Encoder-decoder architecture is a higher-level concept that builds on the encoding steps to make a high dimensional output in the decoding step by up-sampling the compressed representations. The encoder and decoder parts of the network can be completely different in form from each other. Application of the encoder-decoder network includes semantic segmentation and machine translation. Figure 7-6 shows the general structure of an encoder-decoder network. In the left side of the network (encoder), the network consists of several convolutional and activation layers accompanied by pooling layers that down-sample feature maps to extract more detailed features. The right side of the network has several convolutional layers accompanied by up-sampling layers to return to original image resolution.

If the network is part of a larger model of combined neural networks that are considering multiple classifications, there will be a Softmax function at the end of the decoder functions. The Softmax function acts as the activation function for further neural networks.

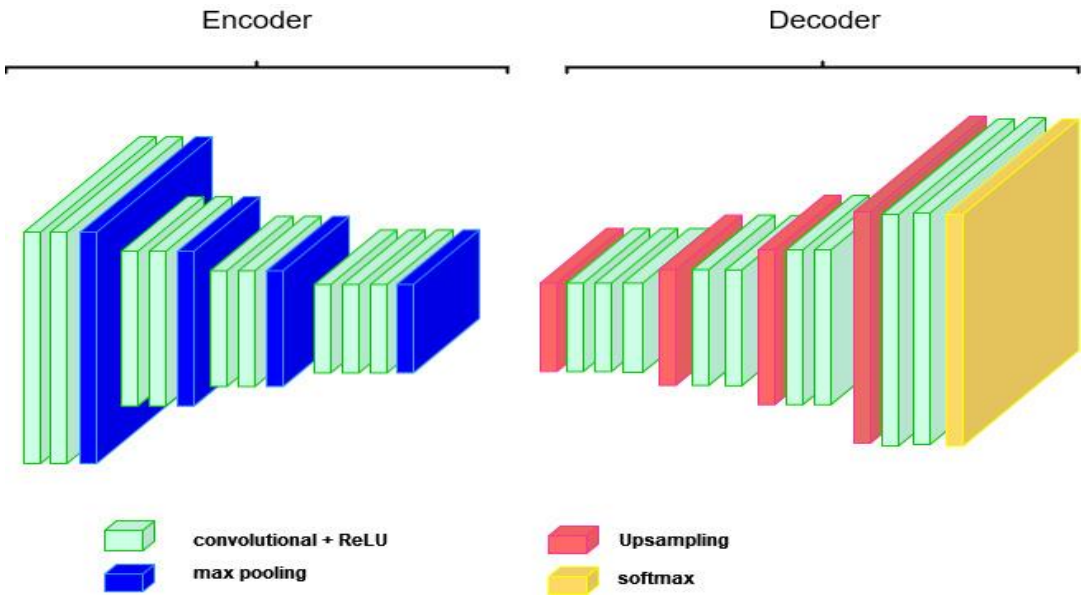


Figure 7-6: General structure of encoder-decoder network

7.4 Semantic segmentation

Semantic segmentation can be assumed as a classification problem in which each pixel in the image is classified with a semantic label, for example, human, sky, car, tree. The goal of image segmentation is to classify each pixel in the image with its corresponding class. Since the prediction is done for each pixel, the task is referred to as a dense prediction. Figure 7-7 shows an example of semantic segmentation.

In this example, there are 9 pre-defined classes: Kids, calendar, lamp, book, carpet, box, doll, chair, and background. The classes are labeled with different colors and the whole pixels in the image are classified into one of the nine classes. In semantic segmentation, there is no differentiation between objects in the same class and they are all shown with the same color. For example, all of the bicycles have the same green color.

Figure 7-7: semantic segmentation example

Semantic segmentation is also very useful in the field of medical image diagnostics. For example, the analysis time performed by a radiologist when reviewing a diagnostic scan can be reduced by only having to examine a specific part of the image. Figure 7-8 illustrates an example of medical image segmentation of a radiology image.

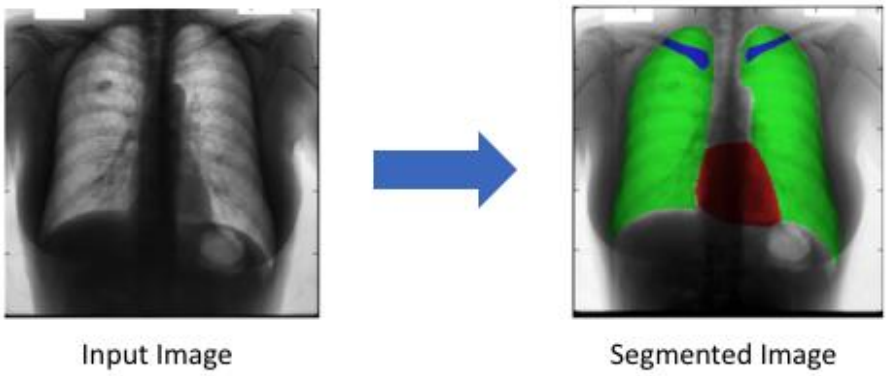


Figure 7-8: Medical image diagnosis with semantic segmentation[86](Reproduced from: Novikov AA, Lenis D, Major D, Hladůvka J, Wimmer M, Bühler K. Fully convolutional architectures for multiclass segmentation in chest radiographs. IEEE Trans Med Imaging. 2018;37(8):1865-1876.)

The goal of semantic segmentation is to take an image (dimensions W x H) as an input and then process the image to produce an output segmentation map (with the same W x H dimensions), where each pixel element has been given a class label. Figure 7-9 shows an example of semantic segmentation. The output has five different classes, and therefore each pixel in the image must be classified into one of these five different classes. On the right-hand side of the image, the pixels belonging to each class are labelled using a class number. The image is labeled as a low-resolution output map in this example for visual clarity, but the prediction map should be the same size as the input image.

Figure 7-9: representation of semantic segmentation task

7.4.1 Deep learning for semantic segmentation

The simplest way of a constructing a DNN for semantic segmentation is to stack a series of convolutional layers, each of which has the same padding to preserve dimensions, to produce an output segmentation map. In so doing, the network gradually learns to identify similar areas as the image passes through each layer. Image resolution is changed during the encoding/decoding stages, and this changes the image dimensions – padding adds columns or rows of black pixels to the image to maintain the same image dimensions. The black pixels do not affect the image processing. However, by requiring the preservation of the image dimensions, the network becomes computationally expensive. Figure 7-10 shows an example of a CNN structure.

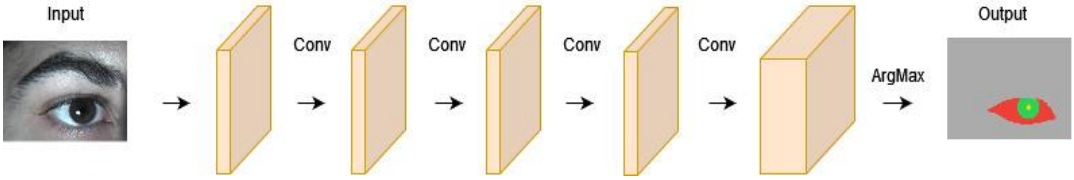


Figure 7-10: Structure of a CNN with padding to preserve image dimensions

CNNs are designed for image classification, but to reduce the computational expense of the network, the network can be customized and used for semantic segmentation.

7.4.2 CNNs for segmentation

In this kind of network, the neural architecture contains convolutional layers, non-activation layers, batch normalization, and pooling layers. Spatial information is not preserved. Spatial information is defined as the relative position of pairs of pixels in the image that can be defined by distance and orientation.

Figure 7-11 shows a CNN composed of several layers for the task of classification. The neurons in the initial layers look for and learn low-level concepts, such as edges and colors, for small regions of the image. The neurons in the next layers look for and learn higher level concepts, such as identifying different objects, from large regions of the image. Having identified these concepts, a process of down-sampling is done by pooling layers. The pooling layer output is then passed through another series of convolutional layers which perform further analyses to look for patterns within the pooled layer data. By adding more layers to the network, a greater refinement in the analysis can be undertaken that assists in network learning. Thus, with each new layer, the size of the image decreases, and the number of processing channels (image refinements) increases.

For the output classification task, the network looks for an output where all of the inputs from one layer are connected to every activation unit of the next layer. This will mean that every feature in the image has been categorised. However, fully connected layers destroy all of the spatial information. To preserve this spatial information, the spatial tensor (data describing the location of the feature in the image) in the final convolutional layer must be mapped to a fixed length vector. After learning, the vectors are used by the network to classify the objects in the image.

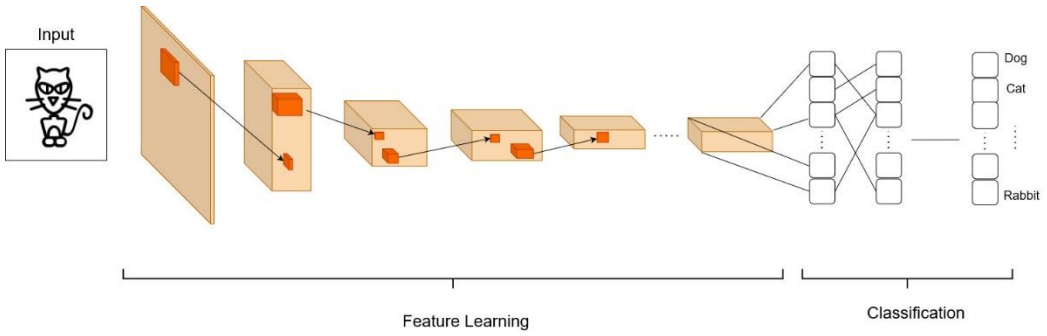


Figure 7-11: Spatial tensor is down-sampled and converted to a vector

However, for the task of semantic segmentation, the spatial information should be preserved. Therefore, while a fully convolutional network is used, the fully connected layer is not used within an encoder-decoder network. The encoder down-samples the input and the decoder up-samples the output from the encoder to return to the original image resolution. The down-sampling (encoder) convolutional layers create a low-resolution tensor that is used to produce high resolution segmentation output by adding up-sampling (decoder) convolutional layers to the network. The output of the encoder is a spatial tensor containing the shape and size of the object. The decoder takes this information and produces a segmentation map. The segmentation map has the same dimensions (width and height) as the original image.

If the encoder-decoder layers are stacked in a DNN, the low-level information from the original image will be lost. This will result in the final decoder for the network incorrectly identifying the segmentation boundaries produced in the output map. To resolve this error, the decoder is given access to the low-level features by using a skip connection.

With sufficient training images that are combined with their semantically labeled mask, the DNN learns a mapping between the training images and their masks. The learning process make the network aware of various semantic concepts that consistently relate to the images.

7.4.2.1 Base model

Different DNN architectures can be used to perform the semantic segmentation task, but all networks consist of a base model (sometimes called a standard model) for the encoder part of the

network and a segmentation structure for the decoder part. The output analyses of the image by the base model form the initial layers of the network and are used as an encoder for subsequent models. The rest of the network is built on top of the base model layer(s). For most applications, a model pre-trained on ImageNet would be an ideal case. ImageNet is a dataset of over 15 million labeled high-resolution images belonging to roughly 22,000 categories.

Commonly used base models include:

- The ResNet (Residual Network) base model software (Microsoft Inc, Redmond, WA, USA) is a multi-layer approach to developing a DNN. It is composed of a large number of layers (up to 152) along with residual connections. The multi-layer approach improves accuracy at the cost of a slower training time period. An accuracy of 96.4% was reported for the ImageNet dataset.
- VGG-16 is convolutional neural network base model, proposed by Simonyan and Zisserman from the University of Oxford, UK. It has less layers (16) compared to ResNet, which makes the training much faster. An accuracy of 92.7% was reported for the ImageNet dataset.
- MobileNet (Google Inc, Mountain View, CA, USA) is a smaller sized base model that enables a higher speed. The model was designed for memory and power constrained devices, such as smart phones. The accuracy of the model was reported as 70.6%.

7.4.2.2 Segmentation models

After selecting the base model, the segmentation architecture should be considered.

7.4.2.2.1 Fully convolutional network (FCN)

A fully convolutional network (FCN) was the first proposed semantic segmentation model, and it has become the standard semantic segmentation model. In this model, standard image classification base models, such as Resnet or VGG, were converted to fully convolutional models by changing the fully connected layers to 1 x 1 convolution layers that produces a probability map for arbitrary input size. The spatial information from down-sampling is recovered by adding up-sampling layers to a standard CNN. Different FCN versions are reported in literature including FCN8, FCN16, and FCN32. Figure 7-12 shows an example of FCN architecture.[87]

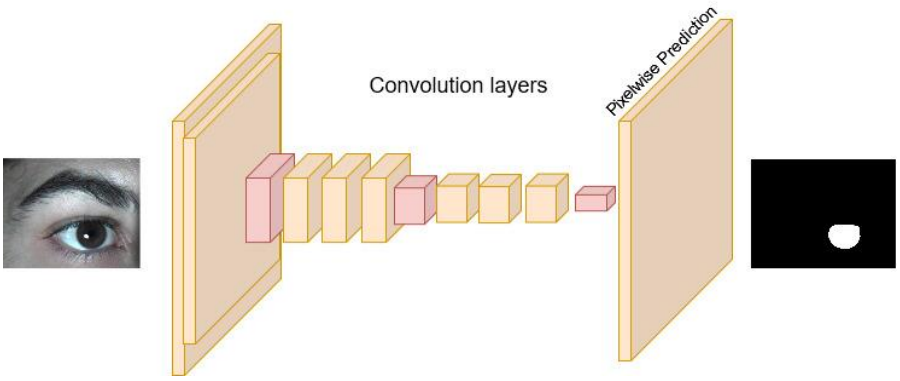


Figure 7-12: Fully convolutional network architecture

7.4.2.2.2 Segnet

The Segnet model is based on fully convolutional and symmetrical encoder-decoder layers. The trainable part of the network consists of an encoder that is identical to the first 13 layers of the VGG16 model. The decoder part of the network consists of pixel-wise classification layers. The indices of the max-pooling layers are used for the corresponding up-sampling layers to produce a sparse feature map, which uses convolution with a trainable filter bank to densify features. Figure 7-13 shows an example of Segnet architecture.[85]

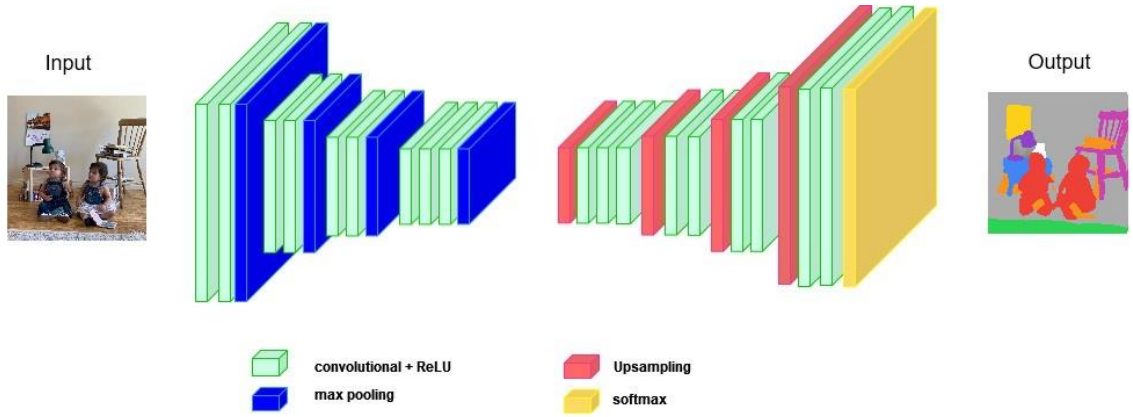


Figure 7-13: An example of Segnet architecture

7.4.2.2.3 U-net

The U-net network was first proposed for biological microscopy image segmentation. The network consists of two parts: a contracting path for capturing the context, and a symmetric expanding path which provides precise localization. The training strategy of the U-net is based on data

augmentation to learn from small size dataset. Figure 7-14 shows an example of U-net architecture.[88]

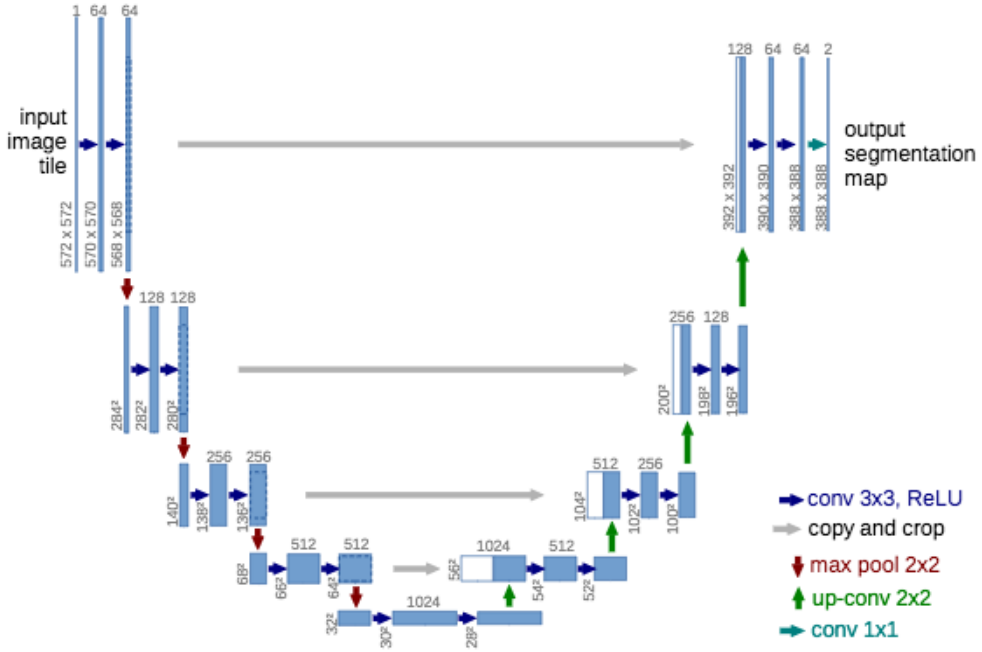


Figure 7-14: An example of U-net architecture [88] (Reproduced from: Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. 2015:234-241)

7.4.2.2.4 Pyramid Scene Parsing Network (PSPNet)

The Pyramid Scene Parsing Network (PSPNet) is a multiscale network optimized for learning global contexts in a scene. The Resnet network is used in this model as a base model feature extractor to extract multiple patterns. The extracted patterns are fed into a pyramid pooling module to separate patterns of different scales. The maps are pooled at four different pyramid levels and processed by 1 x 1 convolutional layers. The pyramid levels outputs are up-sampled and linked together by initial feature maps to detect both local and global context. At the end, the pixel-wise prediction is generated by a convolutional layer. Figure 7-15 shows an example of PSPNet architecture. [89]

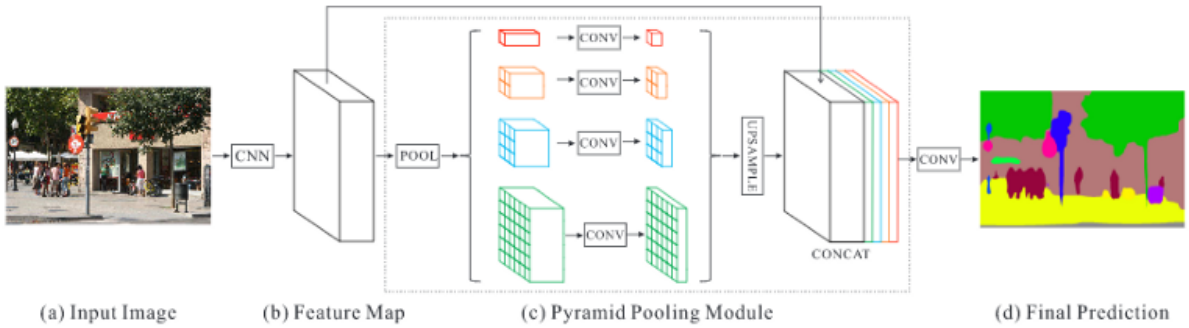


Figure 7-15: An example of PSPNet architecture[89] (Reproduced from: Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. . 2017:2881-2890)

7.4.2.3 Data Augmentation

One of the most important steps of each learning-based segmentation project is to gather enough images in the dataset, since a higher number of images in the dataset will produce more accurate results. Data augmentation is a way of increasing the number of labeled images in the sample dataset and is especially useful for small datasets, such as medical image datasets. This method increases the number of labeled images resulting in a higher performance for DNN segmentation models. In data augmentation, a set of transformations are applied to either the data space, feature space, or both. The typical transformations are translation, warping, rotation, flipping, resizing, color space shifting, scaling, and reflection. Data augmentation provides faster convergence, generalization enhancement, and less chance of overfit.

7.4.2.4 Transfer learning

A CNN trained for image classification produces meaningful information that can be used in the task of image segmentation. The pre-trained convolution layers of a network can be used in the encoder part of a DNN segmentation model. In this way, the accuracy of the segmentation can be improved significantly. Transfer learning is very useful when the number of training images is very low.

7.4.3 Metrics for image segmentation models

Different metrics are used to evaluate semantic segmentation models. Although multiple aspects, such as speed, storage usage, visual, and quantitative accuracy, should be calculated, most methods focus on quantitative measures of accuracy. The most frequently used metrics are described below.

7.4.3.1 Accuracy

The accuracy of the segmentation achieved can be calculated by using the total number of correctly detected pixels over the total number of pixels. It can be calculated using the following equation:

$$Accuracy = \frac{TP + TN}{P + N}$$

where *TP* represents the first class (true positive) pixels detected correctly, *TN* is the number of second class (true negative) pixels detected correctly, *P* and *N* are the total numbers of pixels of first class and second class in the ground truth image, respectively.

7.4.3.2 Mean Error Rate (MER)

Mean error rate (MER) was proposed in the NICE-I competition and is widely used in binary segmentation tasks. The error rate can be predicted by considering the ratio of all false predictions in the whole image. The error can be calculated using the following equation:

$$MER = \frac{1}{N} \times \frac{1}{w \times h} \sum_{x=1}^{w} \sum_{y=1}^{h} M(x, y) \oplus G(x, y)$$

where *N* is the number of testing images, *w* and *h* are the width and height of the test images, and *M* and *G* are predicted and ground truth masks, respectively. $\oplus$ is XOR operator calculating the dissimilar pixels between *M* and *G*.

7.4.3.3 Intersection over Union (IoU)

Intersection over Union (IoU) is an evaluation metric to compare the result of the predicted attention boundary with the ground truth boundary. It can be calculated using the following equation:

$$IoU = \frac{area(p) \cap area(G)}{area(p) \cup area(G)}$$

where *p* represents the predicted attention boundary, and *G* is the ground truth boundary.

7.4.3.4 Mean True Positive Rate (mTPR)

Mean True Positive Rate (mTPR) computes the average ratio of the true predicted pixels over the whole ground truth pixels. It can be calculated as follows:

$$mTPR = \frac{1}{N} \times \frac{TP}{TP + FN}$$

where N is the total number of the testing images and TP and FN are the true positives and false negatives, respectively.

7.4.3.5 F1 score

The F1 score evaluates the ability of the model to recall true positive pixels in the positive cases and not miss any cases. This metric is given by following equation:

$$F1 = \frac{2TP}{2TP + FN + FP}$$

where TP are the true positives, FN are the false negatives, and FP are the false positives.

7.5 Applying machine learning to iris segmentation

As reported earlier, the transparent cornea is the outermost anterior layer of the eye that covers the iris and pupil of the eye on the surface. However, by being transparent, identification of its boundary with the sclera is difficult, but an approximate boundary that is suitable for this project can be obtained by detecting and localizing the outer boundary of the iris. Different datasets and types of networks for deep learning-based iris segmentation have been reported in the literature.

7.5.1 Dataset used for iris segmentation

Using a DNN for image segmentation requires a lot of training data for the DNN to perform the task accurately. Different datasets are available for use in training an iris recognition system. The main problem with these datasets is a lack of ground truth. However, some of the groups have manually segmented the images and thus provided a ground truth mask. These masks are publicly available and can be used for research.

The UBIRIS.V*2* dataset (acquired in a visible light environment) consists of 11102 images from 261 subjects. Image resolution is 400 x 300 pixels. The images were taken on-the-move and at different distances by a high-quality visible camera. Therefore, the images are accompanied by different noises such as poor focused, off-angle, motion blur, eyelid occlusion, irregular specular reflection, and partial iris images. Two sub-sets of the database were used in the NICE-I and NICE-II competitions. The images were manually selected from the dataset by considering the high availability of noises in the images. In total, 1000 images were selected and used in both competitions.

Figure 7-16: Examples of UBIRIS V2 datasets[90] *(Image reproduced from: Proença H, Filipe S, Santos R, Oliveira J, Alexandre LA. The UBIRIS. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance.* IEEE Trans Pattern Anal Mach Intell. *2009;32(8):1529-1535)*

The CASIA.V*4*-distance dataset consists of 2567 near infrared (NIR) images taken from 142 subjects. Each image was captured using a high-quality NIR camera at 3 meters distance from the eye. The ground truth masks were labeled manually and are publicly available. CASIA thousand is a sub-set of the dataset containing 1000 low quality and noisy images for segmentation.

Figure7-17: Sample images from CASIA V4 dataset [91] *(Image reproduced from: Tan T, He Z, Sun Z. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition.* Image Vision Comput. *2010;28(2):223-230)*

The MICHE dataset was created to evaluate iris recognition methods for mobile devices. The visible images were taken by an iPhone5, Samsung Galaxy S4, or Samsung Galaxy Tab2 in uncontrolled conditions. The ground truth labels for the images are publicly available for the dataset.

Figure 7-18: Sample images from MICHE dataset[92] (Image reproduced from: De Marsico M, Nappi M, Riccio D, Wechsler H. Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols. Pattern Recog Lett. *2015;57:17-23)*

7.5.2 Previous methods of iris segmentation

A multi-stage fully convolutional network (MFCN)[93] was the first DNN iris segmentation model that could predict the respective iris and non-iris areas over the whole image. The MFCN was a fully convolutional network consisting of convolutional and pooling layers that could take arbitrary-sized images for input. The final convolutional layer was also able to up-sample the output of the previous layer to the size of the input image, so providing a mask for segmentation. The MFCN consisted of 31 convolutional and 6 pooling layers. Six layers with different depths from shallow to deep were used in the network structure to capture both local and global information. In order to have more accurate results for the segmentation, all of the local and global information was required. The detailed features were needed for those areas of the iris that were occluded by the eyelashes, specular reflection, or hair. Since both shallow and deep layers were used for the segmentation, a fusion layer was used to cover coarse to fine segmentation, which produced more accurate results. The designed model was an end-to-end network with no requirement of pre- or post-image processing.

After fine-tuning by the datasets, the MFCN produced an iris segmentation MER of 0.9% for the UBIRIS dataset, making it the first method to obtain an error of <1%. The results showed that the network outperformed all previous methods in terms of accuracy. The model was robust to illumination variations and noises, such as specular reflection, off-axis image, and occlusion. However, the method did not perform well on dark skin images and images without an iris. Also, some small areas were mis-classified and mis-localized in the images as part of the iris.

Arsalan et al., 2017, proposed a CNN-based method to detect iris boundaries (inner and outer boundaries) accurately in an unconstrained imaging environment in two stages. In the first stage, the rough iris boundary was obtained from the input image to define the region of interest (ROI) for the next stage. The resultant image from Stage 1 was a rough estimation of the iris boundaries that included parts of the upper and lower eyelids and other areas, such as skin, eyelashes, and sclera. In the second stage, the CNN was applied to the segmented iris area to provide more precise iris boundaries with the help of learned features. The approximate position of the pupil was also found by considering the standard pupil/iris ratio between pupil contraction and dilation. For network training, the pre-trained VGG-face model was fine-tuned by the data to produce a 100% training accuracy. However, the need for pre-processing of the images using non-CNN algorithms increased the whole image processing time.

Jalilian et al.[94] introduced a fully CNN for iris segmentation. The network implemented and trained using the stochastic gradient descent (SGD) back-propagation method. In SGD, instead of the whole dataset, a few samples were randomly selected for each iteration. Five different datasets, visible and NIR, were used in the training stage. The network was implemented in two different structures, namely basic variant and Bayesian variant. The Bayesian variant algorithm outperformed the basic variant in terms of accuracy. The method also compared with non-deep learning models on all of the databases. The results showed better performance of the proposed method compared to previous methods. However, the method did not perform well on the UBIRIS dataset. Consequently, the method is not suitable for low quality or noisy images, especially those taken by visible sensors.

In 2018 Arsalan et al introduced a method using a dense layer connection concept and an encoder-decoder structure named IrisDenseNet[95]. The model is a densely connected, fully-convolutional network that doesn't need any pre-processing and that uses gradient flow to prevent overfitting and vanishing gradients. The network was tested over challenging datasets and showed more robust results in noisy areas, such as eyelashes and eyelids. The network used both visible and NIR images in the training stage. The training accuracy reached 100% in the network. For the performance evaluation, the NICE-I evaluation protocol was used and the method was compared

to other methods, and an MER of 0.69% was found. The method out-performed previous non-deep learning methods with all datasets.

The IrisDenseNet segmentation was performed on different iris datasets and the output of the proposed network was compared with the SegNet network (one of the most successful DNN approaches in semantic segmentation) and was found to have a better performance. The reason that IrisDenseNet outperformed the SegNet network was the re-using of features by dense connections. The method was conducted on five different datasets and showed a higher accuracy compared with previous methods for both visible and NIR images. However, the dense layers means that it needs a powerful GPU and a large memory. Therefore, it is not an optimal method for low powered devices, such as mobile phones.

A fully convolutional DNN was proposed by Bazrafkan[96] to target low-resolution images taken in an unconstrained environment. In network training and testing, the CASIA, UBIRIS.V2, and MICHE datasets were used. To prepare suitable data for training, different augmentation levels were done on the datasets. To test the generalization ability of the network, the first two datasets were used in the training stage, and all of the datasets were used for the test stage. The network showed very good accuracy on NIR datasets and a good ability to segment visible iris images. However, the network did not perform well on very noisy images in the UBIRIS dataset, due to the network not being trained for noisy images. When compared to the SegNet-basic network, the SegNet-basic network had a better performance than the proposed network on the CASIA and Bath800 datasets. However, the proposed method outperformed the SegNet-basic while using the UBIRIS and MobBio datasets, showing a better generalization ability for the network.

The attention-guided U-Net was introduced by Lian[97] in 2018 and used two-stage learning. The network estimated an iris area in a rectangle, and this provided a weight map for the second stage. Using the weight map, the network was focused to learn more discriminative features of the iris. The UBIRIS.V2 and CASIA-IrisV4-Distance datasets were used for performance evaluation, and the attention mask estimation reached a high performance in the UBIRIS and CASIA datasets, which means the network was able to accurately assign the attention mask on the iris regions. A performance evaluation for the segmentation part of the network was done and compared with two

other networks. The results showed that the ATT-UNet network, with a performance of 96.8% and 96.3% for the UBIRIS and CASIA, respectively, surpassed the results for the MFCN and U-Net networks.

The IrisParseNet [98] is the most recent study of iris segmentation in which a multi-task learning framework was used to perform iris localization and segmentation. The model also provided a rough estimation of the iris boundary and reserved it for future steps. Two types of attention model, Atrous Spatial Pyramid Pooling (ASPP) and pyramid pooling (PSP), were compared together, with the ASPP achieving better results in the task of iris segmentation. The CASIA-distance, UBIRIS.V2, and MICHE datasets were used for the training stage. The network was compared with previous methods and the ASPP implementation of the network outperformed all of the non-deep methods and MFCN model. The method introduced a network to accurately segment and localize the iris area in visible and NIR images. All of the DNN methods showed better performance than the non-DNN methods. The results show that DNN methods, and especially FCN networks, can be used as an accurate method for an iris segmentation task. The model was time-efficient, and the overall runtime was less than 0.7 seconds. However, the model needed approximately 100MB memory, which limited its application on mobile devices.

7.5.3 Summary of previous methods

Generally, two different methods have been used for iris segmentation methods using DNN. The first method used end-to-end semantic segmentation, which can automatically learn the optimal features. This method does not need any user intervention and can take the whole image as an input and generate the output mask. There is no pre- or post-processing required in an end-to-end network. In the second method, some pre-processing was done on the images to prepare the images for the training. The pre-processing step segments the rough iris area on the images for network training and can be done by DNN or non-DNN methods. Using pre-processed images, the network has the advantage of learning more discriminative features. The network can be trained faster by the segmented area and the accuracy of the method could be improved. However, the overall speed of the model is affected by the number of stages. Both methods are not suitable for mobile devices with low power of processing or memory, which may be relevant for any future development of the system.

Some methods used encoder-decoder models to design an end-to-end network for iris segmentation. There are two paths in these types of networks. The encoder path of the network reduces the feature map size and increases the number of channels to encourage the model to learn local and global features. The decoder path of the network reduces the feature channels by up-sampling to return the image to the original size. By setting some short connections between the encoder and decoder blocks, the network can learn both local and global features. The designed networks based on an encoder-decoder structure are good candidates for iris segmentation task. These networks can learn some geometrical (global) features of the images accompanied by pixel level (local) features. These kinds of learning produce better segmentation results in noisy areas, such as occluded areas by eyelashes or eyelids.

The IrisParseNet method used a multi-task framework to detect the iris boundaries and region in a pixel-wise manner. The method showed better performance than the previous methods. Using localization and segmentation features together resulted in more accurate localization and segmentation separately. However, it was also found to have a high processing time, a high memory usage, and a higher CPU requirement.

7.6 Choice of iris segmentation method

For the task of cornea segmentation in this project, there are several points that were considered. First of all, by paying attention to the problem of cornea segmentation, an encoder-decoder model for designing an end-to-end network was chosen. The encoder-decoder model consists of a base model and a segmentation model. The first priority is accuracy of segmentation.

The base model with the highest accuracy should be chosen and since Resnet50 showed better accuracy on the ImageNet dataset, the base model of Resnet50 was chosen.

The next priority was to choose the decoder part of the network. Different models have been used for iris segmentation. Some networks work better with a noisy dataset and some work better on low-powered devices. However, the most important factor was to gain the highest accuracy. Since the system would process the data offline, processing time was not an issue. Also, the data was

gathered in a controlled illumination room, resulting in less noise in the data. Nevertheless, the cornea and iris appeared in different shapes and sizes in the images, and the iris appeared in different colors. For example, the cornea is not circular when the eyelids are not open wide. Therefore, both local and global features were very important in the segmentation stage of the project. The best network for the segmentation was PSPNet (Pyramid Scene Parsing Network) because it learned global features at different scales.

Thus, the model for the segmentation part of the project was Resnet50-PSPNet pre-trained on the ImageNet dataset. Although the processing time of the files was a bit longer than other networks, it had better accuracy. The accuracy of the FCN32, SegNet, U-Net, and PSPNet were calculated on the sample dataset with same data augmentation technique to choose the best model. Table 7-1 shows the segmentation accuracy using out dataset.

Table 7-1: comparison of the different segmentation models

Model	Mean IoU
FCN32	89.2
SegNet	92.3
UNet	91.7
PSPNet	94.6

Among the models, PSPNet had the highest accuracy for iris segmentation and chosen in this work.

7.6.1 Project dataset

A total of 160 images of the eyes of 10 subjects were captured by the Teledyne FLIR BFS visible camera to produce the dataset used in this project. A total of 250 images were taken of each subject's eye with their gaze directed straight ahead, upwards, downwards, to the left, to the right, and with the eyelid closed and then 10 numbers of the images from each subject were selected for network training and testing. Figure 7-19 shows some sample images from the dataset. The images were captured in a controlled illumination and normal room temperature. The subjects were asked to move their eye and blink naturally during images were taken. For the training step, 120 images were used, and for testing the network, 40 images were used.

As the total number of images for training purpose were low, data augmentation was used in the training step. Both feature and data transformation were used in data augmentation step. The augmentation methods used were: translation, warping, rotation, flipping, resizing, color space shifting, scaling, and changing contrast.

Data annotation was done to create segmentation masks for the data set images. paint.net was used to produce black and white ground truth masks manually, where the white pixels represent the cornea pixels and black pixels represent the non-cornea pixels (Figure 7-20).

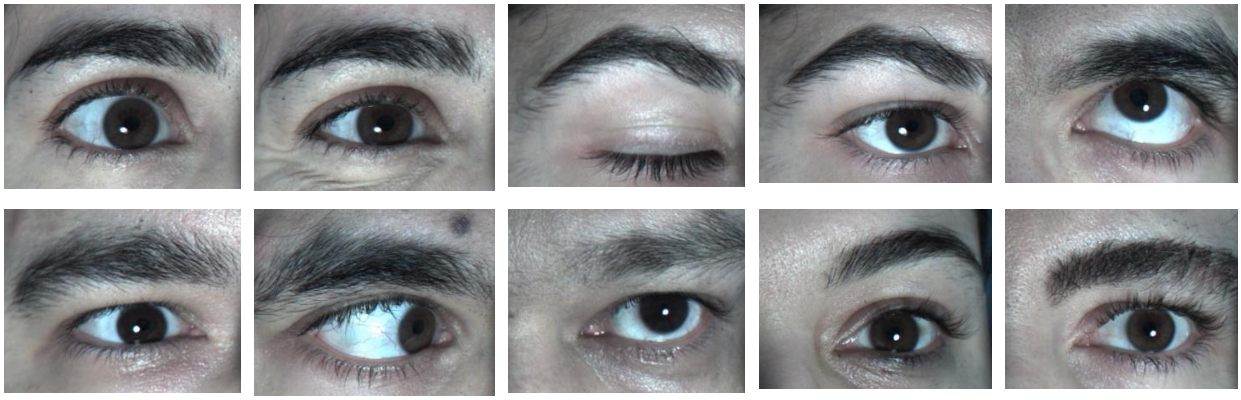


Figure 7-19: Sample images of the eye in different directions of gaze from the training dataset.

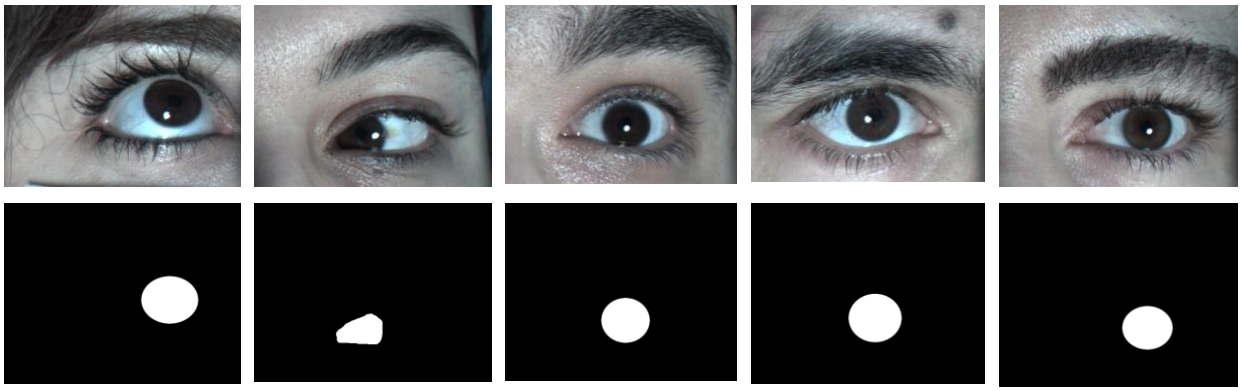


Figure 7-20: Sample images of the eye in different directions of gaze from the training dataset (top) with their respective ground truth masks (bottom).

7.6.2 Iris segmentation results

After training the network using the dataset, the model was created and used for prediction. The 40 images not used in the training step were used for the model test and performance evaluation.

Figure 7-21 shows the result of prediction by the network on 15 sample images. The yellow areas show the cornea pixels detected by the network. After testing the network on all 40 test images, a mean IoU of 94.6% was calculated.

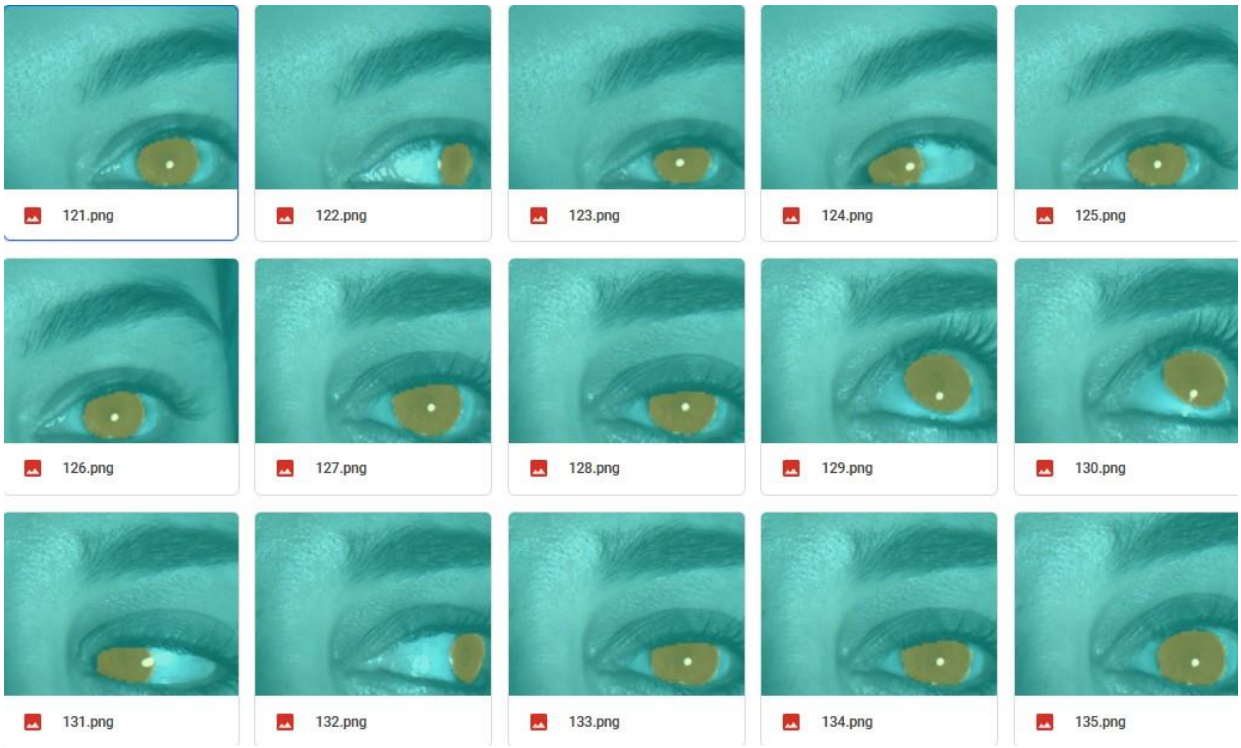


Figure 7-21: Colored predicted masks produced by the network for 15 sample eyes in different directions of gaze from the test dataset.

7.7 Blink detection

When analyzing the video files, artifacts may arise due to the presence of a blink within a frame and these must be removed. Considering again that the purpose of the system is to analyze the surface temperature of the cornea, it is essential to exclude any temperature data from the eyelid that might also be included for analyses when the eyelid is closed. Hence, the blink frames must be removed from the video files.

Fortunately, the segmentation network could also be used as a blink detection tool. When the iris segmentation network failed to detect a cornea in the image (because of the presence of the eyelid during the blink), the network could distinguish between blink and non-blink frames, and the relevant frames removed from the video sequence. Figure 7-22 shows a series of example masks predicted by the trained network.

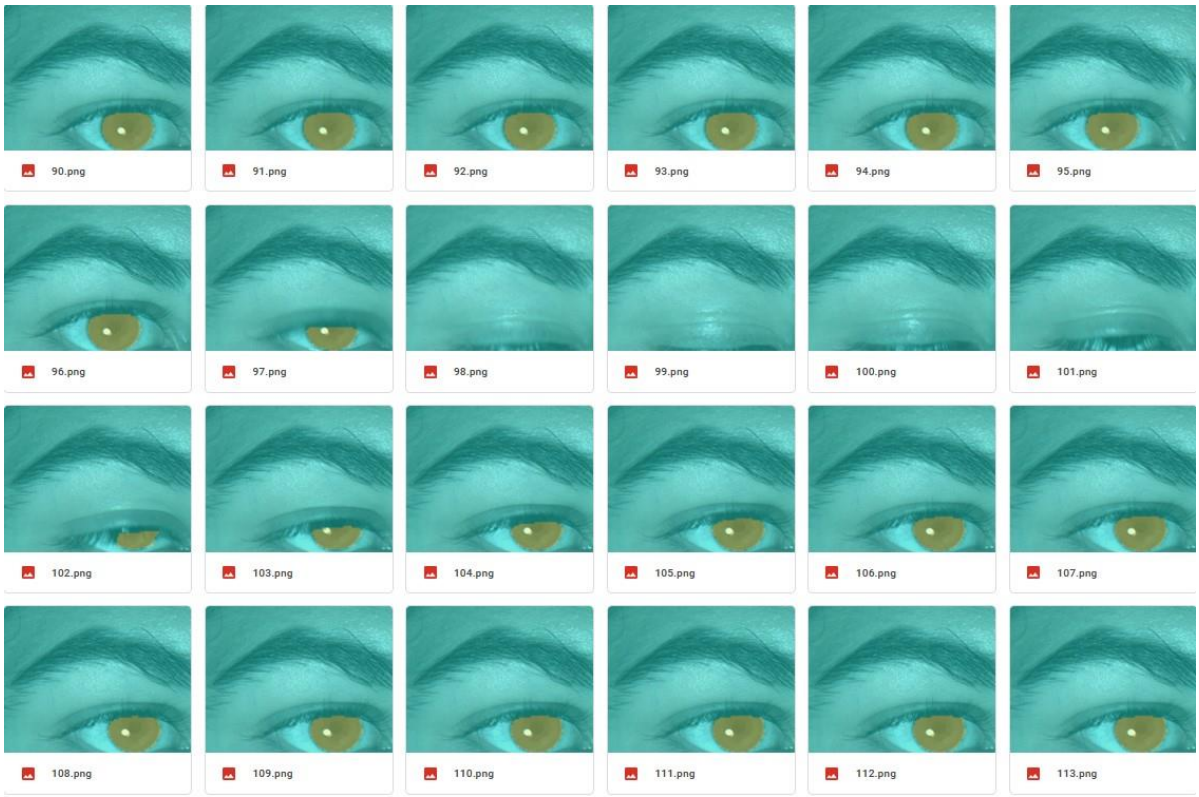


Figure 7-22: predicted masks by the network including video frames in which the subject has blinked

Figure 7-23: BW masks generated by the network Figure 7-23 shows the black and white ground truth masks generated for this image set.

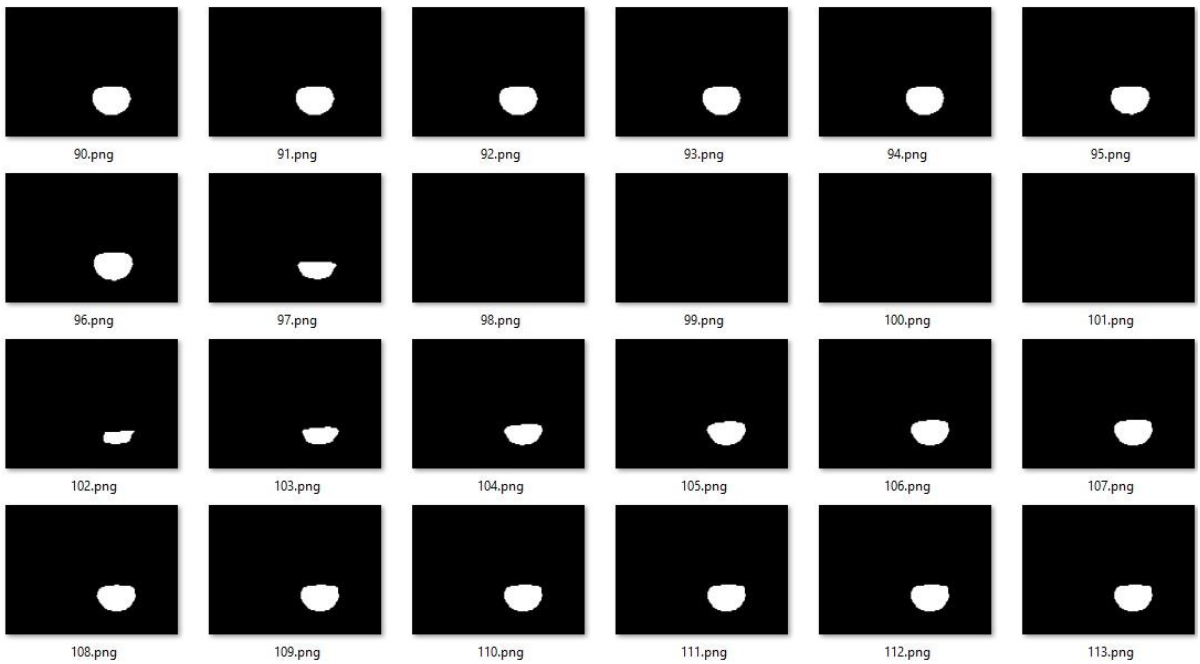


Figure 7-23: BW masks generated by the network

For some frames, the masks were totally black (frames 98-101), which meant that the cornea was absent. Therefore, these frames could be considered as containing a blink. To assess the accuracy of the algorithm to perform this task, two video files, each with 250 frames, were assessed. The two video files had 1 and 2 faulty blink detections over 250 frames, respectively. Figure 7-24 shows an example of two blink frames from the video in which the cornea was detected by mistake.

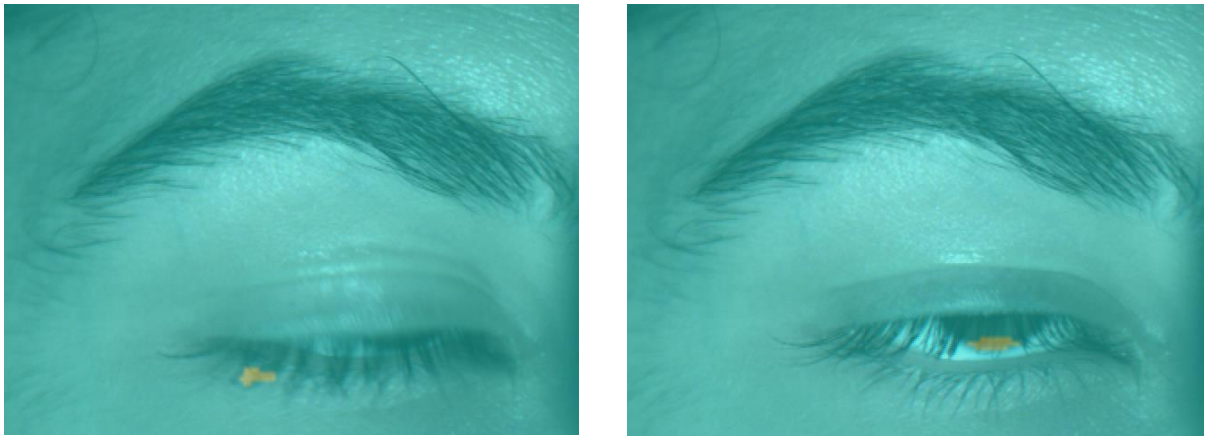


Figure 7-24: An example of the misclassified corneal area in the blink frames

7.8 Eye tracking and Final image processing outcome

A second possible artifact during the measurement was associated with eye movement. One of the major problems of previous systems was the lack of a method for eye tracking and localizing the cornea during eye movement. To track the cornea in a video sequence, the cornea must be localized in each frame of the sequence.

As demonstrated, the cornea could be localized in all of the video frames by using the iris segmentation network to predict the corneal area. When the cornea was localized in all of the frames, the segmentation masks that were produced could be used for eye tracking. In this way, the corneal area could be tracked in both the visible and thermal video files ensuring that surface temperature could be tracked and extracted from the only the cornea for further analysis.

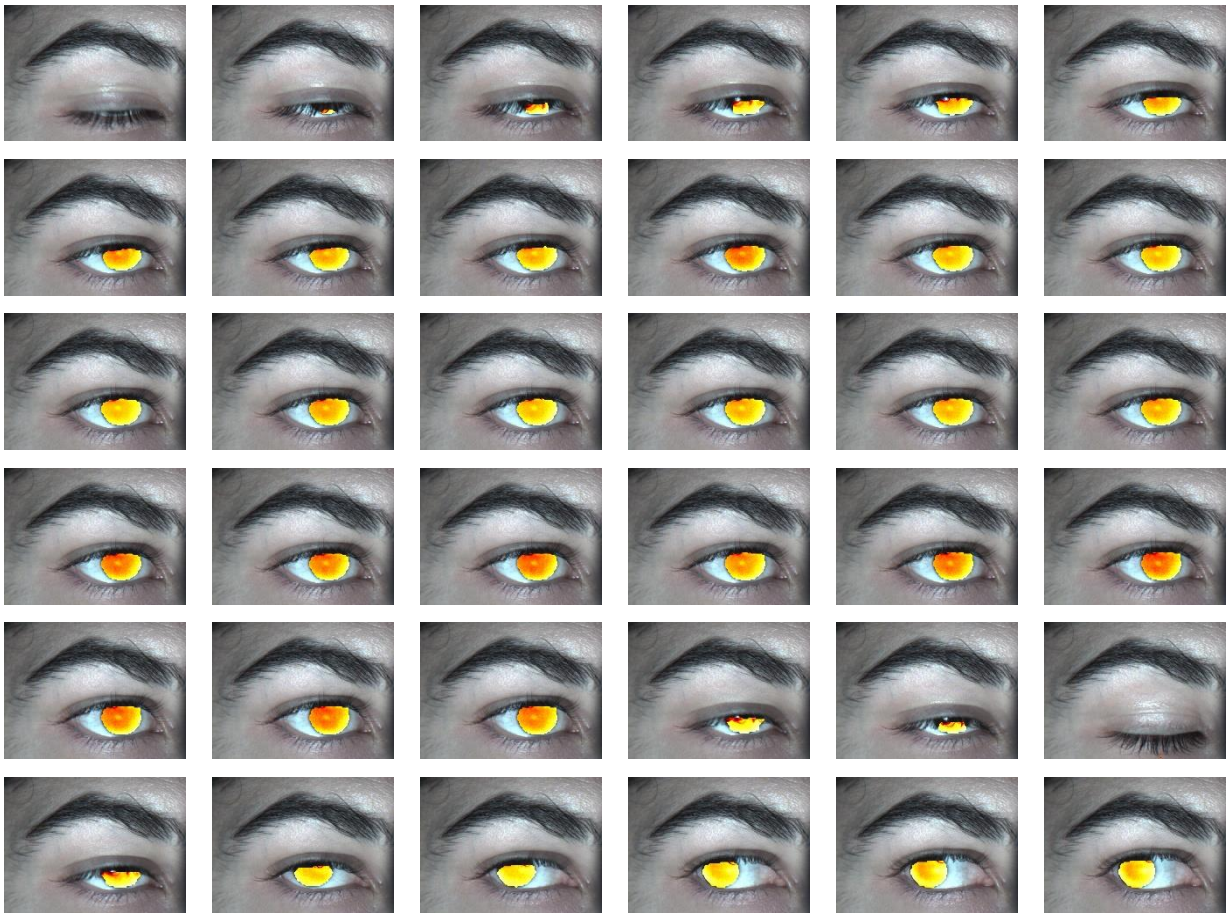


Figure 7-25: A series of frames taken from a sample video file between two blinks

Figure 7.24 shows a sequence of mapped video frames after the whole algorithm implementation. The figure shows extracted consecutive frames from the video file between two blinks. The temperature profile is mapped on the corneal area which gained from the thermal image. The temperature changes are demonstrated by color change in the images. There is no temperature mapped on the blink frames, which means there is no cornea detected in the frame. The last row of the figure demonstrates the eye tracking by the proposed method. When the cornea was moved in the frames, the ROI is segmented correctly, and the temperature profile is mapped on the ROI.

Chapter 8: Data analysis and system application

The system has been shown to be able to register image and video files, to segment the cornea and to deal with eye movement artefacts. However, for the system to be practically useful, it must also be able to extract OST data from each frame and record it in a file for data analysis. This now reveals a new challenge, since each frame contains around 12000 pixels (temperature points) contained within the segmented corneal area, and the data should be analyzed to be useful in clinical application. This is a previously unobtainable level of detail for the cornea, but which also possesses great potential for understanding how the OST varies across the cornea between blinks and in various clinical situations. The full investigation of this data output is beyond the scope of this thesis, but some simple analysis can reveal the potential.

Taking the video sequence illustrated in Chapter 7 as the data source, a series of examples is presented: single point, multiple points, and the average of all data points across the cornea.

8.1 Single pixel point

A single point can be selected on the corneal surface to track the temperature over time. Figure 8-1 shows an example of the temperature trace from a single pixel point location in the center of the cornea. The presence of the blinks is revealed by the positive or negative temperature spikes. These temperature spikes can be detected by the eyelid blinking algorithm and removed from the trace (Figure 8-2).

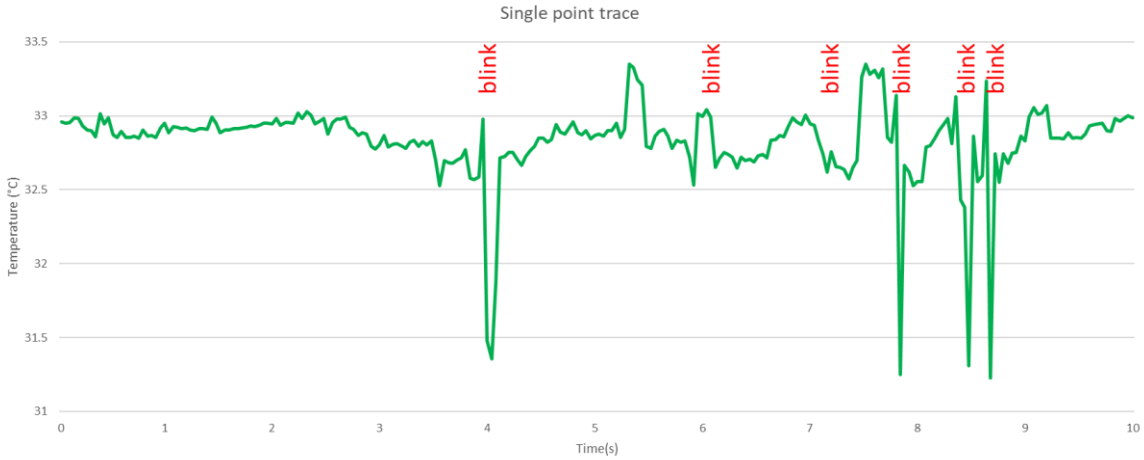


Figure 8-1: An example of single point location

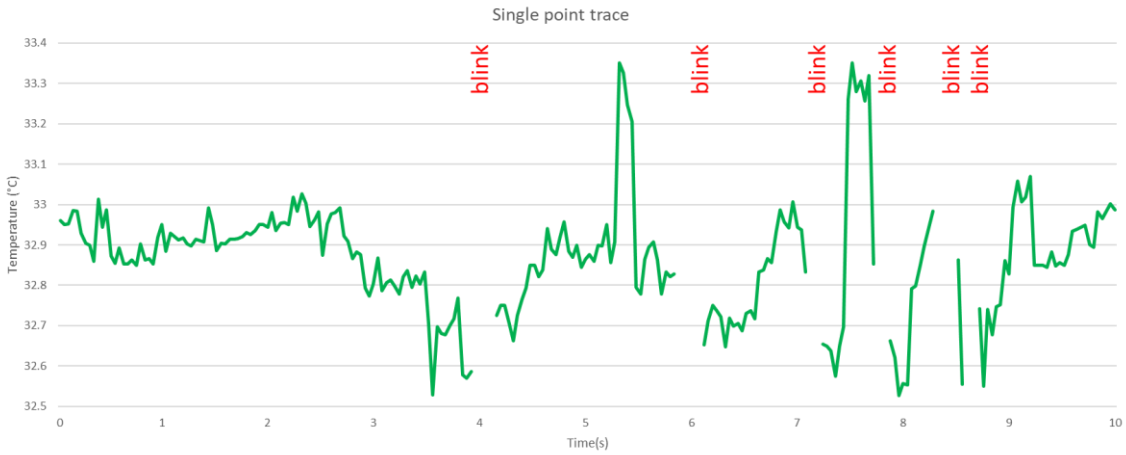


Figure 8-2: Single pixel point temperature tracking over a period of 10 seconds. Recording began immediately after a blink.

Considering Figure 8-2, it is interesting to note how much noise there is in the measurement. This might be due to eye movement or to poor thermal sensitivity in the camera (unlikely since the thermal camera resolution is very good). For this video sequence, the subject was asked to move their eye so that the video tracking and corneal segmentation could be tested, but it may be better to encourage the subject to maintain constant fixation. Alternatively, it may be better to record from a small cluster of pixels (e.g., a 9x9 square) to produce an average of several points. This is a question for further investigation.

8.2 Multiple pixel points

Another way to analyse the data is to select single pixels at different corneal locations and compare the temperature profiles from each location (Figure 8-3). In this way, the OST change can be tracked and compared at different locations.

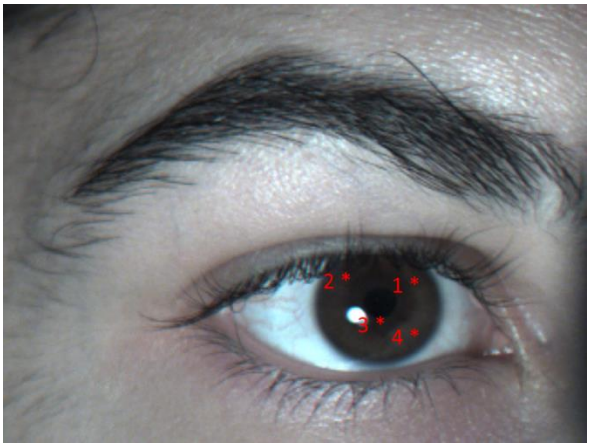


Figure 8-3: Selected pixel location for multi points tracking

Figure 8-3 shows an example of the temperature profiles produced by tracking four locations on the corneal surface. The two central corneal locations (Points 3 and 4) are cooler than the two superior corneal locations (Points 1 and 2). This follows the previously published norms for OST, where the central cornea is known to be cooler than the peripheral cornea, which is assumed to be due to the closer proximity of a warming blood supply for the peripheral cornea. This effect of peripheral corneal warming can be observed by comparing Points 3 and 4, where a cooler temperature is recorded for Point 3 which is located at the approximate corneal center and thus farther from the warming limbal blood supply.

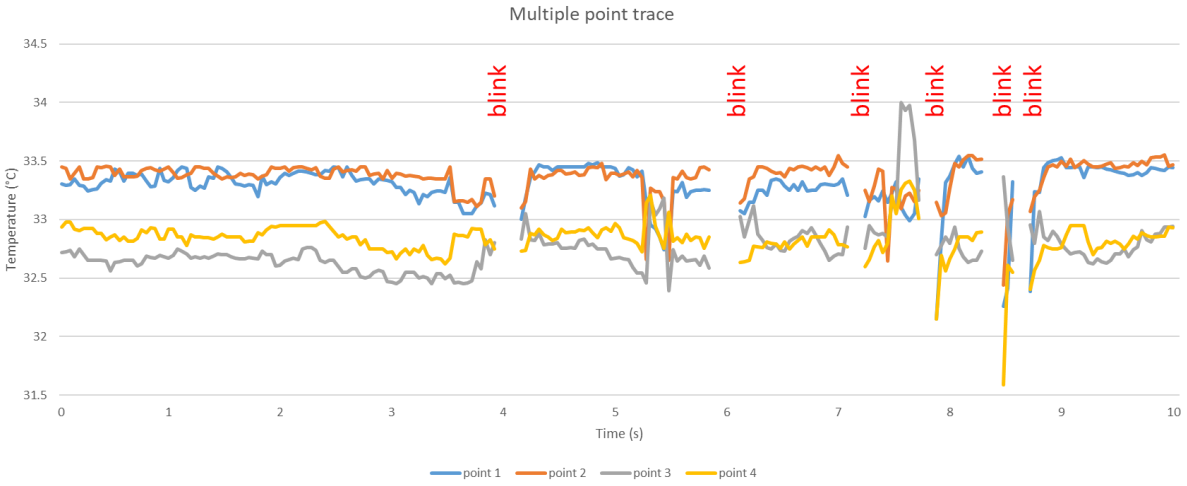


Figure 8-4: Multiple pixel point temperature tracking over a period of 10 seconds. Recording began immediately after a blink.

8.3 Selected region of interest

The system is able to track all pixel points within a selected ROI on the corneal surface and produce an average across all of the frames in a video sequence. Figure 8-6 shows the average temperature within a selected area over the center of the cornea, tracked over time. The benefit of averaging the temperature can be seen in the generation of a smoother plot by the removal of the effect of signal noise. The presence of the unusual sharp dips in temperature are unexpected, but may be due to small errors in the eyelid blinking algorithm that permits images that include the colder eyelashes to be included in the trace. Once again, further investigation and refinement of the system is needed.

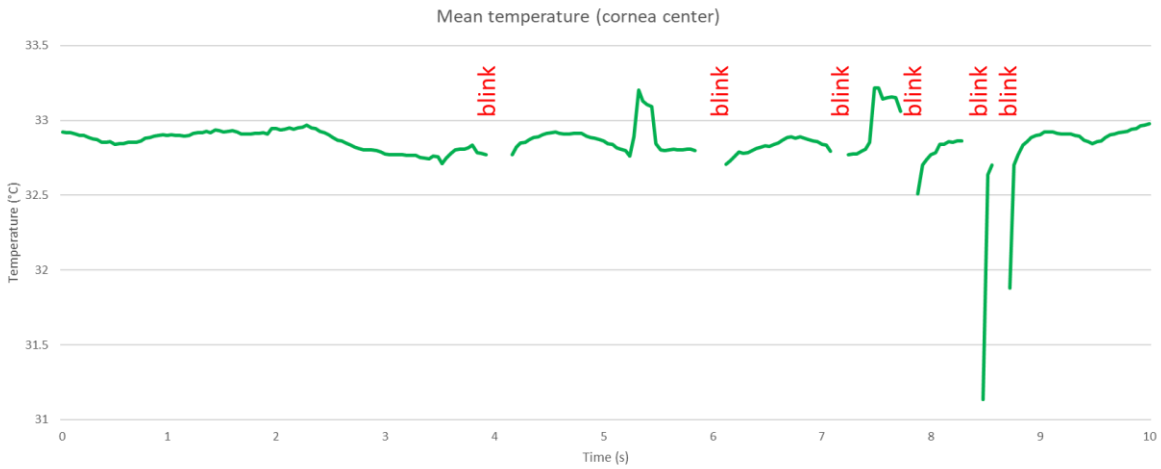


Figure 8-5: Average temperature tracking for a selected region on the cornea over a period of 10 seconds. Recording began immediately after a blink.

8.4 Whole cornea

The developed system is able to detect corneal area and track it on all of the frames. Therefore, all of the pixel points can be tracked together. Figure 8-6 shows the trace produced for the average temperature tracking over the whole cornea.

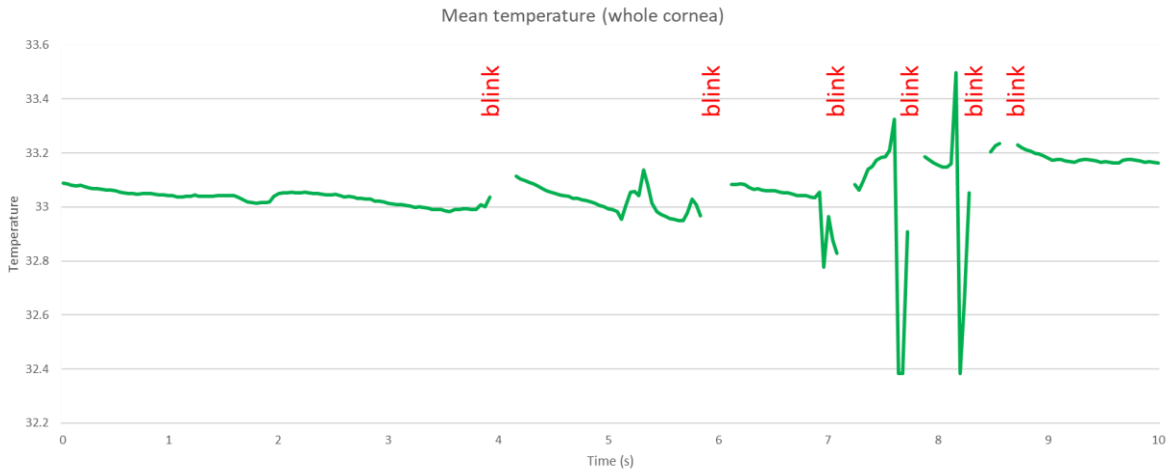


Figure 8-6: Whole cornea average temperature tracking over a period of 10 seconds. Recording began immediately after a blink.

8.5 Sample application

One possible application of temperature tracking is to look at the rate of cooling in the tear film over the cornea between blinks. This cooling is thought to relate to tear stability, and thus the rate of cooling maybe a useful clinical biomarker: a steeper slope would reveal a faster cooling rate and a less stable tearfilm. Figure 8-7 shows the average temperature change over the first 4 seconds after a blink for the whole cornea. The slope of the best fitting line from the start of the trace to just before the first blink is 0.0008. In contrast, if the temperature change slope is calculated from after the first blink to the temperature spike at c.5 seconds, the slope is calculated as 0.002 (Figure 8-7).

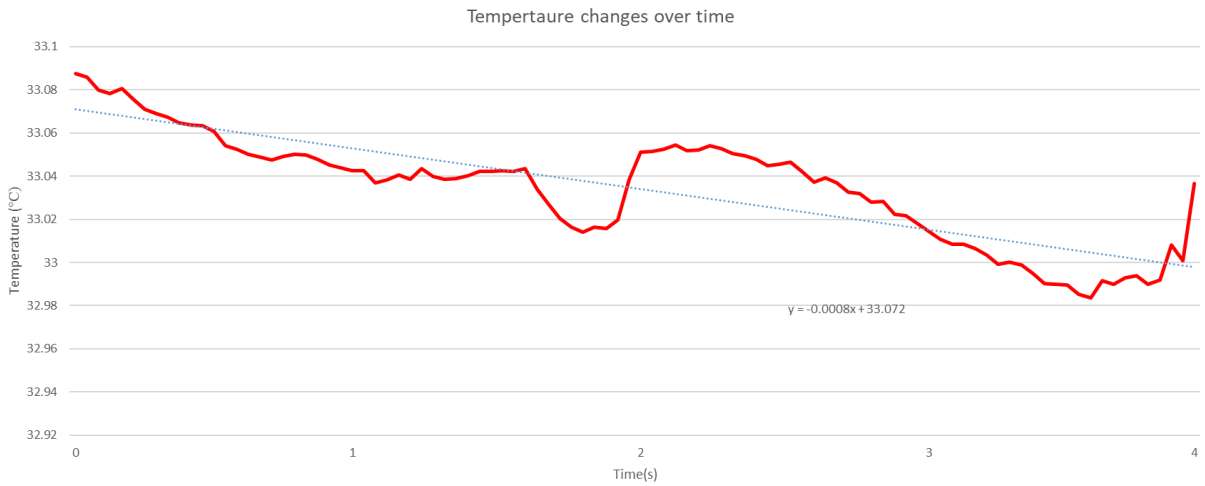


Figure 8-7: Colling rate before blink

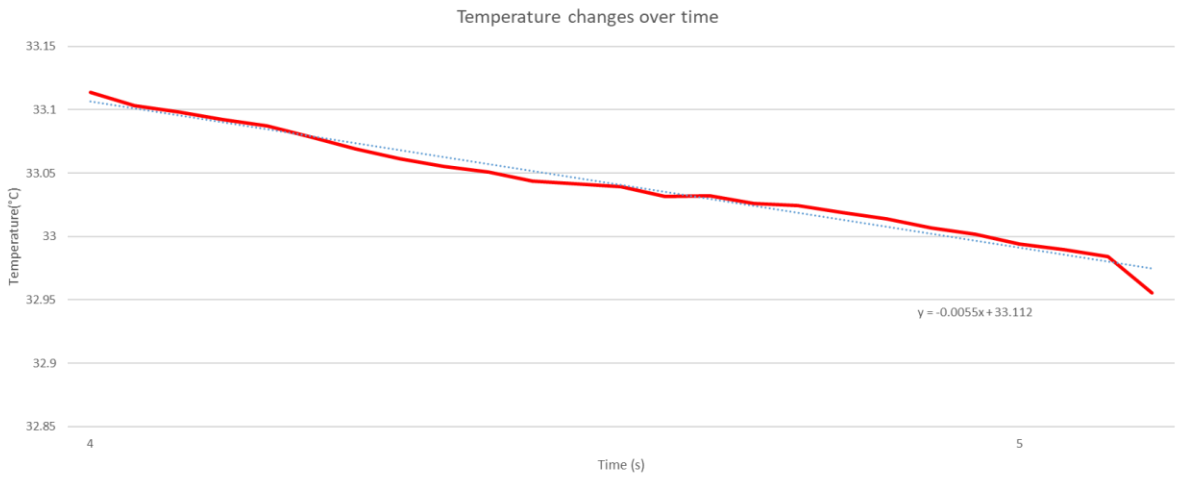


Figure 8-8: cooling rate after blink

The faster cooling rate after the first blink suggests that the tear film is less stable, and is affected by increased tear evaporation, leading to a new blink being triggered, which happens at approximately 6 seconds.

Chapter 9: Discussion and conclusion

This thesis reports on the development of a novel system and method for imaging, segmenting, temporal and spatial tracking, and analysis of visible and infrared images of the ocular surface and eye adnexa. For the first time, a complete system for imaging the ocular surface, synchronizing video sequences, segmenting the cornea, and extracting ocular surface temperature (OST) data from the eye has been developed. US patent (17/236,816) and Canadian patent (#3,116,492) applications have been made for this system.

9.1 Current situation

OST measurement has the potential to provide useful information about ocular surface health. For example, knowledge of the OST can help a physician to diagnose some eye diseases with much improved accuracy.[25-32] It can also be used for a better understanding tear film quality by looking at changes in OST associated with tear film break-up.[33]

However, current methods for assessing OST clinically are severely limited. No commercial systems for assessing OST are available and researchers have instead relied on customized instruments. Of the instruments available, all are designed with either one camera (thermal) or two cameras (one thermal and one visible). The dual camera systems are designed to use the visible camera to help overcome the inherent lack of spatial resolution across the thermal image that makes precise identification of the corneal boundary very difficult. By overlapping the field of vision of both cameras, the corneal boundary seen in the visible camera image can be used to locate the corneal boundary in the thermal image. This lack of spatial resolution in thermal cameras has also meant that researchers must rely on manual selection of individual points or areas of interest on the ocular surface, and to a very restricted set of options in terms of data analysis. Data analysis typically means a comparison over time between individual data points on the ocular surface or between average temperatures within selected areas on the surface across the ocular surface. When one considers that the pixel count across the cornea in the thermal image typically exceeds 12000 pixels per frame of a video sequence, this simplified manual approach inevitably leads to a huge amount of detail being lost.

There is a clear need for a new approach on instrument design and analysis for successful OST measurement.

A series of requirements can be described:

1) The ability to locate the corneal area in the thermogram by including a method for automatically detecting the corneal boundary, rather than requiring the input from the operator to manually select the point or area of interest.

2) The ability to consistently measure from the same location on the ocular surface by tracking and compensating for any head or eye movements. These movements cause relative movements in the areas of interest on the eye and ocular surface during the period of measurement, which degrade the accuracy of measurement over the period of measurement.

3) The ability to track OST changes over time by removing the effect of artefacts in the temporal temperature profile caused by eyelid blinking during a period of measurement.

4) The ability to collect and analyse temperature data from all pixel points across the ocular surface within the image frame over the period of measurement.

5) That all four previously listed requirements should be completed automatically.

No current system for IR imaging of the eye can automatically complete any of these four requirements. Hence, a system to solve these problems was required.

9.2 A novel designed system

In this thesis, a novel combination of instrument hardware design and customized software algorithms was developed to capture a synchronous image/video sequence of the eye surface from two cameras from which the corneal could be tracked in both video files. With the corneal area segmented, the system is capable of measuring and tracking OST at the individual pixel level of detail across the cornea, over any chosen period of time.

9.2.1 Hardware development

The first step in the system development was the physical system design and camera selection. Both cameras needed to have close-up lenses with good fields of view (FOV), high pixel resolution (640x480 pixels), and a high frame rate speed (25 fps) to enable good imaging of details on the

eye surface. The FLIR-Teledyne IR A655sc and FLIR-Teledyne BFS 51S5C cameras were identified as meeting these imaging parameters. Both cameras can support a faster frame rate, but a faster frame rate would have supplied more data which would have needed more processing time and a more powerful processor. 25 fps provided enough data for analysis. The visible camera had a greater FOV than the thermal camera, and so the FOV was cropped to limit the image area to a similar FOV as the thermal camera.

The next requirement for the system installation was to choose a suitable camera base/mount. Slit-lamp biomicroscope are designed with very flexible controls to allow easy alignment of the instrument with the face and eyes of a subject. This flexibility in alignment was ideal for mounting the two cameras, and for positioning the subject. The slit-lamp includes a chin-rest and head-rest for the subject, and two moveable arms on which a binocular microscope and an illumination system are mounted, respectively. For this system, an old slit-lamp was modified by first removing the microscope and illumination systems having, and then a custom mounting was attached to the microscope arm. The camera mount allowed both cameras to be mounted beside each other to ensure a consistent relative angle of photography between the cameras. This mount also gave the ability to adjust the camera positions such that the thermal camera was positioned perpendicular to, and directly in front of, the eye to be imaged.

9.2.2 Algorithm development

A customized algorithm was required to manage the cameras and capture the images. To that end, a software program based on the developed algorithm was developed to control the operation of the cameras and to record image/video from the eye surface using the Python programming language. A graphical user interface was designed to enable simple operation of the hardware. A further series of customized software algorithms (I-VII) were then developed for more detailed control over the hardware and for image processing of the image/video output from the two cameras.

I. With the two cameras linked under software control, the first step was to synchronize their operation. The two cameras was configured in such a way that one of them acted as the primary camera and the other as the secondary camera. To do so, the cameras were connected via a cable

for hardware triggering, with the thermal camera acting as the trigger for the visible camera. A software was then developed for synchronous recorded of the video files/images from the eye surface. The software sent a command to the primary camera to capture an image. The hardware within the camera was activated to capture an image, and simultaneously sent a synchronous signal to the secondary camera to capture an image, and the two images were captured synchronously. Hardware triggering ensured that there was no delay in image capturing between the two cameras. This is an important step, since the corneal limits in the visible image were used to locate the corneal limits in the thermal image. It was therefore essential that the image from each camera represented the same moment in time.

II. The captured images of the eye surface were taken using two different cameras, which meant that they had different sensors and optical qualities. It also meant that only one camera (thermal) could be mounted perpendicularly to the eye, leaving the visible camera with an oblique observation point relative to the eye surface. This latter effect produced a distortion in the visible camera image. Therefore, the images from each camera are not similar and the pixel coordinates of the visible image cannot simply be used to locate the cornea in the thermal image. An algorithm was developed that compared the gathered image from both cameras of a test object with known dimensions, in order to identify the presence of any distortion, and to then transform the images to remove that distortion. This calibrated each camera to produce an optically true image of any object.

III. To be able to use the corneal coordinates from the visible image to locate the cornea in the thermal image, the images must be mapped on top of each other. This technique of image mapping is called image registration. The developed algorithm for image registration used corresponding salient points in both images to calculate the required transformation between the images necessary to warp the thermal image and allow it to be mapped onto the visible image.

IV. In this project, two video files were recorded from the eye surface with the purpose of extracting the corneal OST over a period of time. Having developed algorithm for image registration for a pair of timestamped matched images, the next step was to develop an algorithm that was capable of continuing image registration for each subsequent timestamped matched

images in the video sequences for each camera. As noted, image registration algorithm uses corresponding salient points in each image to calculate the required transformation between the images. In video registration, these points must be continually selected on each new pair of corresponding frames. Manual control point selection for each image pair is cumbersome and time-inefficient. Instead, the problem of point selection for all of the frames should be done automatically. For this system, an optical flow algorithm was developed for point localization. The initial salient point selection was completed manually, but thereafter the points were automatically located on each set of frames (thermal and visible) using the Lucas-Kanade algorithm.

V. After registering the video files together, the pixel coordinates of the visible image were used to extract information from the thermal image on the corneal area. To do so, the first step was to localize the cornea in the visible images. On the eye, the corneal boundary is visible at the limbal junction with the white sclera, but this is not a distinct visible limit. The presence of the underlying iris visible through the transparent cornea provides a stronger visible limit, and so the outer boundary of the iris was assumed as the outer corneal boundary and used for corneal localization. A deep-learning, semantic segmentation method was used for cornea localization. After corneal segmentation in the visible image, the segmented area coordinates were used for temperature extraction of ROI in the thermal images.

VI. The segmentation algorithm allowed for the removal of artefacts in data extracted from the segmented corneal area of the thermal video sequence. These artefacts occurred due to eye movements and eyelid closure. Small eye movements are a normal physiological feature, but larger eye excursions may also occur of the patient loses fixation. With these movements, small errors in the consistency of data collected from the cornea is possible. The developed algorithm for corneal segmentation ensured consistent eye tracking and the removal of eye movement artefacts. The corneal segmentation algorithm was also used to monitor and exclude video frames that included eye lid closure. Eyelid closure will interrupt the observation of the corneal area, and lead to an unwanted replacement of the corneal OST data with eyelid temperature data. Video frames were monitored for absence of corneal segmentation, and then removed from the data stream for analysis.

VII. With completion of these steps (I-VI), the corneal area temperature (OST) could be extracted from the thermal images. For proof of concept, only simple data values were extracted from the available data on each image frame, but with further development, the full data set can be analyzed to extract information on spatial temperature differences across the ocular surface or over time.

9.3 Future work

Further developments are required for the system to be useful in clinical practice and research. These improvements can be considered in four areas:

9.3.1 Algorithm design

The first step would be to improve the accuracy of the point localization and video registration using deep learning techniques. This could be achieved using deep-learning techniques for video registration and point-tracking. Secondly, corneal segmentation accuracy could be improved by using more data for training and using a customized network. Thirdly, the GUI could be improved for better data output and OST reporting. Another, improvement can be made by developing an algorithm to remove any reflection artefacts from the thermal image.

9.3.2 Instrument infrastructure

The overall instrument design and system-set-up also needs to be refined from this basic prototype set-up to provide a simpler arrangement for the user. This includes a housing that fixes the dual camera mounting to avoid interference with the camera asset-up and critical image overlap between the two FOVs.

9.3.3 OST data extraction

Data output from the segmented corneal area will also need considerable work to extract the desired OST data from the thermal images. A criticism of previous systems has been that a very large amount of data is being left un-analyzed. Individual pixel data can be used to detect absolute as well as relative temperature change, in spatial or temporal terms, across the segmented cornea. For example, these relative temperature changes provide new methods for assessing tear film stability which will be of use for dry eye disease diagnosis and treatment. These steps can be added to the current system by developing new algorithms in the software part of the system.

9.3.4 Visual camera image analysis

Further image analysis is possible for the visual image video sequences to develop methods for assessing ocular redness or fluorescein staining or tear film break-up.[70] New algorithms can be developed as part of the software system. The dual camera system permits overlapping image analysis of the same ocular features, (e.g., tear film break-up) to reveal new detail on tear film structure and stability.

Letters of Copyright Permission

Figure 2-1
SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Jan 01, 2022

This Agreement between University of Waterloo -- Ehsan Zare Bidaki ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5220081384888
License date	Jan 01, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Machine Vision and Applications
Licensed Content Title	Thermal cameras and applications: a survey
Licensed Content Author	Rikke Gade et al
Licensed Content Date	Nov 9, 2013
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format electronic	
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no

Title: A System for Ocular Surface Temperature Measurement Using Infrared Thermography

Institution name	University of Waterloo
Expected presentation date	Jan 2022
Portions	Figure 2
Requestor Location	
	University of Waterloo Waterloo, ON N2L 3G1 Canada
Attn:	University of Waterloo
Total	0.00 USD

Figure 2.4

The request you have made is considered to be non-commercial/educational. As the article you have requested has been distributed under a Creative Commons license (Attribution-Noncommercial), you may reuse this material for non-commercial/educational purposes without obtaining additional permission from Springer Nature, providing that the author and the original source of publication are fully acknowledged (please see the article itself for the license version number). You may reuse this material without obtaining permission from Springer Nature, providing that the author and the original source of publication are fully acknowledged, as per the terms of the license. For license terms, please see http://creativecommons.org/

Figure 2.5

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted.

Figure 2.6

ELSEVIER LICENSE
TERMS AND CONDITIONS
Jan 01, 2022

This Agreement between University of Waterloo -- Ehsan Zare Bidaki ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5220101092717
License date	Jan 01, 2022
Licensed Content Publisher	Elsevier
Licensed Content Publication	Biochimica et Biophysica Acta (BBA) - Biomembranes
Licensed Content Title	Tear film lipid layer: A molecular level view
Licensed Content Author	Lukasz Cwiklik
Licensed Content Date	Oct 1, 2016
Licensed Content Volume	1858
Licensed Content Issue	10
Licensed Content Pages	10
Start Page	2421
End Page	2430
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

Title: A System for Ocular Surface Temperature Measurement Using Infrared Thermography

Institution name	University of Waterloo
Expected presentation date	Jan 2022
Portions	Figure 2
Requestor Location	University of Waterloo Waterloo, ON N2L 3G1 Canada
Attn:	University of Waterloo
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

Figure 2.7

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Jan 02, 2022

This Agreement between University of Waterloo -- Ehsan Zare Bidaki ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5220910281348
License date	Jan 02, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Anatomy of the Lacrimal Drainage System
Licensed Content Author	Daniele Marchioni, et al.
Licensed Content Date	Jan 1, 2016
Type of Use Thesis/Dissertation	
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature Content	no

Title: A System for Ocular Surface Temperature Measurement Using Infrared Thermography

Institution name	University of Waterloo
Expected presentation date	Jan 2022
Portions	Figure 1
Requestor Location	University of Waterloo Waterloo, ON N2L 3G1 Canada
Attn:	University of Waterloo
Total 0.00 USD	

Figure 2-9
ELSEVIER LICENSE
TERMS AND CONDITIONS
Jan 01, 2022

This Agreement between University of Waterloo -- Ehsan Zare Bidaki ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5220110458765
License date	Jan 01, 2022
Licensed Content	Publisher Elsevier
Licensed Content Publication	American Journal of Ophthalmology
Licensed Content Title	Screening for Dry Eye With Newly Developed Ocular Surface Thermographer
Licensed Content Author	Tomoyuki Kamao, et al.
Licensed Content Date	May 1, 2011
Licensed Content Volume	151
Licensed Content Issue	5
Licensed Content Pages	11
Start Page	782
End Page	791
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	A System for Ocular Surface Temperature Measurement Using Infrared Thermography
Institution name	University of Waterloo
Expected presentation date	Jan 2022
Portions	Figure 1 and Figure 2
Requestor Location	University of Waterloo Waterloo, ON N2L 3G1 Canada
Attn	University of Waterloo
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

Figure 2.10

Figure 2.11

Figure 2.12
Permission obtained from Author by email.

Figure 4.12
Permission obtained from Author by email.

Figure 5.7
Permission obtained from Author by email.

Figure 5.8
Permission obtained from Author by email.

Figure 5.9
Permission obtained from Author by email.

Figure 5.10
Permission obtained from Author by email.

Figure 5.11

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used

with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted.

Figure 5.12
Permission obtained from Author by email

Figure 5.13
SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Jan 01, 2022
This Agreement between University of Waterloo -- Ehsan Zare Bidaki ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5220181008982
License date	Jan 01, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Journal of Medical Systems
Licensed Content Title	Feature and Intensity Based Medical Image Registration Using Particle Swarm Optimization
Licensed Content Author	Mohamed Abdel-Basset et al
Licensed Content Date	Nov 3, 2017
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Naturecontent	no

Title A System for Ocular Surface Temperature Measurement Using Infrared Thermography

Institution name	University of Waterloo
Expected presentation date	Jan 2022
Portions Figure	8
Requestor Location	University of Waterloo Waterloo, ON N2L 3G1 Canada
Attn:	University of Waterloo
Total	0.00 USD

Figure 7.8

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted.

Figure 7-14
SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Jan 01, 2022

This Agreement between University of Waterloo -- Ehsan Zare Bidaki ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5220190320872
License date	Jan 01, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook

Licensed Content Title	U-Net: Convolutional Networks for Biomedical Image Segmentation
Licensed Content Author	Olaf Ronneberger, Philipp Fischer, Thomas Brox
Licensed Content Date	Jan 1, 2015
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no

Title: A System for Ocular Surface Temperature Measurement Using Infrared Thermography

Institution name	University of Waterloo
Expected presentation date	Jan 2022
Portions	figure 1
Requestor Location	University of Waterloo Waterloo, ON N2J 2N9 Canada
Attn:	University of Waterloo
Total	0.00 USD

Figure 7-15

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted.

References

1. Cheng DK. *Field and wave electromagnetics.* Pearson Education India; 1989.

2. Gade R, Moeslund TB. Thermal cameras and applications: A survey. *Mach Vision Appl*. 2014;25(1):245-262.

3. Wyszecki G, Stiles WS. *Color science.* Vol 8. Wiley New York; 1982.

4. Hunt RWG. *The reproduction of colour*. John Wiley & Sons; 2005.

5. Meola C. Origin and theory of infrared thermography. *Infrared Thermography Recent Advances and Future Trends, Bentham eBooks*. 2012:3-28.

6. Friedman E, Miller JL. *Photonics rules of thumb: Optics, electro-optics, fiber optics, and lasers.* McGraw-Hill Education; 2004.

7. Miller JL. Principles of infrared technology: A practical guide to the state of the art. 1994. *Google Scholar*.

8. Lindsay RB, Lindsay RB, Lindsay RB, Lindsay RB. *Basic concepts of physics.* Van Nostrand Reinhold; 1971.

9. Jones C. Physical aspects of infrared imaging. In: *Webb's physics of medical imaging.* CRC Press; 2016:659-682.

10. Lim S, Leem D, Park K, et al. Organic-on-silicon complementary metal–oxide–semiconductor colour image sensors. *Scientific reports*. 2015;5(1):1-7.

11. Norton PW, Cox S, Murphy B, et al. Uncooled thermal imaging sensor and application advances. . 2006;6206:620617.

12. Purslow C, Wolffsohn JS. Ocular surface temperature: A review. *Eye Contact Lens*. 2005;31(3):117-123.

13. Nishino K, Nayar SK. The world in an eye [eye image interpretation]. . 2004;1:I-I.

14. Holly FJ. Tear film physiology. *Am J Optom Physiol Opt*. 1980;57(4):252-257.

15. Cwiklik L. Tear film lipid layer: A molecular level view. *Biochimica Et Biophysica Acta (BBA)-Biomembranes*. 2016;1858(10):2421-2430.

16. Bron AJ, Tiffany JM. The meibomian glands and tear film lipids. *Lacrimal Gland, Tear Film, and Dry Eye Syndromes 2*. 1998:281-295.

17. Johnson ME, Murphy PJ. Changes in the tear film and ocular surface from dry eye syndrome. *Prog Retin Eye Res*. 2004;23(4):449-474.

18. Tran CH. *Structural and Biochemical Analysis of the Tear Film*. 2003.

19. Korb D, Craig J, Doughty M, Guillon J, Tomlinson A, Smith G. The tear film–structure, function and clinical examination. 2002. *Rochester, UK: Butterworth-Heinemann*. 2002.

20. HOLLY FJ. Tear film physiology and contact lens wear. I. pertinent aspects of tear film physiology. *Optometry Vision Sci*. 1981;58(4):324-330.

21. Lemp MA, Hamill JR. Factors affecting tear film breakup in normal eyes. *Arch Ophthalmol*. 1973;89(2):103-105.

22. Mishima S, Maurice D. The oily layer of the tear film and evaporation from the corneal surface. *Exp Eye Res*. 1961;1(1):39-45.

23. Craig JP, Tomlinson A. Importance of the lipid layer in human tear film stability and evaporation. *Optom Vis Sci*. 1997;74(1):8-13.

24. Mapstone R. Determinants of corneal temperature. *Br J Ophthalmol*. 1968;52(10):729-741.

25. Lemp MA, Foulks GN. The definition and classification of dry eye disease. *Ocul Surf*. 2007;5(2):75-92.

26. Stern ME, Beuerman RW, Fox RI, Gao J, Mircheff AK, Pflugfelder SC. The pathology of dry eye: The interaction between the ocular surface and lacrimal glands. *Cornea*. 1998;17(6):584-589.

27. Rolando M, Zierhut M. The ocular surface and tear film and their dysfunction in dry eye disease. *Surv Ophthalmol*. 2001;45:S203-S210.

28. Milder B. The lacrimal apparatus. *Adler's Physiology of the Eye*. 1987.

29. Trygve S. Ocular anatomy and physiology. . 1993.

30. Marchioni D, Bettini M, Soloperto D. Anatomy of the lacrimal drainage system. In: *Endoscopic surgery of the lacrimal drainage system.* Springer; 2016:1-12.

31. Naidorf H, Prat DL, Moisseiev J, et al. Ocular surface temperature differences in retinal vascular diseases. *Retina*. 2021.

32. Rosenstock T, Chart P, Hurwitz J. Inflammation of the lacrimal drainage system-assessment by thermography. *Ophthalmic Surgery, Lasers and Imaging Retina*. 1983;14(3):229-237.

33. Morgan PB, Tullo AB, Efron N. Ocular surface cooling in dry eye—a pilot study. *Journal of the British Contact Lens Association*. 1996;19(1):7-10.

34. Morgan PB, Smyth JV, Tullo AB, Efron N. Ocular temperature in carotid artery stenosis. *Optometry & Vision Science*. 1999;76(12):850-854.

35. Bourjat P, Gautherie M. Unilateral exophthalmos investigated by infrared thermography. *Mod Probl Ophthalmol*. 1975;14:278-285.

36. Cardona G, Morgan PB, Efron N, Tullo AB. Ocular and skin temperature in ophthalmic postherpetic neuralgia. *The pain clinic*. 1996;9(2):145-150.

37. IuB B, Lenskaia O, Belkina B. Thermography in the diagnosis of retinoblastoma in children. *Med Radiol*. 1985;30(12):19-21.

38. Eremenko AI. Thermography in the diagnosis of vascular neuritis of the optic nerve. *Oftalmol Zh*. 1990;(4)(4):235-239.

39. Shah AM, Galor A. Impact of ocular surface temperature on tear characteristics: Current insights. *Clinical Optometry*. 2021;13:51.

40. Tan JH, Ng E, Acharya R. An efficient automated algorithm to detect ocular surface temperature on sequence of thermograms using snake and target tracing function. *J Med Syst*. 2011;35(5):949-958.

41. Hosaka E, Kawamorita T, Ogasawara Y, et al. Interferometry in the evaluation of precorneal tear film thickness in dry eye. *Am J Ophthalmol*. 2011;151(1):18-23. e1.

42. Chiang HK, Chen CY, Cheng HY, Chen K, Chang DO. Development of infrared thermal imager for dry eye diagnosis. *Infrared and Photoelectronic Imagers and Detector Devices II*. 2006;6294:629406.

43. Mori A, Oguchi Y, Okusawa Y, Ono M, Fujishima H, Tsubota K. Use of high-speed, high-resolution thermography to evaluate the tear film layer. *Am J Ophthalmol*. 1997;124(6):729-735.

44. Tan J, Ng EY, Acharya U R. Automated detection of eye and cornea on infrared thermogram using snake and target tracing function coupled with genetic algorithm. *Quantitative InfraRed Thermography Journal*. 2009;6(1):21-36.

45. Acharya R, Ng E, Yee GC, Hua TJ, Kagathi M. Analysis of normal human eye with different age groups using infrared images. *J Med Syst*. 2009;33(3):207-213.

46. Craig JP, Singh I, Tomlinson A, Morgan PB, Efron N. The role of tear physiology in ocular surface temperature. *Eye*. 2000;14(4):635-641.

47. Mathers W. Evaporation from the ocular surface. *Exp Eye Res*. 2004;78(3):389-394.

48. Dursch TJ, Li W, Taraz B, Lin MC, Radke CJ. Tear-film evaporation rate from simultaneous ocular-surface temperature and tear-breakup area. *Optom Vis Sci*. 2018;95(1):5-12.

49. Hamano H, Minami S, Sugimori Y. Experiments in thermometry of the anterior portion of the eye wearing a contact lens by means of infra-red thermometer. *Contacto*. 1969;13(2):12-22.

50. Mapstone R. Measurement of corneal temperature. *Exp Eye Res*. 1968;7(2):237-IN29.

51. Efron N, Young G, Brennan NA. Ocular surface temperature. *Curr Eye Res*. 1989;8(9):901-906.

52. Hartridge H, Hill AV. The transmission of infra-red rays by the media of the eye and the transmission of radiant energy by crookes and other glasses. *Proceedings of the Royal Society of London.Series B, Containing Papers of a Biological Character*. 1915;89(610):58-76.

53. Prydal JI, Artal P, Woon H, Campbell FW. Study of human precorneal tear film thickness and structure using laser interferometry. *Invest Ophthalmol Vis Sci*. 1992;33(6):2006-2011.

54. Morgan PB, Tullo AB, Efron N. Infrared thermography of the tear film in dry eye. *Eye*. 1995;9(5):615-618.

55. Galassi F, Giambene B, Corvi A, Falaschi G. Evaluation of ocular surface temperature and retrobulbar haemodynamics by infrared thermography and colour doppler imaging in patients with glaucoma. *Br J Ophthalmol*. 2007;91(7):878-881.

56. Sodi A, Giambene B, Falaschi G, et al. Ocular surface temperature in central retinal vein occlusion: Preliminary data. *Eur J Ophthalmol*. 2007;17(5):755-759.

57. Murphy PJ, Morgan PB, Patel S, Marshall J. Corneal surface temperature change as the mode of stimulation of the non-contact corneal aesthesiometer. *Cornea*. 1999;18(3):333-342.

58. Chiang HK, Chen CY, Cheng HY, Chen K, Chang DO. Development of infrared thermal imager for dry eye diagnosis. *Infrared and Photoelectronic Imagers and Detector Devices II*. 2006;6294:629406.

59. Ng E, Tan J, Ooi E, Chee C, Acharya U. Variations of ocular surface temperature with different age groups. *Image Modeling of Human Eye, Artech House*. 2008.

60. Matteoli S, Finocchio L, Biagini I, et al. A thermographic study on eyes affected by age-related macular degeneration: Comparison among various forms of the pathology and analysis of risk factors. *Infrared Phys Technol*. 2016;76:402-407.

61. Tan LL, Sanjay S, Morgan PB. Repeatability of infrared ocular thermography in assessing healthy and dry eyes. *Contact Lens and Anterior Eye*. 2016;39(4):284-292.

62. Sudarshan VK, Koh JE, Acharya UR, et al. Evaluation of evaporative dry eye disease using thermal images of ocular surface regions with DWT and gabor transform. In: *Application of infrared to biomedical sciences.* Springer; 2017:359-375.

63. Ding J, Kim YH, Sarah MY, Graham AD, Li W, Lin MC. Ocular surface cooling rate associated with tear film characteristics and the maximum interblink period. . 2021.

64. Acharya UR, Tan JH, Vidya S, et al. Diagnosis of response and non-response to dry eye treatment using infrared thermography images. *Infrared Phys Technol*. 2014;67:497-503.

65. Matteoli S, Coppini D, Corvi A. A novel image processing procedure for thermographic image analysis. *Med Biol Eng Comput*. 2018:1-10.

66. Tan L, Cai Z, Lai N. Accuracy and sensitivity of the dynamic ocular thermography and inter-subjects ocular surface temperature (OST) in chinese young adults. *Contact Lens and Anterior Eye*. 2009;32(2):78-83.

67. Zheng S, Fu D, Yang T, Luo L, Nan X. A novel method for eye contour extraction from blurred infrared images. . 2016;1:103-106.

68. Kamao T, Yamaguchi M, Kawasaki S, Mizoue S, Shiraishi A, Ohashi Y. Screening for dry eye with newly developed ocular surface thermographer. *Am J Ophthalmol*. 2011;151(5):782-791. e1.

69. Su T, Chang S, Yang C, Chiang HK. Direct observation and validation of fluorescein tear film break-up patterns by using a dual thermal-fluorescent imaging system. *Biomedical optics express*. 2014;5(8):2614-2619.

70. Li W, Graham AD, Selvin S, Lin MC. Ocular surface cooling corresponds to tear film thinning and breakup. *Optom Vis Sci*. 2015;92(9):e248-56.

71. Kricancic H, McNeill H, Titze M, Alonso-Caneiro D, Collins MJ. Instrument for simultaneous assessment of fluorescein and thermal dynamics of the tear film. . 2017.

72. Chiang HK, Chen CY, Cheng HY, Chen K, Chang DO. Development of infrared thermal imager for dry eye diagnosis. *Infrared and Photoelectronic Imagers and Detector Devices II*. 2006;6294:629406.

73. Hartley R, Zisserman A. Multiple view geometry in computer vision 2nd ed., 4th print. . 2006.

74. Saponaro P, Sorensen S, Rhein S, Kambhamettu C. Improving calibration of thermal stereo cameras using heated calibration board. . 2015:4718-4722.

75. Ursine W, Calado F, Teixeira G, Diniz H, Silvino S, De Andrade R. Thermal/visible autonomous stereo visio system calibration methodology for non-controlled environments. . 2012:1-10.

76. Peric D, Lukic V, Spanovic M, Sekulic R, Kocic J. Geometric calibration of multi-sensor image fusion system with thermal infrared and low-light camera. . 2014;9250:92500O.

77. Campo FB, Ruiz FL, Sappa AD. Multimodal stereo vision system: 3D data extraction and algorithm evaluation. *IEEE Journal of Selected Topics in Signal Processing*. 2012;6(5):437-446.

78. Harguess J, Strange S. Infrared stereo calibration for unmanned ground vehicle navigation. . 2014;9084:90840S.

79. St-Laurent L, Mikhnevich M, Bubel A, Prévost D. Passive calibration board for alignment of VIS-NIR, SWIR and LWIR images. *Quantitative InfraRed Thermography Journal*. 2017;14(2):193-205.

80. Zitova B, Flusser J. Image registration methods: A survey. *Image Vision Comput*. 2003;21(11):977-1000.

81. Xie Z, Farin GE. Image registration using hierarchical B-splines. *IEEE Trans Visual Comput Graphics*. 2004;10(1):85-94.

82. Fitzpatrick JM, Hill DL, Maurer CR. Image registration. *Handbook of medical imaging*. 2000;2:447-513.

83. Abdel-Basset M, Fakhry AE, El-Henawy I, Qiu T, Sangaiah AK. Feature and intensity based medical image registration using particle swarm optimization. *J Med Syst*. 2017;41(12):1-15.

84. Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. . 1981.

85. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(12):2481-2495.

86. Novikov AA, Lenis D, Major D, Hladůvka J, Wimmer M, Bühler K. Fully convolutional architectures for multiclass segmentation in chest radiographs. *IEEE Trans Med Imaging*. 2018;37(8):1865-1876.

87. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. . 2015:3431-3440.

88. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. . 2015:234-241.

89. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. . 2017:2881-2890.

90. Proença H, Filipe S, Santos R, Oliveira J, Alexandre LA. The UBIRIS. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Trans Pattern Anal Mach Intell*. 2009;32(8):1529-1535.

91. Tan T, He Z, Sun Z. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image Vision Comput*. 2010;28(2):223-230.

92. De Marsico M, Nappi M, Riccio D, Wechsler H. Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols. *Pattern Recog Lett*. 2015;57:17-23.

93. Liu N, Li H, Zhang M, Liu J, Sun Z, Tan T. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. 2016:1-8.

94. Jalilian E, Uhl A. Iris segmentation using fully convolutional encoder–decoder networks. In: *Deep learning for biometrics.* Springer; 2017:133-155.

95. Arsalan M, Naqvi RA, Kim DS, Nguyen PH, Owais M, Park KR. IrisDenseNet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors. *Sensors*. 2018;18(5):1501.

96. Bazrafkan S, Thavalengal S, Corcoran P. An end to end deep neural network for iris segmentation in unconstrained scenarios. *Neural Networks*. 2018;106:79-95.

97. Lian S, Luo Z, Zhong Z, Lin X, Su S, Li S. Attention guided U-net for accurate iris segmentation. *Journal of Visual Communication and Image Representation*. 2018;56:296-304.

98. Wang C, Zhu Y, Liu Y, He R, Sun Z. Joint iris segmentation and localization using deep multi-task learning framework. *arXiv preprint arXiv:1901.11195*. 2019.