Segmentation and Extraction of Regions of Interest for Automated Detection of Anomalies in Clinical Thermal Infrared Images

by

Christophe L. Herry

A thesis submitted to
The Faculty of Graduate Studies and Research
In partial fulfilment of the requirements for the degree of
Doctor of Philosophy
in Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
Department of Systems and Computer Engineering
Carleton University
Ottawa, Ontario, Canada

April 2008

Your file Votre référence
ISBN: 978-0-494-40523-9
Our file Notre référence
ISBN: 978-0-494-40523-9

The undersigned hereby recommend to the Faculty of Graduate Studies and Research acceptance of the thesis

Segmentation and Extraction of Regions of Interest for Automated Detection of Anomalies in Clinical Thermal Infrared Images

Submitted by Christophe L. Herry, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

---

Dr. Victor Aitken, Chair, Department of Systems and Computer Engineering

---

Dr. Robert Aime Leblanc, External Examiner

Dr. Monique Frize, Thesis Co-supervisor

Dr. Rafik A. Goubran, Thesis Co-supervisor

Carleton University, April 2008

Text incomplete; leaves iii omitted.

Abstract

This doctoral thesis presents a complete framework for automated detection of anomalies in clinical thermal IR images. Within that framework, it addresses two fundamental problems: the automated segmentation of a main object from the background; and the automated extraction of relevant, potentially abnormal Regions of Interest. A first segmentation approach called cued morphological processing of edge maps (CMpEm) is introduced, that uses a minimum amount of *a priori* information about a region to constrain the morphological processing of edge components. It is shown to outperform existing segmentation methods, especially when dealing with faint regions in challenging conditions. A second segmentation approach based on the classification of edge components is presented that is able to recover missing contours when other methods fail. A third segmentation strategy based on a rule-based fusion and morphological post-processing of segmented contours from different techniques is proposed and shown to improve significantly the performance of the global segmentation task. Furthermore, a new approach for automated extraction of anatomical Regions of Interest in clinical IR images is introduced. It is based on a robust landmark identification algorithm that produces more accurate landmark locations on smooth contours than state-of-the-art algorithms. Various aspects of the anomaly detection framework, including blind extraction of potentially abnormal Regions of Interest, are tested on a collection of thermal IR databases acquired in the course of this thesis, in order to demonstrate the key advantages of the framework.

Acknowledgements

I would like to express my gratitude to my thesis supervisors Dr. Monique Frize and Dr. Rafik A. Goubran for their guidance, for their support, both academic and financial, and for their most helpful advice throughout my doctoral studies.

I am thankful to Dr. Victor Aitken and Dr. Richard Dansereau from the Department of Systems and Computer Engineering, Dr A.O. Abd El Halim from the Department of Civil and Environmental Engineering, Dr. Robert Aime Leblanc from the Université de Montréal and Dr. Pierre Payeur from the University of Ottawa for their willingness to act as committee members and for their helpful comments and advice.

I would also like to thank the many friends and fellow graduate students I have had the privilege to know while working on my thesis, and the helpful technical and administrative staff from the Department of Systems and Computer Engineering. Last but not least, I am truly grateful to my family, my wife Melanie and my son Louis-Gabriel for their love, support and infinite patience.

Contents

List of Tables

List of Figures

List of Acronyms

AAM Active Appearance Model

ACLS Active Contour Level Set

ANN Artificial Neural Network

ART Adaptive Resonance Theory

ATD Automated Target Detection

ATR Automated Target Recognition

CT Computed Tomography

CMpEm Cued Morphological Processing of Edge Maps

CSS Curvature Scale-Space

C-SVC C-Support Vector Classification

FCM Fuzzy C-Means

FFT Fast Fourier Transform

FPA Focal Plane Array

FLIR Forward Looking Infrared

GMLP Gaussian Multilayer Perceptron

GMM Gaussian Mixture Modelling

GWR Grow When Required

IR Infrared

KLT Karhunen-Loeve Transform

KNN K-Nearest-Neighbour

MLP Multilayer Perceptron

MRI Magnetic Resonance Imaging

PCA Principal Components Analysis

PNN Probabilistic Neural Network

PRMD Playing-Related Musculoskeletal Disorder

RBF Radial Basis Function

ROI Region of Interest

ROIs Regions of Interest

RSI Repetitive Strain Injuries

SNR Signal-to-Noise Ratio

SOM Self-Organizing Map

SVM Support Vector Machine

TRF Thermoregulatory Frequencies

UCI University of California Irvine

USPS United States Postal Service

Chapter 1

Introduction

1.1 Motivation

Thermal Infrared (IR) imaging has been used extensively in a wide range of applications, ranging from military to non-destructive testing or veterinary applications. It has also been used as a clinical imaging modality in the 1960's-1980's. Several clinical applications were investigated: pain assessment, breast cancer and rheumatoid arthritis, among others. Promising results were reported in the early literature [1–3]. Other studies pointed out some limitations of the technique for clinical purposes [4] [5]. Indeed, a number of issues were identified early on that prevented the widespread use of thermal IR imaging for medical applications until recently: Poor understanding of the limitations of the modality, technological limitations of early systems (size and weight, cooling requirements), subjectivity of the interpretation of thermograms, lack of sensitivity and the cost of IR systems. However, advances in IR imaging technology and standardized protocols have led to a renewed interest in the technique for medical applications. Standards for thermographic examination

protocols are now agreed upon [6] [7]. New IR systems have higher thermal and spatial resolutions and faster frame and processing rates. As a result the quality of the digital IR images produced is usually very good.

On the other hand, the issue of interpreting the resulting IR images remains. The interpretation of thermograms is still mostly done "visually" by a clinical expert trained in the interpretation of thermograms. It is assumed that the information contained in the image presented to the clinical expert is an unbiased and full representation of the actual information obtained from the patient. Unfortunately, this is not always the case. The interpretation of thermograms relies on being able to differentiate subtle temperature variations and on recognizing typical patterns associated with a particular pathology or a physiological problem. The ability of the eye to detect those changes in a grey-level image, where each grey level corresponds to a temperature or a range of temperatures, is not very good. Using pseudo-colours is not a good alternative due to the psychological effect of warm and cold colours [8]. A quantitative assessment of thermograms is a good way to circumvent the subjectivity of the current method of assessment. Currently, quantitative assessment consists of manual identification of potential regions of interest, detection of hot/cold spots, simple asymmetry analysis with first-order statistics and subtraction of images within a sequence (see for example [9]). Few automated computer-aided analysis and processing systems for clinical thermal IR images have been proposed to date. A general strategy for addressing the specificities of clinical IR imaging has not been proposed either. In addition, some of the key building blocks of an automated anomaly detection architecture for clinical IR imaging are still missing.

1.2 Thesis Objectives and Definition of the Problem

This thesis proposes a general framework for automated detection of anomalies in clinical infrared images. Within that framework, it addresses two fundamental problems: the segmentation of a main object from the background; and the extraction of relevant, potentially abnormal, Region of Interests (ROIs).

Thermal IR imaging indirectly collects the surface temperature distribution of an object. As such, it is dependent not only on object temperature, but also on its surface properties, surface orientation, on the wavelength used and overall on the environmental conditions that could affect the collecting of IR radiation. This typically leads to a narrow range of intensities and sometimes strong overlapping of temperature distribution between the object and its background. In addition, thermal IR imaging tends to produce smooth image regions and no colour information is available, which limits or prohibits the use of segmentation techniques based on texture or colour. Segmenting even simple regions within an IR image can be challenging, as for example in the case of the segmentation of cold extremities (hands and feet). Segmentation techniques developed for other imaging modalities or other IR applications typically do not work well when applied to IR images of the human body. Few automated segmentation approaches have been proposed that work consistently well with this modality. For a proper assessment of anomalies in IR images of the human body, a robust segmentation approach is necessary.

The identification of potentially abnormal regions of interest in IR images of the human body can be attempted blindly, if no prior information is available about the

problem at hand. However, a more focused search can be performed if a specific clinical context is known, by looking for specific anatomical structures or physiologically active areas. As mentioned previously, IR thermal images tend to be smooth and lack overall texture and colour information. Temperature variations within an object or a body region are usually not indicative of structural regions such as bones in X-rays or white matter in a brain Magnetic Resonance Imaging (MRI). As a result, extracting anatomical information is difficult. This thesis proposes a novel approach to extract anatomical regions based on the detection of anatomical landmarks.

In order to achieve these objectives, the following work will be presented. First, a novel algorithm that uses *a priori* knowledge to constrain a morphological processing of edge maps is introduced. A second novel segmentation approach using a Multilayer Perceptron (MLP) and a Support Vector Machine (SVM) to classify edge components is proposed, that is able to recover contours of body regions for a variety of situations. When trained on a specific database, this second algorithm will be shown to generalize well to other similar body regions. Finally, several fusion techniques are investigated in order to improve the segmentation results. In particular, a new fusion method is introduced where the agreement between segmentation techniques is used to constrain morphological post-processing operations. This new fusion approach is a global approach as opposed to the pixel-based decision approach of voting scheme or Bayesian based fusion.

The extraction of potential abnormal ROIs is separated in two parallel operations. If expert knowledge is available that allows to focus the search for abnormal regions to specific anatomical locations, then these anatomical locations can be retrieved through the localization of key landmark points. The most robust landmark points

in IR images are located on the contour of body regions and typically correspond to anthropometric landmarks. A novel algorithm to detect these landmark points is implemented and tested. It uses a three stage approach based on curvature analysis of contour, curvature analysis of the average curve between contours for symmetrical regions and region-specific constraints on the location of the landmarks. A clustering procedure allows one to sort through the most relevant landmark points. If expert knowledge is not available, then a blind search for potentially abnormal regions is performed. This is achieved through the fusion of clusters from a thresholded feature map and from an Adaptive Resonance Theory (ART) network. Finally, some applications of the architecture presented in this thesis are given.

1.3 Thesis Outline

The thesis is organized as follows. Chapter 2 presents background theory on IR imaging and IR image processing. In particular, Section 2.1 outlines briefly the physics of IR imaging. Section 2.2 describes the state of the art in the processing of IR images, including preprocessing, segmentation in Section 2.2.2, landmark identification in Section 2.2.3 and feature extraction in Section 2.2.5. The last part of chapter 2 reviews the state-of-the-art in the detection of anomalous regions in IR and optical images.

Chapter 3 presents the methodology used in this work. In particular, Section 3.1 shows an overview of the proposed framework. Section 3.3 details the segmentation strategy introduced including the cued morphological processing of edge maps approach and the various fusion strategies investigated. Sections 3.4 and 3.5 present the proposed algorithms for locating anatomical landmark points and extracting anatom-

ical regions of interest.

Chapter 4 gives an overview of the various databases of IR images that were acquired and used for testing. Chapter 5 presents the segmentation results for the various techniques introduced. Each algorithm was first tested on synthetic images (Section 5.1.2) and real IR images of extremities (Section 5.1.3). Breast segmentation results are covered in Section 5.2. In Chapter 6 the anatomical landmark identification results are presented. Chapter 7 introduces some of the applications of the proposed architecture. Finally Chapter 8 summarizes the contributions of this thesis and highlights potential future work.

Chapter 2

Background Review

2.1 Infrared Imaging

All objects heated at temperatures above absolute zero (zero degree Kelvin (°K) or -273 degrees Celsius (°C)) emit electromagnetic radiation [10–12]. The emissive power of the surface of an object is the total energy transferred from the object to its surroundings. It is proportional to the emissive power E_b of a blackbody, defined as an ideal body that absorbs all incident radiation and radiates in a continuous spectrum according to Planck's law (in vacuum):

$$E_b(\lambda, T) = \frac{2\pi hc^2}{\lambda^5(\exp(\frac{hc}{\lambda kT}) - 1)} \tag{2.1}$$

where $E_b(\lambda, T)$ is the emissive power of a blackbody in watts per steradian per square metre per micron for a particular wavelength λ (m) and temperature T (°K), h is Planck's constant $(6.626 \cdot 10^{34} Js)$, c is the velocity of light (ms^{-1}) and k is the Boltzmann's constant $(1.380 \cdot 10^{23} JK^{-1})$.

The integration of Planck's law over all wavelengths yields the total emissive power $E_b(T)$ $(Watt/m^2)$ of a blackbody [12], which is referred to as Stefan-Boltzmann's law and is given by:

$$E_b(T) = \sigma T^4 \tag{2.2}$$

where $\sigma = \frac{2\pi^5 k^4}{15c^2h^3} \approx 5.670 \cdot 10^8 Wm^2K^4$ is the Stefan-Boltzmann's constant.

The emissivity of an object is defined as the ratio of the emissive power $E_o(\lambda, T)$ from that object to the emissive power $E_b(\lambda, T)$ from a blackbody at the same temperature and wavelength [12]:

$$\epsilon_o(\lambda, T) = \frac{E_o(\lambda, T)}{E_b(\lambda, T)} \tag{2.3}$$

The emissivity of an object is not constant and depends on temperature, wavelength and direction. However, it may be considered as a constant within a certain bandwidth, range of temperature and cone of direction, which simplifies greatly the determination of the emissive power or other related quantities [10–12].

The amount of energy reflected by an object is determined by the reflectivity coefficient ρ_λ, which is given by the simple relationship [12]:

$$\rho_\lambda = 1 - \epsilon_\lambda \tag{2.4}$$

where ϵ_λ is the total emissivity. The reflectivity of a blackbody is null.

If a gradient of temperature is applied to a blackbody and if only radiative heat transfer is considered, the Stefan-Boltzmann's law becomes:

$$E_b = \sigma(T_b^4 - T_s^4) \tag{2.5}$$

where T_b is the temperature of the body and T_s is the temperature of the surroundings.

It is also interesting to notice that the emissive power reaches a maximum at a wavelength λ_{max} given by Wien's law (found by differentiating Planck's law (Eq. (2.1)) according to λ) [12]:

$$\lambda_{max}T = 0.002898m\,^{\circ}\mathrm{K} \tag{2.6}$$

where λ_{max} is the wavelength (in metres) of the maximum emissive power at temperature T (in °K). As a result, at typical room temperature (300 °K) the maximum emissive power is reached for a wavelength of 9.7μm, which is in the far infrared part of the spectrum.

Equations (2.1) and (2.3) provide a direct link between the measured radiative energy of a body with the temperature of that body for a specific wavelength. Thus, the accuracy of the measurement of the temperature of an object or a body depends on the ability to measure the radiative energy correctly. If an object has an emissivity less than unity, the total energy radiated from that body will be the sum of the energy emitted by the body itself, and the energy that is reflected by that body and comes from the surroundings or other bodies present at the time of the measurement. In such a case, the relationship between measured radiative energy and temperature has to be adjusted to take into account the reflectivity. It is therefore desirable to have an emissivity as close to unity as possible in order to have a high accuracy of temperature measurement [13]. The correction needed to account for the reflectivity of the skin is generally relatively small, especially if the environment has a temperature somewhat lower than that of skin and if the angle of viewing of the measurement is less than $\pi/4$, which is usually the case in clinical settings [13] [14].

The human body is typically at temperatures ranging from 300 °K to 315 °K and therefore emits energy in the form of electromagnetic waves. A major part of the energy radiated by the human body goes through the skin and occurs in the infrared part of the electromagnetic spectrum [15]. In particular, it has been found that the emissive power radiated by the skin reaches a peak around 9.5μm and it is very significant in the ranges $3-5\mu$m and $8-14\mu$m compared to the rest of the spectrum, although it should be noted that the average radiative power levels of the skin are low with respect to typical radiative power values in the visible range of the electromagnetic spectrum [13]. An interesting characteristic of human skin is its high emissivity in the $8-14\mu$m range, typically 0.97 to 0.98, which makes skin temperature measurements in that range particularly accurate. The emissivity of skin in the $3-5\mu$m range is somewhat lower than that in the $8-14\mu$m range and the former is also more prone to environmental artifacts according to [13]. As a result, measurements of skin temperature should be carried out in the $8-14\mu$m region in order to obtain better accuracy and more reliable results.

2.1.1 Clinical Infrared Thermography

Clinical infrared thermography or infrared thermal imaging is defined as the recording of the temperature distribution of the human body (or body areas) using infrared radiation emitted by the surface of that body, i.e., the skin [14]. In other words, medical thermography involves a collection of images of the temperature distribution of the surface of the body. Thermography can therefore provide valuable information on the physiology of the human body. It is important to keep in mind that medical thermography is a functional test and therefore cannot be expected to provide a direct

diagnosis of anatomical or structural pathologies. Thermography can however suggest location and/or cause of dysfunctions of an anatomical or neurological nature [8]. The skin temperature distribution of a healthy human body exhibits a contralateral symmetry. Temperature distribution that shows asymmetrical patterns is usually a strong indicator of abnormality [16], but the converse is not always true since some pathological conditions may exhibit bilateral thermal dysfunction and in such cases other signs of abnormalities in the temperature distribution have to be found [17].

Medical infrared imaging must follow strict protocols and standards in order to achieve accurate measurements [6] [14]. In particular, the control of the environment (temperature, humidity, and reflectivity) and of the thermal characteristics of the skin is critical.

Advocates of medical thermography often put forward that it is non-invasive, low cost (after the initial purchase cost) and that it has a wide range of applications in neurology, vascular disorders, rheumatic diseases, oncology, neonatology or ophthalmology, among others. However, the sensitivity and the specificity of the technique vary greatly depending on the applications considered[1]. One explanation for the variation in performance of thermographic examination as a diagnostic aid might be the difficulty to interpret thermograms (i.e. thermographic or thermal images), due to the complexity of the phenomena involved and to the necessity of understanding all the underlying processes that lead to changes in the thermal distribution of the human body [8].

A main part of the earlier literature on medical thermography focuses on qualitative interpretation of thermograms. It involves the determination of abnormal

[1]The sensitivity can be defined as $sens. = \frac{TP}{TP+FN}$; the specificity as: $spec. = \frac{TN}{TN+FP}$; where TN, TP, FN and FP stand for true negative, true positive, false negative and false positive respectively

thermal variations of the skin by means of a visual assessment of pseudo-coloured or grey-level thermograms with the help of isothermal displays, visual localization of hot or cold spots and visual detection of asymmetries [2, 17–24]. The use of pseudo-colours for mapping the temperatures of a thermogram has been criticized for its subjectivity due to the psychological effect of certain colours, which may skew the observer's performance [8].

In order to improve the assessment of thermograms it is important to quantify thermographic findings. Normal values of asymmetries were investigated by Goodman *et al.*, Uematsu *et al.* and more recently by Herry *et al.* [16, 25–27]. They found that temperature differences between contralateral sides were very small in normal subjects and that the thermal symmetry was stable across regions of the body.

Extraction of the ROIs is typically done manually or in a semi-automated manner. Anbar identified two methods of extracting ROIs from a thermal image [8]. The first method implied the delimitation of a ROI on one side of the body by an operator and the corresponding contralateral ROI was obtained by basic reflection across a vertical line of symmetry. The second method was similar to that used by Goodman *et al.* in [16] and consisted of a grid of squares or rectangles superposed to the thermal image. Lipari and Head manually selected breast contours as ROIs and decomposed each breast into four smaller ROIs using anatomical reference points provided by the user [28] [29]. Several authors also suggested stretching and mapping techniques applied to ROIs to allow a more efficient comparison between different ROIs in dynamic thermography, especially when ROIs are not simple geometric shapes [30–33]. More complex methods of boundary or edge detection were implemented for breast segmentation by Qi, Ng and Scales using Hough transforms [34–36]. They are reviewed

in more detail in Section 2.2.2.

A statistical analysis is typically carried out on the ROIs. This includes hot/cold spot detection and first-order statistics. Another useful statistical tool is the histogram that represents the temperature distribution over a discrete range of values. Some authors varied the size of ROIs to further investigate the effect of scaling on temperature asymmetries and improve the statistical accuracy of temperature differences [16] [29]. Mabuchi *et al.* proposed an image processing program to evaluate and display the temperature difference distribution between one healthy side of the body and the contralateral healthy side [37]. Temperature differences were computed on a pixel-by-pixel basis inside trapezoids of various sizes that divided symmetrically the human body. The authors noted that their method works only for pathologies that do not affect both sides of the body. Also, the ROIs were pre-selected manually by the user. Vavilov *et al.* [38] studied the left-right thermal asymmetry of breasts, thorax and ankles of subjects previously classified as "practically healthy". They computed the signal-to-noise ratio, the Z-statistics and the Kolmogorov-Smirnov statistics of contralateral small rectangular regions of interest chosen manually by an operator to quantify thermal asymmetries. Their results suggest that it is possible to distinguish between normal and pathological thermal distributions with an appropriate choice of threshold for each of the statistical parameters, although they underlined the need for test procedures that are more specific to the area investigated. Koay *et al.* [39] proposed a method to assess the temperature asymmetry between left and right breast using an Artificial Neural Network (ANN). Statistical parameters were extracted from the whole breast and from a quadrant division of each breast. However, only two of them (mean and standard deviation) were used reliably due to the small sample

size. The authors showed that a simple perceptron was able to differentiate between normal and abnormal breast. Other quantitative measures of abnormalities include Collins' thermal index, based on isothermal temperature differences [1], and profile analysis [31]. Both of these methods require user intervention.

Sequences of thermal images can be analyzed with a Fast Fourier Transform (FFT) applied to time profiles [40]. Montoro and Anbar argued that frequencies of temperature oscillations are characteristic and that static thermography cannot diagnose reliably temperature asymmetries of less than 0.1 °C because the variation of temperature asymmetries over a short period of time can be of the order of that limit or even higher. They identified Thermoregulatory Frequencies (TRF) across the back ranging from 2 to 50 MHz but were unable to determine if all their frequencies were fundamental physiological frequencies or artifacts created by reflected higher frequencies. Anbar also mentioned several methods for an objective analysis of TRF, such as cluster analysis, feature extraction algorithm, or autocorrelation analysis [13].

Another original method to analyze a sequence of thermograms is the use of the Karhunen-Loeve Transform (KLT). In the case of sequential thermograms, it reduces the temporal and spatial information contained in the sequence in a small number of significant eigen-images, with specific weights to take into account the time dimension. Unser *et al.* [41] claimed that the analysis of the main eigenvalues enabled an easier detection of abnormalities because all the relevant information was present with reduced noise. Varga and DeMuynck [42] used a similar method to process sequential thermograms but used a unique colour coding scheme on the first three resultant eigen-images and summed the three resulting images to improve the discriminatory value of the final image. Varga and Hanka [43] proposed a complex method of classifi-

cation of the response to a cold stress test for normal and abnormal conditions of the hands. The images were standardized and presented as a dual representation before being processed by KLT in order to identify ROI for the feature extractor, which was based on Kittler and Young transformation. A 7-Nearest-Neighbour classifier was then used and the results were presented as a colour-coded image.

2.1.2 Applications and Specificities of Infrared Imaging

IR images differ significantly from visible images. The intensities of the former depend on object temperature, object surface properties, surface orientation, wavelength, etc. and are not necessarily uniform across an object, for which the temperature is uniform. The environmental conditions at the time of imaging also have great influence on the resulting images. In addition, the range of intensities of infrared images is typically much less than that of visible images, leading to low contrast, poor resolution, less texture information, etc. [44] [45].

These effects may be somewhat attenuated when imaging conditions are controlled (see imaging protocol for clinical applications in Appendix A) and when the object of interest has a high emissivity. For example, IR imaging of the body under controlled environmental conditions can produce higher quality images in terms of contrast, Signal-to-Noise Ratio (SNR) and resolution than typical IR imaging, such as military targets or predictive maintenance [46]. Yet, key features that are successfully used in the processing of visible images cannot be used reliably when dealing with IR images [45].

Military and remote sensing IR images usually exhibit a large amount of background clutter and their resolution varies widely depending on the range and the

environmental conditions that cannot be controlled [46]. In contrast, images of the human body have very good resolution and less background clutter, but the contrast can still be problematic depending on the body regions considered. Also, clinical IR thermal images have very little texture. Another major difference between medical IR images and other IR images is the nature of the target of interest. For example, in military IR imaging applications, such as Automated Target Recognition (ATR), the target is mostly rigid (e.g. tanks or airplanes) and the target shape is usually known, although the target presence might not be. A model can be developed to increase the accuracy of ATR systems in challenging conditions. On the other hand, medical targets, such as tumours, change shape easily. The size, shape and orientation of the target can vary over time and modelling of the biological processes causing the changes is very challenging [46].

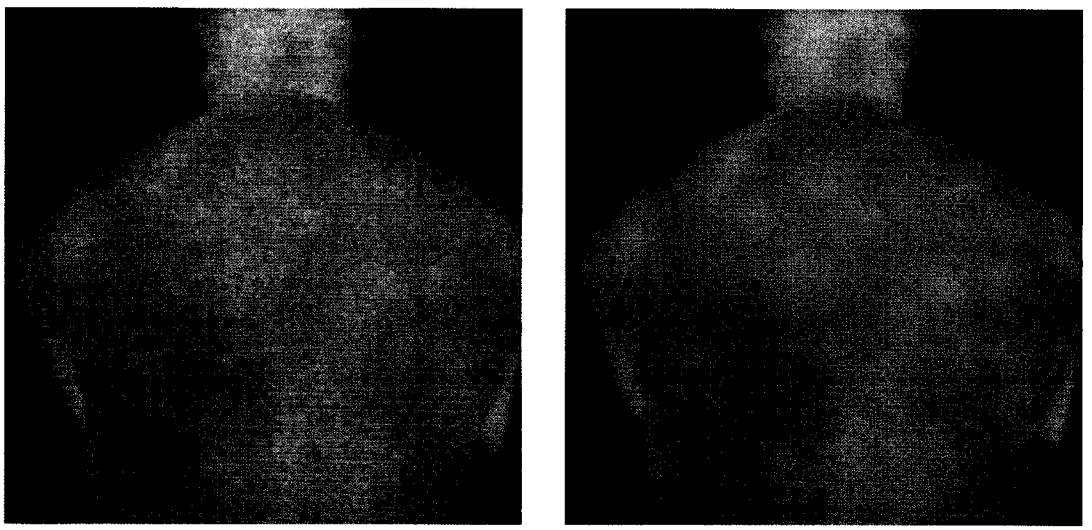
2.2 Infrared Image Processing

The following subsections describe the state-of-the-art in IR image processing. The review is extended somewhat outside of the range of clinical IR image processing since the literature on the topic is scarce, and some of the methods proposed in the broad IR image processing literature are of interest.

2.2.1 Preprocessing

The pre-processing of IR images is not intrinsically different from that of visible images. It involves correcting for possible artifacts created by the imaging system, including the removal of any noise. The choice of a noise removal method is linked to the type of imaging system used, its quality and the type of noise created. The noise

in modern IR systems is usually well characterized by a Gaussian distribution, but is not always additive. IR imaging systems based on photodetectors are better modelled by Poisson noise and appropriate noise removal procedures must be considered (for example using a squared wavelet transform as in [27]). In general, for the applications that are considered in this thesis (i.e. IR images of the human body in closed and controlled environment), the SNR is relatively high and noise removal is not a major issue. However, clinical images acquired with older technology have sometimes a much lower SNR as shown in Fig. 2.1.



(a) Original image (b) Denoised image

Figure 2.1: Example of the processing of a noisy thermal IR image of the upper-back, using a modified version of the Discrete Wavelet Squared transform method.

Figure 2.1 shows an example of processing of a noisy thermal infrared image of the upper-back. This image was taken using a first generation infrared camera with a photovoltaic sensor by M. Frize and her team from the pain clinic and the Moncton Hospital, NB. The noise is modelled as Poisson noise. The denoising method was a modified version of the Discrete Wavelet Squared Transform method [27] [47].

2.2.2 Segmentation

Of greater importance however is the removal of unwanted background in the images. This is the first stage of the segmentation process and is meant to isolate the regions of interest on the body. Note that for most of the applications considered in this thesis, the background is not considered complex. However, due to the nature of IR images, the extraction of the main ROI from the background can be challenging as the intensities from the background region(s) may be strongly overlapping with that of the region to be extracted. In other words, the body area considered might have a wide range of temperatures, which overlaps with the background range. Since the images produced are directly related to the temperature of the objects in the field of view of the camera[2], parts of the area of interest might be blended in the background, similar to a local variation of illumination in optical images. As a result, segmentation techniques relying only on the intensity levels (thresholding, standard clustering algorithms) do not perform well in general. Furthermore, methods developed for automated tracking of military targets [48], or for predictive maintenance [49] do not perform well when translated to other applications such as infrared imaging of the human body.

Combination of common segmentation techniques has been proposed in the IR imaging literature to overcome this issue. For instance, Wu *et al.* used a modified version of the standard Fuzzy C-Means (FCM) algorithm taking into account the membership values in a small neighbourhood of each pixel when updating the membership matrix [50]. They show good results on ship detection in Forward Looking

[2]This is a simplified explanation as the intensity level in the infrared images and therefore the temperature returned also depends on various physical properties of the object, such as its surface properties, its emissivity, etc.

Infrared (FLIR) images.

Ernisse *et al.* presented a segmentation method based on a hit-and-miss filter [51]. They thresholded the filtered image, making it binary, and added a locally thresholded version of the original image to produce a final binary image expected to show the location of the targets in the original FLIR image. The thresholds were optimized using a genetic algorithm. They added a derivative of Gaussian filter as an extra step to incorporate some of the shape information from the original image into the final segmented image, showing the potential location of targets. However, knowledge about the target was assumed to be extensive in designing the hit-and-miss filter, which limits the applicability of the method for other applications.

Zhou *et al.* investigated the use of level-set methods to extract boundary on thermal IR images [52]. An initial contour is grown adaptively using a speed function based on the edge/direction map. They show some applications of their techniques on different examples of thermal IR images, including medical IR images. However, they do not compare their results with those from other segmentation techniques, which may have worked just as well in the examples presented in their paper. They also mention that level-set methods may be susceptible to gaps in the expected contour of the object of interest, in which case the final contour might have "leaked" through the gaps. The technique is nevertheless interesting and promising for the segmentation of IR images.

Qi *et al.* [35] presented an automated asymmetry analysis of breast thermograms. Their approach used a Canny edge detector and a Hough transform to detect global edges of the image and breast boundaries. Then they compared the curvature of smoothed versions of the histograms of the two segmented breasts in order to deter-

mine the degree of asymmetry of the temperature distribution of the breasts. Ng *et al.* presented a method of extracting breast shapes from breast thermograms [36]. Gradients of thermograms were computed using the Robert cross-gradient operator in order to detect edges and boundaries. Gradient images were enhanced by contrast stretching and median filtering and then divided into 25 regions before being submitted to a thresholding rule. Both these papers, however, provide few implementation details and do not present extensive results. Scales *et al.* proposed a method that used a series of Canny edge detectors to detect the bottom contour of breasts by linking connected edge components and looking for the longest component that was not part of the outer body contour [34]. The top and side contours were estimated based on a set of empirical rules and the location of the armpit landmark points. They reported qualitative results for a small sample of IR breast images. Poor segmentation was usually attributed to the failure of the edge detection step. Blank and Kargel introduced a dual thresholding method to segment infrared images of the hand [53]. The thresholds were derived from user-selected rectangular regions. Backgrounds were kept uniform and noise-free. Morphological opening was used to isolate palmar regions from fingers, using two automatically chosen structural elements.

Finally, Chan and Pearce applied image processing techniques to enhance the visualization of subcutaneous vessels in thermograms after a stress test [54]. They compared the results for two different methods. The first method used a 9×9 median filter followed by a Frei-Chen edge detector and thresholding. The second method filtered the images with a rule-based adaptive window-size filter. A 9×9 median filtered version of the resulting images was subtracted from it and a root image derived by successive median filtering. The vascular patterns were then extracted by

thresholding.

Morphological processing has been used extensively in the segmentation literature, including in production inspection and other automated measurement systems. However, most methods proposed in the literature rely on texture information that is not present in infrared images. For instance, Lee and Wong introduced an approach based on skeleton operations, thresholding and morphological opening to segment synthetic visible images [55]. Daut and Zhao proposed a similar method to detect flaws in glass [56]. More recently, Mak *et al.* discussed the optimal design of morphological filter for fabric defect detection, based on a semi-supervised morphological processing of texture features extracted from visible images of textile [57]. An interesting approach was proposed by Scotti *et al.* for extracting blood cells on microscopic images [58]. Their approach used a Canny edge detector to extract the simple circular contour of the blood cells. Then morphological dilation, flood filling and erosion operations were performed to classify the red blood cells from other cells, based on the area of the objects.

2.2.3 Extraction of Anatomical Regions of Interest

The segmentation of the body region of interest is a first step that is crucial in focusing the subsequent processing on relevant areas. Once the segmented area of interest has been performed, two options are available for carrying out the thermal imaging analysis. The first one is to search for abnormal thermal patterns over the entire segmented region; the second one is to refine the segmentation further by identifying anatomically relevant regions and carry the analysis on these regions only. The first option is treated in the next section. The second option is explored in this section.

Note that the two options are not mutually exclusive and it is likely that they would complement each other.

The identification of relevant anatomical structures, as defined by a clinical expert or as warranted by the application, can be performed manually although this is impractical and time consuming for large databases. It can also be automated by localizing salient points or areas and reconstructing the structures of interest from there. The use of remarkable points to define anatomical structures is well documented in the medical field and in anthropometric studies (see [59] or [60] for example). Most anthropometric landmarks are linked to skeletal features, which are typically best found by palpation. These bony landmarks correlate with surface geometrical features albeit in a subtle way. A typical clinical examination by a family practitioner, for example, would likely involve locating bony landmarks or other surface markings in order to find a specific medical structure. An interesting and useful aspect of anatomical surface landmark points is that they are a stable set of points on every human body [61].

Previously proposed methods for locating landmarks in 2D images include methods relying on the detection of high curvature points and corners, methods relying on local energy models and methods relying on probabilistic models of the image and shape. Curvature based methods are popular to find important surface geometrical features indicative of bony landmarks, since information on the shape of a curve is mostly concentrated at points of highest curvature. For instance, Le Briquer *et al.* used the extrema of a modified curvature function proposed as a corner detector by Kitchen and Rosenfeld to detect salient points in brain MRI [62] [63]. The modified curvature is essentially a weighed curvature function using the magnitude of the

gradient of the intensity as weights. Curvature-based methods are related to corner detection techniques and the latter can be applied to locating anatomical landmarks as well. Robust corner detection algorithms include methods based on finding the peaks of local autocorrelation functions (estimated for a local neighbourhood around each pixel) as proposed by Moravec [64]. Improved variants based on different measures of "cornerness" using first and higher order differential operators have been proposed by Harris [65] and by Förstner [66]. Corner detection methods relying on differential operators tend to be sensitive to noise. In addition, they require the careful setting of a threshold on the "cornerness" measure to remove false corners (small extrema from the differential operation) and this can be difficult for a wide range of images depending on illumination conditions and contrast.

Local energy models for finding corners and edges postulate that important features in an image occur at a point where local frequency information is maximally in phase (the phase is then said to be congruent at that point) [67]. The phase congruency measure can be viewed as a function of the deviation of the phase from its mean. Local frequency information is obtained through Fourier transform coefficients or more frequently Gabor wavelets tuned to different spatial frequencies. An example of a corner and edge detector using phase congruency was proposed by Kovesi [67]. The author points out that the setting of thresholds is more readily done than with other corner detection methods through the statistics of the filter responses to the image [67].

Another interesting technique for finding high-curvature points is the scale-space curvature approach introduced by Mokhtarian and Mackworth [68], which calculates curvature zero-crossings of curves at multiple scales. They first extract the contours

using an edge detector and calculate the curvature at high scale. Corners are detected at maxima of the curvature and they are then tracked at lower scale to improve their localization. An example of a recent implementation is given in [69].

Another notable corner detector is the SUSAN (Smallest Univalue Segment Assimilating Nucleus) detector proposed by Smith and Brady [70]. It is based on brightness comparisons within a circular mask assuming that within a relatively small circular region, pixels belonging to a given object will have relatively uniform brightness. It is robust to noise and relatively fast but its repeatability is not as good as other popular corner detectors such as the Harris one.

Finally, probabilistic models can be found in [71], where the authors introduced a local image structure feature extractor at a range of scales and trained a statistical model for each feature in order to extract salient points. Active appearance models have also been used for matching a statistical appearance model of a shape to an image, after training the model from a set of shapes sampled with manually selected landmark points [72]. Active Appearance Model (AAM) is a powerful framework but seems better suited for full segmentation and tracking of objects rather than anatomical landmark extraction.

The goal of landmark identification is typically to recognize and/or segment the shape of an object (white matter from brain MRI for example) and/or to use the landmarks found for registering and warping the object in order to compare it with the same object in a different scenario. For example, finding the landmark points in a brain MRI allows the segmentation, registration and comparison of white and grey matter, possibly enabling full 3D reconstruction. It may also be used in a scenario where images from an MRI, Computed Tomography (CT) and other modalities are

fused together in a concise representation of the region of interest, often yielding additional information that would not have been easily extracted otherwise. In the case of clinical thermal IR imaging, it is especially useful when comparing contralateral sides of the body and assessing asymmetries, which as it has been mentioned already is an important aspect in the search for abnormal thermal patterns. Being able to match one region from the left side of the body to its equivalent on the right side of the body is extremely useful.

In spite of similarities, the problem is slightly different in the case of IR images since there is usually no specific visible anatomical structure inside the body region of interest. Rather, the anatomical regions have to be drawn out from landmark points located on the outside contour of the body region. In other words, the landmark detection cannot be performed on the actual contour of the anatomical ROI, e.g. white matter of the brain or a phalanx, since this contour is inexistent in most if not all cases.

One solution to this problem is to place thermally neutral reflective markers on the body, which act as fiducial markers and help locate key anatomical landmarks. This is usually a challenging and time-consuming task, is not always accurate, and can be uncomfortable for the subject. Having an expert locating landmarks on images is equally challenging and time consuming [73]. Another option is to fuse images of the subject from different imaging modalities, such as visible imaging or x-ray radiography. Acquiring an image of the same subject with a conventional camera simultaneously may reveal additional superficial details not visible on the IR image such as the creases of the hand, well defined contours, colour contrast and so on. An x-ray radiograph would further reveal bony details which can simplify greatly the

localization of anatomical landmarks [74]. Unfortunately, taking multiple x-rays is not practical for obvious reasons and the task of detecting landmarks on fused visible and infrared images is not trivial; this is in fact similar to our problem. A third option is to detect the landmarks automatically and use the found anatomical markers to reconstruct the regions of interest, by using anatomical templates for example.

2.2.4 Extraction of Other Regions of Interest

Further segmentation of the body region is difficult as there is little information about the anomalies, such as their shape, extent, etc. Most methods presented in the literature rely on knowledge about the target in order to find the ROIs. However, in the absence of information about the characteristics of anomalous regions, those methods are not directly applicable. On the other hand, some automated methods that do not use extensive prior information have been presented in the IR imaging literature, and more specifically the Automated Target Detection (ATD) problem in FLIR military images. For instance, Khan and Alam used some morphological pre-processing operations on FLIR images and obtained potential regions of interest using maxima of the resulting image [75]. Some *a priori* information about the target size was used in both the pre-processing step and in the subsequent thresholding.

Casasent *et al.* claimed that no one detection technique could provide the high detection rate, low false positives needed for ATD on FLIR military images and therefore introduced a method based on the fusion of several detection techniques [76]. They used a nonlinear combination of morphological closing/opening functions and edge detection based on wavelet filtering (using Gabor filters) to produce a final image where potential targets have higher intensities. The high intensity regions are then

sorted by successively lowering a threshold and removing an ellipse of pre-specified size (based on the known target size) to avoid large regions at lower thresholds. Threshold parameters were determined empirically and *a priori* information about the target was necessary in the intensity peak sorting algorithm. The method was applied successfully to the detection of tanks in FLIR images.

A more comprehensive system was proposed by Greenberg *et al.* [77]. They located the centres of potential ROIs by thresholding a feature image obtained by locally filtering a FLIR image with a standard deviation filter [77]. The thresholding was based on a complexity measure applied to the standard-deviation map. The centre candidates were then pruned using a graph-based approach on the local neighbourhood in the standard-deviation map. Close centres were then aggregated with a simple hierarchical clustering approach. Finally, the actual clusters were found from the centres using local thresholding. The method was tested on satellite IR images and produced good results when attempting to detect targets. Some parameters used in their paper, however, required a rough estimation of the number and size of the targets. Also, no extensive testing of the final step of their method (segmentation of the actual targets) was reported, which would be important for medical applications.

Privitera and Stark presented an interesting approach to compare the performance of different algorithms aimed at detecting regions of interest in visible images [78]. The algorithms consisted of 6 different filters producing 6 distinct types of feature images (symmetry transform, difference in grey-level orientation, edges per unit area, centre-surround receptive fields, entropy and contrast). Local maxima from each filtered image were then clustered to form small ROIs, where the number of clusters was chosen to be 9. The authors compared similarities between the ROIs produced

by the various algorithms. The goal of the paper being to ultimately model the eye-fixation movements when looking at images, the authors stopped short of actually combining the different feature images to evaluate the performance of a fusion of multiple feature images.

2.2.5 Choice of Features

The choice of the features in a classification paradigm is critical for obvious reasons. One can distinguish several categories of features based on what the features are trying to represent. The first and most commonly used are statistical features based on first- or higher-order statistics extracted from the histogram of an image or its 2D equivalent, the co-occurrence matrix. Most papers on thermal imaging of the human body use simple first-order statistical features for discriminating between normal and abnormal patterns. These statistical features typically consist of mean (or median), maximum, minimum, standard deviation or variance, and to a lesser extent skewness and kurtosis, over the entire image or over a selected region of interest, or as a difference between two symmetrical regions. The reason behind the choice of features for a given application is usually not clear and little work has been done on comparing the discriminating power of the chosen features for a given application.

Researchers working on target detection and recognition in IR images prefer local first-order statistical features, which are statistics computed locally over some neighbourhood of a pixel. The result may be presented as a feature map of the same size as the original image but containing the local statistics for each pixel over a specific neighbourhood. For instance, Greenberg *et al.* used the standard deviation feature map over a 5×5 neighbourhood to extract targets in military applications [77]. We-

ber on the other hand used 5 local features (maximum, contrast, average gradient, local variation) over rectangles of various sizes centred on each pixel [79]. All features were then combined in one index for each pixel yielding what the author calls a confidence map, which is then thresholded to provide the potential targets. In addition to first order statistics, several other potential features have been considered in the IR imaging literature. For instance, the entropy is often used to capture some of the textural information of thermal images [77]. Fractals and fractal dimension were also investigated [80] [81].

The second class of features is concerned with characterizing the shape of particular regions. This includes various properties of the region of interest such as distance from centroid to boundary, perimeter, area, eccentricity, orientation, bounding box or polygon, etc. [82–84]. This is especially useful if some shape information is available for the application considered. Another popular way to characterize the shape of a region is by using moment invariants such as geometric, Zernike, Tchebichef or Legendre moments. Most moments are (made) invariant to translation, rotation, scaling and/or contrast [85] [86]. Moment invariants, or a combination of these, can present in a concise manner the discriminative information about an image. Zhu *et al.* considered well-known moment invariants of thermal medical images, arguing that they characterize images fairly well. They used the first seven moments [86].

The third class of features consists of multiscale representations (such as wavelet or Gabor extensions). Multiscale representations have also been proposed either to differentiate between target from background objects in automated target detection and/or recognition in IR images [75]. Multiscale representations can capture localized changes at multiple scales and spatial frequency. Khan and Alam used Daubechies

wavelet features to classify objects as target or background in cluttered FLIR imagery [75]. Morphological operations were used to detect potential targets.

2.2.6 Summary of Infrared Image Processing

The previous sections have covered the state-of-the-art techniques used in the processing of IR images. Segmentation techniques for IR imaging proposed in the literature have several shortcomings. Some techniques such as FCM [50] or thresholding and hit-and-miss operator [51] work well for finding the location of a specific shape and recovering that approximate shape, but they typically fail to fully segment a shape. This is necessary in clinical applications of IR imaging, where potential abnormal regions may occur at any location within the body region being imaged. Level-set active contour methods [52] showed promising results but are still unable to recover full contours when edges are weak and contrast is poor. Other segmentation techniques [53] require user-inputed data and are therefore not relevant for automated analysis. Finally, methods relying solely on morphological processing were presented that perform well for different imaging modality and on simple contours and backgrounds, but they are not adapted to more complex contours and backgrounds. A robust segmentation approach that recovers the full contour or objects in IR images is still missing.

Extraction of ROIs in clinical IR images is difficult due in part to the smooth temperature variations, the lack of visible anatomical landmarks within a body region or the lack of information about the potentially interesting regions. Extraction of anatomical ROIs requires anatomical landmark points in general. Several methods of extracting salient points have been reviewed in Section 2.2.3. They all tend to produce

a large number of landmark points when the contours are very smooth. Also, finding the right combination of parameters for each method is not always straightforward and the optimal parameter value can vary widely from one image to another. An algorithm for selecting only key anatomical landmark points is needed in order to produce specific anatomical ROIs. For blind search of potentially abnormal ROIs in IR images, several approaches were investigated in Section 2.2.4. Although they identified key concepts such as fusion of several basic techniques and the combination of feature maps, they either required knowledge about the target [77] [87] or they focused on centroids of regions instead of full regions [78].

2.3 Anomaly Detection in Infrared Imaging

In the next few sections the state-of-the-art in anomaly detection is presented in an IR imaging context. The literature abounds in papers on the different techniques used for ATD, where a known target is sought after in challenging backgrounds, often changing over time. The automated target detection and recognition problem is difficult in many respects, for instance when comparing a large number of target signatures in real-time. However, the target detection in itself is greatly simplified by the assumption that extensive knowledge about the target available, through large databases of target signatures for example, or knowledge of the shape of the target [88]. Anomaly detection is usually the first stage in an automated target detection and recognition system. It is also a critical one, in which regions containing potential anomalies or unidentified areas are flagged for further analysis.

Another issue with anomaly detection is the availability of training data and labels. Most approaches work by first modelling normal data and then, using some similarity,

distance or other measure against the normal data, they determine whether or not the new data should be deemed abnormal or normal. Having labelled training data allows for good determination of the normal versus abnormal situations and should result in better classification. On the other hand, when a set of unlabelled data is presented, or when no training data is available, the normal class needs to be estimated reliably in order to obtain good classification results.

Anomaly detection techniques can be found in a wide range of applications, from computer network intrusion to fault detection in machines, and character recognition. Many of the methods developed for non-imaging applications such as network intrusion detection are not readily applicable to imaging problems due to their reliance on sequential and time series analysis for example. For completeness, the following sections cover the main techniques for anomaly detection in an optical and IR imaging context.

2.3.1 Statistical Techniques

Statistical novelty detection techniques model data based on statistical properties and then test a new sample to check whether or not it belongs to the estimated model. Modelling data leads to estimating the probability density function of the underlying phenomenon based on data samples. The accuracy of the estimate depends not only on the number of samples available but also on the distribution of the data in the input space. Low density regions of the input space may be difficult to model accurately, which leads to large errors when setting thresholds between normal and abnormal regions unless great care is taken.

Statistical methods are usually divided into parametric and non-parametric approaches. Parametric approaches assume that the data come from a family of known

distributions, for example a Gaussian distribution, and attempt to best estimate the parameters of the assumed distribution. In the ideal case of a known distribution for the data and Bayes classifier, a threshold can be determined that optimizes the trade-off between error rate (probability of erroneous classification) and reject rate (probability of rejecting a sample pattern, i.e., the pattern does not belong to any of the classes). However, it is rare in practical applications that the underlying distribution is known and therefore the choice of a Gaussian distribution may lead to poor results.

Few parametric methods are used in practice for novelty of anomaly detection due to the lack of flexibility of the model once a family of distributions has been chosen. If little or no prior knowledge about the true density is available, there is a great chance that the resulting probability density function estimate will be poor. On the other hand, semi-parametric methods do not restrict themselves to specific functional forms but still allow the size of the model to grow with the complexity of the problem and not simply with the size of the data set [89]. Mixture modelling, and especially Gaussian Mixture Modelling (GMM), is an example of semi-parametric techniques that has been used successfully for various applications. The number of basis functions in the mixture can be treated as a parameter of the model and estimated along with other parameters of the distributions. This leads to models where the number of mixture components is much less that the number of training samples, as opposed to some non-parametric techniques such as Parzen window density estimation.

GMM approaches use Gaussian kernels to estimate probability density functions of patterns in a training set. The number of kernels is less than the number of patterns in the training set, but if the dimensionality of the data is large, a large number

of samples will be required to train the model reliably [90]. GMM is an attractive model since it has a universal approximation property in theory and is analytically simple [91]. The main issue with GMM is setting the appropriate number of kernels as well as the thresholds between classes. Roberts and Tarassenko [92] proposed a method to select automatically the number of kernels using a growth criterion based on the smallest Mahalanobis distance between training sample and each Gaussian kernel. One solution to the thresholding problem for low-density input space regions was addressed by Tarassenko in [93]. He used local novelty thresholds depending on the density of the data in the region of the input space. The space is partitioned using the k-means algorithm and an estimate of the probability density function for each region was calculated using GMM. The algorithm was applied with some success to mammograms. Roberts introduced the extreme value theory for novelty detection, which statistically models abnormally low or high values in the tails of distributions. The density estimation is performed by a GMM approach where the number of kernels is determined automatically based on penalized likelihood methods [91] [94]. The authors applied their technique to several biomedical datasets including a small number of MRI images. Outliers can make the data density estimation erroneous and some papers address these issues by incorporating an estimate of the probability density function of outliers in the model [95].

Nonparametric methods can be used with arbitrary distributions and do not make any assumption about the form of the underlying densities to estimate [96]. However, they tend to be extremely slow to converge and require more data samples to get accurate estimates of the underlying density. Common nonparametric techniques include Parzen windows, k-nearest-neighbour and some fuzzy classifiers.

2.3.2 Artificial Neural Network (ANNs) Techniques

ANNs have been used extensively for anomaly detection as they typically require less parameters to be set and fewer *a priori* assumptions about the data are necessary [90]. Training can be done in a supervised or unsupervised manner depending on the architecture considered and the availability of training data.

2.3.2.1 Multilayer Perceptrons

MLPs are the most widely used class of neural networks. They consist of an input and an output layer of neurons and one or several layers of interconnected hidden units that effectively create complex mappings between inputs and outputs. MLP rely on hidden units which compute a non-linear function of the scalar product of the input vector and a weight vector [89]. The outputs of an MLP network for a one-hidden layer network may formally be computed as (using Bishop's notations):

$$y_k = \widetilde{g}\left(\sum_{j=0}^{M} w_{kj}^{(2)} g\left(\sum_{i=0}^{d} w_{ji}^{(1)} x_i\right)\right) \tag{2.7}$$

where y_k are the outputs, g and $\widetilde{g}$ are activation functions, $w^{(1,2)}$ are the weights between the input units and first layer of hidden units and between the first and second layer of hidden units respectively, and x_i are the input units. This formula can be extended to a greater number of layers if needed. Training the network means finding the weights that provide the best mapping between the input and actual outputs.

Training of MLP is typically done in a supervised manner where target vectors of known classes are presented to the network. Various error back propagation train-

ing procedures exist to train an MLP network. MLP are powerful and flexible tools for finding non-linear mappings between inputs and outputs. Hidden units act as localized representations of the input space and the linear superposition of these localized functions have universal approximation capabilities. As a result, MLP are able to represent complex problems, such as classification or regression. For novelty detection, however, as Markou *et al.* mentioned in their review of novelty detection techniques, they do not generate closed-class boundaries [90]. In other words, a standard MLP will separate two classes with a hyperplane as opposed to a Radial Basis Function (RBF) network for instance, which attempts to fit a Gaussian function around each class. This may be the cause of serious misclassification problems when novel patterns are presented to the network since they will likely be classified as one of the existing classes. Nevertheless, many papers have dealt with the application of MLP to anomaly detection. For example, Bishop applied Gaussian mixture modelling to estimate the density of the training data and then used the estimate to assess the degree of novelty of an input when passed through the network, which in turn gives an estimate of the validity of a network's prediction [97]. Another simple approach is to threshold the output of the network, assuming that low confidence translates into novelty [98]. Multiple thresholds may also be used on the first few winning nodes [99] [100]. Vasconcelos *et al.* suggested a solution to the problem of open class boundaries generated by MLP networks [101]. The authors used the patterns rejected by the network with high confidence as negative training patterns. This arguably reinforces the decision of the networks and worked better than randomly generating negative training patterns. Vasconcelos *et al.* also considered Gaussian Multilayer Perceptron (GMLP), where the sigmoid activation function, in part responsible for

the open class boundary problem, is replaced by a modified Gaussian activation function [101]. GMLPs tend to create more confined regions around the training data in the pattern space. The authors concluded in favour of a third alternative, namely a RBF network, although they omitted to include real data experiments to justify their theoretical claim. A more elaborate technique based on the same idea of using the rejects to retrain the network was presented by Singh and Markou [102]. The authors used fuzzy clustering on the novel (rejected) samples and the clusters are compared to known class to check whether they are outliers or novel patterns. Novel patterns are then manually labelled and the new class is added for the network training. Their method was developed for natural video scene analysis and showed good results in identifying novel objects in sequences (under controlled environmental conditions). Another solution to the drawbacks of MLP was presented by Wilson *et al.* [103], who used regularization methods and sinusoidal activation functions to improve the network training and dynamics. When tested on classification of hand-printed digits and fingerprints, the authors claimed that their methods produced better error-reject performance and smaller neural networks.

2.3.2.2 Radial Basis Function Networks

Radial basis function networks are based on hidden units (called basis functions) which compute the distance between an input and a prototype vector. Gaussian functions are typically used as basis functions because of their smoothness and localized properties. A major difference between MLP and RBF networks is the way they represent (or divide) the input space. Each hidden unit in an MLP corresponds to a hyperplane separating two regions of the input space, which leads to the open bound-

aries problem mentioned previously. In addition, many hidden units contribute to an output value, the process being highly non-linear with potential problems of local minima and therefore possible convergence issues. This often yields complex network architectures whose inner working is not easily understandable. RBF networks on the other hand divide the input space using hyperspheres (with closed boundaries) and produce local representations of the input space since typically only a few hidden units (or basis functions) are activated for a given input vector. Finally, RBF networks typically have a simple architecture consisting of two layers, a Gaussian layer and a linear layer to the output. Training of RBF networks uses a mix of supervised and unsupervised techniques, where the basis functions are determined using unsupervised techniques (for example clustering) and the second layer is optimized with supervised techniques. RBF neural networks are related to nearest-neighbour classifier, Parzen kernel density estimation (or Probabilistic Neural Network (PNN)), GMM and kernel methods in general as they attempt to fit a kernel function, such as a radial basis function or a Gaussian function to input data points. Examples of simple RBF networks applied to anomaly detection can be found in Brotherton and Johnston's paper [104], where the authors presented an interesting discussion of the application of RBF to anomaly detection for military aircraft. They described a RBF network, deemed class-dependent RBF network, where groups of hidden units are associated with a single class, using k-means clustering to cluster the type of input data. The output is a weighted sum of the output of each basis unit and the weights are determined with a variant of least-mean squares algorithm. The authors used Gaussian basis functions but mentioned that other basis functions such as Rayleigh distribution or mixtures of basis functions can also be used depending on the data.

The anomaly detection is carried out by simply assessing whether the input pattern falls into any of the basis units, in terms of Mahalanobis distance. One or several thresholds on the distances can be used for the classification [105] [106].

Other variations of RBF networks have been proposed. Albrecht *et al.* [107] introduced a generalized radial basis function network with self-organizing properties by adding reverse connections from the output layer to the hidden layer. Their network is equivalent to a Bayesian Maximum Likelihood Classifier and uses globally normalized activation functions with a small cut-off parameter to bound the patterns producing a non-zero activation to a specific region. Their neural network also groups neurons of the central layer together that belong to the same class. The anomaly detection is carried-out by looking at small (vanishing) responses of the output. Their method was tested on speech recognition with mixed success [107]. Roberts and Penny looked at committees of RBF classifiers to alleviate the problem of local minima. This effectively allows initializing the networks with different random weights, thereby producing different sets of optima (or minima in the input space) [108]. This in turn allowed them to derive error and confidence values on the classification.

PNN implement Parzen window density estimation and therefore exhibit the same advantages/disadvantages. PNN are fast to train and retrain since they only require one pass through the training set. However, they have high memory requirements, which may be limiting. They can be implemented as a radial basis network in which the second layer is a competitive layer.

2.3.2.3 Support Vector Machines (SVMs)

SVMs were originally proposed by Vapnik [109]. SVMs map input patterns onto a high(er) dimension space where patterns can be separated by an optimal hyperplane. The mapping is achieved through a kernel function, usually non-linear. Typical kernel functions are polynomial and Gaussian. The training of SVMs consists of finding the optimal hyperplane that maximizes the margin between two (or more) classes. The vectors that are at the maximum margin of the optimal hyperplane are called support vectors. SVM classifiers are completely characterized by the support vectors, which greatly reduces their complexity and makes them less prone to over fitting issues [96]. Note that the support vector description formalism can be viewed as a generalized classifier with several other classifiers as special case. For example, if the kernel is chosen to be Gaussian, the SVM classifier created is essentially equivalent to an RBF neural network. The difference however is that the number of centres is determined automatically and only the width of the Gaussian needs be adjusted. This usually results in better performance than classically trained RBF networks [110]. The original SVM method proposed by Vapnik was a binary classification problem (two classes). For anomaly detection (when only one class is available), the objective function and constraints of the optimization problem need to be adapted. Two main techniques were introduced. The first one tries to find the hyperplane in the feature space that represents the maximal margin separating the origin of the space from most of the data [111]. In other words, this method attempts to separate the region containing the data from the region containing no data. A parameter C controls the number of outliers allowed (soft margin). The second technique consists of finding a hypersphere with minimal radius that fits most of the data [112]. A parameter C

controls the trade off between the volume of the sphere and the number of outliers. For both techniques, once the optimal hypersphere or hyperplane has been determined, any input data outside of the region delimited by the hypersphere (or hyperplane) is considered novel. The choice of the kernel depends on the data but is often chosen to be Gaussian or polynomial. However, kernels tailored to specific applications (e.g. images) may provide better performance as suggested by Barla *et al.* [113]. Note that kernel parameter(s), for example, the width of the Gaussian function, greatly affects the performance of the classifier. There is no general method for setting them but they can be optimized by validation process (cross-validation, leave-one-out, bootstrap methods). The two basic SVM methods for anomaly detection have several drawbacks. For example, Tax *et al.*'s method requires a large dataset and does not perform well when large differences in density exist. Creating artificial outliers around the target class improves the performance results if a tight representation of the target data is available [112]. Otherwise it does not scale well to high-dimensional data. On the other hand, Sch olkopf *et al.*'s method relies too heavily on the origin of the feature space, which effectively assumes that abnormal instances will lie at the origin [110] [111]. Also, the method is extremely sensitive to the choice of kernel parameters. Several modifications addressing some of the drawbacks of basic methods have been proposed in the literature. Campbell *et al.* suggested focusing on the centre of the data distribution rather than the origin [114]. Manevitz *et al.* proposed instead to consider all data points close enough to the origin as outliers or abnormal data. This was achieved by thresholding the feature vectors [115].

Applications of SVM-based anomaly detection methods include fault detection, handwritten digits recognition, document classification and network intrusion. A

combination of supervised Principal Components Analysis (PCA) and novelty detection SVM methods was applied to object detection in 3-band medium wavelength infrared images for video surveillance applications in [116]. Chapelle *et al.* applied SVM classifiers with generalized Gaussian, Laplacian and sub-linear kernels to colour image classification based on colour histogram features and obtained very good performance on a large database of different categories of colour images [117].

2.3.2.4 Self-Organizing Networks

The last group of neural networks used for anomaly detection are self-organizing networks, which typically use unsupervised learning algorithms. They include self-organizing maps and adaptive resonance theory. Self-Organizing Map (SOM) may be described as a nonlinear, ordered and smooth mapping of high-dimensional input data domains onto the elements of a regular, low-dimensional array [118]. They are a common technique of unsupervised learning that is an alternative to statistical clustering algorithms. One of the important features of SOM networks is that they preserve the topological information to a certain extent. Indeed, the SOM algorithm performs competitive learning but updates a neighbourhood of the winning node as well. Hence over time, nodes that are close to each other in the low-dimensional array respond to similar inputs and the set of local neighbourhoods self-organize to produce a global ordering [119].

SOM were developed early on as a novelty filter with or without memory (with or without forgetting capabilities). SOM networks have been applied to several anomaly detection problems, in particular fault detection and monitoring in industrial setting [120–122] and network intrusion detection [123]. After training, most approaches

threshold the distance between a test pattern and the activated neuron for novelty (a large distance indicates that the sample is far from all known clusters and is therefore deemed novel) or they keep track of the activated neurons on the basis that novel/abnormal inputs will activate different neurons than normal inputs.

A natural extension to SOM networks as an attempt to better model the brain mechanisms of learning is the process of habituation. Habituation is a reduction in the strength of a response to a stimulus when the stimulus is seen repeatedly without any ill effects [119]. It can be interpreted as a simple novelty filter, since the focus of attention is shifted away from features that are seen often to concentrate on other features that are unusual. Marsland *et al.* added the concept of habituation to self-organizing maps and applied it to anomaly detection for a mobile robot application [124]. Extending their previous work, Marsland *et al.* presented a new self-organizing neural network for novelty detection called the Grow When Required (GWR) network [125] [126]. The GWR network is a self-organizing network that keeps track of how many times a node has fired with an exponentially decreasing variable between 1 and 0. Nodes that fire frequently are trained less than others. If a node has never fired or fires very infrequently, then the input is deemed abnormal.

Other self-organizing networks based on similar principles have been investigated. ART networks attempt to create stable memories from inputs using match-based learning (Match-based learning is similar to template matching in image processing applications) [119]. They are based on a leader-follower strategy where the node (or cluster centre) closest to the input pattern is changed to be somewhat more like the pattern. However, if the input pattern is very different from all the existing nodes, a

new node is added to the network to represent the new cluster. ART networks consist schematically of two layers, the layer of cluster centres and a layer that receives the input data, feedback from clusters and a time-varying bias signal (to keep a constant level of activity in the layer). The comparison between the signal from the feedback connections and the inputs is thresholded with a parameter called vigilance that determines whether to assign the input to an existing cluster or to create a new one. The convergence of the feedback connections over time is called a resonance. Fuzzy ART-MAP and ART-MAP-FD (ART-MAP with Familiarity Discrimination) are supervised variants of ART networks where each weight vector creates a hyper-box around all the input vectors belonging to a class. Anomaly is based on a thresholded familiarity function, which determines whether a pattern belongs to one of the hyper-boxes defined by weight vectors. Setting the threshold for the familiarity function is crucial for good performance. This technique was applied to simulated radar target recognition [127].

2.3.3 Dimension Reduction Methods

Dimension reduction methods attempt to reduce the dimensionality of the input space by extracting the most important discriminatory components. Various algorithms are available that attempt to extract linear or non-linear components from the input patterns in the input space directly, or in a space of higher-dimension, accessible through kernel mapping, where the extraction process is easier. These methods have been tested on various databases (University of Manchester Institute of Science and Technology face database, University of California Irvine (UCI) repository of machine learning databases, United States Postal Service (USPS) handwritten digit

database and Olivetti-Oracle Research Lab face database) with various degrees of success. Among notable papers are Zhang *et al.*'s paper [128] where the authors propose a kernel auto-associator technique consisting of a mapping to a higher-dimensional space with a Gaussian kernel and a class-specific reverse mapping with linear or polynomial functions. Their method is shown to perform as well as or better than various novelty detection techniques (auto-encoder, SOM, MLP, SVM, K-Nearest-Neighbour (KNN)) on several databases (Promoter DNA, Sonar data, faces). An example of a simple auto-encoder (which is in essence an implementation of PCA) used to detect anomalies can be found in [129], where the method is applied to fault detection in helicopter gearbox data and two datasets from the UCI repository. Similar auto-associative strategies have been used for fault detection or structural damage detection [130] [131]. Auto-associators have also been applied to document classification problems and compared favourably with techniques like nearest-neighbour and naive Bayes [132]. Auto-associative methods require a careful setting of the novelty threshold in order to obtain acceptable performance; in most cases, the threshold selection procedure will depend heavily on the quality of the training samples.

Another approach proposed by Tax and Duin [133] combines the outputs of several multi-class linear classifiers and assesses the variation between the outputs for various training sets. A large variation in the outputs indicates that a sample is hard to classify and is likely novel. Ko *et al.* proposed a three stage architecture using 5 auto-associator networks followed by a probabilistic neural network and a back-propagation MLP to monitor structural damage in bridges [134].

2.3.4 Summary of Anomaly Detection Techniques

In the previous sections the state-of-the-art in anomaly detection for a variety of applications was presented. Few methods dealt specifically with infrared imaging and none with infrared imaging of the human body. However, it is more appropriate to put the problematic of this thesis into a more general context, in order to highlight the potential of each basic approach (statistical parametric and non-parametric approaches, MLP, RBF, SVM and unsupervised networks). In order to do so, various techniques derived from each basic approach were described as well as their application to problems such as industrial fault detections, network intrusion or fingerprint and handwritten character recognition. It is unlikely that those methods would translate well to IR image processing applications, since the specificities of the problems considered are quite different. Anomaly detection techniques developed for ATD are also limited by the fact that they assume extensive knowledge of the target. In the case of clinical thermal IR imaging, abnormal thermal patterns for all but a few pathologies have not been well characterized and classified as of yet. As a result, supervised approaches on their own will likely fail.

From the previous review, the most promising approaches are based on multilayer perceptron networks, support vector machines and self-organizing networks. RBF networks and several other networks can be viewed as special cases of support vector machines and therefore need not be included. Using the key advantages of these three approaches, it is not unreasonable to think that a smart combination of the different networks might achieve better overall classification rate than each method on its own.

Chapter 3

Processing Architecture

This thesis proposes a general framework for automated detection of anomalies in clinical IR images. Within that framework, this thesis addresses two fundamental problems: the segmentation of a main object from the background; and the extraction of relevant, potentially abnormal, ROIs.

3.1 Overview of Anomaly Detection Architecture

An overview of the anomaly detection architecture is presented in Fig. 3.1. Note that the entire architecture is shown here but the focus of this thesis is on the segmentation and the ROI extraction blocks only.

The IR anomaly detection architecture can be summarized as follows:

1. A preprocessing block is used to parse the images according to the body regions they represent. This step is currently done manually. Also in this block, denoising is performed if necessary.

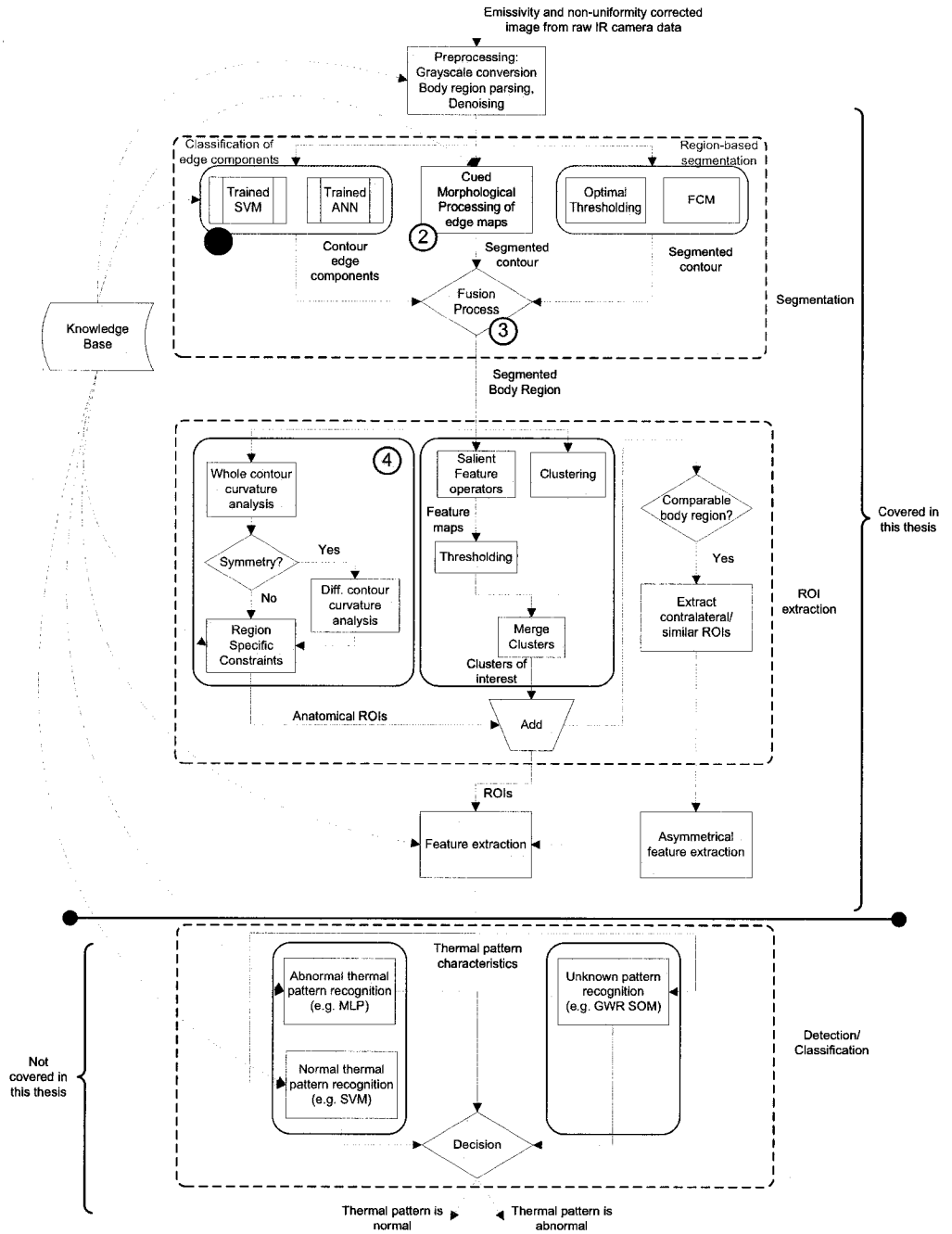


Figure 3.1: Overview of IR anomaly detection architecture. The circled numbers indicate where the major contributions are for this thesis.

2. Once the body region has been identified and noise removed, the image is sent to the segmentation block. The segmentation strategy is based on a fusion of three different types of segmentation techniques:

   - Standard region-based techniques: optimal thresholding and FCM clustering based method. These methods focus on general regions whose grey-level intensities are similar.
   - A novel approach based on the classification of edge components by a classifier. In this work, both a MLP neural network and a SVM classifier were implemented. This approach determines whether an edge component belongs to the contour or not. It is therefore purely edge-based and produces contour edge components.
   - A novel approach based on the morphological processing of edge maps using *a priori* local and global information; This method is a hybrid between region-based and edge-based techniques. It uses region-based information to process the edge components.

   Although each segmentation technique might produce reasonably good results on its own, it is expected that the fusion of all techniques might overcome the shortcomings of each one. It is also expected that the difference between each technique will benefit the fusion process. Several fusion strategies are introduced and evaluated. The resulting image contains only the body region to be analyzed.

3. Potential ROIs (within the main body region) are then extracted. If prior information about the application is available, anatomical ROIs may be deter-

mined. A novel approach is introduced to detect anatomical landmarks on the contour of the body region extracted in the previous block. It relies on a curvature analysis of the contour, region specific constraints and the symmetry of body regions to sort through the potential landmark sites and return a few key anatomical landmarks. From these landmarks, anatomical ROIs are extracted, using anatomical templates.

On the other hand, if no prior information about the application is available, gradient-based, intensity based and entropy-based operators are used to filter the body region and highlight specific characteristics. The resulting feature map is thresholded to return a set of clusters of interest. In parallel, a clustering algorithm is applied to the segmented body region, yielding another set of clusters of interest within the region considered. The two sets of clusters are then merged together to produce the final clusters of interest.

If more than one region is available, for instance contralateral regions or the same body region from a different image, then comparable anatomical regions or clusters of interest should be extracted as well from the other image or region. The extraction of comparable regions is facilitated by the key landmark points previously found. All regions and clusters of interest are pooled together.

4. This step consists of extracting relevant features from the ROIs produced by the previous stage, either anatomical ROIs or clusters of interest. The number and types of features used may be driven by additional information available on the application. For instance, difference statistics or similarity measures between contralateral sides within or between images can be computed if one is looking at asymmetrical trends between one arm and the other arm. Post-processing

of the array of features in the form of feature selection or principal components analysis is application dependent.

5. The final stage consists of classifying each thermal pattern (possibly comprising several ROIs) as normal or abnormal based on the previous features. Based on the amount of *a priori* information available about the application, the classification will likely differ, being either supervised or mostly unsupervised.

The knowledge base block in Fig. 3.1 refers to *a priori* expert knowledge that is available before processing, acquired during processing or extracted after processing. It may consist of:

- Knowledge from the acquisition of the thermal IR images, for instance the type of view and body region or information about the type of noise expected.
- Database of labelled edge patterns for the classification of edge components.
- Anatomical landmark information for specific applications.
- Features relevant to specific applications.
- Database of labelled thermal patterns.

Although not indicated on the diagram of Fig. 3.1, the knowledge base is ideally updated to incorporate information gained through the processing and final classification stage.

The next sections give more details about each processing block of the IR anomaly detection architecture.

3.2 Preprocessing

Pre-processing of thermal images involves correcting for possible artifacts created by the imaging system, including the removal of any noise. The choice of a noise removal method is linked to the type of imaging system used, its quality and the type of noise created. The noise in modern IR systems is usually well characterized by a Gaussian distribution, but is not always additive. IR imaging systems based on photo-detectors are better modelled by Poisson noise and adapted noise removal procedures must be considered [27] [47]. In the application considered here (i.e. IR images of the human body in a closed and controlled environment), the signal to noise ratio was typically high enough so that no extra processing was required on most images. Noisy images from the first generation camera were denoised using a stationary wavelet transform method that kept fine details in the image. The choice of wavelet basis, a (9,7) biorthogonal spline wavelet, was made based on earlier research [135]. This wavelet basis has four vanishing moments (for both the wavelet and scaling functions), a compact support, and the decomposition/reconstruction filters generated from it are symmetrical and have linear phase, which ensures minimal distortion in the resulting denoised image [136–138].

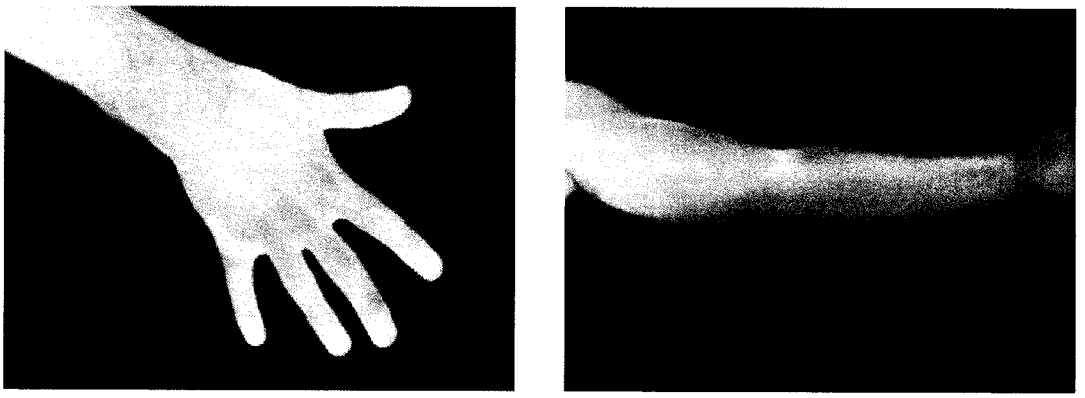
3.3 Segmentation

The typical background of clinical IR images is relatively simple with few objects present. However, the intensities of the objects of interest and that of the background regions may overlap, for example when imaging a person with cold fingers at ambient temperature or when one wants to extract the arm from a side view of the body. Also, the background intensity may not be constant across the image,

due to variations in the emissivity of background objects. This is especially true for images of the extremities (hands and feet), for which temperature variations between one end to another can be significant ($> 10\,^{\circ}\mathrm{C}$) even under normal environmental conditions. As a result, parts of these body regions (e.g. the fingers, toes) may be at the same temperature as the background temperature, which complicates greatly their segmentation. Thresholding techniques (global, locally adaptive) and clustering methods (for example fuzzy C-means), typically do not perform very well on their own in these situations. They often miss important regions of the object of interest (e.g. whole fingers in poorly contrasted images) or produce larger than expected regions where all fine details are lost. For the difficult regions of the extremities, it seems unavoidable to include some *a priori* information about the shape of the region of interest. This additional shape information allows retaining fine details of the region while removing irrelevant information. For example, if it is known that the main region is a hand, expected shape information can be included in the segmentation strategy to help recover missing areas (such as a cold phalanx).

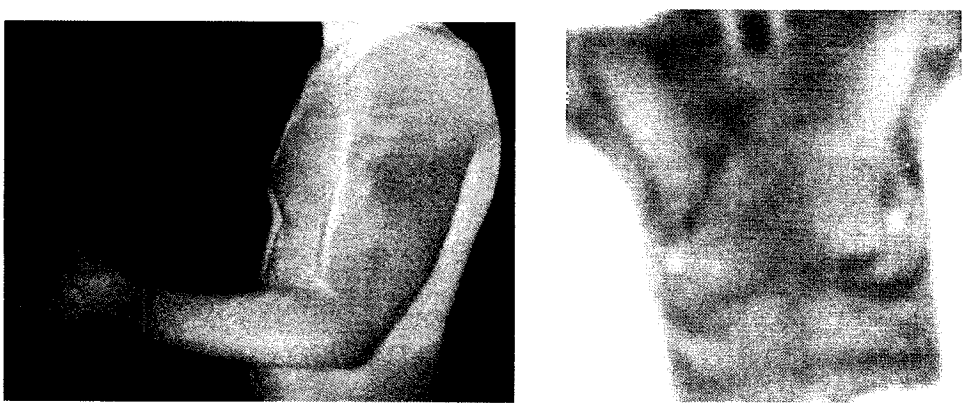
There are two main types of regions to be segmented. The first one is a unique region against a background, such as an IR image of the upper-arm for example. The second type of region to segment is a limited part of a body region, such as a breast or a side view of the shoulder. For the latter type, the segmentation is complicated by the fact that there is no obvious difference in terms of temperature between the body regions one wants to isolate and the rest of the body. For instance, if the whole arm is required from a side view, as in Fig. 3.3 a) the problem of finding the contour of the arm at the shoulder is not trivial. The region boundaries in this case are also somewhat subjective unless an extensive set of markers are placed on the image for

instance. Again, *a priori* information seems necessary to help recover the wanted region of interest in this case. Figs. 3.2 and 3.3 show some examples of the first and second types of regions to segment.



(a) Infrared image of a hand (b) Infrared image of an upper-arm

Figure 3.2: Examples of the first type of segmentation: hand and upper-arm.



(a) Shoulder view (b) Breast thermogram

Figure 3.3: Examples of the second type of segmentation: breast thermogram and side view of the arm.

A general strategy is introduced for segmenting the main region of interest from IR images. It consists of a combination of three different approaches to segmenting IR images, two of which are novel techniques introduced in this thesis. The first novel

approach is based on morphological processing of edge maps using local and global *a priori* information about the region to constrain the segmentation. It is referred to as Cued Morphological Processing of Edge Maps (CMpEm) in the remainder of this thesis. The second novel approach is based on the classification of edge components through a MLP and a SVM classifier, to determine whether an individual edge component belongs to the contour or not, based on its features. Finally, the third one is one or several region based segmentation techniques, either optimal thresholding or FCM clustering. The CMpEm is a key part of the segmentation strategy, as it will be shown to consistently outperform other techniques in a variety of scenarios. However, since an excellent segmentation is crucial for reliable ROI extraction and subsequent processing, if the performance of the segmentation step can be further improved, it should be done. For each of the segmentation groups shown in the diagram of Fig. 3.1, a fusion strategy will improve the total performance of the system by using the complementary contributions of each segmentation technique. This is also desirable because it adds flexibility to the architecture, allowing other segmentation techniques to be integrated as needed.

The two novel approaches are described in the following paragraphs, followed by a description of the fusion process.

3.3.1 Cued Morphological Processing of Edge Maps

The first novel segmentation technique applies morphological processing directly to an edge map comprised of both strong and weak edge components, in order to recover the contour of the object of interest. Working on the edge map allows one to recover edges and contours in very low contrast and high noise situations. In addition, a

small amount of *a priori* information about the expected shape of the region to segment is used in order to constrain the morphological processing of edges. Edge components are extracted using the Canny edge detection technique, which allows the detection of most weak edges thereby capturing fine details [139]. Mathematical morphological processing on the other hand is flexible enough to recover the general contour and shape of the main region. The main steps of the segmentation procedure are summarized in Fig. 3.4:

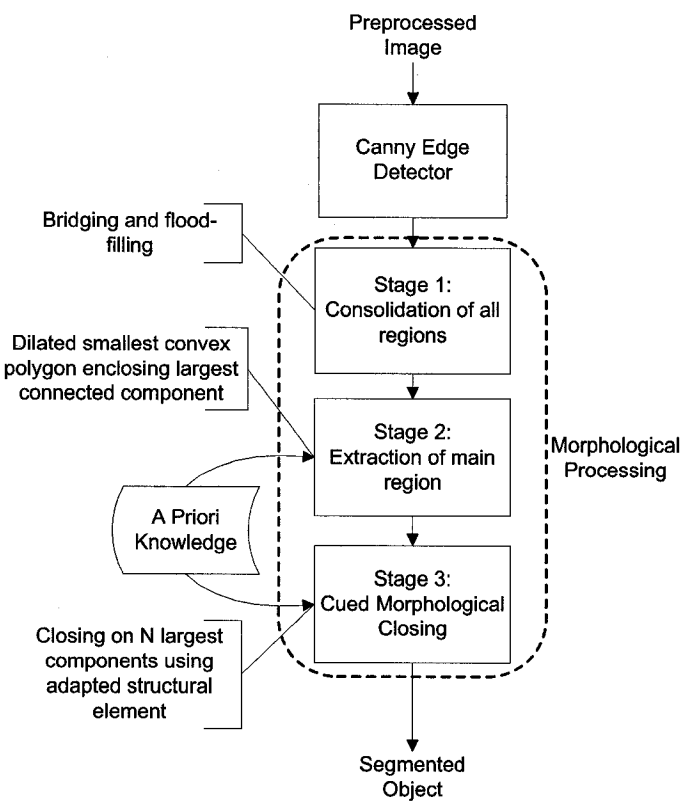


Figure 3.4: Overview diagram of cued morphological processing of edge maps.

Fig. 3.5 shows an example of the segmentation steps for an thermal infrared image of a hand. The segmentation steps are detailed hereafter:

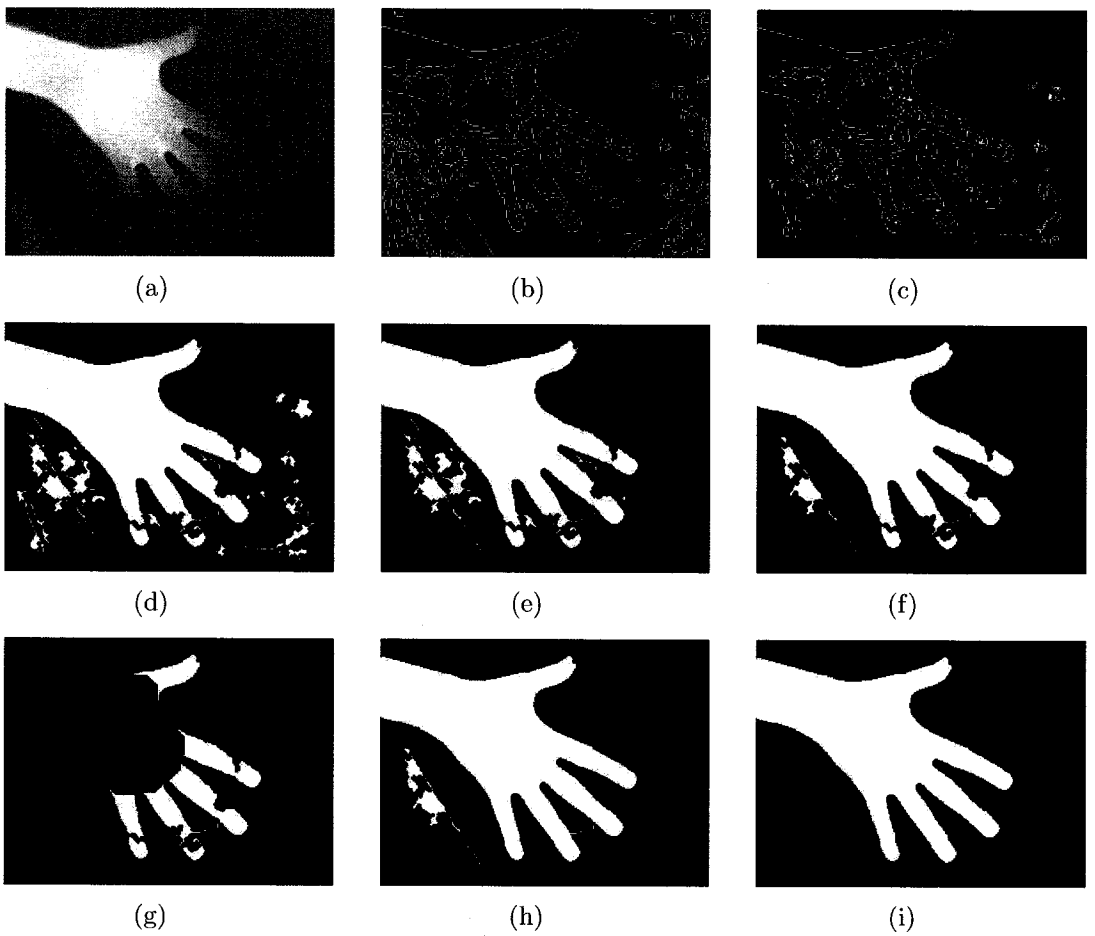


Figure 3.5: Hand segmentation example. a: Original image. b: Canny edge detection. c: Remove border, line-like components. d: Bridging, flood-filling. e: Keep components within main convex hull. f: Majority operation, threshold orientation and keep N largest components. g: Extract potential finger components. h: Cued line closing. i: Final segmentation.

1. Edge detection using Canny's technique (Fig. 3.5(b)). Each connected edge component is labelled.

2. Remove border edge components based on their centroid, and line like components based on their eccentricity (Fig. 3.5(c)).

3. Minor gaps between edge components are bridged and a flood fill operation

consolidates the largest region, under the assumption that there is only one object of interest and that this object consists of a unique region. If the size of the biggest filled region is very small, switch bridge operation to closing operation with increasing disk size until region has a reasonable size. This ensures that the filling is not stuck to a small area of the edge map (Fig. 3.5(d)).

4. The edge components outside of a dilated version of the smallest convex polygon enclosing the largest edge component are discarded. This step ensures that the edge components that are part of the object of interest are kept, although they might be disconnected from the main region from the previous step. It also eliminates most irrelevant edge components that are not part of the object of interest (Fig. 3.5(e)).

5. Majority operation to remove isolated components and spurious edge pixels (Fig. 3.5(f)). For binary images, this simple morphological operation consists of setting a pixel to 1 if a majority of the pixel in its neighborhood are equal to 1 and zero otherwise. A neighborhood of 3×3 was used in this thesis.

6. A maximum of N largest connected components are kept. The other ones are discarded. N is based on *a priori* information about the region. For instance, if the object of interest is a hand, N would be chosen to be 5. Note that this step is optional (Fig. 3.5(g)).

7. Final morphological "cleaning" operation. Filter out components whose orientation is very different from the main filled region. If more than one region remains, a morphological closing using a structural element adapted to the expected shape of the region to segment is performed. For example, if the ex-

pected shape is a hand, a line structural element is used, along the direction of the largest region found so far. This bridges the gap between disconnected regions belonging to the same shape and recovers regions that are typically missing from segmentations using other methods (Fig. 3.5(h)).

8. A final flood fill operation is performed and the contour of the resulting region is extracted by removing interior pixels (Fig. 3.5(i)).

3.3.2 Classification of Edge Components

The second approach consists of classifying a set of features describing the shape, location and orientation of each connected edge component found by a Canny edge detector. This will not yield a fully segmented contour in general, except in simple cases but it can provide an indication of how likely an edge component is to belong to the actual contour. As a result, when used in conjunction with the CMpEm or even with simpler methods, it can help recover some otherwise lost parts of the contour. The set of features used was the following:

- Area, Perimeter
- Lengths of the minor and major axis, eccentricity and orientation of the ellipse with the same normalized second moments as the connected component
- Diameter of the best fitting circle around the connected component
- Solidity and extent: proportion of the pixels in the convex hull (bounding box respectively) that are also in the region
- Centroid (centre of mass of the connected component)

The resulting database of training patterns and labelled targets (whether an edge component belongs to the expected contour or not) is then passed on to one of two classifiers: an MLP feed-forward, back-propagation neural network and a support vector machine classifier. Since most edge components do not belong to the contour of the object, there is a disproportionate amount of negative patterns in the training database submitted to each classifier. This is an issue for most classifiers that often leads to over-fitting of the negative class representing the background. The SVM formulation presented hereafter deals with this elegantly by simply changing the objective function to be minimized. For the MLP, there is no simple way to account for the uneven classes and therefore, the positive patterns were randomly resampled until the number of positive and negative patterns was equivalent.

The first classifier is based on C-Support Vector Classification (C-SVC), one formulation of the support vector machine framework [140]. It is a soft-margin implementation of the SVM formulation, where a weight parameter C is introduced to minimize the error on the margin due to outliers and noisy data points. C-SVC attempts to solve the following optimization problem:

For training patterns $x_i \in \mathbb{R}^n$, $i = 1...l$ from two classes, and a target vector $y \in \mathbb{R}^l$, such that $y_i \in \{-1, 1\}$, solve:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i \tag{3.1}$$

subject to $y_i\left(\mathbf{w}^T\phi(x_i) + b\right) \geq 1 - \xi_i$ and $\xi_i > 0$ for $i = 1...l$. Or equivalently its dual problem:

$$\min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - \mathbf{e}^T\alpha \tag{3.2}$$

subject to $\mathbf{y}^T\alpha = 0$ and $0 \leq \alpha_i \leq C$ for $i = 1...l$, where $\mathbf{w}$ are the support vectors, $\mathbf{e}$ is the matrix whose all components are 1, $C > 0$ is the upper bound on the margin error, Q is an $l \times l$ positive semi-definite matrix where $Q_{ij} \equiv y_i y_j K(x_i, x_j)$ and $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is the kernel mapping the training vectors into a higher dimensional space through ϕ. The decision function is given by:

$$\text{sign}\left(\sum_{i=1}^{l} y_i \alpha_i K(x_i x_j) + b\right) \tag{3.3}$$

The first term in Eq. 3.1 attempts to maximize the margin between the two classes in the high dimensional space, while the second term is a regularization (or smoothing) term that approximates the minimization of the probability of misclassification [140]. The C-SVC formulation is easily adaptable to the case of an uneven database. The objective function to be minimized is adjusted by varying the weight coefficient C for positive and negative patterns.

The kernel can be viewed as a nonlinear similarity measure between patterns [109]. It describes the geometric relationship between input pattern and the support vectors that characterize the difference between two classes. Common nonlinear kernels used in classification include polynomial kernels $K(x_i, x_j) = (\gamma x_i \cdot x_j + 1)^d$ where d is the degree of the polynomial, radial basis function kernels $e^{-\gamma \|x_i - x_j\|^2}$, where γ is proportional to the width of the Gaussian basis function and sigmoid kernels $K(x_i, x_j) = \tanh(a x_i^T x_j + r)$, where a is a gain and r a threshold value. Other kernels can be created so that they fit a particular application, as long as they satisfy the Mercer condition, which means that for all square integrable functions $g(x)$, the following holds:

$$\iint K(x, y) g(x) g(y) dx dy \geq 0 \tag{3.4}$$

Sigmoid kernels have been shown to be equivalent (in terms of decision function) to radial basis function kernels for certain range of values for a and r [141]. However, finding the optimal parameter for sigmoid kernels is not straightforward because the kernel matrix is not positive semi-definite for all values of the parameters a, r, which means that there is no primal-dual relationship between Eq. 3.1 and 3.2 and solving the optimization problem is a lot more difficult [141] [109]. Kernels specifically designed for image processing have also been investigated [113] [117], in the context of object detection and similarity between images. However, they would not benefit our analysis since the input patterns to classify in our scenario are not intensity-based features but feature vectors that are derived from edge components and that characterize the shape of the components. In addition, the classification task differs as well, since the goal is to classify edge components and recover a contour, while image-specific kernels typically try to detect an object or a particular type of image.

In this thesis, both polynomial and RBF kernels were investigated in order to assess the general performance of the C-SVC classification of edge components. The design of kernels specifically tailored to the classification of edge components could improve the correct classification rate and the overall contour detection but in this thesis the focus is on proving that the classification of edge components is a viable alternative even with simple generic kernels.

The second classifier, is a feed-forward back-propagation (1-input layer, 1-hidden layer, 1-output layer) multi-layer perceptron, with logistic sigmoid activation functions and a cross-entropy error function. The cross-entropy error function is better suited to binary classification problems than a sum-of-squares error function, because the latter is derived from the maximum likelihood principle assuming the target data

is from a smooth deterministic function with added Gaussian noise. This is not always the case for binary targets [89]. The cross-entropy error function is defined as [89]:

$$E = -\sum_i (t_i \ln y_i + (1 - t_i) \ln(1 - y_i)) \tag{3.5}$$

where $t_i = 1$, if pattern i belongs to the target contour, 0 otherwise; and y_i is the output from pattern i. One advantage of this formulation is that the output is a probability estimate of an edge component belonging to the expected contour, when a logistic sigmoid activation function is used [89]. The logistic sigmoid activation function is given by:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.6}$$

3.3.3 Region-Based Segmentation

Two types of region-based segmentation techniques were considered: a global optimal thresholding and a FCM clustering procedure. The global optimal thresholding technique is the algorithm proposed by Otsu [142], which separates the histogram of an image in two classes and minimizes the interclass variance between the two classes. It is a simple yet effective method that works best in the case of a bimodal histogram distribution.

The segmentation based on FCM clustering that was implemented in this thesis is a variant of the classical FCM clustering algorithm proposed by Wu *et al.* [50], where information about the local neighbourhood is taken into account when updating the membership values. The classical FCM clustering algorithm is similar to a K-means clustering algorithm where at each iteration data points - in our case pixels - are

assigned to the closest cluster of a pre-specified number of clusters, based on the distance to the centroid of clusters, until convergence. However, in FCM clustering each data point is given a cluster membership value instead of being assigned to a specific cluster. This membership value reveals how much a data point belongs to each cluster. When the algorithm has converged, the final membership values can be thresholded to obtain the final clustering. In Wu's variant, at each iteration, after calculating the membership values for each pixel, the resulting membership matrices (for each cluster) are filtered with a low-pass filter of a given spread. This changes the membership values within a given neighbourhood. This algorithm was tested successfully on infrared images by Wu *et al.* and is therefore a valid candidate for comparison with our approach or for possible integration in the fusion scheme.

A third method was tested and compared with our other approaches. It consists of a level-set active contour method. Active contour methods start from an initial guess of a contour and evolve that contour under internal and external forces until convergence. The stopping criterion is typically gradient based which means that convergence is typically achieved at the boundary of objects. In the level set implementation of active contours found in [143] [144] the contour is represented as the zero level set of a two-dimensional function $z = \phi(x, y)$. This allows one to represent implicitly any arbitrary topology of the object shape and it ensures continuity of the surface $\phi(x, y)$. In particular, it is able to represent an object composed of several discontinuous parts. The level-set is a variational optimization technique that computes the evolution of a contour under a velocity field, which may depend on position, time, geometry of the contour or external forces [143]. Active contours have been used in an IR imaging context by Zhou *et al.* [52].

3.3.4 Segmentation Decision Fusion Process

Throughout this section, the term classifier will be applied loosely to all segmentation techniques, including the CMpEm, Otsu and FCM techniques, which can be viewed as classifying each pixel as contour or background.

Decision fusion techniques are numerous in the image processing literature. As discussed in [145] and [146], when classifiers are very different in nature and the only common parameter is the format of their output (e.g. a class label), then voting schemes are most appropriate to combine the decision from each classifier. Each algorithm produces a binary value for each pixel, either one if it is estimated that it belongs to the contour or zero otherwise. The simplest voting strategy is a majority voting. If K is the number of classifiers, then for each pixel, if at least $\frac{K+1}{2}$, for K odd, or $\frac{K}{2}$, for K even, classifiers agree that the pixel belongs to the contour, that pixel is set to one.

A straightforward variant of the majority voting is to weigh the contribution of each classifier depending on their reliability estimate. These can be obtained from the classification performance of individual methods on test databases, by normalizing the performance criterion with respect to the best performer. The performance criterion in our case is the percentage of segmented contours that are within 90% or more in agreement with the ground truth contour, based on the Haussdorff distance between the two contours. For the case of binary weights, a full search for the best weights is possible since the number of classifiers used in this thesis is relatively small. This basically amounts to choosing the combination of classifiers that minimizes the error on a test database. This constitutes the first fusion alternative. For non-binary weights, the positive predictive value of each segmentation technique was used to

weigh their outcome and the majority voting was applied to the sum of all outcomes. A search for the optimal weights was also performed using the test database and the following objective function:

$$\min_{\mathbf{w}} \left(1 - \sum_{i \in S_{90}} H_m(i, \mathbf{w}) \right) \tag{3.7}$$

where $H_m(i, \mathbf{w})$ represents the distance between the ground truth and the computed contour for image i using weight vector $\mathbf{w}$ to weigh the contour found by each individual segmentation algorithm; The sum is on the images where the distance is greater than 0.9, which in our case means that a contour agrees with the ground truth on 90% of the contour. In addition the weights were chosen positive and the sum of the weights was constrained to be equal to 1.

Note that so far the weights were applied to each method at image level, not at the pixel level. In addition, when using the performance criterion or the positive predictive value mentioned previously as weights for the majority voting scheme, all the information from the negative results of each algorithm (false positives and false negatives) is not used. Using a Bayesian formalism, it is possible to use the information from misclassified patterns in order to estimate the probability that a pixel is on the contour or not [146]. First confusion matrices are derived for each classifier based on their individual performance on a test database. In our case, the confusion matrices are 2×2 matrices, representing the proportion of true positives, true negatives, false positives and false negatives for each classifier. The probability that a pixel belongs to the contour based on the observations from each classifier is

then given by [146]:

$$P(i,j \in C_m) = \frac{\prod_{k=1}^{K} P(i,j \in C_m | obs_k(i,j) = n)}{\sum_{m=1}^{2} \prod_{k=1}^{K} P(i,j \in C_m | obs_k(i,j) = n)} \tag{3.8}$$

where C_m is the class (here 1 or 0), $obs_k(i,j) = n$ is the outcome from classifier k and pixel i,j, $K = 5$ is the number of classifiers. The denominator is a normalizing term. The probability that a pixel i,j belongs to class C_m knowing that the outcome from classifier k is m_k can be estimated from the confusion matrix of classifier k by [146]:

$$P(i,j \in C_m / obs_k(i,j) = n_k) = \frac{p_{mn}}{p_{1n} + p_{0n}} \tag{3.9}$$

where p_{mn} is the number of samples belonging to class m that were classified as belonging to class n by a given classifier. The sum in the denominator reflects the binary class problem considered in this section. The final decision on whether a pixel belongs to the contour or not is as follows:

$$A(i,j) = \begin{cases} 1 & \text{if } P(i,j \in C_1) = \max_m(P(i,j \in C_m)) \geq \alpha, \\ 0 & \text{Otherwise.} \end{cases} \tag{3.10}$$

where α is a constant in $[0,1]$ that allows one to adjust the trade-off between specificity and sensitivity.

Finally, a new fusion technique was introduced. It uses *a priori* information about the agreement between segmentation techniques and morphological post-processing operations. Majority voting schemes discard the information from the classifiers belonging to the minority on a pixel-by-pixel basis. Since majority voting schemes are based solely on the final decision of each classifier, they do not take into account the

local spatial information that may be relevant for a correct final decision. In addition, the previous fusion strategies work best if all classifiers are independent from each other, which is not the case in this thesis.

Intuitively, if we assume that all classifiers perform reasonably well, when there is a large difference between the outcome of two or more classifiers, the correct contour may not be the contour that is most agreed upon but rather a mix of the different contours produced. In an information theoretical context, this would be equivalent to maximizing the entropy. It is also similar to the previous concept of using the errors from each method to increase the accuracy of the final prediction, except that it does not have to be limited to a decision on a pixel basis but can be applied to the entire image at once. In order to combine the contours, *a priori* information about the performance of each algorithm is needed, as well as the general agreement between each method. The latter is an indication of the dependence between each individual method. The performance of individual algorithms is readily available through testing. The agreement can be estimated by computing a robust distance between the contour returned by each algorithm on a test database. A good agreement between contours is not necessarily a perfect match but a reasonably small distance between the two contours. This ensures that minor changes between two contours does not affect the overall measure of agreement. From all this, the following rules were used in combining the different segmentation algorithms:

1. Discard algorithms with a high agreement with other algorithms and worse performance. Sort the remaining algorithms in terms of their performance. Algorithm 1 is the best, algorithm 2 the second best and so on.

2. Compare the contour from algorithm 1 with that of all remaining algorithms.

- If at least one algorithm agrees with algorithm 1 with a certain predetermined percentage of agreement (90% for example), then keep the contour from algorithm 1 as the final contour.
- Else move to the next best algorithm and increase the agreement requirement (to 99% for example) to account for the decrease in accuracy.

3. If there is no match between any of the algorithms, add the contours found by each algorithm together and perform the following:
   - Small morphological closing operation to close small gaps in the resulting binary image.
   - Flood-fill operation to fill holes in the binary image.
   - Small morphological opening operation to smooth out artifacts.

3.3.5 A Particular Case: Breast Segmentation

Isolating breasts in thermal IR images is difficult. The breast boundaries are not easily found on visible images either and their reliable extraction often requires additional active methods such as pushing upward or sideways on the breast to emphasize the top or side boundaries on the breast image [147]. This is not possible in the case of thermal imaging of the breast for obvious reasons. As a result, the detection of the contour of the breast is somewhat subjective and only a qualitative analysis of the segmented results is meaningful.

The lower contour of a breast may be partially or totally found using edge detection techniques. On the other hand, the top portion of the breast needs to be estimated, since there is no clear partition between breast and the upper part of the chest.

The method proposed by Qi *et al.* uses the Hough transform to detect lower breast contours [35]. However, the Hough transform is computationally expensive and is very much dependent on the edges found by the edge detector. It also assumes that the lower contour of the breast follows a specified parametric shape, which is not always true. As a result, this method fails when the lower breast boundary is not well defined.

The following method was introduced to recover the breast regions when the edge detector failed to find strong breast boundaries, in which case both Qi's method and Scales' method do not perform well [35] [34]. First the regions corresponding to the two nipples were found. In most cases they were much colder than the rest of the breast region and were easy to find through simple thresholding. The locations of the armpits and the bottom of the chin were also found by looking at the outermost contours of the body region, which were always well defined against the background. For detecting the outer contours of the body, a Sobel differential operator was used (horizontal and vertical masks) to find the strong edges corresponding to the outer body contour. The location of the armpits were found by looking at the point of highest curvature for left and right outer contours. Note that the thermal imaging of breasts requires the patient to raise their arms so that they do not affect the temperature of the side of the breast. This standard view simplified the armpit localization process. If the chin was present, it was consistently colder than the chest. The chin landmark could therefore be located by looking for the colder region situated in the middle portion of the body, above the armpits. The mid-line between left and right side of the body that is used for determining the middle portion of the body was estimated by calculating the distance between leftmost and rightmost contour of

the body. If the chin was not present in the breast thermogram, the topmost point along the contralateral symmetry line (mid-line previously defined) was used instead. Both armpit and chin (or topmost point on the symmetry line) were used to limit the segmentation to the area most likely to contain the breast.

The breast image was then divided into left and right side. The outer contour of the body was also removed at this stage. The resulting image area was passed through a Canny edge detector. The threshold for the Canny edge detector was determined adaptively by fixing a lower threshold on the density of the edges detected, defined as the ratio of the pixels belonging to edges to the total number of pixels in the region. Although the lower threshold on the density of edge components needs to be determined empirically across a range of breast images, it is a more robust parameter than the upper and lower thresholds on the Canny edge detector, since these can vary greatly from one image to another. Edges outside of the areas likely to contain the breasts were removed. Detected edges were further trimmed of all edge contours corresponding to purely vertical or horizontal lines since they were likely artifacts from the edge detection or from the image itself and did not relate to intuitive breast features.

The centre of mass of all the remaining edge components was determined and edge components closer to this centre of mass were considered more likely to belong to the breast area. An inclusion factor F that determined whether an edge component was close to the centre of mass or not was derived from an estimate of the breast size, using the mean distance from the nipples to the armpits, the mean distance from the nipples to the chin (or topmost point along the contralateral symmetry line) and the torso width as predictor of the size. The rationale behind this estimate is derived

from anthropometric studies of breast measurements [148] [149]. All edge components whose centroid was more than $F \times \sigma$ away from the centre of mass were removed, where σ is the standard deviation of the centroid of all edge components.

Finally, the smallest convex region enclosing the remaining edges was sought and the ellipse that best fitted its contour was chosen as the most likely breast area. The ellipse fitting algorithm was taken from Halir and Flusser [150]. The same procedure was repeated for both sides.

3.4 Identification of Anatomical Landmarks

In this thesis, the interesting landmark points are the surface anatomical interest points on the contour of the body regions. As discussed in section 2.1.2, most medical IR images have a rather narrow range, have little characteristic texture and generally do not reveal anatomical details. This is in contrast to other medical imaging modalities and makes the detection of landmarks challenging and unreliable, if one looks at the distribution of intensity values only. In addition, there is a wide inter-subject variation when looking at the absolute temperature distribution. For example, a smooth variation of temperature on one IR image of the back might indicate the location of the spine but may also indicate some abnormal temperature elevation or be similar to other thermal patterns on the back that do not correspond to anatomical information but rather correspond to physiological phenomena. As a result, relying on landmarks within the body region can lead to false detection and poor segmentation of the internal anatomical structures of interest. One should mention the notable exception of IR face images, where the difference of temperatures from various key features of the face allows the identification of key regions such as the eyes, nostrils and corners

of the mouth [151].

In view of the issues mentioned above, several assumptions need to be made in order to obtain a robust and reliable landmark detection. First, the landmarks should be restricted to the ones found on the outside contour of the body region, since they follow the shape of the body region and are unlikely to be missing, except in the special case of extremities. This in turn assumes that the contour was previously extracted via a segmentation step and that the contour was well recovered. Second, the landmarks found along the contour should be readily detectable on all images. This suggests limiting the landmarks to a few key stable surface landmarks. Anthropometric landmark points are such points, as they have been shown to be a stable set of points in all human beings [61]. Anthropometric landmark points are landmark points that are used for the measurement of human individuals and assess human physical variation. This latter assumption is an issue for many "blind" interest point detectors as they often produce a lot of false positives to avoid missing important feature points. This is made worse by the smoothness of the contour curves we are faced with, which mean detectors that require a threshold on the maximum angle detected will likely produce many more false positives.

Our approach addresses all these issues. It relies on the calculation of the direct curvature of contours but introduces a measure of symmetry to constrain the search for high-curvature points and limits the number of landmarks identified. Symmetry is an important visual tool that helps our visual system in its recognition task. It is also ubiquitous in a lot of natural objects and is prevalent in the human body. Using the symmetry or pseudo-symmetry of contours can help identify landmark points correctly.

The proposed algorithm assumes that a single well segmented object is present in the image. The segmented object is binarized and its contour is extracted using a simple subtraction with an eroded version of itself using: $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ as structural element. As this stage it is also assumed that there is some *a priori* knowledge about the symmetry of the object, either through expert input or through the use of a symmetry finding algorithm. The choice of the algorithm is not important as long as the main axes of symmetry are found. In the following, axes of symmetry were found using an implementation of the fast reflectional symmetry detection algorithm based on orientation histograms proposed by Sun and Si [152].

3.4.1 Stage 1: Curvature Analysis of Entire Contour

The first part of our approach relies on a curvature analysis of the contour of the object. Note that a good corner detector algorithm could be substituted at this stage, although these tend to produce a very high number of false positives when dealing with very smooth curves, which is the case for IR images of body regions.

First the contour coordinates are extracted in an anticlockwise manner starting from the leftmost extremum of the contour and following the curve until back at the original point, or until reaching a border or the end of a contour. The resulting curve is then fitted with a cubic smoothing spline - a smooth piecewise third-order polynomial function - thereby approximately interpolating and smoothing the curve. This facilitates the subsequent differential calculations on the curve since the differentiation of a polynomial function is trivial. The spline fitting is achieved by minimizing

the following objective function:

$$\min_{j} \left(p \sum_{j=1}^{n} |x(j) - f(t(j))|^2 + (1-p) \int \left| \frac{d^2 f(t)}{dt^2} \right|^2 du \right) \tag{3.11}$$

where p is a smoothing parameter, $(x(t))$ is the parameterization of the x coordinates of the contour and f is the spline function. The first term in (3.11) represents the goodness-of-fit in the least square sense to the data while the second term represent the smoothness of the resulting spline function. The smoothing parameter allows one to weigh both terms so that a trade off is achieved between closeness of the fit and desired smoothness of the function. A value of p close to 1 means that the spline interpolates the curve, while a value of p close to zero will give a smooth estimate of the curve. Spline functions are simple yet powerful to represent, approximate and manipulate complex curves and contours [153]. In general, the smoothing parameter for the fitting of the spline depends on the complexity of the contour analyzed and the amount of noise present in the image, which could create artifacts along the contour, thereby returning false landmark points. Measure of the complexity of a contour can be approximated with the square root of the isoperimetric ratio $p = \frac{\sqrt{4\pi Area}}{Perimeter}$ [154] for instance. However, for medical IR images, the contours are usually very smooth and therefore the second term in Eq. (3.11) does not play a significant role until the smoothing parameter is very small. This will be slightly different when considering the difference of contours in the next section.

Once the contour curve has been estimated, its curvature is calculated using:

$$\kappa = \frac{\frac{df_x}{dt} \times \frac{d^2 f_y}{dt^2} - \frac{df_y}{dt} \times \frac{d^2 f_x}{dt^2}}{\left(\left(\frac{df_x}{dt} \right)^2 + \left(\frac{df_y}{dt} \right)^2 \right)^{\frac{3}{2}}} \tag{3.12}$$

where f_x and f_y are the spline functions corresponding to the x and y coordinates of the contour respectively. The curvature is then normalized and a peak searching routine is performed to look for the extrema of the normalized curvature, with a threshold on the minimum peak value and a minimum distance between peaks. The threshold for minimum peak value was determined by specifying the percentile of the distribution of curvature peaks rather than an absolute threshold on the curvature values. The choice of percentile is not as dependent on the image. The minimum distance between peaks is a threshold that depends on the smoothness of the curve and the scale of the object (size and distance from the camera for instance). It should typically be relatively small ($\leq$ 10 pixels in our case) to avoid missing important features of the contour but not too small either to avoid catching irrelevant curvature changes ($\geq$ 3 pixels). The resulting peaks give the location of the extrema of the curvature, corresponding to potential points of interest.

3.4.2 Stage 2: Analysis of the Distance Between Symmetric Contours

One of the limitations of the first stage is the variability in the landmark points returned when only slight modification of the object occurs. This can be a significant issue when dealing with the contour of body regions, because the curvature at a specific location might vary significantly depending on the pose and orientation of these regions. For instance, using the curvature of the contour curve itself typically misses the lateral epicondyle (bottom elbow), which has a more subtle curve than the medial (top elbow), and also tend to miss the radial or ulnar styloid landmarks when the wrist is bent at a large angle with the arm. However, if one looks at the

distance between the top and bottom contours of the arm and wrist, the influence of the angle between the wrist and the arm becomes far less significant, allowing the recovery of the styloid landmarks. Therefore, the second stage of our approach consists of a curvature analysis similar to the first stage but this time on the distance curve between symmetric contours. Here the distance curve refers to the pixel-by-pixel distance between two symmetric contour curves, which is itself a curve. This assumes that the contour of the region exhibits near axial symmetry and it makes use of the symmetry of that shape to restrict the number of key points detected and therefore the number of non-significant landmarks.

The object is first rotated so that its axis of symmetry is horizontal. The pixel-wise distance between the top and bottom contours (top and bottom with respect to the now horizontal symmetry line) is calculated, yielding a new curve showing the evolution of the distance between the two contours. This process is schematically described in Fig. 3.6. If the contour of the object has some concave portions perpendicular to the axis of symmetry, for instance a horizontal arrow shape, then these parts have to be treated separately. The resulting difference curve will then consist of several distinct parts that can be processed iteratively to find the landmarks on each portion. A cubic smoothing spline is then fitted to the difference contour curve and the high curvature points are extracted in the manner described in 3.4.1.

3.4.3 Stage 3: Region Specific Constraints

The two previous methods are quite general and will produce important points along the curve that may not correspond to any significant anatomical details. In order to sort through them, one can look for specific location on the curves and its first- and

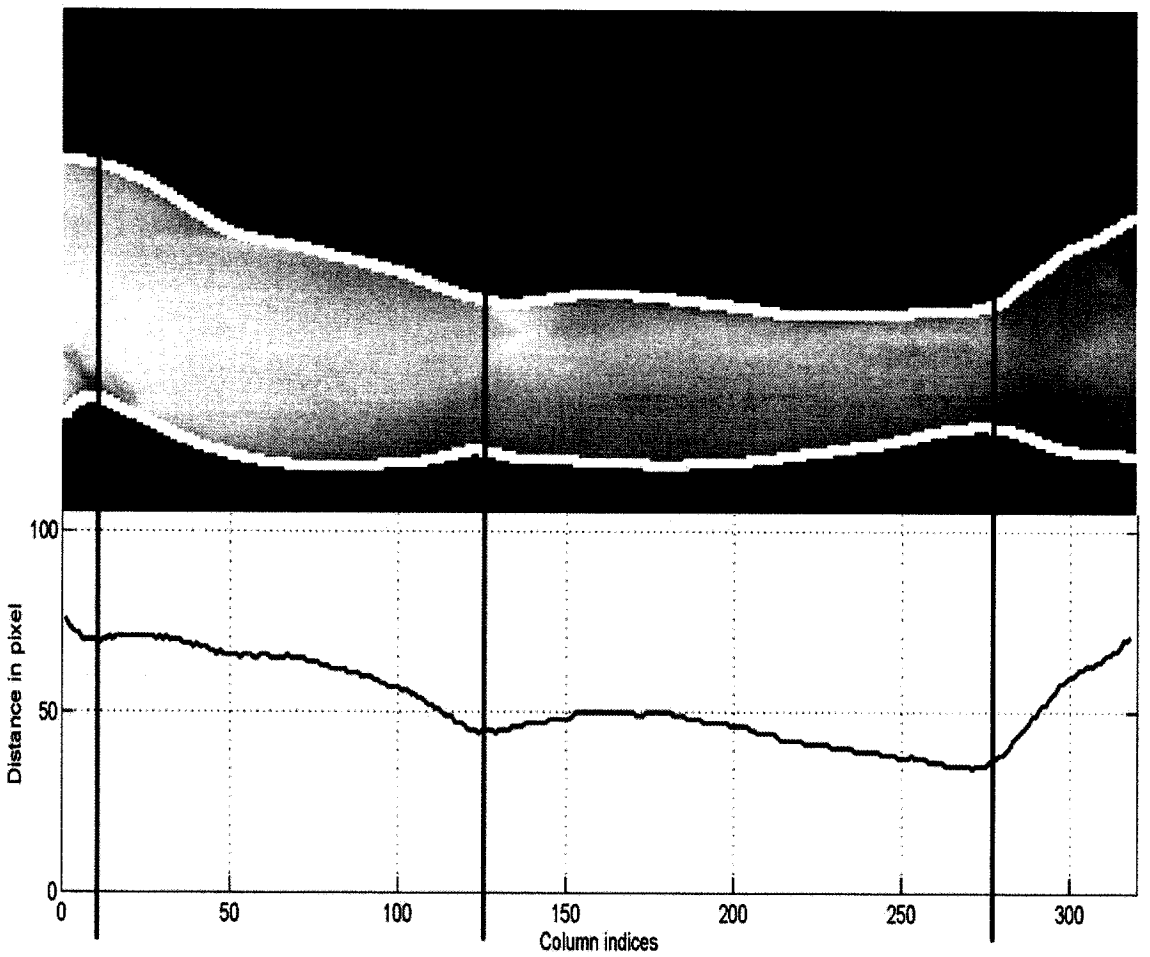


Figure 3.6: Example of a distance curve (in red) from the top and bottom contour of an arm (in bold white). The green lines show some of the salient point of the distance curve and their corresponding locations on the arm.

second-order derivatives. For instance, given a smooth approximation of the difference of contours from an arm, the styloid (wrist junction) and epicondyle (elbow junctions) were consistently found as the first and last zeros of the smoothed difference curve with positive curvature. These "bottlenecks" in the arm correspond anatomically to the ends of the humeral bone and are easy to spot for a human observer. Similarly, the base of the neck and the shoulder tip on images of the back consistently corresponds to

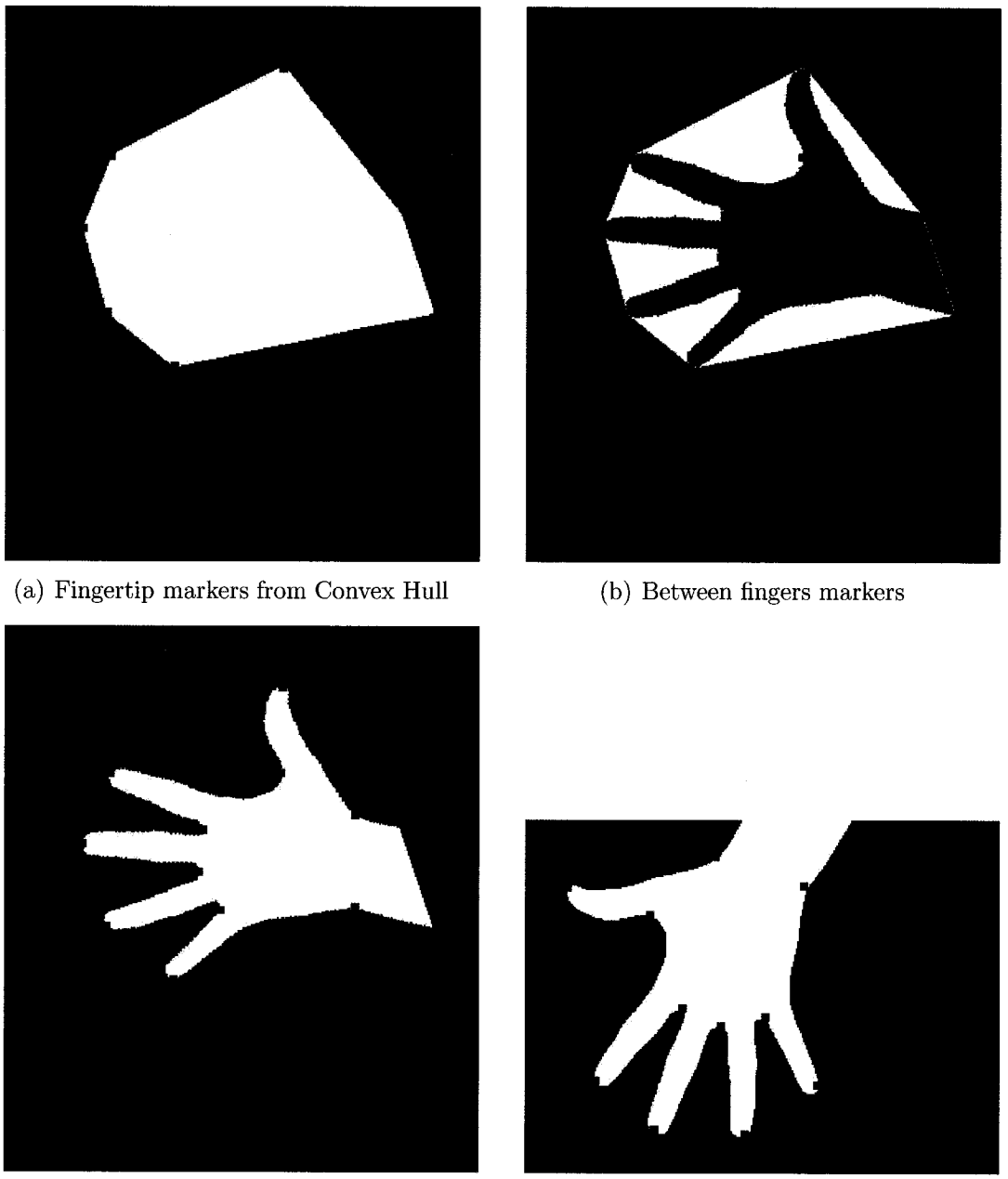
the highest and second highest curvature points respectively on the difference contour curve. For the hand, a high number of high-curvature points may be found by the curvature analysis, especially for the fingers. A quick way to sort through these can be done using the following algorithm:

1. The smallest convex polygon that encloses the entire hand is found and all the binary regions within that polygon that do not belong to the hand are extracted. On most images, this results in 6 distinct regions, 4 regions between the fingers and 2 outside regions. See Fig. 3.7 for an example. A small opening operation is performed to ensure that the regions are not attached to each other.

2. The image is rotated so that the main direction of the hand is horizontal. The location of the extrema points of each of these 6 regions is found and the rightmost (or leftmost depending on the direction of the hand) extremum point is kept for each region. The remaining extrema points are sorted with respect to their horizontal (x-axis) location. The first four points typically correspond to the junction between the palm and the fingers. See Fig. 3.7 for an example.

3. The fingertips are extracted in a similar way using the extrema points of the smallest convex polygon enclosing the hand.

This enables a coarse-to-fine approach to finding the fingertips, where a coarse estimate of the tip is obtained through the maxima of the convex region and a fine local search is then performed on the local curvature, thus preventing multiple landmark points to represent the end of the fingers.



(a) Fingertip markers from Convex Hull

(b) Between fingers markers

(c) All markers on rotated image

(d) All markers on original image

Figure 3.7: Steps showing the extraction of anatomical landmarks for a hand. The final landmarks are shown in red.

3.4.4 Summary of Landmark Detection Approach

For nearly symmetrical regions, the three previous stages are performed in parallel and combined, using a multi-step clustering process. First the landmarks produced by the first two curvature analysis stages are clustered, based on the Euclidean distance between centroids. An estimate of the location of the landmarks sought is given by the region specific constraints. Clusters are retained if they either match the region specific constraints or if they contain more than a set number of high-curvature points. Clusters remaining are then finally pair-matched with their symmetrical counterpart and isolated (non-paired) clusters are discarded. The centroids of these remaining clusters give the location of the landmarks sought. This coarse-to-fine approach improves the localization of the landmarks of interest.

Once the points corresponding to stable anthropometric landmarks have been found, additional landmarks can be derived, including interior landmark points. Most anatomical regions can be inferred from only a few key landmarks, using anatomical templates or atlases and simple affine transform.

3.5 Extraction of Anatomical Regions from Key Landmarks

In this section, some examples of extraction of anatomical regions are described for the arms, hands and back regions. All the anatomical regions presented are superficial anatomical structures from the musculoskeletal system primarily. The key landmarks for these regions consist of bony landmarks that are not affected by interpersonal variability very much. In this case, it is possible to use anatomical templates for the

extraction of muscle anatomy since muscle structure is linked to the bone structure at least at their origin and insertion points. As a result, if the landmarks detected correspond to bony structures, it is reasonable to assume from a biomechanical point of view that the shape and location of the muscles will follow proportionally the location and distance between bony landmarks. As a result, a simple affine transform should be able to match anatomical template regions to a region in a given image.

Figs. 3.8, 3.9, 3.10, 3.11 and 3.12 show examples of anatomical ROI extraction on the arms. Note that once the key landmarks are identified, other divisions are possible than the ones shown. They can be similarly implemented.

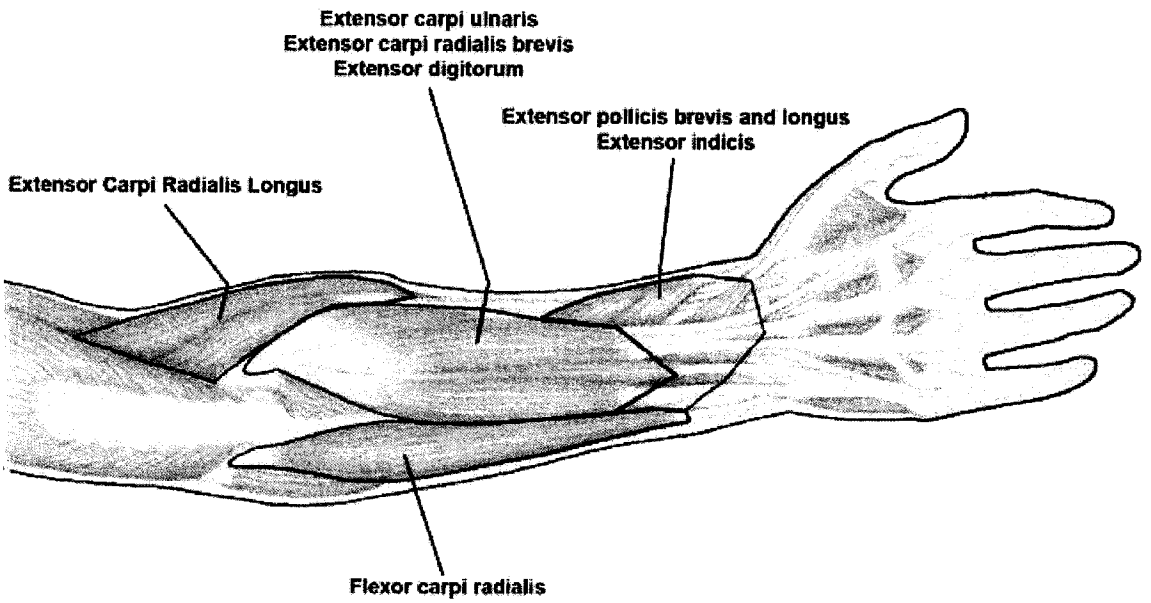
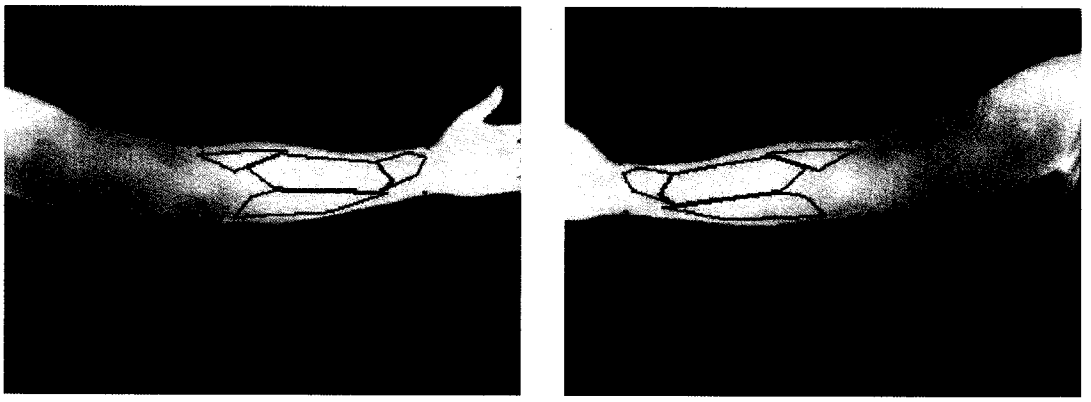


Figure 3.8: Anatomical dorsal view of the superficial muscles of the upper-arm.

Thermal images of the back can also be divided into anatomical segments, for instance areas of the skin originating from the same nerve root. An example of extraction of ROIs approximately corresponding to dermatomes is shown in Fig. 3.12(b).



(a) Dorsal view, right arm superficial muscles (b) Dorsal view, left arm superficial muscles

Figure 3.9: Example of ROI extraction of the dorsal superficial muscles of the upper-arm (a: right arm; b: left arm). The key landmarks are shown as cyan-coloured markers.

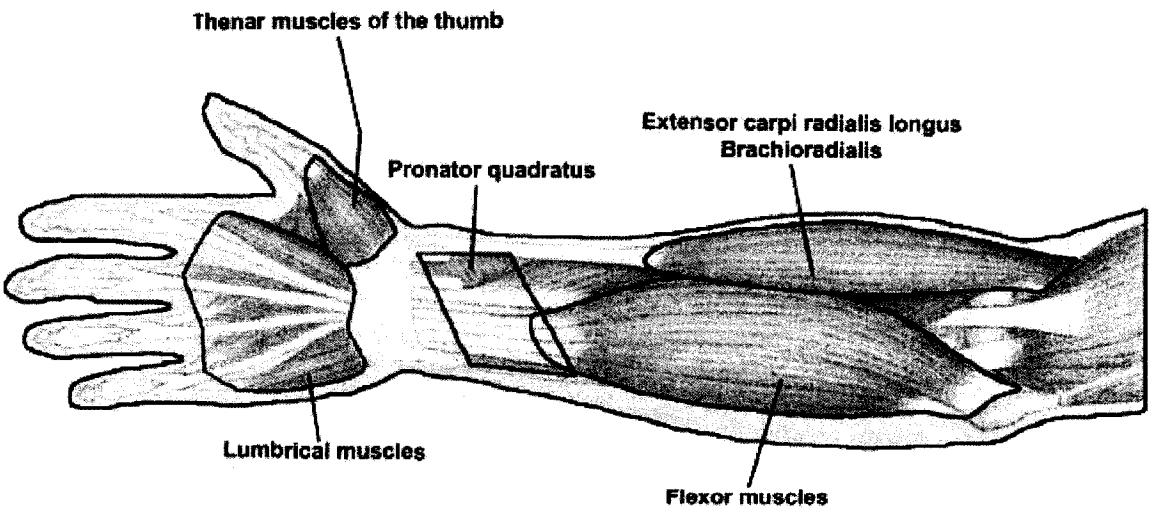
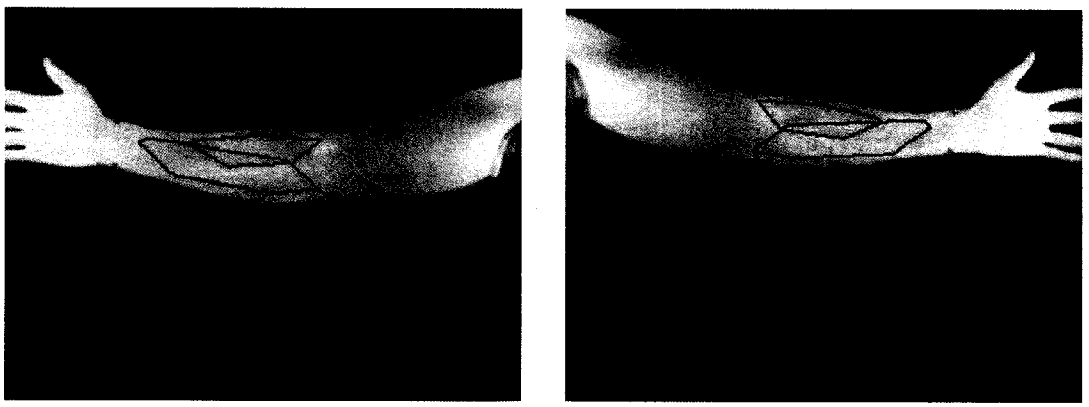
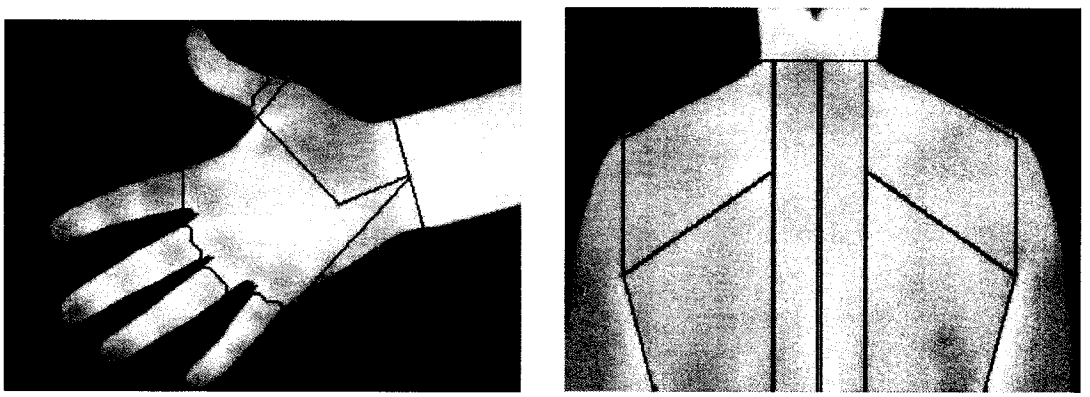


Figure 3.10: Anatomical anterior view of the superficial muscles of the upper-arm.



(a) Anterior view, right arm superficial muscles (b) Anterior view, left arm superficial muscles

Figure 3.11: Example of ROI extraction of the anterior superficial muscles of the upper-arm (a: right arm; b: left arm). The key landmarks are shown as cyan-coloured markers.



(a) Palmar view, right hand (b) Back dermatomes

Figure 3.12: Example of ROI extraction of the superficial muscles of the hand and back (a: palmar view, right hand; b: Back dermatomes with separate region for the spine)Anatomical landmarks used are shown in cyan-coloured markers.

3.6 Determination of Potential Anomalous Regions of Interest

This section discusses the strategies used when no specific information is available about the potential anomalies to be detected. Two parallel approaches are used to highlight "interesting" parts of the region. This helps ensure that the clusters produced are not artificial clusters created by the method used but real salient areas. The first approach uses some well known feature detectors that were combined by Privitera and Stark to define visual regions of interest in natural images [78]. They are normalized global contrast, entropy filter, Laplacian of Gaussian and Sobel operators. Each of them has been shown to correlate somewhat with human vision and the way we identify ROIs in images.

The normalized global contrast is defined as:

$$LA(i,j) = \frac{|L_{nhood}(i,j) - L_{IM}|}{(L_{nhood}(i,j) + L_{IM})} \tag{3.13}$$

where $L_{nhood}(i,j)$ is the local contrast within neighborhood $nhood$ of pixel (i,j), L_{IM} is the mean intensity of the image.

The entropy filter is the local entropy of an image defined as:

$$LocEntA(i,j) = -\sum_{n \in nhood} p_n \log(p_n) \tag{3.14}$$

where p_n is the relative frequency of gray-level n.

The laplacian of Gaussian filtered image is defined as:

$$LoGA = \frac{(l_1^2 + l_2^2 - 2\sigma^2)e^{-\frac{l_1^2+l_2^2}{2\sigma^2}}}{2\pi\sigma^6 \sum_{l_1} \sum_{l_2} e^{-\frac{l_1^2+l_2^2}{2\sigma^2}}} * A \tag{3.15}$$

where l_1, l_2 are the lengths of the 2D filter. In this thesis, $l_1 = l_2$. σ is the standard deviation of the Gaussian, $*$ is the convolution operation and A is the image.

The Sobel operator is the following edge operator:

$$SobA = S * S^T * A \tag{3.16}$$

where $S = \left(\begin{smallmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{smallmatrix}\right)$ is the Sobel filter, S^T is the tranpose of S and A is the image.

Each filter is applied to the image, and the resulting image intensity is normalized. Then the resulting four feature enhanced images are linearly combined to produce the final feature image FA:

$$FA = v_1 \times LA + v_2 \times LocEntA + v_3 \times LoGA + v_4 \times SobA \tag{3.17}$$

where the v_i's are the weights assigned to each feature image. The choice of weights is very much application dependent. In this thesis, the weights were all set to 1. As a result, no emphasis was put on one particular type of feature.

The resulting four feature enhanced images were linearly combined. Different weights can be allocated to each filter, depending on the features to be enhanced. Contrary to Privitera and Stark's approach, single centroid locations corresponding to visual ROIs were not sufficient. Instead the entire ROIs is needed. Instead of clustering the local maxima of the feature image, the feature image was thresholded

at 90% of the estimated cumulative density function (from the histogram of the feature image). Finally, a majority morphological operation was applied to the binary mask resulting from the thresholding operation, in order to eliminate tiny clusters. The size of the local neighborhood was 3×3. This resulted in well localized clusters of interest that corresponded to areas of higher or lower intensity than the rest of the image.

The unsupervised clustering procedure was performed on the body region directly. An ART network was used to segment the region into clusters of interest. Recall that ART networks attempt to create stable memories from inputs using match-based learning [119] [155]. This lead to either a resonant state that triggers stable learning of the candidate pattern or to a self-regulating parallel memory search [156]. If the search ends by selecting an existing class, then the class' center is refined to reflect new information from the input patter. If the search ends by selecting a previously untrained node, then learning of a new category takes place. ART neural networks consist schematically of two layers, the layer of class centers and a layer that receives the input data, feedback from each class and a time-varying bias signal. The comparison between the signal from the feedback connections and the inputs is thresholded with a parameter called vigilance that determines whether to assign the input to an existing class or to create a new one. The convergence of the feedback connections over time is called a resonance. ART networks do not require the setting of the number of classes *a priori*, which was desirable for our problem. Here, a class can contain several disjoint clusters of interest that share the same pixel intensity characeteristics.

A special implementation of ART networks called Fuzzy ART was used. Fuzzy

ART networks are essentially similar to ART1 networks but are meant for analog instead of binary inputs and use fuzzy set theory operation to compute the categories, the match criterion and in the learning phase. Fuzzy ART implementation uses fast learning, a slow rate of forgetting and complement coding. In fast learning, the network weights fully converge to a stable state values for each input pattern, enabling the network to adpat quickly to rare inputs. On the other hand, slow forgetting rate ensures that only persistent changes are taken into account. Finally complement coding is a normalization procedure that accounts for both critically present and critically absent features thereby limiting the proliferation of unnecessary classes.

The clusters of interest from both feature operators and ART clustering are merged together through an AND operation. This step ensures that only the most relevant clusters are kept.

3.7 Choice of Features

The following features are computed for all regions of interest (including whole image, clusters of interest and anatomical regions):

- First-order statistics: Mean, standard deviation, skewness, kurtosis, minimum, maximum, area.
- Entropy measures: Shannon entropy and log-entropy.
- Moments: Hu's moment invariants (7 moments), Maitra's moment invariants (6 moments), Zernike moments (12th order, 47 moments) (See appendix B for the definition of the moments).
- Multiscale features: wavelet features based on the square sum of wavelet coef-

Table 3.1: Summary of features. $X(i)$ stands for a column-wise vector representation of an image

Feature	Definition
Mean	$\mu = 1/(N) \times \sum_i^N X(i)$
Standard deviation	$\sigma = \sqrt{1/(N-1) \times \sum_i^N (X(i) - \mu)^2}$
skewness	$y = E(X - \mu)^3/\sigma^3$
kurtosis	$y = E(X - \mu)^4/\sigma^4$
minimum	
maximum	
area	$\sum_i X(i)$
Shannon entropy	$Ent_{sh} = -\sum_i s_i^2 \log(s_i^2)$
Log-entropy	$LogEnt = \sum_i \log(s_i^2)$
Hu's moment invariants	see appendix B
Maitra's moment invariants	see appendix B
Zernike moments	see appendix B
Wavelet features	for level i, detail j, $wf = 1/N \sum_{k=1}^N \lvert w_{ij}(k) \rvert$

ficients from a (7,9) biorthogonal spline wavelet decomposition with 3 levels of decomposition (for the whole image only).

The features are summarized in Table 3.1. The selection of relevant features depends on the application and the body regions considered. In addition to the previous features, difference statistics are computed if necessary, for example when one is interested in comparing the degree of asymmetry between two contralateral regions of the body or one hand with the other hand. Shape information is extracted from clusters of interest, that is the centroid of the cluster, the eccentricity and orientation of the ellipse with an equivalent second moment, the perimeter and finally the diameter of the best fitting circle enclosing the cluster.

3.8 Summary of Proposed Approaches

In this chapter, we introduced the following new segmentation approaches:

- A segmentation approach called cued morphological processing of edge maps, that uses a minimum amount of *a priori* information about the region to constrain the morphological processing of edge components.

- A segmentation approach based on the classification of edge components by a Support Vector Machine classifier and a Multilayer Perceptron. The classifiers are trained on a database of features characteristic of edge components belonging to the contour of the object of interest.

- A fusion strategy that combines the CMpEm, SVM and ANN classification of edge components, a thresholding and a FCM clustering algorithm in order to increase the overall segmentation performance. Several fusion strategies were described. In particular, a fusion method based on the morphological post-processing of combined segmented contours and on a set of rules derived from the agreement between each individual method was introduced.

Each proposed segmentation method (CMpEm, SVM and ANN) are tested and compared with each other and with an optimal thresholding method, a FCM approach that uses neigbhorhood information and a level set active contour method. The tests are carried on a synthetic database and on several databases of IR images. Similarly, the fusion approach proposed are compared to a majority voting with binary weights and a Bayesian fusion scheme, on several databases of IR images.

A new approach based on morphological processing of the edge components, the edge density and an ellipse fitting was also introduced to deal with the special case

of the segmentation of breast thermograms. This approach are compared with that of Qi *et al.*

A new approach was described to identify anatomical regions of interests in infrared images using automatically found landmarks. This approach is based on curvature analysis of the region contour, a curvature analysis of the distance between contours for symmetrical regions, and region specific constraints to produce clusters of potential locations for anatomical landmarks. It is able to locate accurately key anatomical landmarks on the contours. A comparison with several common landmark point detectors are performed in chapter 6, including a Harris corner detector, a curvature scale-space approach and a phase-congruency based method.

A technique for extracting anatomical regions from anatomical landmarks on IR images was also proposed. It is based on the small number of key landmark points extracted as described previously and a simple affine transform that allows one to recover similar regions in different images. Examples and results are presented in chapter 7.

Finally an approach was presented that enabled one to identify potentially abnormal Regions of Interest (ROIs) by using a combination of ART clustering algorithm and a fusion of feature images to highlight regions that differ from their surroundings. Applications and results of this approach will be presented in Section 6.3 of chapter 6.

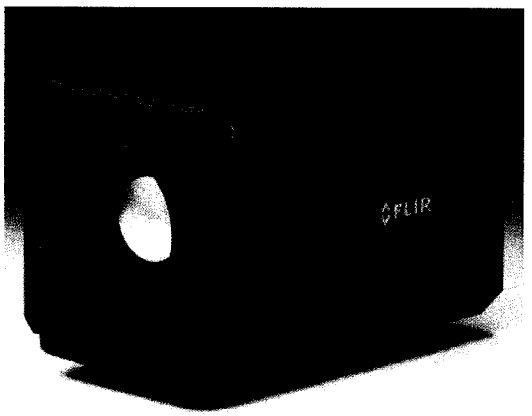
Chapter 4

Data Collection

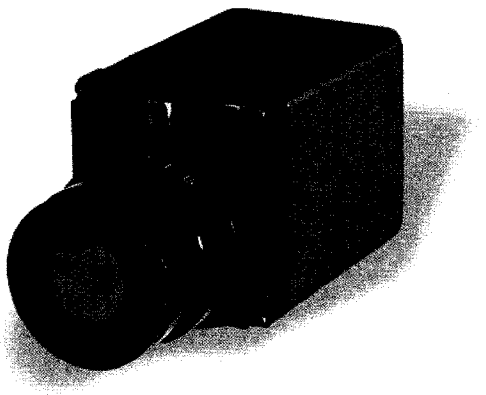
This section presents the IR equipment used in this work as well as the data that was collected from three different potential applications.

4.1 Infrared Technology

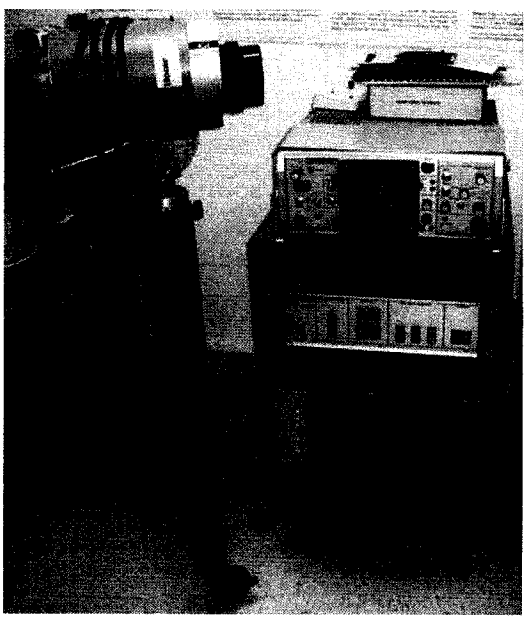
A majority of the IR images collected was acquired with a third-generation Thermovision 320M IR camera from FLIR Systems Inc. The camera used can be seen in Fig. 4.1(a). This infrared camera uses a solid state uncooled microbolometer Focal Plane Array (FPA) of 320×420 pixels, operating in the window $7.5\mu m - 13\mu m$ of the far IR range of the electromagnetic spectrum. It has a 24 mm germanium lens with an anti-reflective coating. This gives a $24° \times 18°$ field of view with a minimum focus distance of 50 cm. Its spatial resolution is 1.3 mrad, which means that at a distance of 1 m, each pixel represents a square of 1.3 mm $\times$ 1.3 mm. Its thermal resolution is 0.07 °C at 30 °C. The camera was connected to a laptop via a PC-Card interface, thereby allowing one to capture sequences of $14 -$ bit digital thermal images at a



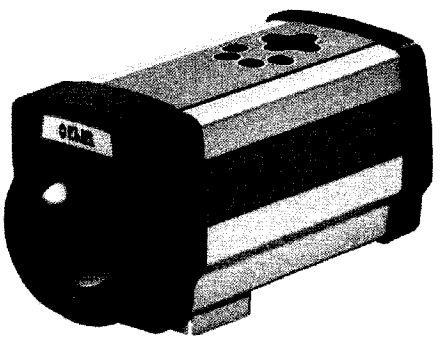
(a) Thermovision 320



(b) Omega



(c) AGA 680M



(d) Thermovision A40M

Figure 4.1: Three infrared cameras used to acquire the data.

maximum speed of 30 fps (frames per seconds).

Images were recorded in FLIR Systems Inc. proprietary format (".seq" for sequences and ".img" for images) using FLIR Thermacam Researcher software and subsequently exported in Matlab as a ".mat" array containing the temperature at each pixel of the image and information about the environmental conditions, such as the room temperature, maximum range of temperatures for each image, etc.

This IR camera is a radiometric thermal camera and therefore produces emissivity measurements that is later converted to absolute temperature measurements by means of the Thermacam Researcher software. The user must input some additional information about the object and general environmental conditions in order for the final temperature reported to represent the actual absolute temperature of the object. These parameters include an estimate of the object emissivity, for instance 0.98 for the skin, the temperature of the background and the ambient temperature. There are emissivity tables for most materials and the temperatures can be measured with a digital thermometer for example.

Three other cameras were used to acquire some of the images in the databases. The first one is a FLIR Systems Inc. Indigo Omega thermal imager camera. The Omega camera is shown in figure 4.1(b). It uses a 164x120 microbolometer FPA with an 18 mm lens giving a field of view of $25° \times 19°$ and a minimum focus distance of 30 cm. The spatial resolution is 2.7 mrad, which means that at a distance of 1 m, each pixel represents a square of 2.7 mm $\times$ 2.7 mm. Its thermal resolution is 0.085 °C at 20 °C. The Omega thermal imager is not a radiometric camera and is not capable of outputting absolute temperature measurements. The thermal images produced by this camera were not used for temperature measurements for this thesis; it was used

to test the image processing algorithms. This camera has the advantage of being extremely small, roughly 10 cm long, and light, roughly 150 g. It has a frame rate of 30 fps.

The second camera is a first-generation AGA Thermovision 680 medical IR camera system along with an AGA OSCAR recording and retrieval system. The camera used can be seen in Fig. 4.1(c). This camera uses an Indium-Antimonide photovoltaic sensor cooled with liquid nitrogen with a 53 mm lens giving a $25° \times 25°$ field of view and a minimum focus distance of 60 cm. Its thermal resolution is 0.1 °C at 30 °C and its spatial resolution is 3.4 mrad, which means that at a distance of 1 m, each pixel represents a square of 3.4 mm $\times$ 3.4 mm. The thermal images produced are analog and can be digitized to $8-$bit, 128×128 pixel images using the AGA OSCAR recording system. Images are stored on tapes and can be transferred to a PC in ASCII format. Temperature information is not available in the images as is, but can be recovered if the camera settings used for the acquisition are known. The thermal sensitivity of this camera is adequate, but the spatial resolution is poor and the size of camera, as well as the fact that it is cooled with liquid nitrogen, makes it difficult to use for the applications considered in this thesis. As a result, this camera was not used for acquiring new data. It is included here since some image processing tests were conducted on a database of IR images of pain patients and of patients from a breast clinic that was acquired by Monique Frize and her team at the Pain Clinic of the Moncton Hospital, Moncton, New Brunswick, Canada, between 1981 and 1984, using this camera.

The third one is an A40M Infrared Camera from FLIR Systems. The camera used can be seen in Fig. 4.1(d). This IR camera uses a solid state uncooled microbolometer

FPA of 320×420 pixels, operating in the window $7.5\mu m - 13\mu m$ of the far IR range of the electromagnetic spectrum. It has a 24 mm germanium lens with an anti-reflective coating. This gives a $24° \times 18°$ field of view with a minimum focus distance of 30 cm. Its spatial resolution is 1.3 mrad, which means that at a distance of 1 m, each pixel represents a square of 1.3 mm $\times$ 1.3 mm. Its thermal resolution is 0.05 °C at 30 °C. The camera was connected to a laptop via a Firewire interface, thereby capturing sequences of 14bit digital thermal images at a maximum speed of 60 fps (frames per seconds). This camera was used to acquire most of the pianist database described in section 4.3.1, where higher acquisition rate was necessary.

4.2 Infrared Backgrounds and Synthetic Images

IR backgrounds were acquired using a Thermovision 320M thermal IR camera system (FLIR systems Inc.) connected to a laptop. They were used to have a basic controlled environment for testing and comparing the segmentation algorithms. The ambient temperature and relative humidity of the room were kept at 23 ± 1 °C and $40\% \pm 2\%$ respectively. A typical experimental setup is shown in Fig. 4.2. Note that the subject was not present when acquiring the IR backgrounds. The camera was placed 50 cm above a table covered with some black squares of felt that had an emissivity close to 1. The squares were loosely overlapped to cover the entire table. The overlapping border between two squares of felt could be seen on the resulting IR images, which added to the difficulty of the segmentation task. Noise was created by using a fan blowing air at the surface of the table at different speeds.

Synthetic images consisted of 6 ellipses representing a rough shape of a hand, placed on different backgrounds. The contrast between object and background as

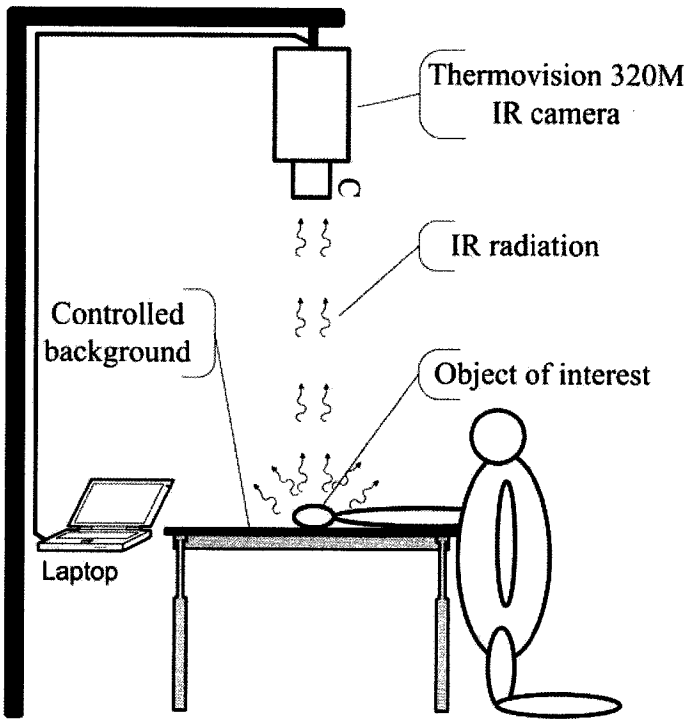
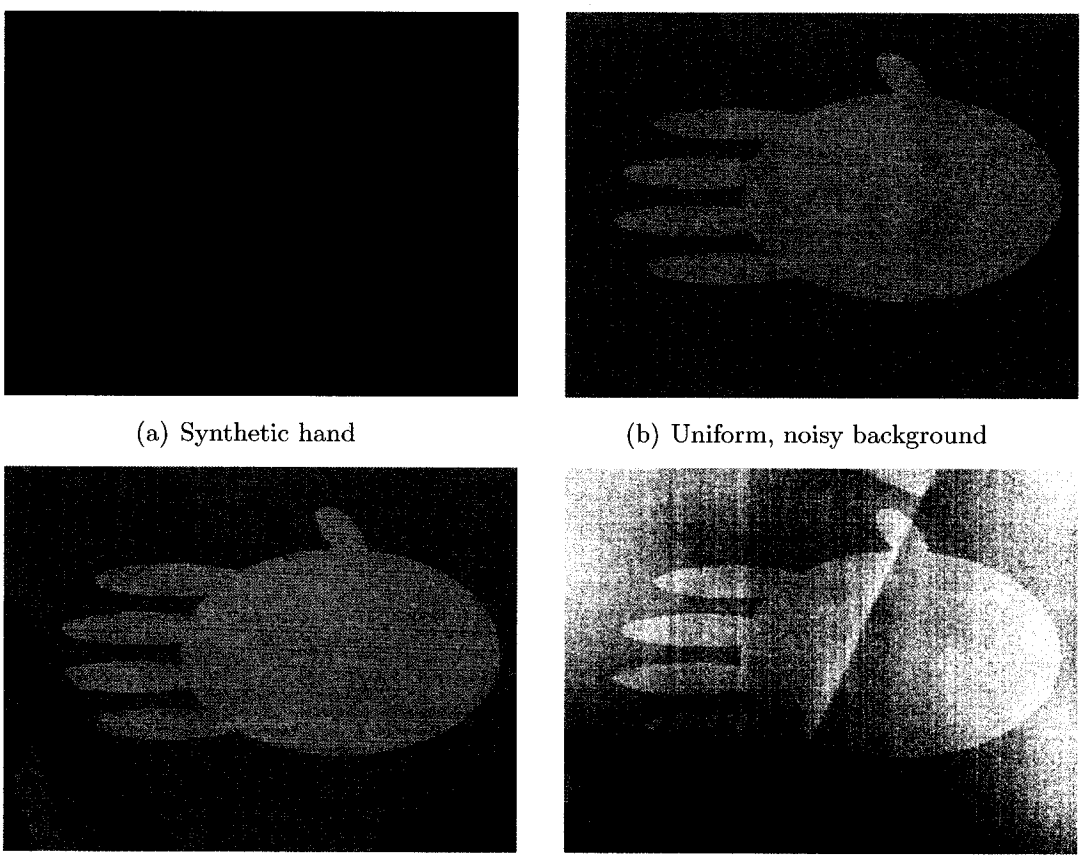


Figure 4.2: Experimental setup showing the imaging of the hand of a subject.

well as the complexity and amount of noise could be changed as desired. An example of a synthetic hand is shown in Fig. 4.3(a). The intensity of each region could be varied as necessary. The default synthetic hand of Fig. 4.3(a) had uniform grey-level intensity within each region. However, in order to test the performance of each algorithm, noisy backgrounds were added to the default image, with varying contrast and noise levels, and different levels of complexity of the background. The three different backgrounds are shown in Figs. 4.3(b),(c) and (d) for fixed contrast ratios. They are IR backgrounds acquired using the setup of Fig. 4.2, with no subject present.



(a) Synthetic hand

(b) Uniform, noisy background

(c) Non-uniform, noisy background

(d) Challenging, very noisy background

Figure 4.3: Synthetic hand and backgrounds used. a: example of a synthetic hand. Weber contrast ratio: 0.33. b: Synthetic hand on a noisy uniform infrared background (background 1). Weber contrast ratio: -0.81. c: Synthetic hand on a more challenging background with a low level of noise (background 2). Weber contrast ratio: -0.81. d: Synthetic hand on a very noisy and challenging background (background 3). Weber contrast ratio: -0.82.

4.3 Databases of Infrared Images

Databases of infrared images of subjects were created and are described below.

4.3.1 Pianist Database

The Pianist Database consists of sequences of thermal IR images from 17 pianists. The data was collected in two phases. The first series of images (9 pianists) consists of images of various views of the hands, upper-arms, arms, neck and face of the subjects taken before, during and after piano practice. The following protocol[1] was followed: The pianists were asked to come 15 minutes prior to their regular practice time wearing a sleeveless shirt to allow their body temperature to reach equilibrium with the ambient temperature of the room, which was kept constant at around 23 degrees Celsius. A first series of images of hands, arms, neck and face were taken, after which subjects started playing the piano. Standard views were used for each region of interest. Every 15 to 20 minutes, another series of images was taken until the end of the practice, which could last from 30 minutes to 2 hours. The duration of the practice was the same as the regular practice time of each subject and they played according to their level.

At the end of the practice, subjects were asked to wait 15 more minutes without playing, after which a last series of images were taken. Each series of images during the practice session took approximately 5 minutes. For each body region, we recorded a series of 10 images on average to reduce the risk of erroneous measurements. Images were recorded using the Thermovision 320M infrared camera from FLIR Systems Inc.

The second series of images (9 pianists, including 1 from the previous series) used the same protocol except that IR image video sequences were taken during the playing as well, which provided more precise temporal information about the local changes of temperature.

[1]The protocol was approved by the Research Ethics Board of the University of Ottawa (file H 06-04-18)

4.3.2 Computer User Database

The population considered for this project consisted of office workers (university staff, faculty, etc) or university students (minimum age of 18 years old) who used a computer regularly. Computer users are believed to be prone to several upper-extremity disorders affecting the hands and wrists due to the repetitive nature of typing. The population was divided into two groups after the data acquisition based on the presence or absence of symptoms of Repetitive Strain Injuries (RSI). The number of subjects at the time of writing was 18. The protocol was as follows: The test session was divided into several parts. First each participant was asked to complete a questionnaire and to show on a hand diagram the location of potential painful areas (see Appendix A, section A.2.1). The questionnaire consisted of multiple choice questions to assess the presence and severity of symptoms related to carpal tunnel syndrome as well as other repetitive strain injuries. The hand diagram represented both sides of the hands and forearms. Subjects would highlight the regions where symptoms appeared (including pain, discomfort, numbness, tingling, and others). Additional information was recorded, such as:

- Previous injuries to hands
- Sleep disorders
- Smoking or drinking
- How much time spent at the computer and types of job on the computer (straight input or editing, intellectual use, mixed use, binge work)
- Most frequent use: mouse or keyboard?

- Type of keyboard used
- Breaks
- Any other heavy work with hands (carpentry, gardening, musical instrument, etc)
- Dominant hand

The second part of the test session was the thermographic imaging. Participants were asked to follow the thermographic guidelines prior to the session to ensure accurate imaging. In addition, the room temperature and humidity were controlled and maintained within acceptable range (room temperature between $19\,^{\circ}\mathrm{C} - 22\,^{\circ}\mathrm{C}$, humidity between $50 - 70\%$). Participants removed any piece of clothing that was on their forearm, wrist or hand and waited 15 minutes to allow for their skin temperature to reach an equilibrium state. After 15 minutes, participants placed their hands on a small table covered with a dark piece of fabric. Several thermal images of their hands and wrists were taken, for each side (palmar and dorsal), using the Thermovision 320M or the Indigo Omega infrared cameras from FLIR Systems Inc. The images were taken continuously for one minute per side.

During the last part of the session, a series of simple short tests were performed with each participant.

- Wrist flexion test: participants were asked to bend and flex their wrists and the angle was measured and recorded for each direction.
- Finkelstein's sign: each participant was asked to make a fist with the thumb placed inside the palm. Then the researcher bent the participant's hand laterally (towards the fifth finger) and asked if it hurt.

- Tinel's sign: the participant's wrist was tapped on lightly by the researcher (over the nerve) to detect nerve irritability.

- Phalen's sign: The participant was asked to hold the forearms vertically and to allow both hands to drop into complete flexion at the wrist for approximately one minute.

- Electromyography: Up to 8 surface electrodes were placed on the participant's skin to acquire necessary biological signals. Some conductive gel was applied between the electrode and the skin to increase the quality of the signal recorded. Participants were asked to contract their muscles.

- Nerve conduction study: Electroneurograms were collected using a specialized data acquisition system. The system consisted of a stimulator, a photoelectric stimulus isolation unit, electrodes, electrode leads, biological signal amplifiers, an analog-to-digital converter, and a digital signal processing computer. Participants were asked to remain still and they received an electrical stimulus. The electrical stimulus was initially applied at the minimum level and slowly increased, in order to ensure participants' safety and comfort. Levels of electrical stimulus required could cause a sensation similar to someone repeatedly tapping the skin surface with the fingertip.

The overall session lasted approximately 60 minutes.

4.3.3 Foot Database

A small database of IR images of feet from two subjects was also acquired using a Thermovision A40M IR camera. The images represented a variety of views of

the foot and ankle. The imaging protocol previously described was not followed for this database, in order to obtain more natural imaging conditions than in a clinical setting. The background consisted of a hardwood floor with significant longitudinal gaps, which created an uneven background when captured by the IR camera. Also, the temperature equilibration period was not always respected, which resulted in some significant temperature variations across the foot for some of the images.

4.3.4 Test Infrared Databases

4.3.4.1 Hand Infrared Database

IR images of hands were acquired using the experimental setup of Fig. 4.2. Test subjects were asked to follow strict guidelines prior and during the imaging, as detailed in [7]. The complete database of IR images of hands consisted of:

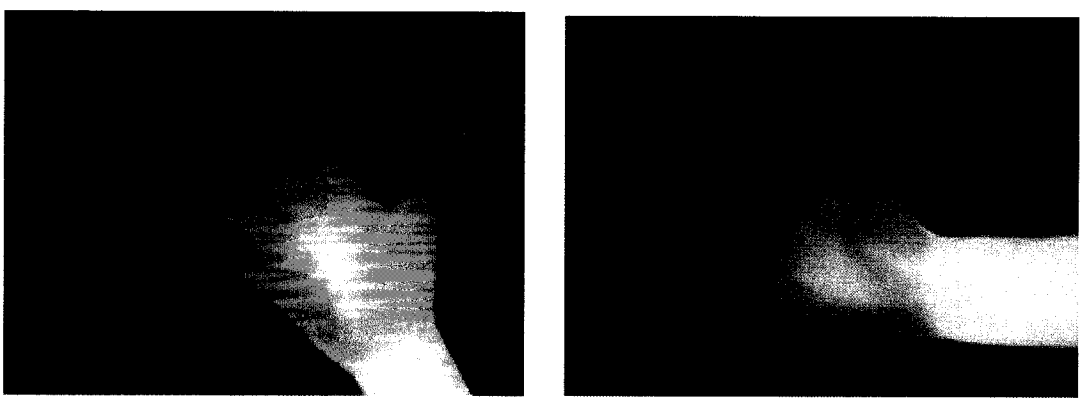
- 4108 images of hands from the piano database (right hand palmar and dorsal views, left hand palmar and dorsal views)
- 1239 images of hands from the computer user database (right hand palmar and dorsal views, left hand palmar and dorsal views)

For each view, from 10 to 30 images were taken at a frame rate between 3 to 12 fps. As a result, for a particular subject, many images are very similar to each other. The segmentation algorithms tested might perform better on some types of images than on other types. This could bias the results towards algorithms that perform better on an image with more frames per views. In order to draw a fair comparison between the different segmentation algorithms to be tested, it was decided to choose only one image per view. In other words, resulting images differed in their orientation, type of

view, location of the hand etc. This yielded 273 distinct IR images, from the piano database and 100 images from the computer user database.

In addition, for some of the results presented hereafter, the three types of background mentioned in Section 4.2 were linearly combined to the hand IR images by varying the contrast level of background and base image. A total of 50 levels of contrast between background and hand IR image were used for each background, ranging from no background added to no hand IR image present (only background present). The number of levels was chosen to ensure a good resolution when plotting the performance curves and to avoid missing the break point at which methods tested fail. This resulted in 150 test images per IR image of hand and a total of 55950 test IR images (3 backgrounds $\times$ 50 contrast levels $\times$ $(273 + 100)$ hand IR images). Note that a similar procedure was used for the synthetic images, yielding a total of 150 synthetic images for each orientation.

Two examples of test images are shown in Fig. 4.4 with their original background. Fig. 4.4 (a) is an IR image of a hand with a gradual decrease of temperature from



(a) Gradual temperature decrease (b) Overlapping temperatures

Figure 4.4: IR hand test images from the computer user database.

the palm to the tip of the fingers. The temperature of the hand is never below that of the background and the intensity across the hand is therefore always greater than that of the background. The background itself is uniform in temperature. Fig. 4.4(b) shows a greater decrease in temperature across the hand, which results in overlapping intensity distribution between the background and the object.

4.3.4.2 Arm Infrared Database

The complete database of IR images of arms consisted of:

- 2853 images of arms from the pianist database (right arm anterior and dorsal views, left arm anterior and dorsal views)
- 919 images of arms from the computer user database (right arm anterior and dorsal views, left arm anterior and dorsal views)

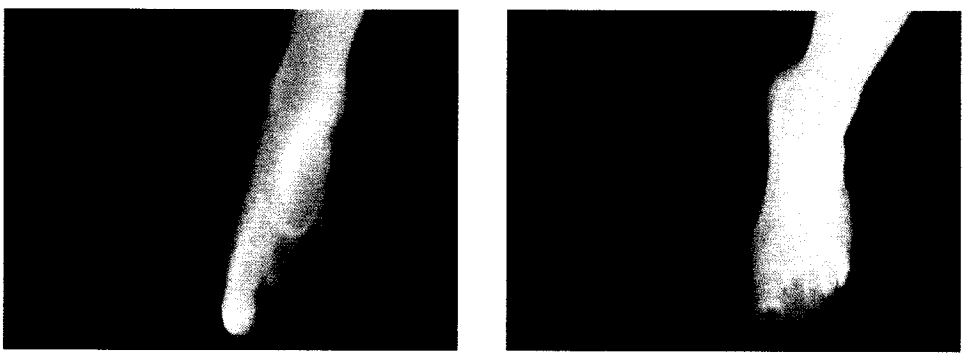
Similar to the hand IR database described in the previous section, a sequence of images was taken for each view, with similar images during the sequence. As a result, 238 distinct images were kept in total. Out of the 238 remaining images, 60 were from the computer user database and 178 were from the pianist database.

4.3.4.3 Foot Infrared Database

This database consisted of 940 images of left and right feet in dorsal and side views and various orientations with respect to the camera. Similar to the test hand databases from Section 4.3.4.1, a sequence of images was taken over a brief period of time for each view. For test purposes, redundant images were discarded and only distinct views were retained. This resulted in 27 distinct thermal IR images.

The original background on the feet images had a prominent parallel line pattern with varying intensity that were expected to be difficult to segment. There were also some phantom shapes of foot that were sometimes hard to differentiate from the actual foot. This made the segmentation quite challenging and not directly comparable with the simpler backgrounds of other databases tested. As a result, a third foot database was tested that had the background removed manually and a uniform background was used instead.

An example of original foot images are shown in Fig. 4.5.



(a) (b)

Figure 4.5: Examples of foot IR images with phantom foot (left) and regular line pattern background (right).

Chapter 5

Segmentation Results

This chapter presents results obtained from the segmentation approach.

5.1 Segmentation of Extremities

Each algorithm was tested on synthetic images and on real IR images of hands and feet. The hands and feet were chosen because they are challenging body regions to segment since their temperature distribution is the most variable and often overlap with that of the background. The true segmented contour was easily retrieved from the generated ellipses in the synthetic image case, thus providing ground truth segmentation for testing and comparing the algorithms. For real IR images, a manual segmentation was performed to obtain the ground truth results. Contours from all algorithms under different scenarios were compared to the ground truth contours using a modified Hausdorff distance:

$$H_d(A,B) = \frac{\sum(A \cap \widetilde{B}) + \sum(\widetilde{A} \cap B)}{2 \times \max(\sum(A), \sum(B))} \tag{5.1}$$

Where A, B are binary images whose ON pixels correspond to the segmented contour, $\widetilde{A}$ and $\widetilde{B}$ are dilated versions of A and B, respectively, using a 3×3 square structural element, $\cap$ is the logical AND operator and the sum is on all the ON pixels. Note that the modified Hausdorff distance as expressed in 5.1 is in fact an inverse distance, *i.e.*, $H_d(A, B) = 1$ when the two contours defined by the ON pixels in images A and B agree perfectly, and $H_d(A, B) = 0$ when the two contours are totally disjoint. As a result, the inverse distance H_d represents the percentage of agreement between two contours. The decrease in contrast is measured by the Weber contrast ratio given by:

$$C_{Weber} = \frac{I - I_b}{I} \tag{5.2}$$

Where I and I_b are the average intensity of the object and the background respectively.

5.1.1 Algorithms Tested and Parameters Used

5.1.1.1 FCM parameters

The modified FCM algorithm used 2 classes, an exponent of 2 for the partition matrix and a 3x3 local neighbourhood. Larger neighbourhoods did not yield better performance for the images tested. The FCM and thresholding method tend to produce multiple regions when the contrast is lowered, which resulted in a poor performance when compared to the ground truth segmented object. To improve on the performance of the FCM and thresholding methods, a post-processing step was added to retain one largest region [157]. Note that this may not be practical to implement and may degrade the segmentation performance for small objects.

5.1.1.2 Level-Set Active Contour

The level set active contour method implemented was a multiscale approach, with respect to the image, and a multiresolution approach, with respect to the active contour. Resolutions of 4, 8, and 2 pixel(s) were used successively in order to produce the final segmented contour. The starting resolution of 4 pixels was found adequate for the initial growth of the contour with respect to the size of the hand or foot. The resolution was then increased to get the main feature of the hand and then decreased to the finer resolution trying to catch the fine details of the hand and recover the fingers. One scale was found to be sufficient for our purpose and a maximum number of 500 iterations per resolution allowed convergence in all cases. The initial seed circular contour was placed within the object. Using multiple seeds within the object as suggested by Zhou *et al.* did not improve the segmentation in our tests [52].

5.1.1.3 SVM and ANN Classification of Edge Components

Training of the C-SVC-based and ANN-based classifiers should ideally be performed on a database containing both background edge components and edge components from the contour of the object that needs to be segmented.

For the hand database, both classifiers were trained on 100 hand IR images from the computer user database for segmentation of the hands, and tested on the 273 hand images from the pianist database. These databases are completely independent, thereby ensuring that the testing results are reliable. The total number of training patterns (edge components) extracted from the 100 training images was 13401.

The foot database was too small for reliably training ANN and SVM classifiers. To remedy this issue, we assumed that edge components belonging to the contour of

a foot were similar to that of a hand so that the trained classifiers from the hand database could be used for this new type of data. An unexpected advantage of using the hand database ANN and SVM classifiers was that we could test the generalization capabilities of these classifiers to similar shapes.

For the C-SVC-based segmentation, the libsvm C++ implementation was used (libsvm version 2.83) [158]. A grid search was performed to find the best parameters C and γ for a radial basis function kernel and C, γ and d for a polynomial kernel, using a 5-fold cross-validation on the RSI database, where each feature was first scaled to [-1, 1]. The parameters C and γ were constrained by $3 < \log_2 C < 25$ and $-1 < \log_2 \gamma < 15$ respectively. The search for the degree d of the polynomial was constrained to $[2, 10]$. The step size for the grid was 1 initially. The optimal neighbourhood was found and a finer search was then performed in that neighbourhood using a step size of 0.5 and 0.1 successively. Further reduction in the step size did not yield significant improvements in the overall classification rate on the training dataset. In addition, a penalty weight was assigned to the negative patterns (edge components that do not belong to the ground truth contour). The penalty weight was 40, based on the ratio of positive to negative patterns in the training set.

The C-SVC classifier was then trained with the best parameters, which were $C = 2^{1.2}$ and $\gamma = 2^{0.8}$ for the RBF kernel and $C = 2^{11.75}$, $\gamma = 1$ and $d = 6$ for the polynomial kernel. The best classification rate on the training patterns was 98.73% and 98.66% of correct classification for the RBF and polynomial kernels respectively. The total number of patterns tested was 13401, including 340 positive patterns (belonging to the true contour) and 13061 negative patterns (not belonging to the true contour).

Table 5.1: Confusion matrix for the C-SVC classifier on the training database, for the best polynomial kernel and the best RBF kernel. $P1$ and $P0$ indicate predicted outcomes (1 and 0 respectively). $Pc1$ and $Pc0$ indicate actual outcomes.

(a) Polynomial kernel SVM confusion matrix

PolySVM	Pc1	Pc0
P1	1.32%	0.02%
P0	1.32%	97.34%

(b) RBF SVM confusion matrix

RBFSVM	Pc1	Pc0
P1	1.31%	0.04%
P0	1.23%	97.43%

The confusion matrices reporting the number of true positives, true negatives, false positives and false negatives, for the best C-SVC classifier on the training database with both kernels are given in Table 5.1, where $P1, P0$ correspond to the positive and negative predicted patterns and $Pc1, Pc0$ correspond to the positive and negative actual patterns. The values are presented as percentage of the total number of patterns.

From the confusion matrix, the specificity, sensitivity, predictive positive and negative value as well as correct classification rate can be estimated. For the best RBF kernel they are: Specificity: 99.96%; sensitivity: 51.47%; positive predictive value: 97.22%; negative predictive value: 98.75% and correct classification rate: 98.73%. For the best polynomial kernel, they are: Specificity: 99.97%; sensitivity: 50.00%; positive predictive value: 98.15%; negative predictive value: 98.66% and correct classification rate: 98.66%. The difference in performance on the training database between the two kernels is not significant and in the rest of this chapter, only the results from the RBF kernel SVM are reported when tested on real and synthetic databases.

For the training of the MLP-based classifier, the features across all patterns were first normalized to lie between 0 and 1, then a PCA decomposition was performed on the 13 normalized features and the principal components that contributed for

less than 2% to the total variance of principal components were discarded. After that step, only 6 feature principal components remained. The hand database was randomly separated into a training set (80% of the database) and a test set (20% of the database). After resampling, the total number of patterns in the training set was 20785 (10336 positive patterns and 10449 negative patterns). The test set was not resampled so that it kept the initial ratio of positive to negative patterns (There were 68 positive patterns and 2612 negative patterns). The training was performed using a Quasi-Newton optimization of weights with a Broyden, Fletcher, Goldfarb, and Shanno (BFGS) update of the weights. In addition, MacKay's Bayesian regularization was used with one regularization parameter for all the weights to avoid over-fitting of the data. Determination of the best structure was achieved by varying the number of hidden nodes from 1 to 50 and choosing the architecture yielding the best performance. More specifically, for a given number of hidden nodes, the ANN was trained on the training set and its performance assessed on the test set. The average classification rate from the test set over 20 randomly initialized ANN with the current number of hidden nodes was recorded and the optimal number of hidden nodes was the one with the best average classification rate. The optimal number of hidden nodes in our case was $N_h = 46$. The output of the MLP is a probability that a given input edge component belongs to the ground truth contour. The final contour consisted of probabilities greater than 0.5.

The confusion matrix for the best MLP architecture on the test set from the training database is given in Table 5.2, where $P1, P0$ correspond to the positive and negative predicted patterns and $Pc1, Pc0$ correspond to the positive and negative actual patterns.

Table 5.2: Confusion matrix for the best MLP architecture ($N_h = 46$) on the test set from the training database. $P1$ and $P0$ indicate predicted outcomes (1 and 0 respectively). $Pc1$ and $Pc0$ indicate actual outcomes.

ANN	Pc1	Pc0
P1	1.72%	1.49%
P0	0.82%	95.97%

The values are presented as percentage of the total number of patterns. From the confusion matrix, the specificity, sensitivity, predictive positive and negative value as well as correct classification rate can be estimated. They are as follows: Specificity: 98.47%; sensitivity: 67.65%; positive predictive value: 53.49%; negative predictive value: 99.15% and correct classification rate: 97.69%.

5.1.1.4 Cued Morphological Processing of Edge Maps

The parameters used for the CMpEm approach are summarized in Table 5.3.

Table 5.3: Parameters used for CMpEm algorithm. *: The length of the line is estimated from the size of the main region and the distance to other components; Typical values are between 5 and 20; The angle is the orientation of the region(s).

Excluded border portion	7% of
Eccentricity threshold for line removal	0.9965
Radius of disk structural element when fill rate after flood filling is low	2 to 5
Radius of disk structural element for dilation for convex hull	5
Threshold on deviation of orientation from main region	$\pm 30°$
Structural element for morphological closing adapted to region	line*

5.1.2 Results on Synthetic Images for Individual Methods

All methods were tested on synthetic images except the SVM and ANN classification of edge components. This is because these two methods require a separate training database in order to assess their real performance. In addition, the synthetic image database of the hand used does not provide enough variability (even with the added backgrounds) to test the performance of the SVM and ANN classifiers once trained. Training and testing on a rotated, scaled version of the synthetic image database yielded near perfect performance for both classification methods, which is certainly not representative of their true performance on real images as we shall see in Section 5.1.3.

Figs. 5.1, 5.2 and 5.3 show the results obtained and how the CMpEm method outperformed the other algorithms as the contrast was decreased and for simple and more challenging backgrounds. The Hausdorff distances are presented as a "percentage of agreement" between calculated and actual contours. An agreement of 90% or above is considered good and resulting segmented regions can be used as is for the next processing stage. Values between 60% and 90% indicate that most of the object contour was recovered, although the extremities are usually lost and resulting segmented regions will require additional manual or automated post-processing steps in order to be useful. Values lower than 60% represent very poor and failed segmented objects. Our method outperformed the three other techniques in most scenarios, when the contrast was decreased in noise-free and in noisy complex backgrounds. The improvement in the ability to segment the object ranged between 2% and 35% with respect to the contrast ratio [157].

When the noise level was low (Fig. 5.1), and the background was uniform, the

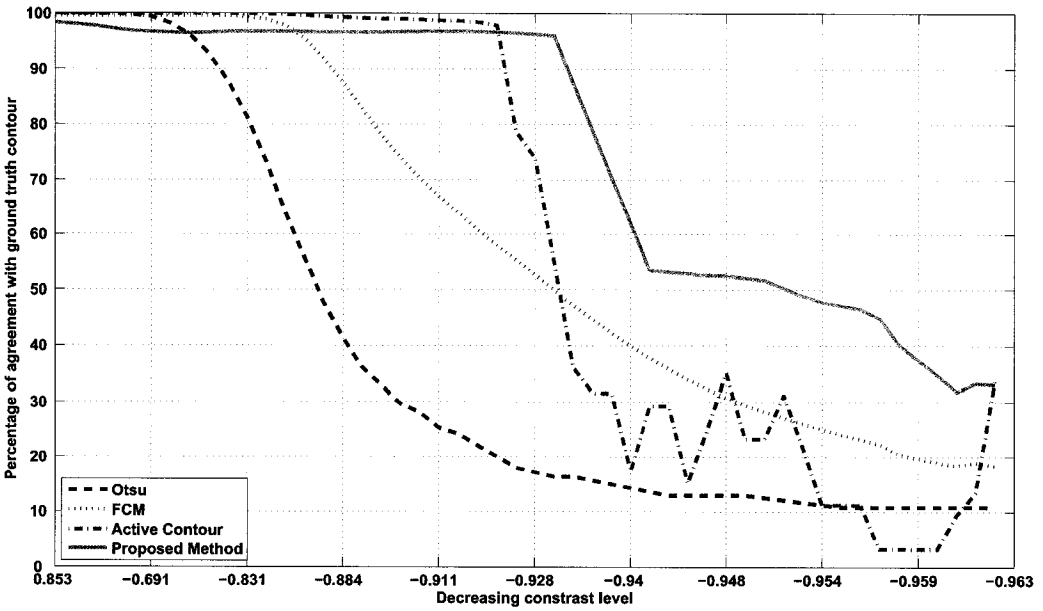


Figure 5.1: Percentage of agreement with ground truth contour for decreasing contrast level between region and background for a uniform noisy background (background #1 from Fig. 4.3)

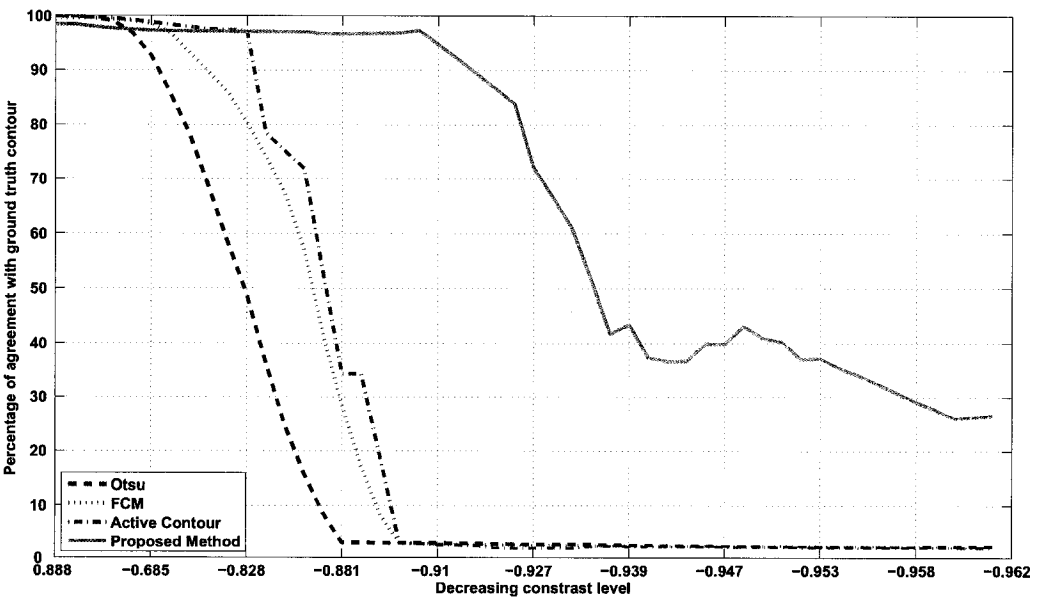


Figure 5.2: Percentage of agreement with ground truth contour for decreasing contrast level between region and background for a non-uniform noisy background (background #2 from Fig. 4.3)

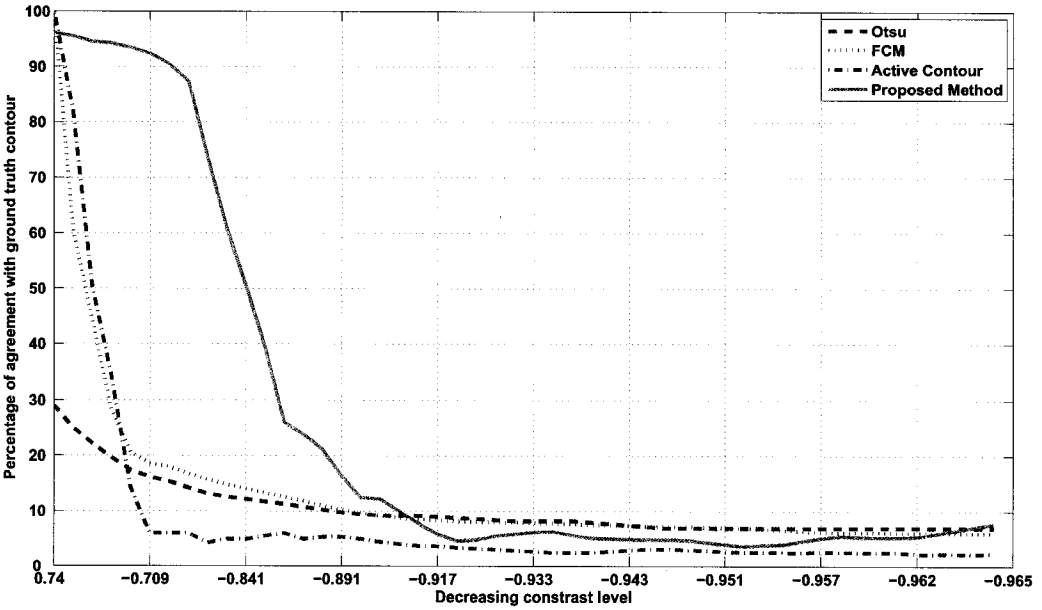


Figure 5.3: Percentage of agreement with ground truth contour for decreasing contrast level between region and background for a non-uniform, noisy and complex background (background #3 from Fig. 4.3)

four methods behaved well as long as the contrast was relatively high. The threshold of 60% for acceptable segmentations was reached for a Weber contrast ratio of -0.845 for the global thresholding and active contour method, -0.916 for the fuzzy clustering method and -0.938 for theCMpEm method. For increased levels of noise and complexity of the background, the difference of contrast at the 60% threshold increased between our method and the three other ones, as can be seen on Fig. 5.2 and Fig. 5.3. This is consistent with the design of the CMpEm algorithm, whose strength is in recovering contours when dealing with challenging backgrounds. This also explains the slight loss of percentage in contour agreement at high contrast between our method and the other three.

The FCM and Otsu thresholding method tend to produce multiple regions when the contrast is lowered, which resulted in a poor performance when compared to the

ground truth segmented object. To improve on the performance of the FCM and thresholding methods, a post-processing step was added to retain one largest region. The results for synthetic images improved significantly as can be seen in Fig. 5.4, 5.5 and 5.6, although for real images and smaller objects, it may not be possible to use this approach and the performance might degrade in those cases. Figs. 5.4, 5.5 and 5.6 show the results with the post-processing step. The CMpEm method still outperformed the other algorithms.

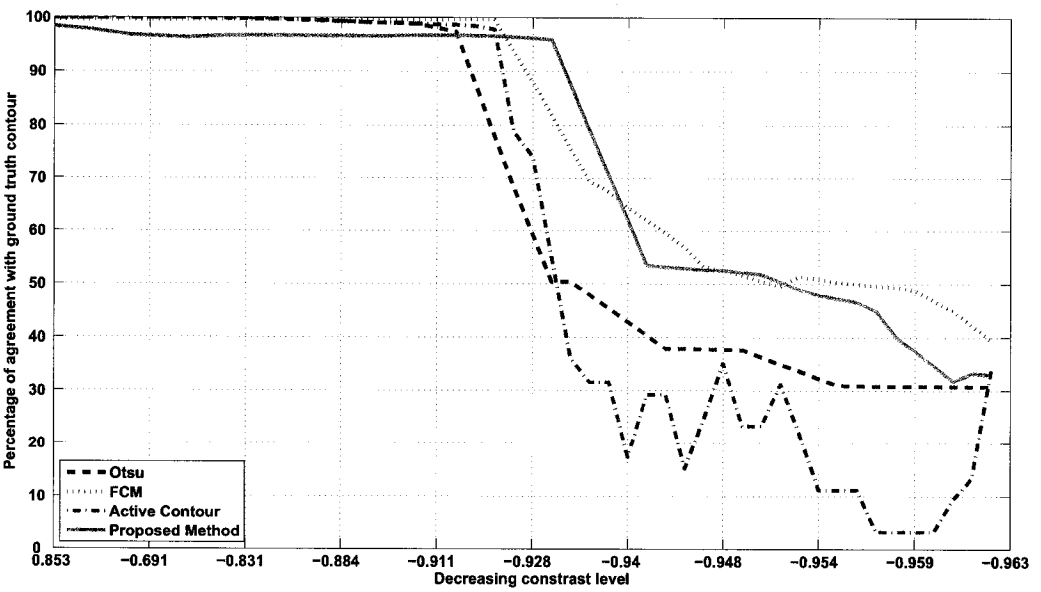


Figure 5.4: Agreement with ground truth contour for decreasing contrast between region and background, with uniform noisy background #1 from Fig. 4.3, and modified FCM algorithm with 2 clusters.

Fig. 5.7 show the segmentation results from the four algorithms, for a contrast ratio of 0.58.

It is possible to improve further on the performance of the FCM algorithm by varying the number of specified clusters. For example, choosing a number of clusters of 3 instead of 2 and keeping only one largest segmented region yielded results similar

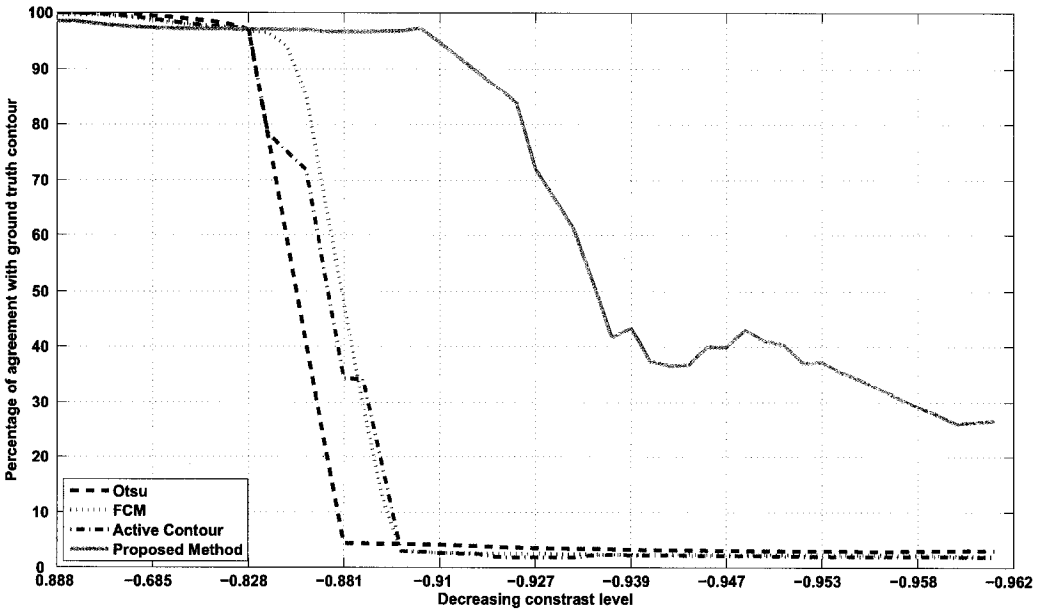


Figure 5.5: Agreement with ground truth contour for decreasing contrast between region and background, with non-uniform noisy background #2 from Fig. 4.3 and modified FCM algorithm with 2 clusters.

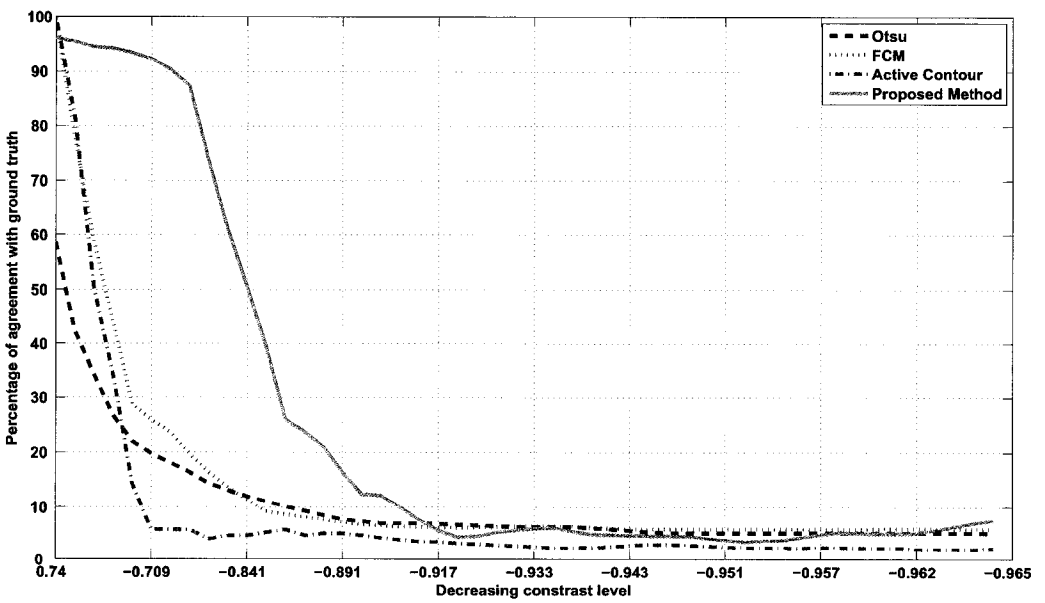
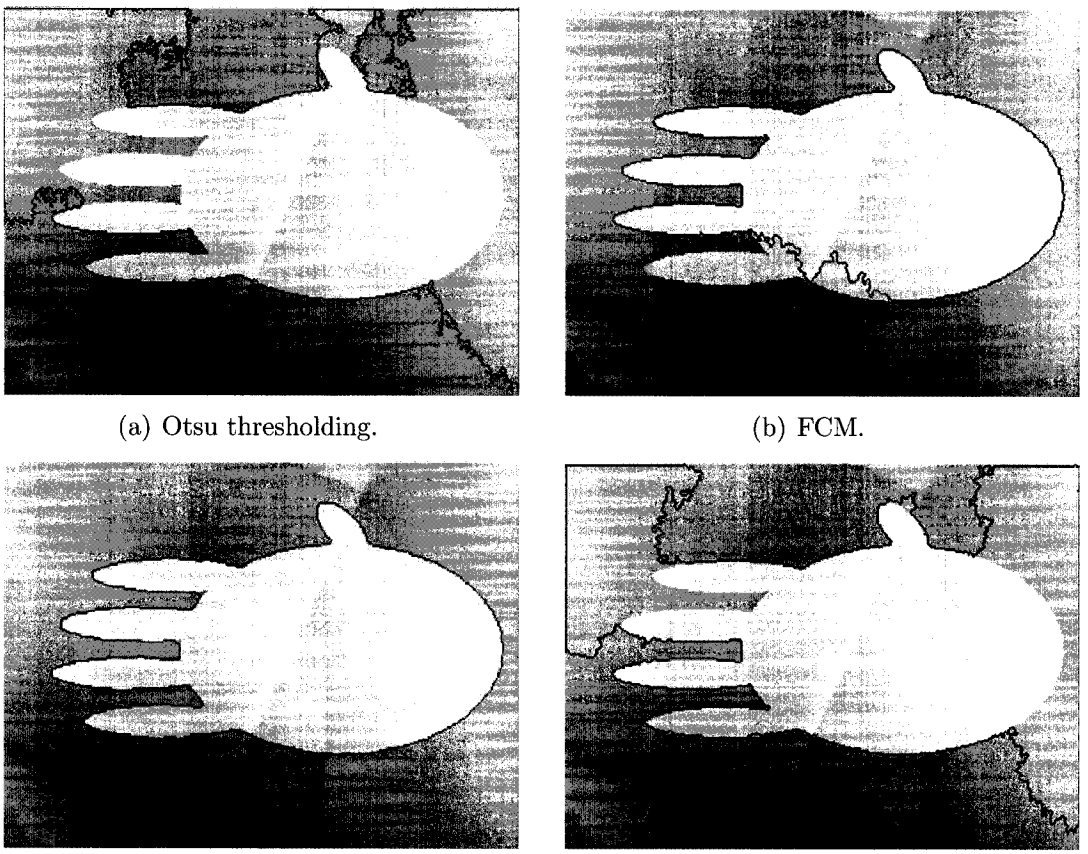


Figure 5.6: Agreement with ground truth contour for decreasing contrast between region and background, with complex background #3 from Fig. 4.3, and modified FCM algorithm with 2 clusters.



(a) Otsu thresholding.

(b) FCM.

(c) Cued morphological processing of edge map.

(d) Active contour level set.

Figure 5.7: Example of segmented synthetic image from algorithms tested on background #3. a: Otsu thresholding method (22.7% agreement with ground truth). b: FCM method (32% agreement with ground truth). c: cued morphological processing of edge map method (97.2% agreement with ground truth). d: Active contour level set method (52.9% agreement with ground truth)

to the CMpEm for simple uniform and non-uniform noisy backgrounds, as can be seen in Fig. 5.8 and Fig. 5.9. However, for more complex backgrounds, the CMpEm algorithm still performed better as shown in Fig. 5.10. In addition, finding the optimal number of clusters is not an easy task as can be seen on Fig. 5.8 at high contrast ratios, where the modified FCM algorithm oscillates around 75% of contour agreement as choosing 3 clusters forces the algorithm to separate some non-homogeneous

regions as belonging to the object. Choosing the number of clusters adaptively and in an unsupervised manner might yield better results for synthetic objects but is a difficult problem for FCM algorithms on anything but simple images, and usually increases the computational burden drastically.

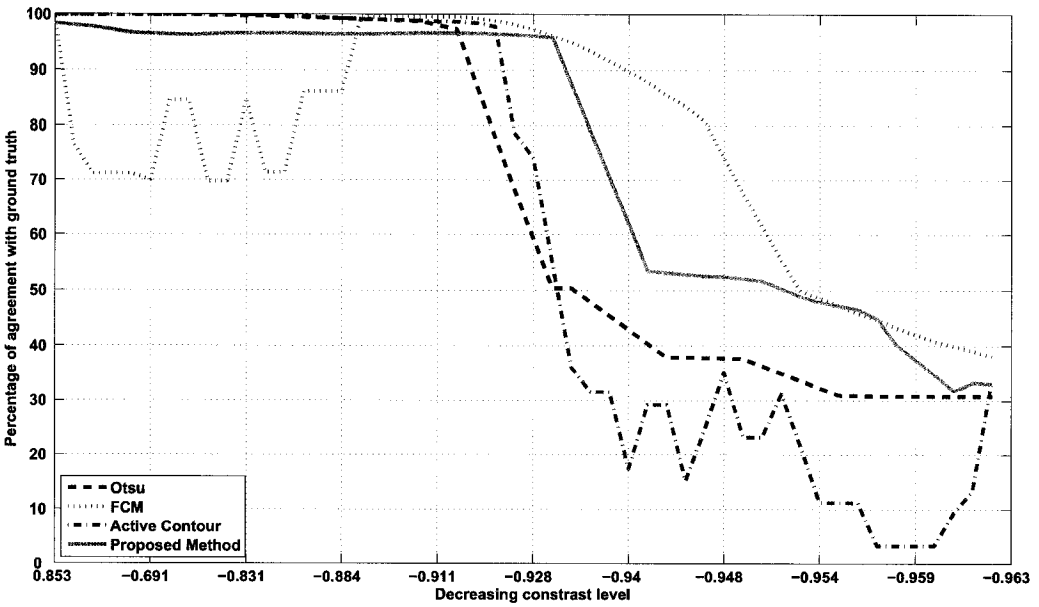


Figure 5.8: Agreement with ground truth contour for decreasing contrast between region and background, with uniform noisy background #1 from Fig. 4.3, and modified FCM algorithm with 3 clusters.

Some oscillations can also be noted at lower contrast ratios for the active contour method on Figs. 5.1, 5.4, and 5.8. This happens when the active contour leaks through holes in the contour of the object or when some regions of the background are included with the object.

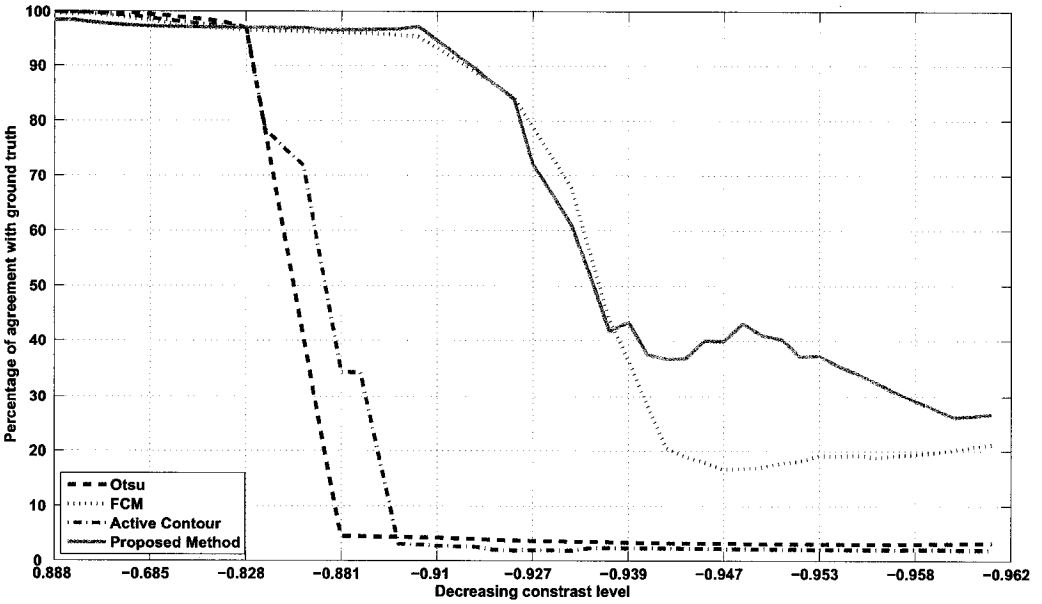


Figure 5.9: Agreement with ground truth contour for decreasing contrast between region and background, with uniform noisy background #2 from Fig. 4.3, and modified FCM algorithm with 3 clusters.

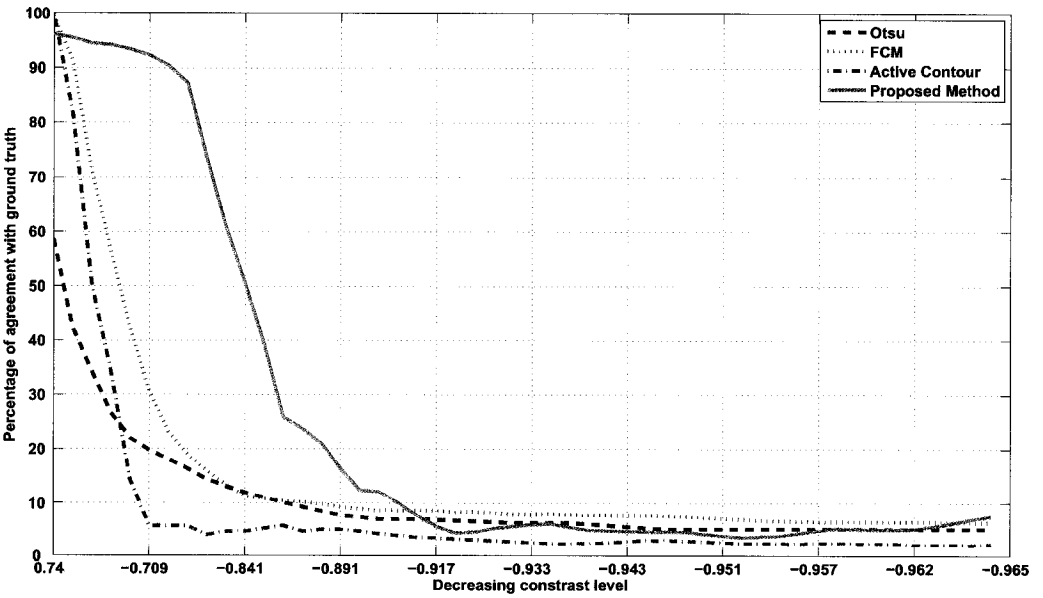


Figure 5.10: Agreement with ground truth contour for decreasing contrast between region and background, with uniform noisy background #3 from Fig. 4.3, and modified FCM algorithm with 3 clusters.

5.1.3 Results on Real Infrared Images

The Otsu thresholding, FCM, CMpEm, Active Contour Level Set (ACLS), ANN and SVM classification of edge components were tested on real IR images of hands and feet. Instead of showing multiple plots of the evolution of the modified Hausdorff distance between computed and expected contours (as in Fig. 5.1), we summarized the results over all backgrounds for each image using the following figure of merit P:

$$P = \frac{\sum_{i=1}^{3} \sum_{j=1}^{N} d_i(j)}{3 \times N} \tag{5.3}$$

where N is the number of backgrounds ($N = 50$ in this thesis) and the d_i's are defined as follows:

$$d_i(j) = \begin{cases} H_{d_i}(j) & \text{if } H_{d_i} \geq C, \\ 0 & \text{if } H_{d_i} < C. \end{cases} \tag{5.4}$$

$H_{d_i}(j)$ is the modified Hausdorff distance for background i and contrast index j, with $H_{d_i} \leq 1$, C is a threshold on the Hausdorff distance. In order to simplify notations, the image arguments A and B from Eq. 5.1 are omitted in Eq. 5.4. The threshold C was successively fixed at 0.7, 0.8 and 0.9 representing 70%, 80% and 90% of agreement between contours, respectively. The figure of merit P effectively reports the average amount of agreement between calculated and ground truth contours for all backgrounds and all contrast ratios, thus giving an idea of how well each algorithm performs over a range of experimental conditions.

Table 5.4: Confusion matrices for all segmentation techniques on the hand database. $P1$ and $P0$ indicate predicted outcomes (1 and 0 respectively). $Pc1$ and $Pc0$ indicate actual outcomes.

(a) ANN confusion matrix

ANN	Pc1	Pc0
P1	1.09%	0.11%
P0	0.18%	98.61%

(b) FCM confusion matrix

FCM	Pc1	Pc0
P1	0.85%	0.36%
P0	0.10%	98.69%

(c) Otsu confusion matrix

Otsu	Pc1	Pc0
P1	0.88%	0.34%
P0	0.14%	98.65%

(d) CMpEm confusion matrix

CMpEm	Pc1	Pc0
P1	1.10%	0.11%
P0	0.12%	98.68%

(e) SVM confusion matrix

SVM	Pc1	Pc0
P1	1.02%	0.19%
P0	0.03%	98.76%

5.1.3.1 Decision Fusion

The majority voting scheme with binary weights was implemented by doing a full search on the 31 possible combinations of techniques. This is excluding the active contour level set method, whose agreement with the FCM and Otsu thresholding method is very high despite its high complexity. As a result, it was excluded from the fusion process. The performance of each combination was assessed on our test databases and the best combination overall for each percentage of agreement previously described was kept. The combination are referred to as COMM1, COMM2 and COMM3 in the tables of results.

The Bayesian fusion framework was applied as described in Section 3.3.4. The confusion matrices were determined for each original test databases and are reported in tables 5.4 and 5.6.

Since the number of contour pixels is very low compared to that of the background

pixels, it is hard to differentiate between the performance of each algorithm. The results are therefore summarized in Table 5.5.

Table 5.5: Summary of confusion matrices results (hand database). SEN: Sensitity ($TP/(TP+FN)$); ACC: Accuracy ($(TP+TN)/(TP+FN+FP+TN)$); SPC: Specificity ($TN/(FP+TN)$); PPV: Positive predictive value ($TP/(TP+FP)$); NPV: Negative predictive value ($TN/(TN+FN)$).

	SEN	ACC	SPC	PPV	NPV
ANN	85.66%	99.70%	99.88%	90.49%	99.81%
FCM	89.57%	99.54%	99.64%	70.53%	99.90%
Otsu	86.54%	99.53%	99.66%	72.31%	99.86%
CMpEm	90.49%	99.78%	99.89%	91.11%	99.88%
SVM	96.83%	99.78%	99.81%	84.16%	99.97%

Similarly for the foot database, tables 5.6 (a), (b), (c) and (d) show the confusion matrices for all algorithms (except the ACLS method). Results are summarized in Table 5.7.

The specificity, accuracy and negative predictive value in tables 5.5 and 5.7 are artificially high due to the high proportion of the background class compared to the contour class. Since we are most interested in the true positives, the sensitivity and positive predictive values are the more interesting numbers to look at. In addition, since the accuracy is high, the threshold α from Eq. 3.10 was set to a value close to one (0.99 in all the following).

Table 5.8 gives the average agreement between calculated contours for both databases (combined).

From Table 5.8, we see that the FCM and Otsu thresholding methods on average return the same contour and including both methods in the fusion would be redundant. The Bayesian fusion of the remaining methods was performed using the values from the confusion matrices according to the description given in Chapter 3, Section

Table 5.6: Confusion matrices for all segmentation techniques on the foot database.$P1$ and $P0$ indicate predicted outcomes (1 and 0 respectively). $Pc1$ and $Pc0$ indicate actual outcomes.

(a) ANN confusion matrix

ANN	Pc1	Pc0
P1	0.52%	0.11%
P0	0.55%	98.81%

(b) FCM confusion matrix

FCM	Pc1	Pc0
P1	0.42%	0.22%
P0	0.12%	99.24%

(c) Otsu confusion matrix

Otsu	Pc1	Pc0
P1	0.43%	0.21%
P0	0.14%	99.22%

(d) CMpEm confusion matrix

CMpEm	Pc1	Pc0
P1	0.57%	0.07%
P0	0.04%	99.32%

(e) SVM confusion matrix

SVM	Pc1	Pc0
P1	0.39%	0.25%
P0	0.07%	99.29%

Table 5.7: Summary of confusion matrices results (foot database). SEN: Sensitity $(TP/(TP+FN))$; ACC: Accuracy $((TP+TN)/(TP+FN+FP+TN))$; SPC: Specificity $(TN/(FP+TN))$; PPV: Positive predictive value $(TP/(TP+FP))$; NPV: Negative predictive value $(TN/(TN+FN))$.

	SEN	ACC	SPC	PPV	NPV
ANN	48.46%	99.22%	99.88%	81.92%	99.44%
FCM	77.67%	99.66%	99.78%	65.72%	99.88%
Otsu	74.93%	99.64%	99.79%	66.86%	99.86%
CMpEm	94.09%	99.90%	99.93%	89.61%	99.96%
SVM	84.00%	99.68%	99.75%	60.85%	99.93%

Table 5.8: Average agreement in terms of modified Hausdorff distance between methods.

	ANN	FCM	Otsu	CMpEm	SVM
ANN	100%	61.32%	63.64%	77.73%	78.94%
FCM	61.32%	100%	95.82%	64.01%	66.23%
Otsu	63.64%	95.82%	100%	66.10%	68.46%
CMpEm	77.73%	64.01%	66.10%	100%	79.56%
SVM	78.94%	66.23%	68.46%	79.56%	100%

3.3.4.

Finally, for the rule-based morphological fusion process, the algorithms were ranked according to their positive predictive value from tables 5.5 and 5.7 and the algorithm described in Chapter 3, Section 3.3.4 was used with the the following predetermined agreement requirements: 90% for the agreement with algorithm 1 and 95% and 99% for agreement with algorithm 2 and 3 respectively.

5.1.3.2 Tables of Results from Hand and Foot Databases

Figs. 5.11, 5.12, 5.13, 5.14, 5.15 and 5.16 show results for both databases tested (hand database and foot database). For each algorithm, two versions of the images were segmented, the original one and the original with an artificial controlled background added, as described previously. The results for the database with backgrounds added show the combined results for all three types of background. These two test databases are referred to in the next tables as HDB and FDBorig for the original databases of hands and feet respectively, and HDBwB and FDBwB for the databases with background added for hands and feet respectively. Note that the active contour level set method was not tested on the databases with backgrounds added due to the large amount of time it would have required. However, the performance for all other methods degraded when IR backgrounds were added so it is expected that the active contour level set method would have followed this trend and would still have been outperformed by most of the other techniques tested.

Figs. 5.11, 5.12 and 5.13 summarize the segmentation results for the hand database. For all figures, ACLS, OTSU, FCM, ANN and CMpEm refer to Otsu's optimal global thresholding,a modified Fuzzy C-Means clustering, Artificial Neural Network classi-

fier and Support Vector Machine classifier, respectively, that were described in section 5.1.1. BF Fusion refers to the Bayesian Framework Fusion with $\alpha = 99\%$ as explained in 5.1.3.1. HDBwB refers to the hand database with backgrounds addeda and HDB refers to the original hand database. Each figure represents the percentage of cor-

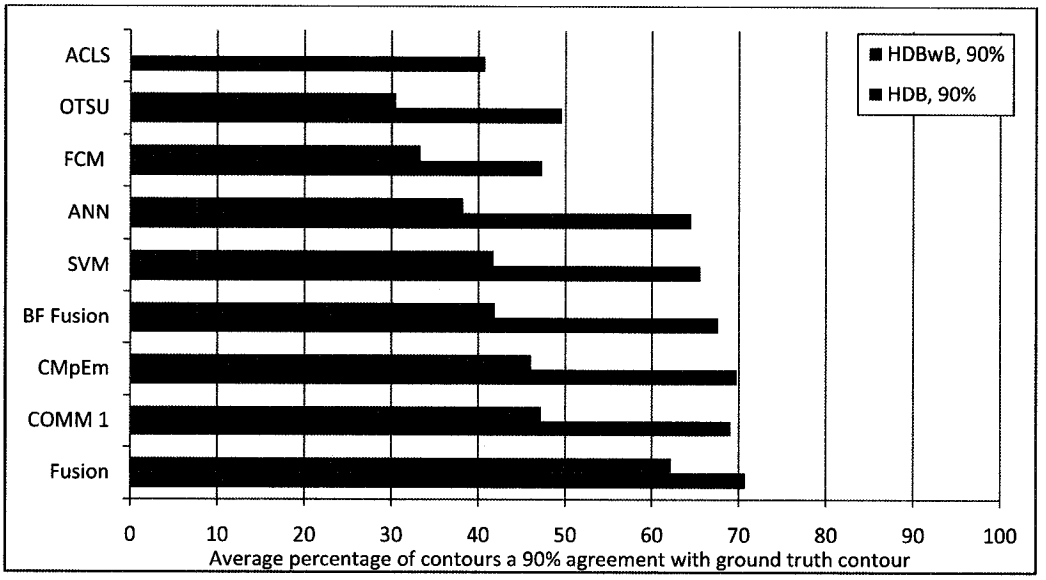


Figure 5.11: Hand database segmentation results at 90% agreement. Fusion: fusion of ANN, SVM, CMpEm and Otsu; COMM1: Best voting committee (ANN+OTSU+CMpEm); .

rectly recovered contours for a given agreement level between calculated and ideal contours. For example, the CMpEm method achieves a 90% agreement between the contour produced and the actual contour for a little over 70% of all the images tested from the original hand database. As the constraint on the quality of the segmented contour is decreased, via the percentage of agreement with the actual contour, the number of segmented contours meeting the quality criterion increases. For example, the CMpEm achieves a 70% agreement between the contour produced and the actual contour for close to 80% of all the images tested from the original hand database.

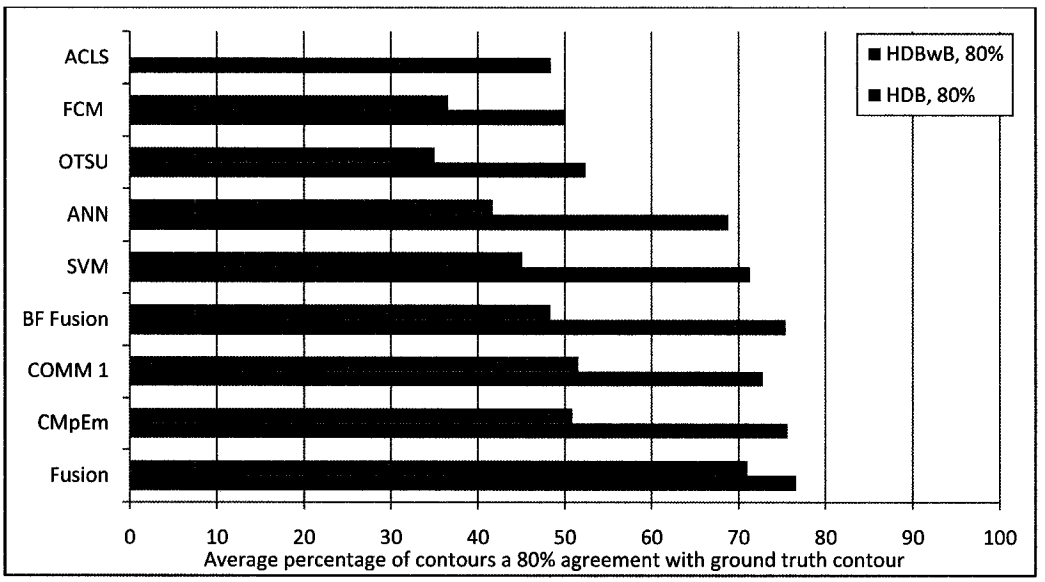


Figure 5.12: Hand database segmentation results at 80% agreement. Fusion: fusion of ANN, SVM, CMpEm and Otsu; COMM1: Best voting committee (ANN+OTSU+CMpEm); BF Fusion: Bayesian Framework Fusion ($\alpha = 99\%$).

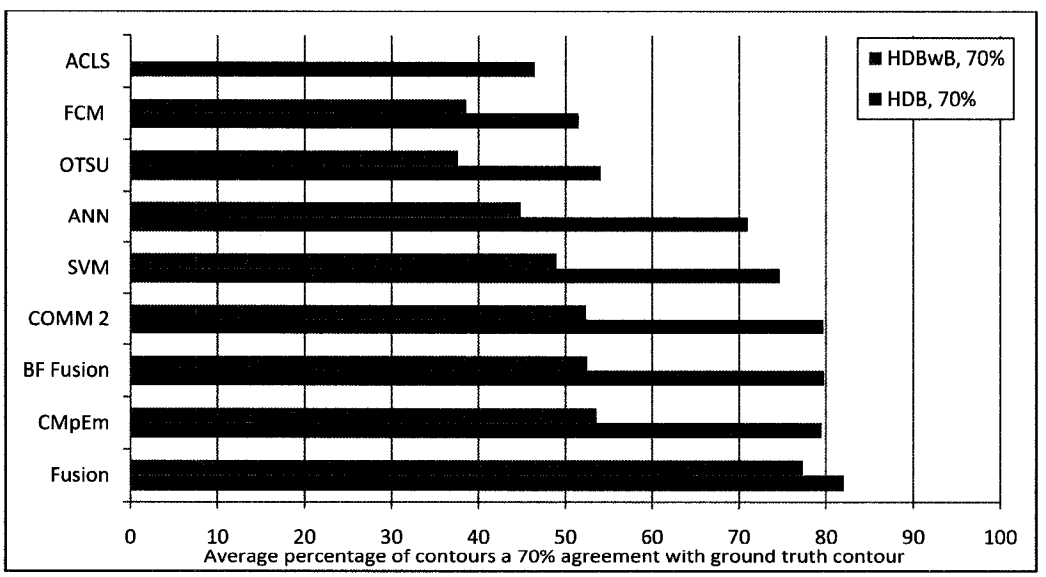


Figure 5.13: Hand database segmentation results at 70% agreement. Fusion: fusion of ANN, SVM, CMpEm and Otsu; COMM2: Best voting committee (ANN+FCM+CMpEm+SVM); BF Fusion: Bayesian Framework Fusion ($\alpha = 99\%$).

The same testing method was applied on a database of infrared images of feet. Results are presented in figures 5.14, 5.15 and 5.16. A missing bar indicates that the algorithm was not tested for that particular database. For all figures, ACLS, OTSU, FCM, ANN and CMpEm refer to Otsu's optimal global thresholding,a modified Fuzzy C-Means clustering, Artificial Neural Network classifier and Support Vector Machine classifier, respectively, that were described in section 5.1.1. BF Fusion refers to the Bayesian Framework Fusion with $\alpha = 99\%$ as explained in 5.1.3.1. FDBwB refers to the foot database with backgrounds added; FDBnB refers to the foot database with background removed and FDBorig refers to the original foot database.

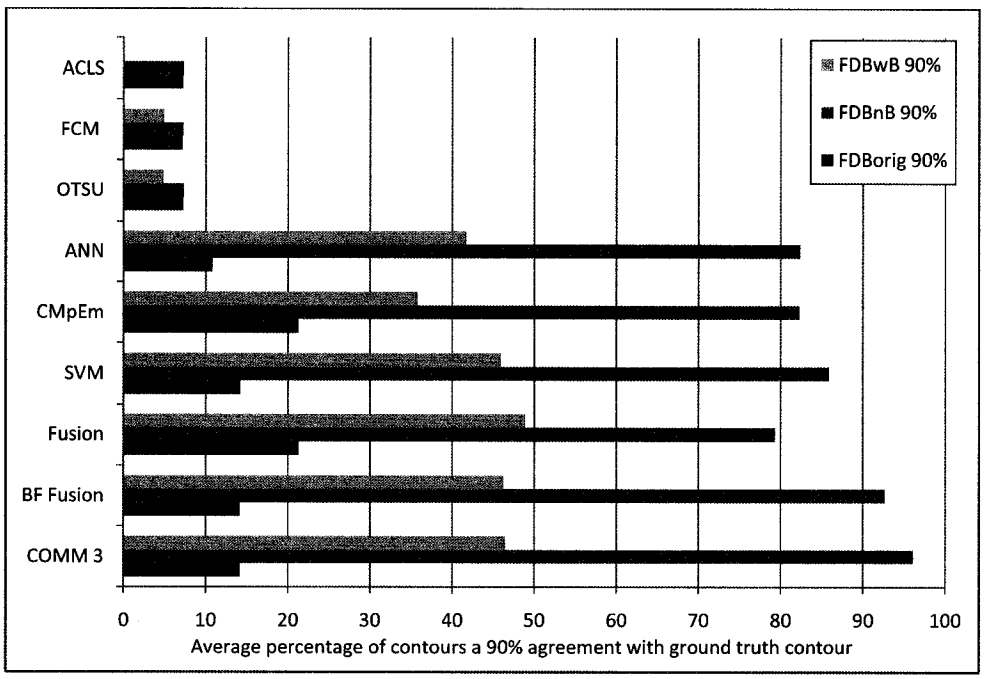


Figure 5.14: Foot database segmentation results at 90% agreement. Fusion: fusion of ANN, SVM, CMpEm and Otsu; COMM3: Best voting committee (ANN+Otsu+CMpEm+SVM).

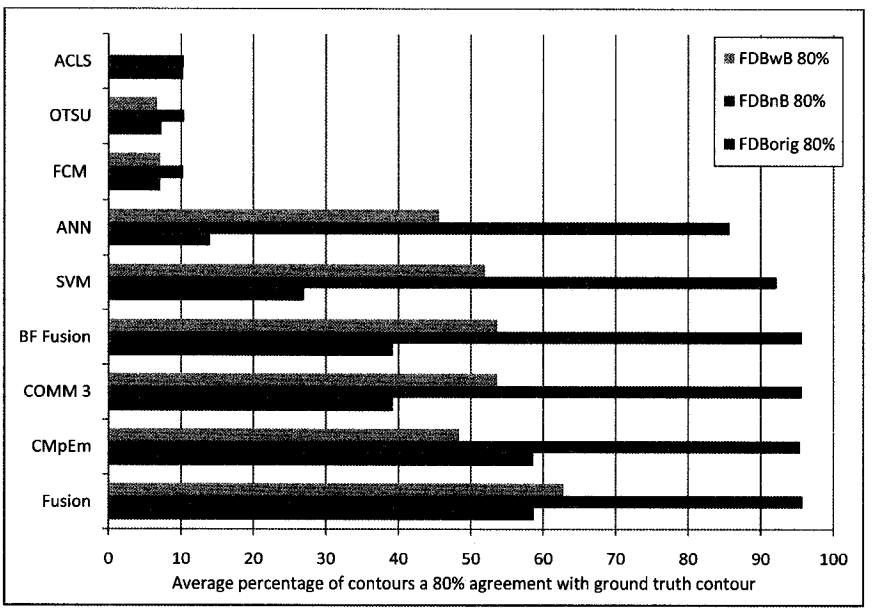


Figure 5.15: Foot database segmentation results at 80% agreement. Fusion: fusion of ANN, SVM, CMpEm and Otsu; COMM3: Best voting committee (ANN+Otsu+CMpEm+SVM).

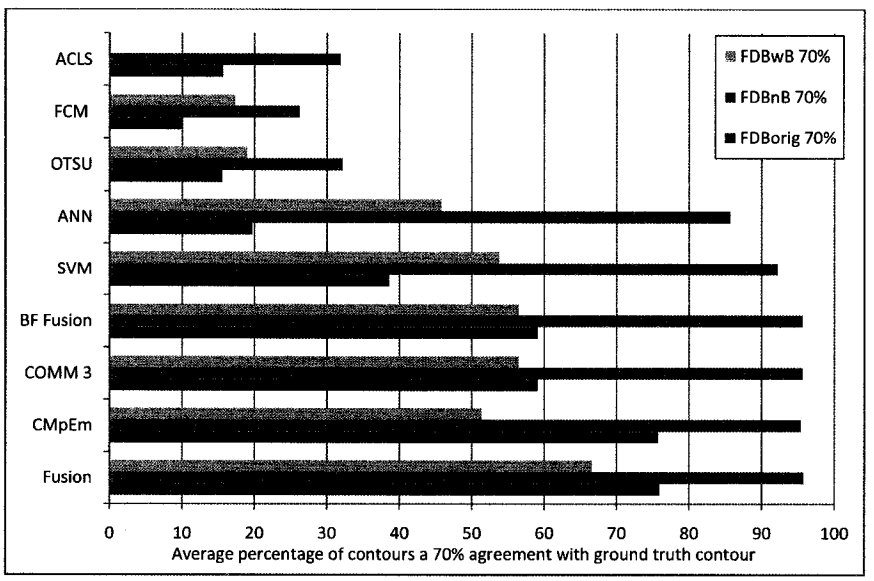
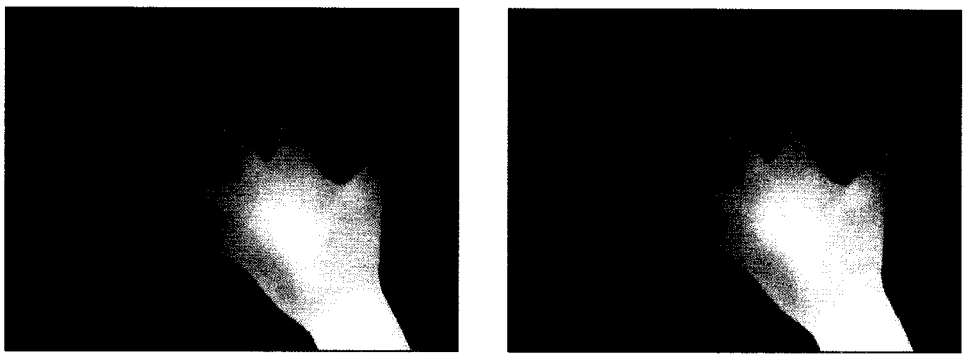


Figure 5.16: Foot database segmentation results at 70% agreement. Fusion: fusion of ANN, SVM, CMpEm and Otsu; COMM3: Best voting committee (ANN+Otsu+CMpEm+SVM).

Figs. 5.17 and 5.18 show examples of segmented hands from the six basic algorithms for two typically difficult images, where the temperature distributions of the hand and background overlap and where the temperature distribution of the background is not necessarily uniform either. These are the original infrared images. The green contour shows the ground truth segmentation while the red contour shows the segmented region found by one of the algorithms tested.

Similarly for the foot database, figure 5.19 shows an example of segmented foot.

Finally, Fig. 5.20 shows an example of the improvement over individual algorithm when using the rule-based morphological fusion.

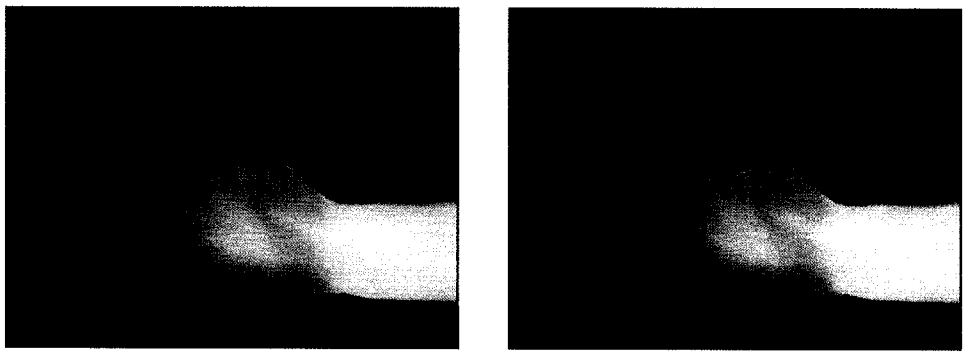


(a) Otsu. (b) FCM.

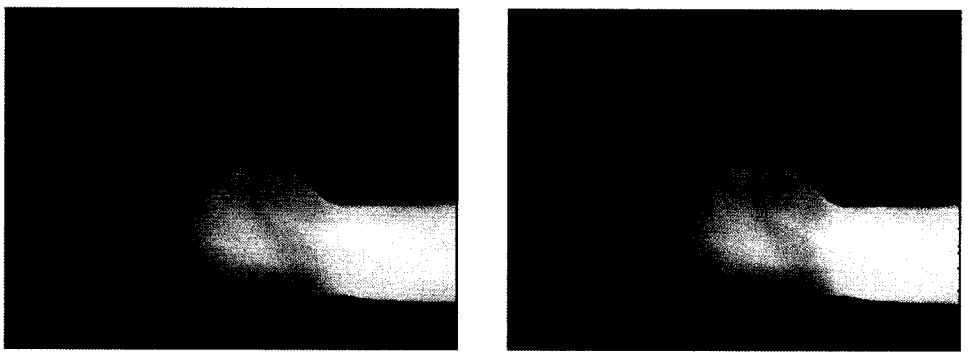
(c) CMpEm. (d) ACLS.

(e) ANN. (f) SVM.

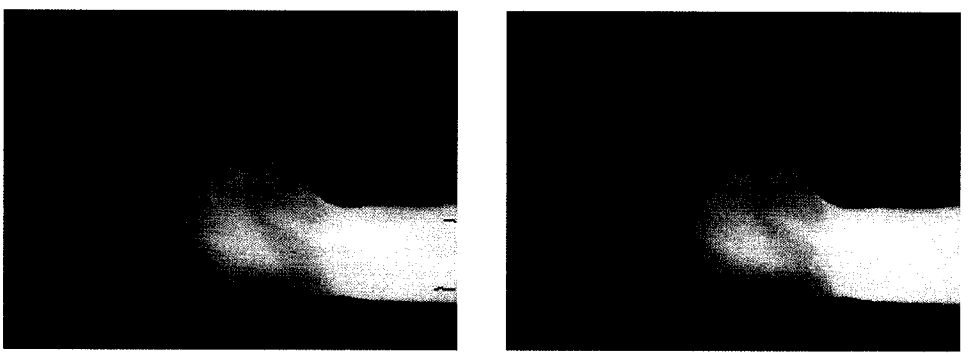
Figure 5.17: Example of segmented real IR image from all algorithms tested. a: Otsu thresholding method (57.4% agreement with ground truth). b: FCM method (50.8% agreement with ground truth). c: cued morphological processing of edge map method (99.8% agreement with ground truth). d: Active contour level set method (55.3% agreement with ground truth). e: ANN classification of labelled edge components (99.8% of agreement between contour and ground truth). f: SVM classification of labelled edge components (99.8% of agreement between contour and ground truth)



(a) Otsu. (b) FCM.

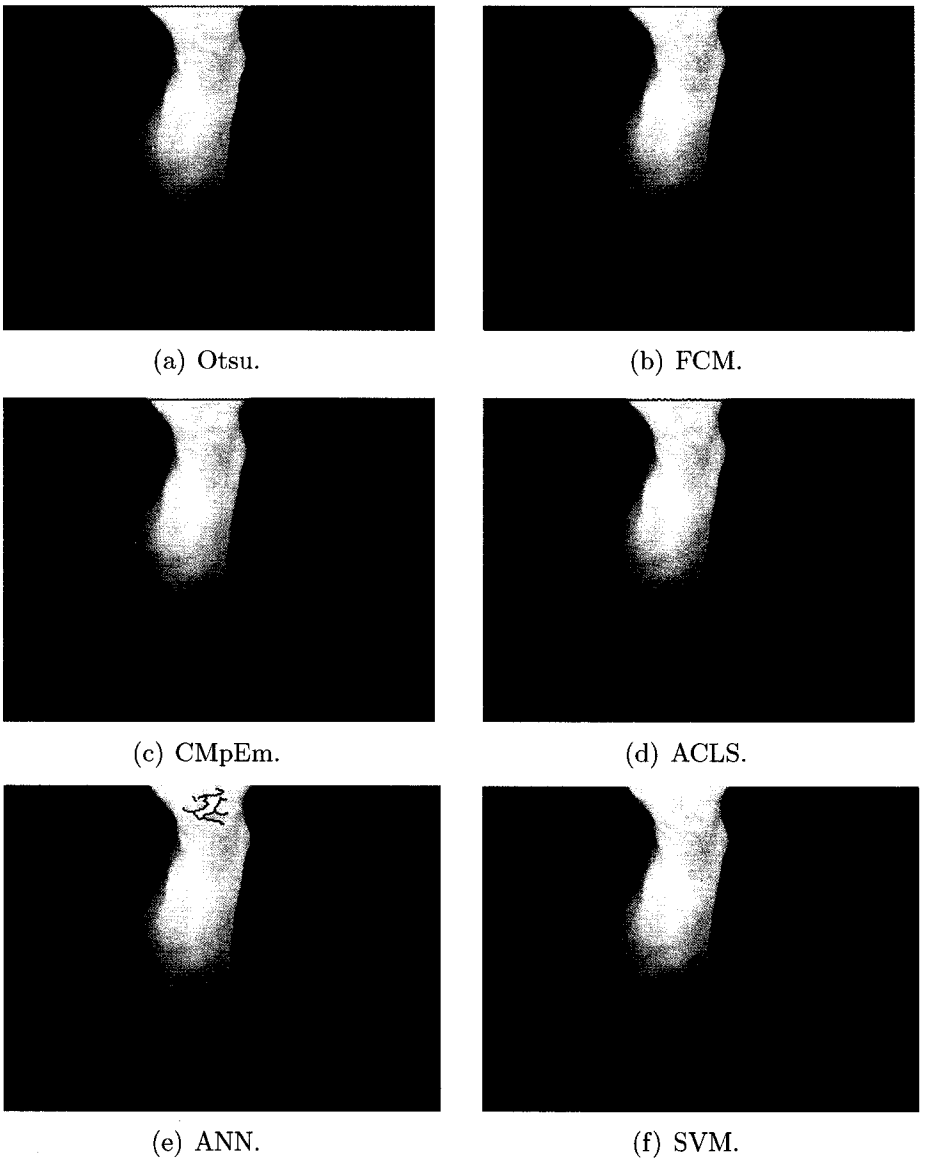


(c) CMpEm. (d) ACLS.



(e) ANN. (f) SVM.

Figure 5.18: Example of segmented real IR image from all algorithms tested. a: Otsu thresholding method (21.5% agreement with ground truth). b: FCM method (20.2% agreement with ground truth). c: cued morphological processing of edge map method (89.6% agreement with ground truth). d: Active contour level set method (20.9% agreement with ground truth).e: ANN classification of labelled edge components (72.8% of agreement between contour and ground truth). f: SVM classification of labelled edge components (52% of agreement between contour and ground truth).

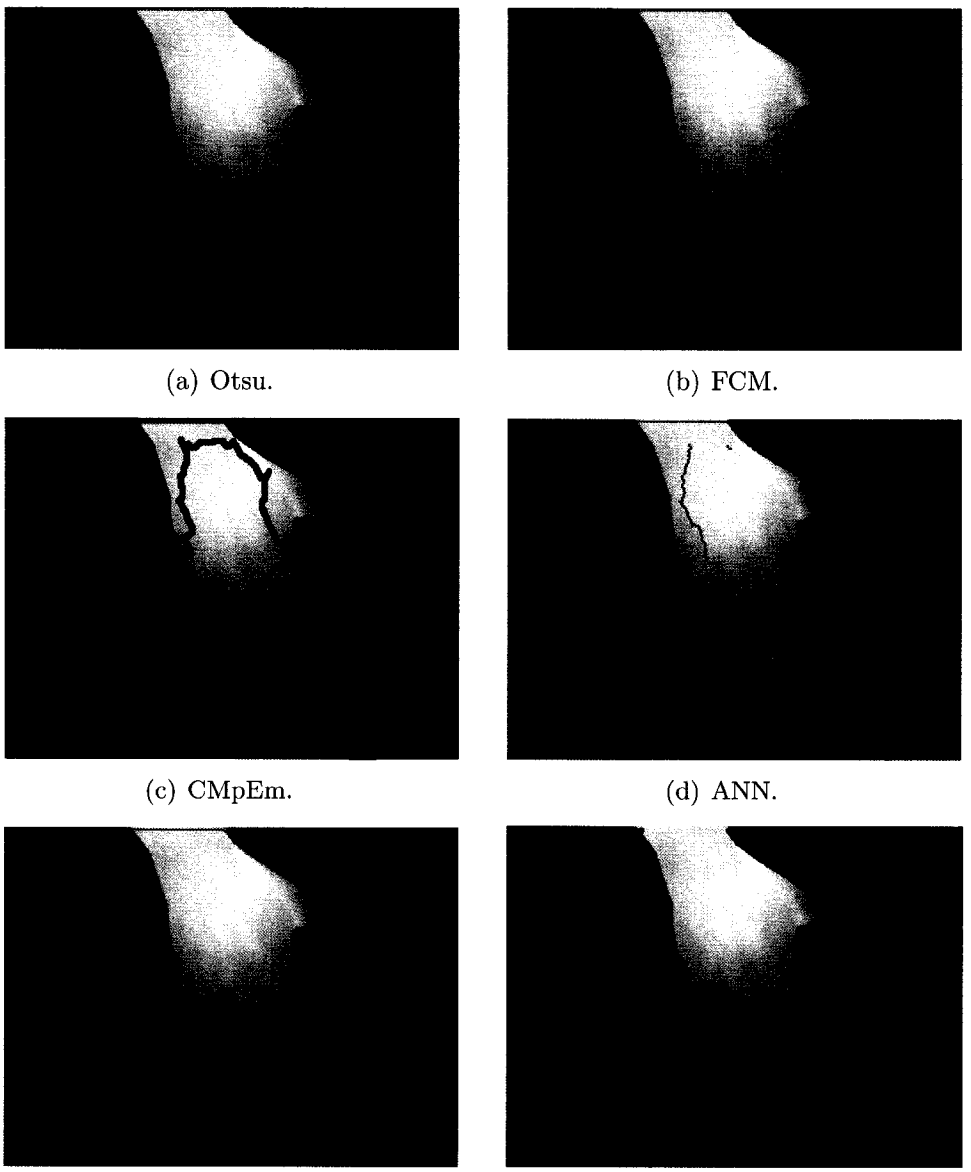


(a) Otsu. (b) FCM.

(c) CMpEm. (d) ACLS.

(e) ANN. (f) SVM.

Figure 5.19: Example of segmented real IR image from all algorithms tested. a: Otsu thresholding method (52% agreement with ground truth). b: FCM method (50% agreement with ground truth). c: cued morphological processing of edge map method (68% agreement with ground truth). d: Active contour level set method (51% agreement with ground truth). e: ANN classification of labeled edge components segmentation method (77% of agreement between contour and ground truth). f: SVM classification of labeled edge components segmentation method (90% of agreement between contour and ground truth).



(a) Otsu. (b) FCM.

(c) CMpEm. (d) ANN.

(e) SVM. (f) Rule-based, morphological fusion.

Figure 5.20: Comparison of algorithms with one of the fusion technique (rule-based morphological fusion).

5.1.4 Discussion

From figures 5.11, 5.12, 5.13, 5.14, 5.15 and 5.16 three groups of algorithms emerge based on their overall performance on the original hand database and the one with backgrounds added. The group performing the worst overall is the one consisting of Otsu thresholding, FCM clustering and active contour level set method. The explanation for their poor performance is exemplified in figures 5.17, 5.18 and 5.19 where they give a region where the intensity is more or less uniform, missing out on the regions where the intensity is close or less than that of the background. This is expected from the Otsu thresholding method, since it uses a global albeit optimal threshold for the entire image. In the case of the FCM algorithm, increasing the number of clusters may in some cases produce better results but the optimal number of clusters is hard to determine and regions whose intensities are too close to that of the background still would not be recovered. For the active contour level set method, the poor performance is in part due to the stopping criterion for the curve evolution, that is hard to set for a wide range of images. Indeed, the evolution of the active contour typically stops at boundary or edges and although it usually works well when small gaps are present in the contour of the object, it typically fails when large portions of the contour have weak or very weak edges. Adding edge orientation information as constraints in the total energy constraining the active contour as suggested by [52] works better with weak edges as long as the intensity of the object is mostly uniform. However, if large intensity variations exist, it fails as well.

The second group of methods consist of the ANN and SVM classification of edge components. The two classifiers work well when the noise level is low to medium and the contrast is high to medium. However, the performance tends to degrade for noisy,

low contrast images. This is partially due to the dependence on the edge detection process. If edge components from the contour are fused with edge components from the background, the two classifiers are unable to recover the edge component parts belonging to the contour.

Although both ANN and SVM classification of edge components require a well labelled training database, the generalization capabilities of both methods was better than expected. They were able to perform well on the foot databases although they were trained on edge components from the training hand database only. For instance, in figure 5.14, the SVM performed slightly better than the CMpEm approach on the foot database with background removed and on the foot database with backgrounds added. On the original foot database however, the CMpEm approach still performed better. The explanation lies with the nature of the contours that were investigated. Indeed, the classification is not done on the shape of the object itself, unless there is only one single clean edge component for the entire contour, but on the edge components that make up the contour. As such, edge components from the contour of smooth objects are often similar. It means that given a more general training database of smooth shapes that are not too dissimilar, the technique based on classification of edge components will be flexible enough to be able to segment different shapes.

One major drawback of the ANN and SVM segmentation methods in addition to the training required is their inability to produce a closed contour in many cases, which means a post-processing step would be required in order for a landmark detection scheme to be performed successfully. As a result, the two methods are perhaps better used in a fusion scheme or in combination with other approaches.

The last group consists of the CMpEm method and the various fusion schemes.

Overall, the CMpEm approach outperformed the approaches from the first two groups and sometimes even fusion schemes. The difference is more noticeable in more difficult cases such as in figures 5.18 and 5.19. The comparison of fusion approaches yielded some interesting results. In most situations, the fusion approach based on agreement rules and morphological processing performed the best out of all the other techniques. The fusion approach based on a Bayesian framework and the weighted majority voting fusion scheme performed globally the same. In most cases they also outperformed the individual techniques, which was expected. Their lower performance with respect to the rule-based morphological fusion scheme can be explained by the fact that they mostly assume that all individual techniques to be combined are independent from each other, which is not likely true in our case.

In terms of computational complexity, the least complex is the Otsu thresholding method. The complexity of the FCM and CMpEm is roughly equivalent as well, for a maximum number of iteration of 100 for the FCM clustering. The active contour level set approach implemented is the most complex of the individual segmentation techniques compared. The number of resolution levels, the number of scales used and the allowed number of iterations all influence the amount of computations required. However, the number of computations was on average two orders of magnitude greater than that of FCM and CMpEm methods. Table 5.9 shows the average time required to process an image, based on an average over 4800 images (300 images for the active contour level set method).

5.2 Selected Results on Breast Segmentation

A precise determination of breast boundaries is not within the scope of this thesis

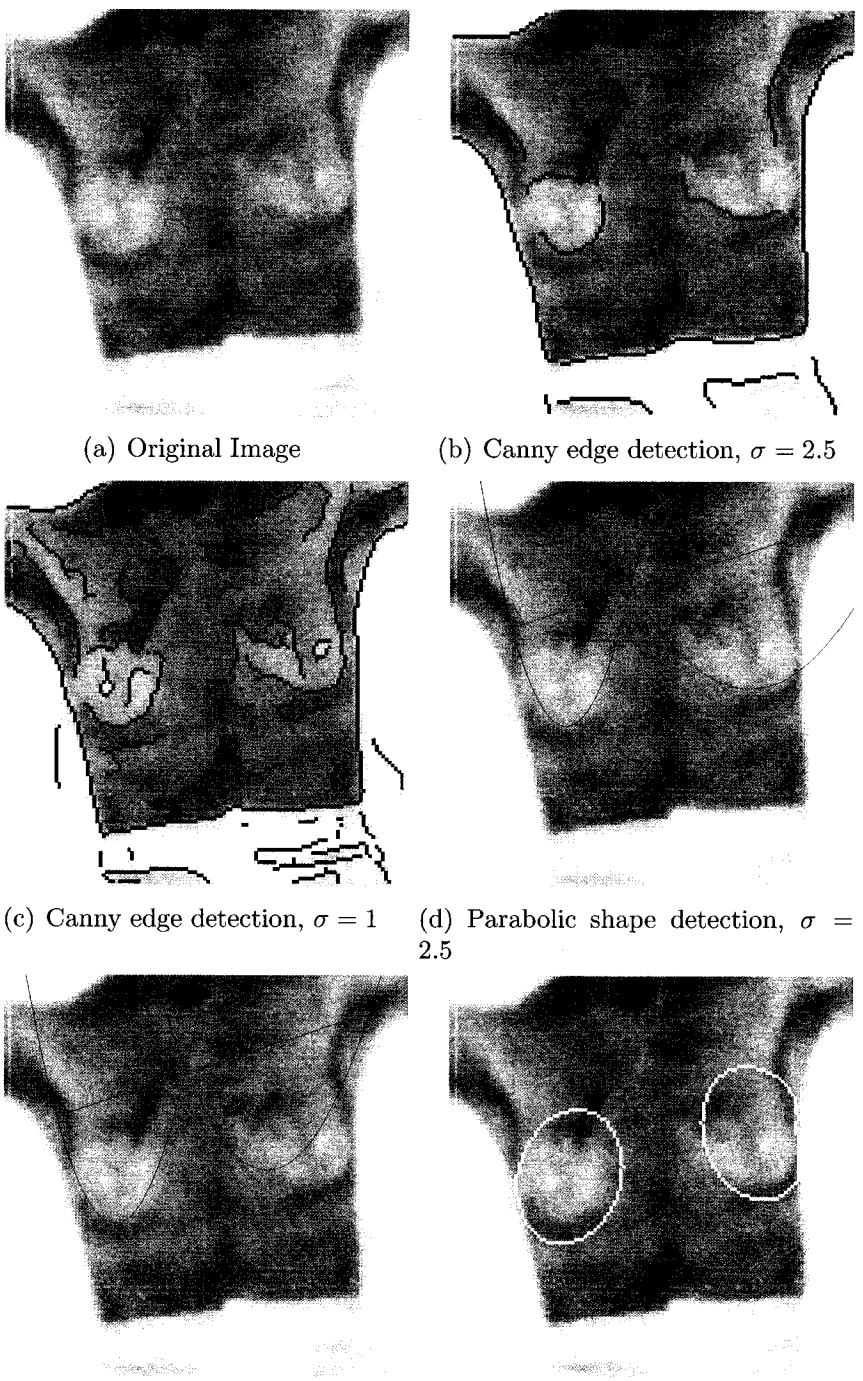
Table 5.9: Average computation time in seconds based on an average over 4800 images (300 for the active contour level set method). Computations were performed on an Intel Centrino 1.66GHz Core Duo processor.

Method	Time (s)
ANN	1.323
FCM	1.953
Otsu	0.063
CMpEm	2.203
SVM	1.363
ACLS	152.045

since the determination of the breast area is not well defined on IR images within additional information (fiducial markers for instance). As a result, no ground truth segmentation is provided in the following but rather qualitative results based on the most probable area comprising the breasts.

Following the procedure outlined in Section 3.3.5, the inclusion factor was determined for each image using a linear combination of the normalized distances from nipples to armpits, from nipples to chin and from the torso width. The coefficients for the linear combination were estimated by regression analysis on 10 representative breast images with different breast sizes and shapes. The resulting inclusion factors ranged from 0.8 to 1.8, depending on the size of the breast.

Figs. 5.21 and 5.22 compare the results of a breast segmentation using the proposed breast segmentation method versus Qi *et al.*'s method [35]. The first image (Fig. 5.21) is from a first generation IR camera (AGA thermovision 680) and was acquired by Monique Frize and her team at the Breast Clinic of the Moncton Hospital, NB. The second image (Fig. 5.22) is from a second generation camera and was acquired by Dr. Keyserlingk and his team at the Ville-Marie Breast Clinic in



(a) Original Image

(b) Canny edge detection, $\sigma = 2.5$

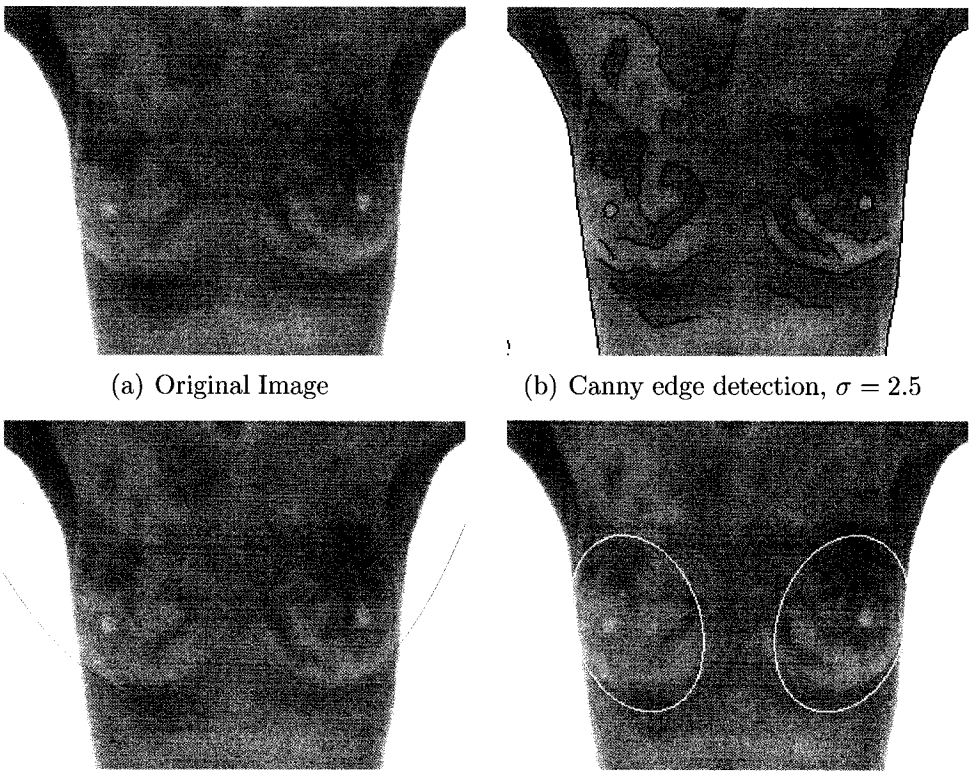
(c) Canny edge detection, $\sigma = 1$

(d) Parabolic shape detection, $\sigma = 2.5$

(e) Parabolic shape detection, $\sigma = 1$

(f) Our segmentation method

Figure 5.21: Segmentation result for a thermal image of the breasts.



(a) Original Image (b) Canny edge detection, $\sigma = 2.5$

(c) Parabolic shape detection for $\sigma = 2.5$ (d) Our segmentation method

Figure 5.22: Segmentation result for a thermal image of the breasts.

Montreal, QC.

The edges returned by a Canny edge detector are shown as well. Qi *et al.* suggested to use a value of $\sigma = 2.5$ for the Gaussian function of the Canny edge detector. Note that the edge components corresponding to the lower contours of the breast have to be selected manually in Qi *et al.*'s method. The quality of the resulting segmentation highly depends on the choice of σ as can be seen in figure 5.21. This is problematic since the selection of σ is not trivial and vary for different images. In addition, an improper choice of σ makes the subsequent parabolic shape detection very difficult, for instance when $\sigma = 1$ in figure 5.21 where the number of edges produced is significantly higher and choosing which edge component to fit with the parabolic

shape is not trivial. Variations of σ change the number and often the shapes of the edge components produced, which in turn change significantly the contour produced by the parabolic shape detection proposed by Qi *et al.*, since the fitting is performed on connected edge components. However, it does not affect our segmentation method significantly, since the value of σ was chosen adaptively by fixing the density of the edge map instead.

Using a Hough transform for detecting the parabolic shape from the lower breast contours, as proposed by Qi *et al.*, was extremely slow. A much faster solution is to use a fast least squares fitting algorithm to fit the parabola from the edges belonging to the lower contour. Each pixel belonging to the edge component from the lower contour is considered to be part of the parabolic curve and the best fitting parabola is found using a least squares fitting algorithm. This was the method used in the following figures. The increase in speed was of several order of magnitudes.

As can be seen, Qi *et al.*'s approach tends to overestimate the breast area and is susceptible to the quality of the edge components returned. Our approach is more robust to the edge components returned although the automated choice of the inclusion factor tends to underestimate the breast area in some cases.

Chapter 6

Landmark Localization Results

This chapter presents results obtained from the ROI identification stage of our proposed architecture.

6.1 Algorithms Tested and Parameters Used

The approach introduced in Section 3.4 was tested on a database of IR images of the arms and hands. The approach was also tested on IR a few images of the back that were acquired using the protocol described for the computer user database. The smoothing parameter p in Eq. 3.11 for the spline fitting was set to 0.1. This choice of parameter puts much emphasis on how smooth the spline has to be, as opposed to how close it should fit the data points from the contours. The rationale for choosing a small value for the smoothing parameter stems from the fact that although the contour of most body regions are rather smooth, the imaging process as well as the segmentation and contour extraction processes result in a spatial discretization that is far from smooth. To avoid the pitfall of overfitting the binary contour discretization

artifacts, it is necessary to choose a lower smoothing parameter. The smoothing parameter may need to be adjusted depending on the type of object considered and the imaging process used prior to the landmark identification.

The minimum peak height for the curvature was determined by the upper 35% percentile of the distribution of high curvature points. This distribution is generated from the normalized absolute value of the curvature plot. If the absolute normalized curvature points are normally distributed, the 35% percentile would be approximately equivalent to keeping a peak if it is more than a standard deviation away from the mean peak value. Choosing a threshold adapted to the distribution of curvature peaks is more robust than a global hard threshold on the peak height to different types of contours that can be encountered depending on the body region or other object of interest.

The minimum distance between peaks as well as the clustering distance for the final stage were set to 5 pixels. This distance can be estimated from the distance from the camera to the object and the approximate size of the object, which give an idea of how close relevant anatomical features would be.

Our algorithm was compared to three different approaches used for landmark point detection. These algorithms are:

- A Curvature Scale-Space (CSS) approach proposed by He *et al.* [69].
- A corner and edge detector based on phase congruency introduced by Kovesi [67].
- An improved version of the original corner detector from Harris *et al.* [65], using a more precise calculation of the derivatives. This was proposed by Schmid *et al.* [159].

As described in Section 2.2.3, each method differs significantly in their approach to extracting corners or high-curvature points. They have all been shown to outperform other techniques from the literature and are a good representation of the state-of-the-art for landmark extraction in 2D images [67, 68, 159]. As a result, they are well-suited for comparison with our approach described in Section 3.4.

For each of the three methods tested, a number of parameters are required to be set. For Kovesi's phase congruency method, the default parameters were used, based on the author's own recommendations included in his code. The parameters are listed in Table 6.1. Parameters for the curvature scale-space method is given in Table 6.2 . Harris' corner detector gave best results for a standard deviation of 1 for the smoothing Gaussian. In addition, the search for landmarks was constrained to the contour either explicitly for our approach and for the CSS method by providing a perfect contour, or implicitly for Kovesi's phase congruency method and for the Harris corner detector by discarding landmark points that were not on the contour.

In addition, Kovesi's and Harris' methods require an additional non-maxima suppression step to extract the final landmark points. The radius of the region considered in non-maxima suppression was set to 2 pixels and the threshold was varied until optimal results were obtained compared to the ground truth location of the landmark points or until a cap on the number of landmark points was reached. In order to enforce the cap in the case of the CSS method, the parameter C in Table 6.2 was varied. The angle θ was kept at $180\,^{\circ}$, otherwise the anatomical landmark points of interest were not detected consistently.

Table 6.1: Parameters used for Kovesi's phase congruency edge and corner detector.

Number of wavelet scales	4
Number of filter orientations	6
Wavelength of smallest scale filter	3
Scaling factor between successive filters	2.1
Ratio of the standard deviation of the Gaussian describing the log Gabor filter's transfer function in the frequency domain to the filter centre frequency	0.55
Ratio of angular interval between filter orientations and the standard deviation of the angular Gaussian function used to construct filters in the freq. plane.	1.2
No of standard deviations of the noise energy beyond the mean at which we set the noise threshold point.	2
fractional measure of frequency spread below which phase congruency values get penalized.	0.5
Sharpness of the transition in the sigmoid function used to weight phase congruency for frequency spread.	10

Table 6.2: Parameters used for the curvature scale-space method.

C, Minimum ratio of major axis to minor axis of an ellipse, whose vertex could be detected as a corner by proposed detector	1.5
θ Maximum obtuse angle that a corner can have when it is detected as a true corner	180
Standard deviation of the Gaussian filter when computing curvature	7

6.2 Landmark Localization Distance Results

Each algorithm was tested on 273 hands and 250 arms in various orientations, from the piano database. Ground truth landmark points were manually extracted by an experienced thermographer, with the help of anatomical and anthropometric templates. They consisted of the radial and ulnar styloid (wrist junction), the fingertips, the interdigital web points (between fingers) and the humeral lateral and medial

epicondyle. Results are compiled in Table 6.3. Values represent the mean distance (in mm) to the ground truth landmarks. Values in bracket are the standard deviations of the distance. For non-specific algorithms such as CSS, phase congruency corner detector and Harris corner detector, the distance was calculated to the landmark closest to the sought location. This is advantageous to the other algorithms tested since they tend to produce a great amount of landmark points for their best parameter settings. As a result, we also show their performance when a maximum number of landmarks is fixed. Iterative increase of the threshold in the non-maxima suppression post-processing step was used to achieve the desired number of landmarks in that case. Fixing the upper cap on the landmark points strictly to the specific landmark points sought for the phase congruency, CSS and Harris detector fails because these methods are intrisically non-specific in the type of landmark points they produce. The upper cap was determined in such manner as to maximize the performance of all algorithms while keeping close to the number of landmark points sought.

Overall our approach performs well compared to other methods tested, giving good localization for all landmarks. The performance of the other algorithms is highly correlated to the number of landmark points returned. If the number of landmark points is capped, their performance drops significantly. This suggests that their good performance in a best parameter scenario is in fact due to the large number of landmarks found, some of which are bound to be close to the ground truth landmarks. In other words, their number of false positives is significantly higher than with our approach. The proposed approach is also more robust with respect to the type and location of landmark points as can be seen in Table 6.3 for the arm landmark points (humeral epicondyle), which represents a very subtle change of curvature on an IR image of the

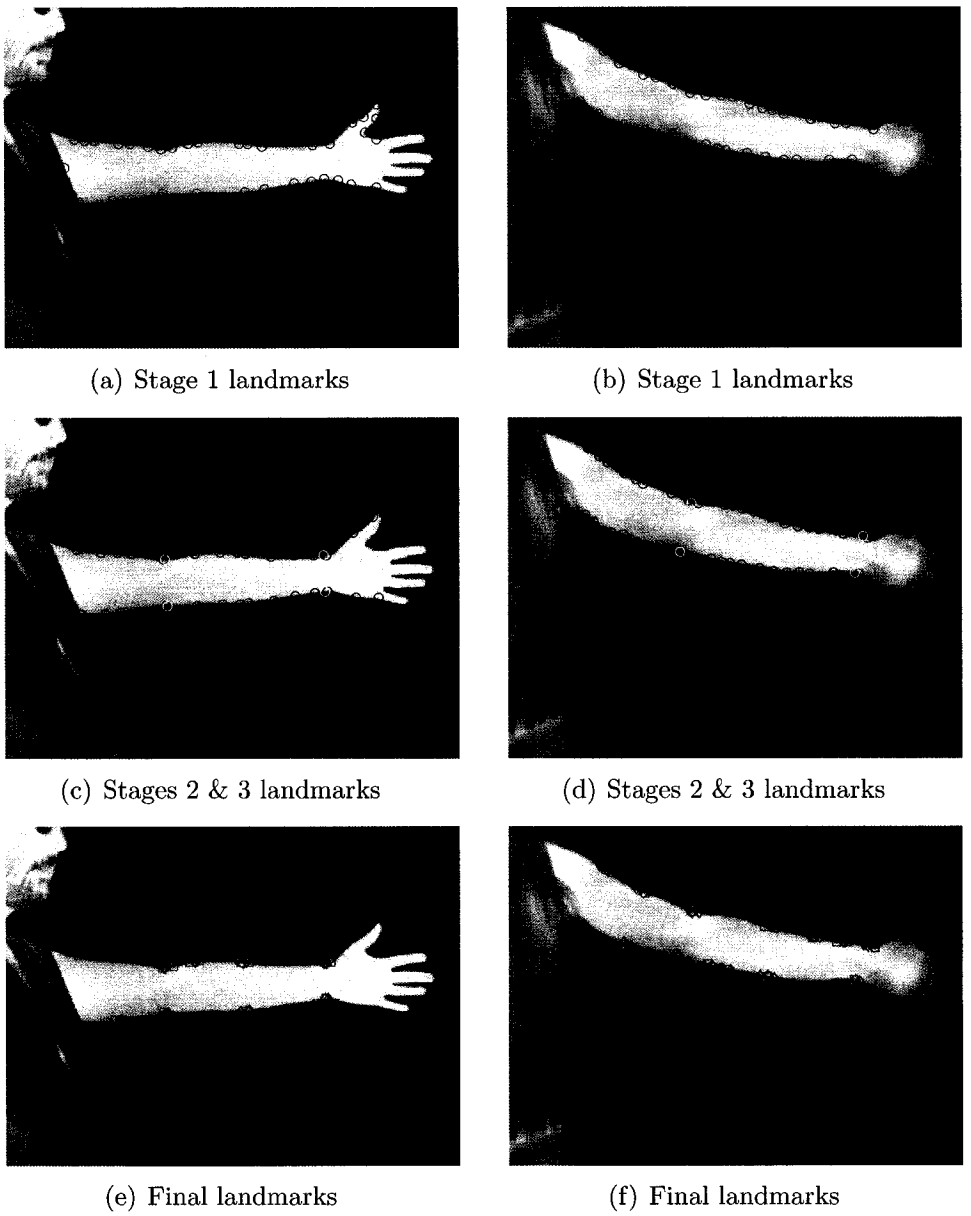
Table 6.3: Mean distance (standard deviation) to ground truth landmarks (in mm) for our approach and both versions of the three other algorithms tested. Legend is as follows: Prop. Appr., proposed approach; CSS best: Curvature scale-space (CSS) with best threshold values, 65 landmarks; CSS cap. (capped): CSS with imposed number of landmarks fixed at 15; Phase cong. best: Kovesis phase congruency with best threshold values, 100 landmarks; Phase cong. cap.: Kovesis phase congruency with number of landmarks fixed at 20; Harris best: Harris corner detector with best parameter values, 40 landmarks; Harris cap.: Harris corner detector with number of landmarks fixed to 20 pts.

Anatomical Landmark Point	Prop. Appr.	CSS Best	CSS Cap. (15)	Phase cong. Best	Phase cong. Cap. (20)	Harris Best	Harris Cap. (20)
Radial Styloid (wrist)	7.2 (6.4)	4.5 (4.5)	21.7 (20)	2.5 (2.1)	8.4 (10)	12.5 (20)	20.3 (26)
Ulnar Styloid (wrist)	5.4 (5.1)	3.6 (3.1)	22.5 (20)	2.2 (1.5)	6.3 (8.3)	4.9 (11)	11.5 (21)
Tip of 1st digit (thumb)	4.8 (2.6)	2.6 (1.5)	2.2 (1.4)	2.5 (2.4)	16.9 (21)	16.8 (23)	19.8 (24)
Tip of 2nd digit (index finger)	3.3 (2.1)	2.5 (1.6)	2.1 (1.5)	3.2 (5.4)	20.0 (22)	23.1 (30)	26.5 (30)
Tip of 3rd digit (middle finger)	2.3 (2.5)	2.8 (1.6)	2.6 (1.8)	3.5 (5.5)	22.3 (23)	26.9 (33)	29.4 (33)
Tip of 4th digit (ring finger)	4.3 (1.9)	2.7 (1.5)	2.3 (2.0)	2.9 (4.5)	19.7 (22)	24.4 (30)	26.0 (30)
Tip of 5th digit (little finger)	5.0 (2.8)	2.1 (1.4)	1.8 (1.3)	2.2 (1.4)	11.5 (18)	16.5 (23)	17.5 (23)
Interdigital web 1 (1st-2nd digit)	3.9 (7.2)	1.8 (1.3)	1.9 (3.8)	1.5 (0.9)	2.1 (3.7)	1.9 (2.2)	2.1 (3.8)
Interdigital web 2 (2nd-3rd digit)	2.7 (5.1)	2.2 (2.3)	2.4 (2.6)	1.7 (1.8)	2.3 (3.6)	2.2 (1.6)	2.2 (1.6)
Interdigital web 3 (3rd-4th digit)	3.1 (5.1)	2.7 (3.9)	2.7 (4.1)	1.7 (1.4)	2.4 (3.6)	2.1 (1.4)	2.1 (1.4)
Interdigital web 4 (4th-5th digit)	4.4 (7.3)	2.1 (2.0)	1.9 (2.2)	1.7 (1.1)	2.7 (4.9)	1.9 (1.1)	2.3 (3.8)
Humeral lateral Epicondyle	1.5 (3.1)	13.1 (8.9)	204.6 (109)	9.3 (5.4)	30.4 (18)	121.2 (114)	393.8 (371)
Humeral medial Epicondyle	1.8 (3.7)	13.3 (11)	194.9 (108)	9.3 (5.2)	30.1 (17)	114.4 (106)	372.0 (346)

arm. This is due in part to the use of the difference of symmetrical contours detailed in 3.4.2.

Figs. 6.1 and 6.2 show the landmarks found at each stage and the final landmarks that are retained by the final clustering. Note how the number of landmark points is reduced to a few key locations. The location of the key landmarks sought for arm and hand images are given in Table 6.3 and shown in figures 3.9, 3.11 and 3.12. The key landmarks on the back include the base of the neck and the shoulder tips. From these key landmark points, anatomical ROIs such as superficial muscles or dermatomes can be recovered.

Fig. 6.3 shows the variation of landmark points returned when varying some of parameters for the Harris, phase congruency and CSS method.

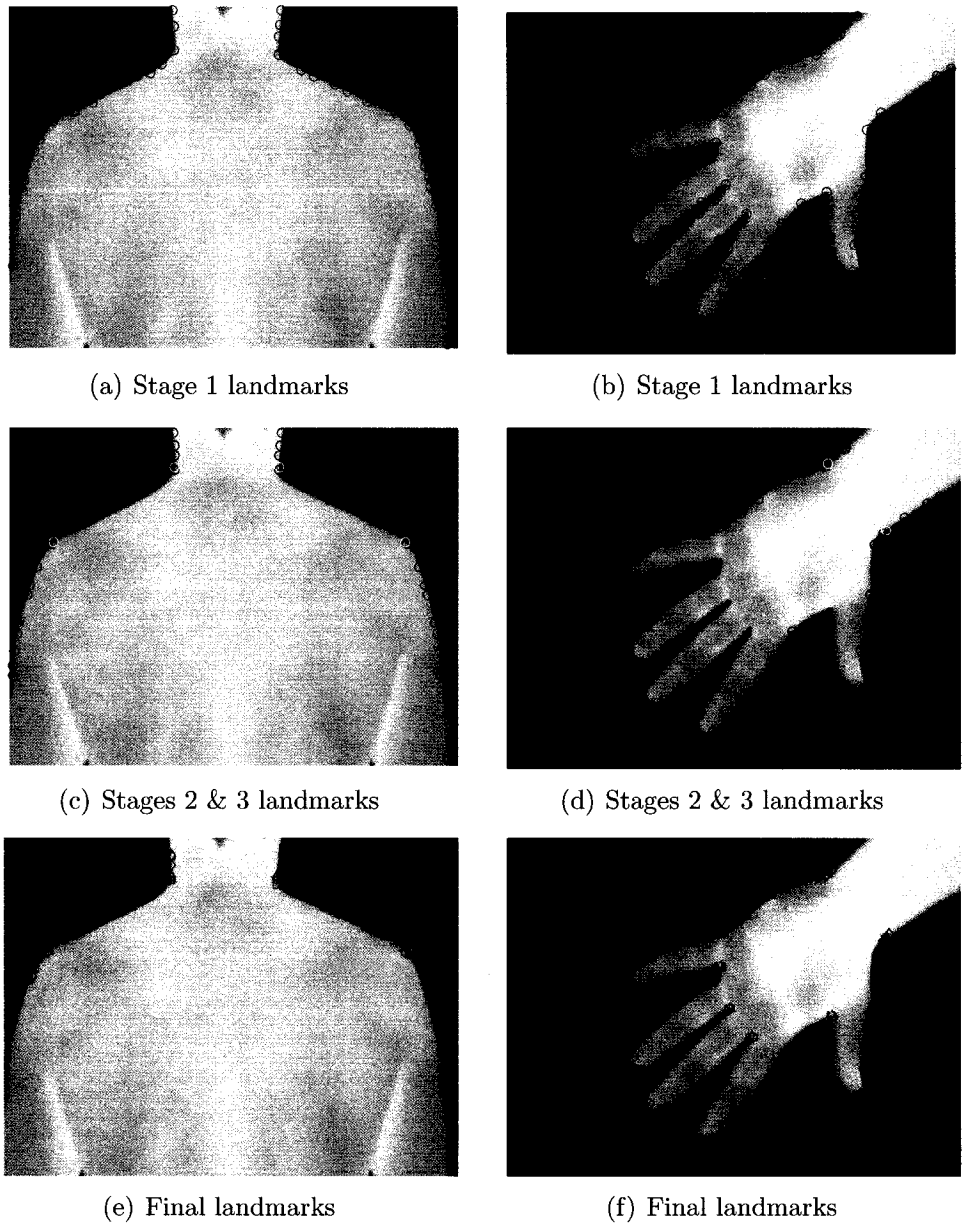


(a) Stage 1 landmarks (b) Stage 1 landmarks

(c) Stages 2 & 3 landmarks (d) Stages 2 & 3 landmarks

(e) Final landmarks (f) Final landmarks

Figure 6.1: a and b: Landmark points returned by the first stage for two infrared images of arms. c and d: Landmark points returned by the second stage (blue points) and region specific constrained landmark points (yellow points). e and f: Landmark points returned after clustering and pair matching of the landmark from previous stages. Star markers indicate the centroid of each remaining cluster. Square markers are remaining landmark after initial clustering; Circle markers are centroid of initial clustering.

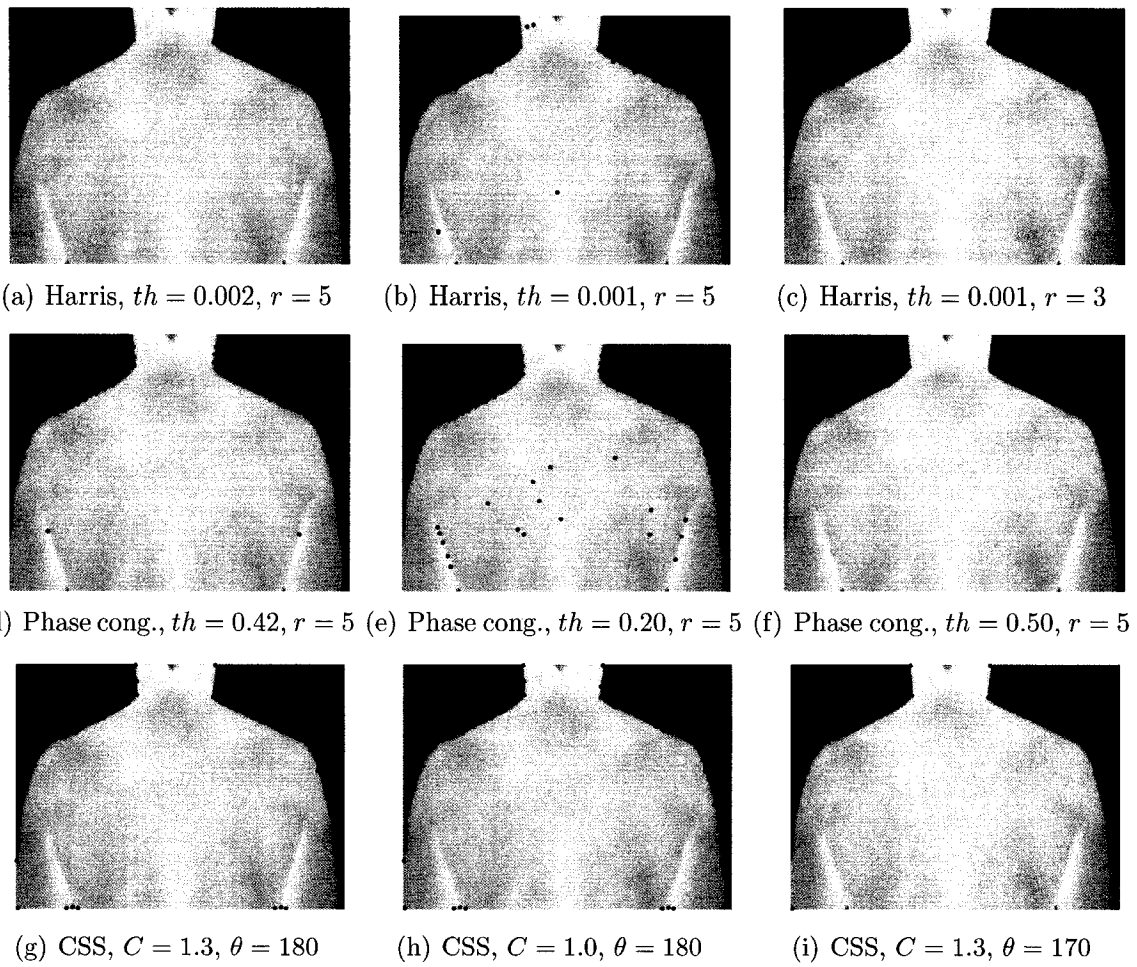


(a) Stage 1 landmarks (b) Stage 1 landmarks

(c) Stages 2 & 3 landmarks (d) Stages 2 & 3 landmarks

(e) Final landmarks (f) Final landmarks

Figure 6.2: a and b: Landmark points returned by the first stage for two infrared images of the back and hand. c and d: Landmark points returned by the second stage (blue points) and region specific constrained landmark points (yellow points). e and f: Landmark points returned after clustering and pair matching of the landmark from previous stages. Star markers indicate the centroid of each remaining cluster. Square markers are remaining landmark after initial clustering; Circle markers are centroid of initial clustering.



(a) Harris, $th = 0.002$, $r = 5$ (b) Harris, $th = 0.001$, $r = 5$ (c) Harris, $th = 0.001$, $r = 3$

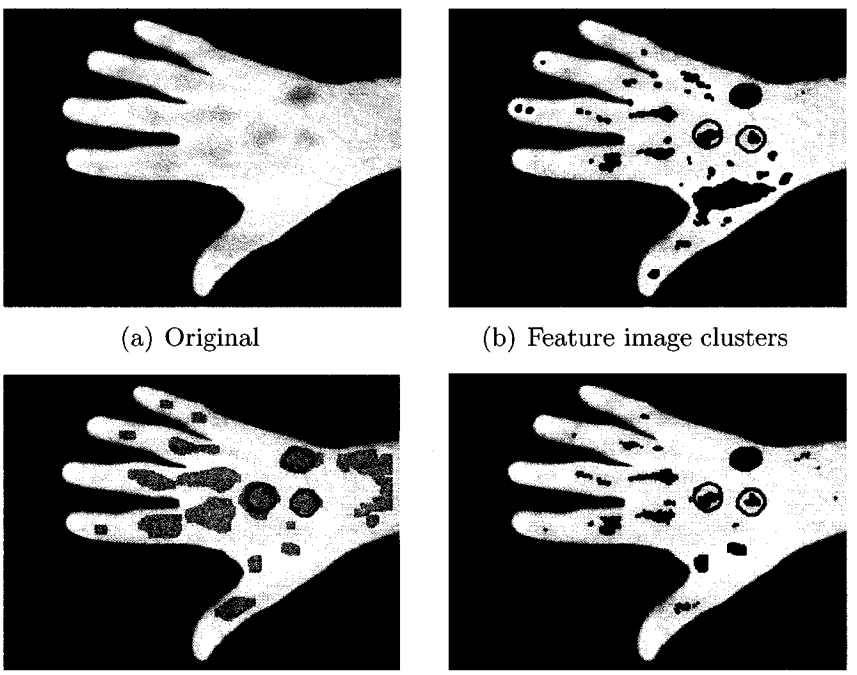
(d) Phase cong., $th = 0.42$, $r = 5$ (e) Phase cong., $th = 0.20$, $r = 5$ (f) Phase cong., $th = 0.50$, $r = 5$

(g) CSS, $C = 1.3$, $\theta = 180$ (h) CSS, $C = 1.0$, $\theta = 180$ (i) CSS, $C = 1.3$, $\theta = 170$

Figure 6.3: a, b and c: Landmark points returned by the Harris corner detector with different thresholds th and radii r for the non-maxima suppression. c, d and e: Landmark points returned by the phase congruency method for different thresholds th and radii r for the non-maxima suppression. f, g and h: Landmark points returned by the curvature scale-space method for different value of C and angles θ.

6.3 Extraction of Clusters of Interest

In order to show how well the extraction of the clusters of interest perform, two images with known anomalies were used. The first one was an IR image of a hand and the second one was an IR image of two knees. The IR image of the knees was acquired at the patient's request and was not part of the ethical approval process decribed in chapter 4. For the IR image of a hand, a cold circular object was pressed on the dorsal part of a right hand sequentially, resulting in three clear circular cold spots on the hand. The location of the circular object was known beforehand and it was therefore a good test to see if each cluster was picked up accordingly. Fig. 6.4 shows the hand with the three artificial anomalies and the clusters of interest found.



(a) Original (b) Feature image clusters

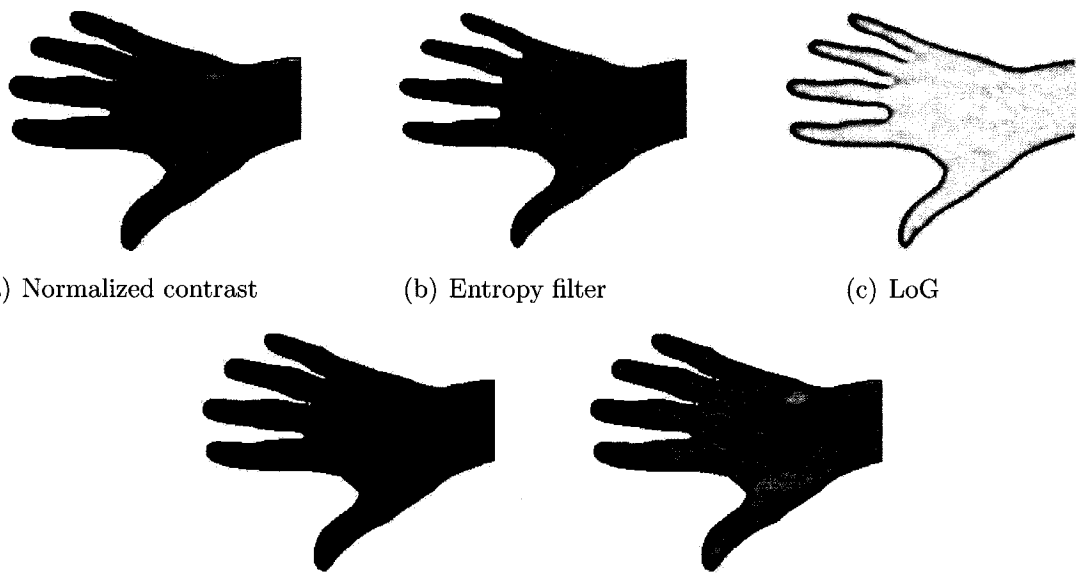
(c) ART clusters (d) Final clusters

Figure 6.4: Extraction of clusters of interest. a: Original image; b: Clusters from thresholded combined feature image. c: Clusters from ART clustering. d: Final clusters. Red circles indicate the location of cold circular objects.

For the Fuzzy ART clustering, the maximum number of class was capped at 10, since a higher number of classes has very little significance for the type of applications considered in this thesis. The initial weights of the ART network were set to 1. The vigilance parameter was 0.75, which represents a compromise between broad class generalization (low vigilance parameter) and highly specific class generalization (vigilance parameter close to 1). The vigilance parameter therefore controls to a great extent the growth of the network. The learning rate was set to 1 to ensure sufficiently fast learning and reasonable time frame for the convergence of the clustering algorithm. The number of epochs was set to 100. The initial bias parameter was set to a small value of 10^{-6}. Note that an optimization procedure to determine optimal parameters was not carried out in this work due to the lack of labelled database of abnormal cases. However, a search for optimal parameters would be desirable for reliable clustering in practical clinical situations.

Both ART and thresholding of the feature image produce the location of three circular object. The ART clustering tended to overestimate the cluster sizes while the feature image thresholding was more conservative. The final image was a compromise between the two. Note the other hot spots in the hand that are visible on the original image. They were also picked up as clusters of interest by our method.

Fig. 6.5 shows the different feature images and the final feature image.

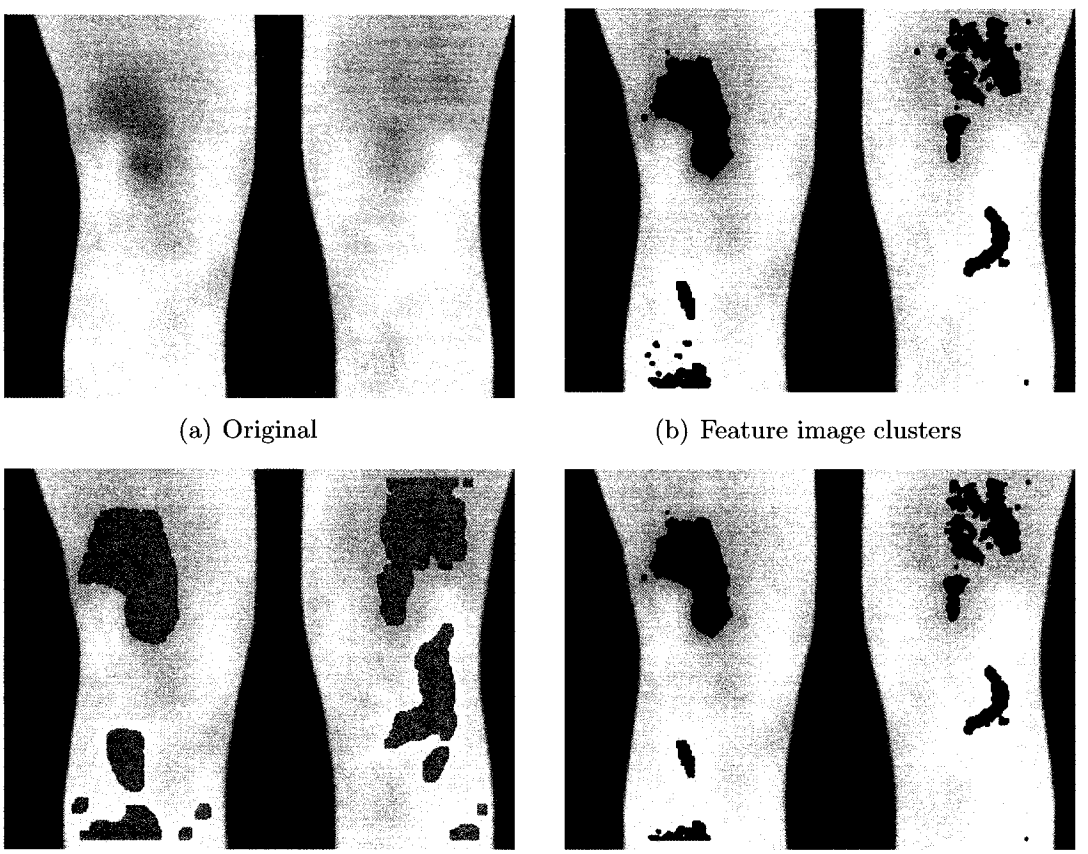


(a) Normalized contrast (b) Entropy filter (c) LoG

(d) Sobel (e) Final feature image

Figure 6.5: Example of feature images. a: Normalized contrast (neighborhood size: 5); b: Entropy filtered image (neighborhood size: 5); c: Laplacian of Gaussian filtered image ($\sigma = 3$, $l_1 = l_2 = 17$); d: Sobel filtered image; e: Final feature image with equal weights for all features.

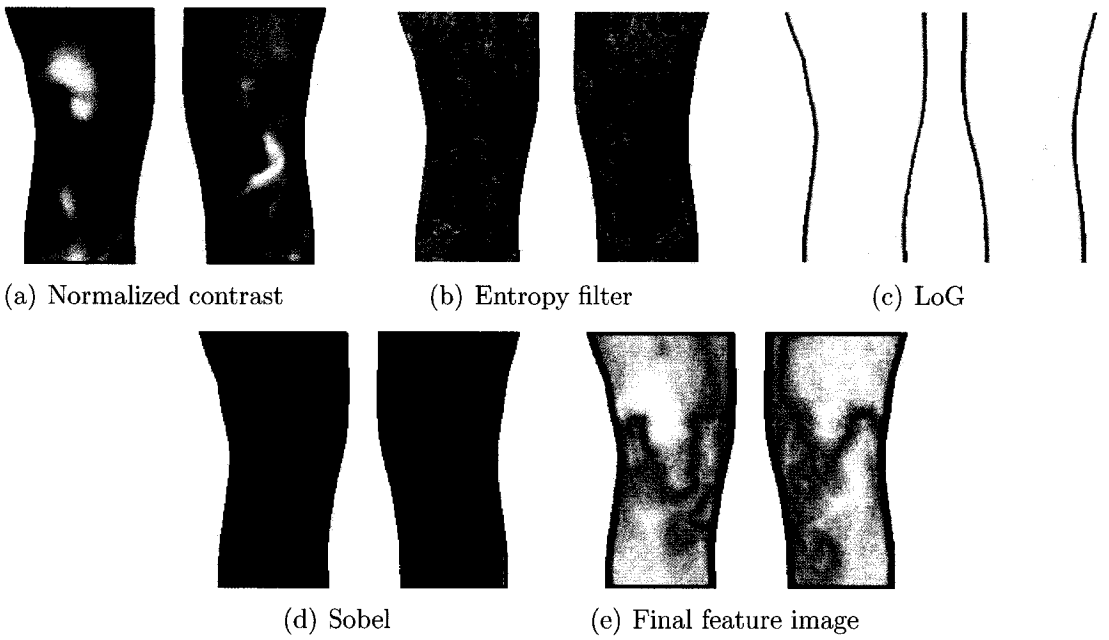
Figs. 6.6 and 6.7 show another example with an image of both knees. No precise contour of abnormal regions were available for this image. However, the subject imaged had a torn ligament below the left knee that had been confirmed by a physician based on MRI scan data. The resulting inflammation can be seen clearly under the left knee as a white region corresponding to an average temperature increase of +2.1 °C compared to the surrounding area and compared to the same region on the right knee (Fig.6.6(a)).



(a) Original (b) Feature image clusters

(c) ART clusters (d) Final clusters

Figure 6.6: Example of extraction of clusters of interest. a: Original image; b: Clusters returned by thresholding the combined feature image. c: Clusters returned by the ART clustering. d: Final clusters.

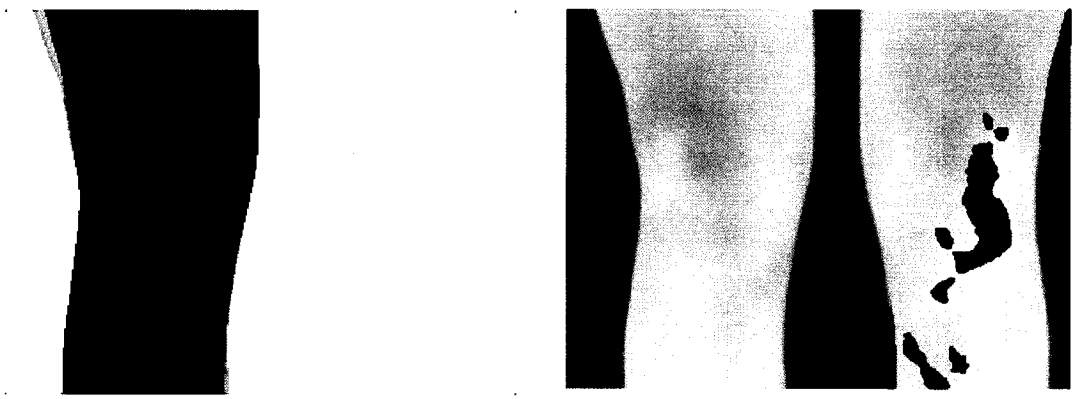


(a) Normalized contrast (b) Entropy filter (c) LoG

(d) Sobel (e) Final feature image

Figure 6.7: Example of feature images. a: Normalized contrast (neighborhood size: 5); b: Entropy filtered image (neighborhood size: 5); c: Laplacian of Gaussian filtered image ($\sigma = 3$, $l_1 = l_2 = 17$); d: Sobel filtered image; e: Final feature image with equal weights for all features.

Since both contralateral sides are present in the image, it is possible to assess the asymmetry between one knee and the other knee. The left knee was mirrored and registered with the right knee through a simple translation in this case. The absolute difference between the two knees was calculated and the results are shown in figure 6.8. The absolute difference image was then thresholded using the 90% point of the cumulative distribution function. The cluster location was then mirrored and translated back to the original location of the left knee. The inflammation region is clearly picked up.



(a) Difference image (b) Clusters found

Figure 6.8: Example of asymmetry analysis. a: Difference image between left and right knee; b: Clusters found by thresholding the difference image.

Chapter 7

Applications of the Segmentation and ROI Extraction

This chapter will introduce some original applications of the techniques described in chapters 5 and 6.

7.1 Musculoskeletal Disorders

7.1.1 Playing-Related Musculoskeletal Disorders

Playing-Related Musculoskeletal Disorders (PRMDs) affect a large number of musicians. Statistics of prevalence vary among experts depending on the criteria used to qualify these problems. However, even the most conservative studies agree that PRMDs typically affect 43% of professional musicians and 17% of young musicians [160]. Experts' estimates on prevalence of PRMDs sometimes are as high as 80% if light pain and aches are taken into account. There is no consensus on the definition of PRMDs,

although it is generally agreed that the symptoms are chronic, uncontrollable and unusual and that PRMDs sensibly influences the playing of the instrument [161]. Furthermore, some researchers consider that the musician is the only one who can define and qualify the severity of the disorders affecting his/her playing [160]. Indeed, PRMDs cover a wide range of symptoms and neuromuscular pathologies (tenosynovitis, focal dystonia, carpal tunnel syndrome, etc.) that primarily affect musculoskeletal tissues of the hands, arms, shoulders, neck, face and back [161] [162] [163] [164]. These symptoms are worsened, if not caused, by the repetitive and often forced motion work of neuromuscular tissues during regular instrumental practice. In addition, a vast majority of PRMDs induce inflammation as well as chronic pain, whose presence is not limited to the piano practice. The myriad of symptoms, their non-specificity and the subjective nature of associated pain make the diagnosis difficult to establish, except at later stages of PRMDs, which require surgical intervention followed by lengthy recovery, or that the musician stops playing temporarily, and sometimes indefinitely [162] [165] [166]. However, the diagnosis is greatly facilitated when the clinical specialist has access to the patient medical history, physical evaluation of the instrument, of contact areas between musicians and instrument as well as appropriate imaging data. Observing the musician in his/her environment is critical for diagnosis [161] [167].

Being able to understand the mechanisms and the evolution of PRMDs is important not only for a proper diagnostic and treatment but also to be able to limit the severity of the disorders or prevent them from spreading to other parts of the body. In addition, once the causes for the appearance of PRMDs are identified, it may be possible to develop preventive measures: better technique, posture, pedagogical

approaches that takes physical characteristics of the musicians into account.

7.1.1.1 Warm-up

Physical warm-up before intense physical activity is a commonly accepted practice in sports. It consists of a series a physical exercises whose duration and intensity vary, which are meant to increase the temperature of muscles, the blood flow to muscles and the oxygen consumption. The temperature increase of the muscles would reduce the resistance of muscles and articulations - lowering the risk of injuries - increase the release of oxygen from hemoglobin and myoglobin, speed metabolism and increase nerve conduction rate [168] [169]. Most warm-up routines in sports are active warm-up but passive warm-ups are sometimes thought to produce similar effects on the body although they are rarely used for practical reasons [168] [169]. Passive warm-up techniques include hot showers, saunas, heating pads, warm rooms.

Physical warm-up is often considered necessary to achieve good sports performance and reduce the risk of injuries. It is interesting to note, however, the lack of scientific evidence demonstrating the benefits of different warm-up techniques. Physiologic consequences of warm-up are still not fully understood, nor are the effects on performance [168] [169].

Surprisingly, physical warm-up is far less systematic in the performing arts, be they professional or amateur musicians. Some musicians, such as wind players, do several minutes of specific exercises to increase the coordination of the muscles of mouth [170]; singers do numerous respiratory exercises. The lack of warm-up is particularly true for pianists for which the topic is still rarely discussed. A musical warm-up is often substituted to a physical warm-up. Various musical pieces exist that

make a specific group of muscles work before the practice starts [171]. The structure of a musical warm-up, however, varies considerably and in many cases musical warm-up does not serve its purpose, sometimes perhaps being even counter-productive (for instance starting a practice by playing fast scales).

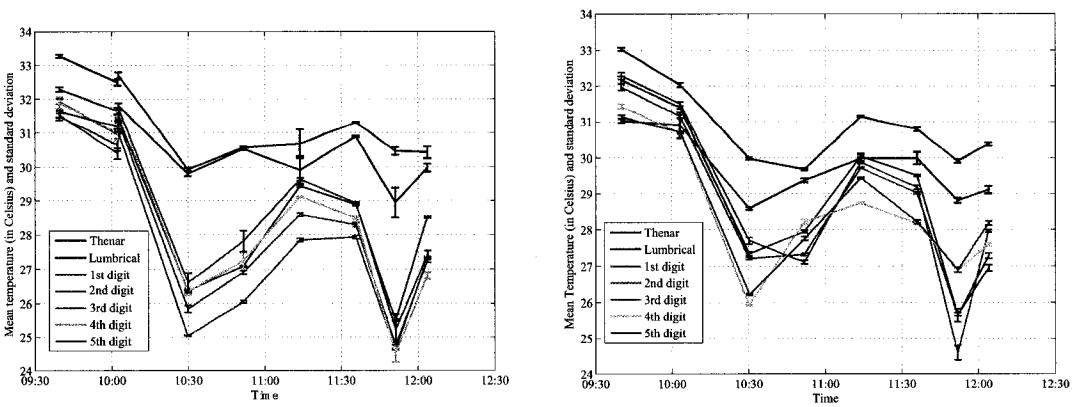
Nevertheless, one could compare the intensity and physical stress of a recital, concert or even of a long practice session with a sports event. Following this premise, one can legitimately think that musicians could benefit from the same positive effects in terms of injury prevention if appropriate warm-up techniques were developed and used systematically. One of the main effects of warm-up is an increase of the muscle temperature, which in turn would lower the viscous resistance of muscles and articulations, increase oxygen levels in muscles, increase anaerobic metabolism and increase nerve conduction rate [168] [169]. Once again, scientific papers on the evolution of the temperature of muscles and neuromuscular tissues are scarce, even in sports medicine literature. In order to understand the effects of warm-up on performance and injury prevention in musicians it is important to study the evolution of muscle temperature in musicians.

7.1.1.2 Evolution of the Temperature of the Palm

Figs. 7.1 and 7.2 show the evolution of the temperature of seven anatomical regions of the hand: the 5 fingers, the thenar muscles and the ulnar part of the lumbrical muscles of the hand. The plots were obtained by using the segmentation, anatomical ROI extraction and feature extraction blocks from figure 3.1. Features extracted were the ones described in 3.7. Only the mean temperature is plotted, along with the standard deviation of the mean temperature for each sequence of images. The

input was the sequence of IR images in Matlab format that were transferred from the FLIR proprietary format through FLIR Thermacam Researcher software. The timing, temperature and scaling information is included in the image data. The only manual input required was the anatomical view (palmar view of the hand). For subject #1, there were 135 right hand images and 82 left hand images. It took an average of 152.4 seconds and 90.5 seconds to process the two sequences of images respectively (from input of the data to the final plot of the evolution of temperature). For subject #2, there were 65 right hand images and 65 left hand images. It took an average of 70.4 seconds to process each sequence of images.

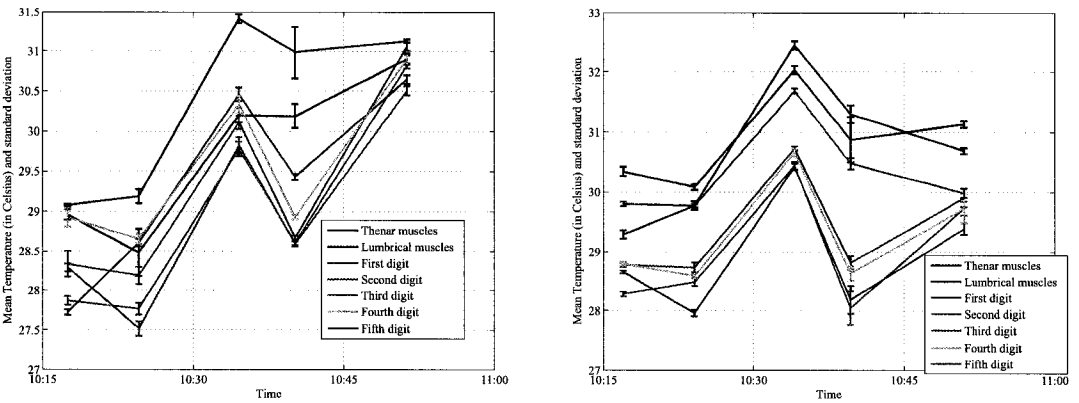
According to the protocol described in Section 4.3.1, each point on the plots corresponds to the average temperature of a region at a specific time during the practice.



(a) right arm, dorsal view, subject #1

(b) left arm, dorsal view, subject #1

Figure 7.1: Evolution of the temperatures of the 5 fingers, the thenar and lumbrical muscles of the palm, during a piano practice for subject #1.



(a) right arm, dorsal view, subject #2

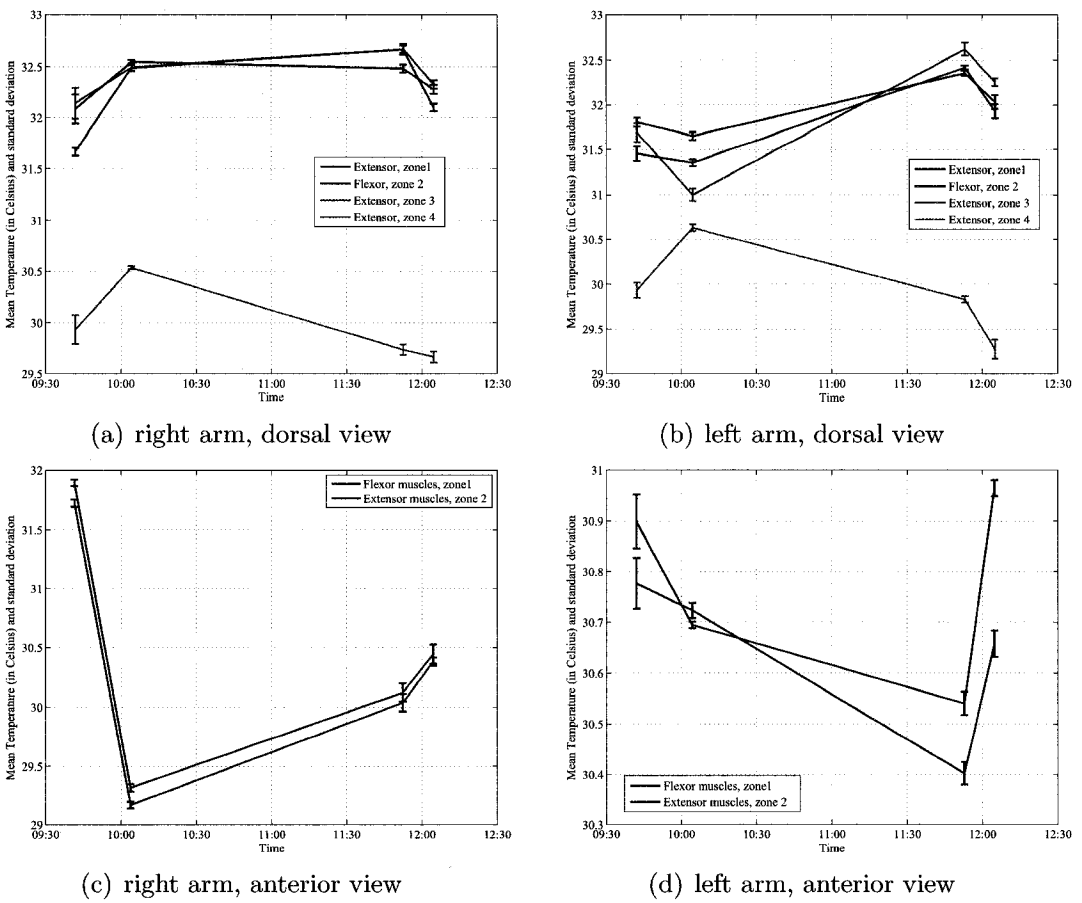
(b) left arm, dorsal view, subject #2

Figure 7.2: Evolution of the temperatures of the 5 fingers, the thenar and lumbrical muscles of the palm, during a piano practice for subject #2.

7.1.1.3 Evolution of the Temperature of the Upper-arm

The plots in figures 7.3 and 7.4 show the evolution of the temperature of 6 anatomical regions of the upper-arm (dorsal view), that were shown in Figs. 3.8 and 3.10. The sequences were from the database of IR images of pianists. Two subjects are shown.

Again, the plots were obtained by using the segmentation, anatomical ROI extraction and feature extraction blocks from diagram 3.1. Features extracted were the ones described in 3.7. Only the mean temperature is plotted, along with the standard deviation of the mean temperature for each sequence of images. The input was the sequence of IR images in Matlab format that were transferred from the FLIR proprietary format through FLIR Thermacam Researcher software. The timing, temperature and scaling information was included in the image data. The only manual input required was the anatomical view (dorsal view of the upper-arm). It took an average of 0.95 second per image to process the sequences of images (from input of the data to the final plot of the evolution of temperature).



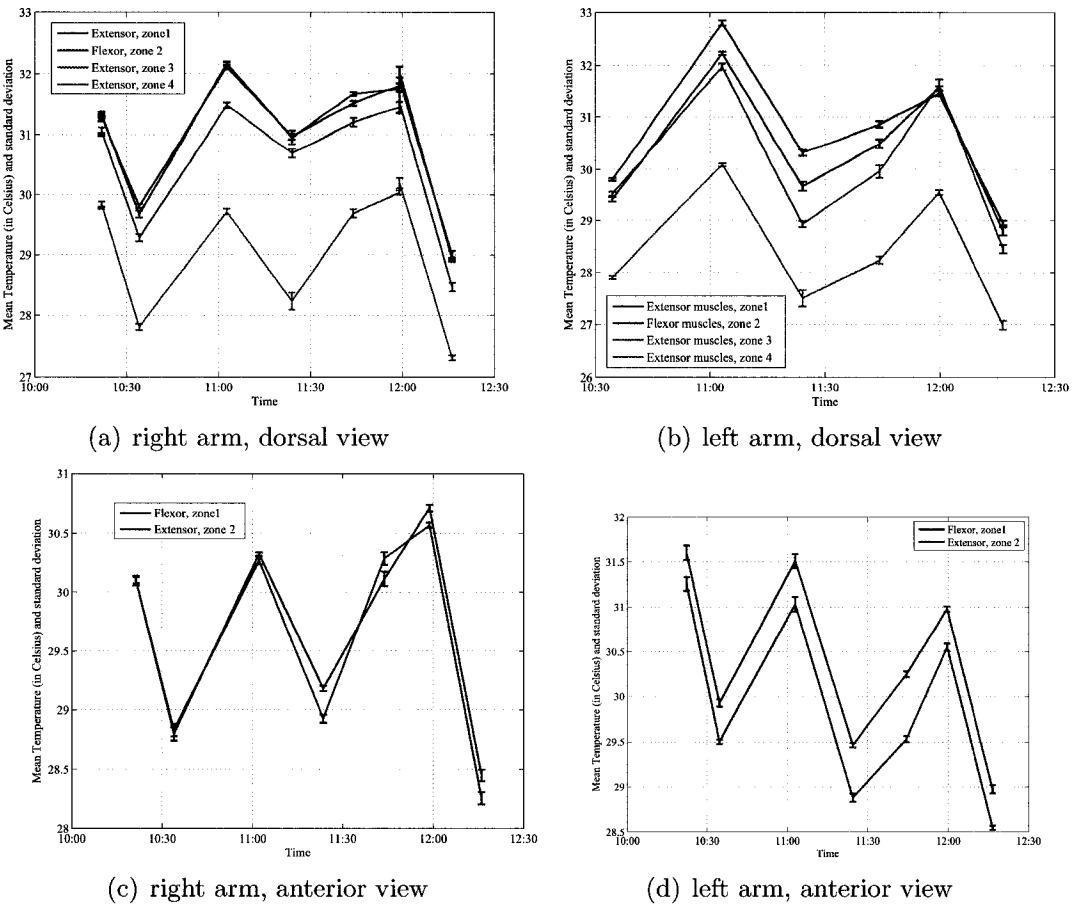
(a) right arm, dorsal view

(b) left arm, dorsal view

(c) right arm, anterior view

(d) left arm, anterior view

Figure 7.3: Evolution of the temperatures of the superficial muscles of the upper arm, during a piano practice for subject #1.



(a) right arm, dorsal view

(b) left arm, dorsal view

(c) right arm, anterior view

(d) left arm, anterior view

Figure 7.4: Evolution of the temperatures of the superficial muscles of the upper arm, during a piano practice for subject #2.

7.1.2 Repetitive Strain Injuries in Computer Users

This application focuses on injuries affecting the hands and wrists (carpal tunnel syndrome, tenosynovitis, ulnar tunnel syndrome, Raynaud's disease, etc). Upper extremity disorders are common among workers who typically perform prolonged repetitive tasks. Diagnosis is difficult due to the multiplicity of symptoms and the subjectivity of the pain involved. Typical tests include physical tests (Phalen's test, Tinel test, Finkelstein's test), questionnaires or diagrams, and electrodiagnostic tests. Most repetitive strain injuries affect the temperature distribution of the musculoskeletal system and by extension of the skin, which can be recorded by infrared imaging, analyzed and classified.

Fig. 7.5 shows a comparison of the left and right palm for the 7 regions mentioned above and an additional anatomical region corresponding to the carpal tunnel area of the wrist. Notice that both hands are on the same image, so the segmentation block was run iteratively in this case, to be able to detect both hands. After one hand has been detected, the corresponding region was removed for the next pass of the segmentation. The subject was complaining about pain in the left wrist and left elbow area and tingling in the ring and index finger of the left hand after several hours of computer use. The subject filled the questionaire from Appendix A but electrodiagnostic tests had not been performed yet at the time of the imaging.

The mean temperature difference between contralateral regions are given in Table 7.1.

The temperature difference of region 8 (carpal tunnel region), is relatively high in this case as well as the thenar and part of the lumbrical muscle regions (regions 1 and 2). This could be indicative of an early inflammation of the carpal tunnel region.

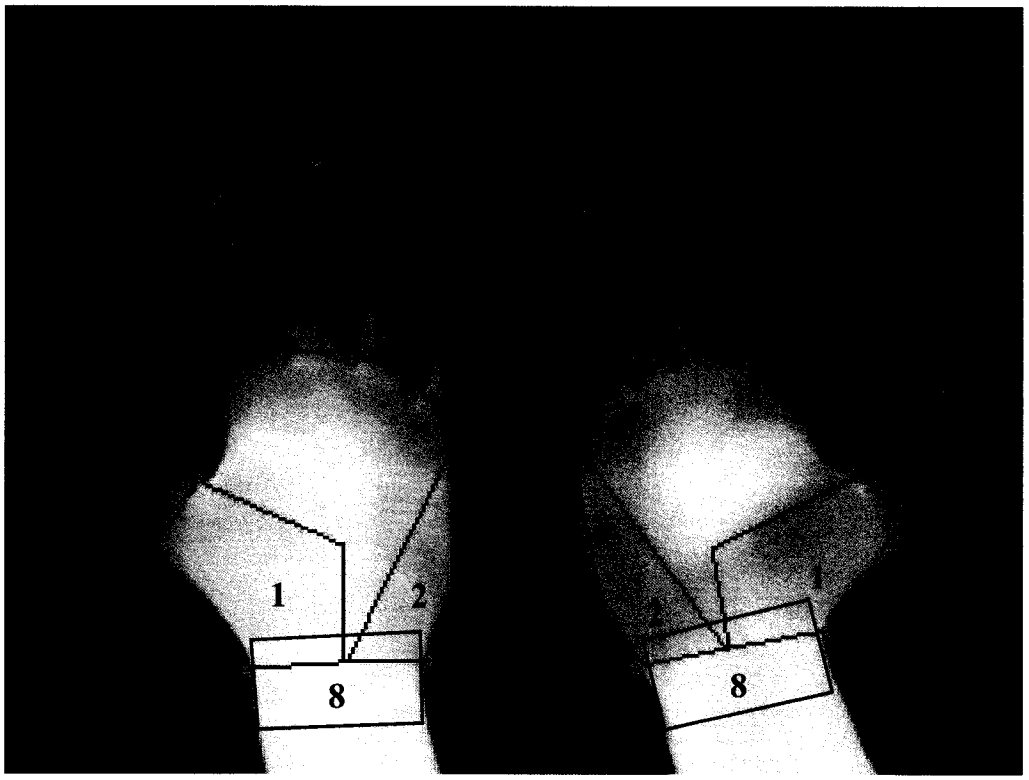


Figure 7.5: Comparison of left vs right anatomical regions of the palm.

Table 7.1: Comparison of temperature of left vs right anatomical regions of the palm.

Region	Mean temp. left (std)	Mean temp. right (std)	Difference
1	30.53 (1.18)	29.00 (0.85)	1.53
2	29.43 (1.34)	28.39 (0.87)	1.04
3	25.21 (1.61)	24.54 (1.32)	0.67
4	25.08 (1.26)	25.19 (1.44)	-0.11
5	25.74 (1.46)	25.49 (1.35)	0.25
6	25.81 (1.40)	25.11 (1.04)	0.70
7	25.55 (1.27)	25.20 (1.62)	0.35
8	31.80 (0.70)	30.70 (0.80)	1.10

Chapter 8

Conclusion and Future Work

8.1 Conclusion

Automated analysis of thermal IR images of the human body is a difficult problem in part due to the characteristics of thermal images, the difficulty in extracting relevant information and the lack of medical IR databases with labelled normal and abnormal data.

The following contributions to knowledge are derived from this thesis. First, a general modular architecture for automated detection of anomalies in clinical IR images was presented. The proposed architecture is flexible and adapts to the amount of information available for specific application. This contribution was published in C. Herry, M. Frize, "Design considerations for a medical thermographic expert system," in Proceedings of the 25th Annual International Conference of the IEEE EMBS, Cancun Mexico, Sept. 17th-21st 2003 [172].

From that general architecture, the segmentation and the extraction of regions of interest blocks were thoroughly investigated. Several novel segmentation approaches

were proposed. The first one called cued morphological processing of edge maps uses a minimum amount of *a priori* information about the region sought in order to constrain the morphological processing of edge components. This approach was shown to outperform significantly existing state-of-the-art segmentation techniques, especially when dealing with faint regions in challenging conditions. This contribution was published in C.L. Herry, R.A. Goubran and M. Frize,"Segmentation of infrared image using cued morphological processing of edge maps", in Proceedings of the 2007 IEEE Instrumentation and Measurement Technology Conference (IMTC07), Warsaw, Poland, 2007 [157]. The second segmentation approach is based on the classification of edge components with a Support Vector Machine classifier or a Multilayer Perceptron. The classifiers are trained on a database of features characteristic of edge components belonging to the contour of the object of interest. Once trained, the classifiers can recover the entire contour in normal situations and at least parts of the contour in difficult situations. These partial contours can in turn be used in conjunction with other segmentation approaches to recover missing contour regions. A third segmentation approach was introduced as a robust segmentation block for the anomaly detection framework. This approach is a fusion strategy that combines several segmentation approaches, including the cued morphological processing of edge maps, the SVM and ANN classification of edge components, optimal thresholding and FCM clustering, in order to increase the overall segmentation performance. The proposed fusion strategy is a rule-based fusion using morphological post-processing of the combined segmented contours. When compared with existing decision fusion schemes, including a fusion scheme based on majority voting with binary weights and a Bayesian fusion scheme, the proposed rule-based fusion strategy demonstrated its

superiority. Finally, a technique was described to automatically segment specific regions such as the breast. This minor contribution was published as part of N. Scales, C. Herry and M. Frize, "Automated image segmentation for breast analysis using infrared images", in Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco California, 2004 [34].

As part of the second main contribution of this work, a new approach to identify anatomical ROIs in IR images using automatically found landmarks was presented. This approach is based on curvature analysis of the region contour, a curvature analysis of the distance between contours for symmetrical regions, and region specific constraints. The results are well localized clusters of landmarks that accurately reflect the true location of key anatomical landmarks on the region contours. This contribution has been accepted for publication in C.L. Herry, R.A. Goubran and M. Frize, "Improving the Detection and Localization of Anatomical Landmark Points in Infrared Images Using Symmetry and Region Specific Constraints", for the 2008 IEEE International Instrumentation and Measurement Technology Conference I^2MTC 2008, Victoria, Vancouver Island, BC, Canada, May 2008. Furthermore, a technique for extracting anatomical regions from the previously found anatomical landmarks on thermal IR images was described. A few key points on the region contour are identified and they allow a simple affine transform to recover anatomical ROIs. This was used to study the evolution of surface temperature of pianists' arm muscles during a piano practice. This contribution was published in C.L. Herry, M. Frize, R.A. Goubran, G. Comeau, "Evolution of the surface temperature of pianists' arm muscles using infrared thermography", in Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, September

2005 [173]. A minor contribution was also presented for the extraction of potentially abnormal ROIs, that is based on a fusion of clusters extracted from a combination of feature images and an unsupervised clustering through a fuzzy ART neural network.

8.2 Suggested Future Work

Several areas of the work presented in this thesis would benefit from further investigation. The first one is the issue the segmentation of multiple overlapping body regions, in which case the contour of each region may not be recovered fully and some of the key landmarks may not be available. Recovery of anatomical regions is likely to be complicated by this aspect and additional processing steps may be needed to obtain partial recovery of anatomical regions of interest.

The second area of research that will need to be addressed is the classification block in our architecture, as presented in the overview diagram of figure 3.1. This will require the help of a medical expert that would be able to link potentially abnormal thermal patterns with specific pathological thermal dysfunction. Once a set of labelled abnormal patterns is available, then each sub-classifier proposed in our architecture can be trained on normal or abnormal thermal patterns and provide a solid basis for finding abnormal thermal patterns. An unsupervised anomaly detection scheme can then be added to detect unseen thermal patterns that may be abnormal.

A third area for future research is the real-time tracking of specific regions over long sequences. The results presented in chapter 7 showed examples of automated analysis of the evolution of temperature of body regions over time, where each frame of a sequence was analysed separately (segmentation and extraction of ROIs). This is not practical for real-time or near real-time sequence analysis in fast moving environment

(for instance a sequence of a pianist playing the piano). The tracking of regions using temporal cues and landmark points could help speed up the processing and enable a real-time analysis that could identify more subtle temperature variations over time for each body region considered.

Finally, the extension of our architecture to incorporate information from several IR cameras or visible cameras could benefit several stages of processing, facilitating the segmentation and extraction of ROIs as well as helping a medical specialist to visualize the abnormal regions.

References

[1] A. Collins, E. Ring, J. Cosh, and P. Bacon, "Quantitation of thermography in arthritis using multiisothermal analysis," *Annals of Rheumatoid Diseases*, vol. 33, no. 2, pp. 113–115, 1974.

[2] M. Aubry-Frize, G. Quartey, H. Evans, and D. LaPalme, "The thermographic detection of pain," in *Proceedings of the 3rd Canadian Clinical Engineering Conference*, Saskatoon, September 1981, pp. 82–83.

[3] M. Gautherie, A. Kotewicz, and P. Gueblez, "Accurate and objective evaluation of breast thermograms: principles and new advances with special reference to an improved computer-assisted scoring system," in *Proceedings of the international Conference on Thermal Assessment of Breast Health*, Washington, DC, USA, 1983, pp. 74–92.

[4] S. Feig, G. Shaber, G. Schwartz, A. Patchefsky, H. Libshitz, J. Edeiken, R. Nerlinger, R. Curley, and J. Wallace, "Thermography, mammography, and clinical examination in breast cancer screening. review of 16,000 studies," *Radiology*, vol. 122, pp. 123–127, 1977.

[5] Y. So, M. Aminoff, and R. Olney, "The role of thermography in the evaluation of lumbosacral radiculopathy," *Neurology*, vol. 39, p. 1154, 1989.

[6] E. Ring, "Criteria for thermal imaging in medicine," in *IEEE 17th Annual Conference of the Engineering in Medicine and Biology Society*, 1995, pp. 1697–1698.

[7] R. Pochaczevsky, "Technical guidelines, edition 2," *Thermology*, vol. 2, no. 2, 1986.

[8] M. Anbar, "Objective assessment of clinical computerized thermal images," in *Medical Imaging V: Image Processing*, ser. Proceedings of SPIE, vol. 1445. SPIE-The International Society for Optical Engineers, 1991, pp. 479–484.

[9] I. Fujimasa, M. Nakajima, K. Mabuchi, T. Chinzei, Y. Abe, Y. Hyakuna, and K. Atsumi, "Development of an on-line real-time computed thermography system (CTS)," in *MEDINFO'86*, 1986, pp. 709–711.

[10] W. Wolfe, *Handbook of military infrared technology.* Office of Naval Research, Department of Navy, 1965.

[11] J. Hudson, R.D., *Infrared system engineering.* Wiley-Interscience, 1969.

[12] G. Gaussorgues, *Infrared thermography.* Chapman and Hall, 1994.

[13] M. Anbar, *Quantitative dynamic telethermometry in medical diagnosis and management.* Boca Raton: CRC Press, 1994.

[14] B. 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.

[15] Y. Houdas and E. Ring, *Human body temperature : its measurement and regulation.* Plenum Press, 1982.

[16] P. Goodman, M. Murphy, G. Siltanen, M. Kelley, and L. Rucker, "Normal temperature asymmetry of the back and extremities by computer-assisted infrared imaging," *Thermology*, vol. 1, pp. 195–202, 1986.

[17] H. Hooshmand, M. Hashmi, and E. Phillips, "Infrared thermal imaging as a tool in pain management - an 11 year study, part I of II," *Thermology International*, vol. 11, no. 2, pp. 53–65, 2001.

[18] J. Hubbard and C. Hoyt, "Pain evaluation by electronic infrared thermography: correlation with symptoms, EMG, myelogram and CT scan," *Thermology*, vol. 1, no. 1, pp. 26–35, 1985.

[19] P. LeRoy, C. Christian, and R. Filasky, "Diagnostic thermography in low back pain syndromes," *The Clinical Journal of Pain*, vol. 1, no. 1, pp. 4–13, 1985.

[20] G. Goldberg, "Infrared imaging and magnetic resonance imaging correlated in 35 cases," *Thermology*, vol. 1, no. 4, 1986.

[21] W. Ellis, J. Morris, and A. Swartz, "Screening thermography of chronic back pain patients with negative neuromusculoskeletal findings," *Thermology*, vol. 3, 1989.

[22] D. Thomas, D. Cullum, G. Siahamis, and S. Langlois, "Infrared thermographic

imaging, magnetic resonance imaging, ct scan and myelography in low back pain," *British Journal of Rheumatology*, vol. 29, pp. 268–273, 1990.

[23] Y.-S. Kim and Y.-E. Cho, "Correlation of pain severity with thermography," in *Proceedings of the 17th Annual International Conference of the IEEE Engineering in Medicine and Biology Society and 21st Canadian Medical and Biological Engineering Conference*, vol. 2, 1995, pp. 1699–1700.

[24] Y. Poussart, M. Frize, and R. Roberge, "The re-evaluation of thermography in breast cancer detection by new image enhancement techniques," Terry Fox Foundation, Fredericton, NB, Tech. Rep., 1988.

[25] S. Uematsu, "Symmetry of skin temperature comparing one side of the body to the other," *Thermology*, vol. 1, no. 1, pp. 4–7, 1985.

[26] S. Uematsu, D. Edwin, W. Jankel, J. Kozikowski, and M. Trattner, "Quantification of thermal asymmetry, part 1: Normal values and reproducibility," *Journal of Neurosurgery*, vol. 69, no. 4, pp. 552–555, 1988.

[27] C. Herry and M. Frize, "Quantitative assessment of pain-related thermal dysfunction through clinical digital infrared thermal imaging," *Biomedical Engineering Online*, vol. 3, no. 1, p. 19, 2004.

[28] J. Head, C. Lipari, F. Wang, and R. Elliott, "Image analysis of digitized infrared images of the breasts from a first generation infrared imaging system," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1997, pp. 681–684.

[29] C. Lipari and J. Head, "Advanced infrared image processing for the breast cancer risk assessment," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1997, pp. 673–676.

[30] L.-M. Gallo, P. Boiger, D. Stucki, and C. Rageth, "Quantitative infrared thermography in varicocele screening," *Thermology*, vol. 2, pp. 554–562, 1987.

[31] J. Montoro and M. Anbar, "New modes of data handling in computerized thermography," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1988, pp. 845–847.

[32] J. Montoro, L. Hershey, and M. Anbar, "Enhancement of interpretation of thermograms through on-line software," *Thermology*, vol. 3, no. 2, pp. 121–124, 1989.

[33] M. Varga and R. Hanka, "An iterative elastic stretching technique applied to thermographic images," in *IEEE 3rd Conference on Image Processing and its Applications*, 1989, pp. 324–328.

[34] N. Scales, C. Herry, and M. Frize, "Automated image segmentation for breast analysis using infrared images," in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, 2004, pp. 1737–1740.

[35] H. Qi, W. Snyder, J. Head, and R. Elliott, "Detecting breast cancer from infrared images by asymmetry analysis," in *Proceedings of the 22nd Annual Inter-*

national Conference of the IEEE Engineering in Medicine and Biology Society, 2000, pp. 1227–1228.

[36] E. Ng, Y. Chen, and L. Ung, "Computerized breast thermography. study of image segmentation and temperature cyclic variations," *Journal of Medical Engineering and Technology*, vol. 25, no. 1, pp. 12–16, 2001.

[37] K. Mabuchi, T. Chinzei, I. Fujimasa, S. Haeno, Y. Abe, and T. Yonezawa, "An image-processing program for the evaluation of asymmetrical thermal distributions," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1997, pp. 725–728.

[38] V. Vavilov, E. Vavilova, and D. Popov, "Statistical analysis of the human body temperature asymmetry as the basis for detecting pathologies by means of IR thermography," in *Thermosense XXIII*, ser. Proceedings of SPIE, vol. 3061. SPIE-The International Society for Optical Engineers, 2001, pp. 482–491.

[39] J. Koay, C. Herry, and M. Frize, "Analysis of breast thermography with an artificial neural network," in *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC04*, vol. 1, 2004, pp. 1159–1162.

[40] J. Montoro and M. Anbar, "Visualization and analysis of dynamic thermographic changes," in *Proceedings of the 1st Conference on Visualization in Biomedical Computing*, 1990, pp. 486–489.

[41] M. Unser, H. Van Hamme, P. De Muynck, E. Van Denhaute, and J. Cornelis, "Karhunen-loeve analysis of dynamic sequences of thermographic images for

early breast cancer detection," in *Proceedings of CVPR'88, Computer Society Conference on Computer Vision and Pattern Recognition*, 1988, pp. 592–596.

[42] M. Varga and P. De Muynck, "Thermal analysis of infra-red mammography," in *Proceedings of the 11th IAPR International Conference on Pattern Recognition, Vol. III Conference C: Image, Speech and Signal Analysis*, 1992, pp. 360–364.

[43] M. Varga and R. Hanka, "Automatic thermal image analysis for medical diagnosis," in *IEEE International Conference on Image Processing and its Applications*, 1992, pp. 526–529.

[44] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn, and I. Masaki, "A shape independent method for pedestrian detection with far-infrared images," *IEEE Transactions on Vehicular technology*, vol. 35, no. 6, 2004.

[45] C. Herry, M. Frize, and R. Goubran, "Segmentation and landmark identification in infrared images of the human body," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, September 2006, pp. 957–960.

[46] J. Irvine, "Targeting breast cancer detection with military technology," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 6, pp. 31–40, 2002.

[47] R. Nowak and R. Baraniuk, "Wavelet-domain filtering for photon imaging systems," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 666–678, 1999.

[48] A. Bal and M. Alam, "Automatic target tracking in FLIR image sequences

using intensity variation function and template modeling," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, pp. 1846–1852, 2005.

[49] S.-Y. Huang, C.-W. Mao, and K. Cheng, "A VQ-based approach to thermal image analysis for printed circuit boards diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 6, pp. 2281–2388, 2005.

[50] J. Wu, J. Li, J. Liu, and J. Tian, "Infrared image segmentation via fast Fuzzy C-Means with spatial information," in *Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics*, August 2004, pp. 742–745.

[51] B. Ernisse, S. Rogers, M. DeSimio, and R. Raines, "Complete automatic target cuer/recognition system for tactical forward-looking infrared images," *Optical Engineering*, vol. 36, no. 9, pp. 2593–2603, 1997.

[52] Q. Zhou, Z. Li, and J. Aggarwal, "Boundary extraction in thermal images by edge map," in *ACM Symposium on applied computing*, March 2004, pp. 254–258.

[53] M. Blank and C. Kargel, "Infrared images to measure changes of the extremities caused by cigarette smoke and nicotine gums," in *Proceedings of the 23rd IEEE Instrumentation and Measurement Technology Conference (IMTC06)*, Sorrento, Italy, 2006, pp. 794–799.

[54] E. Chan and J. Pearce, "Visualization of dynamic subcutaneous vasomotor response by computer-assisted thermography," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 8, pp. 786–795, 1990.

[55] C. Lee and S. Wong, "A mathematical morphological approach for segmenting heavily noise-corrupted images," *Pattern Recognition*, vol. 29, no. 8, pp. 1347–1358, 1996.

[56] D. Daut and D. Zhao, "A flaw detection method based on morphological image processing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 6, pp. 389–398, 2006.

[57] K. Mak, P. Peng, and H. Lau, "Optimal morphological filter design for fabric defect detection," in *IEEE International Conference on Industrial Technology*, December 2005, pp. 799–804.

[58] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," in *Proceedings of the International Conference on Computational Intelligence for Measurement Systems and Applications*, Giardini Naxos, Italy, 2005, pp. 96–101.

[59] NASA, *Anthropometric Source book. Volume 1: Anthropometry for Designers.* NASA Reference Publication, 1978, vol. 1024.

[60] K. Robinette, H. Daanen, and E. Paquet, "The CAESAR project: a 3D surface anthropometry survey," in *Proceedings of the 2nd International Conference on 3D Digital Imaging and Modeling*, 1999, pp. 380–386.

[61] Z. Ben Azouz, C. Shu, and A. Mantel, "Automatic locating of anthropometric landmarks on 3D human models," in *Third International Symposium on 3D Data Processing, Visualization and Transmission*, Chapel Hill, NC, USA, 2006, pp. 750–757.

[62] L. Kitchen and A. Rosenfeld, "Gray-level corner detection," *Pattern Recognition Letters*, vol. 1, pp. 95–102, 1982.

[63] L. Le Briquer, F. Lachmann, and C. Barillot, "Using local extremum curvatures to extract anatomical landmarks from medical images," in *Medical Imaging 1993: Image Processing*, ser. Proceedings of SPIE, vol. 1898, Newport Beach, CA/USA, 1993, pp. 549–558.

[64] H. Moravec, "Towards automatic visual obstacle avoidance," in *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, 1977, p. 584.

[65] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th ALVEY vision conference*, Manchester, UK, 1988, pp. 147–151.

[66] W. Förstner, "A feature based correspondence algorithm for image matching," *Internal Archives of Photogrammetry and Remote Sensing*, vol. 26, no. 3, pp. 150–166, 1986.

[67] P. Kovesi, "Image features from phase congruency," *Videre: A Journal of Computer Vision Research*, vol. 1, no. 3, pp. 2–26, 1999.

[68] F. Mokhtarian and A. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 34–43, 1986.

[69] X. He and N. Yung, "Curvature scale-space corner detector with adaptive threshold and dynamic region of support," in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. II, 2004, pp. 791–794.

[70] S. Smith and M. Brady, "SUSAN - a new approach to low level image processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.

[71] K. Walker, T. Cootes, and C. Taylor, "Locating salient object features," in *Proceedings of the 9th British Machine Vision Conference (BMVC'98)*. Southampton, UK: BMVA Press, 1998, pp. 557–566.

[72] I. Scott, T. Cootes, and C. Taylor, "Improving appearance model matching using local image structure," in *Proceedings of Information Processing in Medical Imaging (IPMI'03)*, ser. Lecture Notes in Computer Science, vol. 2732. Ambleside, UK: Springer, Berlin, 2003, pp. 258–269.

[73] C. Izard and B. Jedynak, "Bayesian registration for anatomical landmark detection," in *Proceedings of the Third IEEE International Symposium on Biomedical Imaging (ISBI 2006)*, Arlington, VA, USA, 2006, pp. 856–859.

[74] J. Duryea, Y. Jiang, P. Countryman, and H. Genant, "Automated algorithm for the identification of joint space and phalanx margin locations on digitized hand radiographs," *Medical Physics*, vol. 26, no. 3, pp. 453–461, 1999.

[75] J. Khan and M. Alam, "Target detection in cluttered forward-looking infrared imagery," *Optical Engineering*, vol. 44, no. 7, pp. 076 404.1–076 404.10, 2005.

[76] D. Casasent, J.-S. Smokelin, A. Ye, and R. Schaefer, "Correlation filter fusion for detection: morphological, wavelet and gabor methods," in *Photonics for Processors, Neural Networks, and Memories*, vol. Proc. SPIE vol. 2026. SPIE-The International Society for Optical Engineers, 1993, pp. 6–16.

[77] S. Greenberg, S. Rotman, H. Guterman, S. Zilberman, and A. Gens, "Region-of-interest-based algorithm for automatic target detection in infrared images," *Optical Engineering*, vol. 44, no. 7, pp. 077 002.1–077 002.12, 2005.

[78] C. Privitera and L. Stark, "Evaluating image processing algorithms that predict regions of interest," *Pattern Recognition Letters*, vol. 19, no. 11, pp. 1037–1043, 1998.

[79] B. Weber, "Comparison of human observer and algorithmic target detection in nonurban forward looking infrared imagery," *Optical Engineering*, vol. 44, no. 7, pp. 076 401.1–076 401.8, 2005.

[80] L. Voicu, M. Uddin, H. Myler, A. Gallagher, and J. Schuler, "Clutter modeling in infrared images using genetic programming," *Applied Engineering*, vol. 39, no. 9, pp. 2359–2371, 2000.

[81] G. Wang, J. Li, and T. Zhang, "Approach to target detection based on scale changing fractal signature," in *Targets and Backgrounds: Characterization and Representation III*, ser. Proceedings of SPIE, vol. 3062, 1997.

[82] B. Nelson, "Automatic vehicle detection in infrared imagery using a fuzzy inference-based classification system," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 1, pp. 53–61, 2001.

[83] B. Southall, T. Hague, J. Marchant, and B. Buxton, "An autonomous crop treatment robot: part I. a kalman filter model for localization and crop/weed classification," *The International Journal of Robotics Research*, vol. 21, no. 1, pp. 61–74, 2002.

[84] S. Sun and H. Park, "Automatic target recognition using boundary partitioning and invariant features in forward-looking infrared images," *Optical Engineering*, vol. 42, no. 2, pp. 524–533, 2003.

[85] C. Wang, H. Lu, and S. Ma, "A synthesized computer recognition system for human hands," in *Third International Computer Science Conference: Image Analysis Applications and Computer Graphics*, vol. 1024. Springer, December 1995, pp. 435–442.

[86] S. Zhu and G. Shaefer, "Thermal medical image retrieval by moment invariants," in *ISBMDA 2004*, vol. 3337. Springer-Verlag Berlin Heidelberg, 2004, pp. 182–187.

[87] D. Casasent and A. Ye, "Detection filters and algorithm fusion for ATR," *IEEE Transactions on Image Processing*, vol. 6, pp. 114–125, 1997.

[88] B. Bhanu and T. Jones, "Image understanding research for automatic target recognition," *IEEE AES Systems Magazine*, vol. 8, no. 10, pp. 15–23, 1993.

[89] C. Bishop, *Neural networks for pattern recognition.* Oxford University Press, 1995.

[90] M. Markou and S. Singh, "Novelty detection: a review-part 1: statistical approaches," *Signal Processing*, vol. 83, no. 12, pp. 2481–2497, 2003.

[91] S. Roberts, "Novelty detection using extreme value statistics," *IEE proceedings on vision, image and signal processing*, vol. 146, no. 3, pp. 124–129, 1999.

[92] S. Roberts and L. Tarassenko, "A probabilistic resource allocating network for novelty detection," *Neural Computation*, vol. 6, no. 2, pp. 270–284, 1994.

[93] L. Tarassenko, "Novelty detection for the identification of masses in mammograms," in *Proceedings of the 4th IEE International Conference on Artificial Neural Networks*, vol. 4, 1995, pp. 442–447.

[94] S. Roberts, D. Husmeier, L. Rezek, and W. Penny, "Bayesian approaches to gaussian mixture modelling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1133–1142, 1998.

[95] M. Lauer, "A mixture approach to novelty detection using training data with outliers," in *Proceedings of the 12th European Conference on Machine Learning.* Springer, 2001, pp. 300–311.

[96] R. Duda, P. Hart, and D. Stork, *Pattern classification.* Wiley-Interscience Publications, 2001.

[97] C. Bishop, "Novelty detection and neural network validation," in *Proceedings of the IEE Conference on Vision, Image and Signal Processing*, vol. 141, 1994, pp. 217–222.

[98] J. Ryan, M. Lin, and R. Miikkulainen, "Intrusion detection with neural networks," in *Advances in Neural Information Processing Systems*, vol. 10. MIT Press, 1998, pp. 943–949.

[99] Y. Le Cun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advance in Neural Information Processing Systems*, M. Kaufman, Ed., vol. 2, Los Altos, CA, USA, 1990, pp. 396–404.

[100] C. De Stefano, C. Sansone, and M. Vento, "To reject or not to reject: that is the question - an answer in case of neural classifiers," *IEEE Transactions on Systems, Man and Cybernetics - Part C*, vol. 30, no. 1, pp. 84–94, 2000.

[101] G. Vasconcelos, M. Fairhurst, and D. Bisset, "Investigating feedforward neural networks with respect to the rejection of spurious patterns," *Pattern Recognition Letters*, vol. 16, no. 2, pp. 207–212, 1995.

[102] S. Singh and M. Markou, "An approach to novelty detection applied to the classification of image regions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 4, pp. 396–407, 2004.

[103] C. Wilson, J. Blue, and O. Omidvar, "Improving neural network performance for character and fingerpint classification by altering network dynamics," in *Proceedings of the World Congress on Neural Networks*, vol. 2, 1995, pp. 151–158.

[104] T. Brotherton and T. Johnson, "Anomaly detection for advance military aircraft using neural networks," in *Proceedings of the 2001 IEEE Aerospace Conference*, vol. 6, March 2001, pp. 3113–3123.

[105] T. Brotherton, T. Johnson, and G. Chadderdon, "Classification and novelty detection using linear models and a class dependent elliptical basis function neural network," in *Proceedings of the IJCNN Conference*, vol. 2, May 1998, pp. 876–879.

[106] Y. Li, M. Pont, and N. Jones, "Improving the performance of the radial basis function classifiers in condition monitoring and fault diagnosis applications

where unknown faults may occur," *Pattern Recognition Letters*, vol. 23, no. 5, pp. 569–577, 2002.

[107] S. Albrecht, J. Busch, M. Kloppenburg, F. Metze, and P. Tavan, "Generalised radial basis function networks for classification and novelty detection: self-organisation of optimal Bayesian decision," *Neural Networks*, vol. 13, pp. 1075–1093, 2000.

[108] S. Roberts and W. Penny, "Novelty, confidence and errors in connectionist systems," in *Proceedings of the IEE Colloquium on Intelligent Sensors and Fault Detection*, vol. 261, 1996, pp. 10/1–10/6.

[109] V. Vapnik, *Statistical learning theory.* Wiley/Interscience, 1998.

[110] B. Sch olkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," Massachusetts Institute of Technology, Tech. Rep., 1996.

[111] B. Sch olkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, pp. 1443–1471, 2001.

[112] D. Tax and R. Duin, "Uniform object generation for optimizing one-class classifiers," *Journal of Machine Learning Research*, vol. 2, pp. 155–173, 2001.

[113] A. Barla, E. Franceschi, F. Odone, and A. Verri, "Image kernels," in *Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines SVM 2002*, vol. 2388. Niagara Falls, Canada: Springer-Verlag Berlin Heidelberg, August 2002, pp. 83–96.

[114] C. Campbell and K. Bennett, "A linear programming approach to novelty detection," in *Advances in Neural Information Processing Systems*, vol. 13. MIT Press, 2001, pp. 395–401.

[115] L. Manevitz and M. Yousef, "One-class SVMs for document classification," *Journal of Machine Learning Research*, vol. 2, pp. 139–154, 2001.

[116] R. Santiago-Mozos, J. Leiva-Murillo, F. Pérez-Cruz, and A. Artés-Rodríguez, "Supervised PCA and SVM classifiers for object detection in infrared images," in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003, p. 122.

[117] O. Chapelle, P. Haffner, and V. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999.

[118] T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas, "Engineering applications of the self-organizing map," *Proceedings of the IEEE*, vol. 84, no. 10, pp. 1358–1384, 1996.

[119] S. Marsland, "Novelty detection in learning systems," *Neural Computing Surveys*, vol. 3, pp. 157–195, 2003.

[120] T. Harris, "Neural network in machine health monitoring," *Professional Engineering*, July/August 1993.

[121] A. Ypma and R. Duin, "Novelty detection using self-organizing maps," *Progress in Connectionist Based Information Systems*, vol. 2, pp. 1322–1325, 1998.

[122] V. Emamian, M. Kaveh, and A. H. Tewfik, "Robust clustering of acoustic emission signals using the kohonen network," in *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00)*. Washington, DC, USA: IEEE Computer Society, 2000, pp. 3891–3894.

[123] K. Labib and R. Vemuri, "NSOM: a real-time network-based intrusion detection system using self-organizing maps," 2002.

[124] S. Marsland, "On-line novelty detection through self-organisation, with application to inspection robotics," Ph.D. dissertation, University of Manchester, 2001.

[125] S. Marsland, J. Shapiro, and U. Nehmzow, "A self-organizing network that grows when required," *Neural Networks*, vol. 15, no. 8-9, pp. 1041–1058, 2002.

[126] S. Marsland, U. Nehmzow, and J. Shapiro, "On-line novelty detection for autonomous mobile robots," *Robotics and Autonomous Systems*, vol. 51, no. 2-3, pp. 191–206, 2005.

[127] G. Carpenter, M. Rubin, and W. Streilen, "ARTMAP-FD: familiarity discrimination applied to radar target recognition," in *Proceedings of the International Conference on Neural Networks*, vol. 3, 1997, pp. 1459–1464.

[128] H. Zhang, W. Huang, Z. Huang, and B. Zhang, "A kernel autoassociator approach to pattern classification," *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, vol. 35, no. 3, pp. 593–606, 2003.

[129] N. Japkowicz, C. Myers, and M. Gluck, "A novelty detection approach to classi-

fication," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence*, August 1995, pp. 518–523.

[130] C. Surace and K. Worden, "A novelty detection method to diagnose damage in structures: an application to an offshore platform," in *Proceedings of Eight International Conference of Off-shore and Polar Engineering*, vol. 4, 1998, pp. 64–70.

[131] K. Worden, G. Manson, and D. Allman, "Experimental validation of a structural health monitoring methodology: part I. novelty detection on a laboratory structure," *Journal of Sound and Vibration*, vol. 259, no. 2, pp. 323–343, 2003.

[132] L. Manevitz and M. Yousef, "Learning from positive data for document classification using neural networks," in *Proceedings of Second Bar-Ilan Workshop on Knowledge Discovery and Learning*, May 2000.

[133] D. Tax and R. Duin, "Outlier detection using classifier instability," in *Advances in Pattern Recognition, the Joint IAPR International Workshops*, 1998, pp. 593–601.

[134] J. Ko, Y. Ni, J. Wang, Z. Sun, and X. Zhou, "Studies of vibration-based damage detection of three cable-supported bridges in Hong Kong," in *Proceedings of the International Conference on Engineering and Technological Sciences*, 2000, pp. 105–112.

[135] C. Herry, "Quantitative assessment of pain through clinical digital infrared thermal imaging," Master's thesis, Carleton University, 2002.

[136] I. Daubechies, *Ten lectures on Wavelets.* SIAM: Philadelphia, PA, 1992.

[137] M. Antonini, M. Barlaud, and P. Mathieu, "Image coding using Wavelet Transform," *IEEE Transactions on Image Processing*, vol. 1, pp. 205–220, 1992.

[138] A. Cohen, I. Daubechies, and J. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Communications on Pure Applied Mathematics*, vol. 45, no. 5, pp. 485–560, 1992.

[139] R. Gonzales and R. Woods, *Digital image processing.* Pearson Education, 2002.

[140] C. Cortes and V. Vapnik, "Support-vector network," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[141] H.-T. Lin and C.-J. Lin, "A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods," Department of Computer Science and Information Engineering, National Taiwan University, Tech. Rep., 2003.

[142] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, March 1979, minimize inter class variance.

[143] S. Osher and S. J.A., "Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulation," *Journal of Computational Physics*, vol. 79, pp. 12–49, 1988.

[144] R. Malladi, J. Sethian, and B. Venuri, "Shape modeling with front propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 158–175, 1995.

[145] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, vol. 20, pp. 226–239, 1998.

[146] L. Xu, A. Krzyzak, and C. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, pp. 418–435, May-June 1992.

[147] H.-Y. Lee, K. Honga, and E. Kim, "Measurement protocol of womens nude breasts using a 3d scanning technique," *Applied Ergonomics*, vol. 35, pp. 353–359, 2004.

[148] D. Smith, W. Palin, V. Katch, and J. Bennett, "Breast volume and anthropomorphic measurements: normal values," *Plastic Reconstructive Surgery*, vol. 78, pp. 331–335, 1986.

[149] J. Vandeput and M. Nelissen, "Considerations on anthropometric measurements of the female breast," *Aesthetic Plastic Surgery*, vol. 22, pp. 348–355, 2002.

[150] R. Halir and J. Flusser, "Numerically stable direct least squares fitting of ellipses," in *Proceedings of WSCG'98*, 1998, pp. 125–132.

[151] D. Socolinsky and A. Selinger, "Thermal face recognition in an operational scenario," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, vol. II, 2004, pp. 1012–1019.

[152] C. Sun and D. Si, "Fast reflectional symmetry detection using orientation histograms," *Real-Time imaging*, vol. 5, no. 1, pp. 63–74, 1999.

[153] C. De Boor, *A practical guide to splines.* New York : Springer-Verlag, 1978.

[154] H. Delingette, "On smoothness measures of active contours and surfaces," in *Proceedings of the IEEE Workshop on Variational and Level Set Methods in Computer Vision*, 2001, pp. 43–50.

[155] G. Carpenter and G. S., *Adaptive Resonance Theory*, 2nd ed., ser. The Handbook of Brain Theory and Neural Networks. Cambridge, MA: MIT Press, 2003, pp. 87–90.

[156] G. Carpenter, S. Grossberg, and D. Rosen, "Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, pp. 759–771, 1991.

[157] C. Herry, R. Goubran, and M. Frize, "Segmentation of infrared images using cued morphological processing of edge maps," in *Proceedings of the 2007 IEEE Instrumentation and Measurement Technology Conference (IMTC07)*, Warsaw, Poland, 2007, pp. 1–6.

[158] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[159] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, pp. 151–172, 2000.

[160] C. Zaza, "Playing-related musculoskeletal disorders in musicians: a systematic review of incidence and prevalence," *Canadian Medical Association Journal*, vol. 158, no. 8, pp. 1019–1025, 1998.

[161] R. Tubiana and P. Amadio, Eds., *Medical problems of the instrumentalist musician.* Martin Dunitz Ltd, 2000.

[162] F. Bejjani, G. Kaye, and M. Benham, "Musculoskeletal and neuromuscular conditions of instrumental musicians," *Archives of Physical Medicine and Rehabilitation*, vol. 77, pp. 406–413, 1996.

[163] A. Brandfonbrener, "Musculoskeletal problems of instrumental musicians," *Hand Clinics*, vol. 19, pp. 213–239, 2003.

[164] C. Wynn-Parry, "Prevention of musicians' hand problems," *Hand Clinics*, vol. 19, no. 2, pp. 317–324, 2003.

[165] R. Lederman, "Neuromuscular and musculoskeletal problems in instrumental musicians," *Muscle and Nerve*, vol. 27, no. 5, pp. 549–561, 2003.

[166] G. Nourissat, P. Chamagne, and C. Dumontier, "Motifs de consultation des musiciens en chirurgie de la main," *Revue de chirurgie orthopédique*, vol. 89, no. 6, pp. 524–531, 2003.

[167] I. Winspur, "Advances in objective assessment of hand function and outcome: Assessment of the musician's hand," *Hand Clinics*, vol. 19, no. 3, pp. 483–493, 2003.

[168] D. Bishop, "Warm up I : Potential mechanisms and the effects of passive warm up on exercise performance," *Sports Medicine*, vol. 33, no. 7, pp. 439–454, 2003.

[169] ——, "Warm up II : Performance changes following active warm up and how to structure the warm up," *Sports Medicine*, vol. 33, no. 7, pp. 483–498, 2003.

[170] M. Bertsch and T. Maca, "Visualization of trumpet players' warm up," in *Proceedings of the 7th European Congress of Medical Thermology*, 1997, pp. 90–91.

[171] A. Grieco, E. Occhipinti, D. Colombini, O. Menoni, M. Bulgheroni, C. Frigo, and S. Boccardi, "Muscular effort and musculo-skeletal disorders in piano students: electromyographic, clinical and preventive aspects," *Ergonomics*, vol. 32, no. 7, pp. 697–716, 1989.

[172] C. Herry and M. Frize, "Design considerations for a medical thermographic expert system," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, September 2003, pp. 1252–1255.

[173] C. Herry, M. Frize, R. Goubran, and G. Comeau, "Evolution of the surface temperature of pianists' arm muscles using infrared thermography," in *Proceedings of the 27th annual international conference of the IEEE Engineering in Medicine and Biology Society*, September 2005, pp. 1687–1690.

[174] M. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[175] Maitra, "Moments invariants," *Proceedings of IEEE*, vol. 67, no. 4, pp. 697–699, 1979.

Appendix A

A.1 Thermographic Protocol

The following guidelines were followed prior to a thermographic imaging session:

- No talcum powder, lotion, drug or deodorant should be used on the skin on the day of the session.
- No alcoholic beverages should be consumed 24h prior to the session.
- No hot beverages should be used at least 1 hour prior to the session.
- Avoid the use of procedure such as electromyography, acupuncture, myelography, transcutaneous electrical stimulation, hot or cold patches, or any other form of physiotherapy at least 24h prior to the session.
- Avoid prolonged sun exposure at least one week prior to the session.
- You should not smoke at least 2 hours prior to the session.
- You should not wear any ring, necklace, and bracelet during the session.

- You should not do any intense physical exercises at least 4 hours prior to the session.

The session will take place in a room with the following characteristics (as much as possible):

- No windows or covered windows to avoid any reflection.
- Constant uniform temperature between 20 and 25 degrees Celsius.
- Constant humidity.
- No shiny or plastic floor.
- Fluorescent lights only.

A.2 Questionnaires

A.2.1 RSI Questionnaire

Part I

Identification Number:

Age:

Sex: [ ] Female [ ] Male

Height: Weight:

The following questions refer to your symptoms for a typical twenty-four-hour period during the past two weeks (circle one answer to each question).

How severe is the hand or wrist pain that you have at night?

1. I do not have hand or wrist pain at night.
2. Mild pain
3. Moderate pain
4. Severe pain
5. Very severe pain

How often did hand or wrist pain wake you up during a typical night in the past two weeks?

1. Never
2. Once
3. Two or three times
4. Four or five times
5. More than five times

Do you typically have pain in your hand or wrist during the daytime?

1. I never have pain during the day.
2. I have mild pain during the day.
3. I have moderate pain during the day.
4. I have severe pain during the day.
5. I have very severe pain during the day.

How often do you have hand or wrist pain during the daytime?

1. Never

2. Once or twice a day
3. Three to five times a day
4. More than five times a day
5. The pain is constant.

How long, on average, does an episode of pain last during the daytime?

1. I never get pain during the day.
2. Less than 16 minutes
3. 10 to 60 minutes
4. Greater than 60 minutes
5. The pain is constant throughout the day.

Do you have numbness (loss of sensation) in your hand?

1. No
2. I have mild numbness.
3. I have moderate numbness.
4. I have severe numbness.
5. I have very severe numbness.

Do you have weakness in your hand or wrist?

1. No weakness
2. Mild weakness
3. Moderate weakness
4. Severe weakness

5. Very severe weakness

Do you have tingling sensations in your hand?

1. No tingling
2. Mild tingling
3. Moderate tingling
4. Severe tingling
5. Very severe tingling

How severe is numbness (loss of sensation) or tingling at night?

1. I have no numbness or tingling at night.
2. Mild
3. Moderate
4. Severe
5. Very severe

How often did hand numbness or tingling wake you up during a typical night during the past two weeks?

1. Never
2. Once
3. Two or three times
4. Four or five times
5. More than five times

Do you have difficulty with the grasping and use of small objects such as keys or pens?

1. No difficulty
2. Mild difficulty
3. Moderate difficulty
4. Severe difficulty
5. Very severe difficulty

Part II

On a typical day during the past two weeks, have hand and wrist symptoms caused you to have any difficulty doing the activities listed below. Please circle one number that best describes your ability to do the activity.

Activity	No Difficulty	Mild Difficulty	Moderate Difficulty	Severe Difficulty	Cannot at all
Writing	1	2	3	4	5
Buttoning of clothes	1	2	3	4	5
Holding a book while reading	1	2	3	4	5
Gripping of a phone handle	1	2	3	4	5
Opening of jars	1	2	3	4	5
Household chores	1	2	3	4	5
Carrying of bags	1	2	3	4	5
Bathing and dressing	1	2	3	4	5

Part III

Did you have any injuries to your arms and forearms requiring surgical procedure?

[ ] Yes (Please list the injuries/surgical procedures)

[ ] No

Do you have sleep disorders?

[ ] Yes

[ ] No

Do you smoke?

[ ] Yes

[ ] No

How much time do you spend at the computer each day?

1. Less than an hour
2. One to two hours
3. Two to three hours
4. Three to four hours
5. More than four hours

What types of job do you do on the computer?

1. straight input or editing
2. intellectual use
3. mixed use

4. binge work

Do you most frequently use the mouse or the keyboard?

[ ] Mouse

[ ] Keyboard

How often do you take breaks when at the computer?

What do you do on a break?

Do you do other heavy work with your hands?

1. Handcrafts (for instance, carpentry, gardening, etc.)
2. Musical instrument
3. Other

Are you:

1. Left-handed?
2. Right-handed?
3. Ambidextrous?

Do you have any of the following conditions?

1. Diabetes
2. Collagen disorders
3. Kidney or thyroid disease
4. Peripheral neuropathy or with onset of symptoms after trauma.

Part IV

Please indicate on the following diagram all the areas where you experience numbness (Put a symbol 'N'), tingling (put a symbol 'T'), pain (put a symbol 'P'), decreased sensation (put a symbol 'S'):

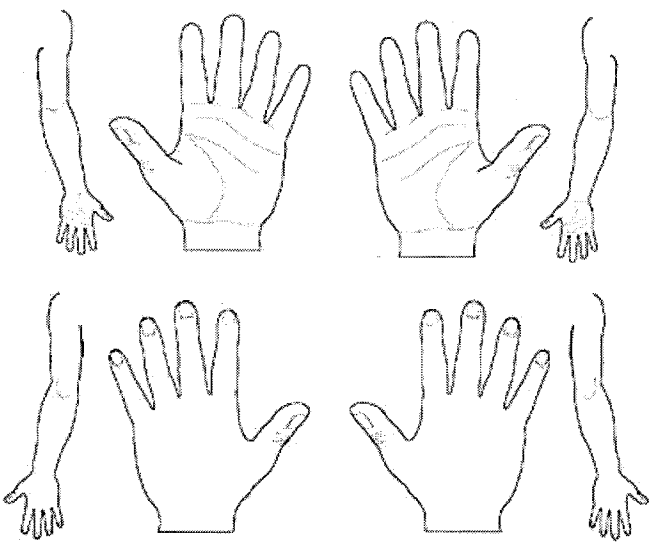


Figure A.1: Pain Hand Diagram

A.2.2 Pianist Questionnaire

Identification Number:

Age:

Gender: [ ] Female [ ] Male

1. How many years have you been playing the piano?

2. How many hours (or minutes) of piano do you play every day?

3. Do (or did) you play another instrument? If yes, please specify which instrument you have (or had) been playing and for how many years.

4. Do you have pain, weakness, lack of control, numbness, tingling, or other symptoms that interfere with your ability to play the piano at the level you are accustomed to?

5. How frequently do you experience these symptoms (pain, weakness, lack of control, numbness, tingling)? Please circle one answer.

   - Once in a while.
   - Regularly.
   - Constantly.
   - Other (specify)

6. Are the symptoms controllable?

7. Do you have any of the following health problems?

   - Joint pain, swelling, erythema

- Muscular pain, swelling
- Bone and joint deformities or infections
- Metabolic bone disease
- Compartment syndrome
- Avascular necrosis
- Osteoporosis
- arthritis
- Other health problems that would affect the musculoskeletal system

8. Did you have an accident resulting in an injury (e.g. broken arm) within the last 12 months?

9. Do you do a musical warm-up (e.g. playing slow scales or else)?

    - If yes, what do you play?

10. Do you do a physical warm-up before playing (e.g. stretching exercises, etc.)? Please circle one answer.

    - Never
    - Sometimes
    - Often
    - Always

11. If yes, what warm up exercises do you do?

12. Are there any other physical parameters that, according to you, influence the way you play the piano? If yes, please mention which ones.

13. Which pieces are you going to play?

14. Other relevant comments:

Appendix B

B.1 Moment Invariants

B.1.1 Central moments

Central moments are used to describe characteristics of an image. They are translation invariant. It is possible to obtain scale, translation and rotation invariance by using a combination of central moments. The definition of the various moments is given below. Central moments of order pq for an image $f(x, y)$ are found using the following formula:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \tag{B.1}$$

Where $\bar{x} = \frac{\sum_x \sum_y x f(x,y)}{\sum_x \sum_y f(x,y)}$ and $\bar{y} = \frac{\sum_x \sum_y y f(x,y)}{\sum_x \sum_y f(x,y)}$.

B.1.2 Hu's Moments

For $i + j \geq 2$, central moments can be made scale invariant using:

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{1+\frac{i+j}{2}}} \tag{B.2}$$

Hu used a combination of the scaled central moments to produce a set of moment invariants that are invariant to scale, translation and rotation changes [174].

$$\phi_1 = \eta_{20} + \eta_{02} \tag{B.3}$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \tag{B.4}$$

$$\phi_3 = (\eta_{30} + 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{B.5}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{B.6}$$

$$\begin{aligned}\phi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \\ & \eta_{03})^2) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)\end{aligned} \tag{B.7}$$

$$\phi_6 = (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \tag{B.8}$$

$$\begin{aligned}\phi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) + \\ & (3\eta_{12} - \eta_{30})(\eta_{21} - \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)\end{aligned} \tag{B.9}$$

B.1.3 Maitra's moment invariants

Maitra went further and proposed a combination of Hu's moments that make them invariant to contrast changes as well [175].

$$\beta_1 = \sqrt{\frac{\phi_2}{\phi_1}} \tag{B.10}$$

$$\beta_2 = \frac{\phi_3 \eta_{11}}{\phi_1 \phi_2} \tag{B.11}$$

$$\beta_3 = \frac{\phi_4}{\phi_3} \tag{B.12}$$

$$\beta_4 = \frac{\sqrt{\phi_5}}{\phi_4} \tag{B.13}$$

$$\beta_5 = \frac{\phi_6}{\phi_1 \phi_4} \tag{B.14}$$

$$\beta_6 = \frac{\phi_7}{\phi_5} \tag{B.15}$$

B.1.4 Zernike moments

Zernike introduced a set of complex polynomials which form a complete orthogonal set over the interior of the unit circle, i.e., $x^2 + y^2 = 1$. These polynomials are of the form:

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta) \tag{B.16}$$

Where n is a positive integer or zero, m is an integer such that $n - |m|$ is even and $|m| < n$, ρ is the length of the vector from the origin (centre) of the image to pixel (x,y) and θ is the angle between the previous vector and the x axis (counterclockwise).

$R_{nm}(\rho)$ is a radial polynomial defined as:

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2} (-1)^s \cdot \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n+|m|}{2}-s\right)!}\rho^{n-2s} \tag{B.17}$$

Zernike moments are the projection of the image function onto these orthogonal basis functions. The Zernike moment of order n with repetition m for an image $f(x,y)$ is:

$$A_{nm} = \frac{n+1}{\pi}\sum_x\sum_y f(x,y)V_{nm}(\rho,\theta) \qquad x^2+y^2 \leq 1 \tag{B.18}$$