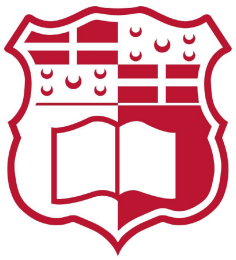
UNIVERSITY OF MALTA

Doctoral Dissertation



Towards the Analysis of Dynamic Medical Thermography: A Framework for Registration, Segmentation and Component Analysis

Ing. Jean Gauci

Supervised by:
Prof. Owen Falzon

Co-supervised by:
Prof. Ing. Kenneth P. Camilleri

A dissertation submitted in fulfilment of the requirements:
for the degree of Doctor of Philosophy
within the

Centre for Biomedical Cybernetics

November, 2024

L-Università ta' Malta

WWW.UM.EDU.MT

First edition, October 4, 2025

L-Università ta' Malta

FACULTY/INSTITUTE/CENTRE/SCHOOL Centre for Biomedical Cybernetics

DECLARATION OF AUTHENTICITY FOR DOCTORAL STUDENTS

(a) Authenticity of Thesis/Dissertation

I hereby declare that I am the legitimate author of this Thesis/Dissertation and that it is my original work.

No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

(b) Research Code of Practice and Ethics Review Procedure

I declare that I have abided by the University's Research Ethics Review Procedures. Research Ethics & Data Protection form code 1498_18042019.

☑ As a Ph.D. student, as per Regulation 66 of the Doctor of Philosophy Regulations, I accept that my thesis be made publicly available on the University of Malta Institutional Repository.

□ As a Doctor of Sacred Theology student, as per Regulation 17 (3) of the Doctor of Sacred Theology Regulations, I accept that my thesis be made publicly available on the University of Malta Institutional Repository.

□ As a Doctor of Music student, as per Regulation 26 (2) of the Doctor of Music Regulations, I accept that my dissertation be made publicly available on the University of Malta Institutional Repository.

□ As a Professional Doctorate student, as per Regulation 55 of the Professional Doctorate Regulations, I accept that my dissertation be made publicly available on the University of Malta Institutional Repository.

To Miriana

For keeping our world running.

Acknowledgements

I would like to express my deepest gratitude to those who have supported and guided me throughout this journey.

First and foremost, I wish to extend my heartfelt thanks to my supervisor, Prof. Owen Falzon, for their invaluable insights, unwavering support, and continuous encouragement. Their expertise and dedication have been instrumental in shaping this research, and I am profoundly grateful for their mentorship.

I am also deeply indebted to my co-supervisor, Prof. Kenneth Camilleri, for their constructive feedback, thought-provoking questions, and constant encouragement. Their guidance has enriched my work significantly.

My sincere appreciation goes to my colleagues and fellow researchers at the University of Malta. The stimulating discussions, shared challenges, and collaborative spirit have made this six year experience both intellectually rewarding and personally fulfilling.

I am also grateful to my friends and family, especially my wife, Miriana, for their boundless love, patience, and understanding. Their encouragement kept me grounded and motivated through the most challenging moments.

This journey has been a collective effort, and I am truly humbled by the support I have received.

Thank you all.

Abstract

Maintaining foot health is vital to overall well-being, especially for individuals with diabetes, who are at an elevated risk of developing lower-limb complications due to neuropathy and poor circulation. These conditions heighten the risk of infections, foot ulcers, and, in severe cases, amputations. Monitoring foot temperature offers a non-invasive and effective means of early detection for such complications. Shifts in temperature can indicate underlying issues: reduced temperatures may point to circulatory disorders such as peripheral arterial disease (PAD), while increased temperatures often suggest inflammation or ulceration, which can quickly worsen if left untreated.

This thesis explores the application of dynamic thermal imaging for foot health monitoring. Thermography, a non-contact technique that detects infrared radiation emitted from the skin, was employed to visualise and analyse temperature distribution across the foot. Unlike traditional static imaging, this study utilises video-based thermal data to capture temporal changes in temperature distribution.

However, incorporating a temporal dimension introduces additional challenges, notably subject movement during video capture. To address this, the study proposes an automated registration method capable of aligning the foot in dynamic thermal sequences without user intervention. A VoxelMorph-trained SynthMorph network was applied to linearly pre-registered videos, and results demonstrate that the method achieves high registration accuracy.

The next stage in the processing pipeline involves isolating the foot region from the background in the thermal data. This was achieved using the Segment Anything Model (SAM), which delivered 100% segmentation success with minimal manual input.

The final step involves analysing the temperature dynamics within the segmented foot region. Due to the complex interplay of physical and environmental factors influencing thermal data, interpretation can be challenging. To address this, Principal Component Analysis (PCA) was used to decompose the temperature dynamics. The findings show that PCA effectively distinguishes between different underlying processes.

Overall, the study demonstrates the viability of dynamic thermal imaging for continuous foot health monitoring and highlights its potential as a preventive healthcare tool. Thermography is shown to be a reliable, efficient, and accessible method for tracking foot health, paving the way for automated solutions in preventative care. By providing early warning signals, dynamic thermal imaging may improve patient outcomes, lower healthcare costs, and enhance overall quality of life.

Contents

List of Figures

List of Tables

List of Abbreviations

CT computed tomography 17
CNN convolutional neural network 48
DMD dynamic mode decomposition 26
EC-CNN Edge-Conditioned CNN 23
FLOPs floating-point operations per second 50
FTNet Feature Transverse Network 23
GELU Gaussian Error Linear Unit 49
HED Holistically-nested Edge Detection 23
ICA independent component analysis 25
IR infrared 3
MAE Masked auto-encoder 48
MAM multi-level attention module 23
MCNet multi-level correction network 23
MEEM multi-level edge enhancement module 23
MI Mutual Information 31
MLP multilayer perceptron 49
MRI magnetic resonance imaging 17
MSA multi-headed self-attention 49
NLP natural language processing 48
NvNet Nightvision-Net 24
PAD peripheral arterial disease 85
PCA principal component analysis 7
ROI region of interest 15
ROIs regions of interest 14
SAM Segment-Anything Model 8
STN Spatial Transformer Network 32
SURF speeded-up robust features 20

1

Introduction

Foot health plays a pivotal role in overall well-being, particularly for individuals with diabetes. Diabetes often leads to complications such as poor circulation and peripheral neuropathy, where nerve damage reduces the ability to feel pain, heat, or cold in the feet. This lack of sensation means that minor injuries, like cuts or blisters, may go unnoticed and untreated, leading to infections or ulcers that heal slowly due to impaired blood flow.

Temperature regulation is an important aspect of foot health, as changes in temperature can indicate underlying issues. Feet that are consistently cold may signal poor circulation, while unusually warm feet can be a sign of infection or inflammation. Monitoring foot temperature regularly can help catch early signs of problems such as Charcot foot, a serious diabetic complication where the bones and joints in the foot weaken and fracture without immediate pain [1]. By paying close attention to foot temperature and other indicators, individuals with diabetes can take preventive measures to avoid severe outcomes such as chronic infections, gangrene, or amputations, all of which significantly reduce quality of life. Therefore, maintaining good foot health, including daily inspection and temperature checks, is essential in preventing the serious, often irreversible complications associated with diabetes.

Body temperature, in general, plays a fundamental role in assessing an individuals health and well-being [2]. Body temperature is tightly regulated by the homeostatic mechanisms of the body, ensuring optimal functioning of metabolic processes. The average normal core body temperature for adults is around 37°C, although slight variations are common [3]. Fever, characterised by elevated body temperature, is the response of the body to infection or inflammation, indicating the activation of the immune system to combat pathogens. Inflammatory conditions like autoimmune diseases and inflammatory responses can also cause temperature fluctuations [4]. Abnormally low body temperatures, on the other hand, can indicate exposure to cold environments or underlying medical conditions such as hypothyroidism [5].

Body temperature can also be a marker of treatment and surgical efficacy such as, the normalisation of body temperature after treatment to infections and inflammations

[6]. It can also be a marker of adverse effect of medication and treatment such as an elevated temperature after an allergic reaction to a medication, signalling the need for adjustments in therapy [7].

During surgery, maintaining normal body temperature in the presence of anaesthesia is crucial to avoid complications like hypothermia. After surgery, body temperature can be monitored for any signs of postoperative fever or for elevated temperature at the surgical site indicating the presence of infection [6]. For example one of the criteria for diagnosing sepsis, a condition that arises when the response of the body to an infection causes widespread inflammation, leading to tissue damage, organ failure, and potentially death, is an abnormal body temperature, either too high or too low [8].

1.1 | Body Temperature Acquisition Modalities

There are three main ways of assessing the body temperature of human subjects:

- **Manual palpation** for assessing body temperature: a traditional and simple method where a healthcare provider uses their hands, usually the back of the hand, to estimate whether a person has an abnormal body temperature [9]. This method is not precise in terms of obtaining the actual temperature of the body, but it can be useful in certain situations, such as in low-resource settings or when immediate access to more specialised equipment is not available. The most common area for manual palpation is the forehead while, the neck and the abdomen are also sometimes used. Feeling the hands and feet can also sometimes be useful but might indicate changes in peripheral circulation rather than core temperature. If the skin feels noticeably warmer than usual, the patient may have a fever. Cool skin, especially in the extremities, might indicate poor circulation, cold exposure, or shock. The presence of sweat along with warmth could suggest a fever breaking, while cold sweats may be a sign of a different condition, such as shock or hypoglycemia. Although the method is simple, effective and non-invasive, it is subjective and can be influenced by environmental factors [9];

- **Thermometry**: is the most reliable and accurate method for assessing body temperature. It involves the use of thermometers to measure the internal or surface temperature of the body [10]. The thermometer is usually placed in the mouth and a temperature reading of about 36.5°C to 37.5°C is expected. Rectal, axillary, tympanic, temporal artery, ingestible, esophageal, bladder and pulmonary artery thermometry all involve the same procedure but different locations and have different levels of invasiveness and accuracy. Thermometry is the gold standard for accurately assessing body temperature but the accuracy and precision of the readings are dependent on the equipment being used [10];

Figure 1.1: A sample thermal image collected for the scope of this study showing the plantar aspect of the foot of a human subject.

- **Thermography**: also known as thermal or infrared imaging: an imaging technique that allows for the visualisation of the infrared radiation emitted by any object or living organism, that has a temperature above absolute zero (0 Kelvin). Thermal cameras use a microbolometer or a photon detector to sense the infrared radiation. The camera then converts the detected radiation into a visual representation called a thermogram or thermal image. Different temperatures are represented as varying colors depending on the colormap being used. Knowing the emissivity of an object, a measure of how efficiently a surface emits thermal radiation compared to a perfect blackbody, a temperature value of the surface of the object can be obtained [11]. Objects with a high emissivity are ideal for thermal imaging as the infrared (IR) originating from the object are primarily emitted from within the same object and there are very little reflections from the surrounding objects due to the low reflection coefficient of the surface. [12] Therefore, accurate temperature readings can be obtained from objects with a high surface emissivity. As compared to the other methods of temperature assessment, thermal imaging is non-invasive and non-contact but requires highly specialised equipment. Thermography can also have high sensitivity. More details about the physics of thermography are provided in Appendix 7. Figure 1.1 shows a sample thermal image collected for this study.

1.2 | Advantages of Thermal Imaging

Manual palpation, thermometry, and thermography have their own strengths, limitations, and specific areas of application. Thermal imaging, in particular, offers unique

Table 1.1: Comparison of temperature acquisition modalities: manual palpation, thermometry, and thermography, showing the advantage of using thermal imaging over the other modalities.

Features	Manual Palpation	Thermometry	Thermography
Accuracy	✗	✓	✗
Detection of Subtle Temperature Changes	✗	✗	✓
Non-contact Measurement	✗	✗	✓
Comprehensive Coverage	✗	✗	✓

advantages. Beyond monitoring skin temperature variations, it provides a visual representation of the surface temperature of the skin, allowing for the identification of patterns that may reveal additional insights. This capability makes thermal imaging a valuable diagnostic tool in certain medical conditions and diseases.

Thermal imaging offers distinct benefits over manual palpation in terms of accuracy and consistency. While manual palpation is a useful initial screening tool, it lacks the precision and density of data that thermal imaging can provide. Thermal imaging offers a clear advantage in clinical settings, especially for high-risk populations like those with diabetes, where early detection of temperature changes can prevent complications like foot ulcers or Charcot foot. Thermal imaging allows for consistent, repeatable results across different healthcare providers and settings, ensuring early intervention can be accurately guided. In contrast, manual palpation, though accessible and cost-free, remains a less reliable method for detailed temperature analysis, particularly when trying to catch early signs of foot complications in diabetic patients.

The primary difference between thermal imaging and thermometry lies in the scope and detail of the temperature data they provide. While thermometry is limited to point measurements, thermal imaging offers a complete, spatially detailed map of temperature distribution. This comprehensive view is crucial for detecting subtle temperature variations that are often early indicators of foot problems, like infections or circulatory issues, in diabetic patients. Thermal imaging excels in detecting asymmetrical temperature differences between the feet or across the foot, an important early sign of diabetic complications. In contrast, thermometry, while useful for spot-checking specific areas, lacks the ability to reveal these broader temperature patterns. However, thermometry is more accessible, easier to use, and requires less specialised equipment, making it suitable for general use and home monitoring. Table 1.1 summarises the advantages and disadvantages of using each of the temperature acquisition methods.

Human skin possesses a relatively high emissivity value, typically $\epsilon = 0.98$, indicating its strong ability to emit infrared radiation. The combination of the importance of temperature in assessing health conditions and the high emissivity value of the skin surface has led to the extensive use of thermography in medical applications. Medical thermal imaging remains an area of active research and development such as in breast cancer detection [13], rheumatoid arthritis [14], and diabetic foot evaluation [15]. Thermal imaging offers a non-invasive and contactless means of assessing skin temperature, making it suitable for diverse populations, including infants, elderly individuals, and patients with mobility limitations. The technology is safe and does not require direct skin contact, enhancing patient comfort and compliance during examinations.

1.3 | Applications of Thermal Imaging

Thermal imaging has been applied across various medical fields, not only in conditions that directly influence skin surface temperature, like fever screening [16] and rheumatoid arthritis [14], but also in those where abnormal skin temperature patterns indicate underlying issues, such as in cancer detection [17] and identifying complications in diabetes [15].

In cancer detection, tumors often stimulate the growth of additional blood vessels, a phenomenon known as angiogenesis, to supply the affected area. This increased blood flow not only nourishes the tumor but also elevates the temperature of the surrounding tissue since blood flow is a primary means of heat distribution in the body [18]. This localised temperature rise can be detected through thermal imaging as a hotspot on the thermal patterns of the affected area [13], [17]. Thermal imaging has thus been used as a supplementary detection tool in various cancer screenings, the most prominent of which are breast cancer [13] and skin cancer [17].

In diabetes, plaque buildup in the arteries restricts blood flow, particularly to the extremities like the hands and feet, which are furthest from the heart. This reduced circulation results in lower surface temperatures in these areas [19]. Podiatrists often manually palpate the plantar aspect of the foot to assess this issue. Thermal imaging can enhance this diagnostic process by detecting temperature variations, as small as 0.1 degrees Kelvin, that may indicate early complications, such as underlying ulcers, in the diabetic foot.

1.4 | Static vs Dynamic Thermal Imaging

The previous sections focused on the most common form of thermal imaging: static thermal imaging, where a single, instantaneous frame is captured. However, another technique known as dynamic thermal imaging captures a sequence of frames over

time, typically analysed as a video. This approach provides additional insights by revealing transient thermal dynamics, offering a more detailed understanding of temperature changes. Dynamic thermal imaging has valuable medical applications, particularly in the assessment of skin burns [20] and the detection of perforating veins [21], where temporal variations in temperature can be crucial for accurate diagnosis and treatment. Additionally, many medical studies that employ dynamic thermal imaging follow thermal provocation [22], where the area of interest is either cooled or warmed, and the subsequent recovery to ambient temperature is monitored. This approach provides valuable insights into the underlying physiological processes. However, a drawback of this method is that the thermal provocation typically induces temperature changes far greater than those caused by natural physiological processes. This significant disparity can overshadow subtle temperature variations related to underlying conditions, potentially masking important diagnostic information.

1.5 | Goals and Objectives

The primary objective of this work is to **study the monitoring of foot health using thermal imaging**. A key goal of this research is to ensure that the developed methods are **broadly applicable to medical thermal data**, while the knowledge gained from temperature dynamics should contribute to **better assessment and management of complications related to diabetic foot conditions**.

Instead of the widely used static thermal imaging approach, we propose **the use of dynamic thermal imaging**, which captures temperature variations over time to provide more detailed information. By analysing these temporal dynamics, dynamic thermal imaging offers deeper insights into underlying foot health issues.

The thermal video exhibits involuntary movement from the subjects, making the analysis of corresponding points challenging. A goal of this work is to develop **a method that accurately aligns corresponding parts of the feet across consecutive frames**, ensuring that each pixel consistently represents the same anatomical location throughout the video.

The registered data consists of thermal video, featuring both the feet in the foreground and the surrounding background. A key objective of this work is to **effectively segment the feet from the background in the dynamic thermal data**, enabling any subsequent analysis to focus exclusively on the foot region.

The extracted temperature dynamics are influenced by various physiological and environmental processes. Another objective of this work is to develop **a method capable of decomposing these dynamics into their constituent components**, allowing for the identification and separate analysis of each effect. The decomposed temperature dynamics are expected to provide insights into how various processes influence the cooling rate of the plantar aspect of the foot.

1.6 | Contributions

The contributions of this work can be split in two categories: first, contributions which **advance the knowledge on assessing thermal imaging of the human foot**, and second, those which **provide technical innovations in the processing of dynamic thermal data**.

The first contribution of this work is the **successful use of dynamic thermal imaging in a medical context**. We successfully extracted valuable insights into the processes influencing the temperature dynamics of the plantar aspect of the foot, highlighting the effectiveness of dynamic thermal imaging for medical assessments.

A second contribution to the advancement of knowledge in assessing thermal imaging of the human foot is our finding that **acclimatisation continues for over an hour after shoe removal**, exceeding the recommended guidelines for acclimatisation time before thermal imaging. This suggests that waiting for the acclimatisation process to fully complete is impractical. Instead, the ongoing acclimatisation should be identified and accounted for in the temperature dynamics to enable accurate assessment.

This work also offers three technical contributions to the processing of dynamic thermal data: (1) the implementation of a two-step registration algorithm for aligning dynamic thermal data, (2) the application of a promptable segmentation method with zero-shot learning on new images for segmenting dynamic thermal data and, (3) the evaluation of principal component analysis (PCA) as a decomposition method for analysing dynamic thermal data.

In dynamic thermal data of human subjects, some degree of movement during recording is unavoidable. By applying **a registration algorithm** before analysing the dynamic thermal data, we can automatically track and extract the temperature at a specific location throughout the entire video. This eliminates the need to manually identify the physical location in each frame, streamlining the analysis process and improving accuracy.

The dynamic thermal data captures both the foreground, the foot, and the background within its field of view. By **segmenting the background from the foreground**, we can isolate and focus exclusively on analysing the foot, ensuring that the analysis remains precise and free from background interference.

Temperature dynamics are influenced by a variety of factors, including both physiological and environmental elements. By **decomposing these dynamics into their constituent components**, we can identify the contributing factors and, in some cases, isolate or eliminate the effects of those that are irrelevant. This approach allows for a more accurate analysis of the temperature data by focusing on the factors that matter most.

The decomposition work together with the findings regarding the acclimatisation of the plantar aspect of the foot have been published in Infrared Physics and Technology [23], while the work and results on the registration algorithm have been accepted for publication at the Quantitative InfraRed Thermography Journal.

1.7 | Thesis Outline

The rest of this dissertation is organised as follows:

- in Chapter 2, background information about thermal data and its different modes of acquisition is given together with a review of pertinent literature in the fields of medical thermal imaging, image registration, image segmentation, and data decomposition;
- in Chapter 3, a registration method for dynamic thermal data is introduced. The registration method is made up of a linear registration part and an elastic registration part with the use of the SynthMorph registration method;
- in Chapter 4 the use of the Segment-Anything Model (SAM) for image segmentation is introduced for the segmentation of dynamic thermal data;
- in Chapter 5, the use of PCA for the decomposition of temperature dynamics is presented;
- in Chapter 6 an overarching discussion of the presented work is provided, and;
- in Chapter 7 a conclusion to the dissertation and the main contributions of the work are provided.

2

Background & Literature Review

Thermal imaging is one way of reading skin surface temperature in humans. This may be used to facilitate the diagnosis or prognosis of certain medical conditions whose present day clinical tests are either unreliable or take too long. It can also be employed as a complementary tool that can provide additional insight on a condition. This approach has been used in various medical fields with varying results [24], [25]. Of particular interest to this work is the application of medical thermography to the diabetic feet. In this chapter, relevant work in the field of medical thermograpghy and medical thermal data analysis and processing is discussed.

2.1 | Medical Thermography

Unlike in non-destructive testing and fault finding applications of thermal imaging, in medical thermal imaging the actual temperature readings of the object of interest are important. In other applications the thermal image is scanned for difference in the patterns obtained by the thermal imaging hardware. For example, in fault finding in electronics, it is sufficient to conclude that part of the circuit is emitting higher amounts of infrared radiation, which is proportional to the surface temperature, when compared to the rest of the circuit. The actual temperature being read by the thermal camera rarely matters in such cases. However, in medical thermal imaging, the actual temperature readings can be important, to detect a fever for example. Therefore, the thermal acquisition setup should reflect this as certain steps should be taken for a precise temperature reading.

For reliable measurements, thermal images of the region of interest should be acquired following a standard view established before any acquisition [26]. The area of interest and the thermal imaging equipment should be placed in such a way that the surface on which the temperature is measured is as parallel as possible to the thermal camera lens. This is required such that the angle of incidence of the radiation is perpendicular to the thermal camera lens. Angles of up to 20° have a negligible effect on

the temperature readings but angles of rotation greater than that should be avoided to obtain a reliable estimate of the surface temperature [27].

It is important that the subjects also take some precautionary measures before being examined by IR thermal imaging systems. Cosmetics, alcohol intake, smoking, ointments, drug treatment, large meals, and exercise, can all affect skin temperature and should be avoided prior to an examination or should be noted to the examiner [24], [26]. When the subjects arrive for examination, they should be told to undo all tight clothing and remove any jewellery. After removing the necessary clothing, the patient should be asked to wait, inside the examination room, for a period of acclimatisation. This period is required for the subjects to achieve thermal equilibrium with the conditions inside the examination room and to achieve adequate stability in skin temperature. A period of 15 minutes is generally considered to be an adequate time for stability [26], [28], [24] although different studies utilise different lengths of acclimatisation periods ranging from no acclimatisation [29] to 20 minutes [30]. Lahiri et al. [24] list different acclimatisation times used for a variety of medical applications. In this work, it is found that these acclimatisation times are too short, as acclimatisation continues for beyond an hour [23] and thus it is argued that waiting for the acclimatisation to be over is unfeasible and that the acclimatisation dynamics should be handled differently as described in Chapter 5. During this acclimatisation time the areas on which thermographic examinations are to be carried out should not touch any cold surfaces, such as metals, or be in contact with wet surfaces.

A distinction can be made between two main methods of acquisition of medical thermal images namely, static thermal imaging and dynamic thermal imaging [22]. Static thermal imaging involves acquiring a single image after a period of acclimatisation of the body region [22]. This is the most commonly adopted approach, as it only requires the acquisition of a single image from which instantaneous thermal patterns can be captured. However, this approach fails to detect any transitional activity of interest caused by physiological processes. On the other hand, dynamic imaging involves capturing thermal images over time in the form of thermal video [22]. This approach allows for the detection of transitional activity of interest caused by physiological processes, making it a valuable tool in medical diagnostics [23].

2.1.1 | Static thermal imaging

One methodology of both medical evaluations and thermal imaging is a population baseline approach in which, data is collected from a large group of subjects, categorised by some criteria based on medical science, and a threshold for each criteria is computed that differentiates between different categories. This threshold is then utilised to evaluate further subjects with ease by comparing the same medical criteria by which the threshold was set and categorising the subject based on whether

their own values for these criteria are above or below the threshold [31]. In thermal imaging a similar approach can be utilised, for example, in non-destructive testing a material is imaged by a thermal camera and if the surface temperature of the material reaches above a pre-determined threshold, a decision can be taken about the validity of the material for that particular application [32]. In medical thermal imaging, this approach has been utilised with most success in fever screening [33]. In this application, a thermal image of the forehead and face of the subject is captured and the temperature in specific locations is computed. If the temperature exceeds a predetermined threshold, the subject is marked as having fever or is referred for further examination. In this application, the non-contact nature of the imaging technique can be advantageous to increase the number of tests which are done and the rapidity with which each test is undertaken. Algorithms which automatically capture and process the thermal data have been developed and have become so efficient that mass, blind fever screening in real-time is now possible [34].

Another methodology for detecting abnormalities using static thermal imaging is pattern analysis, in which the pattern present in the area of interest is analysed to detect a deviation from normal patterns. Here thermal imaging can be utilised to provide provide physiological information that can complement anatomical data obtained from other imaging methods. This method is employed with most success in breast cancer detection, in which a breast thermogram is analysed to detect any abnormal patterns such as hotspots or areas with increased vasculature that may be indicative of the presence of breast cancer. Of all cancers, breast cancer is the most prevalent cancer found in females [35]. Mammography is the gold standard method for breast cancer screening, but this method does not come without its disadvantages. Mammography is a contact-based and sometimes painful method of screening and it also entails exposure to X-rays. Moreover, mammography is not suitable for women with dense breast tissue, a common occurrence in young women [35]. Thus, since 1982, infrared imaging has been approved in the United States as an adjunct imaging modality to mammography. Breast thermography has a high sensitivity (93.6%) and a high negative predictive value (91.2%) [25]. Here again the non-contact and radiation free nature of the imaging technique can be advantageous for more frequent and repetitive screening. Furthermore, breast thermography is not affected by the density of the breast tissue making it a superior modality to mammography in young women [36].

Skin temperature is significantly influenced by blood circulation, as blood carries heat throughout the body. In ischaemic conditions such as those caused by diabetes, a reduced temperature may be observed in the affected areas due to restricted blood supply. These ischemic conditions are more likely to impact peripheral regions, such as the hands and feet. Additionally, diabetes-induced ischemia often affects one side of the body differently than the other [28]. Research on healthy individuals has deter-

mined that the maximum temperature difference between corresponding regions on opposite sides of the body is 2.2°C [37]. Any deviation beyond this threshold may indicate the presence of disease. Consequently, the methodology of asymmetry analysis of thermal imaging in medicine, involves capturing thermal images of both sides of specific body regions, such as the plantar aspects of both feet, and comparing the temperatures in corresponding areas to identify abnormalities. This technique has proven successful in detecting ischemic conditions in the feet of diabetic patients, allowing for early diagnosis and intervention [15].

Another key methodology for applying thermal imaging in medicine is the before-and-after approach. In this method, thermal images are captured both prior to and following a medical intervention. These images are then compared to evaluate the effectiveness of the intervention. One notable application of this approach is in the quality and safety assessment of injectable drugs [38]. By capturing thermal images of the tissue before and after drug administration, practitioners can detect inflammation, which may indicate irritation caused by the drug. Another significant application is in the assessment of rheumatoid arthritis [39]. Here, thermal imaging is used to monitor the affected joints before and after exposure to a cold environment. Joints afflicted by rheumatoid arthritis typically show a reduced response to the cold challenge, which can be detected through comparative thermal imaging. This methodology can also be valuable in heart surgery. By comparing thermal images taken before and after the procedure, medical professionals can monitor whether blood flow in the affected arteries has improved [40].

Although very useful, static thermal imaging has a number of limitations. Static thermography, offers only a single snapshot of the thermal patterns, and any temporary activities of interest driven by physiological regulatory processes go unnoticed. Fernández-Cuevas et al. [41] and Houdas and Ring [42] provide an extensive description of the factors influencing the human body temperature. This means that a single-shot image of the skin surface temperature can be an unreliable source of information because of the multiple process which are affecting the readings at that single time. The before-and-after approach detailed above provides more detailed information about the state of the skin surface temperature because the images are captured under the same conditions with only one factor, medicinal intake in this case, being changed.

Continuous temperature data acquisition can be useful in this scenario because it can give information about the development of the temperature in time resulting from some known process. For example, skin surface temperature oscillates during the day in a cyclic pattern but if the body enters a room which is colder then the ambient temperature in which it was before, the general surface temperature will cool down to reflect the new, lower ambient temperature. Obtaining a single image of this whole process can be misleading as the reading of the surface temperature can be affected

by either one of the processes, the cyclic pattern or the cooling down, or both. In this scenario, a newer kind of thermal image acquisition can be useful and will provide more detailed, context-aware information: dynamic thermal imaging.

2.1.2 | Dynamic thermal imaging

Dynamic thermal imaging overcomes the constraints of static thermal imaging, namely the limitation of having only one image to assess a dynamic process, by recording a series of thermal images across a defined time span as a thermal video. This technique enables the examination of time-dependent variations in skin surface temperature and the observation of relevant transitional processes. Variations in skin temperature associated with cooling and rewarming rates may yield more meaningful information than static measurements. For instance, they can reflect enhanced vascular activity or impaired circulation. Consequently, dynamic parameters can offer further insight into both normal and pathological physiological states.

Dynamic thermal imaging has been applied in various clinical contexts that gain value from the added temporal information including breast cancer detection [43], monitoring of vascular conditions [44], detection of perforating veins [21], [45] and heart surgery [40], [46], amongst others [47].

Dynamic approaches to acquiring thermal data were first introduced in the early 1970s by Patil et al. [45] and Elem et al. [21] for identifying incompetent perforating veins in the lower limbs. These early investigations anticipated contemporary studies that employ dynamic thermal imaging in conjunction with thermal provocation—also referred to as a thermal challenge—applied to the target region. To date, the majority of studies have utilised dynamic thermal imaging in human subjects under such provocations, including cold stress or heat challenge [48]. This acquisition method, commonly known as active dynamic thermography, has proven valuable across multiple clinical domains, such as breast and skin cancer detection and the evaluation of skin burns [49].

Nevertheless, this approach presents certain limitations. The application of thermal provocation to the region of interest may mask subtle physiological or pathological temperature variations that occur naturally, as these are often much smaller in scale than the induced stimulus. To date, only a limited number of studies have examined unprovoked temporal fluctuations in skin temperature [50]. Such naturally occurring thermal dynamics are significant, as they can reveal minor variations that may reflect normal or abnormal physiological processes [23].

2.1.3 | Thermography of diabetic feet

An important clinical application of medical thermal imaging is the monitoring of diabetic feet and their associated complications [24], [51]. In diabetes, thickening of blood vessel walls restricts circulation, particularly in the extremities. As the feet are the most distal part of the body from the heart, they are especially vulnerable to impaired blood flow [52]. This reduction in circulation often results in lower skin surface temperatures at the peripheries [53]. Compared with the conventional method of assessing temperature differences in the diabetic foot through palpation of the plantar surface, thermal imaging can detect much smaller variations in surface temperature. Additionally, it allows for the simultaneous evaluation of multiple regions, unlike palpation which is limited to one area at a time, and provides objective quantitative measurements rather than subjective estimates [53].

Localised increases in temperature within the foot may also serve as an early indicator of ulcer formation [54]. A temperature asymmetry greater than 2.2 °C between contralateral regions of the plantar surface is generally considered abnormal [37], [41]. Owing to its sensitivity in detecting minor temperature differences, thermal imaging offers a simple, inexpensive, and non-invasive tool for identifying neuropathic feet and elevated ulceration risk [55]. For instance, Oe et al. [56] demonstrated in a case study that thermography revealed a region of increased temperature extending from the fourth toe, where an ulcer was present, to the ankle; the patient was subsequently diagnosed with osteomyelitis. In a follow-up study, the same group confirmed the predictive value of thermal imaging by achieving a 100% positive predictive rate for osteomyelitis in 18 diabetic patients [57].

Nagase et al. [58] have monitored thermographic patterns in the plantar surface of the foot and compared the patterns between non-diabetic subjects and diabetic subjects. They have determined 20 different possible patterns which the plantar thermographic pattern can follow using the four different angiosomes in the plantar region of the foot. Both healthy control subjects and diabetic subjects were screened and each of these was categorised to one of the 20 patterns. While the control group subjects were generally categorised in two of the patterns, the diabetic group had a larger distribution across the 20 patterns. This suggests that by monitoring the thermal patterns in the plantar aspect of the foot it is possible to identify any abnormalities and possibly detect complications in the diabetic foot.

Gatt et al. [53] have determined the normative heat pattern distribution in hands, feet and shins in a population of healthy adults. They identified 44 regions of interest (ROIs) across the three body regions from which the thermal patterns were established.

Vardasca et al. [59] have used active dynamic thermal imaging to monitor the plantar aspect of diabetic patients with open ulcers in their feet. They have built a database of 39 subjects in which the plantar aspect of the feet was recorded in a 30-minute thermal video, a cooling challenge, in the form of conduction from an aluminium board, was applied and followed by another 5-minute thermal video. Then they analysed the images by measuring the mean temperature of ROIs which correspond to the most common documented locations of foot ulcers in diabetics. They have found no statistically significant difference between thermal asymmetry values and thermal recovering differences in all ROIs except for one located at the medial forefoot. This finding shows that studying simple temperature dynamics in localised region of interest (ROI)s, might not be sufficient to capture or predict the locations of ulcers in the diabetic foot.

2.1.4 | Approaches to medical thermal imaging acquisition protocol

Various approaches to standardise the acquisition protocol for specific medical applications have been proposed [60], [61], [62], [63]. Ring and Ammer [26] have reviewed various studies which list a medical thermal imaging acquisition protocol and have drawn up a generalised procedure. They have also updated the procedure through subsequent studies [28], [64], [65]. They have examined and addressed various elements of thermal image acquisition, starting from the examination room where thermal imaging is conducted, to maintaining ambient temperature, specifying computer equipment to utilise, outlining the procedure for participants prior to examination, and have suggested a protocol for attaining dependable thermal imaging outcomes.

More recently, the International Academy of Clinical Thermology has published an article which similarly describes standards and guidelines for medical infrared imaging [66].

Thermal imaging sessions should be carried out in a controlled environment in an examination room which needs to meet some basic requirements. The room must have adequate space to locate the thermal imaging equipment at a sufficient distance from the subject and should also have space for one or more patient cubicles. A good estimate of the distance between the thermal camera and the patient can be derived from the optical features of the camera and the size of the feature being imaged. Ring and Ammer [26] state that an examination room with a minimum size of 2×3 metres may be required, but a larger room will be preferable. The room temperature should be kept at a constant temperature between 18°C to 25°C where an average of 22°C to 24°C is ideal [26], [28]. When selecting a room temperature one should keep in mind the examination method. If the examination requires the subject to be disrobed, lower room temperatures may be uncomfortable while for fully clothed subjects higher tem-

peratures may result in sweating. Lahiri et al. [24] have produced a table which shows different room temperatures used for various medical applications. The subject must also be placed away from external heat sources such as light fixtures and direct sunlight which may influence the reliability of the thermographic readings [26], [24].

2.2 | Medical Thermal Data Processing and Analysis

Once captured, thermal images and videos need to be processed further to extract the required information from the data. Traditional image processing techniques play a crucial role in enhancing the quality and interpretability of thermal data. The primary goal is to obtain a clearer, more detailed image of the body part or object being imaged, which is essential for accurate analysis and diagnosis.

Some of the thermal data analysis techniques described in Section 2.1, especially those of dynamic thermal data, necessitate the alignment or registration of the target of interest in the video frames to account for variations due to different camera angles and body proportions. This can facilitate the analysis of changes in time for specific parts of the target object. Image registration techniques are employed for this purpose. In Section 2.2.1, we discuss both traditional and modern techniques used for medical thermal data registration.

In most cases, captured thermal data encompasses not only the body region of interest but also other body regions and background elements. To enhance the performance of analysis techniques, it is sometimes important to segment the body region of interest from the rest of the data. Image segmentation techniques are employed to identify and extract these required regions, allowing for more focused and accurate analysis. In the following section, an overview of classic segmentation techniques and their utility in medical thermal data segmentation is provided

The temperature data obtained through dynamic thermal imaging is influenced by the subjects individual physiological factors, such as height, weight and gender, and the examination room environmental factors, such as temperature and humidity, amongst others, making analysis challenging. Decomposition techniques can play a crucial role by breaking down the thermal data into distinct transients associated with different factors. In Section 2.2.3, we review various data decomposition techniques and their utility in decomposing medical thermal data.

2.2.1 | Image registration

Image registration is a fundamental process in medical image processing, dedicated to achieving an optimal transformation that best aligns the objects of interest between two or more images [67]. In the context of medical image registration, these objects

of interest are typically anatomical structures within the human body [68]. This alignment is carried out between two primary images: the fixed image, often referred to as the reference image, and the moving image, which is subjected to the transformation to achieve alignment with the reference image. The applications of image registration within the field of medical imaging are vast and multifaceted, extending to diverse areas such as computed tomography (CT) and magnetic resonance imaging (MRI) registration [69]. Furthermore, this process finds applications in various medical contexts, including computer-aided diagnosis [70], atlas building [71], anatomy segmentation [72], and motion tracking [73] among others.

Within the domain of medical image registration, there exist three prominent methodologies: feature-based registration, intensity-based registration, and deep learning-based registration [74], [68]. Each methodology offers its own approach to solving the challenge of aligning images.

Feature-based registration involves the extraction of distinctive features from both the reference and moving images. Subsequently, a geometric transformation is computed to minimise the distance between corresponding feature points in these images [67]. This approach relies on the identification of salient landmarks or attributes within the images to guide the alignment process [67].

Intensity-based registration, on the other hand, centers on computing a similarity measure between the fixed and moving images. Through iterative optimisation, a geometric transformation is determined that maximise this similarity measure [67]. In this methodology, the focus is on the overall intensity or pixel values of the images, rather than specific features.

Deep learning-based registration, the most recent and rapidly evolving methodology, incorporates neural networks into the image registration process. Within this methodology, three distinct classes of approaches have emerged, each utilising neural networks in different ways to enhance registration: (i) The first class leverages neural networks as a similarity metric [75]. Neural networks are trained to assess the similarity between images, guiding the registration process by aligning images based on their content; (ii) The second class introduces neural networks to parameterise the transformation model [75]. In this approach, neural networks are employed to directly model the geometric transformations, which are then iteratively adjusted to achieve alignment; (iii) The third class employs neural networks to facilitate various operations that enhance the quality of the registration [75]. This can include using neural networks to correct artifacts, refine transformations, or provide additional context for the registration process.

The results obtained from comparative studies suggest that deep learning-based methods exhibit substantial advantages over traditional methods concerning both the quality of the resulting registration and the computational speed at which this registration is achieved [72], [75]. This observed superiority has significant implications for the field of medical imaging and beyond.

One of the key strengths of deep learning-based methods is their ability to capture intricate patterns and relationships within the images being registered. Traditional methods often rely on handcrafted features and mathematical models, which may not adequately capture the complexity of anatomical structures and their variations across different patients or scenarios [72]. In contrast, deep learning methods can automatically learn and adapt to these intricate features, leading to more accurate and robust image alignment.

Additionally, the computational speed of deep learning-based methods stands out as a significant advantage. Traditional registration techniques can be computationally intensive and time-consuming, particularly when dealing with large datasets or complex transformations. Deep learning models, once trained, are often capable of performing registrations at significantly faster rates, making them more practical for real-time or near-real-time applications such as intraoperative guidance, where speed is of utmost importance [75].

However, deep learning models require substantial amounts of labeled data for training to perform effectively [75]. This data typically consists of pairs of images with known ground-truth transformations. The scarcity of such high-quality labeled datasets poses a significant obstacle to the broader adoption and advancement of deep learning-based registration techniques. Gathering and annotating large medical image datasets is a resource-intensive and time-consuming endeavor, and in some cases, it may be impractical or ethically challenging to acquire sufficient data.

2.2.1.1 | Thermal image registration of human subject

Thermal image registration is a specialised branch of medical image registration. Within this subfield, a particular area of focus is on the registration of dynamic medical thermal image sequences. Unlike static images, dynamic thermal sequences involve capturing a series of thermal images of a human subject over time [76]. This temporal dimension introduces unique challenges and necessitates precise alignment of the body regions within each frame to a designated reference position.

One of the primary challenges to dynamic thermal image processing is the natural movement of human subjects during data acquisition [77]. The inherent movement of test individuals leads to the misalignment of corresponding body regions across sequential frames, which can compromise the accuracy and interpretability of the thermal data. To mitigate this issue, image registration techniques are applied to ensure

that corresponding body regions in successive frames are correctly aligned, enabling the extraction of consistent and meaningful information [77].

Dynamic thermal data presents some challenges to image registration, primarily due to variations in contrast [78]. These variations manifest both between subsequent frames and between the foreground and background regions of the thermal images. The contrast disparities are partly attributed to the different thermal characteristics and transitional processes exhibited by human body regions [78]. This dynamic contrast variation makes the acquisition of salient and useful features for feature-based image registration techniques a particularly challenging task.

Traditional image registration methods may struggle with these variations, as they are designed to handle relatively stable and uniform image properties. To address this, advanced image registration techniques, possibly involving adaptive feature extraction or deep learning approaches, may be required to effectively register dynamic thermal images, ensuring that the relevant body regions are consistently aligned and providing reliable data for medical analysis.

Barcelos et al. [74] propose a combined method for segmentation and registration of static thermal images for an advanced and progressive evaluation. This combined method has been primarily developed with a specific focus on the prevention of muscle injuries in high-performance athletes. Nonetheless, the adaptability of this technique extends beyond sports medicine, making it applicable to other domains. The core innovation lies in the use of optical flow, a deformable image registration technique, on segmented thermal images of athletes' lower limbs. The method is guided by hot or cold isotherms, clusters of pixels with similar temperatures, during the registration process. By leveraging these inherent temperature patterns, the authors argue that their approach allows for a more sophisticated and progressive analysis of thermal images.

However, there are notable challenges associated with the proposed method. Firstly, the segmentation of the body region from the background in thermal imagery is acknowledged as a complex task because the intensities of the background region and the body region usually overlap significantly. In dynamic thermal data, this task becomes even more intricate, as subjects are in motion, leading to variations in their appearance and body geometry. Another limitation to consider is that the method is primarily designed for static thermal images captured on different days. This limitation implies that the method may not be directly applicable to dynamic thermal data, which typically consists of a continuous stream of frames captured in quick succession.

The work by Ciantar et al. [79] introduces an innovative approach for video regis-

tration that is particularly tailored to the challenging domain of dynamic thermal sequences in the abdominal region of pregnant women. The key concept in the technique is the utilisation of a triangulation-based video registration technique that uses local affine transformations to achieve precise alignment of successive frames. The effectiveness of the method is demonstrated through empirical results, which highlight its capability to significantly enhance the alignment of abdominal regions in consecutive frames. This improvement is a crucial advancement, as it enables the tracking of physiological changes and motion in the abdominal area.

However, it is important to acknowledge the limitations of the method, which primarily revolve around the reliance on matched speeded-up robust features (SURF) points. SURF points are distinctive features in images that are commonly used for matching and registration tasks. In the context of dynamic thermal data, especially involving the feet, several challenges arise that make the acquisition of matched SURF points problematic. Firstly, dynamic thermal data from the feet often lacks distinct and consistent physical features that can be used for matching. Unlike traditional visual imagery, thermal images primarily capture variations in temperature rather than intricate patterns or shapes. This limitation makes it difficult to find characteristic points that can be reliably matched across frames. Secondly, thermal imaging is susceptible to variations in contrast caused by environmental conditions, changes in thermal patterns, or even the movement of the subject. This dynamic contrast can affect the detection and tracking of SURF points.

In a recent study, Bouallal et al. [80] compared four methods of image registration for diabetic foot thermal images. Boullal et al. studied the performance of an intensity-based algorithm, an iterative closest point method [81], a subpixel registration algorithm [82], and the pyramid approach for subpixel registration [83]. The performance of the four algorithms was evaluated using several overlap and symmetry metrics and concluded that the intensity-based and pyramid approach for subpixel registration methods gave the best results [80]. The methods outlined by Boullal et al. all rely on prior segmentation of the thermal data to extract the location of the body region from the rest of each frame. However, from our experience with working with dynamic thermal data of the plantar aspect of human feet [84], [85], this segmentation is not easily achievable, as the overlap in temperature, and hence pixel intensity, between the foreground and the background can be substantial.

Registration is only one part of the processing journey of thermal data. One other important process is image segmentation.

2.2.2 | Image segmentation

Medical thermal imaging captures data often including irrelevant regions and background noise that can obscure the area of interest. Image segmentation isolates the relevant body region, which is vital for both improving the analysis accuracy by focusing on the specific area of interest, and for reducing computational load by processing only the segmented region.

Image segmentation is the process of dividing an image into different regions such that each region is somewhat homogeneous in its features, while the union of any two regions is not [86, 87]. The goal of segmentation usually is to separate a ROI from the background for further processing [88]. The challenge of image segmentation in digital processing lies in enabling machines to recognise context, that is naturally discernible to the human eye, such that the machine is able to accurately differentiate between various regions within an image [86]. There are various methods that try to achieve this with the classic classes of image segmentation techniques being: thresholding; edge-based; region-based; clustering-based; watershed algorithm and active contour models [86].

The **thresholding technique** works on grayscale images but colour images can be considered too,. First these are converted into binary image according to one or multiple threshold levels. In the simplest form, a single threshold is set, with any value of intensity above that threshold being considered as foreground, and thus set as binary value of one, and, anything below the threshold being considered as background, and is thus set a binary value of 0. This method is straightforward but can be ineffective for images with varying intensity [86]. Adaptive thresholding takes this idea one step further by calculating different thresholds for several small regions of the image, allowing for better handling of varying lighting conditions [89].

Edges in images are the boundaries where there is a significant change in intensity or color, marking the transition between different regions of the image. **Edge-based segmentation** is a technique which focuses on identifying these edges to segment the image into meaningful regions [90]. The edges can be detected with different methods with the Sobel and Prewitt Operators, which detect edges by measuring the gradient of the image intensity, and Canny Edge Detection, which detects edges by looking for local maxima of the gradient of the image, being the most popular methods. The Canny edge detection algorithm is known for its accuracy in locating the edge and is less likely to be affected by noise [90].

Region-based segmentation techniques can be split into two main categories; region growing and region splitting and merging [91]. Region growing starts from a seed point and grows the region by including neighboring pixels that meet certain criteria. It is effective for images with homogeneous regions [91]. Region splitting and Merging divides the image into regions and merges them based on similarity criteria, creating a more adaptive segmentation [91].

Clustering-based image segmentation is a technique where pixels are grouped into clusters based on their similarity. This similarity can be defined in terms of color, intensity, texture, or spatial proximity. A well known method of clustering-based segmentation is K-means clustering, which partitions the image into K distinct, non-overlapping regions or clusters. Each cluster is defined by its centroid, and the objective of the algorithm is to minimise the variance within each cluster [92]. A variation of this algorithm is the fuzzy C-means algorithm in which each pixel may belong to multiple clusters with varying degrees of membership, providing a more flexible segmentation [93].

The **Watershed Algorithm** treats a grayscale image as a topographic surface and finds the watershed lines, which correspond to the edges of objects. It is particularly useful for separating touching objects in an image [94].

Active contour models or snakes use curves that move within the image to find object boundaries. It is effective for capturing complex shapes and boundaries but can be sensitive to initial placement and noise [95].

In recent years, deep-learning based methods have yielded a new generation of segmentation models with remarkable performance improvements. Often these methods outperform traditional methods of segmentation on popular benchmarks in both speed and accuracy [96]. Minaee et al. [96] have reviewed over a 100 deep learning based methods of segmentation and provide a comprehensive review.

Kirillov et al. [97] have introduced a new task, dataset and model for image segmentation named Segment Anything. They have built a dataset of over one billion segmentation masks on 11 million images and a segmentation model that is designed and trained to be promptable. The model can transfer to new images in a zero-shot manner, which means that it does not need any further training and can be used with its current weights. Moreover, their model and dataset are publicly available [97]. Segment Anything has shown an impressive performance in zero-shot generalisation and is even superior to prior fully supervised methods [97]. Ke et al. [98] have further improved the performance of Segment Anything model with their Segment Anything in High Quality model that has superior performance particularly when dealing with objects that have intricate structures.

Thermal images are grayscale images, meaning that the image has only one channel representing intensity information. In the case of thermal images this intensity is rep-

resentative of the level of temperature of the object being imaged. The fact that objects in the image can only be represented with a varying level of gray, usually white represents the highest intensity and black represents the lowest level of intensity, leads to a high overlap between the intensities of the different objects especially at the edges. This phenomenon, known as thermal crossover, makes the challenge of segmenting grayscale images a difficult one [88]. Furthermore, the variability and complexity of the anatomical structures being imaged make the problem even more complex and has resulted in the area of medical thermal image segmentation being underrepresented in the research community [99].

One possible solution to the thermal crossover problem is to enhance the edges of the thermal image or video prior to the segmentation step. Li et al. [100] present the **Edge-Conditioned CNN (EC-CNN)** which exploits edge prior information to increase the quality of segmentation output in the presence of ambiguous and noisy edges. The proposed network consists of three layers: an edge extractor; the EC-CNN blocks, and; a semantic segmentation network. For the edge extractor layer, Holistically-nested Edge Detection (HED) [101] is used to obtain high-quality edges. EC-CNN blocks consist of convolutional layers and gated feature-wise transform layers to embed edge information properly. The semantic segmentation network employed is DeepLabV3 which employs ResNet as feature extractor and Atrous convolutions [102]. EC-CNN is employed to segment thermal images under varying illumination conditions and achieves better than state-of-the art performance on the authors' own dataset.

Another method which leverages edge information to enhance the segmentation accuracy is proposed by Xiong et al. [103]. They propose a **multi-level correction network (MCNet)** with a multi-level attention module (MAM) and a multi-level edge enhancement module (MEEM). Specifically, MAM selectively captures the inter-class and intra-class contextual dependencies by a multi-level correction process and MEEM continuously combines precise context information and edge prior knowledge in each level to correct the final feature representation [103]. Experiments reported in the paper have shown that the proposed method performs well compared to state-of-the-art methods in semantic segmentation.

Panetta et al. [104] propose **Feature Transverse Network (FTNet)** a novel end-to-end trainable convolutional neural network architecture that aims to adequately capture the context of corresponding pixels without trading off accuracy for speed, or vice-versa. FTNet employs an encoder-decoder architecture and an edge guidance module that conducts reliable pixel-wise classification. FTNet captures and enhances feature representation at multiple scales, which improves its ability to handle high-resolution images and delivers high-quality output with reduced computational cost. The proposed method has been benchmarked on popular thermal datasets and has been shown to be effective in multiple aspects including, quantitative accuracy and

speed [104].

Most of the work related to the segmentation of thermal images, revolves around the autonomous driving application [99]. Wang et al. [105] proposed a method to segment pedestrians from thermal data using a **conditional generative adversarial network (IPS-cGAN)**. The generator part of the network is based on Unet [106] with two modifications: convolutional blocks are replaced by residual blocks to have more efficient connections, and; a dropout rate of 0.5 is deployed so that the network becomes more robust. SandwichNet, a novel convolutional network with symmetrical input, is proposed as the discriminator for the segmentation method. SandwichNet, which is based on PatchGAN [107], needs symmetrical, three-channel, result-image-result segmentation results from the generator and thermal image, and three-channel truth-image-truth segmentation ground truths. The designed generator and discriminator are trained as an end-to-end GAN algorithm with cross-entropy loss. The modifications on Unet and the design of the discriminator provide a robust model against noise for thermal image pedestrian segmentation [105].

For the semantic segmentation of low-resolution thermal images in weak illumination conditions, Chen et al. propose **Nightvision-Net (NvNet)** [108] a network with the architecture of the FCN-8S [109] with an enforcement layer in the front end and a weighted-sigmoid-cross-entropy loss function to calculate the error between the prediction of the network and the ground-truth. The network is trained using transfer learning. The authors report that the proposed method achieves comparable results to other state-of-the-art methods in low illumination environment.

Kutuk et al. [99] have reviewed semantic segmentation methods for thermal images and thermal image datasets while, various other studies exist which review methods of segmentation for medical images [87, 110, 111].

Once the thermal image or video data is segmented and registered, the temperature dynamics of the body region of interest can be extracted. These dynamic are made up of a complex mixture of factors, from physiological to environmental and equipment related and thus, are difficult to understand on their own. Data decomposition techniques can be useful to break down the temperature dynamic into its constituent factors and thus analysing them independently. In the next section, a review of salient and popular decomposition techniques is presented together with their application in medical thermal imaging.

2.2.3 | Decomposition techniques

Decomposition techniques are powerful and versatile methods applied across numerous disciplines to address the inherent complexity of data, systems, and mathematical functions. At their core, these techniques involve breaking down a complex entity; a dataset, a signal, a system, or a function; into smaller, more manageable compo-

nents [112]. Decomposing a system into its fundamental components makes it easier to interpret and understand. For instance, in image processing, decomposing an image into different frequency components using Fourier Transform, allows one to analyse various aspects like edges, textures, and overall patterns separately [113]. Decomposition techniques also reduce the computational burden associated with analysing large or complex datasets and systems. This efficiency can be achieved in two manners: dimensionality reduction in which the number of dimensions of the data is reduced while preserving most of the information [114] and; parallelisation in which, by dividing a large problem into smaller, independent sub-problems, it becomes possible to solve them simultaneously on multiple processors, significantly speeding up computation [115].

Decomposition techniques are also essential for extracting meaningful patterns from data. This capability is particularly valuable in fields like machine learning, where identifying patterns is crucial. Decomposition techniques are used to extract latent features or components from data which can then be used as inputs for machine learning models, often leading to better performance and interpretability [116]. In time-series analysis, decomposition methods separate a time series into trend, seasonal, and residual components. This separation allows for more accurate forecasting and a better understanding of underlying patterns [117].

Some seminal methods of data decomposition include amongst others:

- Singular value decomposition (SVD) - a matrix factorisation technique that decomposes a matrix into three matrices: U, Σ, and V, where U and V are orthogonal matrices and Σ is a diagonal matrix with singular values [118]. SVD is a foundation for many other techniques and is widely used in signal processing, image compression, and collaborative filtering. It provides a way to approximate matrices, which is useful in low-rank approximations but can be computationally intensive for very large datasets [118];

- Principal component analysis (PCA) - a linear dimensionality reduction technique that transforms the original dataset into a new set of uncorrelated variables called principal components [119]. These components capture the maximum variance in the data, with the first few components often retaining most of the important information. It is traditionally used for data compression, noise reduction, and feature extraction. PCA is simple, fast, and effective for reducing dimensionality in large datasets but only captures linear relationships, and the results can be difficult to interpret when the data is highly nonlinear [120];

- Independent component analysis (ICA) - a technique used to separate a multivariate signal into additive, independent components. It assumes that the observed data are linear mixtures of independent source signals [121]. ICA is pri-

marily used in blind source separation, such as for separating mixed audio signals, and in feature extraction. It is effective for finding underlying factors or sources in the data but assumes statistical independence of sources, which may not always hold [121];

- Wavelet transform - a time-frequency decomposition technique that represents a signal in terms of shifted and scaled versions of a wavelet function [122]. It provides a hierarchical decomposition of data at different levels of resolution. The wavelet transform is used in signal processing, image compression, and denoising. It is capable of capturing both frequency and location information, making it useful for analysing non-stationary signals but the choice of wavelet function and decomposition level can significantly affect results [122];

- Factor Analysis - a statistical technique used to describe variability among observed variables in terms of fewer unobserved variables called factors [123]. It assumes that the observed variables are linear combinations of the potential factors, plus noise. It is used in psychology, finance, and social sciences for identifying underlying relationships between variables. Factor Analysis helps reduce the number of variables while retaining the essential structure of the data but assumes linear relationships and that the data fits a normal distribution, which may not always be the case [123];

- Dynamic mode decomposition (DMD) - a data-driven technique used to analyse the dynamics of complex systems. It decomposes high-dimensional data, such as time series data from simulations or experiments, into a set of spatial modes, each associated with a specific temporal behavior (growth/decay rates and frequencies) [124]. DMD provides insights into the underlying structures and dynamics of the system by identifying patterns that evolve over time. Originally developed in fluid dynamics, DMD is now widely used in various fields for system identification, reduced-order modeling, and forecasting, particularly when the governing equations of the system are unknown or difficult to model directly [125]. DMD does not require prior knowledge of the governing equations of the system, making it useful for analysing data where the underlying dynamics are unknown or too complex to model directly but can be sensitive to noise in the data, which may lead to inaccurate modes or frequencies, particularly in the presence of significant measurement noise [124].

The results from these methods will vary depending on the form of the data to be decomposed. Thus, there are a multitude of methods, based on the more basic decomposition techniques, which are application specific [126, 127]. For this reason, we refrain from listing the decomposition techniques which are applied to different applications in this review and refer the interested reader to other literature [128].

Literature related to the decomposition of thermal data is scarce [129, 130] and most of the developed methods utilise the decomposition methods to extract features from the images as opposed to decomposing the dynamics of the thermal data. Mozafari et al. [131] use tensor decomposition to extract the respiration rate from thermal video data. To the best of our knowledge there is no other method which decomposes temperature dynamics obtained from thermal data for the purpose of extracting factors which contribute to the overall thermal pattern.

This thesis investigates the use of dynamic thermal imaging for monitoring foot health. Thermography is utilised to visualise and analyse temperature distribution across the foot. Through an advanced image registration and segmentation framework, the study successfully aligns and isolates the foot from the background, enabling precise analysis of thermal data through PCA.

In the next chapter, the work and results related to the registration algorithm are presented together with a discussion of the results.

3

Dynamic Medical Thermal Data Registration

Medical analysis frequently requires distinct images to be compared either for the assessment of the success of medical procedures or the prognosis of certain diseases. Images can be either from the same source, such as comparing before and after scans, or from different sources, such as comparing T1 weighted and T2 weighted MRI images. Studies on populations and baseline definitions will also require various images to be compared to a base image, known as the atlas image. The process of analysis in these scenarios can be facilitated if corresponding physical structures in different images are aligned as opposed to analysing different images whose corresponding pixels do not correspond to the same anatomical locations.

In image processing, the process of aligning images based on the presence of common structures is called image registration and is a well-known problem with many different methods being proposed for different scenarios as seen in Section 2.2.1 above.

A specialised field of image registration is the registration of thermal data, particularly dynamic thermal data. In a medical scenario of dynamic thermal data acquisition, the human subjects are likely to move, even slightly, during the acquisition of the thermal video. Therefore any particular pixel might not represent the same location on the body throughout the video. This means that either the body region of interest has to be aligned throughout the video via post-acquisition registration or the temperature extraction process has to be movement invariant. The registration of dynamic thermal data is complicated by the fact that features of interest are not only spatially misaligned between the fixed and moving image but can also be of different intensity and contrast. Thus, traditional registration techniques that are applied to the registration of RGB videos do not perform well on dynamic thermal data.

In this work, we adapt a method, SynthMorph [69], a deep learning registration framework, to align video frames representing dynamic thermal data. The method is applied to thermographic videos of the plantar aspect of human feet, an area of

significance in the context of diabetes [23, 41, 84]. The proposed approach is shown to yield highly accurate registration results. The alignment of dynamic thermal data significantly facilitates temperature analysis for specific ROIs and allows for improved visualisation and analysis of results.

3.1 | Methods

The registration process presented here involves two steps. First, a rigid linear pre-registration involving the removal of large linear deformations between frames, such as translation and rotation, is carried out. Subsequently, an elastic registration procedure that removes the smaller elastic deformations, is applied.

Since this work requires alignment of consecutive frames in a thermal video, where frames exhibit similar intensity patterns, an intensity-based registration method is chosen to handle the linear deformations. Feature-based registration is less suitable in this context due to the challenge of extracting corresponding features in situations where contrast between the foreground and background is inconsistent, and the foreground object—such as a foot in this case—lacks distinctive physical features.

For handling elastic deformations, a deep learning-based approach was selected, as such methods have consistently achieved state-of-the-art performance in elastic image registration. The specific method used, SynthMorph, has demonstrated superior results in medical image registration, particularly with MRI data, and was found to be effectively transferable to dynamic thermal video data.

A two-step approach, linear registration followed by elastic registration, was adopted to meet the requirements of the VoxelMorph network, which takes as input images that are linearly aligned but still contain elastic deformations. This approach also aligns well with the characteristics of our data, which exhibit large linear deformations, particularly due to foot rotations in the image plane. These linear deformations are relatively easier to correct. However, smaller, more challenging non-linear deformations arise when the foot rotates out of the image plane, causing perceptible changes in its shape in the data. This effect is illustrated in Figure 3.1, which shows a sample from our dataset.

3.1.1 | Rigid transformation

Let $\mathbf{V}_n$, where $n \in \{0, ..., N-1\}$, be a thermal video with N frames. Consider two consecutive frames, v_j and v_{j+1}, from the sequence, which are denoted by $\mathbf{F}$ and $\mathbf{M}$ respectively. $\mathbf{F}$ denotes the *fixed* image which is transformed to align with the *moving* image, $\mathbf{M}$. Pixel coordinates in the images can be denoted by using a pair of values, $\mathbf{x} = (x, y)$ [132]. If the difference between $\mathbf{F}$ and $\mathbf{M}$ consists of only a translation and rotation in the plane of the image, the relationship between a pixel on the body region

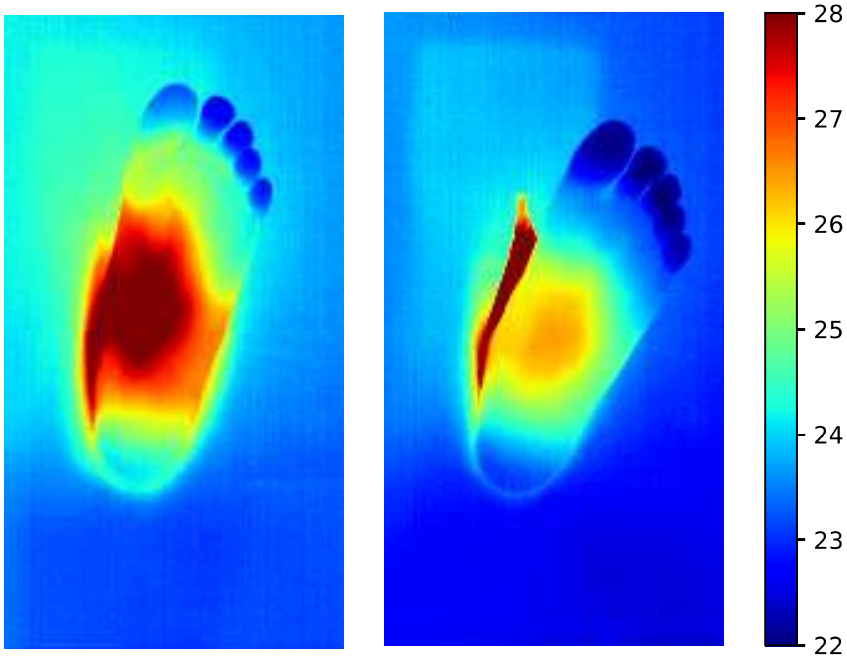


Figure 3.1: Sample data for the registration procedure reveals that misalignment between frames is caused by two primary factors: large linear deformations resulting from in-plane foot rotations, and smaller non-linear deformations caused by out-of-plane rotations, where the foot moves forward, altering its perceived shape in the image.

of interest from frame **F**, and the pixel on the corresponding body region of interest in frame **M**, denoted by $\mathbf{x_F}$ and $\mathbf{x_M}$ respectively, is [132]

$$\mathbf{x_M} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}}_\mathbf{F} \tag{3.1}$$

where $\bar{\mathbf{x}}_\mathbf{F}$ is the augmented vector $(x_F, y_F, 1)^T$, $\mathbf{t} = (t_x, t_y)$ is the translation vector where t_x represents translation in the X-axis and t_y represents translation in the Y-axis, and **R** is the rotation matrix [132]

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.2}$$

The rotation matrix, **R**, becomes an identity **I** if the difference between **F** and **M** is only a translation, where **I** is the (2×2) identity matrix [132]

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.3}$$

Equation (3.1) is known as the rigid transformation and the parameters of the transformation matrix describe the mapping between two images in terms of rotation and translation [132]. By estimating the parameters in Equation (3.1) between

images **F** and **M** and then transforming image **M** using the transformation parameters, **M** is aligned to **F** such that corresponding body regions of interest in the two images occupy corresponding pixel coordinates.

The alignment of a moving image **M** to a fixed image **F** through an estimation of the parameters in Equation (3.1) is done by using an iterative optimisation of a similarity metric [132]. The similarity between **F** and **M** is computed with a metric of choice. A transformation, which can be either based on prior knowledge or an initial guess, is applied to **M** and the measure of similarity between **F** and **M** is re-computed. An optimisation algorithm is employed to iteratively update the transformation parameters to minimise the difference between the fixed and moving image. The iteration continues until a convergence criterion is met. This criterion can be a predefined number of iterations, a threshold on the improvement in the similarity metric, or the stability of the transformation parameters. The transformation parameters are applied to the fixed image in the iteration to align the moving image followed by an interpolation method to estimate intensity values on the image grid, to have continuity in the image.

The similarity metric, optimisation algorithm and interpolation method can be one of a number of methods. Here we define the methods which are later used in our implementation, namely: Mutual Information (MI) metric; gradient ascent optimisation, and; nearest neighbour interpolation.

Mutual Information (MI) measures how much information one image provides about another. Specifically, it quantifies the statistical dependency between pixel intensities in the two images being aligned. When the images are well-aligned, the MI is maximised, indicating that the intensity distributions of the images share the most information. Each image is treated as a random variable where pixel intensity values define the distribution. For the two images, **M** and **F**, MI examines how the intensities of **M** correspond with those of **F**. MI relies on the joint probability distribution of intensity values from the two images. For each pixel in image **M** with intensity $\mathbf{I_M}$ and a corresponding pixel in image **F** with intensity $\mathbf{I_F}$ the joint distribution $P(\mathbf{I_M}, \mathbf{I_F})$ shows how often these intensities co-occur. The goal is to find the transformation that maximises the MI between the two images, which means the two images are most informative about each other and hence are best aligned.

The MI between the fixed image **F** and the moving image **M** is defined as:

$$I(\mathbf{M}; \mathbf{F}) = H(\mathbf{M}) + H(\mathbf{F}) - H(\mathbf{M}, \mathbf{F}) \tag{3.4}$$

where $H()$ is a function which computes the entropy of the image representing uncertainty in its intensity values. $H(\mathbf{M}, \mathbf{F})$ represents the joint entropy of the two images, representing the uncertainty of their combined intensity values. $H(\mathbf{M}, \mathbf{F})$ is defined

as:

$$H(\mathbf{M}, \mathbf{F}) = -\sum_{m \in M} \sum_{f \in F} \mathcal{P}(m, f) log_2 \mathcal{P}(m, f) \tag{3.5}$$

where $\mathcal{P}(m, f)$ is the joint probability of intensity pairs m and f from the two images. When the images are well aligned the joint entropy decreases and the MI increases.

The optimisation algorithm drives the direction in which updates to the transformation parameters are made after each iteration. Since MI is higher with better transformations, the metric needs to be maximised. Gradient ascent optimisation maximises a function by iteratively moving towards the direction of the steepest increase in the value of the function.

Specifically let $\theta = [\mathbf{R}, \mathbf{t}]$ from Equation 3.1 be the transformation parameters. At each iteration, the parameters θ are updated using the gradient ascent rule:

$$\theta^{(t+1)} = \theta^{(t)} + \alpha \nabla I(\mathbf{F}, \mathbf{M}, \theta^{(t)}) \tag{3.6}$$

where $\theta^{(t)}$ represents the current transformation parameters, α is the learning rate, controlling the step size and, $\nabla \mathbf{I}(\mathbf{F}, \mathbf{M}, \theta^{(t)})$ is the gradient of MI with respect to the transformation parameters at iteration t.

After applying the transformation parameters to the fixed image in the current iteration to transform $\mathbf{F}$ to be better aligned with $\mathbf{M}$, most transformed pixels will be off the defined pixel locations, requiring interpolation at every grid point in the image. To address this, we utilise simple nearest neighbour interpolation. Nearest neighbour interpolation is a method used to estimate the value of a function at a given point by selecting the value of the nearest known data point. Nearest neighbour is computationally inexpensive, as it only requires finding the closest pixel and copying its value, with no need for complex calculations.

Next, the elastic deformations of the images need to be catered for. For this we utilise SynthMorph, a training strategy for VoxelMorph, which is a neural network that learns deformable registration parameters using the Spatial Transformer Network architecture. We next introduce the core concepts of the Spatial Transformer Network, VoxelMorph, and SynthMorph.

3.1.2 | Spatial transformer networks

A Spatial Transformer Network (STN) is an architecture of neural networks that allows for the spatial manipulation and transformation of data within the network [133]. An STN is a differentiable module and can thus be inserted into any convolutional architecture enabling the network to actively spatially transform feature maps, conditional

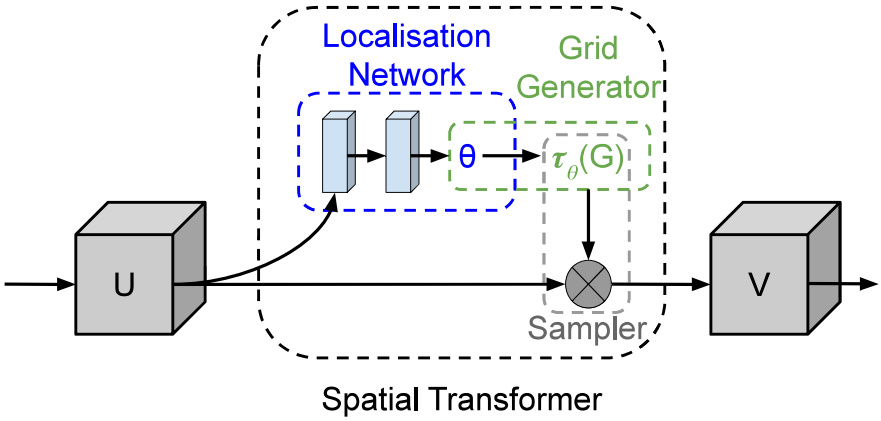


Figure 3.2: Spatial transformer module architecture. The input feature map, U, is passed to a localisation network which regresses the transformation parameters, θ. The regular spatial grid, G over V, is transformed to the sampling grid, $\tau_\theta(G)$, which is applied to U, producing the warped output feature map, V.

on the feature map themselves, without any additional training or supervision [133]. The main purpose of an STN is is to allow neural networks to learn spatial transformations such as scaling, translation, and rotation, in an end-to-end manner. This capability makes them particularly useful in tasks that involve handling and processing images, where the ability to align or transform the input data can significantly improve performance, such as feature extraction and matching, object detection and recognition and image segmentation [133].

The spatial transformer mechanism consists of three parts. A localisation network takes a feature map as an input and through a series of hidden layers outputs the parameters of the spatial transformation that should be applied to the feature map. Then, the grid generator takes the transformation parameters and creates a sampling grid, a set of points where the input map should be sampled to produce the transformed output. Finally, the sampler produces the output map which is sampled from the input feature map at the grid points. The combination in series of these three components forms a spatial transformer as seen in Figure 3.2 [133]. The spatial transformer, which is a self-contained module, can be dropped into a network and performs explicit spatial transformations of features. The STN is learned in an end-to-end fashion, without making any changes to the loss function of the network that it is dropped in [133]. The three STN components are explained in more detail in the next sections.

3.1.2.1 | Localisation network

The localisation network takes as input the feature map $\mathrm{U} \in \mathbb{R}^{H \times W \times C}$ with height H, width W and C channels and, outputs θ, the parameters of the transformation τ_θ to be applied to the feature map via:

$$\theta = f_{loc}(\mathrm{U}) \tag{3.7}$$

where $f_{loc}()$ is the localisation network function [133]. The function can take any form, such as a fully-connected network or a convolutional network, but should include a final regression layer to produce the transformation parameters θ. The size of θ can vary depending on the transformation type that is parameterised, for example for a rigid transformation as in Equation 3.1, θ is 6-dimensional [133].

3.1.2.2 | Parameterised sampling grid

After determining the transformation parameters θ, the grid generator creates a sampling grid, which tells the network how to sample pixels from the input. Essentially, the grid generator defines where each output pixel should come from in the input image. Hereunder, by pixel we mean an element of a generic feature map and not necessarily an image. The output pixels are defined to lie on a regular grid $G = \{G_i\}$ of pixels $G_i = (x_i^t, y_i^t)$, forming an output feature map $\mathrm{V} \in \mathbb{R}^{H' \times W' \times C}$ where H' and W' are the height and width of the grid respectively and C is the number of channel, which is the same in the input and output.

Assume for the time being that τ_θ is a 2-dimensional affine transformation, A_θ, the point-wise transformation can be defined by:

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \tau_\theta(G_i) = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \tag{3.8}$$

where (x_i^t, y_i^t) are the target coordinates of the regular grid in the output feature map and (x_i^s, y_i^s) are the source coordinates in the input feature map that define the sample points. Coordinates are height and width normalised such that $-1 \leq x_i^s, y_i^s \leq 1$ when within the spatial bounds of the input and, $-1 \leq x_i^t, y_i^t \leq 1$ when within the spatial bounds of the output. The transformation can have any parameterised form, provided that it is differentiable with respect to the parameters. This crucially allows gradients to be back-propagated through from the sample points $\tau_\theta(G_i)$ to the localisation network output θ.

3.1.2.3 | Differentiable image sampling

The sampler takes the grid generated by the grid generator and uses it to interpolate the input feature map, producing the transformed output. To perform a spatial transformation of the input feature map, a sampler must take the set of sampling points $\tau_\theta(G)$, along with the input feature map U and produce the sampled output feature map V [133]. Each coordinate, (x_i^s, y_i^s), in $\tau_\theta(G)$ defines the spatial location in the input where a sampling kernel is applied to get the value at a particular pixel in the output V. This can be defined as:

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c k(x_i^s - m; \Phi_x) k(x_i^s - n; \Phi_y) \; \forall i \in [1 \cdots H'W'] \; \forall c \in [1 \cdots C] \tag{3.9}$$

where U_{nm}^c is the value at location (n, m) at channel c of the input, $k()$ is a generic sampling kernel which defines the image interpolation and Φ_x and Φ_y are its parameters and V_i^c is the output value for pixel i at location (x_i^t, y_i^t) in channel c. It should be noted that the sampling is done identically for each channel of the input such that, every channel is transformed in an identical way.

In theory, any sampling kernel can be used as long as gradients can be defined with respect to (x_i^s, y_i^s). For example, a nearest neighbour sampling kernel can be used reducing Equation 3.9 to:

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c max(0, 1 - |x_i^s - m|) max(0, 1 - |y_i^s - n|) \tag{3.10}$$

The combination of the localisation network, grid generator, and sampler creates a spatial transformer, which is a modular component that can be seamlessly integrated into a CNN architecture at any point and in any quantity, forming spatial transformer networks. This module is computationally efficient and does not significantly slow down training, often adding minimal overhead or even speeding up attentive models through downsampling of the output of the transformer. Incorporating spatial transformers into a CNN enables the network to learn how to actively adjust feature maps to minimise the overall cost function during training. The learned transformations for each sample are encoded in the weights of the localisation network during the training process. In some cases, passing the output of the localisation network, θ, forward through the network can be beneficial, as it encodes the transformation or pose of a region or object. Spatial transformers can also down-sample or up-sample feature maps by defining different output dimensions (H' and W') from the input dimensions (H and W), though using fixed, small sampling kernels like the bilinear kernel can introduce aliasing when down-sampling. Multiple spatial transformers can be used in a CNN, either at increasing network depths to handle more abstract representations or in parallel to focus on different objects or regions within a feature map. However, a limitation of this architecture in purely feed-forward networks is that the number of parallel spatial transformers limits the number of objects the network can model.

VoxelMorph and SynthMorph are a pair of neural networks which learn deformable registration parameters using STN architecture. This family of networks was chosen because it represents the state-of-the-art method in elastic registration of medical images when measured on various datasets. The chosen networks also have the ability to register images with varying contrasts, a feature of thermal data which has been a challenge in the registration process of thermal image frames in a video due to the changing temperature along time.

3.1.3 | VoxelMorph and SynthMorph

Deformable or elastic registration is the registration between a pair of n-D image volumes via a dense, non-linear correspondence [134]. VoxelMorph is one method that utilises deep learning for the registration of 3D medical images [134]. Specifically, it is designed for the registration of 3D MRI data but it can be easily adapted to other modalities as well as 2D data. VoxelMorph learns a deformable, pairwise registration and obtains state-of-the-art results while operating orders of magnitude faster than similar models, such as ANTs [135] and NiftyReg [134]. The speed of VoxelMorph comes from the fact that the network does not require an optimisation on each image test pair.

VoxelMorph learns a parameterised registration function from a collection of volumes using a convolutional neural network (CNN), that takes as input the image pair consisting of a fixed reference image and a moving image, and outputs a mapping of each pixel from one image to another [134].

Specifically, let $\mathbf{F}$ and $\mathbf{M}$ be the fixed and moving images, respectively. Thermal images are effectively single-channel grayscale data, which aids in simplifying their representation as $\mathbf{F}$ and $\mathbf{M}$ and the network architecture. Assume that $\mathbf{F}$ and $\mathbf{M}$ are affinely aligned as a pre-processing step, such that the remaining sources of misalignment between the image pair are primarily nonlinear [134]. VoxelMorph is a deep-learning-based method for image registration that parameterises the transformation model using a neural network. In VoxelMorph, a function $g_\theta(\mathbf{F}, \mathbf{M}) = \phi$ is modeled using a CNN, where ϕ is a registration field defined as a vector representing the mapping of each pixel and θ are learnable parameters of g. For each pixel $\mathbf{x}$, $\phi(\mathbf{x})$ is a location such that $\mathbf{F}(\mathbf{x})$ and $\mathbf{M}(\phi(\mathbf{x}))$ define identical anatomical locations [134]. The network takes $\mathbf{F}$ and $\mathbf{M}$ as inputs and computes ϕ using a set of parameters θ. A spatial transformer function [133], as defined in the previous section, is used to warp $\mathbf{M}$ to $\mathbf{M}(\phi)$ which enables the evaluation of the similarity between $\mathbf{M}(\phi)$ and $\mathbf{F}$ [134]. Single-element stochastic gradient descent is used to find the optimal values of parameters θ by minimising an expected loss function using a training dataset [134].

The parameterisation of $g_\theta(\cdot, \cdot)$ is based on a CNN which consists of encoder and decoder sections with skip connections, similar to UNet [106], as depicted in Figure 3.3.

VoxelMorph may use one of two loss functions: an unsupervised loss, that evaluates the model using only the input images and the generated registration fields, and an auxiliary loss that leverages information in anatomical segmentations at training time [134]. For the purpose of this work, only the unsupervised loss will be discussed, since anatomical segmentations are not available and cannot be reliably extracted from our data.

The unsupervised loss, $\mathcal{L}(\cdot, \cdot, \cdot)$ consists of two components and can be expressed

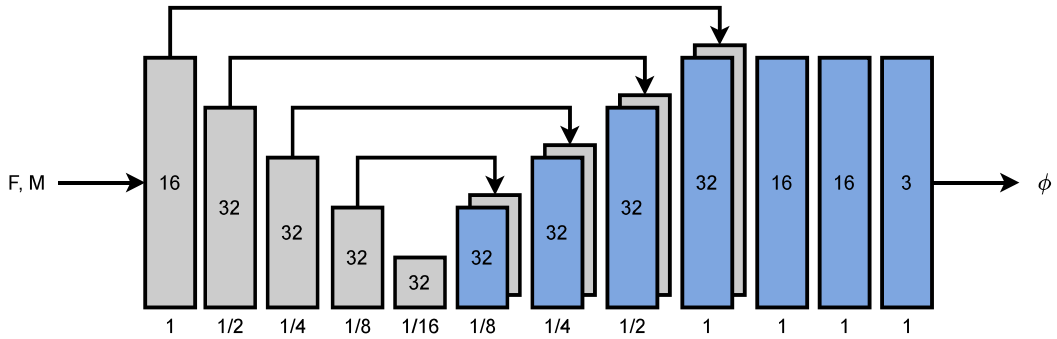


Figure 3.3: Convolutional UNet architecture which implements $g_\theta(\mathbf{F}, \mathbf{M})$ Each rectangle represents a 3D volume, generated from the preceding volume using a 3D convolutional network layer. The depth of the convolution layer is shown inside each block. The spatial resolution of each volume with respect to the input volume is shown underneath. Arrows represent skip connections.

as

$$\mathcal{L}(\mathbf{F}, \mathbf{M}, \phi) = \mathcal{L}_{sim}(\mathbf{F}, \mathbf{M}(\phi)) + \lambda \mathcal{L}_{smooth}(\phi) \tag{3.11}$$

where $\mathbf{M}(\phi)$ is the warped image of $\mathbf{M}$, function $\mathcal{L}_{sim}(\cdot, \cdot)$ measures image similarity between $\mathbf{M}(\phi)$ and $\mathbf{F}$, $\mathcal{L}_{smooth}(\cdot)$ imposes regularisation on ϕ, λ is the regularisation parameter, and ϕ is a displacement vector field, specifying the vector offset from $\mathbf{F}$ to $\mathbf{M}$ for each pixel.

For $\mathcal{L}_{sim}$, the local cross-correlation of $\mathbf{F}$ and $\mathbf{M} \circ \phi$, the warped image $\mathbf{M}$, is utilised, due to its robustness to intensity variations found across different images, which is a feature of our data [134]. Specifically, let $\hat{\mathbf{F}}(\mathbf{P})$ and $[\hat{\mathbf{M}} \circ \phi](\mathbf{P})$ denote images with local mean intensities:

$$\hat{\mathbf{F}}(\mathbf{P}) = \frac{1}{n^3} \sum_{\mathbf{P}_i} \mathbf{F}(\mathbf{P}_i) \tag{3.12}$$

where $\mathbf{P}_i$ iterates over an n^3 volume around a voxel $\mathbf{P}$, where n is set to 9 in our experiments. $\mathbf{M}(\phi)$ is obtained in the same way as defined by Equation 3.12. The local cross-correlation of $\mathbf{F}$ and $\mathbf{M} \circ \phi$ is written as:

$$CC(\mathbf{F}, \mathbf{M} \circ \phi) = \sum_{\mathbf{P} \in \Omega} \frac{(\sum_{\mathbf{P}_i}(\mathbf{F}(\mathbf{P}_i) - \hat{\mathbf{F}}(\mathbf{P}))([\mathbf{M} \circ \phi](\mathbf{P}_i) - [\hat{\mathbf{M}} \circ \phi](\mathbf{P})))^2}{(\sum_{\mathbf{P}_i}(\mathbf{F}(\mathbf{P}_i) - \hat{\mathbf{F}}(\mathbf{P}))^2)(\sum_{\mathbf{P}_i}([\mathbf{M} \circ \phi](\mathbf{P}_i) - [\hat{\mathbf{M}} \circ \phi](\mathbf{P}))^2)} \tag{3.13}$$

where $CC()$ represents the cross-correlation function. A higher CC indicates a better alignment of the two images, yielding the loss $\mathcal{L}_{sim}(\mathbf{F}, \mathbf{M}, \phi) = -CC(\mathbf{F}, \mathbf{M} \circ \phi)$.

Minimising $\mathcal{L}_{sim}$ will encourage $\mathbf{M} \circ \phi$ to approximate $\mathbf{F}$ but may generate a non-smooth ϕ that is not physically realistic. Thus a smoothing loss, $\mathcal{L}_{smooth}$, is utilised. $\mathcal{L}_{smooth}$ is a diffusion regulariser on the spatial gradients of displacement $\mathbf{u}$

$$\mathcal{L}_{smooth}(\phi) = \sum_{P \in \Omega} ||\nabla \mathbf{u}(\mathbf{P})||^2 \tag{3.14}$$

and approximate spatial gradients using differences between neighbouring voxels [134].

The costly optimisation of classical registration methods for each image pair is replaced by a global function optimisation during a single training phase. Registration between a new test image pair is achieved by evaluating the learned function on the given image pair, resulting in rapid registration [134].

The drawback of utilising VoxelMorph is that it requires a substantial dataset for training that consists of test image pairs and the transformation function that is required to align them. Such a large amount of data was not available to us, as there are no datasets for similar applications which are available for use. Collecting such data is also difficult, as a vast amount is required. We therefore opted for using SynthMorph, a strategy for learning deformable image registration that utilises the VoxelMorph network architecture [69]. SynthMorph does not require any pre-existing image data but utilises synthesised data that simulate deformations to train the VoxelMorph network [69]. The SynthMorph strategy produces networks that are agnostic to contrast differences introduced by images acquired using different modalities or from different subjects since the synthesised data on which it is trained has contrast differences whose magnitude far exceeds those expected in real data [69]. SynthMorph enables the registration of real images whose contrasts vary, whilst learning only from synthetically generated data whose contrast far exceeds the range in realistic medical images [69]. Figure 3.4 shows the unsupervised learning strategy that is used for contrast-agnostic registration. SynthMorph creates two label maps which are used to generate intensity images whereby each label defines an intensity distribution, and trains a deformation network, h_θ, to register the label maps (rather than the intensity images themselves), thus making the method robust to the region intensities.

Specifically, paired label maps, $\{s_m, s_f\}$, are generated using a function $g_s(z) = \{s_m, s_f\}$ given random seed z [69]. Next, another function, $g_I(s_m, s_f, \tilde{z}) = \{m, f\}$, generates two intensity images based on the maps $\{s_m, s_f\}$ and random seed $\tilde{z}$ [69]. These two synthesised intensity images are then used as input to a modality agnostic CNN, which is defined using the VoxelMorph architecture as described above [69]. The interested reader is referred to [69] for a detailed explanation of the synthesisation process of training data.

During the training process, hyperparameters are chosen based on visual inspection of results to yield shapes and contrasts between images that far exceed the expected range in medical images. This forces the network to learn generalisable features that are independent of contrast [69].

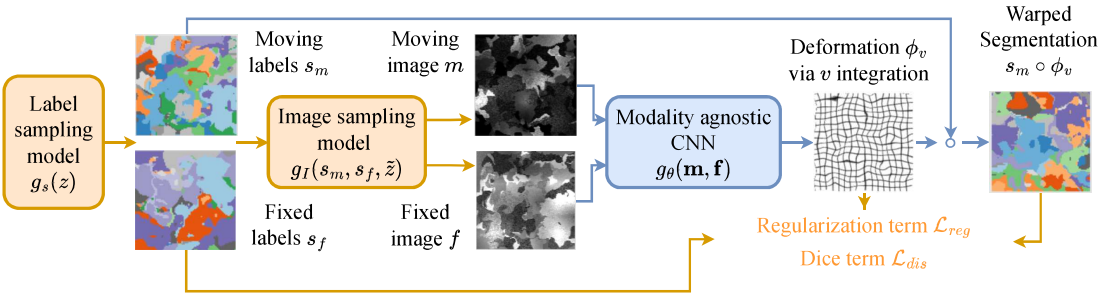


Figure 3.4: Unsupervised learning strategy for contrast-agnostic image registration where the modality agnostic CNN is the VoxelMorph network defined above.

The methods described above have been implemented and tested on data acquired for the purpose of this work. A description of the methodology used for the experimental implementation follows.

3.1.4 | Data acquisition

For the purpose of this study, dynamic thermal data were collected from the plantar surfaces of human feet. Nine healthy adult participants, comprising five males and four females, provided informed consent to take part in the study, which received approval from the University of Malta Research Ethics Committee. A copy of the signed consent form is included in Appendix 7. The mean age of the participants was 26.3 $\pm$4.8 years. In accordance with standard medical imaging protocols [65], participants were instructed to avoid exercise for up to six hours prior to the recording session and to refrain from consuming large meals or excessive warm beverages beforehand.

During data acquisition, participants removed their shoes and were seated in a semi-Fowler's position on a couch, with both feet positioned perpendicular to the thermal camera and within its field of view. They were asked to minimise movement of their legs and feet throughout the one-hour thermal video recording session.

Thermal videos were recorded using a FLIR SC7200 camera, which features a detector resolution of 320 $\times$ 256 pixels and a temperature resolution of 20 mK with an accuracy of $\pm$1 K. The ambient temperature in the examination room was maintained at a constant 24°C $\pm$0.5°C for at least one hour prior to and during data acquisition. Thermal frames were captured at a rate of one frame per minute. To provide a uniform and stable background, a thermally neutral cloth was used behind the feet, reducing background clutter and ensuring consistency throughout the video sequence. Figure 3.5 presents sample data from an actual participant, illustrating both foot movement across frames and the evolving thermal patterns over time.

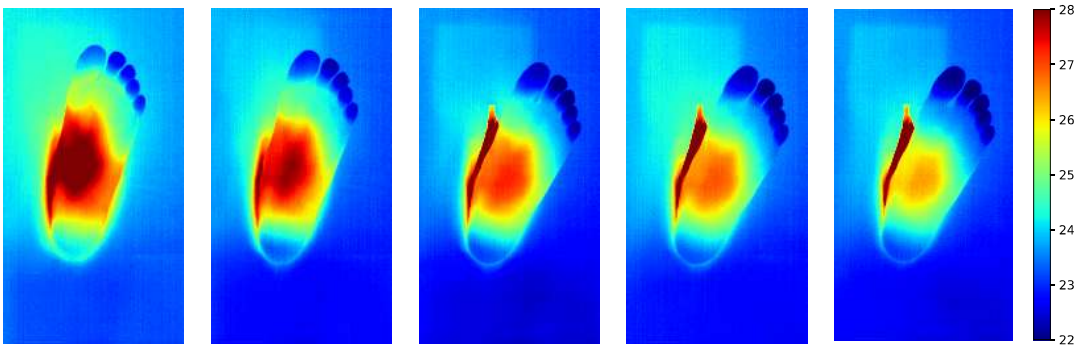


Figure 3.5: Sample of the data acquired for the purpose of this study showing the frames at the beginning of acquisition and at 15 minute intervals from there-after.

3.1.5 | Pre-processing

Each foot in the thermal video was analysed individually. Every frame in each sequence was min-max normalised to the range $[0, 1]$ by using

$$F_{norm} = \frac{F - F_{Min}}{F_{Max} - F_{Min}} \tag{3.15}$$

where F_{norm} is the normalised frame, F is the original frame with temperature values as pixels, and F_{Max} and F_{Min} are the maximum and minimum values in F respectively. This normalisation is a requirement of the SynthMorph and VoxelMorph architectures and is reversed after the registration process is complete.

3.1.6 | Thermal video registration

Consider a dynamic thermal sequence with N frames, V_n, which captures the plantar aspect of a human foot. Since the foot is not static, it does not occupy the same spatial locations throughout the entire video. In medical imaging, the initial position of the body region of interest is typically the preferred one since, the subjects are placed in this preferred position prior to the start of the data collection and in time they are prone to move away from this ideal position, thus necessitating registration of the sequence of images. Therefore, the reference image, F_{ref}, is chosen to be the first frame in the sequence, with all other frames having to be aligned to this reference frame. The differences between subsequent frames in the thermal sequence of a body region are expected to be rotations and translations of the body region, resulting in both linear and non-linear deformations as can be seen in Figure 3.5. The thermal videos acquired for this work represent a particular challenge for registration. Since the videos contain both feet, in the foreground, whose temperature is expected to vary with time, and the background, whose temperature is expected to remain stable throughout the video, the foreground-background contrast is expected to change during the course

of the video. Thus a contrast-agnostic registration framework, SynthMorph, was selected for the registration of our thermal data. However, SynthMorph requires the input images to be linearly aligned prior to registration. Thus the additional step of rigid registration to align the linear deformations was taken.

An iterative process is used to remove the linear deformations. The process starts by aligning the second frame in the sequence to F_{ref} by using a rigid transformation as described in Section 3.1.1. The transformation matrix that achieves the desired registration is extracted and saved. Next, the moving image in the registration process is selected to be the next frame in the sequence, in this case, the third frame, and the fixed image is selected to be the previous, unregistered, moving image, in this case, the second frame. Once again the transformation matrix that achieves the desired registration is extracted and saved. The transformation matrix of the latest transformation is multiplied to the product of all previous transformation matrices and the resulting transformation matrix is used to align the current moving image to F_{ref}. This process is repeated $N-1$ times such that all frames are aligned to F_{ref}. This approach was preferred over the alternative, that of registering each frame to the first frame, because from experimenting with the dataset it resulted that for some subjects the changes between the first frames and the latter frames in the sequence was sometimes too big, resulting in errors in the registration.

The rigid transformation was implemented using the ANTsPY toolbox [136] that implements the ANTs [137] toolbox, a state-of-the-art medical image registration and segmentation toolkit, in Python. As a similarity metric, mutual information (MI) is used with gradient ascent optimisation and the transformation is applied to the images using nearest neighbor interpolation. For each image pair, the optimisation stops at the fifth iteration, as test results have concluded that more iterations offer a negligible improvement in the registration.

Following the linear pre-registration, the deformable registration process is applied, such that the fine details of the feet are aligned throughout the video frames. For this purpose, the SynthMorph registration strategy [69], as described in Section 3.1.3 is utilised. Thus, aligning frames from different parts of the video to each other requires the alignment of images with different contrasts.

SynthMorph network weights were downloaded from this link[1]. These weights were trained on random shapes with varying contrasts on synthetic images and could thus be used directly without the need for training from scratch. The parameters and methods used to obtain these model weights are described in detail in [69]. The linearly aligned thermal video frames are fed one by one to the pre-trained SynthMorph network and all frames are individually aligned to F_{ref}. The SynthMorph network

[1]https://nmr.mgh.harvard.edu/ mu40/synthmorph/

outputs the aligned video frames and the displacement field related to each frame pair.

3.1.7 | Post-processing

After the registration process is complete, the resulting sequence has frames that are min-max normalised. In order to recover the temperature values in the registered frames, Equation (3.15) is reversed and applied to each frame. This leads to

$$F_{reg} = (F_{Norm} \times (F_{Max} - F_{Min})) + F_{Min} \tag{3.16}$$

where F_{reg} is the registered frame, scaled back to the original temperature values, F_{Norm} is the normalised and registered frame, and F_{Max} and F_{Min} are the maximum and minimum values in the original frame used in Equation (3.15).

3.1.8 | Testing methodology

To analyse the results obtained from the registration algorithms, a visual inspection of the registered thermal sequences, side-by-side with the original sequence and overlaid with a grid, was done. Such an inspection gives a good indication of the performance of the algorithms and can give insight as to which scenarios prove to be difficult for registration. The registered thermal sequences were classified as being the results of a correct registration or not. A correct registration was one which results in a thermal sequence in which any region of the frame contained the same anatomical location throughout the thermal video. That is, anatomical locations, such as the toes, occupied the same pixels in all frames of the thermal sequence. The resulting registered thermal sequence was placed side-by-side with the corresponding original thermal sequence and saved in form of a video. A rater, with experience in thermal imagingm, played the video frame-by-frame. The rater, with the help of the grid, determined whether the foot occupied the same pixel locations and whether the structure and ratio of the foot were maintained.
Another test was devised in which the difference over time of the position of points placed manually on salient points of the foot was computed to quantify the amount of remaining movement of the foot after the registration process.

Five points were selected on salient points of the foot, as shown in Figure 3.6. The points were selected on the edges at the base between the hallux and the second toe, the edges at the top between the second and third toes, third and fourth toes, and, fourth and fifth toes, and the center of the arc at the edge of the heel. These five points were selected on the first frame of the registered sequence and at each frame at 15-minute intervals until the last frame. The total cumulative position difference for each point across all frames is then computed as a measure of how much these

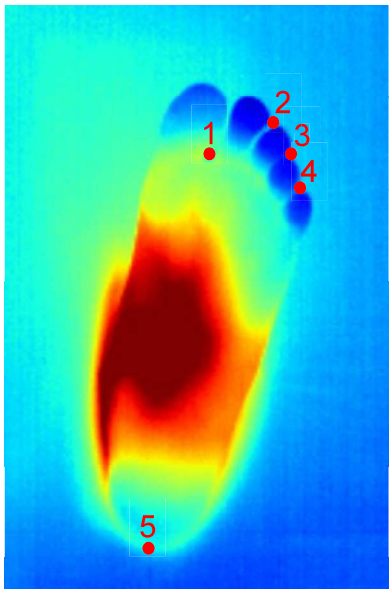


Figure 3.6: The five points selected on the foot for carrying out the second test.

salient points move throughout the registered thermal video. In an ideal scenario, the distance between corresponding points is zero, since an anatomical point should occupy the same pixel coordinates throughout the registered video. As a comparison, the process was carried out on the unregistered video sequences, as well as the registered ones, and, the before- and after-registration results are compared.

During the thermal sequence, any displacement of the foot can be considered an error, as ideally it should remain perfectly aligned throughout the entire sequence, resulting in a displacement of 0. This error has two components: the human labeling error, which is the error resulting from repeatedly, manually locating a point of interest, and the registration error, which is the error resulting from the methods described in this work.

Thus, a final test was devised to quantify the mean human labeling error. Fifteen frames were selected at random from the whole dataset and divided into groups of five. Three raters were given five frames each and were asked to demarcate a point, selected at random from the five points shown in Figure 3.6, on each frame. The demarcation process was repeated three times, such that each rater demarcates the same random point on the same frame, three times. The mean displacement between the three points on each frame for the three raters was then computed.

3.2 | Results

The result of the registration process on a sample thermal sequence is shown in Figure 3.7. The original, unregistered thermal sequence (top row) shows signs of movement of the foot throughout the sequence, mainly as a rotation in the plane of the image. The

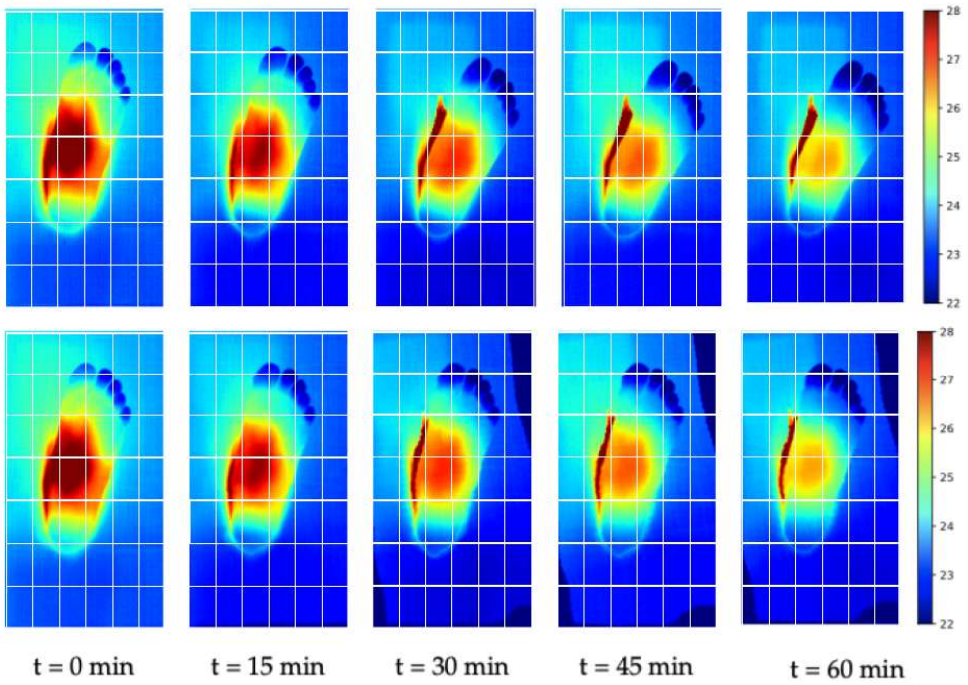


Figure 3.7: Results of the registration process. The top row shows the original, unregistered thermal sequence with the evident movement of the foot as the sequence progresses. The bottom row shows the same thermal sequence, registered using the rigid registration and SynthMorph network described in this work. By comparing the top images to the bottom images it is evident that the movement of the foot has been rectified by the registration algorithms.

registered sequence can be seen in the bottom row of the same figure. It is clear from this figure that the movement of the foot throughout the video has been corrected, and that the resulting sequence contains pixels that correspond to the same anatomical location throughout.

On a test dataset of 18 thermal sequences of human feet, 16 sequences were judged to have been registered correctly as there was no perceptible movement of the foot region, one sequence had a misalignment in the registration of the toes and in one sequence the overall structure and ratio of the foot were not maintained. This means that using the visual inspection test a success rate in registering thermal sequences of 88.8% was achieved.

Table 3.1 gives the results achieved from the five-point test. The mean displacement before the registration process was 22.56 pixels ($\pm$7.94) whilst the displacement after the registration was 6.43 pixels ($\pm$2.99), a reduction of 71.5%.

For the second test, the mean displacement between the three points on each frame for the three raters was computed to be 0.79 pixels ($\pm$0.44). This means that effectively around 8% of the remaining displacement after registration is due to human error.

Table 3.1: Test results. Each row represents the result from one particular thermal sequence. The first column represents the mean distance moved by the five points for each thermal sequence before the registration process. The second column represents the mean distance of the five points for the thermal sequences resulting from the registration process.

Sequence number	Participant Mean Displacement - Original (pixels)	Participant Mean Displacement - Registered (pixels)
Foot 1	11.76	3.48
Foot 2	15.27	4.86
Foot 3	28.78	7.14
Foot 4	32.01	4.62
Foot 5	7.31	4.27
Foot 6	11.22	5.11
Foot 7	25.74	12.91
Foot 8	30.19	13.65
Foot 9	21.90	2.93
Foot 10	35.64	5.13
Foot 11	19.11	5.89
Foot 12	19.43	4.86
Foot 13	22.79	5.85
Foot 14	28.79	11.63
Foot 15	25.92	6.30
Foot 16	24.44	5.62
Foot 17	13.94	5.24
Foot 18	31.86	6.17
Mean (STD)	22.56 ($\pm$7.94)	6.43 ($\pm$2.99)

These results are discussed in more detail in the next section.

3.3 | Discussion

Visual inspection of the registration results revealed that two sequences of the 18 were not satisfactorily registered. Both sequences exhibited the same issue: the toes in the images had similar intensity to each other and to the background. This made it challenging to distinguish the toes from the background and from each other, even visually. As a result of this overlapping intensity, the VoxelMorph network struggled to compute accurate transformation parameters, leading to incorrect warping of the toes as seen in Figure 3.8.

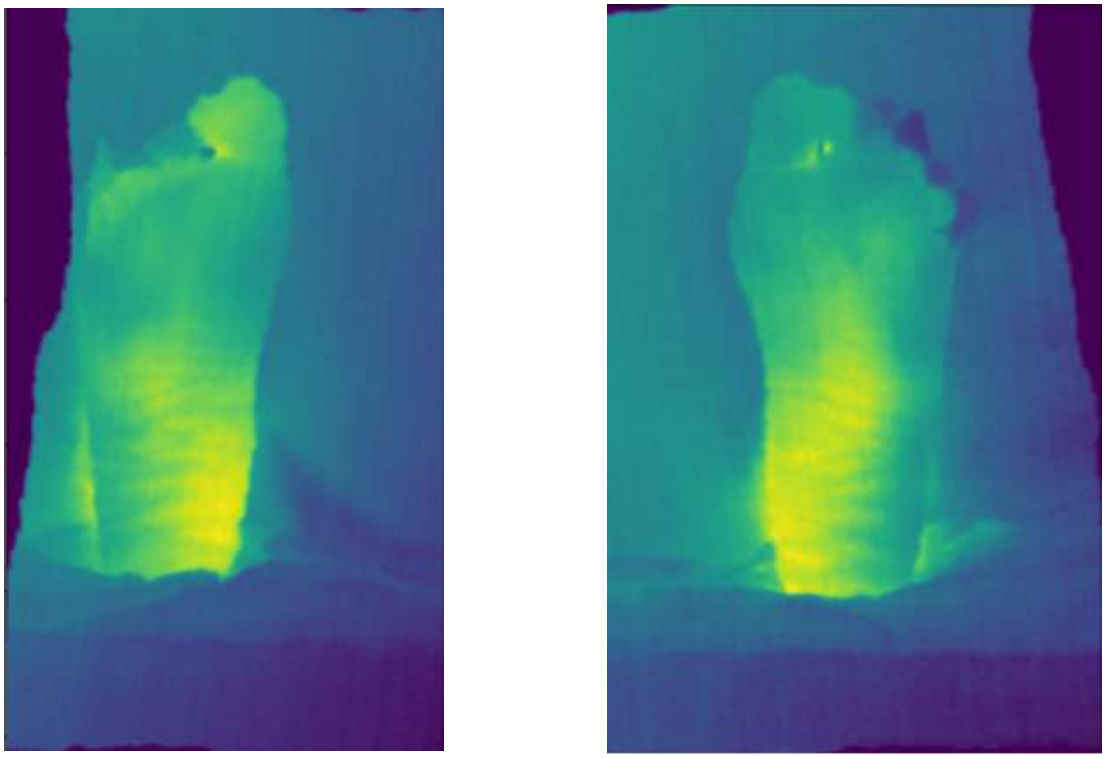


Figure 3.8: Errors in the registration process led to the toes becoming warped and misaligned in the final registration in two cases from the dataset of 18 videos.

It is worth noting that in both cases the majority, around 80%, of the sequence was registered correctly and thus the need for manual intervention to fix the registration issues was minimal.

The thermal neutral background used during the acquisition of the data also has other implications for the registration algorithm. The thermal-neutral background does not have a varying intensity throughout the thermal sequence, while the foot does vary due to the temperature change arising from acclimatisation to the room temperature. This leads to a varying background-foreground contrast throughout the video and hinders the accuracy of the registration when using classical registration techniques. SynthMorph, and its architecture VoxelMorph, were designed such that MRI images and sequences with varying contrasts could be registered. The problem of registering dynamic thermal sequence with varying background-foreground contrast is similar to the problem of registering MRI data captured with different contrasts and thus why SynthMorph was selected as the registration method of choice. From the results obtained in the previous section, it is clear that the registration of dynamic thermal videos benefits from the application of the SynthMorph network as accurate results are obtained.

The work and results on the registration algorithm have been accepted for publication at the Quantitative InfraRed Thermography Journal.

4

Dynamic Medical Thermal Data Segmentation

The image registration pipeline outlined in the previous chapter successfully aligns data such that each pixel corresponds to the same anatomical location of the foot across the videos. This alignment ensures that pixel-wise comparisons are anatomically consistent. However, the thermal data contains not only the foot but also the background of the video. To proceed with any meaningful analysis, it is crucial to differentiate between the foreground, the foot, and the background. Without this distinction, any subsequent thermal analysis risks being confounded by irrelevant data that does not pertain to the foot itself.

Before analysing the temperature dynamics of the foot, an essential step is to segment the foot from the background. This segmentation process identifies which pixels correspond to the foot and which belong to the background. By isolating the foot, further analysis can focus solely on the relevant anatomical region, ensuring that temperature readings and trends are derived from the actual subject of interest.

Beyond simple foreground segmentation, it is often advantageous to further subdivide the foot into key anatomical regions, such as the toes, arch, and heel.

In this study we suggest the use of Segment-Anything Model (SAM) [97], a state-of-the-art model designed for general-purpose image segmentation. SAM is designed to handle a wide variety of image segmentation tasks without needing task-specific tuning. It can generalise across different objects, contexts, and image domains. The model can segment objects based on text prompts, clicks, or bounding boxes provided by a user, making it a highly flexible tool for interactive segmentation.

We develop prompts which can be used to segment the registered thermal video data of the feet and assess how well SAM is able to perform the task at hand.

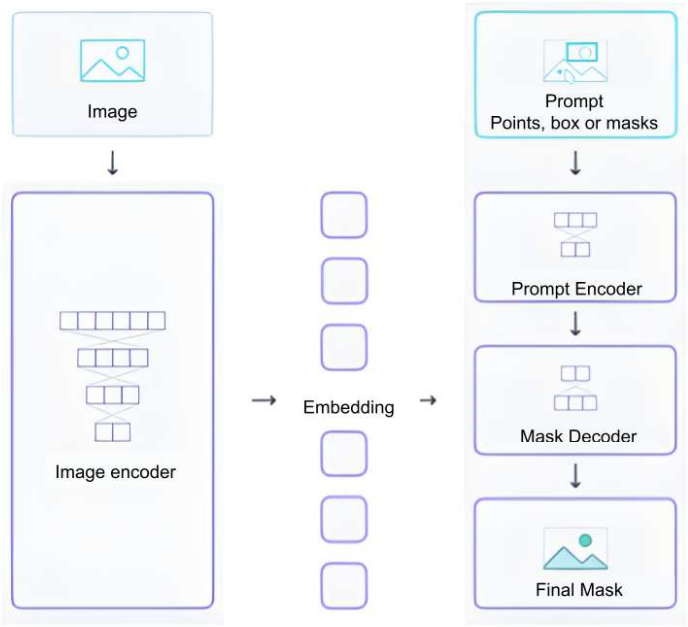


Figure 4.1: The architecture of the Segment Anything Model utilised for the scope of segmenting the registered dynamic thermal data.

4.1 | Segment Anything Model

Segment-Anything Model (SAM) has three components: an image encoder; a flexible prompt encoder, and; a fast mask decoder as shown in Figure 4.1.

4.1.1 | SAM - Image encoder

The image encoder can be any network that outputs a $C \times H \times W$ image embedding, where C is the number of channels of the input image and, H and W are the height and width respectively. Due to a strong pre-training, the authors of SAM use a Masked auto-encoder (MAE) [138] pre-trained Vision Transformer (ViT) [139] with minimal adaptations to process high resolution inputs. That is the same approach that we use in this work.

4.1.1.1 | Vision transformer model

The Vision Transformer (ViT) model is a deep learning architecture for image recognition. Inspired by the success of transformer models in natural language processing (NLP) (like BERT [140] and GPT [141]), ViT applies transformers to computer vision tasks, such as image classification, challenging the conventional reliance on the convolutional neural network (CNN) for this domain. Instead of processing the entire image as a single entity, ViT divides the input image into a grid of smaller, non-overlapping patches, often 16×16 pixels in size, but 14×14 in this work. Each patch

is then flattened into a vector. This approach lets the transformer model work with a sequence of fixed-sized inputs, similar to word embeddings in NLP.

The standard transformer receives as input a one-dimensional sequence of token embeddings. To handle the two-dimensional input image $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, it is reshaped into a sequence of flattened two-dimensional patches:

$$x_p \in \mathbb{R}^{N \times (P^2 \cdot C)} \tag{4.1}$$

where (H, W) is the resolution of the input image, C is the number of color channels, (P, P) is the resolution of each image patch, and $N = \frac{HW}{P^2}$ is the resulting number of patches, which also serves as the effective input sequence length for the Transformer.

The Transformer maintains a constant latent vector size D across all its layers, so the image patches are flattened and mapped to D dimensions using a trainable linear projection. The output from this projection is called the patch embeddings and can be defined as z_0 as:

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \cdots x_p^N E] + E_{pos}, \qquad E \in \mathbb{R}^{(P^2 \cdot C) \times D}, E_{pos} \in \mathbb{R}^{(N+1) \times D} \tag{4.2}$$

where $z_0^0 = x_{class}$ is a learnable token similar to BERT's $[class]$ token, whose state at the output of the Transformer encoder z_L^0 serves as the image representation y, where $y = LN(z_L^0)$ and $LN()$ is a Layernorm function which is applied before every block in the network, and; E_{pos} are one-dimensional position encoding used to retain positional information.

In both pre-training and fine-tuning, a classification head is attached to Z_L^0. During pre-training, this classification head is a multilayer perceptron (MLP) with one hidden layer, whereas in fine-tuning, it consists of a single linear layer. The resulting sequence of embedding vectors serves as input to the encoder. The resulting sequence of embedding vectors serves as input to the encoder.

The Transformer encoder consists of alternating layers of multi-headed self-attention (MSA) and MLP blocks. Layernorm is applied before every block, and residual connections after every block. The MLP contains two layers with a Gaussian Error Linear Unit (GELU) non-linearity.

4.1.1.2 | Masked auto-encoder

Masked auto-encoder (MAE) training is a self-supervised learning technique used primarily for training vision transformers. The goal is to reconstruct missing parts of an input image given only partial information, which forces the model to learn meaningful representations without the need for labelled data.

During training, random patches from the input image are masked out. For instance, up to 75% of the image patches may be masked out. This means that only a

small subset of the patches, usually around 25% is available to the model, creating a context where the model must infer the missing information.

The unmasked patches are fed into a ViT encoder as defined in the previous section. The encoder learns to represent the available patches while trying to capture enough context to reconstruct the missing ones. Since fewer patches are processed, training becomes more efficient than if the entire image were used.

A lightweight transformer-based decoder takes the encoder's output and the masked positions as input. Its role is to reconstruct the masked patches, producing a completed version of the image. The model is trained to minimise the difference between the reconstructed and the original unmasked patches, focusing on learning effective representations for the masked areas. By forcing the model to fill in missing information, MAE training enables the ViT to learn high-level semantic features that are not reliant on labelled data. This can lead to better generalisation and robustness across tasks.

SAM utilises an MAE pre-trained ViT with 14×14 windowed attention and four equally-spaced global attention blocks. The image encoder's output is a 16× downscaled embedding of the input image. Since goal of SAM is to process each prompt in real-time, a high number of image encoder floating-point operations per second (FLOPs) can be afforded because they are computed only once per image, not per prompt. SAM uses an input resolution of 1024×1024 obtained by rescaling the image and padding the shorter side. The image embedding is therefore 64×64. To reduce the channel dimension, a 1×1 convolution is used to get to 256 channels, followed by a 3×3 convolution also with 256 channels. Each convolution is followed by a layer normalisation.

4.1.2 | SAM - Prompt encoder

SAM can be prompted with two sets of prompts: sparse, example points, boxes or text, and; dense masks. In this work, the text and mask prompts were not utilised and thus we focus on the points and boxes prompts.

Points and boxes prompts are represented by positional encodings and then mapped to by 256-dimensional vectorial learned embeddings. A point is represented as the sum of a positional encoding of the point's location and one of two learned embeddings that indicate if the point is either in the foreground or background. A box is represented by an embedding pair consisting of: the positional encoding of its top-left corner summed with a learned embedding representing "top-left corner", and; the same structure but using a learned embedding indicating "bottom-right corner".

4.1.3 | SAM - Mask decoder

The mask decoder efficiently the image embedding and a set of prompt embeddings to an output mask. The standard Transformer decoder [142] is modified to combine the two inputs, image and prompt embeddings, and output a mask with the segmentation. Before the mask decoder a learned output token embedding is inserted into the set of prompt embeddings. This token will be used at the output of the decoder. For simplicity, the prompt and output token embeddings together are referred to as tokens. Each layer of the mask decoder performs 4 steps:

1. self-attention on the tokens;
2. cross-attention from tokens to the image embedding;
3. a point-wise MLP to update each token, and;
4. cross-attention from the image embedding to the tokens which updates the image embedding with prompt information

During the cross-attention step, the image embedding is treated as a set of 642 256-dimensional vectors. Each self/cross-attention and MLP has a residual connection, layer normalisation, and a dropout of 0.1 at training. The next decoder layer takes the updated tokens and the updated image embedding from the previous layer as input. SAM utilises a two-layer decoder.

To provide the decoder with essential geometric information, positional encodings are added to the image embeddings whenever they are involved in an attention layer. Furthermore, the complete set of original prompt tokens, including their positional encodings, are reintroduced to the updated tokens each time they participate in an attention layer. This ensures a strong reliance on both the geometric location and type of the prompt tokens.

After passing through the decoder, the updated image embedding is upsampled by $4\times$ using two transposed convolutional layers. At this stage, the image embedding is downscaled by $4\times$ relative to the input image. The tokens then attend to the image embedding once more. The updated output token embedding is subsequently fed into a small 3-layer MLP, producing a vector that matches the channel dimension of the upscaled image embedding. Finally, a mask is predicted by applying a spatially point-wise product between the upscaled image embedding and the MLP's output.

The transformer operates with an embedding dimension of 256. Its MLP blocks have a large internal dimension of 2048, though the MLP is applied solely to the prompt tokens, which are relatively few in number. However, in cross-attention layers that involve a 64×64 image embedding, the channel dimension of the queries, keys, and values are halved to 128 to improve computational efficiency.

All attention layers use eight heads. The transposed convolutions used to upscale the output image embedding are 2×2, stride 2 with output channel dimensions of 64

and 32 and have GELU activations. They are separated by layer normalisation.

4.2 | Methodology

SAM as explained in the previous sections was used to segment the dynamic thermal image sequences. Since the sequences were already registered, it was only required for the first frame to be segmented and, the rest of the sequence could be segmented using the same mask.

The segmentation process requires a manual input, to select a box roughly around the foot in the image. When the segmentation results need to be refined further a point can be added on the region to be added or removed from the segmentation mask.

Since SAM requires colour images (not grayscale with one channel) the first frame of each sequence was colourised by applying a colour palette to it and was then fed to the network. The code [1] for the network handles the user interface which accepts prompts from the user, in forms of a box or points and overlays the resulting mask on the input image. The mask is then made available to the user in form of a binary image for each of the resulting regions (background and foot in our experiments).

The same data from the nine participants which was used for the registration process in Chapter 3 was used for the segmentation method experiments and the left and right feet were analysed individually, resulting in 18 input sequences to be segmented.

Results were visually inspected for alignment of the mask with the foot boundaries. A satisfactory segmentation was one in which the segmentation mask overlapped all of the foot stopping at the foot boundaries and not overlapping with the background.

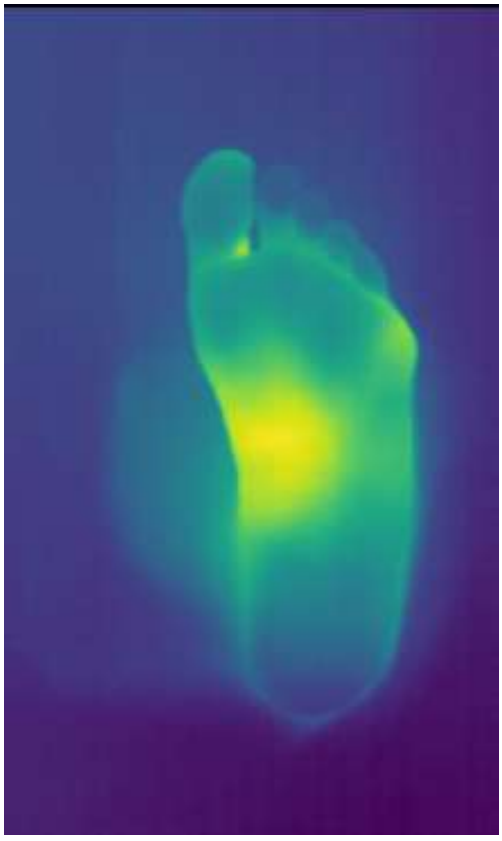
In the next section, the results for the segmentation algorithm are provided and a discussion of the results ensues.

4.3 | Results

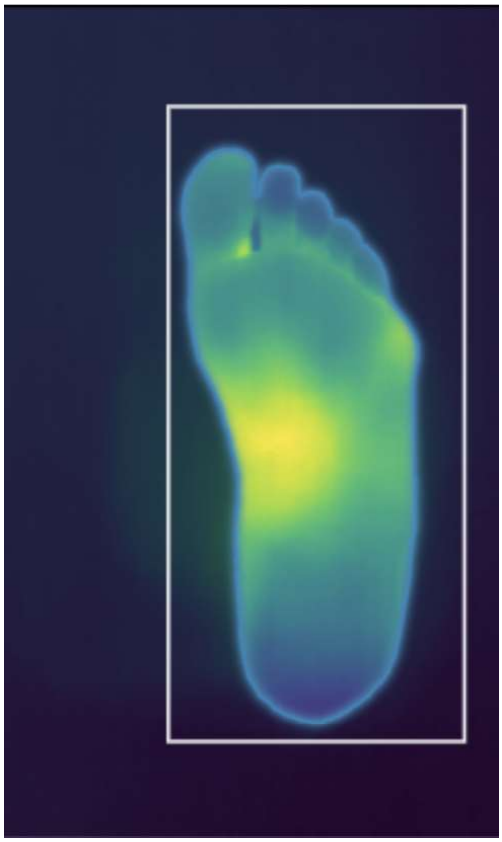
A sample of the registration results is shown in Figure 2 while, the full results are presented in Appendix 7

From 18 dynamic thermal sequences which were segmented using the method presented above, all of the sequences were satisfactorily segmented, with three of the sequences requiring a further refinement by way of a point prompt after the box prompt was inputted (See Figure 2(d)).

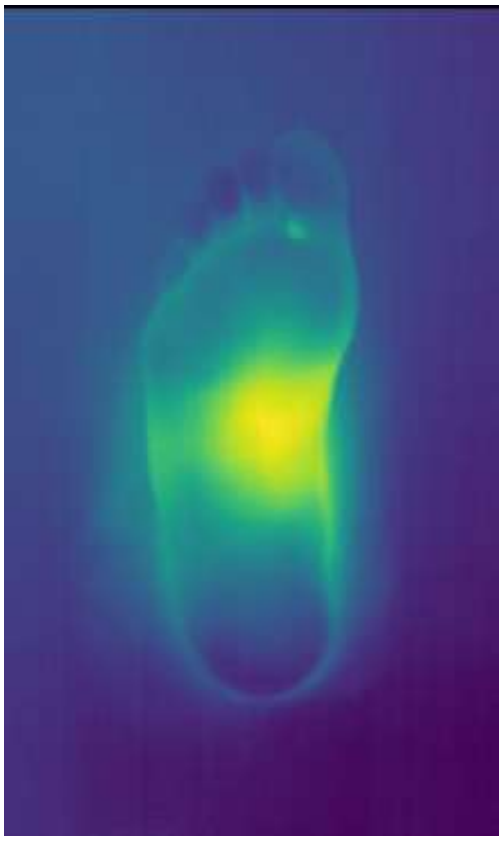
[1] The code for the network, user interface and, the pre-trained weights is freely available at `https://github.com/facebookresearch/segment-anything`



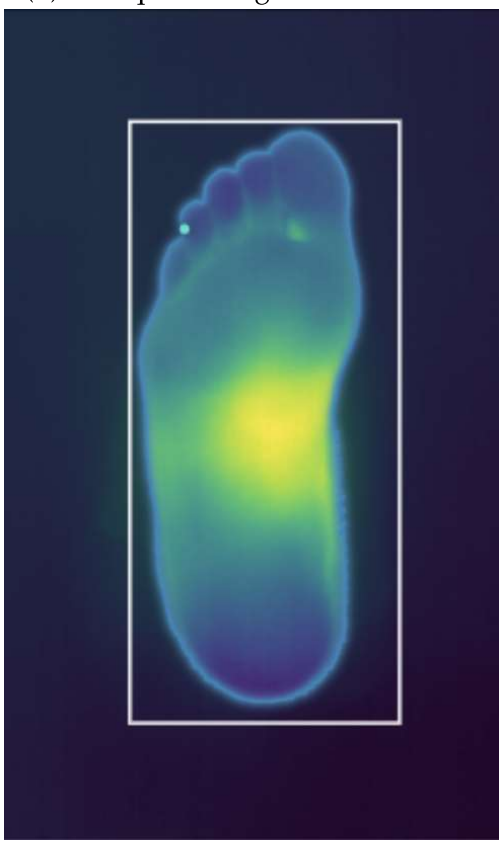
(a) Input image



(b) Prompt and segmentation mask

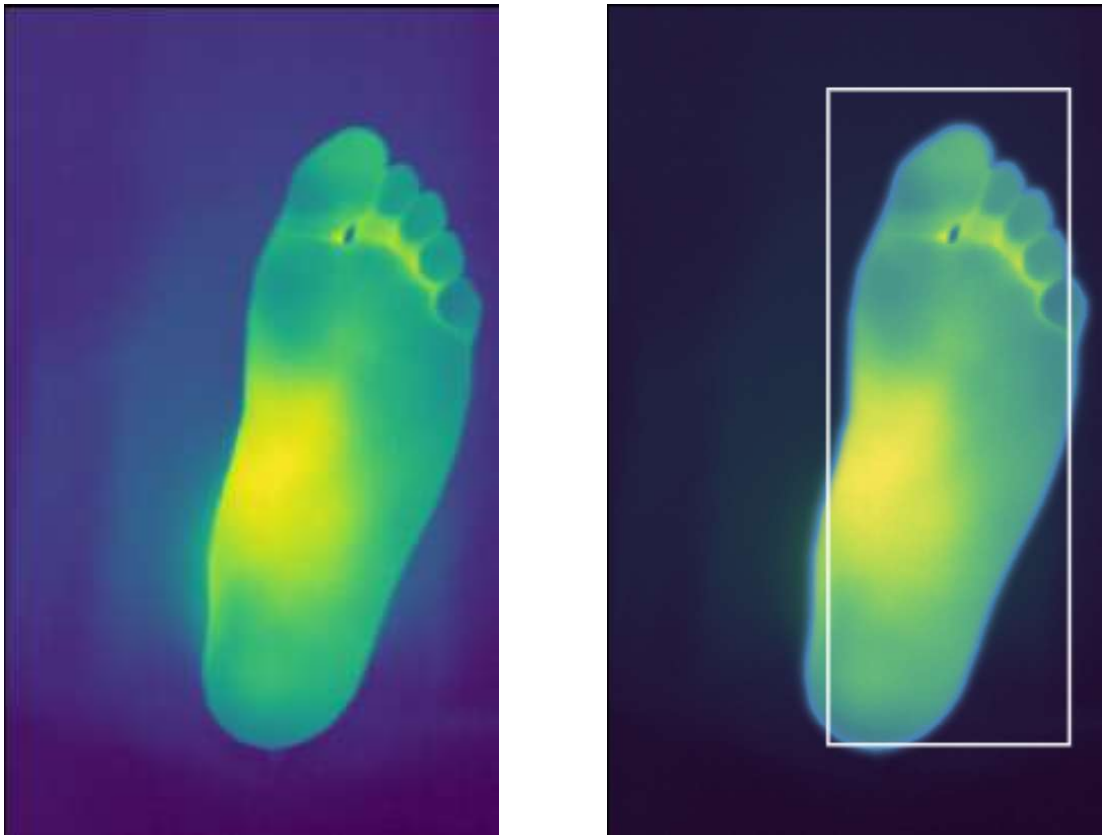


(c) Input image



(d) Prompt and segmentation mask

Figure 4.2: A sample of the results from the segmentation algorithm showing the input images on the left and the input prompts and output segmentation mask overlayed on the input image on the right.



(a) Input image

(d) Inaccurate prompt and accurate segmentation mask

Figure 4.3: The bounding-box-like box promt does not need to be very accurate as the model is resilient to some errors in its placement.

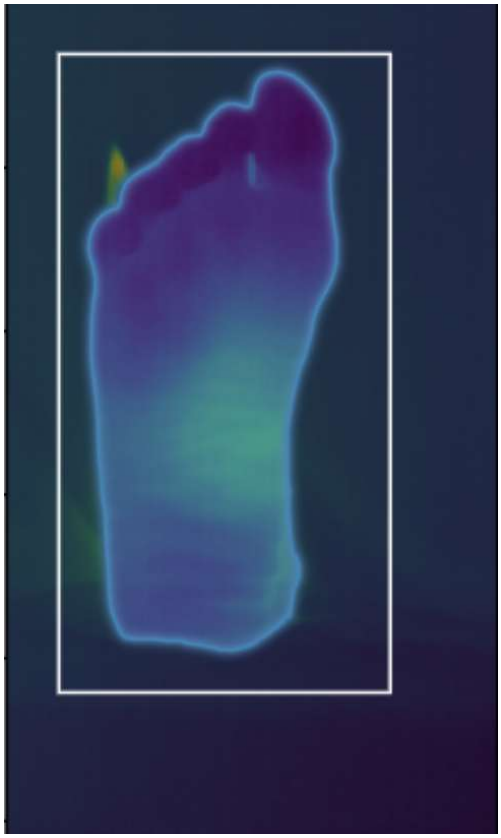
4.4 | Discussion

The Segment-Anything Model (SAM) operates in real-time, needing only a few seconds to generate the input image embeddings. Following this, users can interact with the model using prompts. In this work, box and point prompts were utilised: box prompts to define an approximate bounding box around the foot and point prompts to refine the segmentation mask when needed. As prompts are inputted, SAM outputs the segmentation mask instantly, providing real-time feedback.

User interaction with the model is straightforward and minimal, reducing the need for expert involvement in segmentation tasks. The user defined box prompt does not have to be placed very accurately but a rough bounding-box-like prompt suffices. In a particular case, the bounding box was drawn within the limits of the food boundaries, with the model still being able to accurately segment the foot from the image, as shown in Figure 4.3.

Given the high efficiency of the model, the results are very promising, with all 18

(a) Input image with slight background noise

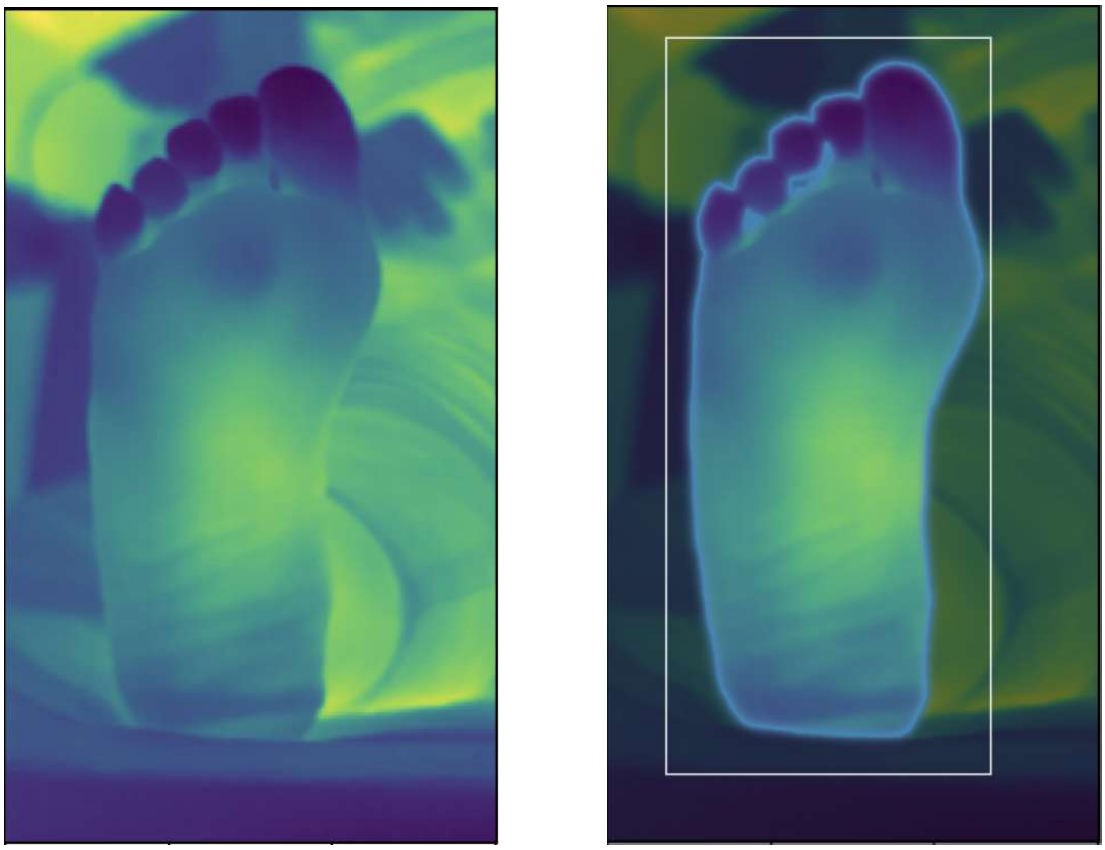


(d) Prompt and accurate segmentation mask

Figure 4.4: The segmentation model is resilient to slight background noise, such as that seen in this figure, with the segmentation mask still being extracted accurately.

frames segmented successfully and to a high standard. The model can also be integrated in the image processing pipeline of any application with relative ease.

The model is also resilient to some background noise as shown in Figure 4.4. As a further experiment an additional thermal sequence was captured without the background, thermal-neutral cloth which is present in the other images and its first frame was fed to the SAM network. The results from this segmentation are as accurate as the previous ones as can be seen in Figure 4.5.



(a) Input image without background cloth and prominent background noise

(d) Prompt and accurate segmentation mask

Figure 4.5: The segmentation model is also resilient to full background noise, such as that seen in this figure, with the segmentation mask still being extracted accurately.

5

Dynamic Medical Thermal Data Decomposition

Skin surface temperature in humans is affected by a multitude of factors, including physical, intake and, environmental factors. When the skin surface temperature is captured using a thermal camera, one also has to consider the technical factors which affect the readings being obtained [41].

Physical factors which play a role in the overall skin surface temperature are the basic characteristics of the subject and are primarily related to biological and anatomical parameters such as: gender, age, height, weight, hair density and metabolic rate among others [41]. Thus, it is necessary that when obtaining skin surface temperature data these parameters are reported for every subject and that when obtaining data from a number of individuals that the subjects are grouped based on some of these parameters.

Intake factors are factors that affect individuals for a limited period of time and are extrinsic to the physical features of the individual. Some of these intake factors are: drug and food intake factors; applications factors of ointments and cosmetics; water intake; exposure to sunlight; therapies and physical activity [41]. It is important that prior to the acquisition of skin surface temperature these factors are limited as much as possible, for example by suggesting that no physical activity is undertaken for 6 hours prior to the acquisition.

Environmental factors which may affect the temperature at the skin surface are related to the place where the reading is being performed. Some example of these type of factors are: room size and its ambient temperature; relative humidity; atmospheric pressure and background radiation [41]. As already described in Chapter 2, Section 2.1.4 considerable effort has been made to standardise the acquisition protocol of medical thermal data to be able to obtain reliable readings in which the impact of environmental factors are minimised. Some technical factors that affect the precision and accuracy of the thermal data are; the validity and reliability of the data; the distance of the subject from the lens; the angle of subject to the thermal camera; resolution of the equipment and the interference of background radiation among oth-

ers [41]. Most of the environmental factors can be catered for in the computation of the temperature values, in degrees Celsius for example, from the intensity of the infrared radiation received by the thermal camera.

Therefore, when analysing temperature values from thermal data, it is essential to account for all relevant factors while recognising that the spatial and temporal thermal patterns observed on the skin are the result of a complex interplay of physiological, environmental, and individual influences [42]. In dynamic thermal data, these considerations remain critical, and the process of temperature extraction is further complicated by the overlapping effects of these factors. Additionally, the impact of a physiological process of interest on skin temperature may be relatively small compared to other influences [84], [143]. Consequently, careful interpretation of thermal video data is necessary to draw accurate and meaningful conclusions.

In this study, we propose the application of principal component analysis (PCA) to decompose dynamic thermal data. A data-driven principal component analysis (PCA)-based method for analysing such recordings is presented. This approach separates the dynamic thermal data into multiple uncorrelated spatio-temporal components, each potentially associated with distinct sources of thermal variation. By isolating these components, it becomes possible to examine their spatial distributions and temporal dynamics individually. PCA thus provides an effective framework for disentangling the effects of independent factors influencing skin temperature. Current literature does not typically consider methods for exploring the contributions of different underlying factors to surface skin temperature. We suggest that analysing the decorrelated components of temperature changes can yield deeper insights into the factors driving spatial and temporal thermal variations. The proposed method demonstrates the ability to differentiate between diverse sources of dynamic thermal variation—including environmental and physiological influences—enabling independent analysis of each source.

First, the proposed PCA method is applied to the original, unregistered, and unsegmented dynamic thermal data to demonstrate the approach and expected outcomes. However, due to the lack of registration and segmentation, it is challenging to extract PCA components for the entire foot. This difficulty arises from subject movement, necessitating the computation of the proposed method on manually selected ROIs.

Next, the method is applied to registered but unsegmented thermal sequences. This step highlights that even the background, covered in this work with a thermally neutral cloth, can provide valuable insights into the observed foot dynamics.

Finally, and most significantly, PCA components are extracted from fully registered and segmented thermal sequences. This demonstrates the results of the complete registration-segmentation-decomposition framework presented in this dissertation.

5.1 | Principal component analysis of dynamic thermal data

Based on observations from a set of variables, PCA performs an orthogonal transformation that converts the original, potentially correlated variables into a new set of uncorrelated variables, known as principal components [119]. These components are arranged so that the first principal component accounts for the greatest portion of variance in the data, with each subsequent component explaining progressively smaller proportions of the remaining variance.

Consider a thermal video sequence of S frames. Each frame, $F_i \in \mathbb{R}^{m \times n}$ where i $\in \{1, \ldots, S\}$, is converted to a column vector $\mathbf{x_i} \in \mathbb{R}^P$, where $P = m \times n$. Then $\mathbf{X}$ can be defined as:

$$\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_S}] \tag{5.1}$$

where the rows of $\mathbf{X} \in \mathbb{R}^{P \times S}$ represent the temporal temperature sequences for P pixels across S frames. In this representation, each row of $\mathbf{X}$ corresponds to the temperature values of a single pixel over time. The temporal sequences of each pixel can be viewed as a set of highly correlated variables, which are then decorrelated through PCA, resulting in the PCA transformation:

$$\mathbf{Z} = \mathbf{W}^\top \mathbf{X} \tag{5.2}$$

where the rows of $\mathbf{Z} \in \mathbb{R}^{P \times S}$ comprise the decorrelated P temporal sequences, represented as principal components, while the columns of $\mathbf{W} \in \mathbb{R}^{P \times P}$ consist of spatial filters that map the original temporal sequences onto these principal components. The coefficients within $\mathbf{W}$ indicate the contribution and influence of each principal component on the corresponding spatial regions. Each column of $\mathbf{W}$ corresponds to a specific component, such that $W = [\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_P}]$, where $\mathbf{w_1}$ contains the weights for component $\mathbf{Z_1}$, and so forth.

5.2 | Dynamic data reconstruction

Since $\mathbf{W}^\top$ is orthogonal [120] then $\mathbf{W}^\top = \mathbf{W}^{-1}$. Therefore, the retained components can be reprojected on the original image space using the transformation:

$$\mathbf{WZ} = \mathbf{X}_R \tag{5.3}$$

where the columns of $\mathbf{Z}$ represent the principal components arranged in order of decreasing variance, the rows of $\mathbf{W}$ contain the associated coefficients, and $\mathbf{X}_R$ denotes

the reprojection of the selected components into the original image space. $\mathbf{X}_R$ is identical to $\mathbf{X}$ only when all components are reprojected. More specifically, the reconstructed data highlights the dynamics of chosen components, allowing unwanted effects such as, environmental influences, to be suppressed and specific dynamic behaviours to be isolated.

5.3 | PCA on original data through ROIs

The PCA process outlined above is first used to extract components from manually extracted ROIs from the original, unregistered and unsegmented dynamic thermal videos. This method of analysis is a novelty in the way that dynamic thermal data is analysed. Until now studies which utilise this kind of data in the medical field have relied on simple average thermal profiles to analyse the data [144], [145]. In this section it is shown that by utilising the unprocessed dynamic thermal video and extracting simple manually demarcated ROIs, PCA can already yield richer information about the temperature dynamics present in the ROIs.

5.3.1 | Methodology

The dynamic thermal videos of the plantar aspect of the human feet which were utilised in the previous chapters are also utilised here. The data consisted of a $320 \times 256 \times T$ matrix, where T is the number of frames collected at one frame per minute, and each element of the matrix corresponds to the temperature value of the pixel. Since, the data used in this section was not registered, any participants whose feet moved significantly during the acquisition could not be utilised for this exercise, as the resulting components would be corrupted by this movement. Thus, a dynamic thermal video in which movement of the foot throughout was minimal was selected for this experiment. As a proof of concept and to provide initial results, seven regions of interest (ROIs) were manually selected on the foot, one on each of the toes, one on the ball of the foot and, one on the heel, as shown in Figure 5.1. These locations were selected as they are prime locations for the presence of ulcers in diabetic patients. Since, the thermal data was not registered, each ROI has to be extracted on every frame to remove the influence of any movement of the foot. This was facilitated by replicating the position of the ROIs on the previous frame to the current frame and then modifying the position of any misplaced ROIs.

The mean temperature of each ROI in each frame was computed and seven lists each of length T were extracted. Finally, the lists were appended in a matrix of dimension $7 \times T$, where each row of the matrix represents the fluctuation in time of the mean temperature of a particular ROI.

The PCA procedure as described in the previous sections was computed on the matrix defined above and the first three components were noted to explain over 90% of

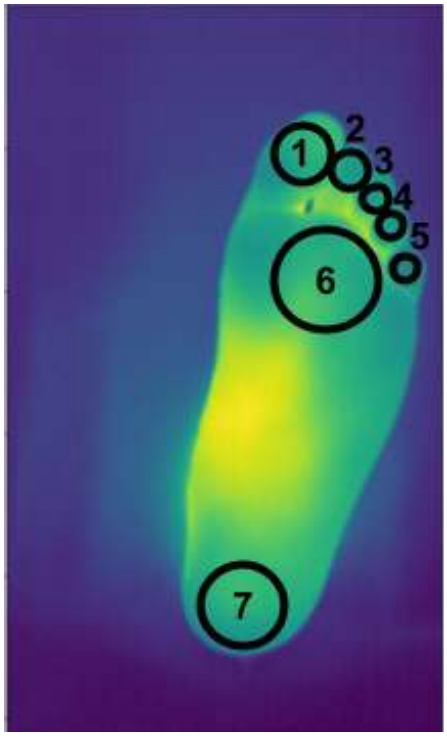


Figure 5.1: The manually selected ROIs used for the scope of showcasing the utility of PCA on dynamic thermal data. The numbers next to the ROIs represent the ROI number which is used later in the results section.

the variance in the original video. Thus these three components are presented and discussed in the next sections.

5.3.2 | Results

The first three principal components of the PCA decomposition on the matrix as defined above, are shown in Figure 5.2. Each figure shows the principal components, ordered by decreasing amount of variance explained (in percentage of the whole video variance) on the left hand side of the figure and, the corresponding coefficients on the right hand side. The components represent the fluctuations in time of the ROIs while the coefficients show how much of an effect each ROI had in obtaining the component.

5.3.3 | Discussion

The first component in Figure 5.2 explains 91.64% of the whole video variance and can be explained as a decay in temperature throughout the video with a partial, and sudden, increase halfway through. This component is influenced mainly by ROIs 1 and 6, the hallux and the ball of the foot, with the other ROIs showing a negative coefficient for this component. This negative coefficient means that those ROIs do not follow well the trend extracted from that particular component but actually have the inverted behaviour, a rise in temperature overall. This is unexpected since prior to the

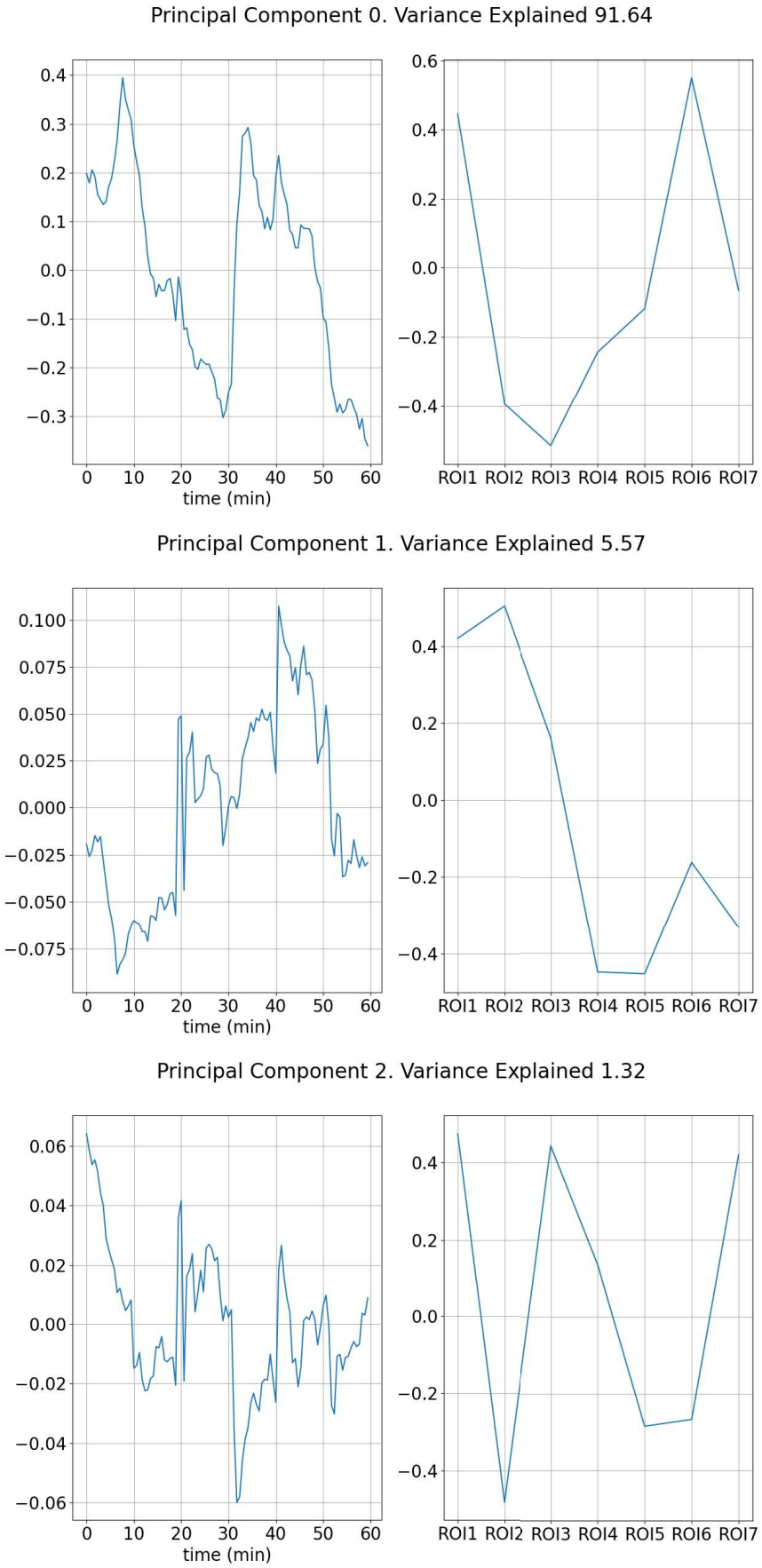


Figure 5.2: First three principal components (left, in order of decreasing variance explained) and corresponding coefficients (right) resulting from proposed PCA through the seven ROIs.

acquisition the foot was covered with shoes and socks and has just been uncovered to the colder room temperature and thus was expected to cool down.

The results in Figure 5.2 also highlight a limitation of applying PCA to ROI data. The coefficients for ROIs with close physical proximity vary widely for the same components. This can be due to the movement of the foot that is not suitably compensated for and, may be leading to artifacts being introduced in the data. This problem further shows the need for the introduction of a registration algorithm prior to the analysis of the data, as was done in this work.

Similarly, the last two components, which have a much lower rate of variance explained (5.57% and 1.32% respectively) seem to have unpredictable patterns both in the components themselves, and thus in time, and on the coefficients, and thus spatial patterns. However, one should also note that overall the components seem to indicate a downward pattern, meaning cooling of the ROIs over time. This cooling effect is however influenced by noise, which is expected for components with such little variance explained as well as for ROIs which spatially occupy a small region on the image.

This unpredictable behaviour can also be due to the manual placement of the ROIs and the fact that the foot moves between one frame and another, since the thermal sequence is not registered. Thus exact placement of the ROI on corresponding anatomical regions frame-by-frame is impossible to achieve.

These results thus showcase that PCA applied to dynamic thermal data can yield informative results, such as the decay pattern observed in the first component, but doing this through manually extracted ROIs without registered video frames is limiting. Thus the registration algorithm utilised in this work will enable the same PCA procedure to be carried out but, on the whole image while compensating for the effect of movement. This procedure is outlined in the next section and results from such a decomposition are presented and discussed.

5.4 | PCA on registered data

5.4.1 | Methodology

As observed in the previous chapters, movements during the recording of dynamic thermal data in human participants is unavoidable due to breathing and other involuntary movement. For various methods, including the PCA approach proposed here, the realignment of images such that each pixel represents the same location throughout the whole video sequence, is fundamental [146]. Fluctuations in temperature caused by the movement of the subject may otherwise confound with the

physiological effects. For the purpose of aligning the dynamic thermal data, the rigid transformation followed by SynthMorph registration method which was presented in Chapter 3, was utilised. Fig. 5.3 shows a sample of frames from a registered dynamic thermal sequence of the plantar aspect of the foot. The temporal changes of the thermal patterns can be readily observed.

5.4.1.1 | Thermal video of control object

Dynamic thermal data of a thermally passive, inanimate object were acquired using a FLIR SC7200 thermal camera, which has a detector resolution of 320×256 pixels and a temperature resolution of 20 mK with an accuracy of ±1 K. Specifically, a thermal video of a leather ball was recorded over one hour. Leather, with an emissivity of 0.95, closely approximates human skin. The aim of this experiment was to test and validate the proposed method on a static object with controlled and predictable temperature variations. Unlike conventional studies, dynamic thermography data were collected immediately from the ball's surface, including the initial phase typically considered an acclimatisation period, in order to monitor the effects of surface equilibration with ambient temperature.

The ambient temperature of the examination room was maintained at a constant 24°C ± 0.5°C for at least one hour prior to and during data acquisition, controlled via an air-conditioning system and monitored with a high-accuracy temperature logger. Frames were captured at a rate of one per minute. Prior to recording, the ball was placed in a warmer environment for one hour and then positioned on a couch in front of the thermal camera. Consequently, the surface of the ball was initially warmer than the ambient temperature, generating the temperature variations captured by the thermal video.

To simulate a localized physiological process during recording, heat was applied to a specific region of the ball's surface midway through the session, producing a temperature differential independent of the normal thermal transients. The outcomes ob-

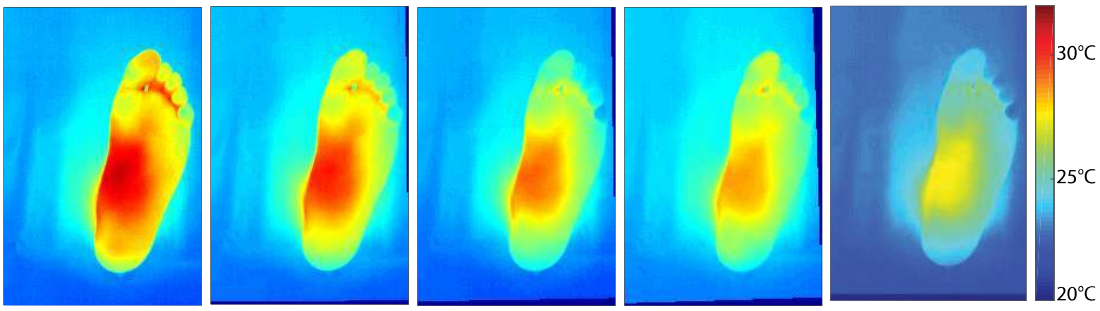


Figure 5.3: Sample of registered frames from thermal image sequence at 15 minutes interval showing evolution of temperature pattern.

tained using the proposed method are presented in the following sections.

5.4.1.2 | Thermal video of human participant

In addition to the video of the thermally passive object, the previously collected thermal videos of the plantar surfaces of human feet were also included in this experiment. The PCA methodology described earlier was applied to both the thermal sequences of the leather ball and those of the human participants. The data were structured as a $320 \times 256 \times T$ matrix, where T represents the number of frames captured at a rate of one frame per minute, and each matrix entry corresponds to the temperature of a specific pixel. Each frame included the object of interest, either the leather ball or a foot of the participant, as well as the surrounding background pixels. By analysing the complete frames, including both object and background regions, it is possible to identify sources of thermal variation originating from the object itself as well as from the surrounding environment.
In the following sections, we demonstrate that the proposed approach can: (1) detect passive cooling associated with acclimatisation and separate it from other temperature variations; (2) distinguish background environmental effects from foreground or skin region processes; (3) identify and differentiate distinct processes within localised regions of interest; and (4) reconstruct the thermal video from the principal components while eliminating background influences. This is illustrated through the application of PCA to both the control thermal video (Section 5.4.2) and the thermal videos of human participants (Section **??**).

5.4.2 | Results – Thermal Video of Control Object

Figure 5.4 presents the first three principal components derived from the PCA of the control thermal video. The left-hand subfigures display the temporal temperature fluctuations of each principal component, while the right-hand subfigures show the corresponding coefficient maps. Each coefficient map includes the percentage of total variance explained by the component. By multiplying a principal component (left plot) with the coefficients of its corresponding map (right image), the contribution of that component to the original image in absolute temperature values can be obtained. Summing the contributions of all components reconstructs the original thermal video.

5.4.2.1 | Identification of Passive Cooling Activity

The first principal component of the control thermal video captures a gradual decrease in temperature. This cooling pattern was expected for the leather ball, which had been placed in a warmer environment outside the examination room before recording began. Using PCA, this passive cooling activity was successfully isolated from other

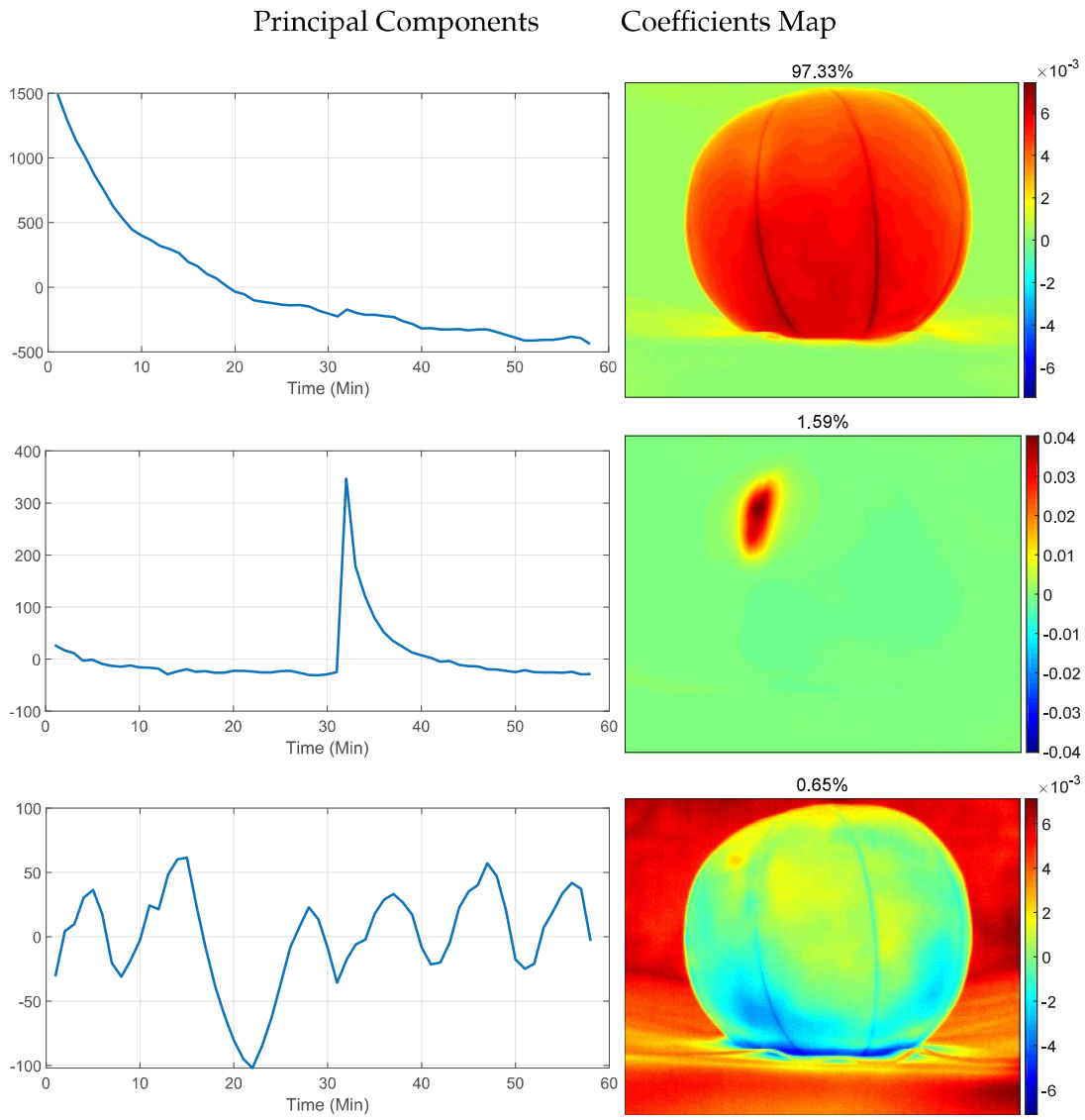


Figure 5.4: First three principal components (left, in order of decreasing variance explained) and corresponding coefficients (right) resulting from proposed PCA on the thermal video data of the static object. The percentage of variance explained by each components is provided above the corresponding coefficients.

temperature fluctuations on the surface of the object. As anticipated, the corresponding coefficients for the first component are highest for pixels in the ball (foreground) region, reflecting that the cooling trend occurs primarily on the ball itself.

5.4.2.2 | Identification of Local Activity

The second principal component (Figure 5.4) highlights activity in local foreground regions. Notably, the temporal signal shows a sharp increase around the 30-minute mark, corresponding to an artificially induced temperature rise on a localized area of the ball, designed to simulate a physiological process.

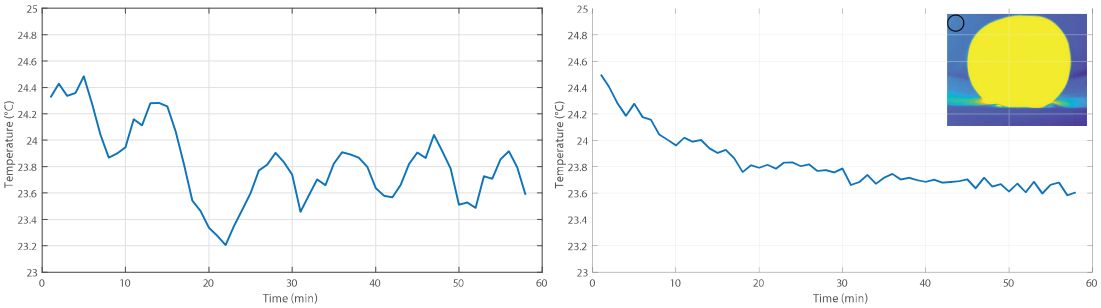


Figure 5.5: Mean temperature of region on background before (left) and after (right) reconstruction for control thermal video. The selected region of interest is shown in black on the image inserted on the plot of the mean temperature after the reconstruction. The selected region is the same for before and after reconstruction cases.

5.4.2.3 | Separation of Background Activity

The third principal component displays an oscillatory pattern predominantly in the background regions, as indicated by larger coefficient values, while the foreground shows minimal correlation with this pattern. This behaviour reflects environmental influences, specifically the air-conditioning in the room.

5.4.2.4 | Reconstruction of the Thermal Video

In this experimental scenario, the background temperature oscillations captured in the third component represent an unwanted artefact and do not convey relevant information. Therefore, it is desirable to exclude this component and reconstruct the thermal video without its influence. As described in Section 5.2, thermal video data can be reconstructed using selected principal components while omitting one or more undesired components. Here, the component associated with background activity is removed, and the video is reconstructed. Figure 5.5 illustrates the mean temperature of a background region across all frames, before and after reconstruction. The results demonstrate that removing this component effectively eliminates the oscillatory behaviour from the thermal video.

Figures 5.6 and 5.7 present the first three principal components derived from the thermal videos of one participant's right and left feet, respectively. The participant was a 24-year-old healthy male with no history of lower limb vascular issues. The left column in each figure displays the principal components, while the right column shows the corresponding coefficient maps along with the percentage of total variance explained.

5.4.2.5 | Identification of Passive Cooling Activity

The first principal component in the participant's thermal video exhibits a gradual decline in temperature throughout the recording, consistent with acclimatisation to

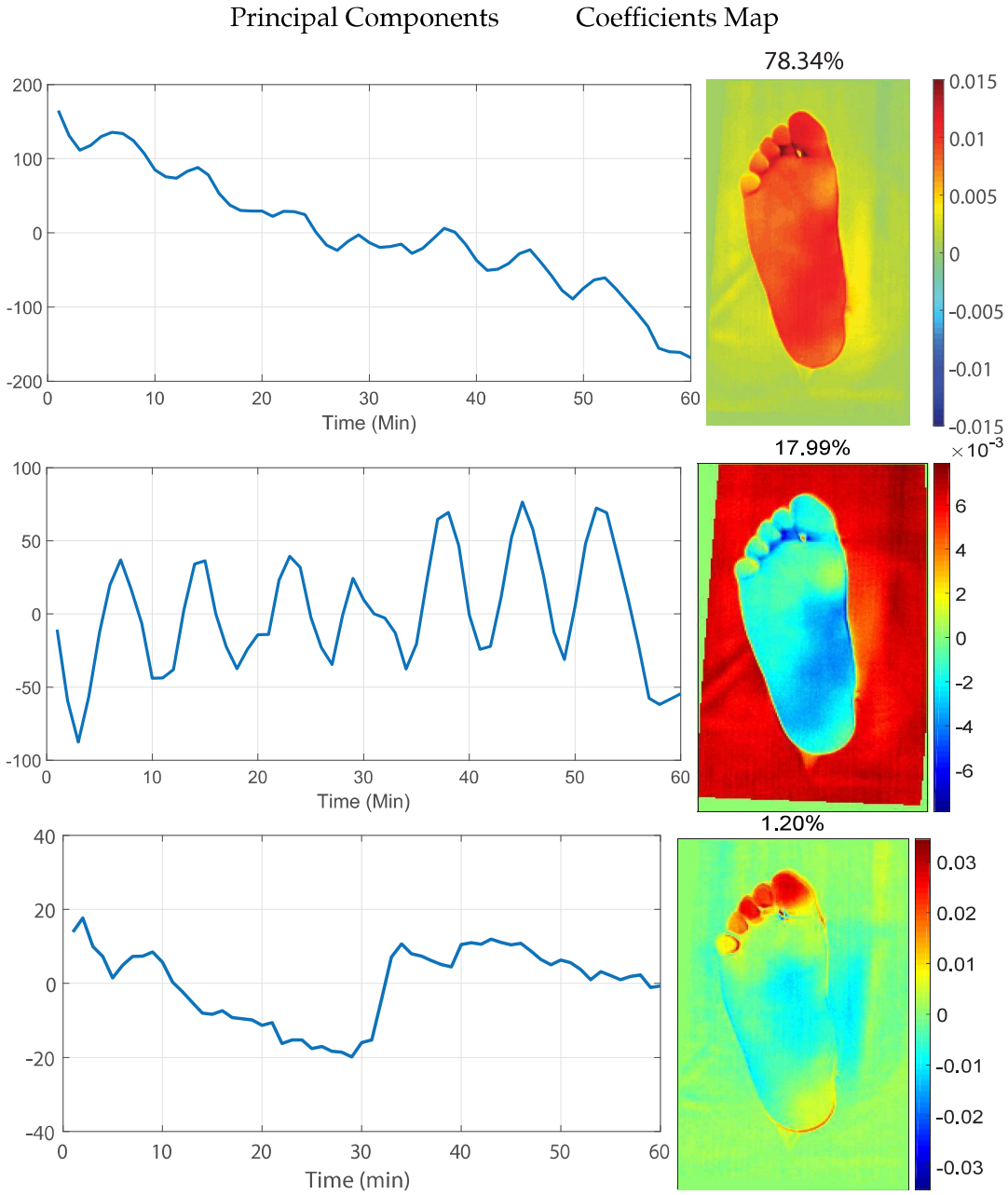


Figure 5.6: First three principal components (left, top to bottom in descending order of variance explained) and corresponding coefficients explained (right) resulting from proposed PCA on the thermal video data of a human participant (Right foot). Each component's percentage variance explained is listed above the corresponding coefficients.

the controlled ambient temperature. The coefficient maps indicate that this cooling effect is concentrated on the foot regions, as pixel values in these areas are higher compared with the background.

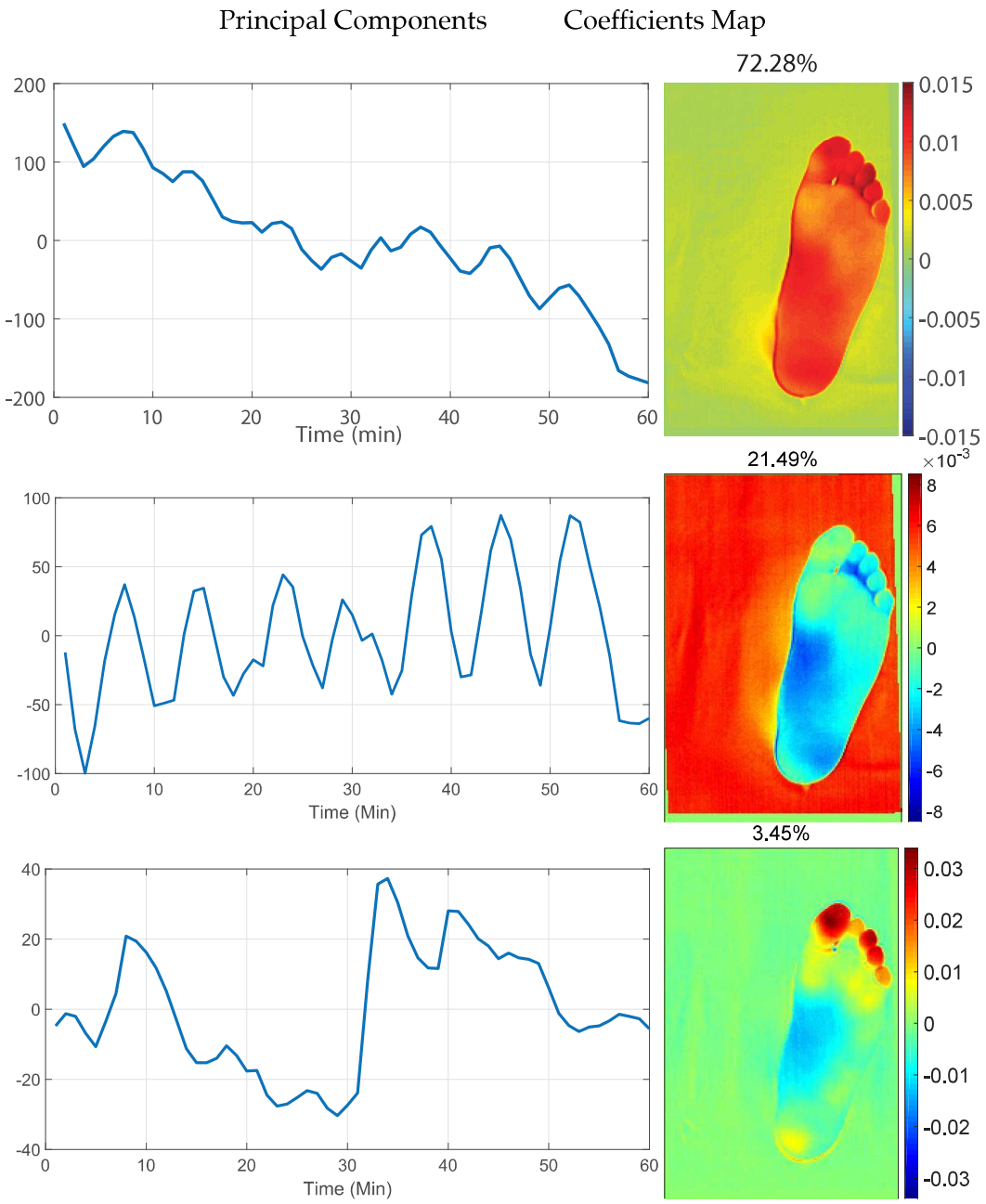


Figure 5.7: First three principal components (left, top to bottom in descending order of variance explained) and corresponding coefficients explained (right) resulting from proposed PCA on the thermal video data of a human participant (Left foot). Each component's percentage variance explained is listed above the corresponding coefficients.

5.4.2.6 | Separation of Background Activity

Similar to the third component observed in the control thermal video, the second principal component in the participant's thermal data displays an oscillatory pattern that is primarily evident in the background regions. The relatively higher coefficients for background pixels compared to those of the foot suggest that this component reflects

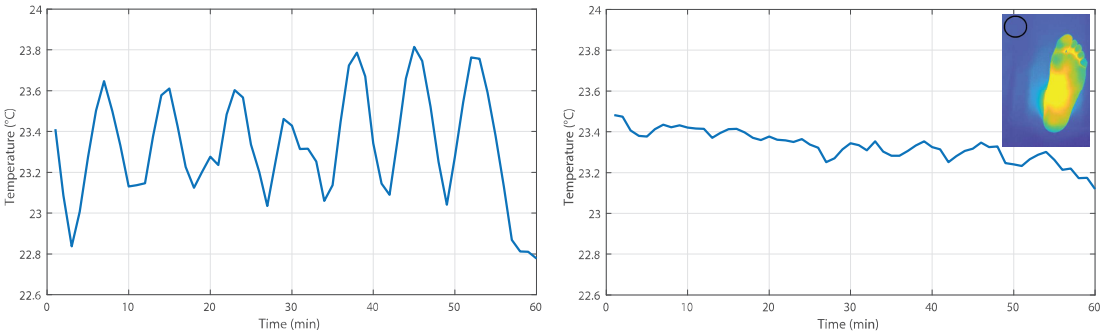


Figure 5.8: Mean temperature of region on background before (top) and after (bottom) reconstruction for the human participant thermal video. The selected region of interest is shown in black on the image inserted on the plot of the mean temperature after the reconstruction. The selected region is the same for before and after reconstruction cases.

environmental temperature fluctuations. Some influence of these atmospheric variations on foot temperature is also apparent, as indicated by fluctuations in the first principal component.

5.4.2.7 | Identification of Local Activity

The third principal component (Figure 5.7) highlights activity within local foreground regions. Notably, a sharp temperature increase occurs around the 30-minute mark. The physiological origin of this change is currently unknown and requires further investigation.

5.4.2.8 | Reconstruction of the Thermal Video

As performed with the control thermal video, the component associated with background activity was removed, and the thermal video was reconstructed. Figure 5.8 shows the mean temperature of a background region across all frames, before and after reconstruction. The results demonstrate that eliminating the background-related component effectively removes the oscillatory behaviour from the thermal video.

5.4.3 | Discussion

In this study, PCA was applied to dynamic thermal data of both a control object and human feet. The analysis and interpretation of the resulting components, as well as the conclusions drawn, are discussed below, following the structure outlined in Section **??**.

5.4.3.1 | Identification of Passive Cooling Activity

The first principal components in Figures 5.4 and 5.7 demonstrate a general decline in temperature over the course of the recording, as reflected in the decaying shape of the components. Coefficient maps indicate that this behaviour primarily occurs in pixels

corresponding to the surface of interest, the ball or the foot, while background pixels have values near zero, indicating minimal or no contribution from these areas. These patterns are consistent with acclimatisation of the object to ambient conditions.

To validate the dynamics identified by PCA, reconstructed videos using the components of interest can be examined. For instance, by reviewing the reconstructed video of the acclimatisation component at a slower playback speed, the temperature decline in the foreground region was observable frame by frame, confirming the component's interpretation.

This acclimatisation component was evident in seven of the nine human participants. Notably, the acclimatisation period extended well beyond the 5–20 minutes commonly recommended in medical thermography literature [147]. Most studies initiate recordings after a brief acclimatisation, often between five and thirty minutes, but PCA in this study successfully isolated the component associated with passive cooling of the skin. Removing this component allows underlying temperature dynamics to be analysed without interference from the cooling process, eliminating the need for prior participant acclimatisation.

5.4.3.2 | Separation of Background Activity

The third principal component in Figure 5.4 and the second components in Figures 5.6 and 5.7 correspond to oscillatory temperature changes in the background. Despite the use of controlled air-conditioning and precise temperature monitoring, small fluctuations in ambient temperature occurred, which can influence surface temperatures.

Interestingly, foot regions displayed similar fluctuations, as seen in the first principal component in Figure 5.6, but with a 90° phase shift relative to the background. This suggests a physiological response of the foot to ambient temperature changes. PCA made this response explicit, a phenomenon that would be difficult to detect with standard thermographic analysis, demonstrating the ability of the method to separate background processes from the data.

5.4.3.3 | Identification of Local Activity

PCA also identified temperature dynamics related to local activity on the surface of interest, as observed in the second principal component of Figure 5.4 and the third principal component of Figure 5.6. For example, a sudden rise in temperature around the toes in Figure 5.7 was examined by visually inspecting both the original and registered thermal data to ensure it was not an artefact. This highlights that PCA can detect small, localized, uncorrelated processes that might otherwise be overlooked.

Static thermography would not reveal such temporal patterns; it might only show a higher temperature at the toes with no information about its dynamics. This underlines the advantage of dynamic thermography for clinical applications and demon-

strates how PCA provides deeper insights into underlying processes than raw data analysis.

Even components accounting for a small fraction of total variance can reflect important local physiological processes. For example, the toe-related activity accounted for only 3.45% and 1.2% of variance in the left and right foot, respectively. While such small percentages might typically be considered insignificant, detailed inspection of coefficient maps can reveal meaningful physiological or environmental dynamics. Hence, careful analysis of principal components alongside their coefficients and the original thermal data is essential.

5.4.3.4 | Other Observations

The PCA results from the thermal videos of human feet (Figures 5.6 and 5.7) show consistent patterns between the right and left feet, especially regarding background activity. This similarity confirms the reliability of the analysis, as the background was shared between both feet.

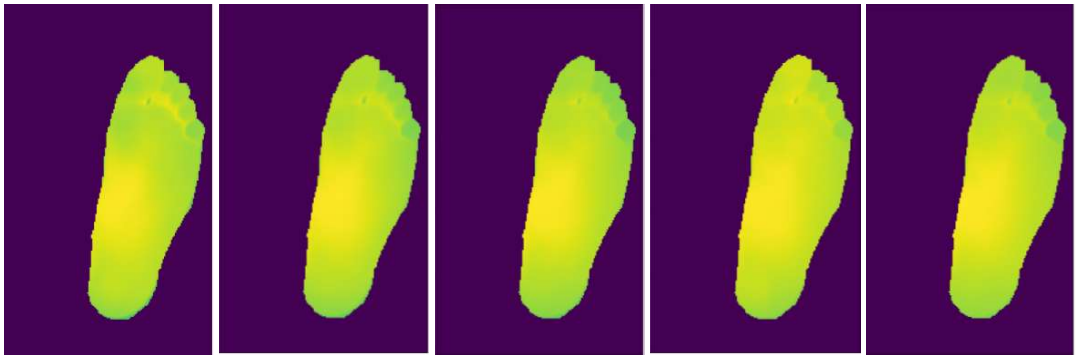
Moreover, this symmetry could have clinical applications: significant differences between corresponding components of the left and right foot might indicate vascular abnormalities. Identifying and analysing such differences in contra-lateral regions could provide a basis for further investigation and diagnostic development.

5.5 | PCA on registered and segmented data

Throughout the previous section we have shown that by utilising the registered, but unsegmented dynamic thermal data, useful results in terms of phenomena on the foot and the background can be extracted. In some cases the background might be retained to assess what factors might be affecting the foot as well as the environment (indicating an external influence), in comparison to something physiological that would only affect the foot. However, for such clean results as those obtained in that section, the background needed to be covered with a thermal neutral cloth, which might not be necessarily possible in many situations. Thus, PCA might need to be carried out on a segmented version of the original dynamic thermal data which contains only the foot or the body part of interest. In this section, we propose the use of PCA on the registered and segmented data obtained for this study.

5.5.1 | Methodology

The registered dynamic thermal videos used in the previous section were segmented by using the segmentation method outlined in Chapter 4 to obtain registered and segmented dynamic thermal videos as shown in Figure 5.9. From this figure, one can observe that the foot is stationary as it occupies the same pixels throughout as well as



t = 0 min t = 15 min t = 30 min t = 45 min t = 60 min

Figure 5.9: Sample of registered and segmented frames from thermal image sequence at 15 minutes interval showing evolution of temperature pattern but no movement of the foot.

that the background in the video has been removed and replaced by zeros. The zeroed background will not add anything to the variance of the video and thus will not affect the PCA computation. The resulting video from the registration-segmentation process has dimensions of $320 \times 256 \times T$ where T is the number of frames captured at 1 frame per minute. As described previously, the input matrix for the PCA algorithm consists of a matrix with dimensions $(320 \times 256) \times T = 81920 \times T$ where effectively each row of the matrix describes the fluctuations in times of one particular pixel.

The PCA procedure as outlined above and as utilised in the previous sections, has been applied to this matrix and once again the first three components explained well above 90% of the total variance and are thus presented in the next section.

5.5.2 | Results

Figures 5.10 and 5.11 show the first three components of the PCA decomposition on the registered and segmented data of a left and corresponding right foot respectively. Once again the left part of the figure shows the components while the right hand side of the figure shows the corresponding coefficient maps. The variance explained by each component is described above. A discussion of these results follows.

5.5.3 | Discussion

From Figure 5.10 and Figure 5.11 the benefits of isolating the foot for the PCA computation can be readily observed. All of the components show activity related to the foot with clear patterns originating in both the components and the coefficients.

The first component, which describes 92.14% of the total variance on the left foot and 96.96% of the total variance on the right foot, as before, describe the acclimatisation process of the foot to the atmospheric temperature. Although the first component

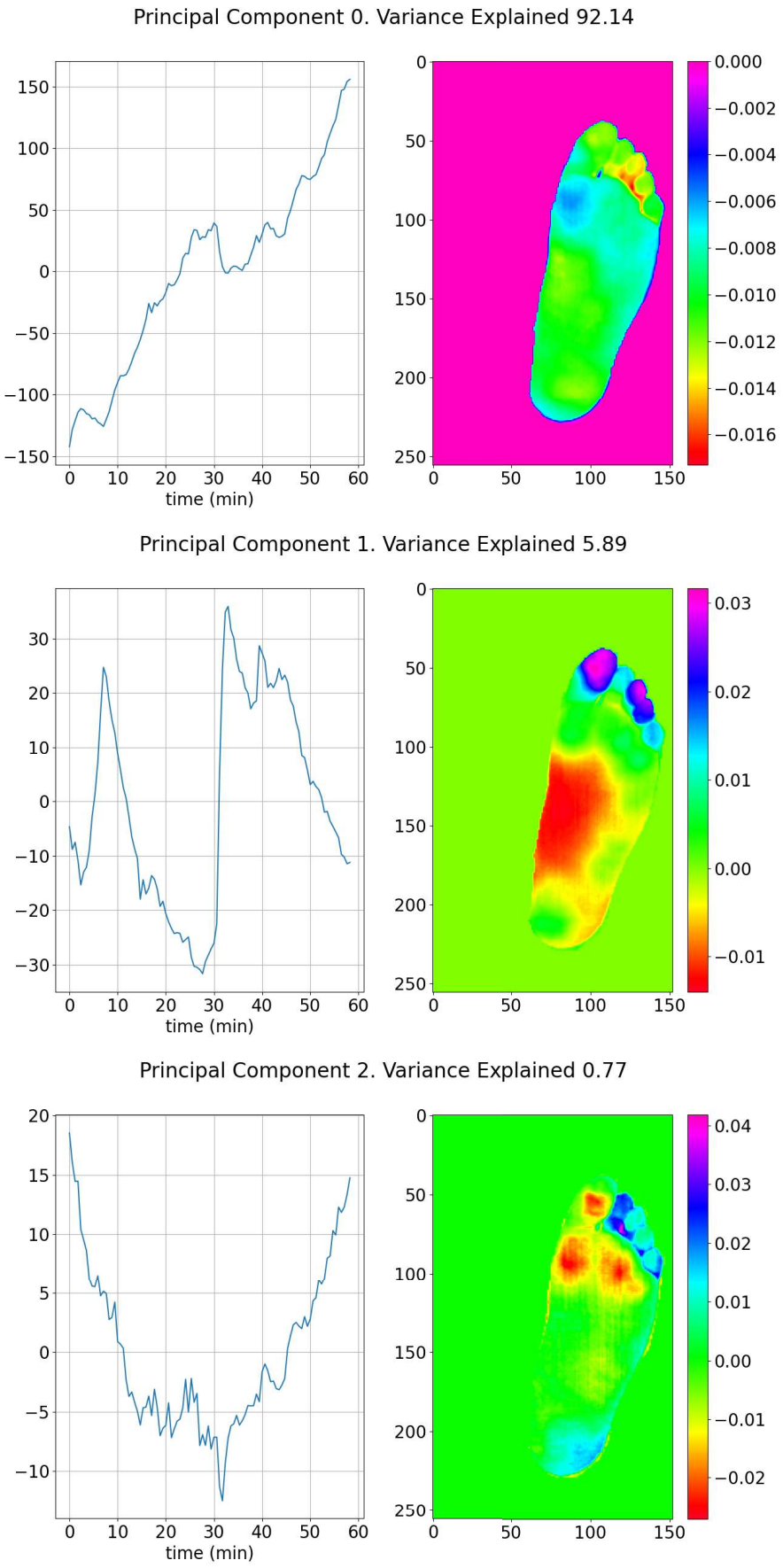


Figure 5.10: First three principal components (left, in order of decreasing variance explained) and corresponding coefficients (right) resulting from proposed PCA on registered and segmented dynamic thermal data of a left foot.

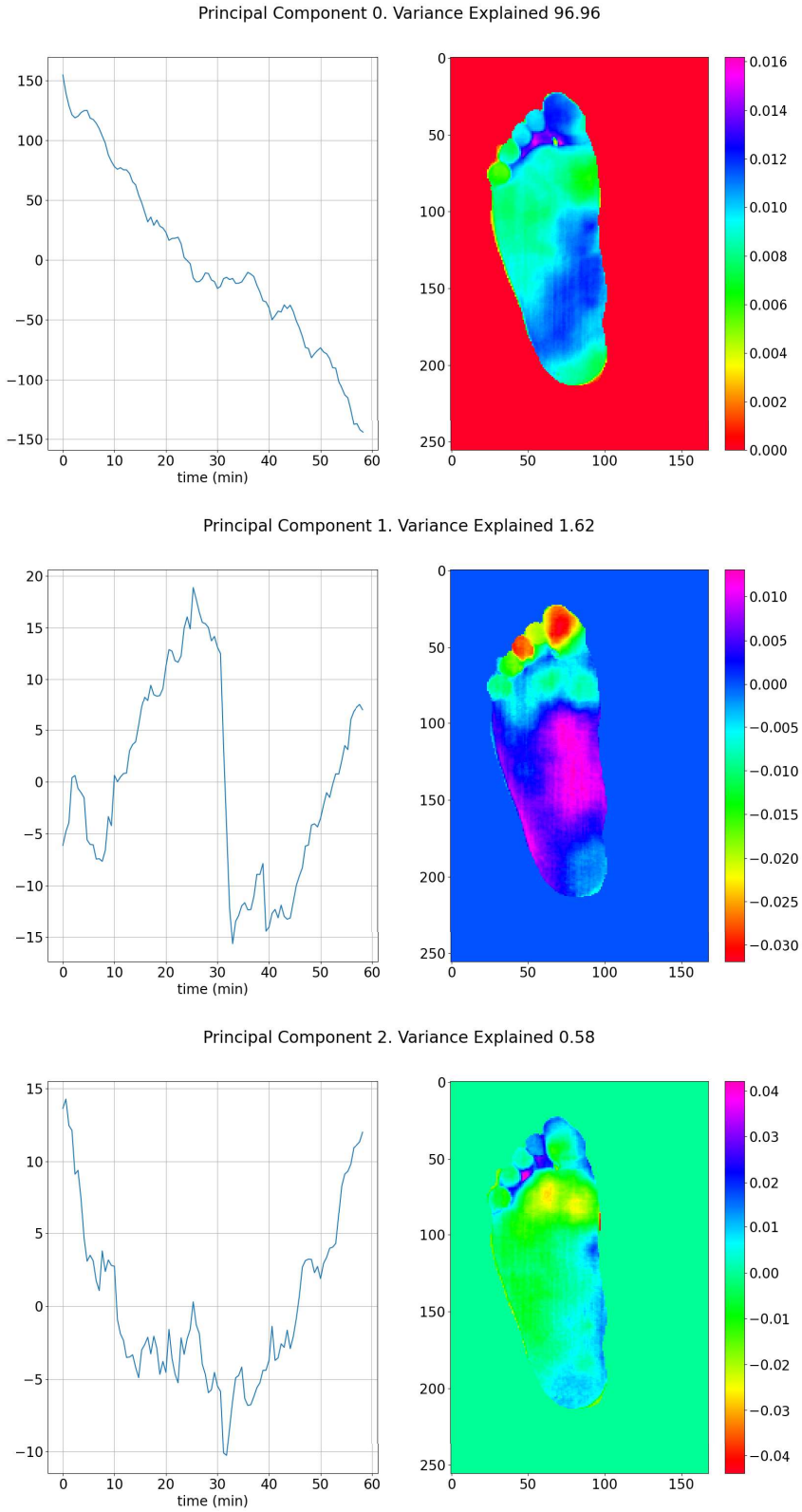


Figure 5.11: First three principal components (left, in order of decreasing variance explained) and corresponding coefficients (right) resulting from proposed PCA on registered and segmented dynamic thermal data of a right foot.

in Figure 5.10 appears to be rising, the coefficients on the foot are all negative and thus the trend in the component has to be inverted for actual interpretation. The decrease in temperature is more obvious in this case as, the undulations which were present in the previous results, which were a contribution of the background, are no longer visible due to their removal using the tools presented in this work. It is also clearer to see the pattern of cooling on the foot then it was previously possible. The left foot shows a more uniform cooling with only a minor region on the left side of the ball of the foot and, the middle toe showing slight deviation from the pattern while, on the right foot the midsection of the foot and the first three toes show a stronger similarity to the pattern observed in the component than the rest of the foot.

These first components on the registered and segmented dynamic thermal video, strengthen the claim that the acclimatisation of the foot goes on for longer than the time claimed in previous literature, with the cooling down of the foot still ongoing, one hour after the commencement of the recording.

The second components in this decomposition match with the third components in the previous section and show a sudden increase in temperature at around minute 30 on the first, third and, fourth toes. The arch of the foot appears to have the opposite behaviour, due to the negative coefficients. This component is once again proof of the utility of utilising PCA for analysing dynamic thermal data, as local activity on the foot could be observed which would have been difficult to notice had the data not been decomposed as the effect of this phenomenon is quite limited on the overall temperature changes as evidenced by the percentage of variance explained, 5.89% for the left foot and 9.62% for then right foot.

An even smaller component, with 0.77% for the left foot and 0.58% for the right foot, is the third component in Figures 5.10 and 5.11. This component shows a decay in temperature until 30 minutes, followed by a rise in temperature until the end of the video. Interestingly, the coefficients show the the areas mostly affected by this component are the regions of the foot which were not affected by the second component.

Temperature in humans is known to be affected by physiological processes which have a cyclic rhythm [148]. The process identified in the third component of this decomposition could be one such process. Further study of this process is still required to confirm this however, this still shows the utility of PCA in extracting such processes.

In the next chapter, the results obtained from the presented framework are discussed together with the implications which such results can have on the study of dynamic thermal imaging, particularly that of the plantar aspect of human feet.

6

Evaluation & Discussion

The introduction of dynamic thermal data as a means to assess foot health presents a significant contribution in the field of foot health assessment using thermal data. This approach diverges from traditional methods, which have primarily focused on static thermal data measurements, and instead provides a more detailed analysis by incorporating changes in temperature over time. These dynamic insights may yield a better understanding of foot health, particularly in the early detection of disorders that impact blood flow, tissue integrity, or nerve function in the lower limbs. This work also aligns with and enhances existing research on thermography for foot health monitoring, especially valuable for populations with compromised foot health, such as those with diabetes or vascular disease.

A registration-segmentation-decomposition framework was presented in this work which has made possible the advanced analysis of dynamic thermal data. Although the application in this project was the plantar aspect of human feet, all of the methods utilised in this framework are adaptable to other body regions, potentially opening up interesting new areas of research. Results obtained from the implemented methods have shown that the expertise required for processing dynamic thermal data have moved from the tedious job of manually aligning the images and extracting the relevant information from regions of interest, to the job of analysing the data. This is a step forward for dynamic thermal data in the medical field and it enables more complex analysis and thus better informed decisions to be achieved.

Figure 6.1 shows a complete result for one dynamic thermal sequence together with the sequence of events in the framework. The top most row shows the original dynamic thermal sequence with movement of the foot apparent. This sequence is fed into the linear registration first and then the SynthMorph network for elastic deformation alignment resulting in the second row of images. In this row it can be seen that the misalignment of the foot has been handled satisfactorily even if the original movement of the foot had resulted in a background artifact. Next, the registered ther-

mal sequence is fed into the Segment Anything Model and the third row of images is obtained with only the foot in the image and the background segmented out. PCA is finally applied to the sequence obtaining results like those seen in the last row of the figure, showing the first component of this decomposition. These results are discussed in detail below.

6.1 | Dynamic Medical Thermal Data Registration

The registration of dynamic thermal data is challenging due to the varying contrast throughout the sequence between the body region of interest in the foreground and the rest of the background. The similarity in intensity between the two regions makes for additional complexity. In some cases it is difficult even for human eyes to discern the difference between the foreground and the background and to detect the edge of the body region of interest. Prior to the application of this method, no other method has been able to register dynamic thermal sequences with comparable accuracy and without any user input as done in this work.

The registration algorithm, made up of a linear transformation and a SynthMorph trained VoxelMorph network for elastic deformations, provided good results with a success rate of 88.8%. The experiments carried out show that the utilised methods can be successfully used on dynamic thermal data of the plantar aspect of the feet.

The registration method was tested only on 18 thermal sequences all of which has the background covered with a thermal neutral cloth. Thus, this registration method needs to be tested further on more sequences and with a wider range of scenarios being represented, such as more background noise, to confirm the results obtained.

One advantage of the implemented method is that the registration process is fully automated, only requiring user input when the registration algorithm fails and, thus, doing away with the need of experts during the processing of the thermal data sequence. It should also be noted, that in the case of the failed registrations, around 80% of the thermal sequence was still registered correctly, while the remaining frames only had minor deformations on the toes.

The method has been observed to be resilient to some background artifacts, such as the movement of the background cloth uncovering a part of the actual background, as well as large movements of the foot which lead to both linear and elastic deformations.

6.2 | Dynamic Medical Thermal Data Segmentation

The segmentation algorithm utilised for this work, Segment-Anything Model (SAM), on the other hand, has been shown to provide an accurate segmentation in 100% of the

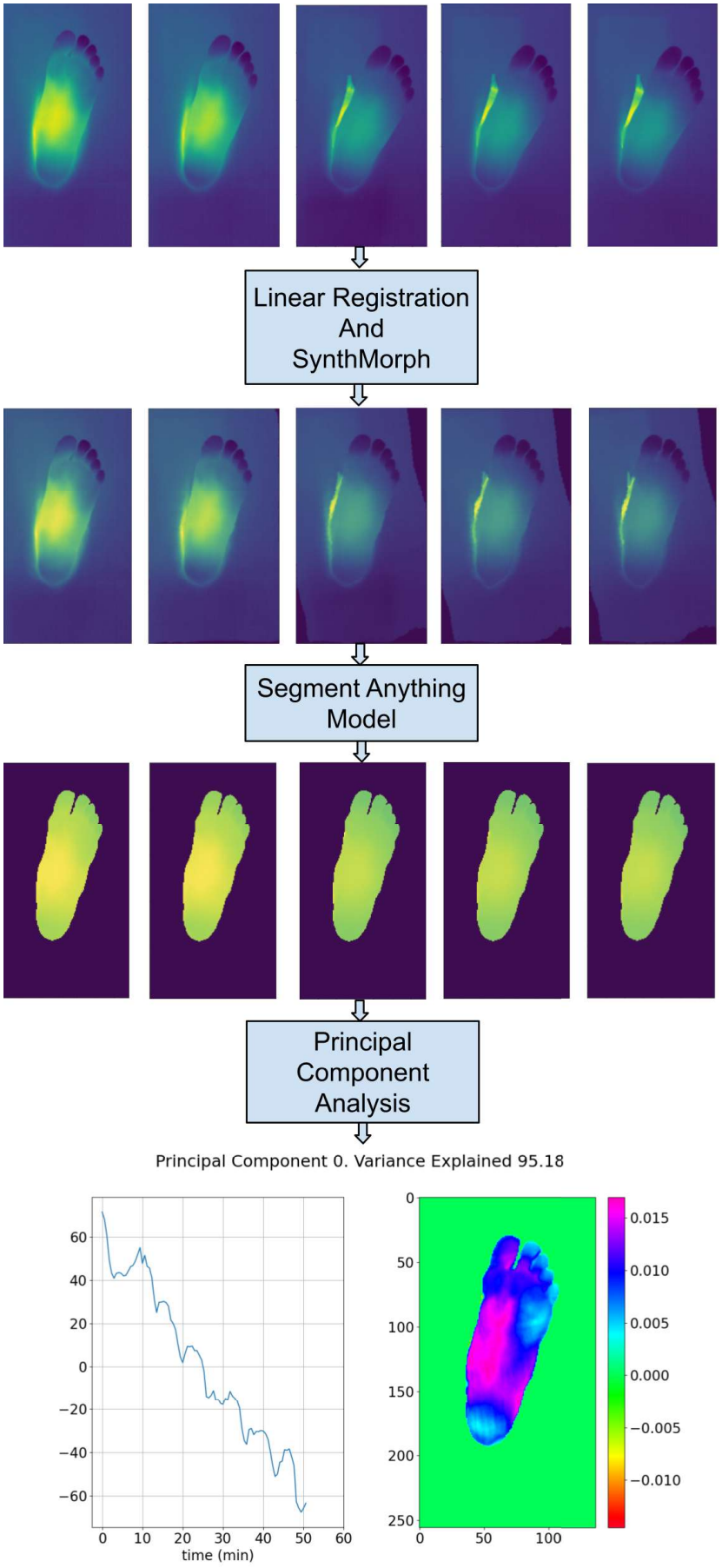


Figure 6.1: The result from the registration, segmentation and decomposition framework on one particular dynamic thermal data sequence.

test images. This was also achieved with minimal user input and in very little time per segmentation. The user input was limited to a bounding-box-like box prompt and point prompts when regions of the segmentation mask needed to be added or subtracted from the first resulting mask. Once again this means that an expert will not be necessary for processing dynamic thermal data in a medical aspect and could be used for the analysis of such data.

The segmentation algorithm only required around 10 seconds per image to extract the image embedding. From then on the segmentation mask could be extracted and modified in real time. This means that in the full pipeline of processing dynamic thermal data, the segmentation would take very little processing time. Since the dynamic thermal sequences were registered prior in the pipeline, the segmentation algorithm effectively needs to segment one image per sequence, with the mask being generalisable for the whole sequence.

Prior to utilising SAM for the segmentation part of the processing pipeline, a background-foreground segmentation and ROI segmentation algorithm was being purposely built for the scope of this work. However, with the release of SAM, that was made freely available and that could be used on medical images with zero-shot generalisation, the implementation of a new segmentation method became ineffective and unproductive. This is due to the strengths of SAM, namely its zero-shot generalisation, its real-time segmentation, its ease of use and, its overall accuracy and, due to the fact that the method being implemented was not providing as good results as SAM. Segmentation of thermal images is complicated by the fact that thermal images do not provide very clear features due to the overlapping intensity between the foreground and the background. In some instances it is difficult even for the human eye to assess where the boundary between the two regions lies.

The segmentation method was also utilised in a case study in which the background of the thermal sequence was not covered by a thermal neutral cloth and thus the background contained a lot of artifacts. As shown in Chapter 4 this thermal sequence was also segmented with high accuracy and thus suggesting that the method is resilient to such artifacts.

6.3 | Dynamic Medical Thermal Data Decomposition

Dynamic thermal data is made up of a number of physiological and environmental processes whose effects add up tom produce the temperature dynamics observed on the foot. In this work, it is shown that PCA can be utilised to decompose the effect from these processes and that each process can be studied individually. The findings presented in this dissertation demonstrate that dynamic thermal data offers unique insights that static measures cannot capture.

Two-way PCA is primarily suited for analysing two-dimensional datasets, such as matrices where rows represent samples and columns represent variables [149]. In PCA, the data is decomposed into orthogonal principal components that maximise variance and reduce dimensionality, simplifying complex datasets by capturing the most significant patterns in a smaller number of components. In this work, the multidimensional array resulting from dynamic thermal imaging, was unfolded in a two-way matrix for use in PCA.

In the case studies presented in Chapter 5, both of an inanimate object, as well as the human foot, it can be observed that PCA works well in decomposing the dynamic thermal data into constituent processes such as the acclimatisation component and local activity components.

6.3.1 | Acclimatisation component findings

One of the component which were identified was that of the acclimatisation of the foot to the room temperature. The main finding from this component is the discovery that the acclimatisation period for the plantar aspect of the foot, is much longer than what is suggested in the current literature. In that literature, acclimatisation for the foot region ranges from no acclimatisation to 20 minutes. With the use of principal component analysis (PCA), the component related to the acclimatisation of the foot to room temperature has been extracted and identified. This acclimatisation is shown to last for the full duration of the thermal video, 60 minutes, and possibly even longer.

This finding is important both for dynamic thermal imaging as well as static thermal imaging. Traditionally the static thermal image is captured after the period of acclimatisation is over but, this finding suggests that a rethinking of this concept is needed as the acclimatisation is usually still ongoing when the images are captured. This means that both static and dynamic data are still being influenced by the acclimatisation and potentially each individual is at a different stage of the acclimatisation process when the data is captured. One solution to this can be dynamic thermal imaging and the application of PCA decomposition to identify, extract, and subtract the acclimatisation component from the rest of the data, as was done in this work. This would reduce the need to keep the participant waiting for a period of time prior to acquiring the data and would rid the data from unknown variables such as the acclimatisation rate.

In Chapter 5 three scenarios where PCA could be utilised were presented. These were; prior to the registration-segmentation framework; on registered but unsegmented data and; on registered and segmented data.

The PCA decomposition on unregistered and unsegmented data could only be carried out through ROIs which were manually selected on each individual frame. This means that an expert in analysing such data would need to tediously select all of the

required ROIs on each frame to be able to extract any meaningful interpretation from the data. Manually selecting ROIs and extracting thermal profiles of the same ROIs is already an established procedure in dynamic thermal data analysis but this method is not without its problems. Firstly, the placement of the ROIs although guided by physical markers, such as the edge of the toes, will always be subjective and thus repeatable results are unlikely. Secondly, the foot is not static throughout the thermal sequence and thus each ROI has to be modified on each frame to counter for the movement of the foot. This means that the problem with the subjectivity of the placement of the ROIs is further compounded. It also means that the subject has to try an be very still during the acquisition or that when analysing the data, sequences that contain a lot of movement have to be omitted from the study.

The results from this scenario, were still useful as the acclimatisation component was still captured. However, the coefficients were difficult to interpret and there was not a clear structure to the patterns extracted from the coefficients.

On the other hand, the PCA decomposition on the registered but unsegmented data, was a step forward in terms of interpreting the results obtained. This decomposition led to the discovery of both a background process which oscillates the room temperature, as well as the discovery that the acclimatisation in the foot goes on for the whole period of the video acquisition. Furthermore, a local activity component could also be identified as well as the physical region from which it originated.

The same findings were then confirmed, and in some cases enhanced, in the third case where, PCA was implemented on registered and segmented thermal sequences. In this decomposition the effect of background processes was missing and thus more local activity which was previously missed has been uncovered.

It is worth noting at this point, that should an analysis of a particular body region of interest be required, such as for example the heel, the registered and segmented data would be ideal as the desired ROI can be selected on the first frame of the sequence and applied on the rest of the frames. PCA components which are free from background activity could then be extracted from this data.

Although with the registered and segmented data, the background processes were removed and the components were made up only of processes originating from the foot, the effect of the oscillating room temperature on the foot itself was not removed. A solution for this is suggested as part of any future work on this project.

Initially a PCA decomposition can be carried out on the registered but unsegmented thermal sequence. Any oscillatory components originating from the background or foreground can be removed and the data be reconstructed without these components. This reconstruction process is shown in Chapter 5. Next, another PCA

decomposition can be carried out on the reconstructed data without the oscillatory behaviour.

6.3.2 | Thermal camera specifications

The thermal camera which was used for this work, Flir SC7200, has a high spatial and temperature resolution. The downside to this is that the thermal camera is large and heavy and has limited portability. During acquisition the camera is restricted to a tripod and cannot be moved around like a handheld camera could. This limits the practicality of the system and, therefore, for a clinical tool as described above, small, portable thermal cameras should be used for better portability. However, small thermal imaging systems do not yet provide the temperature sensitivity of the high end thermal cameras. The camera used for this work has a thermal sensitivity which is less than 25 mK and a temperature measurement accuracy of ±1°C. These specifications have not yet been matched by handheld systems. If an implementation of these algorithms is to be used in a clinical tool, the effect which this drop in temperature sensitivity has on the performance of the algorithms, especially the decomposition work, has to be studied and catered for in the implementation of the algorithms.

6.4 | Future work - Towards an Integrated Pipeline

The long term aim of this work is to enable the creation of a self-contained device which utilises dynamic thermal data to image the feet of healthy and diabetic patients. The registration-segmentation-decomposition framework could then be integrated in the device for the processing of the collected thermal data. The device would then be able to decide whether the patient requires any immediate care or further tests.

However, to reach this goal some more work is required. Firstly the technical algorithms require further testing and confirmation of results on a larger and more diverse dataset. Also, it is required that the decomposition techniques be run on more data to collect more knowledge about the temperature dynamics of the plantar aspect of human feet. Moreover, these method need to be also shown to work on diabetic patients feet and to assess whether there are differences between the healthy and diabetics patients temperature dynamics. Such difference could be utilised for the final decision on whether the patient requires further care.

One should note that, the analysis technique utilised in this work, PCA, is a standard method of decomposition, and more advanced and modern techniques exist. The aim of this work was to introduce the framework for processing dynamic thermal data and to show that the registration-segmentation pipeline enables more advanced processing techniques than currently possible. However, the framework also enables

the user to select any analysis technique which is deemed as most suitable and which benefits from registered and segmented data.

The decomposition work together with the findings regarding the acclimatisation of the plantar aspect of the foot presented in this chapter have been published in Infrared Physics and Technology [23]

7

Conclusions

Foot health is a critical aspect of overall well-being, especially for individuals with diabetes. Diabetes often leads to complications in the lower limbs, largely due to nerve damage (neuropathy) and reduced blood circulation. These issues increase the risk of developing infections, foot ulcers, and in severe cases, can even result in amputations. Monitoring foot temperature is an effective, non-invasive method to detect early signs of such complications.

Temperature changes in the feet can serve as valuable indicators. A decrease in foot temperature often suggests poor blood flow, potentially pointing to circulatory problems like peripheral arterial disease (PAD). Conversely, an increase in foot temperature may signal inflammation or ulceration, both of which can quickly progress if left untreated.

In a broader sense, body temperature is an essential marker of health and well-being. Fluctuations in temperature can reveal underlying infections, inflammatory responses, or issues with blood flow, each of which has implications for a person's health. By tracking both general and foot-specific temperature changes, individuals, particularly those with diabetes, can catch potential issues early, allowing for timely medical intervention and reducing the risk of serious complications. Regular monitoring, combined with proper foot care and lifestyle management, can greatly enhance quality of life and overall health.

Thermography is a leading method for assessing body temperature in human subjects and plays an important role in medical diagnostics and health monitoring. This imaging technique operates by capturing infrared radiation naturally emitted from the body, converting it into a visual map that represents temperature variations across the surface of the skin. By analysing these variations, thermography can identify areas of abnormal heat distribution that may indicate underlying issues such as inflammation, infection, circulatory problems, or nerve-related complications.

One of the greatest advantages of thermography is its non-invasive and non-contact nature. Unlike traditional methods, such as using a thermometer, thermography does

not require physical contact with the patient, making it a comfortable option. The process is also swift, allowing clinicians to gather the necessary information in seconds and promptly identify areas that may require further examination or intervention.

The primary objective of this research was to explore the effectiveness of thermal imaging in monitoring foot health, particularly for early detection of potential complications of diabetes. The study demonstrated that dynamic thermal imaging, a process that captures temperature changes over time, can be successfully utilised in this context, offering valuable insights into foot temperature dynamics that were previously underexplored. This new knowledge can in the future be used in the management of foot health in diabetics although, in this project this area remained unexplored.

To extract this new knowledge, it was essential to process the dynamic thermal data in two key ways. First, an image registration technique was employed to ensure that the foot remained aligned throughout the video, enabling further processing without the need to track salient points on the foot. Secondly, a segmentation process was applied to isolate the foot from the background, allowing the subsequent analysis to focus exclusively on the region of interest.

The registration algorithm, which was presented in Chapter 3, is a combination of a linear registration together with a SynthMorph trained VoxelMorph network for countering elastic deformations. Using this method, a successful registration rate of 88.8% was achieved with no user input.

The segmentation method presented in Chapter 4 was the Segment-Anything Model (SAM), with which all of the dynamic thermal videos could be segmented with high accuracy without the need of any expert or manual user input.

In Chapter 5, it was shown that the registration-segmentation framework enables advanced analysis of the thermal data to be performed on the whole foot. principal component analysis (PCA) was utilised to decompose the temperature dynamics into its constituent factors and it was shown that this analysis can provide richer information after the thermal video is registered and segmented than was the case before.

In Chapter 6, a discussion of the findings was presented and potential future build-up work for the project was suggested.

The decomposition work together with the findings regarding the acclimatisation of the plantar aspect of the foot have been published in Infrared Physics and Technology [23], while the work and results on the registration algorithm have been accepted for publication at the Quantitative InfraRed Thermography Journal.

In conclusion, this research has successfully demonstrated the potential of dynamic thermal imaging as a powerful tool for monitoring foot health, with particular promise

for applications in preventive care. By employing advanced image registration and segmentation techniques, we were able to process and analyse thermal data with high precision, revealing critical insights into the temperature dynamics of the foot. These insights can provide a foundation for developing early intervention strategies, especially valuable for individuals at risk of diabetic foot complications.

The findings underscore the value of continuous monitoring in healthcare, as dynamic thermal imaging can enable real-time assessment and ongoing risk evaluation. Future research could extend these findings by exploring fully integrated and automated systems.

Ultimately, this thesis affirms that dynamic thermal imaging holds great promise as a reliable, efficient, and accessible method for foot health assessment. By facilitating early detection and informed decision-making, it has the potential to improve patient outcomes, reduce healthcare costs, and enhance quality of life for at-risk populations.

References

[1] O. Rabe, M. Winther-Jensen, K. Allin, and O. Svendsen, "Fractures and osteoporosis in patients with diabetes with charcot foot," *Diabetes Care*, vol. 44, pp. 2033 – 2038, 2021.

[2] E. Grodzinsky and M. Levander, "Assessment and evaluation of body temperature," *Understanding Fever and Body Temperature*, 2019.

[3] M. Sund-Levander and E. Grodzinsky, "Time for a change to assess and evaluate body temperature in clinical practice.," *International journal of nursing practice*, vol. 15 4, pp. 241–9, 2009.

[4] D. Hanson, "Fever, temperature, and the immune response," *Annals of the New York Academy of Sciences*, vol. 813, 1997.

[5] A. Garami, A. Steiner, and A. A. Romanovsky, "Fever and hypothermia in systemic inflammation.," *Handbook of clinical neurology*, vol. 157, pp. 565–597, 2018.

[6] V. B. Poveda, R. A. Oliveira, and C. M. Galvão, "Perioperative body temperature maintenance and occurrence of surgical site infection: A systematic review with meta-analysis.," *American journal of infection control*, 2020.

[7] M. Cuddy, "The effects of drugs on thermoregulation.," *AACN clinical issues*, vol. 15 2, pp. 238–53, 2004.

[8] A. Drewry, B. Fuller, T. Bailey, and R. Hotchkiss, "Body temperature patterns as a predictor of hospital-acquired sepsis in afebrile adult intensive care unit patients: a case-control study," *Critical Care*, vol. 17, pp. R200 – R200, 2013.

[9] D. Levine, J. R. Walker, D. Marcellin-Little, R. Goulet, and H. Ru, "Detection of skin temperature differences using palpation by manual physical therapists and lay individuals," *Journal of Manual & Manipulative Therapy*, vol. 26, pp. 101 – 97, 2018.

[10] D. Niven, J. Gaudet, K. Laupland, K. Mrklas, D. Roberts, and H. Stelfox, "Accuracy of peripheral thermometers for estimating temperature," *Annals of Internal Medicine*, vol. 163, pp. 768–777, 2015.

[11] E. Barreira, R. Almeida, and M. L. Simões, "Emissivity of building materials for infrared measurements," *Sensors (Basel, Switzerland)*, vol. 21, 2021.

[12] A. Sharma, S. Sharma, P. Vasistha, and J. P. Mangalhara, "Estimation of effect of emissivity on target detection through thermal imaging systems," *Defence Science Journal*, vol. 67, pp. 177–182, 2017.

[13] A. Hakim and R. Awale, "Thermal imaging-an emerging modality for breast cancer detection: a comprehensive review," *Journal of Medical Systems*, vol. 44, no. 8, pp. 1–18, 2020.

[14] Y. Tan, C. Hong, H. Li, J. Allen, and J. Thumboo, "A novel use of combined thermal and ultrasound imaging in detecting joint inflammation in rheumatoid arthritis.," *European journal of radiology*, vol. 134, p. 109421, 2020.

[15] J. J. van Netten, J. G. van Baal, C. Liu, F. van Der Heijden, and S. A. Bus, "Infrared thermal imaging for automated detection of diabetic foot complications," *Journal of Diabetes Science and Technology*, vol. 7, no. 5, pp. 1122–1129, 2013.

[16] N. Aggarwal, M. Garg, V. Dwarakanathan, N. Gautam, S. S. Kumar, R. Jadon, M. Gupta, and A. Ray, "Diagnostic accuracy of non-contact infrared thermometers and thermal scanners: A systematic review and meta-analysis," *Journal of Travel Medicine*, 2020.

[17] S. E. Godoy, M. Hayat, D. A. Ramirez, S. A. Myers, R. S. Padilla, and S. Krishna, "Detection theory for accurate and non-invasive skin cancer diagnosis using dynamic thermal imaging.," *Biomedical optics express*, vol. 8 4, pp. 2301–2323, 2017.

[18] C. Song, J. Rhee, and S. Levitt, "Blood flow in normal tissues and tumors during hyperthermia.," *Journal of the National Cancer Institute*, vol. 64 1, pp. 119–24, 1980.

[19] E. Suzuki, A. Kashiwagi, Y. Nishio, K. Egawa, S. Shimizu, H. Maegawa, M. Haneda, H. Yasuda, S. Morikawa, T. Inubushi, and R. Kikkawa, "Increased arterial wall stiffness limits flow volume in the lower extremities in type 2 diabetic patients.," *Diabetes care*, vol. 24 12, pp. 2107–14, 2001.

[20] J. Ruminski, M. Kaczmarek, A. Renkielska, and A. Nowakowski, "Thermal parametric imaging in the evaluation of skin burn depth," *IEEE Transactions on Biomedical Engineering*, vol. 54, pp. 303–312, 2007.

[21] B. Elem, B. A. Shorey, and K. L. Williams, "Comparison between thermography and fluorescein test in the detection of incompetent perforating veins," *British Medical Journal*, vol. 4, no. 5788, pp. 651–652, 1971.

[22] M. Kaczmarek, "A new diagnostic ir-thermal imaging method for evaluation of cardiosurgery procedures," *Biocybernetics and Biomedical Engineering*, vol. 36, no. 2, pp. 344–354, 2016.

[23] J. Gauci, K. P. Camilleri, and O. Falzon, "Principal component analysis for dynamic thermal video analysis," *Infrared Physics & Technology*, vol. 109, p. 103359, 2020.

[24] B. B. Lahiri, S. Bagavathiappan, T. Jayakumar, and J. Philip, "Medical applications of infrared thermography: a review," *Infrared Physics & Technology*, vol. 55, no. 4, pp. 221–235, 2012.

[25] N. Jasti, S. Bista, H. Bhargav, S. Sinha, S. Gupta, S. Chaturvedi, and B. Gangadhar, "Medical applications of infrared thermography: A narrative review.," *Journal of Stem Cells*, vol. 14, no. 1, 2019.

[26] E. Ring and K. Ammer, "The technique of infrared imaging in medicine," *Thermology international*, vol. 10, no. 1, pp. 7–14, 2000.

[27] S. Westermann, H. H. Buchner, J. P. Schramel, A. Tichy, and C. Stanek, "Effects of infrared camera angle and distance on measurement and reproducibility of thermographically determined temperatures of the distolateral aspects of the forelimbs in horses," *Journal of the American Veterinary Medical Association*, vol. 242, no. 3, pp. 388–395, 2013.

[28] F. Ring, "Thermal imaging today and its relevance to diabetes," *Journal of Diabetes Science and Technology*, vol. 4, no. 4, pp. 857–862, 2010.

[29] D. Rusch, M. Follmann, B. Boss, and G. Neeck, "Dynamic thermography of the knee joints in rheumatoid arthritis (RA) in the course of the first therapy of the patient with methylprednisolone," *Zeitschrift fur Rheumatologie*, vol. 59, no. SUPPL. 2, pp. 131–135, 2000.

[30] C. Purslow, J. S. Wolffsohn, and J. Santodomingo-Rubido, "The effect of contact lens wear on dynamic ocular surface temperature," *Contact Lens and Anterior Eye*, vol. 28, no. 1, pp. 29–36, 2005.

[31] M. Antoniak, M. Pugliatti, R. Hubbard, J. Britton, S. Sotgiu, A. D. Sadovnick, I. M. Yee, M. A. Cumsille, J. A. Bevilacqua, S. Burdett, *et al.*, "Vascular factors and risk of dementia: design of the three-city study and baseline characteristics of the study population," *Neuroepidemiology*, vol. 22, no. 6, pp. 316–325, 2003.

[32] R. Marani, D. Palumbo, V. Renò, U. Galietti, E. Stella, and T. D'Orazio, "Modeling and classification of defects in cfrp laminates by thermal non-destructive testing," *Composites Part B: Engineering*, vol. 135, pp. 129–141, 2018.

[33] Y. Zhou, P. Ghassemi, M. Chen, D. McBride, J. P. Casamento, T. J. Pfefer, and Q. Wang, "Clinical evaluation of fever-screening thermography: impact of consensus guidelines and facial measurement location," *Journal of Biomedical optics*, vol. 25, no. 9, p. 097002, 2020.

[34] D. Perpetuini, C. Filippini, D. Cardone, and A. Merla, "An overview of thermal infrared imaging-based screenings during pandemic emergencies," *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, p. 3286, 2021.

[35] D. Singh and A. K. Singh, "Role of image thermography in early breast cancer detection-past, present and future," *Computer methods and programs in biomedicine*, vol. 183, p. 105074, 2020.

[36] S. Ekici and H. Jawzal, "Breast cancer diagnosis using thermography and convolutional neural networks," *Medical hypotheses*, vol. 137, p. 109542, 2020.

[37] D. G. Armstrong, K. Holtz-Neiderer, C. Wendel, M. J. Mohler, H. R. Kimbriel, and L. A. Lavery, "Skin temperature monitoring reduces the risk for diabetic foot ulceration in high-risk patients," *The American journal of medicine*, vol. 120, no. 12, pp. 1042–1046, 2007.

[38] A. Urakov, N. Urakova, A. Samorodov, P. Shabanov, I. Yagudin, A. Stolyarenko, D. Suntsova, and N. Muhutdinov, "Thermal imaging of local skin temperature as part of quality and safety assessment of injectable drugs," *Heliyon*, vol. 10, no. 1, 2024.

[39] J. Pauk, M. Ihnatouski, and A. Wasilewska, "Detection of inflammation from finger temperature profile in rheumatoid arthritis," *Medical & Biological Engineering & Computing*, vol. 57, pp. 2629–2639, 2019.

[40] A. Casha, O. Falzon, and J. Gauci, "A preliminary Investigation into the use of Thermography during Cardiac Surgery," in *Proceedings of the 11th Malta Medical School Conference*, (Malta), 2015.

[41] I. Fernández-Cuevas, J. C. B. Marins, J. A. Lastras, P. M. G. Carmona, S. P. Cano, M. Á. García-Concepción, and M. Sillero-Quintana, "Classification of factors influencing the use of infrared thermography in humans: A review," *Infrared Physics & Technology*, vol. 71, pp. 28–55, 2015.

[42] Y. Houdas and E. Ring, *Human body temperature: its measurement and regulation*. Springer Science & Business Media, 2013.

[43] J. L. Gonzalez-Hernandez, A. N. Recinella, S. G. Kandlikar, D. Dabydeen, L. Medeiros, and P. Phatak, "Technology, application and potential of dynamic breast thermography for the detection of breast cancer," *International Journal of Heat and Mass Transfer*, vol. 131, pp. 558–573, 2019.

[44] S. Jankovic, S. Stankovic, S. Borjanovic, L. Tenjovic, and M. Bogdanovic, "Cold stress dynamic thermography for evaluation of vascular disorders in hand-arm vibration syndrome," *Journal of Occupational Health*, vol. 50, no. 5, pp. 423–425, 2008.

[45] K. D. Patil, J. R. Williams, and K. L. Williams, "Thermographic Localization of Incompetent Perforating Veins in the Leg," *British Medical Journal*, vol. 1, no. 5690, pp. 195–197, 1970.

[46] M. Kaczmarek, A. Nowakowski, M. Suchowirski, J. Siebert, and W. Stojek, "Active dynamic thermography in cardiosurgery," *Quantitative InfraRed Thermography Journal*, vol. 4, no. 1, pp. 107–123, 2007.

[47] M. Kaczmarek and A. Nowakowski, "Active Dynamic Thermography in Medical Diagnostics," in *Application of Infrared to Biomedical Sciences*, pp. 291–310, Springer, Singapore, 2017.

[48] K. Ammer, "Medical Thermology 2018 - A computer-assisted literature survey," *Thermology International*, vol. 29, no. 1, pp. 7–39, 2019.

[49] Y. V. Gulyaev, A. G. Markov, L. G. Koreneva, and P. V. Zakharov, "Dynamical Infrared Thermography in Humans," *IEEE Engineering in Medicine and Biology Magazine*, vol. 14, no. 6, pp. 766–771, 1995.

[50] O. Falzon, A. Ciantar, L. Sammut, M. Schembri, Y. M. Baron, J. Calleja-agius, P. Demicoli, and K. P. Camilleri, "Principal Component Analysis of Dynamic Thermography Data from Pregnant and Non-Pregnant Women," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5664–5667, IEEE, 2018.

[51] M. Etehadtavakol and E. Y. Ng, "Assessment of foot complications in diabetic patients using thermography: a review," *Application of infrared to biomedical sciences*, pp. 33–43, 2017.

[52] E. Jude, I. Eleftheriadou, and N. Tentolouris, "Peripheral arterial disease in diabetes—a review," *Diabetic medicine*, vol. 27, no. 1, pp. 4–14, 2010.

[53] A. Gatt, C. Formosa, K. Cassar, K. P. Camilleri, C. De Raffaele, A. Mizzi, C. Azzopardi, S. Mizzi, O. Falzon, S. Cristina, *et al.*, "Thermographic patterns of the upper and lower limbs: baseline data," *International journal of vascular medicine*, vol. 2015, no. 1, p. 831369, 2015.

[54] D. G. Armstrong, L. A. Lavery, P. J. Liswood, W. F. Todd, and J. A. Tredwell, "Infrared dermal thermometry for the high-risk diabetic foot," *Physical Therapy*, vol. 77, no. 2, pp. 169–175, 1997.

[55] S. J. Benbow, A. W. Chan, D. R. Bowsher, G. Williams, and I. A. Macfarlane, "The prediction of diabetic neuropathic plantar foot ulceration by liquid-crystal contact thermography," *Diabetes Care*, vol. 17, no. 8, pp. 835–839, 1994.

[56] M. Oe, R. Yotsu, H. Sanada, T. Nagase, and T. Tamaki, "Thermographic findings in a case of type 2 diabetes with foot ulcer and osteomyelitis," *Journal of wound care*, vol. 21, no. 6, pp. 274–278, 2012.

[57] M. Oe, R. R. Yotsu, H. Sanada, T. Nagase, and T. Tamaki, "Screening for osteomyelitis using thermography in patients with diabetic foot," *Ulcers*, vol. 2013, no. 1, p. 284294, 2013.

[58] T. Nagase, H. Sanada, K. Takehara, M. Oe, S. Iizaka, Y. Ohashi, M. Oba, T. Kadowaki, and G. Nakagami, "Variations of plantar thermographic patterns in normal controls and non-ulcer diabetic patients: novel classification using angiosome concept," *Journal of Plastic, reconstructive & aesthetic Surgery*, vol. 64, no. 7, pp. 860–866, 2011.

[59] R. Vardasca, C. Magalhaes, A. Seixas, R. Carvalho, and J. Mendes, "Diabetic foot monitoring using dynamic thermography and ai classifiers," in *Proc. QIRT Asia*, (Tokyo, Japan), 2019.

[60] J. Engel, "Thermography in locomotor diseases-recommended procedure-locomotor-diseases-group report 1978 of the anglo-dutch-thermographic-society," *European Journal of Rheumatology and Inflammation*, vol. 2, no. 3, pp. 299–306, 1979.

[61] E. Ring, "Standardisation of thermal imaging in medicine: physical and environmental factors," *Gautherie M, Albert E, Keith L. Thermal assessment of breast health. Lancaster/Boston/The Hague: MTP Press Ltda*, pp. 29–36, 1983.

[62] E. Ring, J. Engel, and D. Page-Thomas, "Thermologic methods in clinical pharmacology-skin temperature measurement in drug trials.," *International journal of clinical pharmacology, therapy, and toxicology*, vol. 22, no. 1, pp. 20–24, 1984.

[63] E. Ring, N. Aarts, C. Black, P. Boesiger, *et al.*, "Raynaud's phenomenon: assessment by thermography," *Thermology*, vol. 3, pp. 69–73, 1988.

[64] E. Ring and K. Ammer, "Infrared thermal imaging in medicine," *Physiological Measurement*, vol. 33, no. 3, p. R33, 2012.

[65] E. Ring and K. Ammer, "The technique of infrared imaging in medicine," in *Infrared Imaging: A casebook in clinical medicine*, IoP Publishing, 2015.

[66] W. Amalu, "International academy of clinical thermology medical infrared imaging standards and guidelines," *International Academy of Clinical Thermology: Foster City, CA, USA*, 2018.

[67] F. P. Oliveira and J. M. R. Tavares, "Medical image registration: a review," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 17, no. 2, pp. 73–93, 2014.

[68] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: a review," *Physics in Medicine & Biology*, vol. 65, no. 20, p. 20TR01, 2020.

[69] M. Hoffmann, B. Billot, D. N. Greve, J. E. Iglesias, B. Fischl, and A. V. Dalca, "Synthmorph: learning contrast-invariant registration without acquired images," *IEEE Transactions on Medical Imaging*, vol. 41, no. 3, pp. 543–558, 2021.

[70] X. Huang, J. Ren, G. Guiraudon, D. Boughner, and T. M. Peters, "Rapid dynamic image registration of the beating heart for diagnosis and surgical navigation," *IEEE Transactions on Medical Imaging*, vol. 28, no. 11, pp. 1802–1814, 2009.

[71] A. Gooya, G. Biros, and C. Davatzikos, "Deformable registration of glioma images using em algorithm and diffusion reaction modeling," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 375–390, 2010.

[72] Y. Fu, S. Liu, H. H. Li, and D. Yang, "Automatic and hierarchical segmentation of the human skeleton in ct images," *Physics in Medicine & Biology*, vol. 62, no. 7, p. 2812, 2017.

[73] X. Yang, P. Ghafourian, P. Sharma, K. Salman, D. Martin, and B. Fei, "Nonrigid registration and classification of the kidneys in 3d dynamic contrast enhanced (dce) mr images," in *Medical Imaging 2012: Image Processing*, vol. 8314, pp. 105–112, SPIE, 2012.

[74] E. Z. Barcelos, W. M. Caminhas, E. Ribeiro, E. M. Pimenta, and R. M. Palhares, "A combined method for segmentation and registration for an advanced and progressive evaluation of thermal images," *Sensors*, vol. 14, no. 11, pp. 21950–21967, 2014.

[75] X. Chen, A. Diaz-Pinto, N. Ravikumar, and A. F. Frangi, "Deep learning in medical image registration," *Progress in Biomedical Engineering*, vol. 3, no. 1, p. 012003, 2021.

[76] A. Mashekova, Y. Zhao, E. Y. Ng, V. Zarikas, S. C. Fok, and O. Mukhmetov, "Early detection of the breast cancer using infrared technology–a comprehensive review," *Thermal science and engineering progress*, vol. 27, p. 101142, 2022.

[77] T.-Y. Cheng and C. Herman, "Motion tracking in infrared imaging for quantitative medical diagnostic applications," *Infrared physics & technology*, vol. 62, pp. 70–80, 2014.

[78] J. V. Hajnal and D. L. Hill, *Medical image registration*. CRC press, 2001.

[79] A. Ciantar, O. Falzon, L. Sammut, M. Schembri, Y. M. Baron, J. Calleja-Agius, P. P. Demicoli, and K. P. Camilleri, "Registration of dynamic thermography data of the abdomen of pregnant and non-pregnant women," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5668–5671, IEEE, 2018.

[80] D. Bouallal, H. Douzi, and R. Harba, "Registration methods for thermal images of diabetic foot monitoring: A comparative study," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, 2022.

[81] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611, pp. 586–606, Spie, 1992.

[82] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics letters*, vol. 33, no. 2, pp. 156–158, 2008.

[83] P. Thevenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE transactions on image processing*, vol. 7, no. 1, pp. 27–41, 1998.

[84] A. Gatt, O. Falzon, K. Cassar, C. Ellul, K. P. Camilleri, J. Gauci, S. Mizzi, A. Mizzi, C. Sturgeon, L. Camilleri, *et al.*, "Establishing differences in thermographic patterns between the various complications in diabetic foot disease," *International Journal of Endocrinology*, vol. 2018, 2018.

[85] J. Gauci, O. Falzon, C. Formosa, A. Gatt, C. Ellul, S. Mizzi, A. Mizzi, C. Sturgeon Delia, K. Cassar, N. Chockalingam, *et al.*, "Automated region extraction from thermal images for peripheral vascular disease monitoring," *Journal of Healthcare Engineering*, vol. 2018, 2018.

[86] H.-D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern recognition*, vol. 34, no. 12, pp. 2259–2281, 2001.

[87] K. Ramesh, G. K. Kumar, K. Swapna, D. Datta, and S. S. Rajest, "A review of medical image segmentation algorithms," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 7, no. 27, pp. e6–e6, 2021.

[88] Y.-Y. Chen, W.-S. Chen, and H.-S. Ni, "Image segmentation in thermal images," in *2016 IEEE International Conference on Industrial Technology (ICIT)*, pp. 1507–1512, IEEE, 2016.

[89] F. H. Chan, F. K. Lam, and H. Zhu, "Adaptive thresholding by variational method," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 468–473, 1998.

[90] R. Muthukrishnan and M. Radha, "Edge detection techniques for image segmentation," *International Journal of Computer Science & Information Technology*, vol. 3, no. 6, p. 259, 2011.

[91] S. Gould, T. Gao, and D. Koller, "Region-based segmentation and object detection," *Advances in neural information processing systems*, vol. 22, 2009.

[92] K. P. Sinaga and M.-S. Yang, "Unsupervised k-means clustering algorithm," *IEEE access*, vol. 8, pp. 80716–80727, 2020.

[93] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[94] I. Levner and H. Zhang, "Classification-driven watershed segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 5, pp. 1437–1445, 2007.

[95] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.

[96] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021.

[97] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4026, 2023.

[98] L. Ke, M. Ye, M. Danelljan, Y.-W. Tai, C.-K. Tang, F. Yu, *et al.*, "Segment anything in high quality," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[99] Z. Kütük and G. Algan, "Semantic segmentation for thermal images: A comparative survey," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 286–295, 2022.

[100] C. Li, W. Xia, Y. Yan, B. Luo, and J. Tang, "Segmenting objects in day and night: Edge-conditioned cnn for thermal image semantic segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 7, pp. 3069–3082, 2020.

[101] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 1395–1403, 2015.

[102] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[103] H. Xiong, W. Cai, and Q. Liu, "Mcnet: Multi-level correction network for thermal image semantic segmentation of nighttime driving scene," *Infrared Physics & Technology*, vol. 113, p. 103628, 2021.

[104] K. Panetta, K. S. Kamath, S. Rajeev, and S. S. Agaian, "Ftnet: Feature transverse network for thermal image semantic segmentation," *IEEE Access*, vol. 9, pp. 145212–145227, 2021.

[105] P. Wang and X. Bai, "Thermal infrared pedestrian segmentation based on conditional gan," *IEEE transactions on image processing*, vol. 28, no. 12, pp. 6007–6021, 2019.

[106] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 234–241, Springer, 2015.

[107] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.

[108] S. Chen, Z. Chen, X. Xu, N. Yang, and X. He, "Nv-net: Efficient infrared image segmentation with convolutional neural networks in the low illumination environment," *Infrared Physics & Technology*, vol. 105, p. 103184, 2020.

[109] A. K. Bhandari and K. Rahul, "A context sensitive masi entropy for multilevel image segmentation using moth swarm algorithm," *Infrared Physics & Technology*, vol. 98, pp. 132–154, 2019.

[110] D. D. Patil and S. G. Deore, "Medical image segmentation: a review," *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 1, pp. 22–27, 2013.

[111] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of digital imaging*, vol. 32, pp. 582–596, 2019.

[112] J. Knittel, A. Lalama, S. Koch, and T. Ertl, "Visual neural decomposition to explain multivariate data sets," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1374–1384, 2020.

[113] E. O. Brigham, *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.

[114] C. O. S. Sorzano, J. Vargas, and A. P. Montano, "A survey of dimensionality reduction techniques," *arXiv preprint arXiv:1403.2877*, 2014.

[115] S. Prema and R. Jehadeesan, "Analysis of parallelization techniques and tools," *International Journal of Information and Computation Technology*, vol. 3, no. 5, pp. 471–478, 2013.

[116] L. Wang and Y. Shao, "Fault feature extraction of rotating machinery using a reweighted complete ensemble empirical mode decomposition with adaptive noise and demodulation analysis," *Mechanical systems and signal processing*, vol. 138, p. 106545, 2020.

[117] V. Zarnowitz and A. Ozyildirim, "Time series decomposition and measurement of business cycles, trends and growth cycles," *Journal of Monetary Economics*, vol. 53, no. 7, pp. 1717–1739, 2006.

[118] V. Klema and A. Laub, "The singular value decomposition: Its computation and some applications," *IEEE Transactions on automatic control*, vol. 25, no. 2, pp. 164–176, 1980.

[119] H. Hotelling, "Analysis of a complex of statistical variables into principal components.No Title," *Journal of educational psychology*, vol. 24, no. 6, p. 417, 1933.

[120] S. Wold, K. Esbensen, and P. Geladi, "Principal Component Analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[121] A. Hyvärinen, "Independent component analysis: recent advances," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, p. 20110534, 2013.

[122] D. Zhang and D. Zhang, "Wavelet transform," *Fundamentals of image data mining: Analysis, Features, Classification and Retrieval*, pp. 35–44, 2019.

[123] J. C. Cooper, "Factor analysis: an overview," *The American Statistician*, vol. 37, no. 2, pp. 141–147, 1983.

[124] P. J. Schmid, "Dynamic mode decomposition and its variants," *Annual Review of Fluid Mechanics*, vol. 54, no. 1, pp. 225–254, 2022.

[125] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Dynamic mode decomposition with control," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142–161, 2016.

[126] T. Chao and R. Sanzolone, "Decomposition techniques," *Journal of Geochemical Exploration*, vol. 44, no. 1-3, pp. 65–106, 1992.

[127] M. Agarwal and R. Mehra, "Review of matrix decomposition techniques for signal processing applications," *Int. Journal of Engineering Research and Applications*, vol. 4, no. 1, pp. 90–93, 2014.

[128] J. Lu, "Matrix decomposition and applications," 2023.

[129] B.-S. Yang, F. Gu, A. Ball, *et al.*, "Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 601–614, 2013.

[130] R. Vidhya, P. V. Ranjan, and N. Shanker, "Transformer breather thermal image decomposition for fault diagnosis," in *2021 7th International Conference on Electrical Energy Systems (ICEES)*, pp. 448–453, IEEE, 2021.

[131] M. Mozafari, A. J. Law, J. R. Green, and R. A. Goubran, "Respiration rate estimation from thermal video of masked and unmasked individuals using tensor decomposition," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–5, IEEE, 2022.

[132] R. Szeliski, *Computer vision: algorithms and applications*. Springer Nature, 2022.

[133] M. Jaderberg, K. Simonyan, A. Zisserman, *et al.*, "Spatial transformer networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[134] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[135] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ants similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.

[136] B. B. Avants, "Antspy." https://github.com/ANTsX/ANTsPy, 2017.

[137] B. B. Avants, N. Tustison, G. Song, *et al.*, "Advanced normalization tools (ants)," *The Insight Journal*, vol. 2, no. 365, pp. 1–35, 2009.

[138] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

[139] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[140] J. Devlin, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[141] A. Radford, "Improving language understanding by generative pre-training," 2018.

[142] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[143] A. Gatt, K. Cassar, O. Falzon, C. Ellul, K. P. Camilleri, J. Gauci, S. Mizzi, A. Mizzi, C. Sturgeon, N. Chockalingam, *et al.*, "The identification of higher forefoot temperatures associated with peripheral arterial disease in type 2 diabetes mellitus as detected by thermography," *Primary care diabetes*, vol. 12, no. 4, pp. 312–318, 2018.

[144] D. Hernández-Contreras, H. Peregrina-Barreto, J. Rangel-Magdaleno, F. Orihuela-Espina, and J. Ramirez-Cortes, "Measuring changes in the plantar temperature distribution in diabetic patients," in *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–6, IEEE, 2017.

[145] A. Wasilewska, J. Pauk, and M. Ihnatouski, "Image processing techniques for roi identification in rheumatoid arthritis patients from thermal images," *acta mechanica et automatica*, vol. 12, no. 1, pp. 49–53, 2018.

[146] V. Agostini, S. Delsanto, F. Molinari, and M. Knaflitz, "Evaluation of feature-based registration in dynamic infrared imaging for breast cancer diagnosis," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006. EMBS'06.*, pp. 953–956, IEEE, 2006.

[147] I. Fernández-Cuevas, J. C. Bouzas Marins, J. Arnáiz Lastras, P. M. Gómez Carmona, S. Piñonosa Cano, M. Á. García-Concepción, and M. Sillero-Quintana, "Classification of factors influencing the use of infrared thermography in humans: A review," *Infrared Physics and Technology*, vol. 71, pp. 28–55, 2015.

[148] J. Bass, "Circadian topology of metabolism," *Nature*, vol. 491, no. 7424, pp. 348–356, 2012.

[149] I. T. Jolliffe, "Principal component analysis: a beginner's guide—i. introduction and application," *Weather*, vol. 45, no. 10, pp. 375–382, 1990.

[150] M. F. Modest and S. Mazumder, *Radiative heat transfer*. Academic press, 2021.

[151] B. F. Jones, "A reappraisal of the use of infrared thermal image analysis in medicine," *IEEE transactions on medical imaging*, vol. 17, no. 6, pp. 1019–1027, 1998.

The Physics of Thermal Imaging

A blackbody is defined as an object which absorbs all incident radiation, without any reflection or transmission, and emits electromagnetic radiation according to Planck's law [150] which gives the body's spectral emissive power at wavelength, λ, E_λ by [24]:

$$E_\lambda(\lambda, T) = \frac{C_1}{\lambda^5 e^{\frac{C_2}{\lambda T}} - 1} \tag{1}$$

where $C_1 = 3.742 \times 10^8\ W \cdot \mu m^4/m^2$ is the first radiation constant, $C_2 = 1.439 \times 10^4$ $\mu m \cdot K$ is the second radiation constant and T is the temperature of the body in Kelvin [150]. Theoretically, objects radiate electromagnetic radiation in all wavelengths but in many practical objects the IR band is the spectral band with the highest emitted energy. Figure 1 plots the spectral emissive power at different wavelengths computed using Planck's law for various blackbody temperatures.

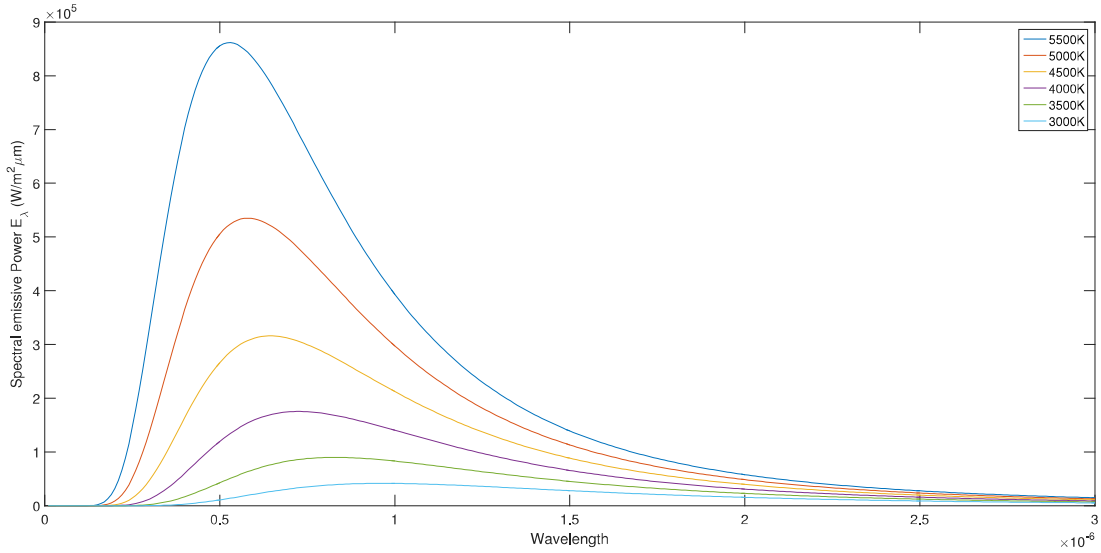


Figure 1: Planck's Law for blackbodies at different temperatures. As temperature rises, the peak emissive power moves to a shorter wavelength but increases in magnitude.

The Stefan-Boltzmann law integrates Plank's law over all wavelengths to describe the total emissive power of a blackbody source [24]:

$$E = \sigma T^4 \tag{2}$$

where $\sigma = 5.67 \times 10^{-8}\ W/m^2 \cdot K^4$ is the Stefan-Boltzmann constant and T is the blackbody temperature in Kelvin. The total emissive power of the blackbody increases with temperature. According to Wien's law, the wavelength at which maximum emissive power occurs, λ_{max}, shifts to longer wavelengths as temperature decreases and can be calculated using:

$$\lambda_{max} T = b \tag{3}$$

where $b = 2.90 \times 10^{-3}\ m \cdot K$ is Wien's displacement constant [151]. Wien's law can be observed in the peaks in Figure 1 in which the peak emissive power shifts to longer wavelengths as temperature decreases.

The absorption coefficient, α, of a body is defined as the fraction of incident radiation which is absorbed by the body [151]. For a body in thermal equilibrium the energy radiated must be equal to the energy absorbed by the body. Using the definition of a body in thermal equilibrium and Kirchhoff's law, the emissivity, ϵ, of a body can be defined as [151]:

$$\alpha = \frac{E}{E_{BB}} = \epsilon \tag{4}$$

The emissivity, ϵ, of a body defines the ratio of the body's surface emissive power, E, to the emissive power of a blackbody, E_{BB}. A blackbody thus has an emissivity value of $\epsilon = 1$ and an absorption coefficient $\alpha = 1$ [150]. For any real surface, the emissivity, and by extension the absorption coefficient, are always less than unity. The definition of the emissivity of a body also modifies Stefan-Boltzman law, given by Equation 7, for real bodies as follows [150]:

$$E = \epsilon \sigma T^4 \tag{5}$$

A blackbody thus, radiates with maximum power, over the whole spectrum, at temperature T, while real surfaces radiate with a fraction of this power according to the surface emissivity value. The emissivity of a surface is also related to the surface reflection coefficient, ρ, and surface transmission coefficient, τ , by [150]:

$$\epsilon + \rho + \tau = 1 \tag{6}$$

where ρ is the reflectance of the surface, which shows how much of the incident radiation is reflected by the surface and τ is the transmittance of the surface, which shows

how much of the incident radiation is transmitted through the surface. A blackbody, which absorbs all incident radiation, thus $\alpha = 1$, and has unity emissivity, thus $\epsilon = 1$, has $\rho = \tau = 0$ while real surfaces with $\epsilon < 1$ have some degree of reflectance and transmittance [150].

Applications of thermal imaging include:

- Electrical and Mechanical Inspections: Identifying overheating components and potential issues in electrical systems and machinery for preventive maintenance.
- Search and Rescue: Locating people or animals in challenging conditions, such as smoke, darkness, or dense foliage.
- Building Diagnostics: Detecting thermal anomalies in construction, insulation, and HVAC systems, facilitating energy efficiency and building safety.
- Medical Diagnosis: Monitoring and diagnosing various health conditions, including fever detection and the identification of inflammation.
- Wildlife Observation: Observing and studying animal behaviou; tracking and locating wildlife in natural habitats.
- Military and Defence: Enhancing situational awareness, identifying concealed threats, and improving night vision capabilities.
- Firefighting: Locating hotspots and trapped individuals in smoke-filled environments.
- Agriculture: Assessing crop health and optimising irrigation by detecting variations in plant temperature.
- Scientific Research: Studying diverse phenomena, including volcanic eruptions, climate change, and astronomical observations.

Ethics Approval

For the purpose of this study, ethics approval was sought from the University of Malta Research Ethics Committee. This approval was granted in April 2019 and the below is the consent form which was presented to the subjects taking part in this study.

L-Università ta' Malta
Centre for Biomedical Cybernetics

Centre for Biomedical Cybernetics
University of Malta
Msida MSD2080 Malta
Tel: 23402865
jean.gauci@um.edu.mt
www.um.edu.mt/cbc

Analysis of temperature transient patterns using dynamic infrared thermography.

Informed Consent Form

Before you decide to participate in this research study it is important that you understand the purpose of the research and the procedures involved. Please take time to read the following information carefully and to decide whether or not you wish to participate.

Purpose of the research: In this research, a set of thermal cameras will be used in order to measure and analyse individuals' skin temperature fluctuations in time.

Procedure: If you decide to participate in this research, you will be asked to follow the procedures described in the participant information sheet prior to the commencement of the study. On the day of the study you will be positioned on a couch in front of a thermal camera with your shoes off in a controlled environment. The expected duration of the study is one hour.

Risks and benefits: The procedure is completely painless and there are no foreseeable risks or discomfort associated with taking part in the study. There are no direct benefits to participants in the study. Any diagnostic tool developed from this study will be beneficial to society.

Confidentiality of records: All the information which you provide during the research will be kept strictly confidential. All personal data will be stored separately. Your data will be completely anonymised and there is no way that it can be linked to your personal data. Since the data is fully anonymised and therefore you will not be identifiable, your data cannot be deleted on request. Your data will be stored till the end of the study and deleted thereafter. With your consent your data may be used for additional or subsequent research (Tick below to confirm).

Voluntary participation: Your participation in this research is entirely voluntary. You are free to stop at any time. Refusal to participate or stop the study will involve no penalty or loss of benefits. Your personal information will be deleted immediately if you do not participate or if you decide to withdraw from this study.

Contact: In case of further questions kindly contact the researcher: Jean Gauci (Email: jean.gauci@um.edu.mt, Tel: 23402865) who is under the supervision of Dr. Owen Falzon (Email: owen.falzon@um.edu.mt, Tel: 23403651) and Prof. Ing. Kenneth Camilleri (Email: kenneth.camilleri@um.edu.mt, Tel: 23402070)

By ticking this box you agree that your data may be stored after the study has been completed and may be used for additional or subsequent research.

Tick to confirm ☐

By ticking this box you agree that your data may be used and published anonymously in an online repository.

Tick to confirm ☐

By ticking this you confirm that a copy of this consent form together with the participant information sheet has been provided to you.

Tick to confirm ☐

By ticking this box you confirm that you have read and understood the information above. You did have opportunity to ask questions about the study and your questions have been satisfactorily answered. You agree to participate in this research.

Tick to confirm ☐

Participant's name _____ Researcher's name _____

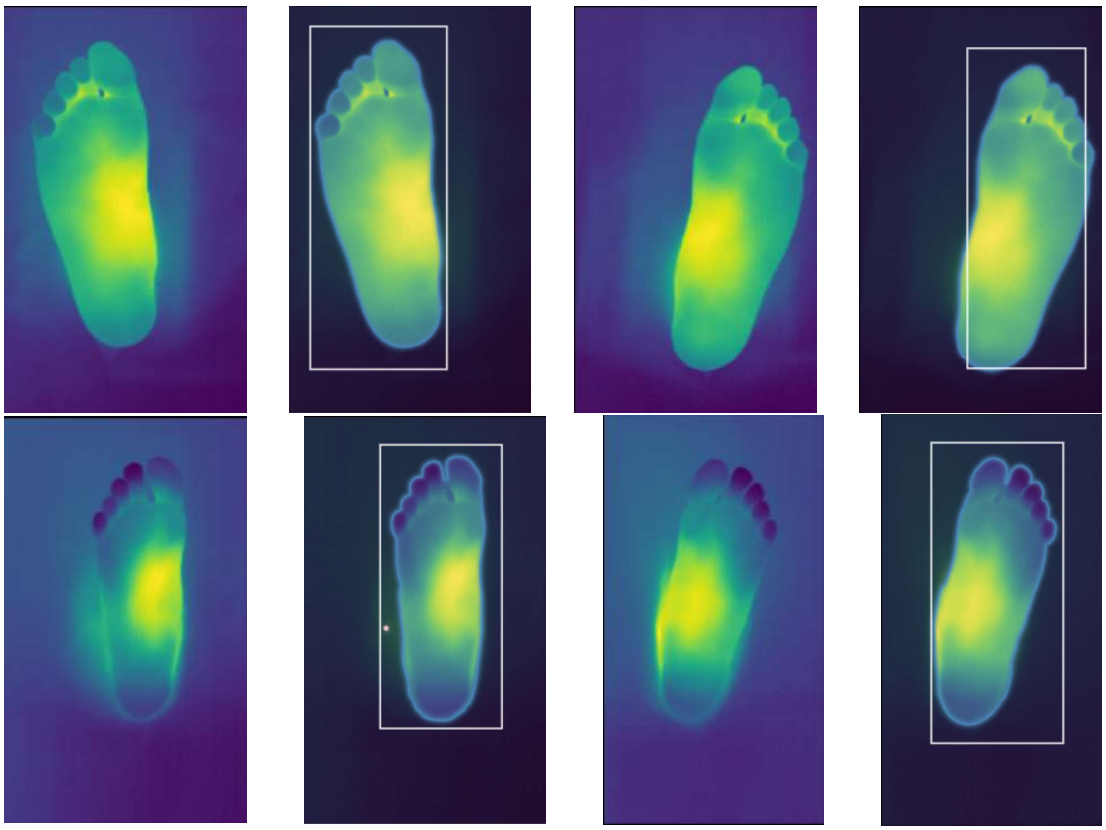
Participant's signature _____ Researcher's signature _____

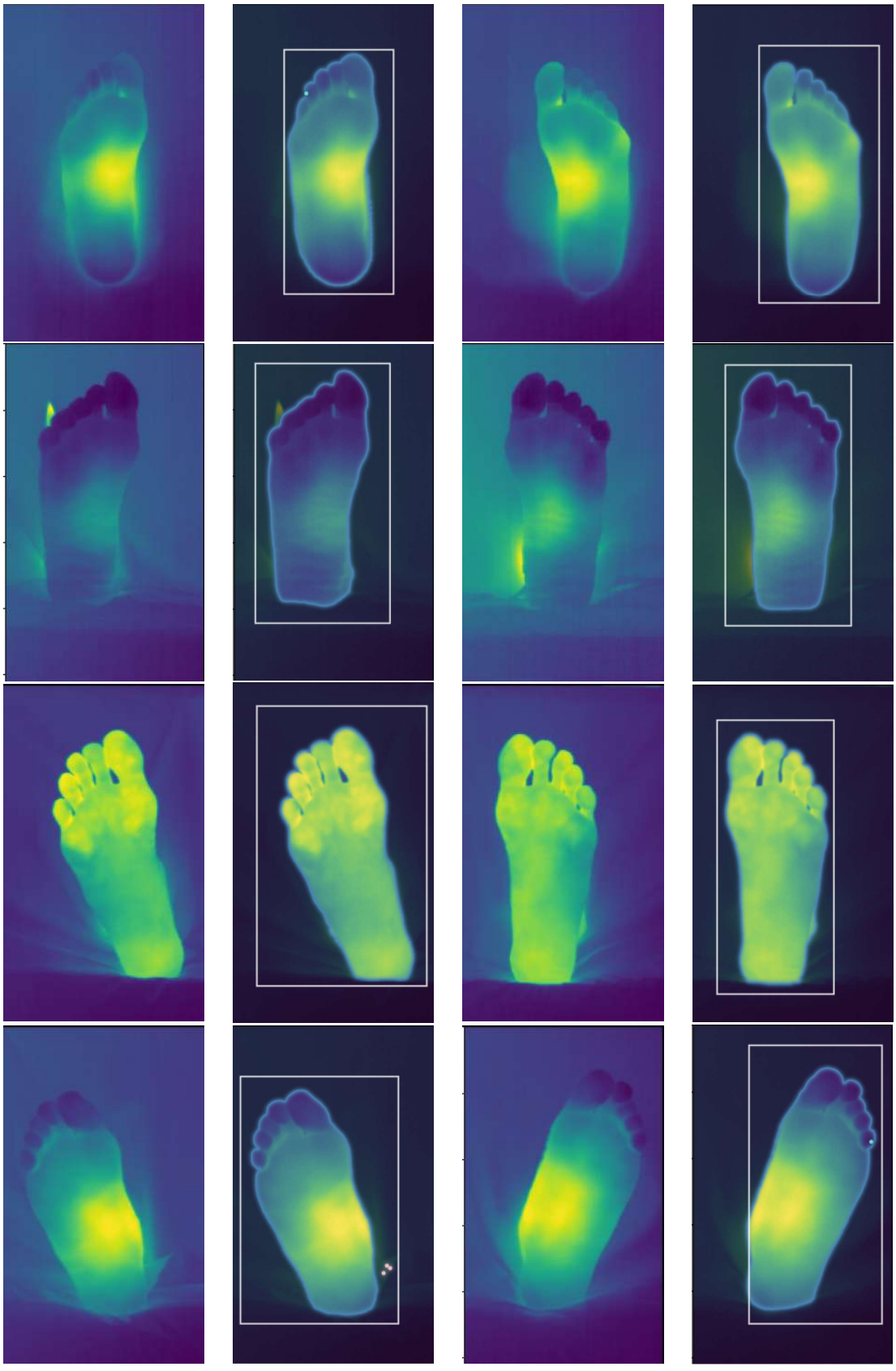
Date of consent _____

Full Segmentation Results

The full results from the segmentation method presented in Chapter 4 are presented in this Appendix. Each row of the figure contains the input and output images from one participant split by right and left feet. On the left side of the figure, the input image is shown while on the right hand side, the prompts, in form of a box and where required a spot prompt are overlayed on the output image together with the segmentation mask.

It is worth noting that all of the images used for this experiment have been segmented satisfactorily.





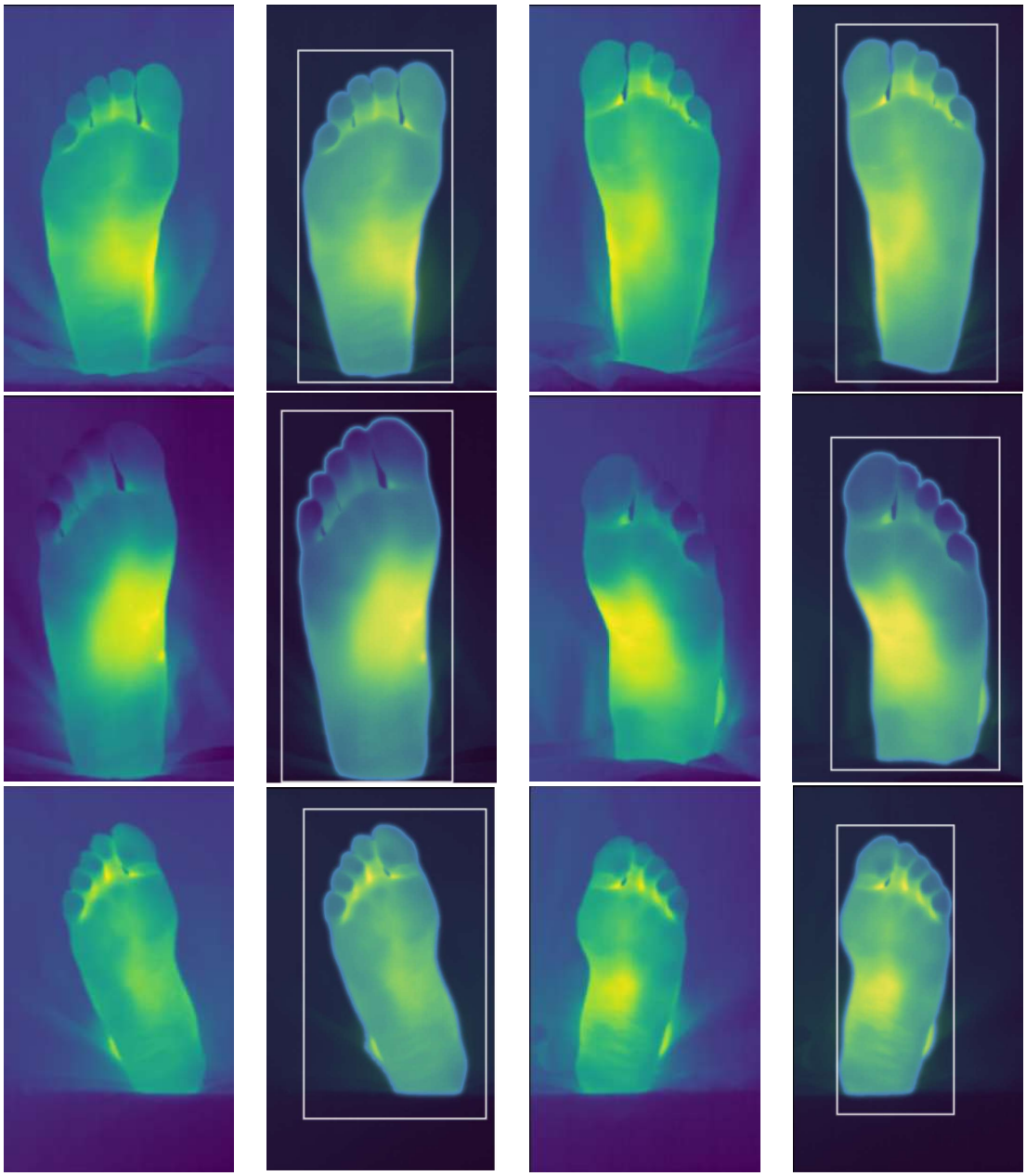


Figure 2: Full segmentation results