Infrared Imaging Decision Aid Tools for Diagnosis of Necrotizing Enterocolitis

By

Yangyu Shi

Thesis submitted in partial fulfillment of the requirements for

Master of Applied Science

In

Biomedical Engineering

University of Ottawa

June 2020

Abstract

Neonatal necrotizing enterocolitis (NEC) is one of the most severe digestive tract emergencies in neonates, involving bowel edema, hemorrhage, and necrosis, and can lead to serious complications including death. Since it is difficult to diagnose early, the morbidity and mortality rates are high due to severe complications in later stages of NEC and thus early detection is key to the treatment of NEC. In this thesis, a novel automatic image acquisition and analysis system combining a color and depth (RGB-D) sensor with an infrared (IR) camera is proposed for NEC diagnosis. A design for sensors configuration and a data acquisition process are introduced. A calibration method between the three cameras is described which aims to ensure frames synchronization and observation consistency among the color, depth, and IR images. Subsequently, complete segmentation procedures based on the original color, depth, and IR information are proposed to automatically separate the human body from the background, remove other interfering items, identify feature points on the human body joints, distinguish the human torso and limbs, and extract the abdominal region of interest. Finally, first-order statistical analysis is performed on thermal data collected over the entire extracted abdominal region to compare differences in thermal data distribution between different patient groups. Experimental validation in a real clinical environment is reported and shows encouraging results.

Acknowledgements

First and foremost, I would like to pay my deepest gratitude and special regards to my supervisors, Dr. Payeur and Dr. Frize. They are respectable, responsible and resourceful scholars, who have provided me with valuable guidance in every stage of my master's study. Without their enlightening instructions, impressive kindness and patience, I could not have completed my thesis. Their keen and vigorous academic observation enlightens me not only in this research but also in my future study.

In the meanwhile, I shall extend my thanks to Dr. Bariciak for all her kindness and help during ethical application and experimental validation in hospitals. My sincere appreciation also goes to her for helping me revise my thesis, especially in the biomedical aspects of my work.

Finally, I would like to show my special regards to my parents and friends for their encouragement and support all through these years.

Table of Contents

List of Figures

List of Tables

Table of Acronyms

ANN	Artificial Neural Networks
APIs	Application Programming Interfaces
CHEO	Children's Hospital of Eastern Ontario
CT	Computed Tomography
CV	Chan-Vese
DLT	Direct Linear Transformation
ECG	Electrocardiograms
FCM	Fuzzy C-means algorithm
FOV	Field of View
GA	Gestational Age
HOG	Histogram of Oriented Gradients
IFOV	Instantaneous Field of View
IQR	Interquartile Range
IR	Thermal Infrared
MD	Mahalanobis Distance
MIEA	Multi-modal Immune Evolutionary Algorithm
MRI	Magnetic Resonance Imaging
NEC	Necrotizing Enterocolitis
NICU	Neonatal Intensive Care Units
OpenNI	Open Natural Interaction
PCNN	Pulse-Coupled Neural Network
RAC	Radial Alignment Constraint
RGB-D	RGB and depth
SLIC	Simple Linear Iterative Clustering algorithm
SVM	Support Vector Machine
TOF	Time-of-Flight
TOH-GC	Ottawa Hospital General Campus
UL	Upper-to-Lower

Chapter 1 Introduction

1.1 Motivation

Neonatal necrotizing enterocolitis (NEC) is a relatively common, devastating digestive system illness which primarily affects neonates and involves intense pain, inflammation, and even necrosis of the intestine. Although there is a wide spectrum of illness severity, the most severe cases can even be life threatening [1]. The incidence of NEC in preterm and very low birth weight (<1500g) infants is about 7%, however the mortality rate of severe NEC is as high as 20 to 30% [1,2]. In recent years, the incidence of NEC has increased along with the survival rate of very low birth weight infants. NEC not only increases the cost of medical care due to long-term hospitalization, but also increases the risk of long-term adverse outcomes, such as delayed development, short bowel syndrome, repeated infections, and long-term parenteral nutrition-related liver disease. The pathogenesis of NEC is still not fully understood, further contributing to the difficulty in making an early, accurate diagnosis [1-3].

Various medical imaging technologies have been applied to the diagnosis of NEC, with radiographic imaging being the most widely used. Physicians use abdominal x-ray images to determine whether there are typical clinical features of NEC, such as bowel wall thickening, gas accumulation within the wall of the intestine, portal venous gas, and peritoneal effusion [4, 5]. This requires that physicians are highly experienced in reading x-rays and that symptoms of NEC must be significantly advanced to be apparent on x-rays. According to the modified Bell staging criteria for NEC [6], the early signs of NEC are non-specific in radiographic images. Other non-NEC entities, such as benign feeding intolerance related to immaturity, can also present clinically and radiographically very similar to early stages of NEC. As a result, although x-ray imaging can detect more advanced stages of NEC, it is not ideal for its early detection.

Thermal Infrared (IR) imaging technology is a non-contact, non-radiative modality that has been studied in early detection of NEC in recent decades by measuring the differences in abdominal temperature profiles in babies with and without NEC [7, 8]. Since NEC is usually accompanied by the intestinal ischemic injury or even necrosis of the tract [9], it is expected that the abdominal temperature of NEC infants will change significantly compared to healthy babies.

Comparing and analyzing the different distributions of data in the thermogram between NEC babies and normal babies, it is postulated that early detection of NEC may be feasible.

In previous work done by our group using IR imaging in neonates [9, 10, 11], significant thermal distribution differences were found between normal babies and babies with NEC. This research, however, involved some limitations. First, the position of the IR camera was not standardized between patients, leading to variations in image size and quality that affect the ability to directly compare images. Second, the processing of the IR images involved a lot of repetitive work. It was time consuming when researchers needed to manually select the area of interest to analyze, which also introduced the possibility of biased results and a risk of missing crucial information. Furthermore, it is unrealistic for medical personnel in hospitals to spend much time on image processing. Therefore, the benefits of automatic procedures are significant as automation can help detecting NEC more efficiently and effectively, while also providing support for continuous observation of patients with IR sensors with minimal manual intervention.

1.2 Objectives

The solution proposed in this thesis attempts to provide an automated IR image acquisition and analysis system for NEC detection, which is more accurate, more uniform and easier to use by healthcare personnel. An RGB-D sensor is combined with a thermal IR camera to help provide additional and more accurate information that might be blurred in IR images or even missing in some cases. The system captures three sets of calibrated images synchronously (color and depth images from an RGB-D sensor, and IR images from an IR camera). It automatically segments and extracts regions of interest based on color and depth information, maps the selected areas to the corresponding IR images, and finally applies a thermal distribution analysis to help find significant differences in the thermal distribution over the abdominal region between groups of subjects (NEC vs healthy babies). In order to develop the solution, the following specific objectives are pursued:

i. Develop a fully integrated image acquisition system for recording image data, including color, depth, temperature information, and group categories, while ensuring that all the data are fully documented. Meanwhile, ensure that the RGB-D sensor and the IR camera operate with the same frame rate, as well as start and end recording at the same time.

ii. Design calibration and verification methods between the RGB-D sensor and IR camera. The geometrical correspondence among the color, depth, and IR images within the same set should be accurately estimated while remaining robust to varying surface shapes at different angles and positions. As a result, perform data alignment to ensure that images are in one-to-one correspondence.

iii. Make full use of color information and depth information to automate the segmentation of human body images. Segmentation should accurately separate the target from the background, identify the difference in skin color, distinguish each part of the human body, and extract the human body area of interest (i.e. the abdomen in this study). The segmented region is meant to guide the extraction of thermograms from IR data over the entire area of interest to ensure data completeness.

iv. Organize data and apply statistical analysis methodologies over the temperature distribution in the region of interest in order to detect significant differences between NEC babies and healthy babies.

1.3 Thesis Organization

Chapter 2 reviews the research work related to the use of infrared camera technology to detect NEC. Classical calibration and image segmentation techniques are also described. Chapter 3 presents an overview of the entire framework and methodology used in this work, including the design of the image acquisition assembly, calibration between devices, image processing and segmentation, extraction of regions of interest, analysis, and comparison of temperature distributions. Chapter 4 reports on an experimental verification conducted in an actual hospital environment. It details the formulation and rules followed as part of the ethical protocol for data collection in the hospital environment and provides extensive results and analysis from the available sample data. Finally, Chapter 5 concludes the thesis, highlights the contributions and discusses potential future work.

Appendix A provides the complete set of results for all ten normal babies considered in this study, where the identity of baby subjects is blurred by masking their face to address the privacy and confidentiality issue.

Chapter 2 Literature Review

This chapter presents a detailed introduction about the NEC disease (section 2.1), as well as its diagnosis methods (section 2.2), among which infrared thermography technology is highlighted due to converging evidence on its potential for NEC detection (section 2.3). Relevant work in the literature about calibration methods is also reviewed to support the manipulation of imaging acquisition devices (section 2.4). Furthermore, current studies on processing RGB-D and IR images and segmenting objects from complex background are presented and compared (section 2.5). Following that, classical data analysis methods on thermal data extracted over the abdominal region are introduced, including statistical representations that describe the temperature distribution on the entire abdomen, and boxplots that are used to compare the different trends of temperature data between NEC and normal babies (section 2.6). Finally, section 2.7 summaries the literature review.

2.1 Background Information about NEC

The clinical manifestations of NEC are diverse, and the symptoms are not specific. The early signs include abdominal distension, vomiting, abdominal wall swelling, and blood in the stool [12]. Therefore, the diagnosis of NEC cannot be made only by clinical manifestation, necessitating the addition of medical imaging methods. Some imaging methods being explored for NEC research include radiographic imaging, ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and infrared (IR) imaging, among which abdominal x-ray imaging is the most widely used method in the clinical environment [9,13].

According to the modified Bell's staging criteria, NEC can be divided into three stages [6, 14, 15]. In stage 1, the abnormal clinical and imaging findings are mostly non-specific. The early symptoms can be described as unstable body temperature, sudden apnea, mental state wilting, abdominal swelling, vomiting of bile, positive fecal occult blood, intestinal inflation, or signs of intestinal obstruction in x-ray images. In stage 2, there may be some mild metabolic acidosis and mild thrombocytopenia, with the possibility of being accompanied by the disappearance of bowel sounds, blood or black stool, obvious flatulence, functional bowl obstruction, and x-ray findings of intestinal wall thickening with air tracking into the intestinal wall (pneumatosis intestinalis). In

stage 3, clinical manifestations are more dramatic and include hypotension, severe apnea, severe respiratory, or metabolic acidosis. There are signs of peritonitis, abdominal tenderness, and severe abdominal distension. X-ray imaging may show signs of intestinal perforation as well as ascites. Although significant thickening of the intestinal wall gas accumulation within the intestinal wall, portal vein gas accumulation, and pneumoperitoneum can be easily seen in the middle and late diagnosis of NEC, they are still not evident in early stages of the disease.

2.2 Methods to Diagnose NEC

In recent years, studies have also shown that ultrasound imaging can assist in the diagnosis of NEC, as it is more sensitive to portal gas accumulation than x-ray images [16]. Ultrasound imaging can also be used to describe intestinal wall thickness and peritoneal effusion. Studies have shown that when ultrasound examination reveals that the intestinal wall is getting thinner and the blood flow is decreasing, it may indicate intestinal necrosis [17]. However, ultrasound examination of babies with bloating may be impaired due to the air obscuring the ultrasound signal. The acquisition of ultrasound images also requires the presence of an experienced technician and interpretation by a radiologist, who may not always be available.

Determination of the iohexol density value in urine samples using CT imaging after the administration of this water-soluble iodide contrast agent in the intestine is another method that has been investigated for diagnosing NEC [15, 18]. This contrast material is absorbed into the bloodstream through the damaged intestinal wall and then excreted into the urine when NEC is present, however it cannot reach the bloodstream or urine in normal babies who have an intact intestine. Despite the ability of this modality to detect NEC, it exposes patients to contrast material which may lead to unintended complications. MRI imaging technology can also be used as a non-invasive diagnostic tool for intestinal necrosis, which helps with the timing of NEC surgery and provides information on the location of intestinal necrosis during surgery [4], however this modality cannot be performed at the bedside, necessitating the transportation of very sick, unstable patients to the radiology suite for imaging.

Infrared image technology has been studied in recent years for NEC detection because the occurrence of NEC is accompanied by inflammation of the small intestine or colon, resulting in abnormal abdominal temperature [8]. By analyzing the differences between the heat maps of sick

and healthy infants, studies reveal that is possible to distinguish the occurrence of NEC disease [7, 8, 10,11,19]. Although infrared thermal imaging technology cannot replace existing diagnostic methods, it brings complementary resources for the detection of NEC.

2.3 Infrared Thermography

It is known that any object in the natural world above absolute zero (equal to -273 degrees Celsius) radiates energy outward through ongoing molecular vibrations. An infrared camera can passively collect the infrared radiation of the human body and convert this radiation information into an infrared thermal image that can be observed by a human. Because of the different anatomical structures, tissue metabolism, and blood circulation, the temperature of different body parts is not the same, and thus, the infrared image will form different thermal features [20].

Through the optical-electronic system, a thermal infrared camera converts the far-infrared light wave radiated by the human body into power signals through filtering aggregation. It converts the signal into a digital quantity after modulation and photoelectric conversion, which is processed by a multimedia image processing technology to display a heat map of the human body [21]. When there is a particular disease or functional change in a specific part of the human body, the blood flow and cellular metabolism change, resulting in a local temperature change, which is revealed as higher or lower temperature. If a person's whole body or local temperature deviates from normal levels, it may indicate a disease or some damage [22]. As a result, the heat maps corresponding to the standard level and to the abnormal level are compared to help physicians analyze and determine the location of the patient's lesion, the nature of the disease, and the extent of the lesion.

As early as 1800, British astronomer William Herschel discovered the relationship between heat and infrared radiation [23]. The evaporative thermal imager developed by Czerny et al. [24] in 1929 marked the birth of infrared thermal imaging technology. Infrared thermal imaging emerged in the 1960s, but this advanced technology was prematurely aborted due to technical and other reasons [25]. In the 1990s, due to the breakthrough and practical application of uncooled infrared focal plane array technology, it began to be used in the civilian field, covering many aspects of social life [22, 26].

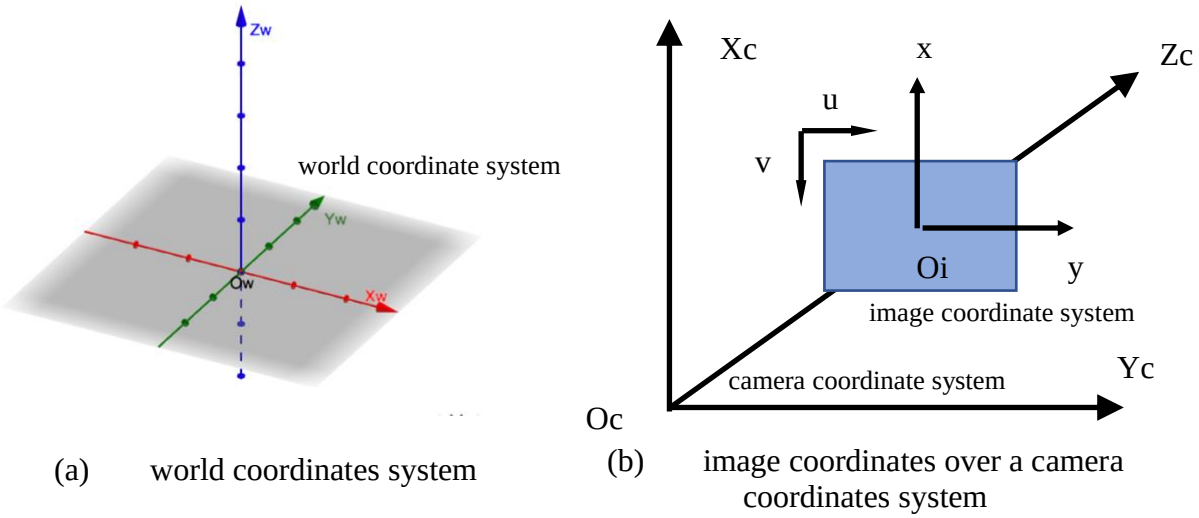
In Herry and Frize's work [27], a computerized assessment of thermal images was applied to patients with nociceptive and neuropathic pain. Data of digital thermal infrared images from

patients from the Moncton Hospital Pain Clinic were analyzed, and the results were very promising when compared to normal healthy volunteers at Carleton University. Similar processing techniques were also successfully applied to breast thermal images to detect cases affected by cancer [28].

When applied to the field of NEC diagnosis, Rice et al. [7, 8] were the first to publish about using thermography for NEC detection. They used medical thermography to help detect NEC in newborn infants by comparing temperature distribution over abdominal segments to those of the chest, and found that IR imaging can accurately detect the temperature differences between these anatomic regions of extremely low birthweight newborns with and without NEC. Herry et al. [29] compared thermal images of premature infants who were diagnosed with NEC to those of normal infants who did not have symptoms of intolerance. They found that the rate of change in the abdominal surface temperature of the two groups of infants was significantly different. During cooling, the abdominal temperature drop in the NEC group was slower than the other group, which supports the hypothesis that the IR imaging technique can be used to distinguish between NEC and healthy babies. Ntonfo et al. [11] manually selected a region on the abdominal temperature map, excluding the central area around the navel, to retrieve thermal features represented by sample distributions from values of the 8-bit grayscale palette. The distinction between NEC infants and healthy infants was made by comparing the differences in first-order statistical features in the sample data. In the study of Nur and Frize [9], the temperature characteristics of the baby's abdominal upper-to-lower (UL) region were used for detection. The study found that there were significant differences in the UL region of NEC infants, whereas normal infants had no such characteristics in the UL region. Although existing studies have demonstrated that the thermal distribution map generated by the infrared imaging of the abdomen of premature infants can be analyzed and classified to distinguish infants with and without NEC with the help of appropriate statistical methods and decision trees, research to date has overlooked the limitations introduced by manual selection of regions of interest. This requires input from research staff rather than being an autonomous tool for bedside use, and can lead to important areas on the thermogram being missed if they are not selected by the researcher.

2.4 Calibration

The purpose of camera calibration is to determine the relationship between the geometric position of one point on the surface in a three-dimensional space or object and its corresponding point in an image, and then to establish a geometric model of the imaging process. These geometric model parameters represent camera parameters. In a general camera model, there are three major coordinate systems: the world coordinates (X_W, Y_W, Z_W), the camera coordinates (X_C, Y_C, Z_C), and the image coordinates (u,v) [30]. Figure 2.1 illustrates the general representation.



(a) world coordinates system (b) image coordinates over a camera coordinates system

Figure 2.1 Three major coordinates systems in a general camera model (inspired from [30]).

The world coordinates, camera coordinates, and image plane coordinates systems do not coincide. Also, other factors influence the calibration [30, 31]:

(1) the distortion error of the camera lens, by which the imaging position on the image plane deviates from the perspective transformation projection result calculated by the linear transformation formula;

(2) the image coordinates unit in the computer is the number of discrete pixels in the memory, so the continuous coordinates on the image plane also need to be rounded;

The camera's internal parameters describe the relationship between camera coordinates and image coordinates. The external parameters describe the position and orientation of the camera in the world coordinates system.

Camera calibration methods are mainly divided into three categories: camera self-calibration methods, active vision-based methods, and traditional methods. At present, self-calibration algorithms mainly use the constraints of camera movement, however the camera's motion constraints are too strong, making it impractical in practice. The use of scene constraints mainly refers to some parallel or orthogonal information in the scene. The intersection point of the spatial parallel line on the camera image plane is called the vanishing point. This is a very important feature in projective geometry, so many scholars have studied the camera self-calibration method based on the vanishing point [32, 33, 34]. The self-calibration method is flexible and can be used to calibrate the camera online, but because it is based on the method of absolute quadratic curve or surface, its algorithm is not robust [35].

The camera calibration method based on active vision refers to the calibration of the camera with certain motion information of the known camera. This method does not require a calibration object but needs to control the camera to make some special movements, and the internal parameters of the camera can be calculated by using the particularity of this movement. The advantage of active vision-based camera calibration is that the algorithm is simple and often leads to linear solutions, so it is more robust. The disadvantages are high system cost, expensive experimental equipment, and high experimental conditions [36, 37].

Conventional camera calibration methods use known scene structure information to complete calibration, where calibration blocks are commonly used. It has high calibration accuracy and can be used in any camera model. Although the traditional camera calibration method always requires a calibration object in the calibration process, and the accuracy of the calibration object's production will affect the calibration result, compared to the other two calibration methods, the traditional camera calibration method is still widely adopted [38].

In 1971, Abdel-Aziz et al. [39] proposed the direct linear transformation method (DLT), that directly relates the linear coordinates between the world and the image coordinates of corresponding object points. Shapiro [40], on this basis, developed a method for the field of 3D high-speed photography, where two high-speed cameras were used to take points with known spatial coordinates in order to determine the spatial coordinates of unknown points. The method

of Abdel-Aziz et al. [39] was then applied to the reduction of photogrammetric data. It did not involve iterations and was therefore unaffected by computational errors due to iterations. This method provided good accuracy when the traditional colinear method failed due to the lack of unknown parameters (internal or external parameters).

The two-step method based on a radial alignment constraint (RAC) proposed by Tsai [41] first uses the linear transformation method to solve the camera parameters, and then considers the obtained parameters as initial values. It uses a nonlinear optimization method to further improve the calibration accuracy by considering the distortion factor. More recently, Deng et al. [42] proposed a combination of differential evolution and particle swarm optimization algorithms to determine the internal and external parameters of the camera with both image features and corresponding 3D features. This method can better avoid local optimum and improve calibration precision.

Zhang's camera calibration method [43] relies on a single-plane checkerboard. Before this approach was introduced, traditional calibration methods involved a three-dimensional and very precise calibration board, which was challenging to manufacture. Zhang’s approach overcomes the shortcomings of high-precision calibration while only requiring the use of a printed checkerboard. As a result, when compared with self-calibration, the accuracy is improved, and the operation made convenient. Figure 2.2 illustrates the overall flow chart of the calibration procedures proposed by Zhang [43].

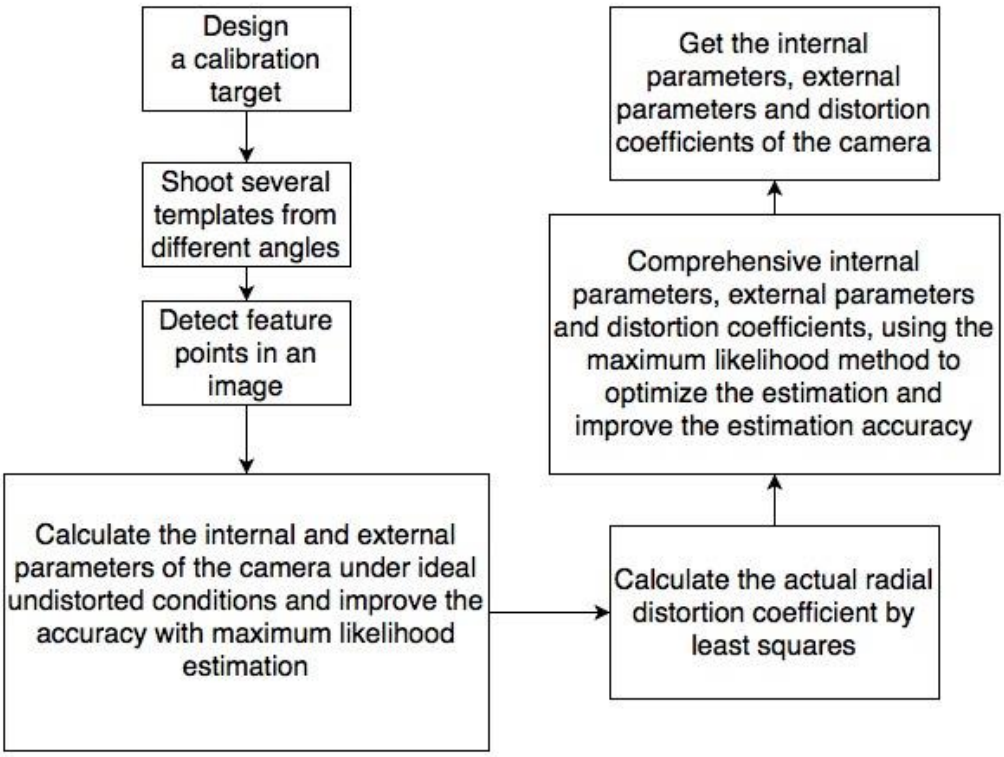


Figure 2.2 Overall flow chart of the calibration procedures proposed by Zhang (inspired from [43]).

Zhang et al. [44] also proposed a technique for performing binocular calibration using both environmental and calibration object information. The method accurately estimates the available matching epipolar geometry from the scene and calibration objects for projective reconstruction. According to prior knowledge available on the calibration object, the method performs the calibration of the stereo device in the Euclidean space. Recently, Lu et al. [45] applied Zhang's camera calibration method to 3D reconstruction. With the help of the pinhole camera model and the three-coordinate relationship, Zhang 's calibration method was used to figure out the internal and external parameter matrix of the camera, the camera distortion coefficient, radial and tangential distortion. This method showed higher calibration accuracy and proposed a novel way to seek depth from binocular stereo vision. Zhao et al. [46] took the actual pixel coordinates of the camera image as the input value and used the re-projection point obtained by the world coordinates as the output value according to Zhang's calibration method. A "black box system" was proposed to build a connection between the ideal image point and the actual image point.

Wu et al.'s calibration method [47] uses a set of parallel circles as calibration objects. The intersection of the quadratic curves on the image is calculated from the minimum number of parallel circles to complete the calibration based on the quasi-affine invariance of camera imaging. Because the method for calculating the circular point image is simple enough, and only needs to start from the fitted quadratic curve without any matching instead of calculating the center of the circle, Wu's calibration approach has been widely applied, not only in the case of a plane, but also for turntable based reconstruction [47].

2.5 Segmentation

Medical image segmentation is a complex and critical step in the field of medical image processing and analysis. The purpose is to segment the parts of the medical image with some specific meanings and extract relevant features, to provide a reliable basis for clinical diagnosis and pathology research, and to assist doctors in making a more accurate diagnosis. Due to the complexity of medical images, a series of problems such as unevenness and individual differences need to be resolved during the segmentation process, therefore general image segmentation methods are difficult to apply directly to medical image segmentation. The latter is still developing from manual segmentation or semi-automatic segmentation to fully automatic segmentation.

Threshold segmentation is the most common segmentation method for parallel direct detection of regions [48]. If only one threshold is selected, it is called single-threshold segmentation, which divides the image into targets and backgrounds [49]; if multiple thresholds are required, it is called the multi-threshold method, and the image will be segmented into multiple target regions and background areas [50]. Differentiating targets requires marking each area. The threshold-based segmentation method is based on a hypothesis of grayscale images: the grayscale values between adjacent pixels in the target or background are similar, but the pixels of different targets or backgrounds differ in grayscale. That is, different targets and backgrounds correspond to different peaks when reflected in the image histogram. The threshold chosen should be in the valley between the two peaks to separate the peaks. The advantage of threshold segmentation is that it is relatively simple to implement. When the gray value or other feature values of different types of objects are very different, it can effectively segment the image. Threshold segmentation is usually used as pre-processing on medical images, and then a series of other segmentation

methods are applied for post-processing [51]. The disadvantage of threshold segmentation is that it is not suitable for multi-channel images or images with little difference in eigenvalues [52]. It is difficult to obtain accurate image segmentation results since there is no obvious gray difference in the image or the gray value range of each object has a large overlap. In addition, because it only considers the gray information of the image and not the spatial information of the image, threshold segmentation is sensitive to noise and gray unevenness.

Most medical images are blurry, with low image quality and high noise. The fuzzy clustering method combines the fuzzy set theory with the clustering algorithm applied to the field of medical image segmentation since the fuzzy set theory has a good ability to describe the uncertainty of the image. This method does not rigidly divide pixels into a certain area in a "one-size-fits-all" manner, but introduces the concept of "membership" in fuzzy theory to divide pixels into areas with a high degree of membership, improving the accuracy of segmentation. Currently the most commonly used is the fuzzy C-means algorithm (FCM) [53], which obtains the optimal boundary through two iterations. However, there are problems in using FCM directly for medical image segmentation: the segmentation result is more sensitive to the initial value, and it is easy to fall into a local minimum during iteration. Based on these issues, Haldar et al. [54] used Mahalanobis Distance (MD) to obtain the initial cluster center, which replaces the randomly initialized cluster center in the original algorithm and improves accuracy when processing original fuzzy electrocardiograms (ECG) images due to the limitations of scanning or shooting conditions.

When it comes to IR human detection, due to the low signal-to-noise ratio and contrast of the infrared image, the halo effect around the target, and blurred edges, it is difficult to segment the human target from complicated background in the infrared image. Tan et al. [55] improved the background subtraction method based on the mixed Gaussian model. During the binarization phase, an improved pulse-coupled neural network (PCNN) was used for fine segmentation, and the multi-modal immune evolutionary algorithm (MIEA) was used to determine the PCNN parameters automatically. Simulation results show that the algorithm has high accuracy in human thermal image segmentation. Ma et al. [56] also used the background subtraction method to segment infrared human targets, described multiple changing backgrounds through multiple Gaussian processes with weights, and updated the number of models, weights, and learning rates. Then, the segmented area of interest was characterized by the accumulation of oriented edge and histogram of oriented gradient. Finally, the support vector machine was used to realize the classification and

detection of human targets. Experiments show that this algorithm can correctly detect human targets in complex scenes, and also has a good effect on situations where multiple targets are close or even partially adhered. Although the background subtraction method based on the mixed Gaussian model can effectively solve the problem of background disturbance [57, 58], the disadvantage is that it has a large computational cost. Under the Intel Xeon 5150 processor hardware environment, the algorithm can only process 7 frames of 640 * 480 resolution video images per second [59]. Model parameters (especially variances) can also be problematic in a noisy environment in the real world. That has led to people to using a fixed variance when implementing algorithms in hardware [60]. Finally, it should be noted that many natural images do not actually correspond to Gaussian models [61].

Teaney et al. [62] used static human targets in single-frame infrared image as research objects and conducted research on human target detection techniques in complex environments. Firstly, based on the principle of infrared imaging and the changes of human skin temperature, an empirical formula for estimating the central gray level of the human target was obtained by statistical regression analysis. Then, based on the principle of clustering, a new target clustering method was proposed. This method used the gray value corresponding to the ambient temperature and that corresponding to the human target skin temperature as the clustering center and set the clustering threshold to one third of the Euclidean distance between the two clustering centers. Although the experimental results show that this method can quickly and effectively detect human targets in infrared images with complex backgrounds, the recognition of specific parts of the human body has not been well solved.

Aiming at the problem of inaccurate segmentation of the human body in complex scenes only by infrared images, a Kinect sensor was proposed to obtain better segmentation results with the help of color and depth information. For the human body segmentation method based on Kinect depth camera, Nava et al. [63] proposed a human body segmentation algorithm combining RGB-D and skeleton information in a graph theory optimization framework. The algorithm considers the mutual promotion of the color, depth, and skeleton information of the image in the 3D scene to obtain more accurate human segmentation results. First, the edge-guided filtering algorithm was used to repair low-quality depth maps to obtain high-quality depth maps. Then they used a K-means-like clustering algorithm to cluster RGB-D data to obtain super-pixels. Finally, super-pixels were regarded as nodes in the graph model, and the corresponding human skeleton information

was combined to improve the ability to distinguish between humans and backgrounds with similarly colored areas. In order to pursue faster detection speed, Sun et al. [64] proposed a human detection algorithm based on the histogram of oriented gradients (HOG), using integral histograms to calculate HOG features, and training the samples with a cascaded support vector machine (SVM) classifier. This method, however, requires a large amount of sample data for training. Also, the generation process of the HOG feature descriptor is lengthy, resulting in slow training speed and high equipment requirements. In addition, it is difficult to deal with occlusion problems based on HOG features. Moreover, due to the nature of the gradient, this descriptor is quite sensitive to noise.

The color information provided by Kinect sensor can be effectively used to detect human contours based on skin color. There are many problems with the skin color detection technology proposed in the past: the skin color model with a fixed threshold value lacks the adaptability to different illumination conditions and different image backgrounds [65, 66, 67]; lack of effective removal methods for a large number of confused backgrounds in the image, leading to an excessively high false detection rate of skin color [68, 69]. In response to the above problems, Nitsuwat [70] proposed an adaptive threshold skin color detection algorithm based on a statistical color model. According to the skin color probability distribution histogram of the image and four heuristic rules, used the artificial neural networks (ANN) algorithm to obtain the optimal threshold value. Also, an image relation operation was proposed to remove the confused background to a certain extent. Bergasa et al. [71] proposed a detection method for dynamic modeling of human skin color. Based on the results of face detection, a targeted human skin color model was established online combined with the mixed Gaussian model. Ban et al. [72] combined the skin color probability variance histogram and the mixed Gaussian model to remove a variety of confused backgrounds in the image by using a piecewise processing method. Shi et al. [73] first extracted skin color samples and established an adaptive Gaussian model to extract skin color regions, and then extracted clothing regions in HSV color space segmentation, and finally linearly fused the above two detection regions to achieve detection of human body regions. The skin color model can assist in extracting relatively accurate and complete human body regions, which is a crucial step for completing human body segmentation in medical images.

After extracting the human body, many algorithms used the obtained image sequences for body segmentation and motion analysis. For that matter, Chen et al. [74] proposed a skeleton model-based method. Leung et al. [75] proposed a two-dimensional model based on human

contours, and Rohr [76] proposed a method based on a three-dimensional human body model consisting of a cylinder. To support motion analysis, low-level features, such as the position of each joint, were extracted from successive image frames. These features were mainly obtained by artificially pasting color markers on various parts of the human body. Then a correspondence relationship between the features of each frame was established, and finally the three-dimensional human body structure from the correspondence between the features was recovered. The limitation of these algorithms is that they can only be applied to a simple background.

In order to improve the above problems, Fan et al. [77] proposed a joint position determination method based on human contours. This method first extracted the contour of the human body from the image, then used the energy function to extract a virtual skeleton similar to the real human skeleton, and then used the knowledge of human anatomy to determine the position of the joint. This method does not need to restrict the motion and color information of the research object, and also has better performance on more complex backgrounds. Junior et al. [78] proposed an algorithm based on a classifier and a graph structure. First, they used the edges as features to train the model, so that the initial distribution probability of each part can be obtained separately, and then they used the classifier for preliminary segmentation. After that, they extracted the regional features on the segmentation results obtained in the first step, retrained the model, and then performed positioning and segmentation procedures for each part. In this way, the edge and area features were combined together through the cascade of the two to complete the human segmentation. This method is computationally intensive because the detection templates need to be trained for each part and the parameters for each template need to be calculated through a conditional random field algorithm. More recently, Ren et al. [79] aimed at the problem of simple linear iterative clustering algorithm (SLIC) that needs to specify the number of pixel blocks when performing super pixel block segmentation. Based on the Chan-Vese (CV) energy model, an adaptive super-pixel block was constructed by minimizing the picture into multiple regions and performing level set iterative segmentation so that each super-pixel block, after the segmentation, fits more closely to a single-color block in the image. Then combined with the average human body template, the standard pose area of the human body was marked on the picture, which improved the anti-interference ability for complex backgrounds. Finally, a k-means clustering algorithm was used to cluster each super-pixel block as a node to achieve standard human image segmentation. This algorithm improves the segmentation accuracy of the human body's standard

pose while ensuring the efficiency of image segmentation and has strong anti-interference ability for complex backgrounds with rich chromaticity.

Although much is known about medical IR image processing and human body segmentation based on color and depth information, segmenting the special shape of a baby from the background with inputs from a variety of sources (color, depth, and IR) has remained widely unexplored, and untested.

2.6 Statistical Analysis

2.6.1. Statistical representation and data analysis

In order to grasp the nature of the data as a whole, it is imperative to choose appropriate statistics to represent its characteristics, especially the concentration, dispersion, and distribution shape of the sample data, which can determine the cardinal properties of the whole data set and give an excellent reference for subsequent data analysis. When applied to distinguish between the NEC and normal groups, due to the rarity of NEC infant cases ($n < 10$), some common medical statistical tests (such as t-test, chi-square test, and analysis of variance) can lead to biased results. Therefore, some basic statistics can better describe the trend and dispersion of the data [80].

There are three main categories in the basic statistics used to describe the data: central tendency, dispersion, and distribution shape.

2.6.1.1 Central tendency

The central tendency [81] is the statistic that represents the location of the specific data among the whole, that is, intuitively, where does its value mostly fall when given an attribute? This parameter is typically represented by the mean, median and mode variables, which are defined as follows:

(1) Mean

Mean, also known as the arithmetic mean, describes the data that guides the average position.

$$\bar{X} = (\sum x_i)/ n \tag{2.1}$$

Situationally, each value in a set of data can be associated with a weight, w_i, which reflects the importance of the attached value or the frequency of occurrence. That is termed the weighted arithmetic mean.

$$\text{weighted arithmetic mean} = (\sum(x_i w_i))/ n \quad (2.2)$$

Although the mean is the most helpful statistic to describe the trend of the dataset, it is not always the best way to measure the data center because mean is sensitive to extreme values (outliers). To counteract the effects of a few extreme values, researchers also use the trimmed mean that is the average of data values after dropping a certain number of the extremes.

(2) Median

For oblique (asymmetric) data, the data center statistic is better described as its median, which is the median of the ordered data values. The median can keep data set from being affected by extreme values and measure the overall medium condition of the data. Since the median is the representative value of the total unit values determined by its position among all the values, it is not affected by the maximum or minimum value of the distribution series, thereby increasing the median's representation of the distribution series to some extent. Nevertheless, in some monomial series of discrete variables, if the frequency distribution is skewed, the representation of the median will be affected.

(3) Mode

Mode is the most frequently occurring value in a variable. It is often used to determine the mode for qualitative data. It is less reliable to use only a mode to quantify a set of data, although extremes rarely influence it. The mode is especially beneficial when the values or the observers do not have a definite order (often occurring in non-numerical data), as the arithmetic mean and median may not be well defined.

2.6.1.2 Dispersion

The statistics that measure the degree of data dispersion are mainly the standard deviation and the quartile range.

(1) Standard deviation (or variance)

The standard deviation is used to measure the degree of dispersion of the data distribution. A low standard deviation means that the data observation tends to be close to the mean, and a high standard deviation means that the data spread over a broad value range.

$$\text{Standard deviation} = \sqrt{\frac{1}{n-1}\sum(X_i - \bar{X})^2} \quad (2.3)$$

(2) Quartile range

A quartile is a type of quantile in statistics, that is, when all values are arranged in ascending order and divided into four equal parts, the first quartile is recorded as Q1, also known as the "minor quartile", which is equal to 25% of the numbers in the sample. The second quartile (Q2), also known as the "median", is equal to 50% of the numbers. The third quartile (Q3), known as the “larger quartile”, means data on the 75th percentile. The difference between the third quartile and the first quartile is also called the interquartile range, which gives the range covered by the middle half of the data, indicating a simple measure of the degree of dispersion

$$\text{Interquartile range (IQR)} = Q3 - Q1 \quad (2.4)$$

2.6.1.3 Distribution shape

The distribution shape is measured using the skewness and the kurtosis coefficients.

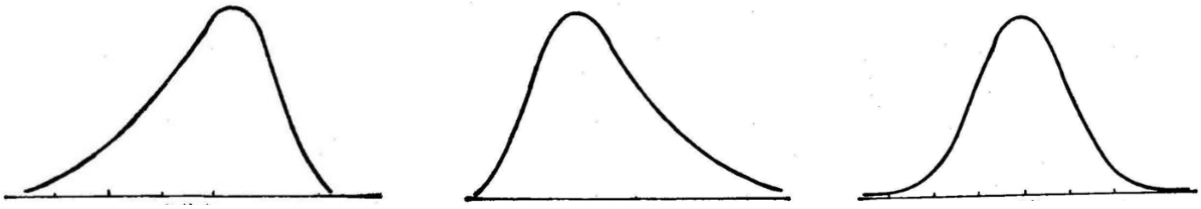
(1) Skewness

Skewness is a measure of the direction and extent of statistical data skew and is a numerical feature of the degree of asymmetry in the distribution of statistical data. Intuitively, it is the relative length of the tail of the density function curve. The definition of the skewness is the third-order normalized moment of the sample, which is defined as follows, where k_2 and k_3 represent the second order and third order central moments, respectively:

$$Skew(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{k_3}{\sigma^3} = \frac{k_3}{k_2^{3/2}} \quad (2.5)$$

The skewness of a normal distribution is 0, and the lengths of the tails on both sides are symmetrical. Assuming that sk is used to indicate skewness, if sk<0, then the distribution has a negative deviation, also called the left-skewed state, where the data are mainly located on the right side of the mean value and less on the left side. The visual representation is that the left tail is longer than the right tail because there are a few variables with small values causing that the tail on the left side of the curve is dragged longer. If sk>0, it reveals that the distribution has a positive deviation, also known as the right-skewed state, in which more data are located on the left side of the mean than that on the right side, visually expressed as the right tail is longer relative to the left

tail, because there are a few variables with large values that make the tail on the right side of the curve drag long. If sk is close to 0, the distribution is considered symmetric. If it is noted that the distribution may deviate from the normal distribution in the skewness, the deviation can be used to test the normality of the distribution [82]. In the right deviation, arithmetic mean > median > mode, while in the left bias, it is the opposite such that the mode > median > mean. All three values are equal in the normal distribution. Figure 2.3 depicts data distributions with various skewness [82].



(a) sk < 0 (b) sk > 0 (c) sk = 0

Figure 2.3 (a) If the skewness is negative, the dispersion on the left side of the x-mean is stronger than the right side; (b) If the skewness is positive, the dispersion on the left side of the x-mean is weaker than the right side; (c) For a normal distribution (or strictly symmetric distribution) skewness equals to 0 [82].

(2) Kurtosis

Kurtosis studies the steep or smooth distribution of data. By measuring the kurtosis coefficient, we can determine whether the data distribution is steeper or more gradual relative to the normal distribution.

The kurtosis is contrasted with respect to the normal distribution. The kurtosis coefficient of the normal distribution is 0, that of the uniform distribution is -1.2, and that of the exponential distribution is 6 [83]. When the kurtosis > 0, it is more steep or thicker than the normal distribution in terms of morphology, while when the kurtosis < 0, it is more gradual than the normal distribution, or the tail is thinner. In the actual environment, if one part is thick tailed, this distribution tends to have a more substantial "quality" than the tail of the normal distribution because it contains more extreme values. Based on D'Agostino [82], the tail and outliers have a positive and the most significant impact on the kurtosis. The high probability zone also has a positive but lighter effect on the kurtosis, while in the mountainside position, the medium probability zone has a negative influence.

2.6.2. Box plot

A box plot is a kind of statistical graphs used to display the data dispersion. It is mainly used to reflect the characteristics of the original data distribution and also to compare the distribution characteristics of multiple sets of data. The box plot was proposed in 1974 by Friedman and Tukey [84]. The box plot provides a way to make a simple summary of the data set with only five points, including the midpoint, Q1, Q3, and the high and low positions of the distribution state, and thus, it calculates the relevant statistical points with the percentile calculation method.

An example of a boxplot is given in Figure 2.4. where the line across the box is the median value. The upper and lower limits of the box are the upper quartile and the lower quartile of the data, which means that the box contains 50% of the data. Therefore, the height of the box reflects the degree of fluctuation of the data to some extent. If the boxes are all relatively long, this indicates that the data are more discrete, while if the boxes are shorter, then the data are more concentrated. From this boxplot, it is clearly seen that most data are in the range of 175 to 250, with median around 245. In addition, the strength of the data distribution skewness can be judged according to the position of the median line, and the median deviation from the center position of the upper and lower quartiles indicates that the distribution skewness is stronger than average. Since the median line is closer to the upper quartile, it indicates that most of the data are on the side with larger values and distributed on the right side of the X-axis, the tail of the curve extends to the left, and the data are left-biased. Besides, the outliers are concentrated on the side of the smaller value, also indicating that the data distribution is left-skewed. In this example, there are not many outliers, indicating that the tail is light and the degree of freedom, that is, the number of freely varying amounts, is small;

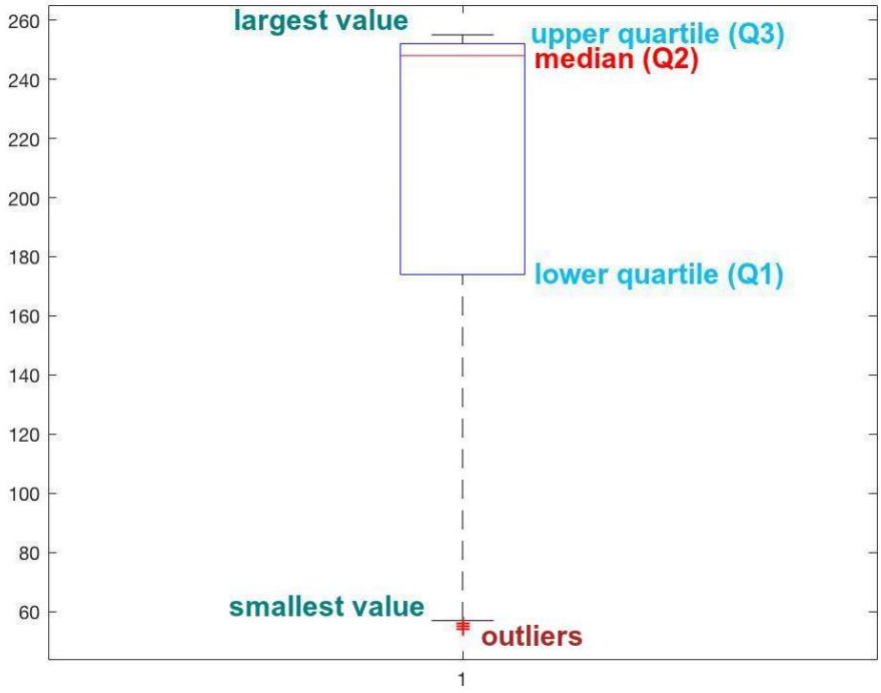


Figure 2.4 An example of a boxplot.

The box plot provides a standard for identifying outliers, which are defined as values less than Q1-1.5IQR or higher than Q3+1.5IQR. Although this standard is somewhat arbitrary, it comes from empirical judgment, and experience shows that it performs well in dealing with data that requires special attention, which is quite different from the classical method of identifying outliers. It is well known that the 3σ rule or the z-score method based on normal distribution presupposes that the data is assumed to follow a normal distribution, but the actual data often do not strictly obey the normal distribution. Besides, their criteria for judging outliers are based on the mean and standard deviation of the data batches, but the resistances of the mean and standard deviation are extremely small, and the outliers themselves have a massive impact on them, so the number of outliers generated is no more than 0.7% of the total [85]. The application of this method to determine outliers in non-normally distributed data is limited in its effectiveness.

In contrast to that, drawing box plot relies on the actual data. It is not necessary to assume that the data is subject to a specific distribution form, and there is no restriction on the data. It is only a pure and intuitive representation of the original shape of the data. On the other hand, the

criteria for determining the outliers of the box plot are based on quartiles and interquartile ranges. Quartiles have some resistance given that up to 25% of the data can be arbitrarily far without disturbing the quartiles, so outliers cannot make a difference on this standard, and the results to identify outliers are objective. Hence one can see that the box plot has certain advantages in identifying outliers [86]. When studying the temperature of the baby's abdomen, due to the complex symptoms of NEC, the appearance of abnormal values may not be obvious. Thus, boxplots can well avoid the interference of unknown outliers on experimental results and have been applied in the field of NEC detection [87].

On the same axis, the box plots of several batches of data can be arranged in parallel, and thus the comparisons of the median, tail length, outliers, and distribution intervals of several batches of data are manifest at a glance. In one batch of data that belongs to one group of infants, it is obvious to see which data points outstanding, which data points are not as reliable as usual, where these data points are placed in other similar groups, and so on. If the box plots of several batches of data of the same group are compared, the analysis and evaluation is the visual illustration of the norm-referenced method; if the box plot of two different groups (e.g. NEC and normal groups) are compared, it is the visual illustration of the validity criterion. In this thesis, the box plot helps the analysis process to be fast and straightforward when discriminating between NEC and normal babies.

2.7 Summary of Literature Review

This chapter briefly introduces the pathological background of NEC, followed by a detailed review of existing related infrared technologies. Also, when it comes to the calibration of sensors and infrared cameras, a comparison of the advantages and disadvantages of classical algorithms is described. Then the categorization and summary of current image segmentation techniques based on IR and RGB-D images are discussed. In the last section, background information on elementary statistical concepts and boxplots are supplied to convey some ideas related to data analysis.

Chapter 3 Methodology

Based on the findings discussed in chapter 2, the purpose of the present study is to conduct a more nuanced examination of the distribution of thermal data over infants' abdominal area with the help of a RGB-D sensor. An automated IR imaging acquisition and processing system is proposed to synchronously collect IR, depth, and color images of the complete infant body, then to segment a region of interest related with NEC disease from the complex background, and finally to analyze the differences between the thermograms of NEC and normal groups.

In this study, a novel multispectral acquisition platform is designed where a Microsoft Kinect Xbox One sensor is bundled and calibrated with a FLIR A320 infrared (IR) camera to synchronously acquire information on the human body in real-time, and then to automatically extract more accurate abdominal area thermographic data. The Kinect sensor is used to capture color and depth images, while the IR camera records the temperature distribution over the human body. After the former is accurately calibrated with the latter, three sets of images (color, depth, and IR) are simultaneously acquired at the same frequency (30 frames per second). With the help of the Kinect sensor, the system automatically locates and extracts the relevant abdominal area without requiring manual intervention since symptoms of NEC disease usually only occur in the abdominal area. Moreover, due to the different nature of the sensors, the additional color and depth information succeeds to compensate for fuzzy contours that are typically observed in infrared images, which helps to segment the subject from the background, to clear environmental interferences, and to segment and extract the trunk region over the human body. As a result, the approach improves the accuracy of the temperature distribution analysis within the region of interest.

Figure 3.1 shows a flow chart of the proposed automated multispectral acquisition and processing system. It consists of three main components: acquisition platform design and calibration, segmentation and feature extraction from multispectral images, and thermal distribution analysis from infrared images. After the Kinect sensor and the FLIR IR camera are installed at a fixed position and angle, a total of three cameras embedded in those two devices are calibrated to achieve simultaneous acquisition of corresponding pixels in between the three sets of images (color, depth, and IR). Subsequently, an algorithm is proposed for image segmentation and processing procedures, including background removal, skin color segmentation, human body parts

recognition and separation, and data extraction from the region of interest in the IR image. Finally, pixel values related to temperature statistics are assembled into thermal signatures and the latter are analyzed to find significant variations between the NEC group and the normal group, based on classical statistical analysis methods.

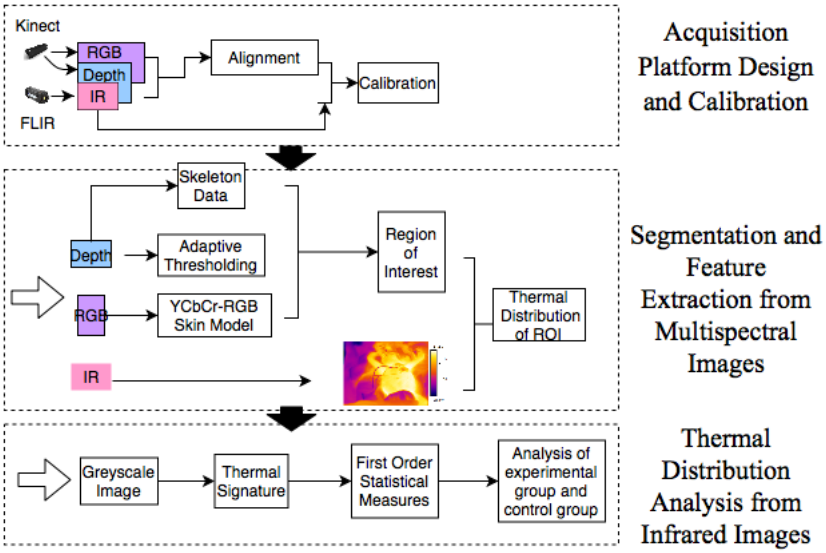


Figure 3.1 Flow chart of the proposed automated multispectral acquisition and processing system.

3.1 Design of Multispectral Acquisition System

3.1.1. Microsoft Kinect Xbox One sensor

The Kinect Xbox One sensor is a body-sensing device launched by Microsoft in 2013. Because it provides a simple and efficient communication interface between people and machines, the depth sensor has been a focal research topic since then. The sensor's core chip is designed by PrimeSense, which is especially qualified for complex computing and algorithm design [88].

The Kinect sensor consists of one USB3.0 bus interface, one color camera, one three-dimensional depth sensor and a microphone array. Figure 3.2 shows the overall hardware specifications of the Kinect sensor [89], in which the one on the left is a RGB camera that collects 1920 x 1080 color images at up to 30 frames per second. The depth sensor is made up of one IR camera and one IR projector, which are both placed in the middle to detect the relative position of

the subjects in front of the Kinect sensor. The IR projector is an infrared emitter, which emits an infrared signal to the surroundings, while the IR camera is an infrared receiver, which obtains the depth information by returning time from reflected infrared rays to estimate the surrounding scene depth information. During imaging, the data stream returned by the depth sensor encodes depth information in pixels organized in a 512 x 424 image map. By interpolation, the Kinect sensor provides near-synchronized color and depth images with the same resolution and aligns color with depth images using a simple calibration process. Table 3.1 gives the parameter information of the Kinect sensor in detail, which shows the Kinect sensor has a detection range of 0.5 - 4.5m while the depth error is only 0.002 - 0.004m, making it well suited for detecting a newborn in the incubator in natural scenes [90].

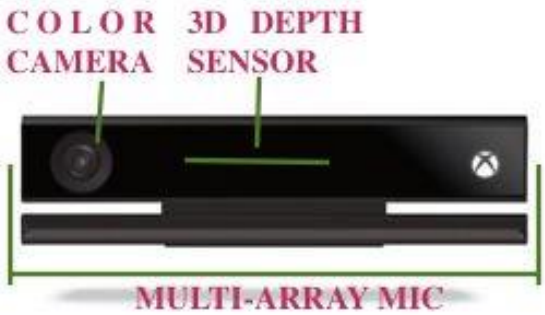


Figure 3.2 Overall hardware specifications of Kinect Xbox One sensor [89].

Table 3.1 Internal characteristics of Kinect Xbox One sensor [89].

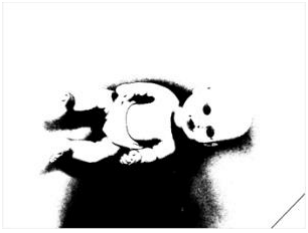
Resolution of Color image	1920 x 1080
Fps of Color image	30 fps
Resolution of Depth image	512 x 424
Fps of Depth image	30 fps
Range of detection	0.5 - 4.5m
Average depth accuracy error	0.002 - 0.004m
Horizontal angle range	70°
Vertical angle range	60°

The color camera of the Kinect sensor is used to capture RGB images, as seen as Figure 3.3(a), which contains color information for human body contour detection based on skin color segmentation. The Kinect sensor driver can provide the synchronous RGB image stream and depth image stream, which helps users to acquire the primary image data.

The Kinect sensor is unique to ordinary cameras in that it uses time-of-flight (TOF) technology to obtain the depth of field data to generate a depth image. Each pixel value in a depth image represents the distance between the Kinect sensor and the corresponding area of a pixel on the subject facing the camera. When taking pictures with the Kinect sensor, the requirements for the environment around the object to be measured are very low, which means imaging will not be affected by conditions such as light, shadows, object occlusion, and so on. The reason why the Kinect sensor has such a unique advantage is that the pixel gray value of the depth image acquired by the Kinect sensor is only related to the distance. In other words, the image information collected by the sensor is only the distance to the measured object, which is independent of other external factors. That makes the Kinect sensor robust enough to provide valid RDB-D data.

Open Natural Interaction (OpenNI) is a multi-language, cross-platform framework that defines Application Programming Interfaces (APIs) for writing applications and leveraging their natural interactions [91]. The OpenNI 2.0 APIs provide a compatible depth sensor interface for accessing PrimeSense [92], which allows an application to initialize the sensor and to receive depth, color and infrared video streams from the device. The OpenNI2 class provides a static entry to the APIs, which is OpenNI2::initialize() [92]. This method initializes all sensor drivers and scans all available sensor devices in the system. All applications using OpenNI2 should call this method before using other APIs. Once the initialization method is completed, it is possible to use Device objects to interact with the real sensor hardware. The OpenNI2::enumerateDevices() method returns a list of available sensor devices that are already connected to the system [92]. Based on that, the Kinect sensor can be initialized, and the color and depth image frames can be acquired as shown in Figure 3.3.

(a)



(b)

Figure 3.3 Acquisition of color and depth images by OpenNI2: (a) original color image (1920 x 1080); (b) original depth image (512 x 424).

3.1.2. FLIR A320 IR camera

The FLIR A320 thermal imaging camera is designed to provide accurate thermal images and reproducible temperature measurements for a wide range of automation applications. It can establish a local area network with a personal computer through a 100 Mbps Ethernet network and perform operations, including camera control, configuration, detection, and data collection via network protocols. The A320 also provides images at 320 x 240 resolution, rendering high-precision temperature hotspot measurements of small objects over long distances to improve data quality. Figure 3.4 displays the appearance of the FLIR A320 IR camera, and Table 3.2 provides a summary of the camera's key features [93]. Since it can detect small temperature differences, it is highly effective for medical applications.

Figure 3.4 FLIR A320 IR camera.

Table 3.2 Summary of FLIR A320 IR camera's key features [93].

IR resolution	320 × 240 pixels
Field of view (FOV)	25° × 18.8°
Minimum focus distance	0.4 m (1.31 ft.)
Focal length	18 mm (0.7 in.)
Spatial resolution (IFOV)	1.36 mrad
Image frequency	30 Hz
Spectral range	7.5–13 µm
Object temperature range	–20 to +120°C (–4 to +248°F) 0 to +350°C (+32 to +662°F)
Accuracy	±2°C (±3.6°F) or ±2% of reading

The Atlas SDK for .NET developed by the FLIR Company is applied to the system to enable developers to get full control of the IR camera and create applications as needed. Figure 3.5 gives a sample of IR image obtained by the A320 camera on a baby.

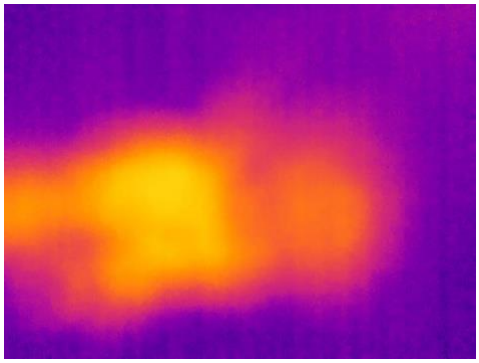


Figure 3.5 An example of IR image on a baby.

3.1.3. Multispectral acquisition platform design

In response to the need for extracting the abdominal area from the original RGB-D image and processing the corresponding thermal distribution from the IR image, the purpose of the study is to design a novel custom acquisition platform consisting of both a Kinect sensor and an IR camera to take three sets of images (color, depth, and IR images) of the newborns in a clinical setting. In the neonatal intensive care units (NICU) of the Children's Hospital of Eastern Ontario (CHEO) and the Ottawa Hospital General campus (TOH-GC), where we recruited the eligible infants, the newborns usually lie in transparent closed incubators for vital signs monitoring and disease prevention and treatment. The incubator box is made of plexiglass and closed all around the baby, leaving only 1-2 access ports for the convenience of the medical staff to take care of the baby. Figure 3.6(a) shows one example of the incubators currently used in the CHEO and TOH-GC NICUs.

During the imaging process, the Kinect sensor and IR camera are designed to take images from above the incubators to get the complete and flat body images of the babies. Although IR rays can pass through glass based on common sense, in actual operation the plexiglass of the incubator has a non-negligible effect on the IR camera's acquisition of infrared rays, including affecting the incident angle of infrared light, increasing the error of detecting human temperature, and detrimental effects on IR camera imaging. Thus, it is required that there is no obstacle between the cameras and the babies, which means the incubators should be open during the imaging process. According to different types of incubators, there are two main ways to open them: one can only be

opened from the side, as shown in Figure 3.6(b); the other is to open the top directly, seen in Figure 3.6(c). In the former case, the baby and the bed cushion can be moved slightly closer to the outside, so the cameras can take pictures obliquely from above the plexiglass cover; in the latter case, since the top cover is removed directly, the cameras can collect images directly from above.

(a) (b) (c)

Figure 3.6 An example of incubator and two different opening ways: (a) closed incubator; (b) incubator opening from the side; and (c) incubator opening from the top.

In order to satisfy the imaging demands, the requirements of this design are listed below:

1) Both imaging devices are mounted on the same stand, and the lenses point in the same direction and are parallel to each other;

2) One device does not appear in the field of view of the other device；

3) The two devices do not interfere with each other during the imaging process;

4) The devices can be tilted slightly to cater to the side-opening incubators;

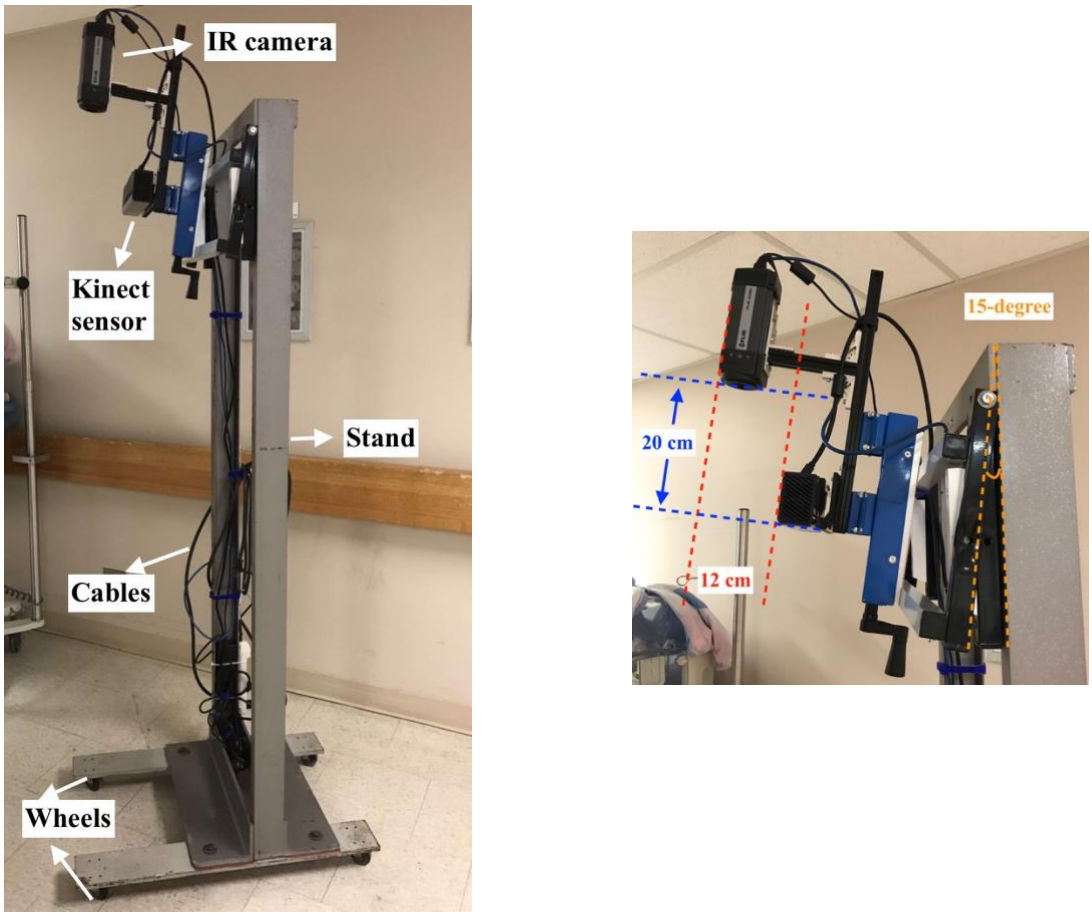
5) The height of the devices relative to the babies should be variable to cater to different types of incubators and cribs;

6) The stand is sturdy, durable and movable, suitable for easy movement between hospital beds.

In order to meet those criteria, a steel rack was used as a stand for the Kinect sensor and the IR camera, which were fixed on the top, facing down towards the incubators, as shown in Figure 3.7(a). Since the image captured by Kinect has a broader viewing angle and a larger field of view than that of the IR camera, the Kinect sensor must be positioned closer to the subject

compared with the IR camera. A distance of 20 cm between the sensors was empirically found to be appropriate. Meanwhile, in order to prevent the Kinect sensor occluding parts of the subject in the IR images, the Kinect sensor was assembled 12 cm aside from the IR camera, as shown in Figure 3.7(b). Moreover, the imaging planes of the two devices need to remain parallel without tilt, and the optical axes of the two devices should be parallel. Also, since the Kinect depth sensor emits a signal in a spectral range of 827 - 850 nm while the IR camera's infrared light is sensitive to a spectrum range from 7.5 μm to 13 μm. As a result, the FLIR infrared sensor operation is not interfered by the Kinect depth sensor. Furthermore, taking into account of some side-opening incubators, the cameras are designed to have a 15-degree inclination to facilitate the full-body imaging of the baby, as seen in Figure 3.7(b). There is also a translation plate that attaches the Kinect sensor and IR camera to the stand, which is convenient for controlling the two devices to adjust the height to adapt to different types of incubators, as shown in blue in Figure 3.7(b).

There are also four wheels mounted under the steel rack so that the whole acquisition platform can be moved around. All cables are included on the side of the rack to prevent interference with the imaging process.



(a) (b)

Figure 3.7 Assembly of Kinect sensor and IR camera in a custom multispectral acquisition platform: (a) the assembly of the full setup; (b) detailed illustration of IR camera and Kinect sensor.

In order to make the image acquisition process more user friendly, a companion multispectral acquisition software was developed based on Python and Qt that supports the real-time acquisition of image data captured for the study. As shown in Figure 3.8, it provides an intuitive graphic interface for data entry and image display, which makes the acquisition system more convenient to use. It allows medical staff to input the subject study identification number and subject group (NEC or Normal) to distinguish samples in the database. Then by clicking the "CONNECT" button, both the Kinect sensor and IR camera are activated and start streaming at the same time. This step also sets the transmission conditions of the two devices, including synchronizing the camera frequency of the two devices to ensure the synchronization of image

acquisition (30 fps). The sequence numbers of the images collected by both sensors are integers starting from 0. And the images are mirrored to facilitate the observation. Then, a depth image, RGB image, and IR image are transmitted and shown on the screen, respectively. The “RECORD” button is used to record each subsequent frame of all three image streams (depth, RGB, and IR) when the subject is ready to avoid collecting too many invalid images, which would make the database redundant. Finally, clicking the “EXIT” button terminates the imaging process, and the system will shut down.

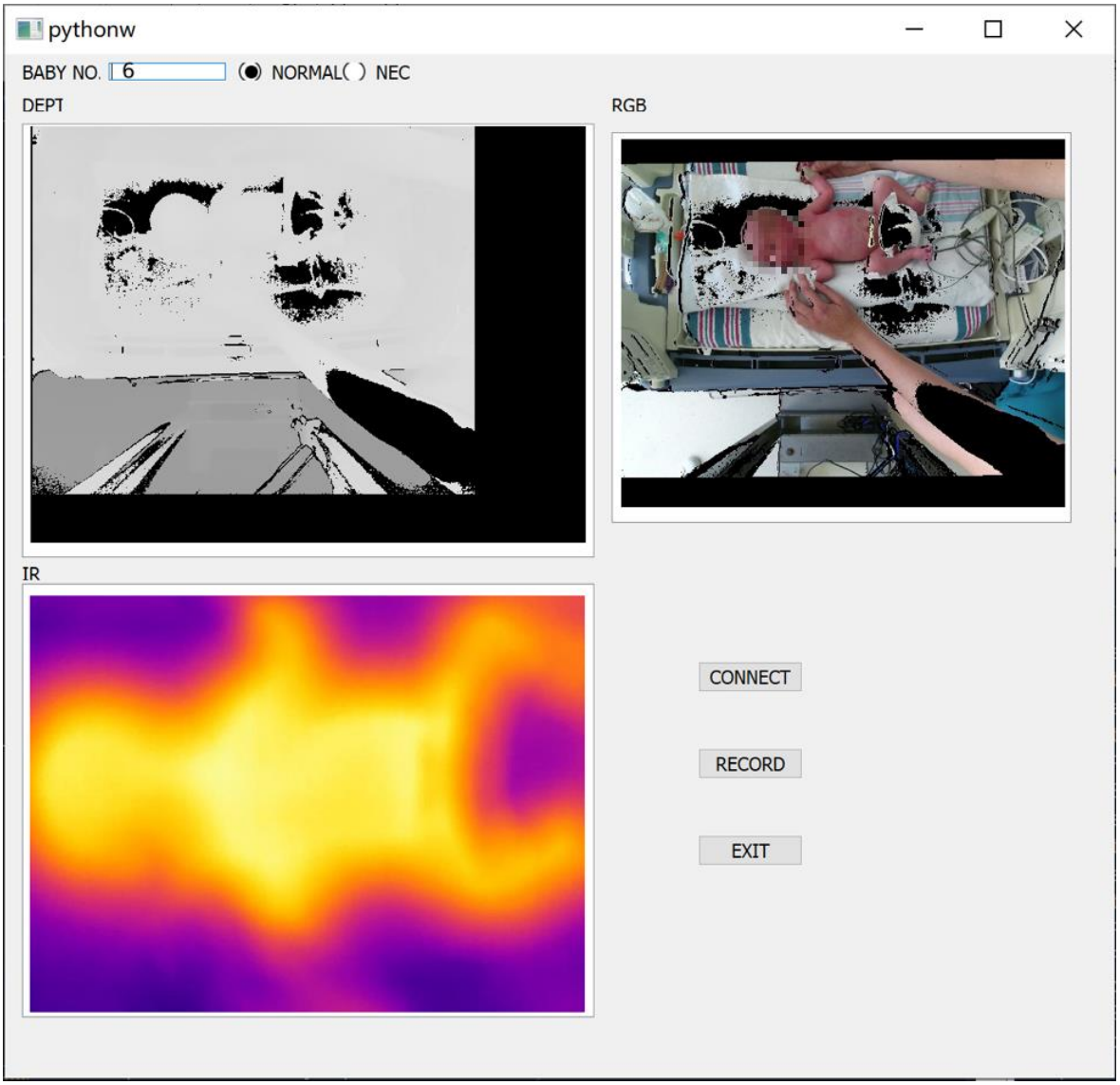


Figure 3.8 Graphic interface of the automated multispectral acquisition system.

3.2 Data Processing and Multi-camera Calibration

The proposed custom multispectral acquisition platform consists of one Kinect sensor and one IR camera, but the exact positions of the cameras inside these two devices are unknown, so the main purpose of multi-camera calibration is to derive the relative position relationship between the two cameras. Furthermore, since the analysis relies on comparisons between the distribution of thermal data among different groups, the study does not involve thermal calibration procedures to determine the exact correspondence between the grayscale levels encoded in IR images and the actual temperature in Celsius. Thus, the temperature measurement calibration is not required and not performed in this thesis.

The most common method of camera calibration is Zhang's method [43], which is leveraged in this research. That is, first the two cameras to be calibrated are used to take a plurality of checkerboard pictures at different viewing angles, from which the internal and external parameters of each camera are estimated. Then the rotation and translation matrices of the two cameras are calculated relative to the same coordinates system based on the same sets of calibration images. During the calibration process, there are the following requirements:

1) The calibration checkerboard should be visible both to Kinect sensor and IR camera;

2) The Kinect sensor and IR camera should start and end imaging at the same time;

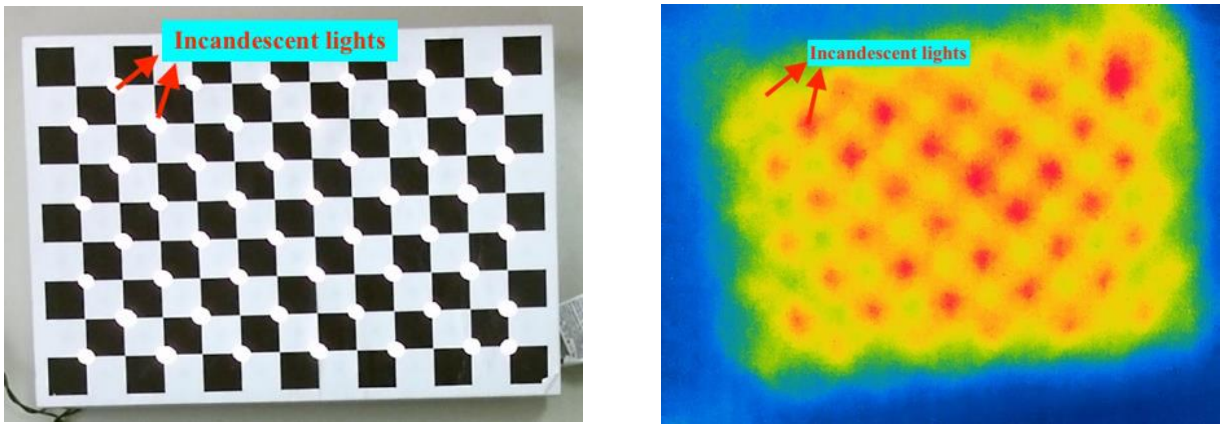
3) The Kinect sensor and IR camera should be adjusted to the same frequency (30fps) to ensure that the calibration photos are taken at the same time and that sequence numbers correspond one to one;

4) The calibration images with an angle of 15-degree between the plane of the calibration board and the image plane should be included to simulate the actual situation of the side of the incubator being opened;

5) Take enough pictures for calibration to ensure accuracy;

For camera calibration, calibration checkerboards with known shapes and sizes are widely used [43, 44]. However, because the IR camera can only sense the temperature information and cannot directly obtain the texture of the calibration board like the RGB camera, it is impossible to use the regular checkerboard. To address this constraint, a customized checkerboard was designed for feature detection, as seen in Figure 3.9(a). Embedded incandescent light bulbs are integrated

at intersection points in the grid. They dissipate heat and form hot spots to be detected by the IR camera. Meanwhile, the checkerboard consists of a 13 x 9 black and white grid made of square areas measuring 28 x 28 mm, and where two adjacent regions are different in color, so that the Kinect sensor can directly detect the black and white checkerboard patterns. Moreover, in order to avoid mutual interference between the warmer spots, only 50% of the intersections in the grid receive an incandescent bulb (49 bulbs in total). Furthermore, in order to distinguish the directivity and the accuracy of calibration, the numbers of rows and columns of the bulb array are not equal. The three light bulbs in the lower right corner are closely connected to each other compared to the others to help orient the calibration board. In this way, feature points can be detected synchronously by the Kinect color sensor and the IR camera to support matching between the RGB-D and the IR images, thereby achieving calibration between the multispectral sensors.



(a) (b)

Figure 3.9 Custom multispectral calibration checkerboard made up of 13 x 9 black and white squares with 49 embedded incandescent light bulbs at intersection points: a) color image of calibration board under normal ambient lighting conditions, and b) IR image of calibration board with incandescent bulbs hot spots.

With the help of the custom checkerboard, the multi-camera calibration process is completed in the following steps:

1) Preprocessing the original images collected by the Kinect sensor, and aligning the depth image and the RGB image through OpenNI2 to generate a RGB-D image (section 3.2.1.1);

2) Calibrate the Kinect sensor based on the RGB-D image to obtain the internal parameters of the color camera after being aligned with the depth camera (section 3.2.1.2);

3) Pre-process the original IR image (3.2.2.1);

4) Calibrate the IR camera based on the IR feature map to obtain the internal parameters of the IR camera (section 3.2.2.2);

5) Stereo calibration to register the IR camera with the RGB-D sensor (section 3.2.3).

3.2.1. Kinect sensor calibration based on RGB-D image

3.2.1.1 Preprocessing of original images and alignment to form RGB-D images

Unlike color images with a resolution of up to 1920 x 1080, depth images contain a lot of noise, especially at the edges of objects, as seen in Figure 3.10(a). This is because the measurement principle of the depth camera is TOF, which easily forms noise proportional to the square of the distance [94]. In view of this kind of noise characteristics, a denoising method based on a pass-through filter is used to remove the data of the edge from the field of view, as it offers poor quality.

Suppose the depth map is represented as (u, v, d), where u, v represents the pixel coordinates of the depth map, d represents the depth, the coordinates of the center point of the depth map is (u_0, v_0), and the width of depth map is w, height is h, then we can select the depth value around the center of the depth image by equation (3.1) based on a pass-through filter [95]:

$$\begin{cases} u_0 - (w/2) \times r \le u \le u_0 + (w/2) \times r \\ v_0 - (h/2) \times r \le v \le v_0 + (h/2) \times r \end{cases} \tag{3.1}$$

Where r represents the ratio of retaining the depth map, r = 1 means the entire depth value is retained. Assuming max represents the maximum allowed depth and min represents the minimum, then a valid range of depth values can be selected through min < d < max.

Take each depth value (u_i, v_i, d_i) in the depth map as the center of the window, select the window area with a window size of 3 × 3, and calculate the standard deviation, σi, over the depth, d, distribution in this window area. Let n = 3 × 3.

$$\sigma_i = \frac{1}{n-1} \sqrt{\sum_{j=0}^{n} (d_{ij} - u)^2} \tag{3.2}$$

Where d_{ij} represents the depth value of adjacent elements in the window centered on d_i. μ is the average value of the depth values in the window. All standard deviations are combined into a standard deviation set $s = \{\sigma_1, \sigma_2, ..., \sigma_k\}$. Thus, each depth value corresponds to a σ_i. Calculate the mean, μ_s, and standard deviation, σ_s, of this standard deviation set, as shown in the equations (3.3) and (3.4).

$$\mu_s = \frac{1}{n}\sum_{i=0}^{k}\sigma_i \tag{3.3}$$

$$\sigma_s = \frac{1}{n-1}\sqrt{\sum_{i=0}^{k}(\sigma_i - \mu_s)^2} \tag{3.4}$$

$$\mu - t \times \sigma < \sigma < \mu + t \times \sigma \tag{3.5}$$

Keep the corresponding depth value that satisfies the variance of formula (3.5), where t is a threshold defined as a floating-point number. Finally, the denoised depth image is shown in Figure 3.10(b). It is clearly seen that in the denoised depth image, the edges of the objects are smoother and more complete, and the outlines of the persons are clearer while the depth details on the edges of the objects and the junction areas are well retained, compared to the original depth image.

(a) (b)

Figure 3.10 Denoising of depth images: (a) original depth image; (b) depth image after denoising.

In order to register the color image and depth images to obtain a RGB-D image, a point cloud is often used to perform the back-projection from the depth image space to the camera space. However, if back-projection is performed point by point, the cost of space and time is too large, which affects the real-time performance of the imaging process. Therefore, the Coordinate mapping class in OpenNI2 is used to help align the depth and RGB images. The Coordinate mapping class mainly accomplishes two functions: 1) project the coordinates in the 3D camera coordinates space to the 2D depth map, or back-project from the depth map to the camera coordinates space; 2) find the correspondence of pixel positions between the depth map and the color map. With these two functions, the depth map is mapped to the color map, that is, for a pixel on the depth map, a corresponding pixel on the color map is found. The specific steps are as follows:

1) Obtain a depth frame and save the depth data to an array;

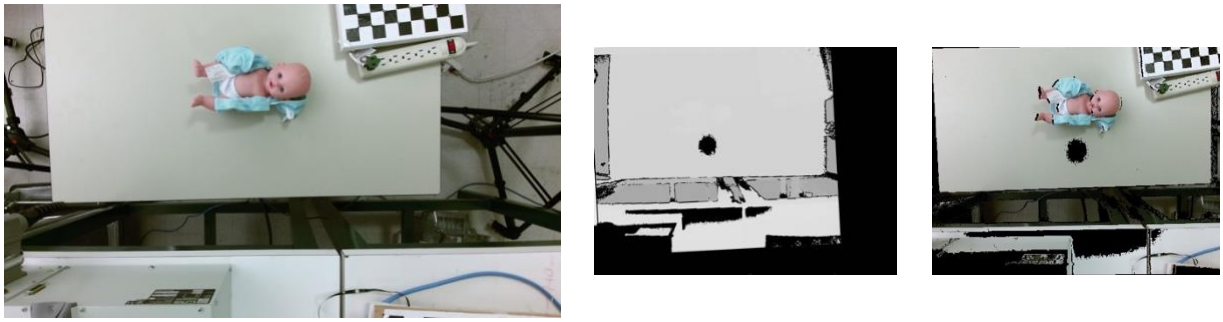
2) Get the coordinate mapper;

3) Traversing the pixels of the color image.

If the mapping is correct, that is, the coordinates of the point mapped to the depth image are valid in the depth map, the color pixel value of the point is assigned to the corresponding pixel in the depth array and displayed.

Therefore, the alignment result of the color map and the depth map is actually the appearance of a broken color map, where the pixels displayed indicate where the alignment is

successful, and the black pixels indicate locations where the alignment failed. As shown in Figure 3.11(c), the incomplete areas in the RGB-D image mainly come from the incompleteness of the depth map (such as the black hole in the center of the table). In addition, all the color information in the range of the depth map is basically retained in the RGB-D image. Hence, the RGB-D image can be drawn with the same size as the depth image, as it covers a narrower field of view than the color image. Figure 3.11 shows an example of alignment between an original color image and a depth image to form an RGB-D image.



(a) (b) (c)

Figure 3.11 Acquisition of raw RGB-D image: (a) original color image (1920 x 1080); (b) depth image (512 x 424); and (c) aligned RGB-D image (512 x 424).

3.2.1.2 Color camera calibration based on RGB-D image

The calibration procedure for the RGB camera calibration can be described as follows:

1) Acquire a number of checkerboard images by pointing to the calibration target from various angles, as shown in Figure 3.12. It is known from Zhang's calibration method [43] that at least three images are required to make the internal parameter formula have a unique solution. It is empirically known that the number of images is usually about 10 for a better calibration result. Conversely, the impact of acquiring too many images on accuracy is not as great as expected [96]. The most important aspect to pay attention to in this process is the relative position of the calibration checkboard and the camera in the image. Generally, the calibration board must be facing up, looking down, left and right, and so on.

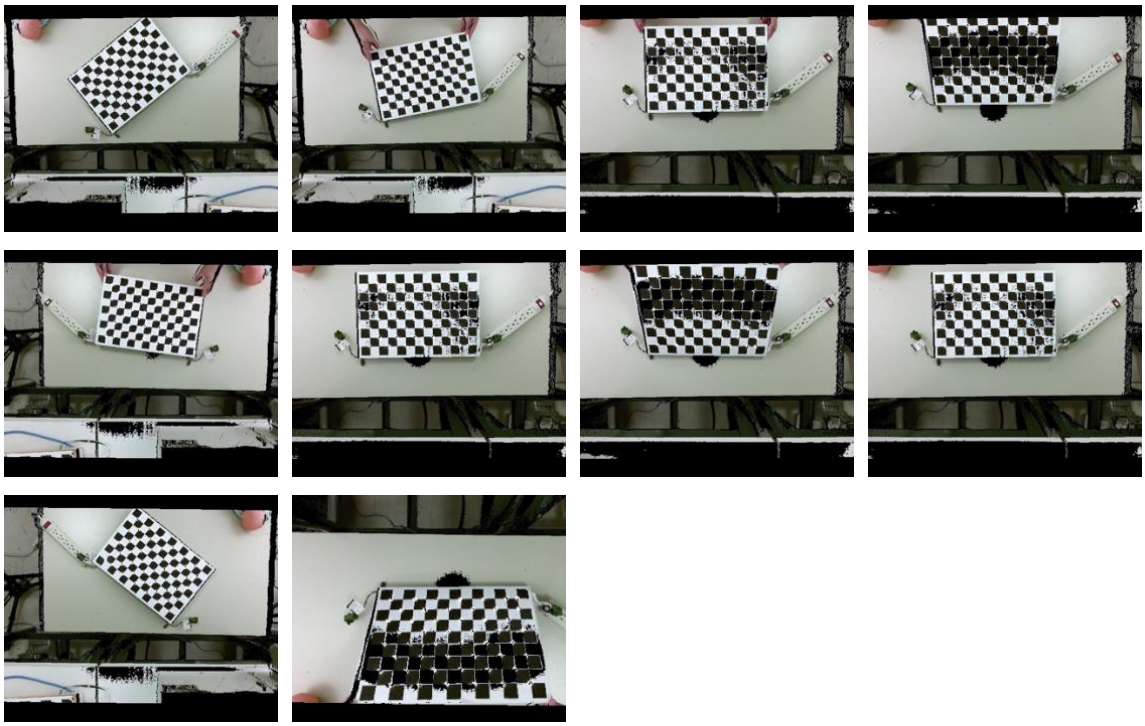


Figure 3.12 Ten checkerboard RGB calibration images from various angles.

2) Extract grid corners with actual window size, which is a 11 x 7 grid. Considering the corner where three incandescent light bulbs are closely connected as the origin, the short side of the board is in the X direction, and the long side is in the Y direction, according to the number of corner points in each row and column of the board, the coordinates of the corner points can be obtained. As can be seen from Figure 3.13, red crosses represent the extracted feature corners.

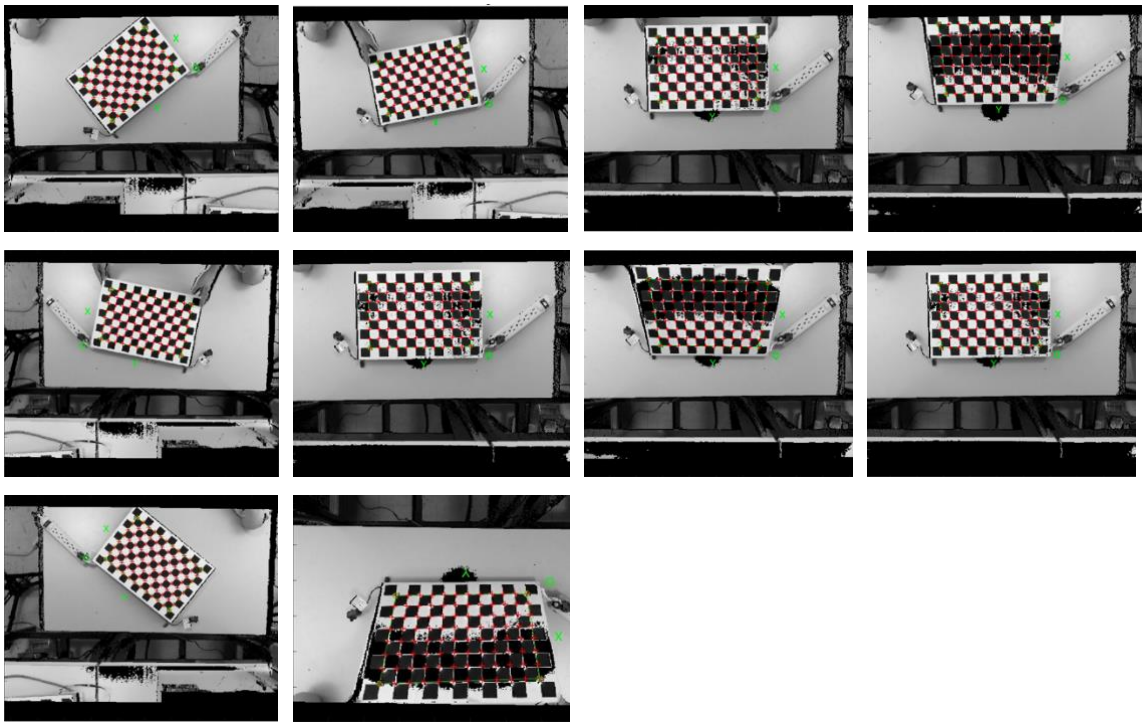


Figure 3.13 Extracted grid corners

3) Estimate calibration parameters after initialization using the OpenCV:calibrateCamera() function and then calculate intrinsic parameters (including distortion coefficients) of RGB-D camera according to Zhang's method [42] (Table 3.3). Table 3.4 shows intrinsic parameters of both the color and depth cameras from Burrus's work [97], where the calibration results based on RGB-D images are closer to that of the depth camera because the RGB-D imaging principle fills color information into the corresponding position of the depth image, which makes the calibration feature points read in the RGB-D image actually in the position of the depth image. The calibration parameters based on the RGB-D images can be seen as that on the visible depth images, and the comparison of the calibration results of the Kinect sensor's depth camera shows that the intrinsic parameters are basically consistent and reasonable.

Table 3.3 Intrinsic calibration results of RGB-D camera after optimization.

Focal Length	fc = [349.81558 350.99530] +/- [37.53429 38.18605]
Principal point	cc = [244.21208 201.39782] +/- [8.57758 23.25039]
Skew	alpha_c = [0.00000] +/- [0.00000] => angle of pixel axes = 90.00000 +/- 0.00000 degrees
Distortion	kc = [0.07345 -0.04673 0.01521 -0.00662 0.00000] +/- [0.08630 0.18522 0.01595 0.00691 0.00000]
Pixel error	err = [0.84874 0.82762]

Table 3.4 Intrinsic calibration parameters of Kinect sensor reported in Burrus's work [97]

	Color camera	Depth camera
Focal Length	fc = [1094.03583 1087.37528] +/- [55.02816 51.42175]	fc = [379.40726 378.54472] +/- [40.73354 34.75290]
Principal point	cc = [942.00992 530.35240] +/- [13.00131 31.27892]	cc = [263.73696 201.72450] +/- [9.17740 30.29723]
Skew	alpha_c = [0.00000] +/- [0.00000] => angle of pixel axes = 90.00000 +/- 0.00000 degrees	alpha_c = [0.00000] +/- [0.00000] => angle of pixel axes = 90.00000 +/- 0.00000 degrees
Distortion	kc = [0.06857 -0.10542 0.00233 0.00092 0.00000] +/- [0.02206 0.02884 0.00379 0.00492 0.00000]	kc = [0.03377 -0.04195 0.00519 0.00734 0.00000] +/- [0.07368 0.25678 0.01111 0.00965 0.00000]
Pixel error	err = **[** 0.49343 0.67737 **]**	err = **[** 0.88997 0.92779 **]**

3.2.2. IR camera calibration

3.2.2.1 Pre-processing of IR feature images

A typical IR image of the individual calibration target is shown in Figure 3.14(a), in which the areas in red indicate the positions of higher temperature incandescent bulbs, while the regions in yellow are the surface areas radiated by the incandescent bulbs, and the areas in blue correspond

to the background where the temperature is lower. The purpose of preprocessing procedures is to identify accurate feature points (incandescent light bulbs) without interference from the board surface that is simultaneously heated. In order to input valid feature points to the calibration procedure, pre-processing of the raw image ought to be carried-out:

1) Apply a gray-level threshold within the grayscale representation of the original IR image, as shown in Figure 3.14(b);

2) Erode the highlighted areas of the image to eliminate noise, segment out individual image elements and connect adjacent elements in the image (Figure 3.14(c));

3) Detect lines of the calibration checkerboard using Hough line transform after applying Canny edge detection [98]. It can be seen in Figure 3.14(e) that estimated lines are drawn with a blue or red color that represents two different perpendicular sets connecting intersection points on the calibration checkerboard that correspond to hot spots formed by the embedded incandescent bulbs;

4) For each line of the red group, traverse its intersection with each line of the blue group and record all the points of intersection within the image range; Figure 3.14(f) shows a total of 49 points corresponding to 49 incandescent lights.

5) Given that incandescent lights are embedded only at every second corner of the black-and-white checkboard pattern, and with the aim of achieving one-to-one correspondence between feature points in the RGB-D image and those in the IR image, the remaining feature points are supplemented by finding the midpoint of any two adjacent points (Figure 3.14(g)).

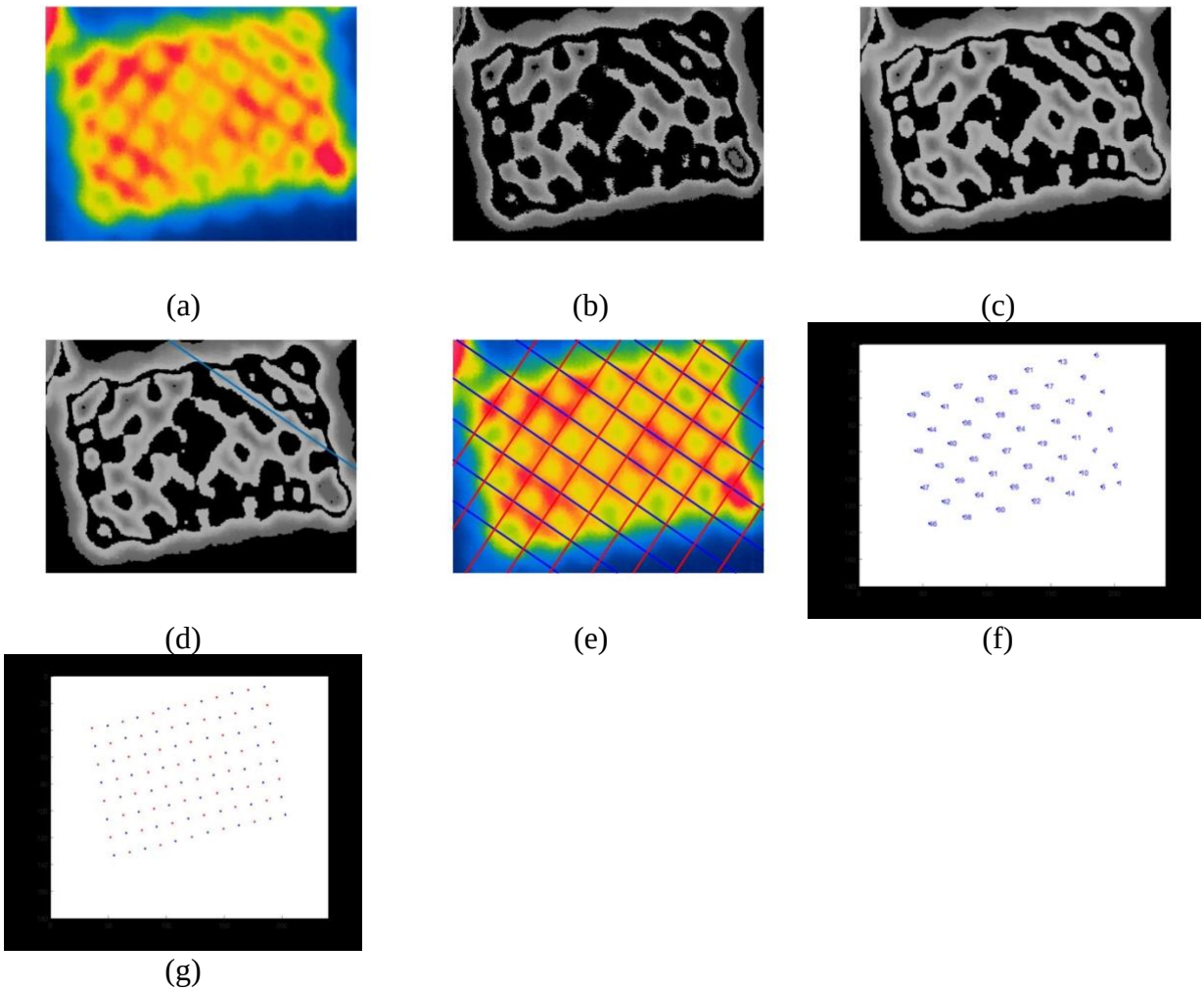


Figure 3.14 Pre-processing procedures of IR images for calibration with RGB-D sensor: (a) original IR image of the individual calibration target; (b) after applying a gray-level threshold; (c) erosion of the highlighted areas; (d) draw one estimated line; (e) draw all the estimated lines in blue or red; (f) recognition of 49 feature points; (g) supplement feature points to match all the RGB-D feature points.

3.2.2.2 IR camera calibration with RGB-D sensor

The procedure above converts the IR image into a registration map that can be detected by the calibration tool. Thereafter, a set of input images corresponding to RGB-D images is processed the same way as shown in section 3.2.1.2 for the purpose of IR camera calibration, as depicted in Figure 3.15. The results obtained for the initial calibration parameters are summarized in Table 3.5. It can be seen that the focal length from IR images is about 690 pixels. Since the cell size of

each pixel is about 26mm [99], the focal length of the FLIR IR camera is calculated as in equation (3.6):

$$\text{Focal length} = f_c \times \text{cell size} = 0.026 \text{ mm/pixel} \times 690 \text{ pixels} = 17.94 \text{ mm} \quad (3.6)$$

According to the parameters given in Table 3.2, where the focal length of the IR camera provided by the manufacturer is 18 mm, it indicates that the calibration results obtained from processed IR feature maps are reasonable.

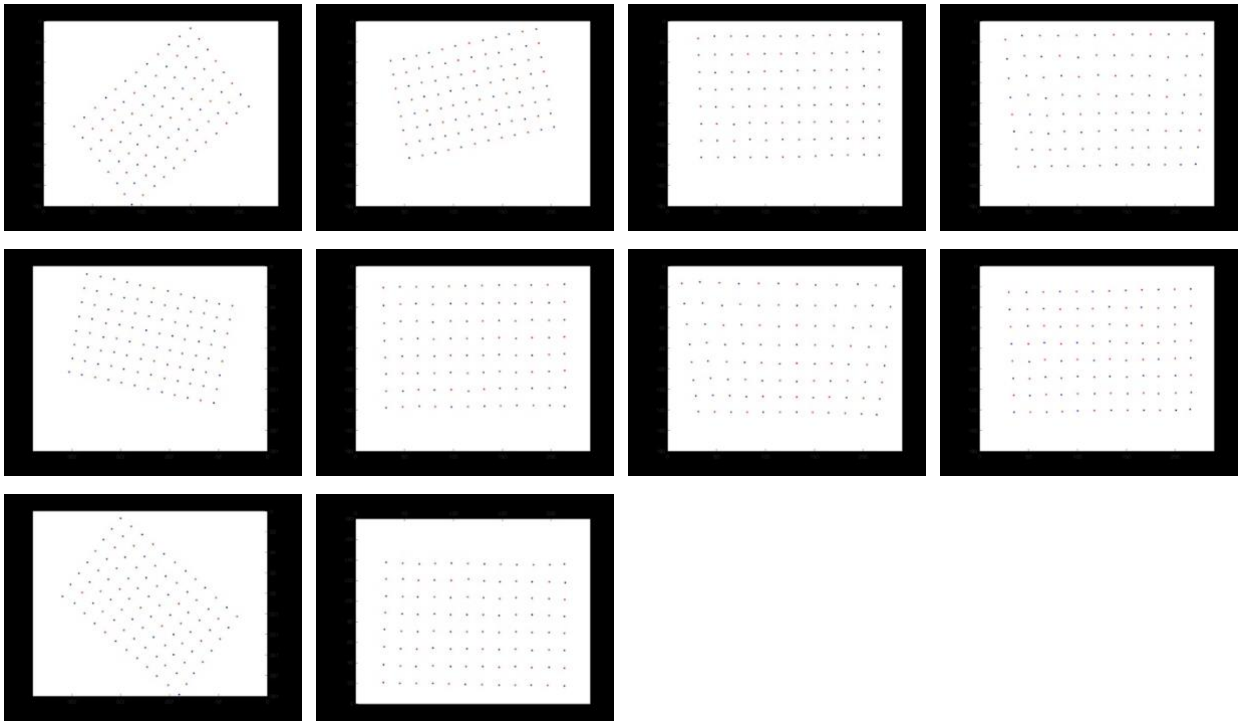


Figure 3.15 Ten IR calibration images for IR camera calibration.

Table 3.5 Intrinsic calibration results of IR camera after optimization.

Focal Length	fc = [689.96207 783.84902] +/- [18.46342 20.98384]
Principal point	cc = [281.66859 424.07136] +/- [0.00000 0.00000]
Skew	alpha_c = [0.00000] +/- [0.00000] => angle of pixel axes = 90.00000 +/- 0.00000 degrees
Distortion	kc = [-0.06006 0.04283 -0.01687 -0.00447 0.00000] +/- [0.06991 0.11585 0.01462 0.00178 0.00000]
Pixel error	err = [0.85692 1.04921]

3.2.3. Stereo calibration

Stereo calibration is done to calculate the relative pose between the two camera coordinates systems, that is, on the premise of the obtained intrinsic parameter matrix for each camera, solve the relative relationship between the RGB-D and IR cameras by calculating IR camera's translation vector, T, and rotation matrix, R, relative to the RGB-D camera.

Figure 3. 16 shows the coordinates system setting of the relative position of IR camera and Kinect sensor. With the IR camera as the center, the direction away from the stand horizontally is the positive X-axis direction, the direction parallel to the body of the Kinect sensor and facing away in the image is the positive Y-axis direction, and the vertical downward direction is the positive Z-axis direction.

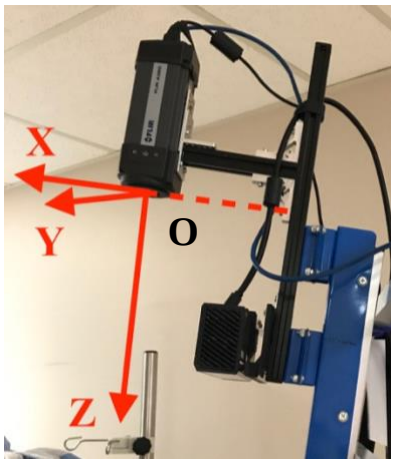


Figure 3.16 The coordinates system settings for the relative positions of the IR camera and the Kinect sensor.

Set P_{rgbd} as the 3D spatial coordinates of a point in the RGB-D camera, p_{rgbd} as the projection coordinates of the point on the depth plane (the units of (x, y) are pixels, and z is equal to the depth value whose unit is mm), and H_{rgbd} as the intrinsic parameter matrix of the RGB-D camera (Table 3.3). From the pinhole camera model, the following relationships are established:

$$p_{rgbd} = H_{rgbd} P_{rgbd} \tag{3.7}$$

$$P_{rgbd} = H_{rgbd}^{-1} p_{rgbd} \tag{3.8}$$

Also, let's consider P_{ir} as the 3D spatial coordinates of the same point in the IR camera, p_{ir} as the projection coordinates of the point on the IR image plane, and H_{ir} as the intrinsic parameter matrix of the camera (Table 3.5). Since the coordinates of the RGB-D camera and the

coordinates of the IR camera are different, they can be linked by a geometrical transformation, namely:

$$P_{ir} = RP_{rgbd} + T \tag{3.9}$$

where R is the rotation matrix, and T is the translation vector. Finally, use H_{ir} to project the 3D point P_{ir} to the IR image plane and estimate its IR image coordinates:

$$p_{ir} = H_{ir}P_{ir} \tag{3.10}$$

It should be noted that both p_{rgbd} and p_{ir} use homogeneous coordinates, so when constructing p_{rgbd}, the original pixel coordinates (x, y) should be multiplied by the depth value, and the final pixel coordinates must be divided by p_{rgbd}, i.e. (x/z, y/z). The value of z is the distance (in millimeters) from the point to the RGB-D camera.

Moreover, to determine the rotation matrix and the translation vector of the two coordinate systems, the external parameters of the two sensors are used. The external parameter matrix is composed of a rotation matrix $R_{rgbd}(R_{ir})$ and a translation vector $T_{rgbd}(T_{ir})$, which means that the point P in a global coordinate system is transformed into the camera coordinate systems of the depth camera and the IR camera respectively. If it is satisfied, then:

$$P_{rgbd} = R_{rgbd}P + T_{rgbd} \tag{3.11}$$

$$P_{ir} = R_{ir}P + T_{ir} \tag{3.12}$$

Then (3.11) and (3.12) lead to:

$$P_{ir} = R_{ir}R_{rgbd}^{-1}P_{rgbd} + T_{ir} - R_{ir}R_{rgbd}^{-1}T_{rgbd} \tag{3.13}$$

Given that

$$P_{ir} = RP_{rgbd} + T \tag{3.14}$$

Thus,

$$R = R_{ir}R_{rgbd}^{-1} \tag{3.15}$$

$$T = T_{ir} - R_{ir}R_{rgbd}^{-1}T_{rgbd} = T_{ir} - RT_{rgbd} \tag{3.16}$$

Consequently, we only need to obtain the external matrix of the checkerboard relative to the RGB-D camera and IR camera in the same scene. Then we can calculate the transformation matrix that connects the two sensor coordinate systems. Since all rotation matrices are orthogonal arrays, transposition operation instead of inverse operation can be directly applied. Although the external parameter matrices obtained in different acquisition procedures are slightly different, resulting in different calculated R and T, using several calibration images of different views of the

target and then calculating the average value can generate better results according to the actual experimental results. The stereo calibration results are listed in Table 3.6:

Table 3.6 Extrinsic parameters (position Kinect depth sensor of wrt IR camera):

Rotation vector	R = [-0.11806 0.09681 0.00154] ± [0.00729 0.00818 0.00015] rad
Translation vector	T = [-117.23159 14.68501 188.79907] ± [6.67013 2.11307 16.07277] mm

The T matrix describes the relative relationship between the origins of the two device coordinates. According to the coordinates system of pre-set relative positions of the two devices (Figure 3.16) and the actual physical distances between them (Figure 3.7(b)) , the stereo calibration results show the actual x-direction is approximately -120mm, measured as -117 ± 7mm; the z-direction is approximately 200mm, measured as 189 ± 16mm; the Y-direction is measured as 15 ± 2mm. Although the specific position of the depth camera inside the Kinect sensor is unknown, it can be seen that it is near the centerline of the Kinect sensor, and since the centerline of the IR camera and Kinect sensor is the same when installed, so the Y-direction data can be inferred to be accurate. The accuracy of the data in the three directions also basically meets the requirements. The R matrix describes the relative orientation relationship between the two camera coordinates systems. Since the orientation of the experimental camera coordinates systems are almost the same, the results are reasonable.

3.3 Automated Region of Interest Extraction

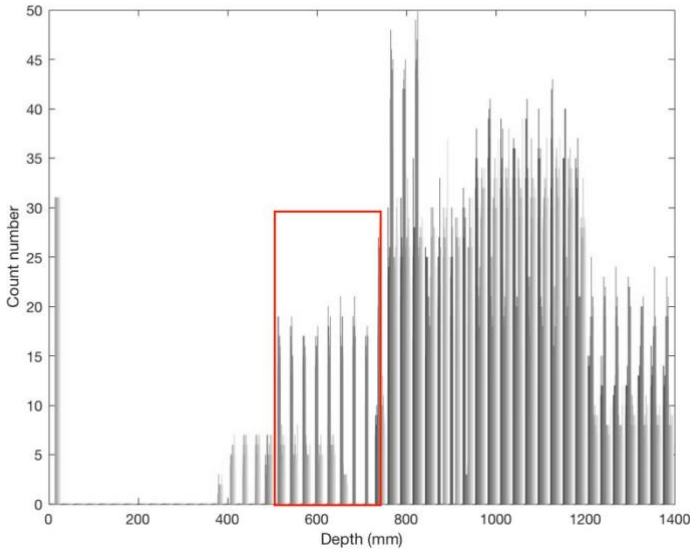
With the custom acquisition system described above, three sets of RGB-D, depth, and IR data are collected to extract a meaningful representation to help detect NEC. In the previous studies [9, 10, 11], the IR images were usually manually segmented directly by research staff based on the usual anatomical landmarks of the abdomen. However, there are certain problems with this dependency, such as being time-consuming, reliance on research staff input to complete the imaging, potentially missing important areas involved in the NEC process, and segmentation inaccuracy due to the subtle variations in temperature of the human body and surrounding area. Thus, it is challenging to directly extract the region of interest in a single IR image. Hence, a

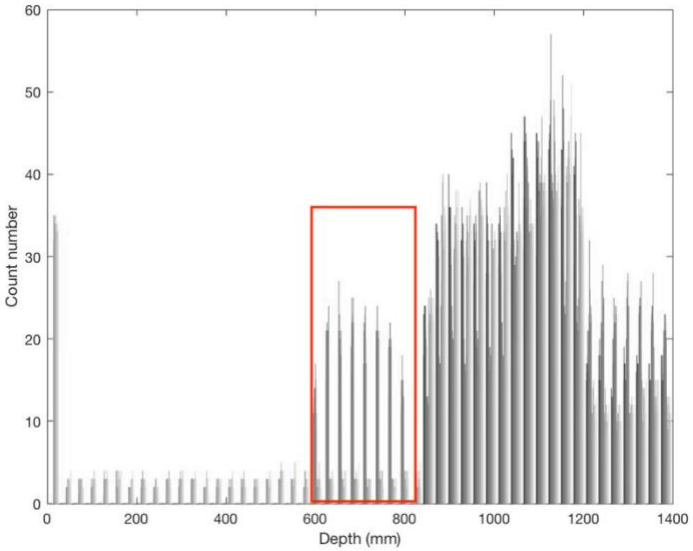
segmentation method for automated region of interest extraction is proposed and detailed in this section, which relies on the extra color and depth information to distinguish body parts.

The segmentation procedure starts from a depth threshold applied of the D channel of the registered images collected over a subject. The objective is first to isolate the body from the background, which is the bedding surface since the acquisition of baby images requires that the cameras face down to take a full body image of the baby, as described in section 3.1.3. Segmentation is then refined using a skin color retrieval process to remove any visible apparatus such as probes, tubes, wires, and clothes. Next, skeleton data is extracted and analyzed to locate the joints that connect different parts of the baby's body. Finally, the region-of-interest area is mapped from the RGB-D image to the IR image to retrieve the temperature distribution only over a specific area, which is meant to reduce bias in the thermal distribution analysis that may result from considering only subparts of the torso, as in previous research [9, 11], or including background items in the considered thermal map. The whole process is illustrated in Figure 3.1.

3.3.1. Background removal and fine-tuning of the detected human body regions

As mentioned earlier, each point of the RGB-D image is expressed by a vector of coordinates (x,y,z), in which z means the distance from the subject to the depth sensor. Figure 3.17(b) shows the corresponding histogram of a depth map in Figure 3.17(a). According to the prior information about the imaging conditions given in section 3.1, the baby's body area is always closer to the camera than the background, as highlighted in the red box in Figure 3.17 (b), while the depth data larger than this area belongs to the background areas (bed, ground) that are farther away from the camera. Also, there is a small part of the data on the left side of the red box that represents some accidental interference items such as the arms and hands of the nurse who holds the baby. However, considering that the proportion of these interference terms is low, and the depth value is significantly smaller, they can be easily excluded. In addition, there are several data with the value of 0 that correspond to pixels where the sensor is unable to collect depth information, as illustrated by black areas in Figure 3.17(a). The latter do not affect the thresholding as long as all relevant data are successfully collected.

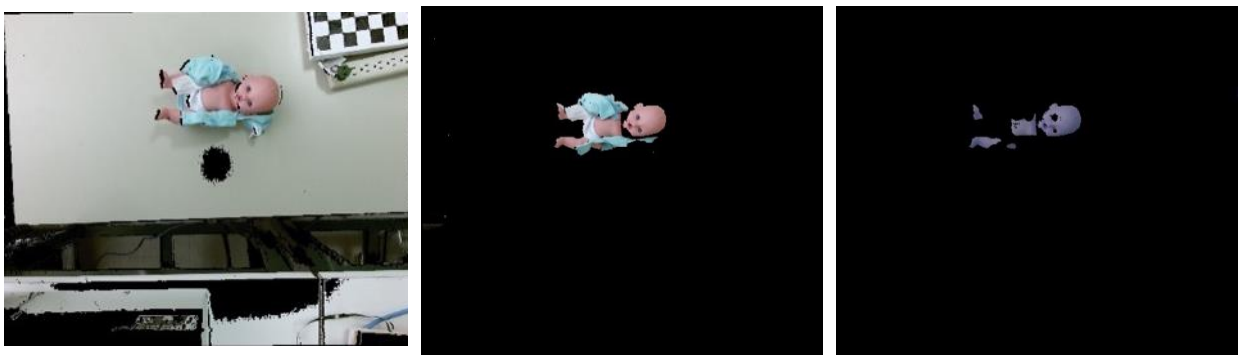




(a) (b)

Figure 3.17 Depth maps and their corresponding histogram: (a) original depth images; (b) histograms of the depth maps, where the red box highlights the threshold range of the target human body.

By applying a distance threshold based on the largest cluster with smallest value, it is easy to remove background information from the original image. The original RGB-D image and background removal image are displayed below in Figure 3.18(a) and (b) for a typical scene created with a baby doll.



(a) (b) (c)

Figure 3.18 Results of background removal and fine-tuning of the human body contour: (a) original RGB-D image; (b) background removal image; (c) after applying skin color detector.

Skin color is one of the most prominent features of the body surface, which is the least sensitive to positions of the body, rotation of the image, and any other changes, so it has advantages in body detection. Three predominant color spaces widely considered in computer vision are RGB, HSV, and YCbCr. By comparing those three spaces, it is well known that other objects or illumination interfere less significantly with the YCbCr and HSV spaces. However, the conversion from RGB to HSV space is more complicated than that from RGB to YCbCr. Therefore, the YCbCr color representation is chosen as the skin color segmentation model [100]. The flow chart of the YCbCr skin color-based filter is given in Figure 3.19.

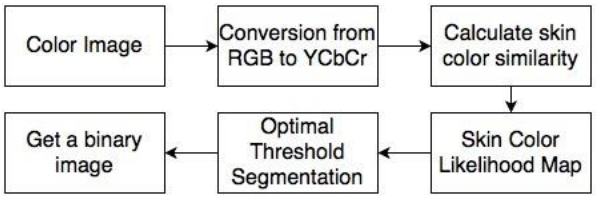


Figure 3.19 Flow chart of a YCbCr skin color-based filter.

This method mainly uses statistical principles, assuming that the random samples of skin color conform to the normal distribution and satisfy a Gaussian distribution [101] . The Gaussian model composes the continuous data information by calculating the probability value of the pixels

and obtains a skin color probability map, leading to confirming the skin color. A two-dimensional Gaussian function can be written as:

$$P(Cr, Cb) = \exp\left[-0.5(x-m)^T C^{-1}(x-m)\right] \quad (3.17)$$

where x is the value of the sample pixel in the YCbCr space, which only emphasizes the chromaticity: Cb, Cr, and not the intensity, Y, resulting in less sensitivity to illumination:

$$x = [Cb,Cr]_T \quad (3.18)$$

M is the mean of samples of the skin color in the YCbCr space:

$$M = E(x) \quad (3.19)$$

and C is the covariance matrix of the skin color similarity model:

$$C = E((x-M)(x-M)T) \quad (3.20)$$

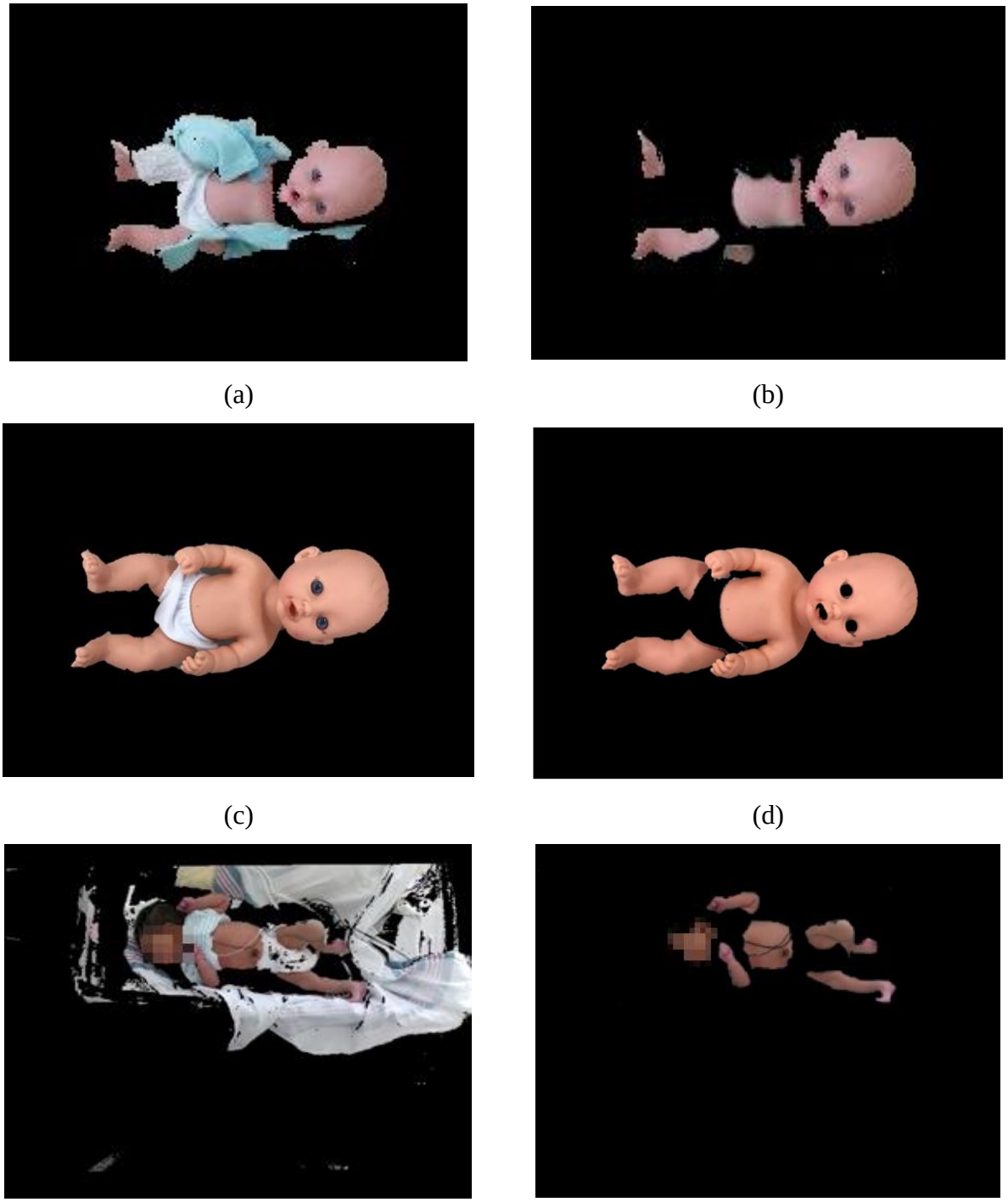
To determine the parameters in the function, a large number of skin color samples need to be collected, on which the statistical characteristics are calculated in order to estimate the values of M and C. Tan et al. [102] offer a human skin detection dataset that contains human skin data randomly downloaded from Google under different illumination conditions and complex backgrounds. Phung et al. [103] created the face and skin detection database made up of over 4000 images of four skin types: 1) whitish, pinkish; 2) yellowish, light brownish; 3) reddish, darkish, dark brownish; 4) other skin types. Both databases were used as the skin color samples to make the skin color filter robust to different illumination, multiple skin colors, and complex backgrounds. M and C are then substituted into (3.17) to find each $P(Cr, Cb)$ value and to normalize by using:

$$Pi(Cb, Cr)/\max\left(Pi(Cb, Cr)\right) \quad (3.21)$$

Considering this quotient as the similarity value of the point, in order to view the similarity image, we can convert the scale [0,1] to [0, 255] by applying:

$$\frac{Pi(Cb, Cr)}{\max\left(Pi(Cb, Cr)\right)} * 255 \quad (3.22)$$

As shown in Figure 3.20, color information helps adjust the regions of the body parts, where three different skin color levels (whitish, yellowish, dark brownish) do not affect the filtering results, as well as different levels of clothing. And as seen from Figure 3.20 (e) and (f), the wires lying on the chest are rejected by the skin color filter as expected.



(a) (b)

(c) (d)

(e) (f)

Figure 3.20 Fine-tuning of the human body regions: (a), (c), (e): baby body after applying the depth thresholding; and (b), (d), (f): after applying the skin-color filter.

3.3.2. Body parts detection with skeleton data

Anatomical knowledge defines how the human body's bones are connected into a whole through joints, muscles, ligaments and other tissues, and support the body. As a result, the bones form the scaffold of the human body, dividing it in different parts. For this reason, a skeleton model can well represent the connectivity and topology of the shape of the object, and the human skeleton data supported by a depth sensor can be used to recognize body joints and to describe relevant information about different areas of the human body.

Inspired by the work of Zhi-guo et al. [104], a skeleton recognition method based on depth images is proposed. However, the work of Zhi-guo et al. relies on three-dimensional human motion data or scan data, which leads to problems including ineffective skeleton extraction, lack of the original topology of the body, and redundant skeletons when data are lacking. The proposed method rather extracts the external contours of the human body from the depth image, and uses polygonal approximation algorithms and detailed bone extraction constraints to extract human skeleton information in external polygons. Based on these segments, the abdominal region of interest that we intend to automatically retrieve over the baby is defined as the intersection of a quadrilateral area, bounded by the two feature points of the elbows and two feature points of the hips, with the segmented area obtained with skin color filters described previously. The resulting area of interest is shown in yellow in Figure 3.21(b).

(a) (b)

Figure 3.21 Skeleton models and definition of the abdominal region: (a) extracted skeleton models; (b) the abdominal region of interest defined in yellow.

The skeleton recognition method includes the following steps:

(1) Retrieve the depth image of the subject after applying background removal and fine-tuning of the human body region operations (Figure 3.22(a));

(2) Starting from the edge of the image, use Halcon template matching function [105] to find the closest round part, considered as the baby's head, which is marked in blue circle in Figure 3.22(b);

(3) Detect all contours by Canny edge detection in the depth image (Figure 3.22(c));

(4) Discretize the external contour curve into an external polygon using a polygonal approximation algorithm, which proceeds as follows;

(a) Connect a straight-line AB between the two points A and B at the beginning and end of the curve, which is the chord of the curve, as shown with the yellow line in Figure 3.22(d);

(b) Obtain the point C on the curve with the largest distance from the straight-line segment, and calculate its distance d from AB;

(c) Compare the distance with a predetermined threshold range. If it is within the threshold range, the straight-line segment is used as an approximation of the curve, and the curve is processed; if it is smaller, then discard it;

(d) If the distance is greater than the threshold range, use C to divide the curve into two sections AC and BC, as shown with the blue line in Figure 3.22(d), and perform (a)-(c) on the two sections of the curve, respectively;

(e) When all the curves have been processed, the polylines formed by connecting the divided points in turn can be used as an approximation of the curve (Figure 3.22(e)).

(5) Select a point p within the external polygon and two points q_1, q_2 on the external polygon closest to p (q_1, q_2 are on different sides) to determine their local maximum. As shown in Figure 3.23, the dotted line is the central axis (skeleton) of the long side of the rectangle. There is also a horizontal solid line that intersects vertically at the intersection point c. The horizontal solid line and the two vertical edges of the rectangle also intersect perpendicularly, and the intersection points are r1 and r2. Another four points a, b, c, and d are on the solid line in order. According to the definition of the Euclidean distance transformation, the Euclidean distance of point a is the length of a to r1 | ar1 |, and the distance of point b is | br1 |. The distance of point c is | cr1 | = | cr2 |. Point d to the right of point c has the closest background point to r2, so its Euclidean

distance is | dr2 |, and e is | er2 |. The Euclidean distance values of all 5 points are shown in Figure 3.23(b). Therefore, in the horizontal direction, point c has experienced a local maximum, and point c is the point on the central axis, which is the skeleton point, so all the skeleton points have this property. However, because of the uncertainty of the shape, size, and orientation of the skeleton to be extracted, the skeleton points are different from the regular graphics in that they experience a local maximum distance in a certain direction but are local maxima in an irregular direction. The skeleton points can be determined by this property [106], which is that if p satisfies the following equations (3.23) and (3.24), then p is the skeleton point (Figure 3.22(f)):

$$\begin{cases} D^2(q_1,p) - D^2(q_2,p) \le \max(\|x_1 - x_2\|, \|y_1 - y_2\|) \\ D(p,q_0) \le D(q_1,q_2) \end{cases} \tag{3.23}$$

$$\begin{cases} D(\overline{q_1}, \overline{q_2}) > T_1 \\ L_1 \cap L_2 = \emptyset \end{cases}, if\ q_1 \in P_e, q_2 \in P_e \tag{3.24}$$

Where q_0 represents the midpoint of q_1 and q_2; D() represents the Euclidean distance; L_1, L_2 respectively represent the edge of the polygon where q_1, q_2 is located; if q_1 is the vertex of the polygon, then $\overline{q_1}$ equals to the vertex q_1 while if $q_1$is not the vertex of the polygon, then $\overline{q_1}$ equals to L_1; similarly, if q_2 is the vertex, then $\overline{q_2}$ equals to q_2, otherwise $\overline{q_2}$ equals to L_2; T_1 is a preset parameter;

(6) Consider the baby's head part determined in step (2) as one anatomical point, and the remaining skeleton lines are further optimized by the polygonal approximation algorithm to keep the bone lines in the largest connected domain and remove the other noise bone branches, as shown in Figure 3.22(g);

(7) Consider the endpoints of the bone lines as joint points. Combined with the baby doll's color image shown in Figure 3.22 (h), the complete skeleton data and joint point distribution can be shown in Figure 3.22 (i). Among them, the number of each joint point is marked from the head. The rules for serial numbers are:

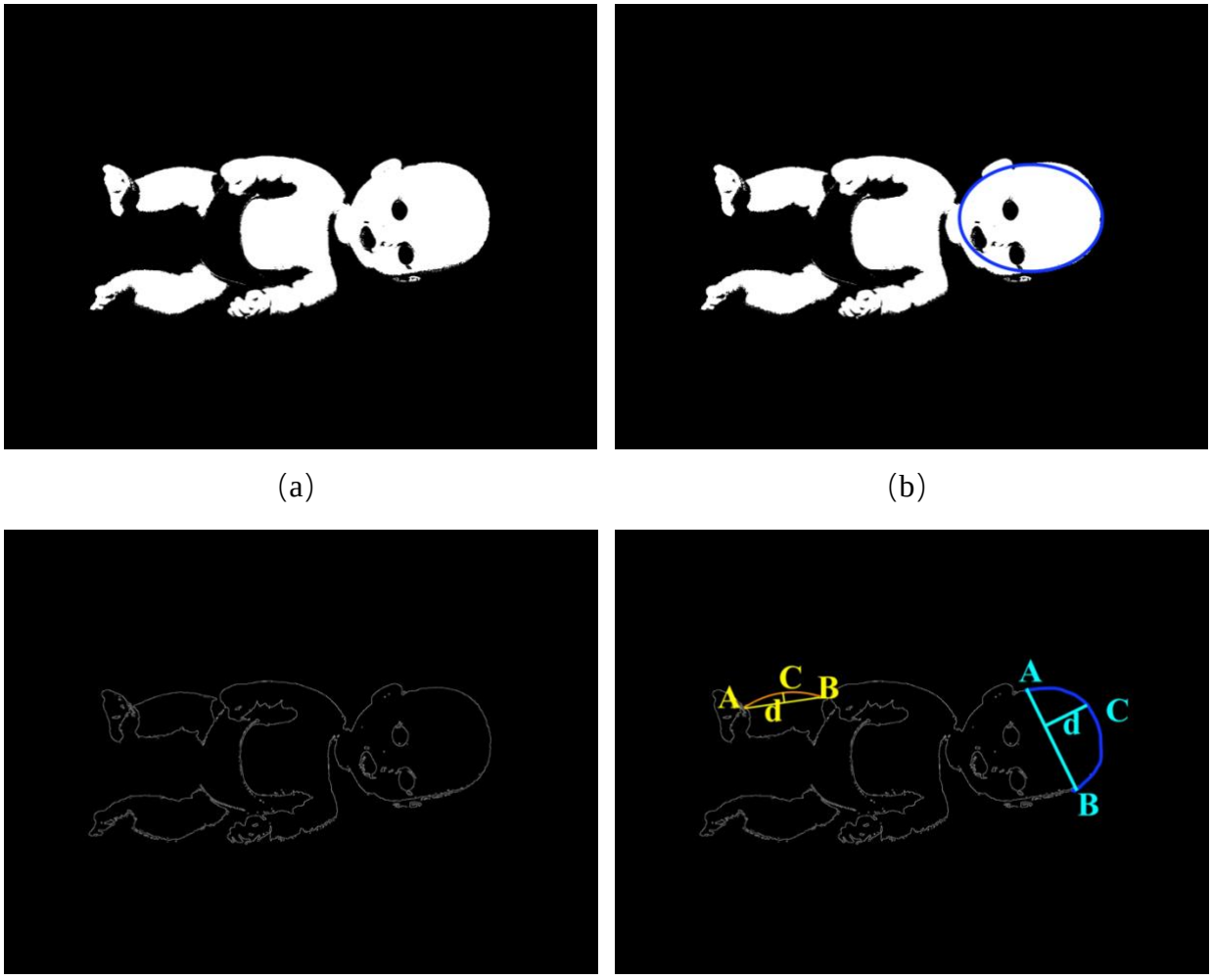
(a) Mark the serial number from 1, where 1 is always the head position;

(b) Find the next unlabeled joint point adjacent to the joint point of which each serial number is located in sequence;

(c) For multiple joint points adjacent to the same joint point, the joint point to the left or above the line between the joint point number 1 and the current joint point takes precedence, to the right or below joint point is next, and the joint point opposite to the

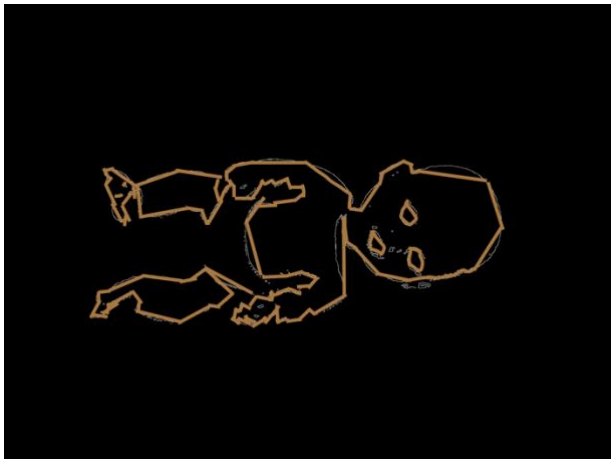
direction in which the current joint point points to the joint point numbered 1 has the lowest priority;

(8) Define the initial abdominal region as the quadrilateral areas surrounded by the four feature points 6, 7, 10, 11, as shown in yellow in Figure 3.22(j), which still include some background areas that do not belong to the body. So, the final segmentation is the intersection of the initial segmented abdominal area defined by joints 6,7, 10, 11, and the skin color filter result described previously, as shown in Figure 3.22(k).

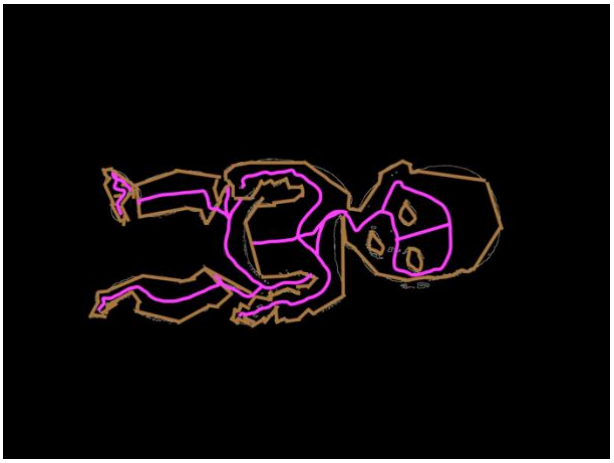


(a) (b)

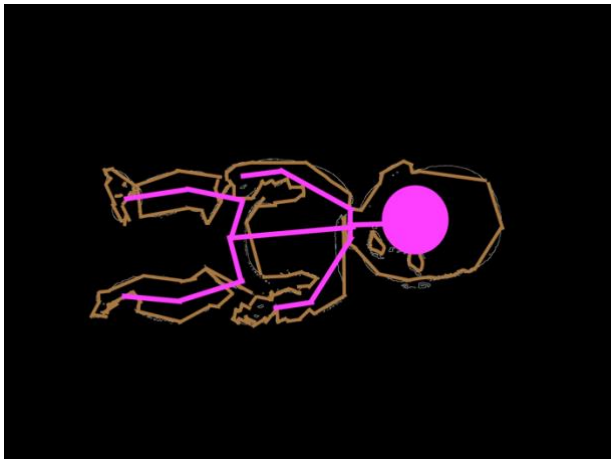
(c) (d)



(e)

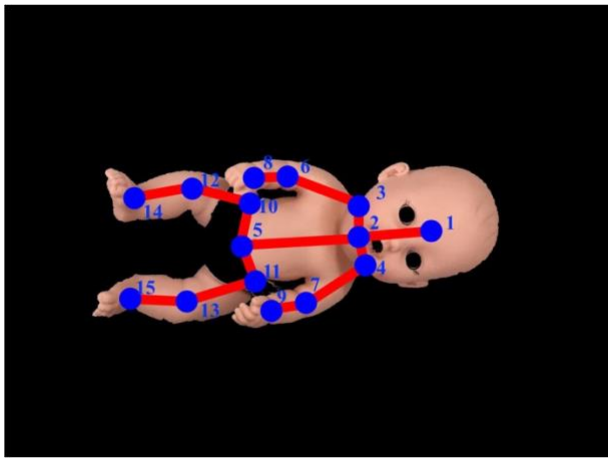


(f)

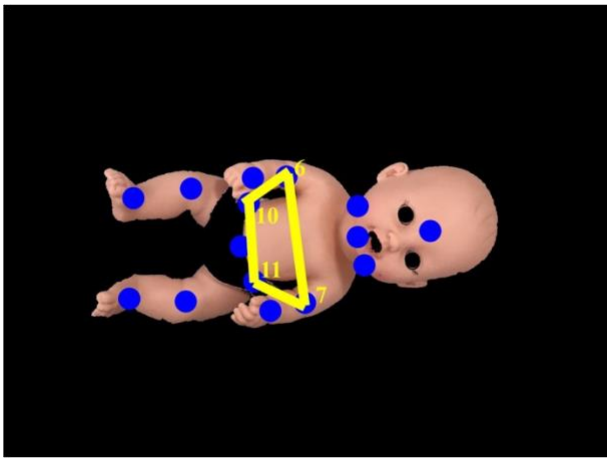


(g)

(h)



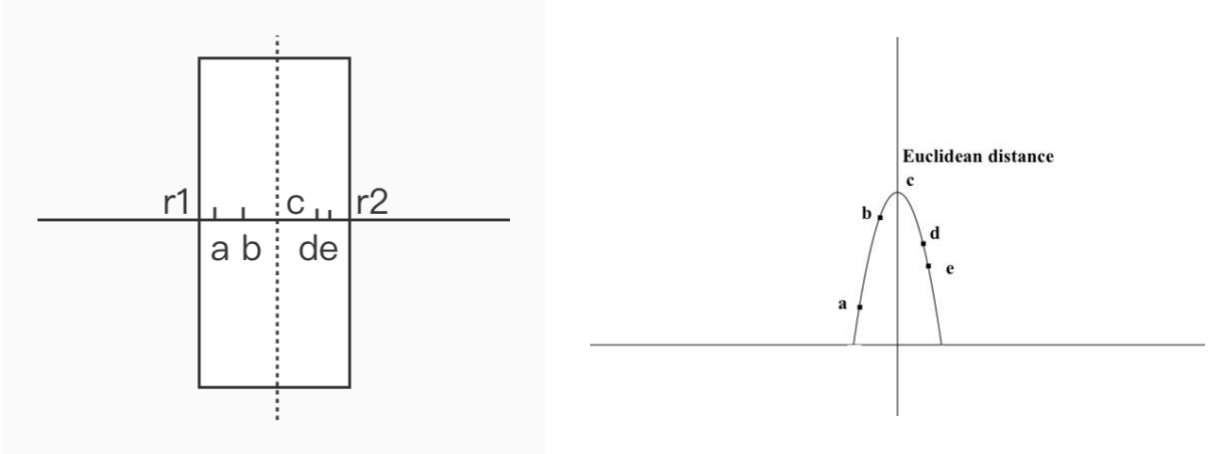
(i)



(j)

(k)

Figure 3.22 Segmentation results performed over a mock baby doll: (a) depth image after applying background removal and fine-tuning of the human body region operations; (b) find the closest round part from the edge of image, considered as the baby's head within the blue circle; (c) Canny edge detection; (d) polygonal approximation algorithm; (e) discretize the external contour curve into an external polygon; (f) find skeleton points based on local maximum algorithm; (g) keep the skeleton lines in the largest connected domain and remove the other noise bone branches; (h) original color image of the baby doll; (i) the complete skeleton data and joint points distribution with order numbers; (j) initial definition of the abdominal region; and (k) final segmentation result of the abdominal region of interest.



(a) (b)

Figure 3.23 Local maximum definition: (a) the positions of the 5 points; (b) Euclidean distance of those 5 points.

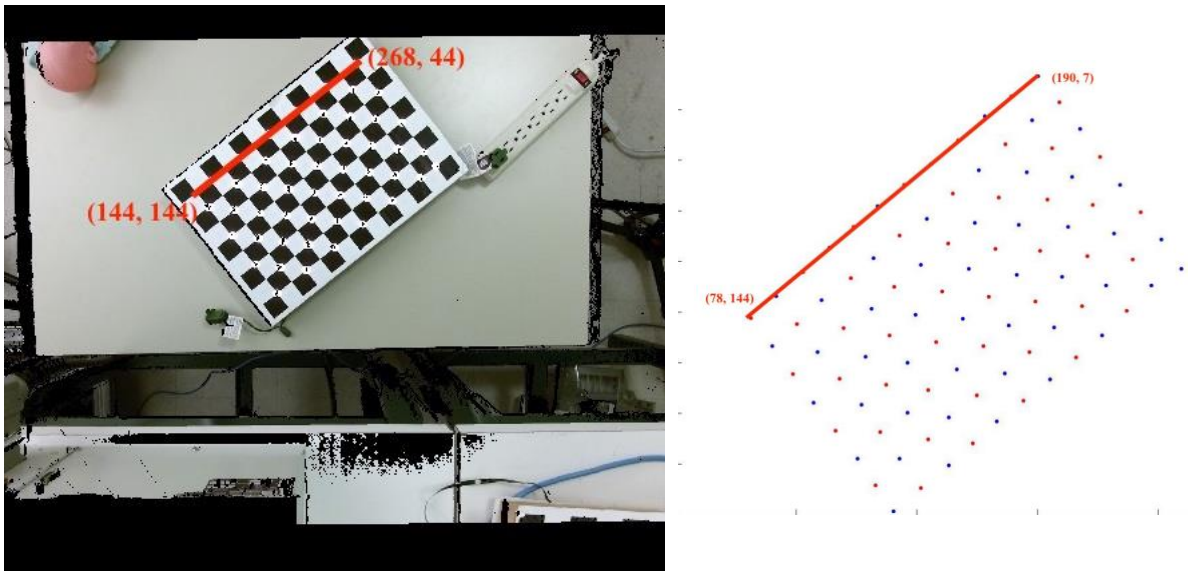
3.3.3. Mapping of region-of-interest between RGB-D image and IR image

Since the resolutions of the RGB-D and IR images are not the same, the segmentation results in the RGB-D image obtained previously cannot directly be used to extract the corresponding region in the IR image. Based on the relationship between pixels in the region of interest, the mapping operation is completed by calculating the proportional relationship between the RGB-D image and the IR image. The procedure is defined as follows:

We use a regular checkerboard pattern (such as a calibration target mentioned in section 3.2) to measure the ratio of the length of the same side in the RGB-D map and the IR map. As seen from Figure 3.24, for the same segment length, the abscissa difference in the RGB-D image is 268 - 144 = 124, and the ordinate difference is |44 - 144| = 100; while in the IR image, the abscissa difference is 190 - 78 = 112, and the ordinate difference is |7 - 144| = 137. Therefore, the ratio of the abscissas between the IR and RGB-D images is : IR_{ab} : $RGB\text{-}D_{ab}$ = 0.9032 : 1, and the ratio of ordinates is: IR_{or} : $RGB\text{-}D_{or}$ = 1.37 : 1. In order to obtain more accurate ratio value, the final ratio of the RGB-D and the IR images is estimated by averaging several sets of data, which results in the following experimental scaling ratios, which are then used to support the projection of the extracted region of interest from the RGB-D map to the IR map:

$$IR_{ab} : RGB\text{-}D_{ab} = 0.9053 : 1 \quad (3.25)$$

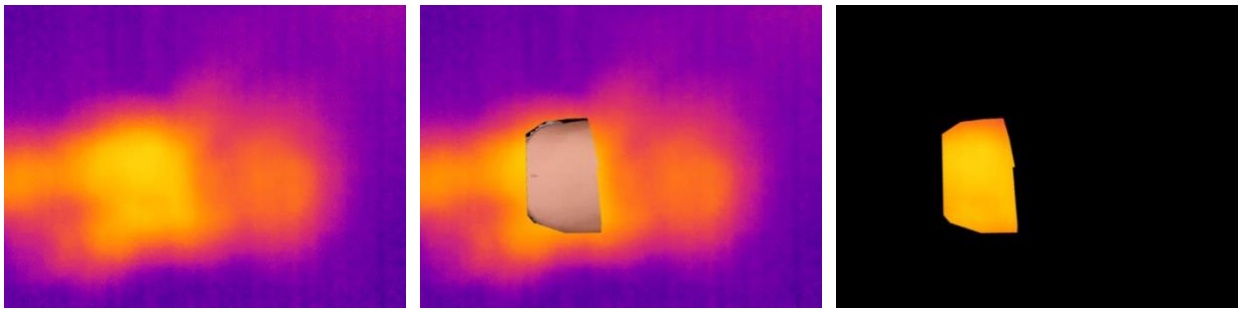
$$IR_{or} : RGB\text{-}D_{or} = 1.3674 : 1 \quad (3.26)$$



(a) (b)

Figure 3.24 Ratio of the length of the same side in the RGB-D map and the IR map: (a) same length in the RGB-D map with resolution of 512 x 424; (b) same length in the IR map with resolution of 320 x 240.

Figure 3.25 illustrates a segmentation result obtained by applying the methods proposed in this chapter over a sample infrared image acquired on the baby doll. It can be seen that the method reliably extracts a temperature distribution map over the region of interest to detect NEC from a complete human body. It properly removes the areas that may create interference in thermal data analytics, such as clothing, background, hands of nurses assisting in the image collection process, etc., that have a similar temperature to the baby's body. Therefore, it maintains the original shape structure of the body, making the segmentation results more accurate, complete and robust in comparison to previous unconstrained IR imaging approaches [9, 11].



(a) (b) (c)

Figure 3.25 Illustration of segmentation and region of interest extraction results projected over an IR image of a baby while considering scaling factors; (a) original IR image; (b) mapping of segmented area from RGB-D image to IR image; and (c) extracted abdominal area of interest from IR image.

3.4 Thermal Distribution Analysis and NEC Detection

After the region of interest is extracted, the temperature values collected within the entire region are directly analyzed. The IR image is first converted to a grayscale image representing the temperature distribution of the abdominal region. Then first-order statistics (mean, median, variance, standard deviation, quartile range, etc.) and box plots are used to compare temperature distribution differences between different groups, as in Ntonfo et al.'s previous work [11]. Based on known physiology, it is sufficient to describe and distinguish between NEC and normal infants by analyzing the thermal differences statistically because, under normal circumstances, infants in the NICU rarely have a fever. Even if they do, they are usually due to severe diseases such as sepsis, in which case they will be excluded from our research.

In order to fully illustrate the methodology, a comparison of thermal distributions was initially performed in a laboratory environment on a mock baby doll with artificially generated abdominal temperatures. One sample was considered a "normal" baby doll and was warmed to body temperature by holding it in human hands to allow it to be detected by the IR camera. The other sample represented a baby doll with "simulated NEC" which was reproduced by first warming the doll in human hands to bring it to body temperature, and then heating its abdomen with an incandescent bulb for 3 seconds prior to data acquisition. Although this cannot replicate the actual temperature distributions seen in a normal baby and one with NEC, it was an attempt to create a simulated situation where one subject had a different abdominal temperature compared to

the other case, to allow for testing of the proposed methodology. Figure 3.26 and Table 3.7 shows the results of the two different samples, using Matlab [107] for data analysis. It is important to mention that the values presented refer to grayscale levels in the IR image, and not to absolute temperature.

Table 3.7 Statistical analysis of the abdominal temperature (grayscale levels in IR image) distribution over a mock baby doll.

	Simulated NEC	Normal
Mean	157.1	157.4
Median	172	168
Standard deviation	24.2	29.5
Min	81	71
Max	180	190
Skewness	-1.2	-0.5
Median absolute deviation	20.0	27.3
Interquartile range	30	57
Kurtosis	3.2	1.6

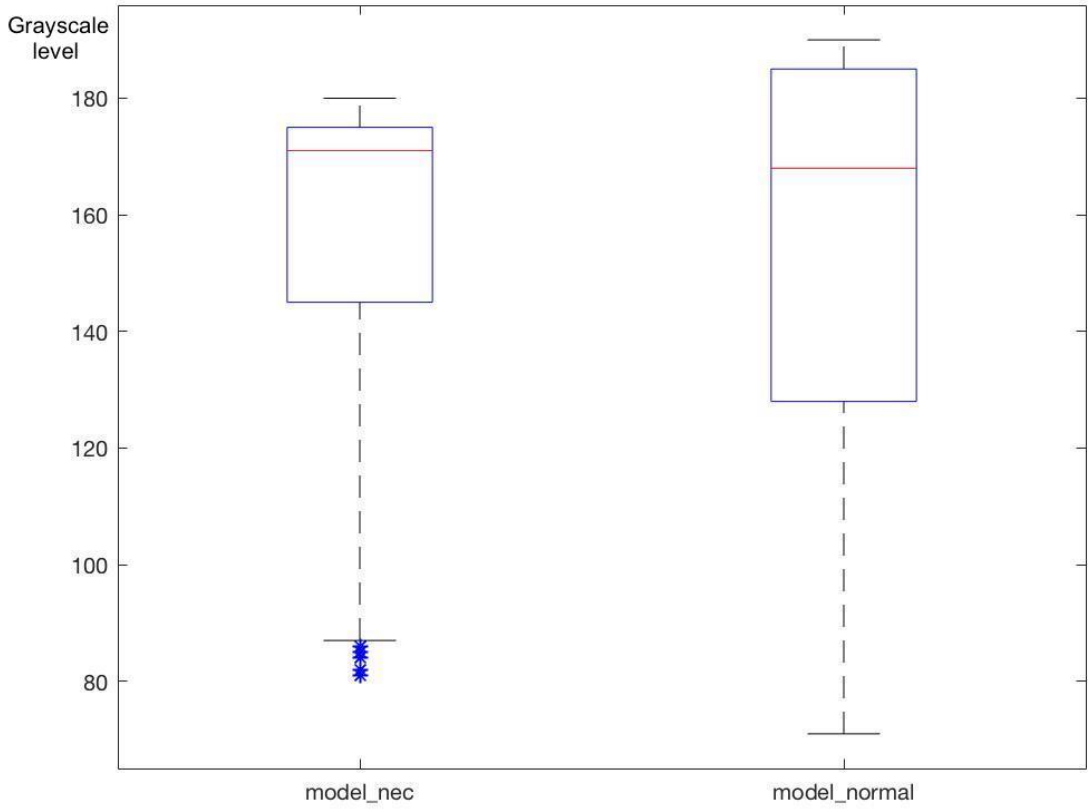


Figure 3.26 Box plot of grayscale distribution over abdominal region of baby model.

From this case, we can observe that the mean of the abdominal temperature grayscale values are similar, with the median value of the "simulated NEC" doll slightly higher than that of the "normal" case, which is coherent with the heating process. The minor differences in between the two groups are related to the imperfection in the simulation of the actual temperature distributions that can be observed in a normal baby and one with NEC. Furthermore, by observing the length of the box plots, it can be seen that the data distribution of the "simulated NEC" case is more concentrated, while the "normal" case is more dispersed. Although the median lines of both "normal" and "simulated NEC" dolls are closer to the upper quartile, suggesting the data distributions are both left-biased, it is clear to see that the "simulated NEC" data is more left-biased, which can also be quantified from the skewness value in Table 3.7. This simulation of thermal distribution analysis cannot accurately replicate what might be seen in the case of NEC in an actual baby, but rather it is an attempt to assess the proposed methodology's ability to distinguish two

different groups by their central tendency, dispersion, and distribution shape. More realistic testing scenarios, performed in a clinical environment on real subjects, are reported in Chapter 4.

3.5 Summary of Methodology

Chapter 3 details the design and implementation of a custom acquisition platform that meets specific requirements for NEC detection and the operational environment in the clinical setting by medical personnel who have limited time and are not trained in image processing and statistical analysis. Based on this set-up, a calibration method between individual multi-spectral sensors is proposed. An adaptive RGB-D based human body areas segmentation method with experimental results is also presented, as well as the projection mechanism of the region of interest over a corresponding thermal IR image map that supports abdominal temperature distribution analysis needed to detect NEC in newborns.

Chapter 4 Experimental Validation

Beyond the development phase detailed in chapter 3, this chapter reports on an experimental validation of the acquisition platform, data processing and analytics that was performed on a number of live babies acquired in clinical settings. Section 4.1 introduces the recruitment of eligible normal and NEC babies in this research and the essential ethical considerations. Then the standard image data acquisition process is completed on babies accompanied by medical staff (section 4.2). Afterwards, section 4.3 illustrates data analysis and presents experimental results obtained on the two groups of babies. Finally, section 4.4 summarizes the findings in this chapter.

4.1 Ethical Protocol

4.1.1. Subject population

According the protocol submitted for ethics approval, a sample size of 10 normal babies was recruited to form the normal group, who were at 26+0 - 42+0 weeks corrected gestational age (GA) without any clinical, radiological or pathological signs commonly associated with NEC. They had no diagnosis of clinical sepsis or hypotension and were in stable condition (with respect to respiratory status, heart rate, blood pressure, oxygen saturation, and pain control) as determined by the bedside and study physicians. The participants were recruited at the Children's Hospital of Eastern Ontario (CHEO) and the Ottawa Hospital General Campus (TOH-GC) neonatal intensive care units (NICU), in close collaboration with Dr. Bariciak, who is affiliated with both units.

Since there are limited patients diagnosed with NEC each year at both NICUs, and considering that the estimated consent rate of eligible babies with NEC is about 75%, the target enrollment of NEC Group was five babies who were between 26+0 - 42+0 weeks corrected gestational age, diagnosed with signs and symptoms of definitive NEC and Bell's stage 2 or higher x-ray findings [10]. Of note, the occurrence of NEC was very low in both NICUs during the experimental validation phase of this study, and thus only 2 babies with NEC were enrolled.

4.1.2. Inclusion and exclusion criteria

For both groups, any baby with a known congenital anomaly involving the intra-abdominal organs, with umbilical lines, tapes, or dressings applied to the abdomen, with active suspected or proven sepsis, and any baby deemed to be clinically unstable by the bedside nursing/physician team or the study physician are excluded from enrollment.

4.1.3. Ethical considerations

Under the ethics approval obtained, the following considerations were taken into account.

(1) Known or possible risks and benefits to subjects

There are no known risks associated with thermography and infrared-based range imaging. Thermal imaging modality is non-invasive and non-ionizing. It does not send any electromagnetic waves to the subject. It merely records the infrared radiation emitted naturally by all bodies. A Kinect sensor is equipped with a passive color camera and an active range sensor that uses low power infrared light-emitting diodes to obtain the depth information of the surrounding scene by estimating signal travel time to and from the subject. The technology has been developed by the gaming industry and is dedicated initially to perform motion capture over human beings. Therefore, it represents no risk to the baby.

However, the population considered in this study is typically cared for in an incubator, often involving an overbed warmer, as shown in Figure 3.6. Plexiglas walls do not let infrared radiation pass to reach the thermal camera; therefore, it is necessary to remove the subjects from the incubator, as discussed in section 3.1. The heat being emitted by the overbed warmer may also interfere with the recording of infrared radiation emitted from the baby. Thus, the warmer is turned off while recording the images, of which the overall process should not exceed two minutes. The brief period during which the subject is outside of the protective environment of an incubator or overbed warmer could induce a decrease of the subject's body temperature if appropriate precautions are not taken. The brief removal of the subject from the incubator may also increase the subjects' stress level but not more than during regular medical care. Handling may cause transient changes in heart rate, blood oxygen saturation, or blood pressure. However, there is no evidence that these changes are significant or harmful during brief gentle handling in stable babies.

The procedure will not be performed or will be immediately discontinued in any baby who manifests any signs of instability.

There are no immediate medical benefits for the subjects involved in this study, as this is a further study to improve the infrared imaging tools used in the assessment and management of NEC. Being able to understand better the physiological mechanisms associated with NEC could, however, lead to improved management of the disease and reduce the mortality and morbidity rates in the future.

(2) Precautions taken to deal with known or possible risks

The entire procedure takes place in the neonatology care unit and under the immediate supervision of medical personnel. The overall process outside the incubator/with the heater turned off takes approximately 2 minutes: 60 seconds for cooling and 60 seconds for imaging. Nurses assess the pain level of the babies during the imaging procedure.

Neonatal health care specialists are present at all times to ensure that the subject's vital signs are within acceptable limits. Any degradation of a subject's vital signs results in the immediate stopping of the imaging procedure, and the health care staff attend to the baby at that time.

The investigators provided in-service the nursing staff before the study. NICU nurses are familiar with signs of instability. However, as a precaution, a list of specific changes that require discontinuation of the procedure were drawn up to include changes in respiratory status, heart rate, blood pressure, color, and oxygen saturation particularly. Additionally, the nurse's decision to discontinue for any other reason, including her/his perception that the baby is not tolerating the procedure, will be final.

(3) Subject recruitment procedures, information to be collected and sources of information

The research physician identified the subjects who fit the criteria for inclusion in our study and approached the parents with a preliminary outline of the study and its purpose only. The research physician and any person not at arm's length from the clinical care of the infants were not otherwise involved in subject recruitment. If the parents indicated to the research physician that they are not interested in receiving any further information about the study, no further attempts to recruit the subject were made. If the parents indicated a willingness to see or hear further information, an investigator who does not have any relation to the subject then approached the potential subject's parents. He/she gave all the information necessary for the parents to make an

informed decision on whether or not to participate in this study. This process of informing the parents included declaring all the concerns proposed by parents, notifying the parents that declining the study does not have any effects on the baby's care. Moreover, the staff gave as much time as possible for the parents to consider when presenting the consent. No information concerning the subject were collected unless the parents expressly agreed to it and the consent form had been signed. Parents were also assured that they can withdraw their baby from the study at any time without explanation and that this does not in any way compromise the baby's care.

(4) Conditions under which a baby would be withdrawn from the study

Any baby who appeared not to tolerate the imaging procedure was withdrawn. Babies who were already physiologically unstable were not imaged unless and until stability returned. Babies who went to the operating room for abdominal surgery were also withdrawn to ensure no unnecessary handling of the abdominal wound, because the incision generates inflammation and heat, thus rendering thermographic images uninterpretable. Babies in whom a non-treatment decision was made by the parents and team were also withdrawn.

4.2 Data Acquisition

The following information was collected for each imaged subject in the database:

- Subject Study Identification number
- Subject Group: NEC vs. Normal
- RGB-D and IR images sequence of baby's whole body lasting 1 minute
- GA at birth, days of life, and corrected GA
- Weight at birth and current weight
- Presence of feed intolerance, stool changes, abdominal distention, apnea, lethargy, metabolic acidosis, thrombocytopenia (clinical indicators of NEC)
- Timing of last feed
- Timing of last bowel movement
- Radiographic changes suggestive of NEC, including date and time of x-ray and regions where pneumatosis is seen and regions where bowel thickening is seen
- Subject's skin temperatures from incubator/warmer bed probe
- Other comments

Permission was obtained from some parents to take photographs of the babies, that would be de-identified using pixilation of the head region, and used for study reporting purposes.

4.2.1. Data collection

The proposed experimental acquisition setup used to collect data is described in Section 3.1. Medical thermography requires a period of slight cooling of the subject in order to stabilize the body's surface temperature. The room temperature was maintained slightly below thermoneutrality, as is usually the case in a neonatal intensive care unit, to ensure accurate imaging without inducing any sustained cold stress. Before imaging, all clothing was removed except for clothes covering extremities such as the head, hands, or feet. While imaging, the baby was removed out of the incubator by medical staff for less than or equal to 60 seconds. The distance from the baby was determined to about 60 cm, which is the shortest distance that ensures both Kinect sensor and IR camera can collect valid information. The baby was lying flat, imaged with the camera facing downward, as shown in Figure 4.1. During the 60 sec imaging process, the clothes and beddings were removed from the baby' abdomen, avoiding the interference with subsequent segmentation operations. No attempts were made to relocate probes and wires that are within the main imaging area as the goal of this new technology is to detect and remove those areas from the thermal datasets automatically.

Figure 4.1 Data collection at the TOH-GC NICU.

As seen from Figure 4.1, all sensors are mounted on a single tripod rack that can be rolled over the incubator. Once the apparatus height is set and is positioned over the location where the baby is to be imaged, the thermal, color, and depth images are displayed to fine-tune the position of the sensors. Then, the images start recording for 60 seconds after the over bed warmer is shut off and the bedside nurse lifts the hood of the incubator or pulls out the mattress on which the baby lies for a brief time. After approximately 60 seconds, the recording is stopped. Then, the mattress is pushed back into the incubator, the hood is closed, or the overbed heater is turned back. The bedside nurse also records an axillary temperature before and after the imaging to ensure the baby's temperature has not decreased significantly.

4.2.2. Confidentiality and privacy protection

The images and non-identified data collected during the study are stored on protected recordable media and a password protected, secure personal computer for analysis. An independent study number was assigned to each subject, and the link between the subject and the study number is kept in separate encrypted, password protected databases at TOH-GC and CHEO. Informed consent forms are kept separately from the data and are stored in a locked office at the

respective hospital where the subject was enrolled. Only the research physician has access to the link and the consent forms. Moreover, only the researchers directly involved in the study have access to the images and other data collected. In this thesis, regions of images depicting a subject have been blurred manually to also ensure privacy and confidentiality.

4.3 Data Analysis and Experimental Results

4.3.1. Segmentation and region of interest extraction

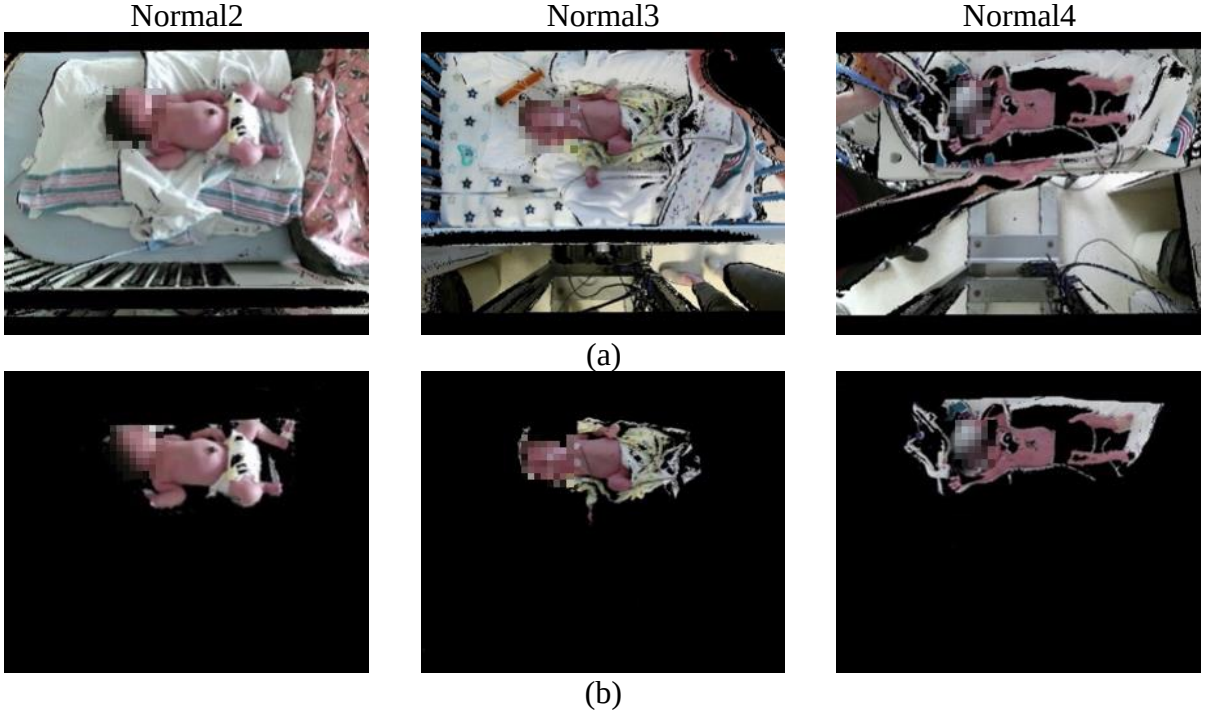
In this section, three different sets of results from subjects Normal2, Normal3, and Normal4 are presented as an illustration of performance and relative robustness of the proposed solution with respect to different baby positions, postures, and overlying clothing/monitoring equipment in the normal group. First, the original RGB-D images are presented in Figure 4.2(a), followed by background removal results after applying an adaptive depth threshold in Figure 4.2(b). During this process, it is clearly seen that besides the baby's body, which we are interested in, there is also some bedding, clothing, probes, and wires remaining, which may affect the body parts recognition because their temperature as they sit near the body is similar to that of the body itself, making it difficult to segment in the IR image.

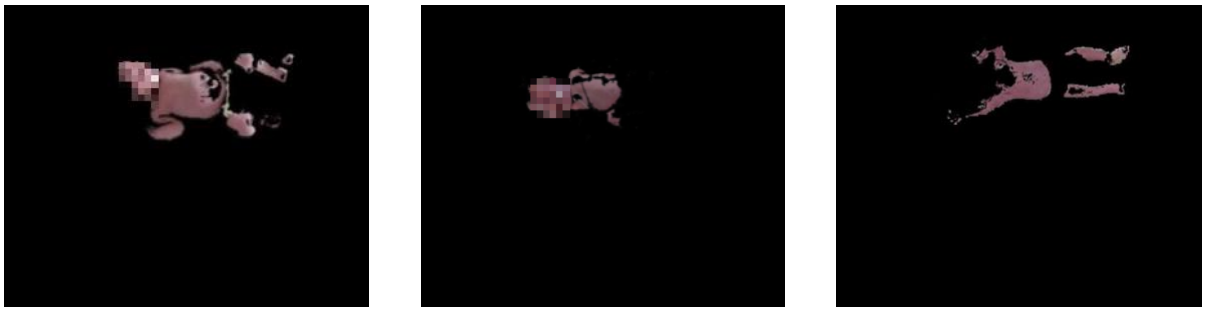
The skin color filter procedure was performed on each image to remove the other interferences such as clothing, bedding, tubes, probes, etc. (Figure 4.2(c)). Afterward, the skeleton recognition method based on depth image was applied to recognize the skeleton distribution and 15 joint points on the baby's body (Figure 4.2(d)). The latter were then used to isolate the relevant abdominal area bounded by blue quadrilaterals, as shown in Figure 4.2(e).

As can be seen from subject Normal2, who exhibits a relatively complete and regular body posture, the procedure reliably detects all the skeleton distribution and relevant anatomical joints as was seen for the baby model described in Chapter 3. However, processing data from the Normal3 baby produces an incomplete skeleton map due to the legs being covered by beddings and clothes. Although the covered parts bring some difficulties for leg recognition, the four joints used to locate the potential abdominal area are not affected, resulting in a consistent segmentation result in the end. The Normal4 baby is another special case because of his raised arms. In this case, the midpoints of the line segment connecting the elbow joints and the ipsilateral hips are considered as feature points to form the blue quadrilaterals instead of the actual elbow joints,

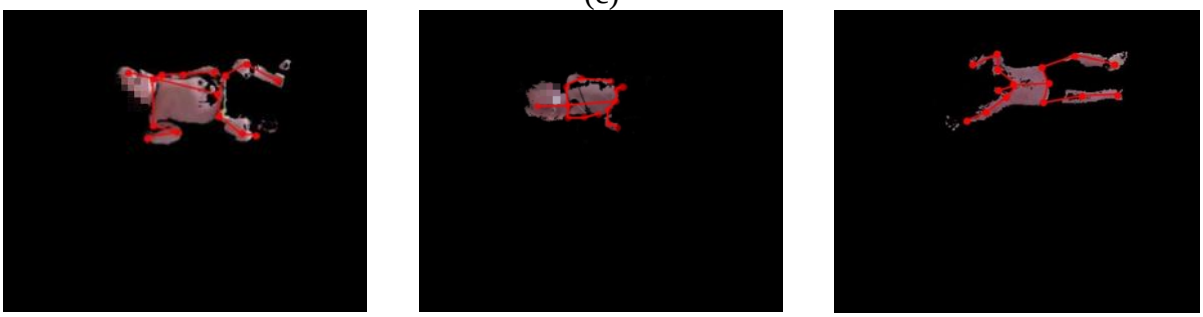
resulting in the initial segmentation shown in Figure 4.2(f). Since the initial segmentation is restricted by four anatomical joints, it contains some erroneous areas which are not part of the human body, leading to segmentation of extra parts in the IR image that do not belong to the region of interest. Therefore, it is necessary to refine the initial results with the skin color filter.

The intersection region of blue quadrilateral and skin color filtered segment was extracted as the final region of the abdomen, as seen in Figure 4.2(g), leading to much more accurate extraction of only surfaces belonging to the baby's abdomen and dropping any remaining background regions that may have been retained in the initial segmentation from anatomical feature points only. The final segmentation results are then mapped onto the corresponding IR image for each subject, (Figure 4.2(h), to generate the extracted thermal map over only the abdominal region of interest, shown in Figure 4.2(i). The latter will be analyzed using first-order statistics in the following section. Appendix A provides the complete set of results for the other seven normal babies considered in this study.

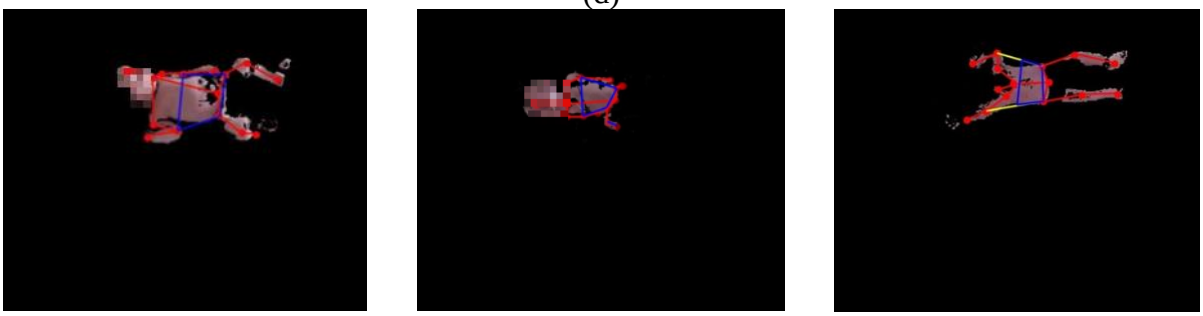




(c)



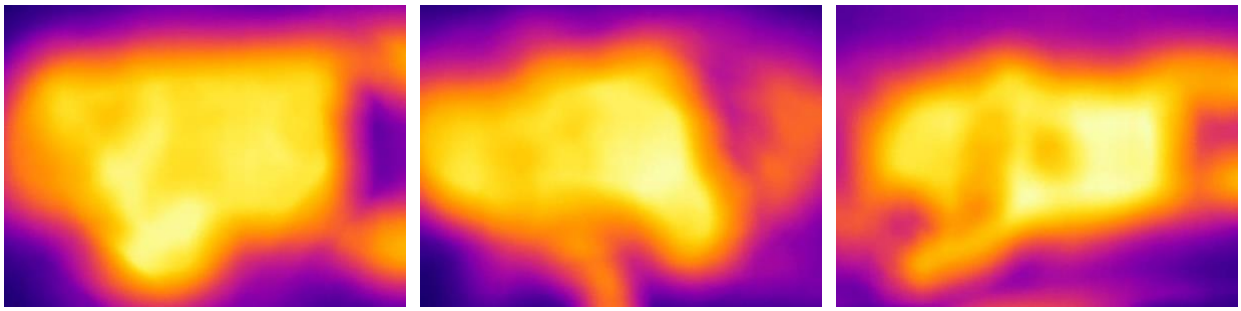
(d)



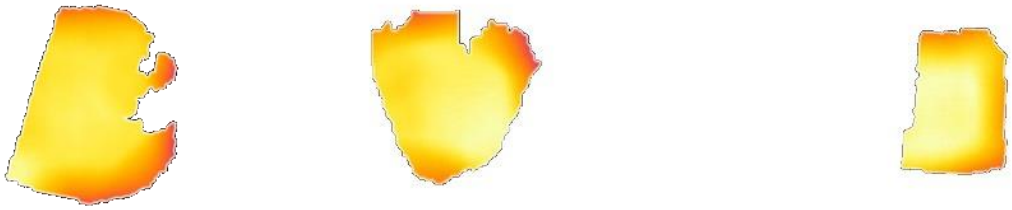
(e)

(f)

(g)



(h)

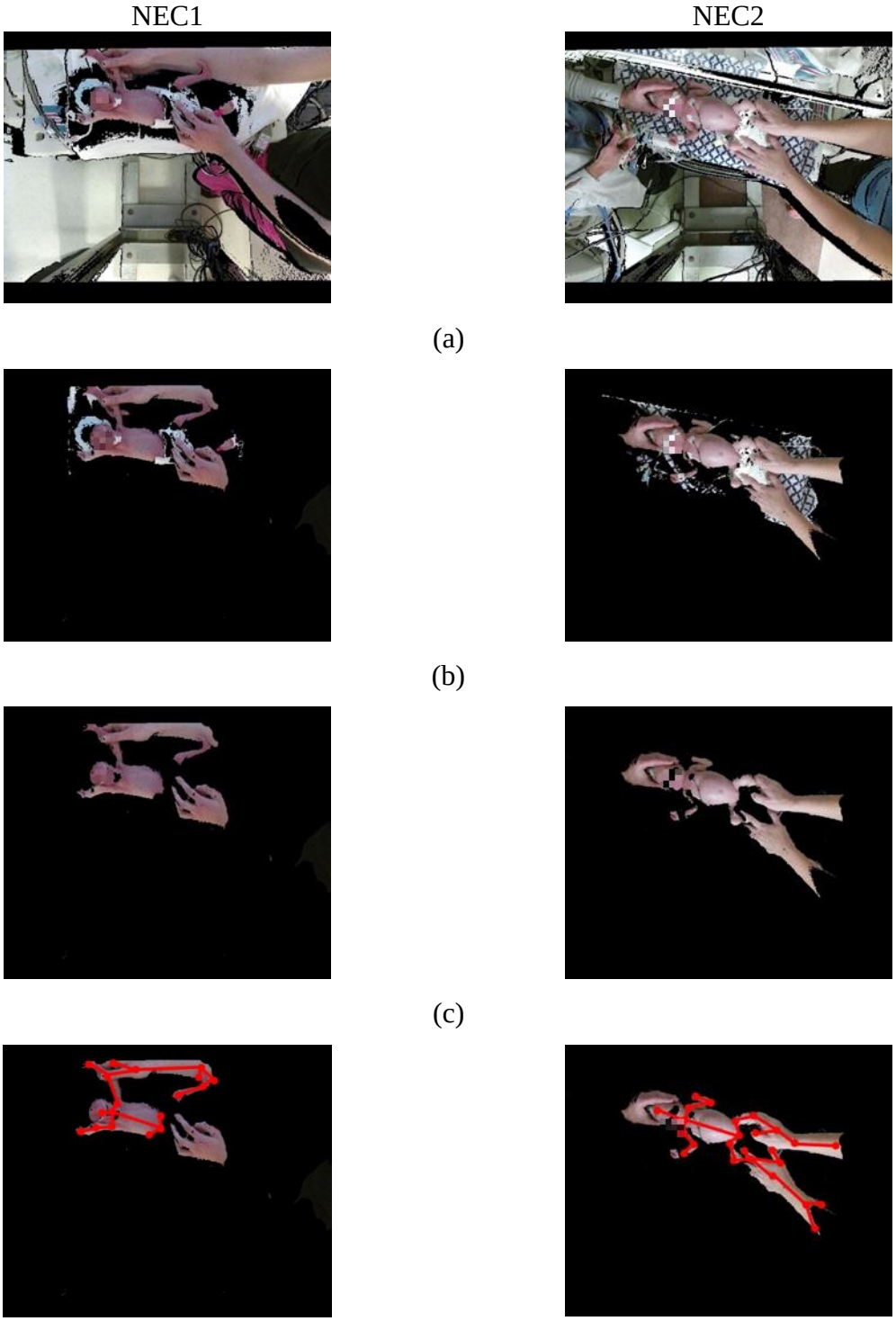


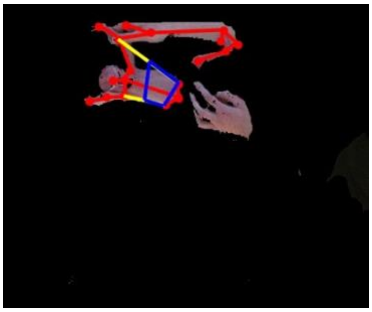
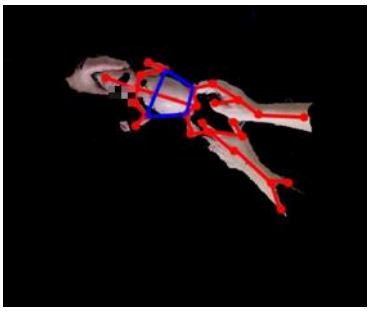
(i)

Figure 4.2 Segmentation and region of interest extraction results on multispectral data collected on real normal babies in a clinical environment, for three representative cases: (a) original RGB-D images; (b) background removal results; (c) after applying skin color filter; (d) recognized anatomical joints; (e) potential abdominal area restricted by blue quadrilaterals; (f) initial estimation of abdominal region; (g) refined extracted region of abdomen with skin filter segment; (h) original IR images; (i) final region of interest extracted in IR images.

Additionally, Figure 4.3 illustrates the complete segmentation and region of interest extraction results for the two subjects in the NEC group. Unlike those shown in Figure 4.2, both NEC cases were imaged with the hands and arms of the medical staff who helped secure the babies' position. Since the nurse's hands are basically at the same depth as the baby's body, the hands are retained in Figure 4.3(b). They are also kept after applying the skin color filter in Figure 4.3(c), as expected. However, the interferences do not actually affect the skeleton points recognition because the nurse's hands and arms are either connected to the baby's body so that they are considered an extension of the baby's body, or are separated from the baby's body and then abandoned during the identification process, as seen in Figure 4.3(d). In the former case, the number of the established skeleton points has not changed before extending to the hands, so the four certain skeleton points used to restrict the initial abdominal region are still reasonable and feasible, as shown in Figure

4.3(e). The final results are similar with that obtained on the normal group, as seen from Figure 4.3(f) to (i).

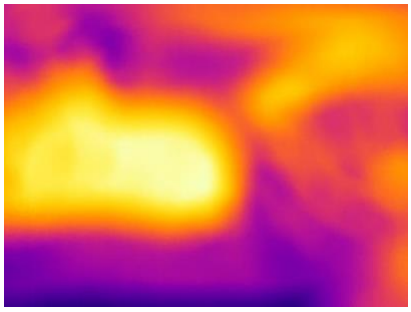
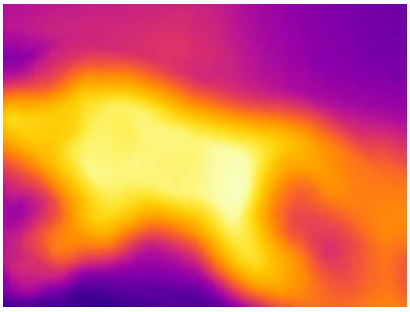


(e)


(f)


(g)

(h)

(i)

Figure 4.3 Segmentation and region of interest extraction results on multispectral data collected on real NEC babies in a clinical environment: (a) original RGB-D images; (b) background removal results; (c) after applying skin color filter; (d) recognized anatomical joints; (e) potential abdominal area restricted by blue quadrilaterals; (f) initial estimation of abdominal region; (g) refined extracted region of abdomen with skin filter segment; (h) original IR images; (i) final region of interest extracted in IR images.

4.3.2. Thermal distribution analysis

For each individual subject in the normal and NEC groups, there were six segmented IR images selected randomly to minimize the overall impact of biased data. Then the grayscale values were computed from the combined data, which contain all pixels within the region of interest from the six selected sample IR images corresponding to a given subject. Table 4.1 lists first-order statistical estimates extracted from the IR data over only the previously segmented region of interest for each of the ten normal babies on whom data was acquired at CHEO and TOH (documented in Section 4.3.1 and Appendix A). Table 4.2 lists similar statistics for the two NEC babies. After comparing the data of the individual subjects (grayscale pixel values for 6 images per subject were combined and then analyzed) in the normal group and that in the NEC group respectively (Figure 4.4 and 4.5), we observed that the grayscale distribution range within a same group is similar, as well as the distribution trend of the individual subjects as measured by median values.

In order to further synthesize the data of the normal and NEC groups\, we then amalgamated the data of all subjects within each group to formulate representations for the overall normal and NEC groups. The amalgamated first-order statistics of the normal and NEC groups,

respectively, and corresponding boxplots are presented in Table 4.3 and Figure 4.6 to support the analysis. While the estimated mean and median values exhibit similar trends, median values are closely examined in this study rather than mean values given that data distribution is not necessarily normal.

Table 4.1 Individual first-order statistics over ten samples of babies in the group of Normals.

	Mean	Median	Standard deviation	Min	Max	Skewness	Median absolute deviation	Interquartile range	Kurtosis
Normal2	129.7	151	48.1	26	175	-0.7	42.9	92	1.9
Normal3	134.7	158	45.1	35	174	-0.8	39.7	78	2.1
Normal4	116.2	147	62.5	1	173	-0.7	55.6	101	2.0
Normal5	131.4	158	46.5	27	175	-0.7	41.7	77	2.0
Normal6	128.6	152	48.8	18	175	-0.8	43.0	85	2.2
Normal7	126.8	144	49.0	24	173	-0.7	42.6	87	2.0
Normal8	126.8	150	50.8	16	174	-0.7	44.7	90	2.1
Normal9	128.6	156	49.4	17	175	-0.7	43.8	82	2.1
Normal10	130.6	157	47.7	30	175	-0.7	42.9	85	1.9
Normal11	131.5	155	45.6	33	175	-0.7	40.6	77	2.1

Table 4.2 Individual first-order statistics over two samples of babies in the group of NECs.

	Mean	Median	Standard deviation	Min	Max	Skewness	Median absolute deviation	Interquartile range	Kurtosis
NEC1	162.3	180	38.9	27	195	-1.4	30.9	42	3.8
NEC2	159.5	180	39.1	26	194	-1.2	32.4	54	3.4

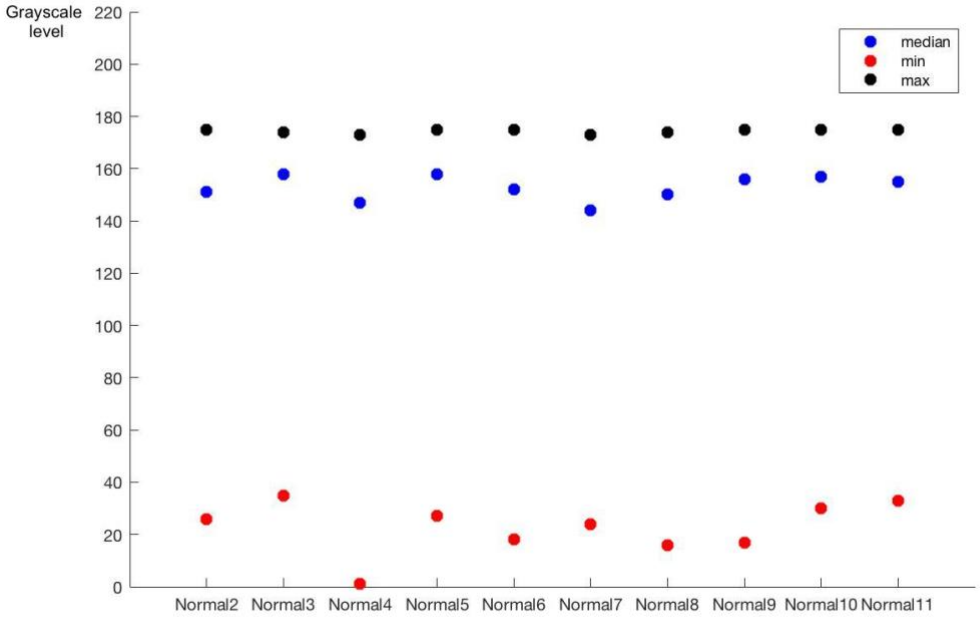


Figure 4.4 Distribution of individual grayscale level representing temperature over abdominal region of the normal group.

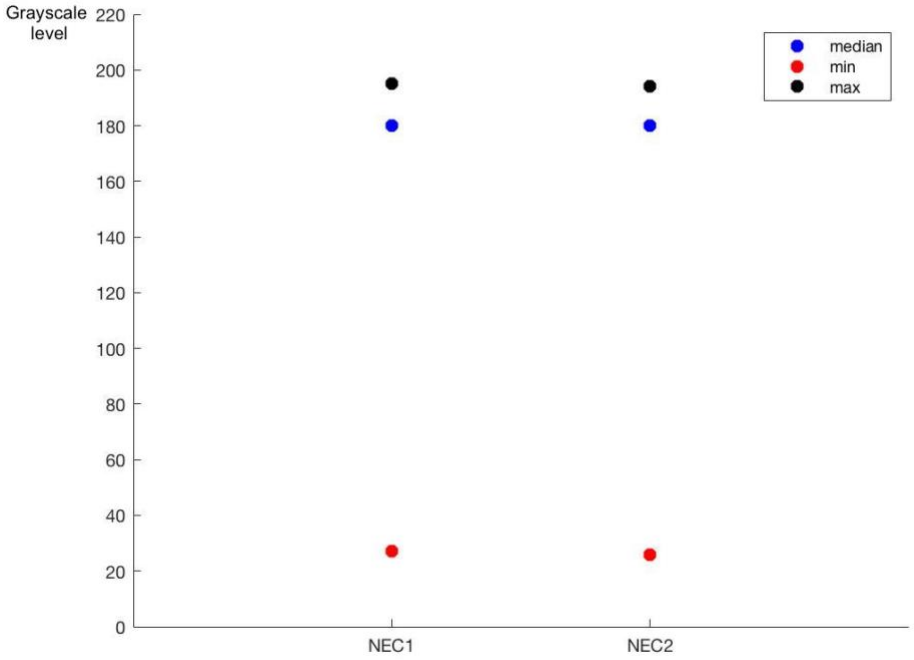


Figure 4.5 Distribution of individual grayscale level representing temperature over abdominal region of the NEC group.

Table 4.3 Amalgamated first-order statistics over the normal and NEC groups.

	Mean	Median	Standard deviation	Min	Max	Skewness	Median absolute deviation	Interquartile range	Kurtosis
Normal	128.6	153	49.9	1	175	-0.8	44.0	86	2.2
NEC	160.6	180	39.1	26	195	-1.3	31.8	52	3.5

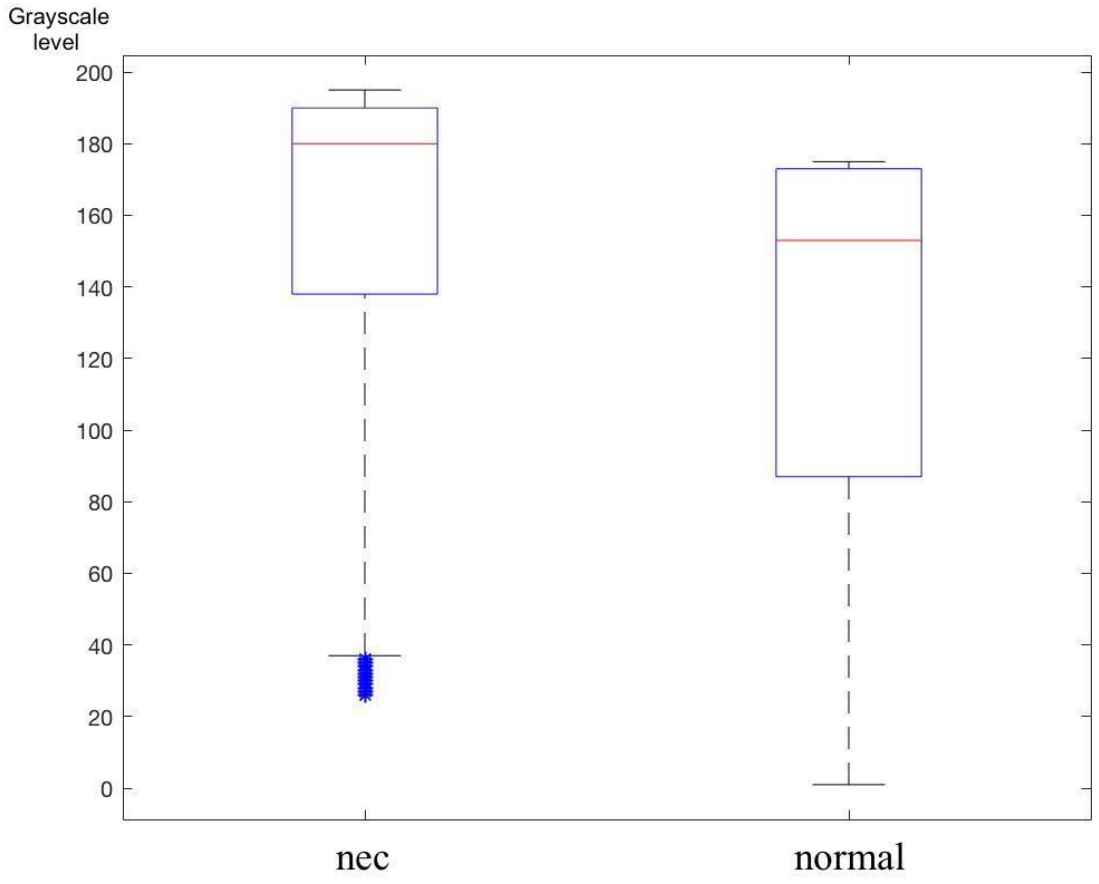
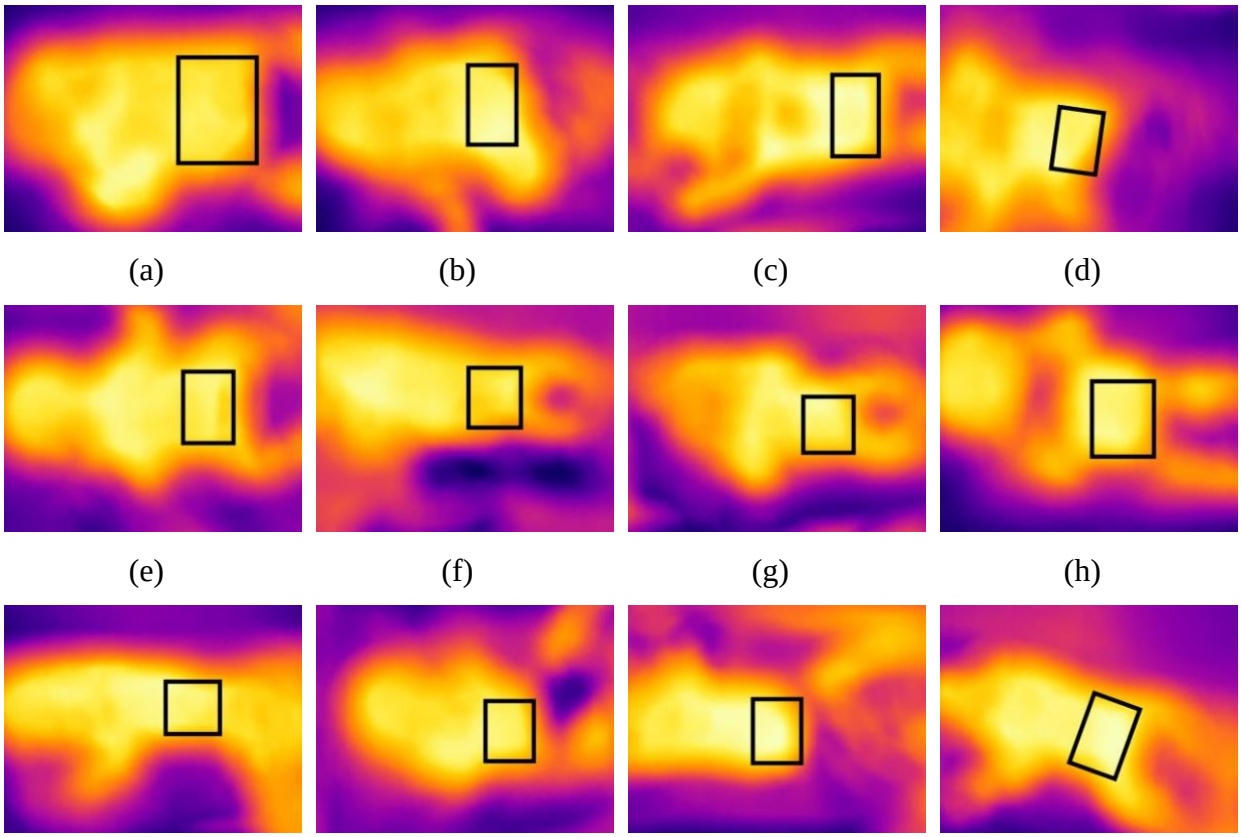


Figure 4.6 Boxplot of temperature distribution, represented by grayscale levels, over abdominal region of the NEC and normal groups.

As can be seen above, the median grayscale values are higher on the NEC group than on the normal (red lines in Figure 4.6), potentially representative of intestinal inflammation in the NEC babies, which triggers a rise in abdominal temperature. The box of the NEC group is also shorter than that of the normal one, showing that the temperature distribution throughout the abdomen is more concentrated in general, which may reflect that the abdominal temperature rise due to inflammation elevates the temperature of the entire NEC abdominal region vs the normal

abdomen which may have some more varied temperature distribution across the various underlying structures. Additionally, it can be seen from the NEC group that the median is closer to the upper quartile, revealing that most data are distributed on the right side of the X-axis, the tail of the curve extends to the left, and the data are left-biased, which is also seen for the normal group, but is less pronounced. The degree of left-biased can also be quantified from the skewness value in Table 4.3, indicating that the data distribution of NEC group is more left-biased than that of normal groups.

In order to compare the results above with that obtained when using the methodology proposed in a previous study by Ntonfo et al. [11], a rectangular area surrounding the umbilical stump was manually selected over each RGB-D image, and then mapped to the corresponding thermal image, as shown in Figure 4.7. As there were no specific protocols provided in the previous study, we hypothesized that the size and orientation of the rectangle do not affect the results, and related experimental error can be ignored. Table 4.4 and Figure 4.8 detail the statistical analysis performed using manually segmented thermal maps of the abdominal region of the same two NEC and ten normal babies, in accordance with the segmentation method used by Ntonfo et al. [11].



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 4.7 Rectangular areas manually selected over the abdominal regions on IR images acquired with the proposed imaging system, respectively on: (a) Normal2; (b) Normal3; (c) Normal4; (d) Normal5; (e) Normal6; (f) Normal7; (g) Normal8; (h) Normal9; (i) Normal10; (j) Normal11; (k) NEC1; and (l) NEC2.

Table 4.4 Amalgamated first-order statistics over normal and NEC groups based on previous segmentation method proposed in [11].

	Mean	Median	Standard deviation	Min	Max	Skewness	Median absolute deviation	Interquartile range	Kurtosis
Normal	131.7	161	50.6	1	175	-1.0	44.2	75	2.7
NEC	150.8	180	55.1	1	195	-1.3	44.9	69	3.5

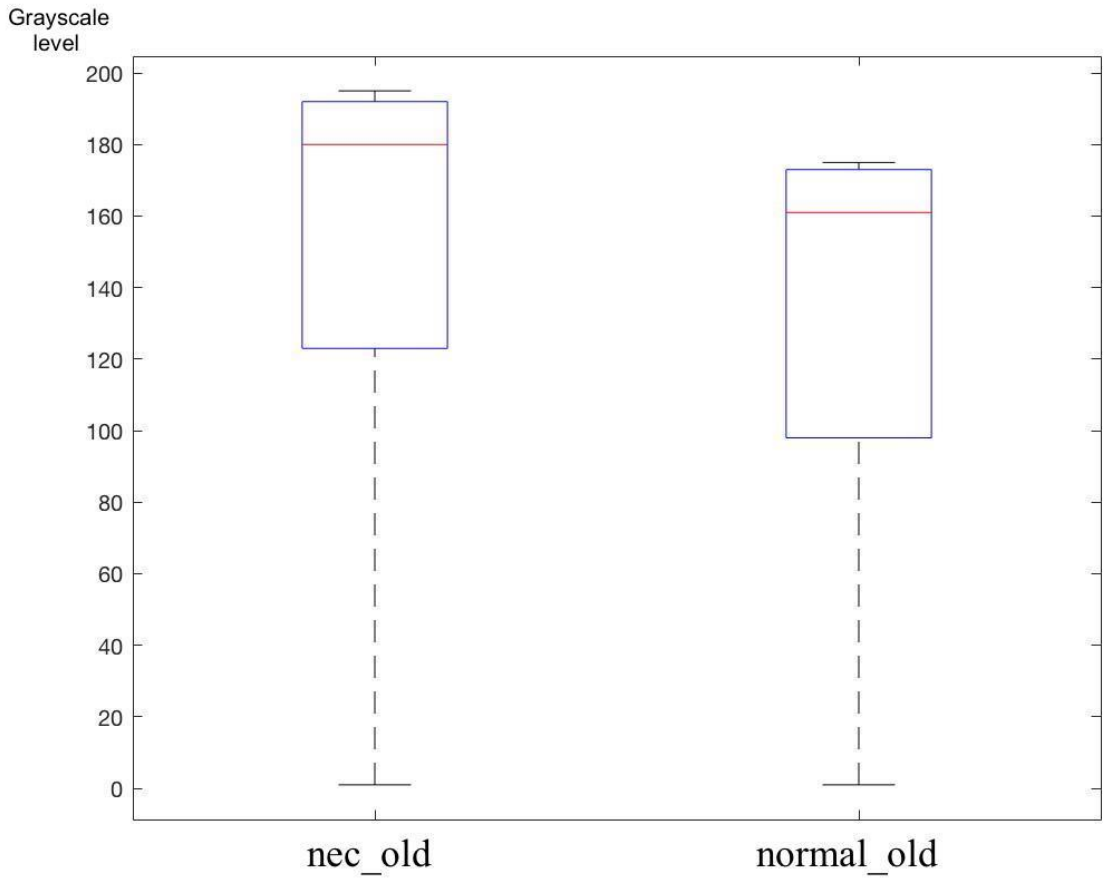


Figure 4.8 Box plot of temperature distribution, represented by grayscale levels, over abdominal region of the NEC and normal groups based on previous segmentation method [11].

As shown in Figure 4.8, when using the previously described segmentation and analysis method, the median of grayscale values distribution over the abdominal region of the NEC group is also higher than that of the normal group, and both two groups are left-biased as well. However, the spread of the grayscale distribution for each group is less pronounced compared to the results from the method proposed in this thesis. The relative positional relationship between the median and the quartile also does not show a large difference between the two groups either from the boxplot or the skewness values.

Furthermore, when compared to Ntonfo's work [11], our study has yielded conflicting findings on the overall differences in grayscale values between the two groups, where the median grayscale value of NEC group is higher than that of normal group in the present dataset but the median grayscale value of NECs is lower than that of normals based in Ntonfo's work. The reasons

can be related to limited number of subjects studied and differences in the IR thermal cameras. Considering that there was only one normal and one NEC baby reported on in Ntonfo's work [11] and there are ten normal and two NEC babies in this study, it is possible that the babies recruited in each study had different stages of NEC, with the babies in Ntonfo's study possibly having necrosis and dead tissue which has less inflammation and heat generation than tissue in earlier stages of NEC that has not died. Moreover, the different IR thermal cameras with different imaging principles and different resolutions used in the two studies could affect the collection of grayscale values of the IR images, which then could result in the inverted thermal data direction (higher-lower median temperature) reported.

Although there is ambiguity in the use of grayscale values to describe data from the different datasets, it still can be concluded that statistical analysis is feasible and attainable and show differences between NEC and normal groups in both Ntonfo's [11] and this work. However, the limited number of subjects makes it impossible to quantitatively summarize whether the comparison of data between the two studies is significantly different. Further study on an extended dataset are needed. Meanwhile, the present method provides a completely automated and more standardized segmentation process for extracting thermal data of the abdominal region compared with Ntonfo's method [11], which ensures full use of all available and relevant information while minimizing manual interference.

4.4 Summary and Discussion

Chapter 4 starts with a detailed description of the study protocol, as approved by the research ethics boards, that has been used for recruitment and data collection. Next, experimental results with the segmentation and region of interest extraction method proposed in Chapter 3 are presented. It was demonstrated that the proposed multi-spectral acquisition and automated analysis approach has improved the application of IR-based detection of NEC while operating in a real clinical environment. Several situations, such as interference from the bedside nurse's hands and arms, pieces of medical equipment in the field of view, or a baby lying in a variety of configurations are properly handled by the image processing framework. As such, data acquisition can take place quickly and without imposing severe constraints on the bedside team or the patient during the data acquisition process. As such, the method has been demonstrated to be robust in a

variety of representative clinical conditions, however for optimal results, it remains that it is ideal for babies to be well-positioned and naked without other additional human body parts within the sensors' field of view. Finally, an analysis of the available sample data is presented to show how a classification between NEC and healthy subjects could be performed, with minimal intervention from the medical personnel. While the dataset acquired for this study remains limited, clear trends in the first-order statistics have been observed. Furthermore, a comparison between the proposed automated segmentation and abdominal region of interest extraction method and alternative ways to manually retrieve information directly for IR thermal maps suggests that the automated method may be more effective.

Chapter 5 Conclusion

5.1 Summary

In this study, a combined multispectral sensor system is proposed in response to the need for automating the image processing and analysis procedures to support IR imaging-based NEC detection in newborns. The research also aims at filling an existing research gap by providing insight into leveraging RGB-D sensor data to further refine IR medical imaging technology.

A novel design of a multispectral imaging platform is introduced, as well as the corresponding calibration process for three sets of RGB, depth, and IR images collected from different viewing angles and resolutions. Calibration procedures among three sensors are customized to achieve correspondence between the different image coordinate systems and in preparation for subsequent segmentation and data alignment. The additional data channels provided by color and depth information are fully utilized to increase accuracy on segmentation compared to the traditional direct segmentation of IR images. The skeleton data inherently supported by the depth information is leveraged to extract the abdominal region of interest from the entire body image by means of anatomical joints recognition. To achieve this, an innovative skeleton points recognition method is introduced based on separate images from each frame instead of a sequence of images. As an outcome of automated segmentation, a complete temperature distribution analysis over the entire abdominal area replaces a more restricted and error prone analysis over manually selected regions in thermal maps, as proposed in earlier work, with the goal to increase accuracy of NEC detection. The new method provides the means to avoid missing key information regarding NEC.

Experiments validated the proposed framework on mock baby subjects in the laboratory as well as with representative sample data collected on real babies in the clinical environment of the CHEO and TOH-GC hospitals' NICUs. Experiments demonstrated that the proposed acquisition system and automated multispectral image processing methods offer the potential to improve detection of NEC and are adapted for operation in realistic clinical conditions with minimal imaging experience required from the medical personnel, and while minimizing the introduction of bias from manual operation, as observed in previous studies.

5.2 Contributions

This research contributes to the field of biomedical instrumentation for the detection of NEC with IR imaging technology integrated in a multispectral imaging approach with data amalgamation from RGB-D sensors. There are four main contributions described in this thesis:

i. The design of a novel multispectral acquisition platform composed of a calibrated thermal IR camera and Kinect sensor that work synchronously to collect color, depth and IR images.

ii. A custom calibration process to align data between the RGB and depth sensors with that of the IR camera, involving the use of a dedicated type of calibration target and preprocessing method for IR features extraction.

iii. A novel automated segmentation and region of interest extraction method that uses data amalgamation between color, depth, and IR information, leading to a more accurate and comprehensive extraction of relevant thermal information distribution when compared to manual selection as found in previous studies. A specific skeleton joints recognition approach was introduced to achieve this goal.

iv. An experimental study was conducted on real data collected in a clinical environment to demonstrate the feasibility of the approach, its robustness to realistic operational constraints, and potential benefits for medical personnel. An exclusive and new dataset containing rich color, depth and IR thermal data on 10 normal babies and 2 babies with NEC emerged from this research, which can be of benefit to future studies and other researchers.

This research led so far to two publications [19, 108].

5.3 Future work

Experimental results obtained with the proposed imaging framework suggest that the automated multispectral imaging approach works well, demonstrating the potential of an innovative way for detection of NEC. In consequence, future work will involve further thermal data analysis associated with NEC cases. More subjects with NEC symptoms should be recruited when available and consenting to participate in order to extend the sample database. Moreover,

subjects from various ethnicities should be recruited as well to supplement the database. Furthermore, more comprehensive data analytics methods can be selected to optimize the extraction of subtle variations in thermal data distribution between the normal and the NEC groups, in the pursuit to find ways to detect NEC in patients who are experiencing minimal clinical symptoms in the earliest phases of the disease. Besides, it will be desirable to compare and analyze the temperature distribution between relative parts of the body within the baby itself for future work. Further study involving machine learning could also provide unique insights. Although a large population of NEC and normal babies will be required during the learning process, should a machine learning-based solution be investigated, it is still worthy of exploring this path in the next phases of the research given the accuracy that can be achieved in data analysis, strong robustness to noise, and the potential to distinguish among subtle nuances in complex nonlinear relationships.

References

[1] Neu J, Walker W. (2011) Necrotizing Enterocolitis. *New England Journal of Medicine*, *364* : 255-64.

[2] Niemarkt, H. J., de Meij, T. G., van de Velde, M. E., van der Schee, M. P., Van Goudoever, J. B., Kramer, B. W., Andriessen. P., & de Boer, N. K. (2014). Necrotizing enterocolitis: a clinical review on diagnostic biomarkers and the role of the intestinal microbiota. *Inflammatory Bowel Diseases*, *21*(2), 436-444.

[3] Neu, J. (2015). Preterm infant nutrition, gut bacteria, and necrotizing enterocolitis. *Current* O*pinion in Clinical Nutrition and Metabolic Care*, *18*(3), 285.

[4] Oh, S., Young, C., Gravenstein, N., Islam, S., & Neu, J. (2010). Monitoring technologies in the neonatal intensive care unit: implications for the detection of necrotizing enterocolitis. *Journal of Perinatology*, *30*(11), 701.

[5] Epelman, M., Daneman, A., Navarro, O. M., Morag, I., Moore, A. M., Kim, J. H., Faingold, R., Taylor, G., & Gerstle, J. T. (2007). Necrotizing enterocolitis: review of state-of-the-art imaging findings with pathologic correlation. *Radiographics*, *27*(2), 285-305.

[6] Bell, M. J., Ternberg, J. L., Feigin, R. D., Keating, J. P., Marshall, R., Barton, L., & Brotherton, T. (1978). Neonatal necrotizing enterocolitis. Therapeutic decisions based upon clinical staging. *Annals of Surgery*, *187*(1), 1.

[7] Knobel, R. B., Guenther, B. D., & Rice, H. E. (2011). Thermoregulation and thermography in neonatal physiology and disease. *Biological Research for Nursing*, *13*(3), 274-282.

[8] Rice, H.E., Hollingsworth, C.L., Bradsher, E., Danko, M.E., Crosby, S.M., Goldberg, R.N., Tanaka, D.T., Knobel, R.B. (2010). Infrared thermal imaging (thermography) of the abdomen in extremely low birthweight infants. *Journal of Surgical Radiology*, *1*(2), 61-122.

[9] Nur, R., & Frize, M. (2013, March). Image processing of infrared thermal images for the detection of necrotizing enterocolitis. In *Medical Imaging 2013: Image Processing* (Vol. 8669, p. 86692M). International Society for Optics and Photonics.

[10] Frize, M., Nur, R., Bariciak, E., & Herry, C. (2013). Infrared Imaging and Classification of Neonates with Necrotising Enterocolitis. In *World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China* (pp. 1309-1312). Springer, Berlin, Heidelberg.

[11] Ntonfo, G. K., Frize, M., & Bariciak, E. (2015, May). Detection of Necrotizing Enterocolitis in newborns using abdominal thermal signature analysis. In *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings,* pp. 36-39.

[12] Kliegman, R. M., & Fanaroff, A. A. (1984). Necrotizing enterocolitis. *New England Journal of Medicine*, *310*(17), 1093-1103.

[13] Oh, S., Young, C., Gravenstein, N., Islam, S., & Neu, J. (2010). Monitoring technologies in the neonatal intensive care unit: implications for the detection of necrotizing enterocolitis. *Journal of Perinatology*, *30*; 701-708.

[14] Walsh, M. C., & Kliegman, R. M. (1986). Necrotizing enterocolitis: treatment based on staging criteria. *Pediatric Clinics of North America*, *33*(1), 179-201.

[15] Gordon, P. V., Swanson, J. R., Attridge, J. T., & Clark, R. (2007). Emerging trends in acquired neonatal intestinal disease: is it time to abandon Bell's criteria?. *Journal of Perinatology*, *27*(11), 661.

[16] Muchantef, K., Epelman, M., Darge, K., Kirpalani, H., Laje, P., & Anupindi, S. A. (2013). Sonographic and radiographic imaging features of the neonate with necrotizing enterocolitis: correlating findings with outcomes. *Pediatric Radiology*, *43*(11), 1444-1452.

[17] Staryszak, J., Stopa, J., Kucharska-Miąsik, I., Osuchowska, M., Guz, W., & Błaż, W. (2015). Usefulness of ultrasound examinations in the diagnostics of necrotizing enterocolitis. *Polish Journal of Radiology*, *80*, 1.

[18] Hiorns, M. P. (2011). Gastrointestinal tract imaging in children: current techniques. *Pediatric Radiology*, *41*(1), 42-54.

[19] Shi, Y., Payeur, P., Frize, M., & Bariciak, E. (2019). Infrared Imaging Tools for Necrotizing Enterocolitis (NEC) Diagnosis Guided by RGB-D Sensing. *Canadian Medical and Biological Engineering Society Conference Proceedings*, *42*.

[20] Ring, E. F. J., & Ammer, K. (2012). Infrared thermal imaging in medicine. *Physiological Measurement*, *33*(3), R33.

[21] Diakides, M., Bronzino, J. D., & Peterson, D. R. (Eds.). (2012). *Medical Infrared Imaging: Principles and Practices*. CRC press.

[22] Jones, B. F. (1998). A reappraisal of the use of infrared thermal image analysis in medicine. *IEEE Transactions on Medical Imaging*, *17*(6), 1019-1027.

[23] Rowan-Robinson, M. (2013). *Night vision: Exploring the infrared universe*. Cambridge University Press.

[24] McDaniel, G. W., & Robinson, D. Z. (1962). Thermal imaging by means of the evaporograph. *Applied Optics*, *1*(3), 311-324.

[25] Ring, E. F. J. (2004). The historical development of thermal imaging in medicine. *Rheumatology*, *43*(6), 800-802.

[26] Ring, E. F. J. (2006). The historical development of thermometry and thermal imaging in medicine. *Journal of Medical Engineering & Technology*, *30*(4), 192-198.

[27] Herry, C. L., & Frize, M. (2004). Quantitative assessment of pain-related thermal dysfunction through clinical digital infrared thermal imaging. *Biomedical Engineering Online*, *3*(1), 19.

[28] Scales, N., Kerry, C., & Frize, M. (2004, September). Automated image segmentation for breast analysis using infrared images. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 1, pp. 1737-1740.

[29] Herry, C. L., Frize, M., & Bariciak, E. (2011). Assessment of abdominal skin temperature change in premature newborns with NEC compared to healthy controls. In *5th European Conference of the International Federation for Medical and Biological Engineering*, pp. 191-194. Springer, Berlin, Heidelberg.

[30] Sturm, P. (2005). Multi-view geometry for general camera models. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1, pp. 206-212.

[31] Weng, J., Cohen, P., & Herniou, M. (1992). Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10), 965-980.

[32] Duan, S., Zang, H., Xu, M., Zhang, X., Gong, Q., Tian, Y., Liang, E. & Liu, X. (2015). Camera self-calibration method based on two vanishing points. In *AOPC 2015: Image Processing and Analysis* (Vol. 9675, p. 96752M). International Society for Optics and Photonics.

[33] Feng, X. F., & Pan, D. F. (2018). A camera calibration method based on plane mirror and vanishing point constraint. *Optik*, *154*, 558-565.

[34] Aihua, C., Chenghui, G., & Bingwei, H. (2012). Camera calibration method based on orthogonal vanishing point pair. *Chinese Journal of Scientific Instrument*, 1.

[35] Yin, S., Ren, Y., Zhu, J., Yang, S., & Ye, S. (2013). A vision-based self-calibration method for robotic visual inspection systems. *Sensors*, *13*(12), 16565-16582.

[36] Hu, Z. Y., & Wu, F. C. (2002). A review on some active vision-based camera calibration techniques. *Chinese Journal of Computer*, *25*(11), 1149-1156.

[37] Diraco, G., Leone, A., & Siciliano, P. (2010, March). An active vision system for fall detection and posture recognition in elderly healthcare. In *Design, Automation & Test in Europe Conference & Exhibition* (pp. 1536-1541).

[38] Song, L., Wu, W., Guo, J., & Li, X. (2013, August). Survey on camera calibration technique. In *5th International Conference on Intelligent Human-Machine Systems and Cybernetics* (Vol. 2, pp. 389-392).

[39] Abdel-Aziz, Y. I., Karara, H. M., & Hauck, M. (2015). Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. *Photogrammetric Engineering & Remote Sensing*, *81*(2), 103-107.

[40] Shapiro, R. (1978). Direct linear transformation method for three-dimensional cinematography. *Research Quarterly. American Alliance for Health, Physical Education and Recreation*, *49*(2), 197-205.

[41] Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal on Robotics and Automation*, *3*(4), 323-344.

[42] Deng, L., Lu, G., Shao, Y., Fei, M., & Hu, H. (2016). A novel camera calibration technique based on differential evolution particle swarm optimization algorithm. *Neurocomputing*, *174*, 456-465.

[43] Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(11), 1330-1334.

[44] Zhang, Z., Faugeras, O., & Deriche, R. (1995, April). Calibrating a binocular stereo through projective reconstruction using both a calibration object and the environment. In *Europe-China Workshop on Geometrical Modelling and Invariance for Computer Vision* (pp. 253-260).

[45] Lu, P., Liu, Q., & Guo, J. (2016). Camera calibration implementation based on Zhang Zhengyou plane method. In *Proceedings of the 2015 Chinese Intelligent Systems Conference* (pp. 29-40). Springer, Berlin, Heidelberg.

[46] Zhao, Y. H., & Bai, J. M. (2018, December). A novel camera calibration method without distortion model. In *Optical Precision Manufacturing, Testing, and Applications* (Vol. 10847, p. 1084709). International Society for Optics and Photonics.

[47] Wu, H., Chen, Q., & Wada, T. (2004, August). Visual line estimation from a single image of two eyes. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* (Vol. 3, pp. 290-293).

[48] Sonka, M., Hlavac, V., & Boyle, R. (2014). Image Processing, Analysis, and Machine Vision. *Nelson Education*, 178-254.

[49] Sahoo, P. K., Soltani, S. A. K. C., & Wong, A. K. (1988). A survey of thresholding techniques. *Computer Vision, Graphics, and Image processing*, *41*(2), 233-260.

[50] Wang W., & Wang Q. (2007). Two-dimensional threshold segmentation method based on image edge information. *Journal of Image and Graphics, 12*(1), 78-81.

[51] Cuevas, E., & Sossa, H. (2013). A comparison of nature inspired algorithms for multi-threshold image segmentation. *Expert Systems with Applications*, *40*(4), 1213-1219.

[52] Brink, A. D. (1992). Thresholding of digital images using two-dimensional entropies. *Pattern Recognition*, *25*(8), 803-808.

[53] Chuang, K. S., Tzeng, H. L., Chen, S., Wu, J., & Chen, T. J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics*, *30*(1), 9-15.

[54] Haldar, N. A. H., Khan, F. A., Ali, A., & Abbas, H. (2017). Arrhythmia classification using Mahalanobis distance based improved Fuzzy C-Means clustering for mobile health monitoring systems. *Neurocomputing*, *220*, 221-235.

[55] Tan, J., & Pan, B. (2011). Algorithm of human thermal image segmentation under complicated background. *Journal of Computer Engineering and Applications*, *47*(14), 13-16.

[56] Ma, Y, Chang, Q, & Hu, M. (2017). Research on Infrared Human Target Detection Algorithm in Complex Background. *Infrared Technology*, 1038-1044.

[57] Stauffer, C., & Grimson, W. E. L. (1999, June). Adaptive background mixture models for real-time tracking. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* (Vol. 2, pp. 246-252).

[58] Stauffer, C., & Grimson, W. E. L. (2000). Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 747-757.

[59] Varcheie, P. D. Z., Sills-Lavoie, M., & Bilodeau, G. A. (2010). A multiscale region-based motion detection and background subtraction algorithm. *Sensors*, *10*(2), 1041-1061.

[60] Elgammal, A., Harwood, D., & Davis, L. (2000, June). Non-parametric model for background subtraction. In *European Conference on Computer Vision* (pp. 751-767). Springer, Berlin, Heidelberg.

[61] Srivastava, A., Lee, A. B., Simoncelli, E. P., & Zhu, S. C. (2003). On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, *18*(1), 17-33.

[62] Teaney, B. P., Du Bosq, T. W., Reynolds, J. P., Thompson, R., Aghera, S., Moyer, S. K., Flug, E., Espinola, R. & Hixson, J. (2012, May). Human target acquisition performance. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIII* (Vol. 8355, p. 835510). International Society for Optics and Photonics.

[63] Nava, A., Garrido, L., & Brena, R. F. (2014, November). Recognizing activities using a kinect skeleton tracking and hidden markov models. In *13th Mexican International Conference on Artificial Intelligence* (pp. 82-88).

[64] Sun, H., Li, T., Pu, B., Zhang, Q., Wang, H. (2013). Human Detection Algorithm in Complex Background. *Journal of Application of Computer Systems*, *122*(4), 136-139.

[65] Phung, S. L., Chai, D., & Bouzerdoum, A. (2001, November). Skin colour based face detection. In *The Seventh Australian and New Zealand Intelligent Information Systems Conference,* 171-176.

[66] Tan, K. W., & Stephen, I. D. (2013). Colour detection thresholds in faces and colour patches. *Perception*, *42*(7), 733-741.

[67] Gomez, G. (2002, August). On selecting colour components for skin detection. In *Object Recognition Supported by User Interaction for Service Robots* (Vol. 2, pp. 961-964).

[68] Milczarski, P., & Stawska, Z. (2014). Complex Colour Detection Methods Used in Skin Detection Systems. *Information Systems in Management*, *1*(3), 40-52.

[69] Kakumanu, P., Makrogiannis, S., & Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Pattern Recognition*, *40*(3), 1106-1122.

[70] Nitsuwat, S. (2007, November). Face recognition in video, a combination of eigenface and adaptive skin-color model. In *International Conference on Intelligent and Advanced Systems* (pp. 742-747).

[71] Bergasa, L. M., Mazo, M., Gardel, A., Sotelo, M. A., & Boquete, L. (2000). Unsupervised and adaptive Gaussian skin-color model. *Image and Vision Computing*, *18*(12), 987-1003.

[72] Ban, Y., Kim, S. K., Kim, S., Toh, K. A., & Lee, S. (2014). Face detection based on skin color likelihood. *Pattern Recognition*, *47*(4), 1573-1585.

[73] Shi, L., Tang, W., Wan T., Zhu Y., & Wu T. Algorithm of Human Body Contour Extraction in Complex Background. *Science Technology and Engineering*, *19* (09), 143-147.

[74] Chen, Z., & Lee, H. J. (1992). Knowledge-guided visual perception of 3-D human gait from a single image sequence. *IEEE transactions on Systems, Man, and Cybernetics*, *22*(2), 336-342.

[75] Leung, M. K., & Yang, Y. H. (1995). First sight: A human body outline labeling system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(4), 359-377.

[76] Rohr, K. (1994). Towards model-based recognition of human movements in image sequences. *CVGIP: Image understanding*, *59*(1), 94-115.

[77] Fan, B., & Wang, Z. (2006). A contour-based method for determining human joint positions. *Pattern Recognition and Artificial Intelligence*, *19* (1), 67-72.

[78] Junior, J. C. J., Jung, C. R., & Musse, S. R. (2012, September). Skeleton-based human segmentation in still images. In *2012 19th IEEE International Conference on Image Processing* (pp. 141-144).

[79] Ren, Y., Li, C., Liu, H., & Yang, C. (2019). Image segmentation of human body with standard pose based on adaptive slic. *Application of Computer System*, *28* (05), 104-111.

[80] Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486.

[81] Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: measures of central tendency. *Psychological Methods*, *8*(3), 254.

[82] D'Agostino, R. B. (2017). Tests for the normal distribution. In *Goodness-of-Fit-Techniques*, pp. 367-420. Routledge.

[83] Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(1), 183-189.

[84] Friedman, J. H., & Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, *100*(9), 881-890.

[85] Horn, P. S., Feng, L., Li, Y., & Pesce, A. J. (2001). Effect of outliers and non-healthy individuals on reference interval estimation. *Clinical Chemistry*, *47*(12), 2137-2145.

[86] Dawson, R. (2011). How significant is a boxplot outlier?. *Journal of Statistics Education*, *19*(2).

[87] Beichel, R., Bischof, H., Leberl, F., & Sonka, M. (2005). Robust active appearance models and their application to medical image analysis. *IEEE Transactions on Medical Imaging*, *24*(9), 1151-1169.

[88] Tach, D. 'Kinect is Always Listening on Xbox One, but Privacy is a "Top Priority" for Microsoft,' Polygon. May. 21 2013. Accessed on: Nov.30, 2019. [Online]. Available: http://www.polygon.com/2013/5/21/4353580/kinect-always-listening-on-xbox-one-privacy-is-a-top-priority.

[89] "All about the Xbox One Kinect sensor", Microsoft 2019. Accessed on: Nov.30, 2019. [Online]. Available: https://support.xbox.com/en-US/xbox-one/accessories/kinect-sensor-info#Setting-up-the-Kinect-sensor.

[90] Pagliari, D., & Pinto, L. (2015). Calibration of kinect for xbox one and comparison between the two generations of microsoft sensors. *Sensors*, *15*(11), 27569-27589.

[91] OpenNI, N. (2010). OpenNI User Guide. *OpenNI organization*, 646.

[92] Xiang, L., Echtler, F., Kerl, C., & Blake, J. (2016). libfreenect2: Release 0.2. *2016-04-28)[2018-02-22]*. [Online]. Available*: http://github.com/OpenKinect/libfreenect2*.

[93] "FLIR A320 Tempscreen", April 27, 2016. Accessed on: Nov. 30, 2019. [Online]. Available: https://support.flir.com/DsDownload/Assets/48201-1201-en-US.html#styler-id1.20.

[94] Li, S., Li, Q., Li, H., Hou, P., Cao, W., Wang, X. & Li, H. (2016). Kinect v2 based real-time accurate 3D reconstruction system. *Journal of Software, 27* (10), 2519-2529

[95] Satheesh, S., & Prasad, K. V. S. V. R. (2011). Medical image denoising using adaptive threshold based on contourlet transform. *arXiv preprint arXiv:1103.4907*.

[96] Mundhenk, T. N., Rivett, M. J., Liao, X., & Hall, E. L. (2000, October). Techniques for fisheye lens calibration using a minimal number of measurements. In *Intelligent Robots and Computer Vision XIX: Algorithms, Techniques, and Active Vision* (Vol. 4197, pp. 181-190). International Society for Optics and Photonics.

[97] Burrus, N. (2014， May 26). Kinect Calibration. Nicolas Burrus homepage website. Retrieved January 13, 2020. [Online]. Available: http://burrus.name/index.php/Research/KinectCalibration.

[98] Duda, R. O., & Hart, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, *15*(1), 11-15.

[99] Chernov, G., Chernov, V., & Flores, M. B. (2017). 3D dynamic thermography system for biomedical applications. In *Application of Infrared to Biomedical Sciences* (pp. 517-545). Springer, Singapore.

[100] Phung, S. L., Bouzerdoum, A., & Chai, D. (2002, September). A novel skin color model in ycbcr color space and its application to human face detection. In *Proceedings. International Conference on Image Processing* (Vol. 1, pp. 289-292).

[101] Bari, C., & Ghorude, T.N. (2016). Face Recognition Using Skin Colour Segmentation of YCbCr and RGB Color Models. *Intl Journal of Emerging Technology in Computer Science & Electronics* (IJERCSE), *23*(1), 976-1353.

[102] Tan, W. R., Chan, C. S., Yogarajah, P., & Condell, J. (2011). A fusion approach for efficient human skin detection. *IEEE Transactions on Industrial Informatics*, *8*(1), 138-147.

[103] Phung, S. L., Bouzerdoum, A., & Chai, D. (2005). Skin segmentation using color pixel classification: analysis and comparison. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (1), 148-154.

[104] Zhi-guo, L. V., Yan, L. I., & Han-gen, H. E. (2008). Method of Poser Model-based 3D Human Body Modeling. *Journal of Computer Engineering*, 13.

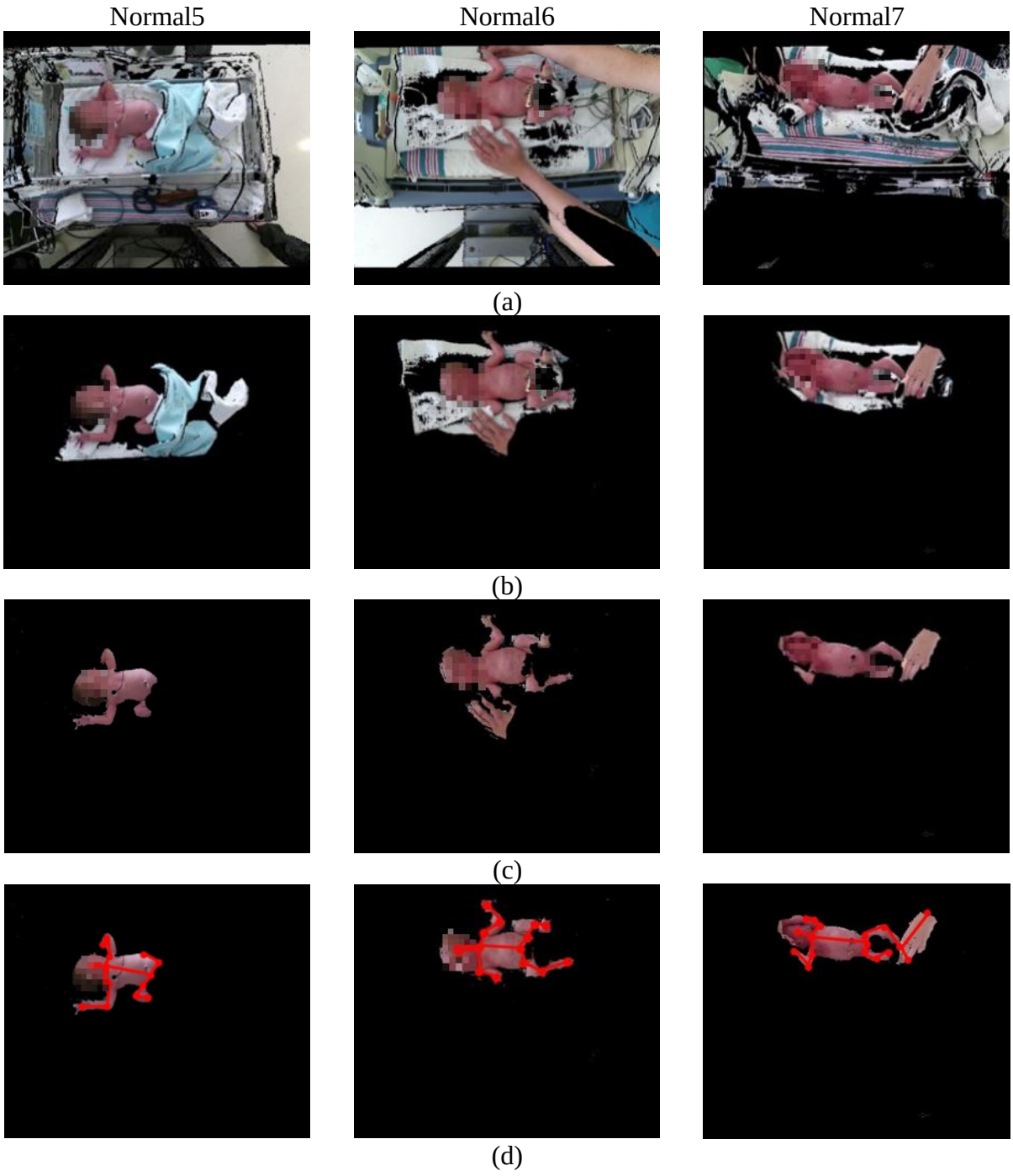
[105] "Summary of template matching method based on Halcon". Aug 1, 2018. Access on: Jan 22, 2020. [Online]. Available: https://topic.alibabacloud.com/a/summary-of-template-matching-method-based-on-halcon_8_8_10275122.html

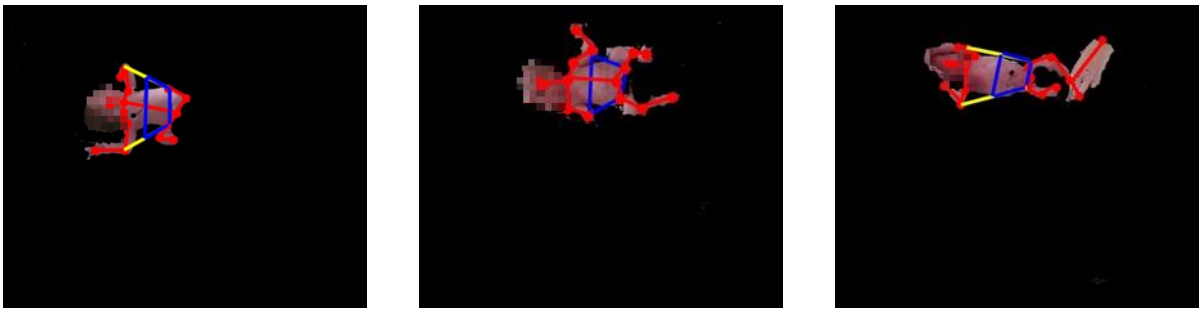
[106] Hesse, N., Stachowiak, G., Breuer, T., & Arens, M. (2015). "Estimating Body Pose of Infants in Depth Images Using Random Ferns". *Proc. of the IEEE Intl Conf. on Computer Vision Workshops*, 35-43.

[107] MATLAB. (2017). version 9.3.0.713579 (R2017b). Natick, Massachusetts: The MathWorks Inc.

[108] Shi, Y., Payeur, P., Frize, M., & Bariciak, E. (2020). "Thermal and RGB-D Imaging for Necrotizing Enterocolitis Detection", *Proc. of the IEEE International Symposium on Medical Measurements & Applications*, 1-6.

Appendix A

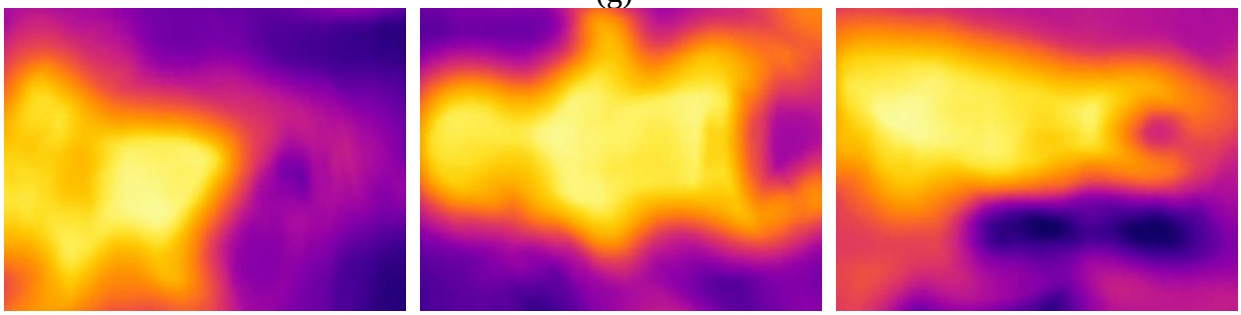




(e)

(f)

(g)



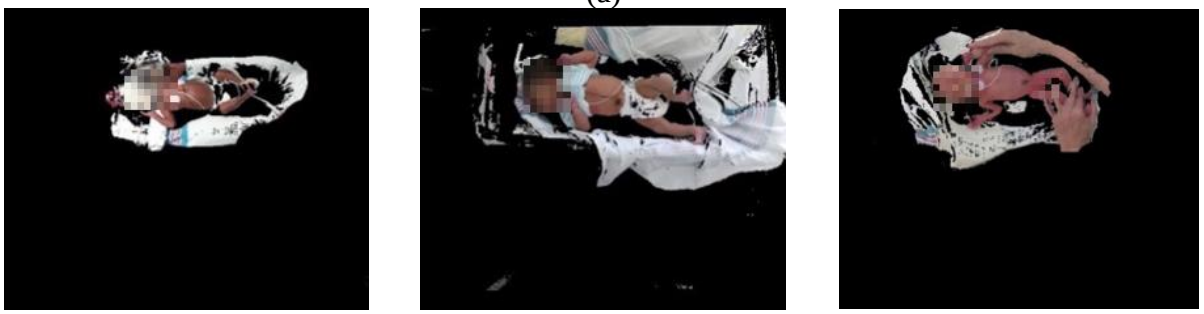
(h)

(i)

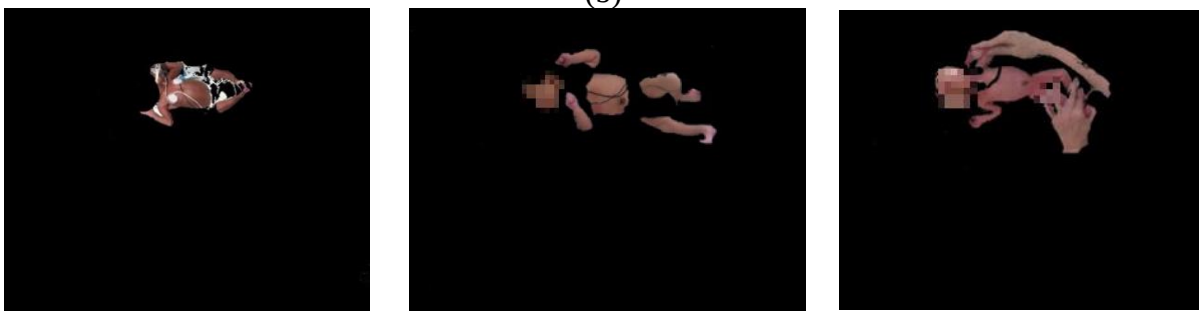
Figure A.1 Segmentation and region of interest extraction results on multispectral data collected on real normal babies (subjects 5-7) in a clinical environment, for three representative cases: (a) original RGB-D images; (b) background removal results; (c) after applying skin color filter; (d) recognized anatomical joints; (e) potential abdominal area restricted by blue quadrilaterals; (f) initial estimation of abdominal region; (g) refined extracted region of abdomen with skin filter segment; (h) original IR images; and (i) final region of interest extracted in IR images.

Normal8 Normal9 Normal10

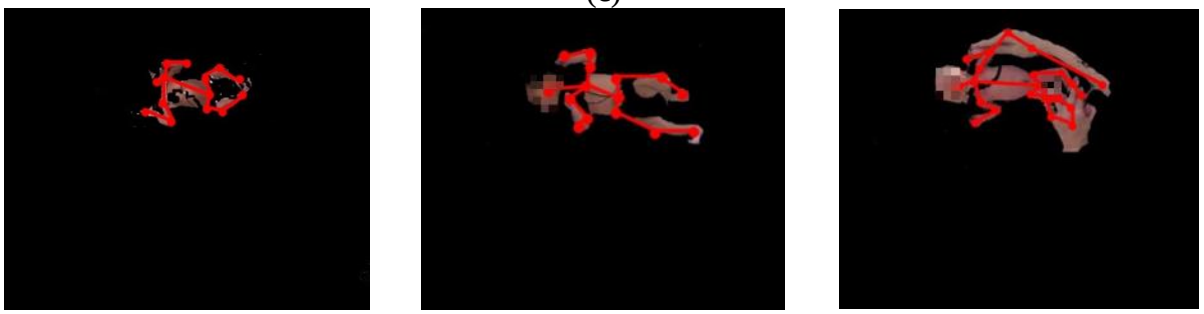
(a)



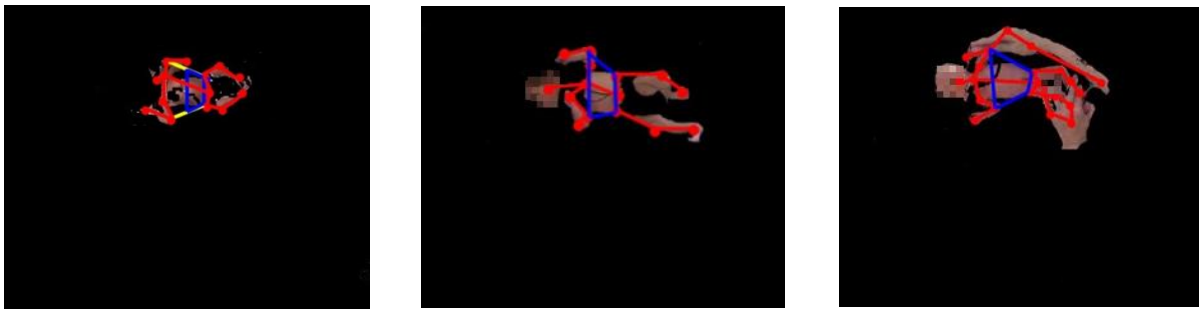
(b)



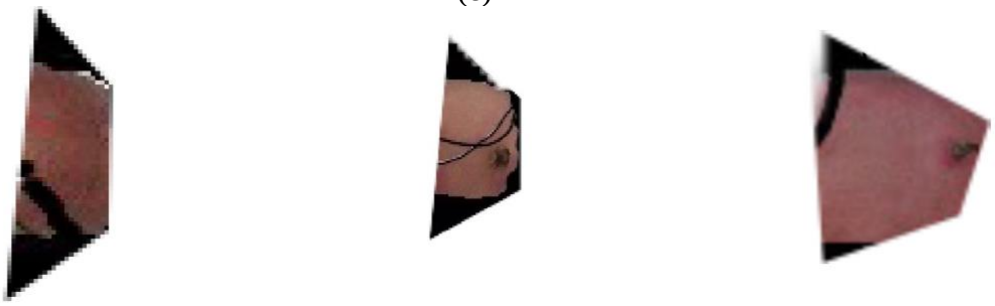
(c)



(d)

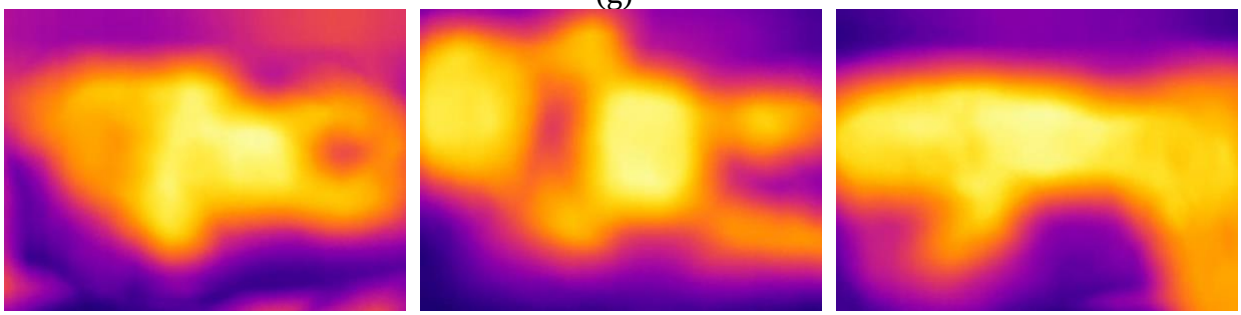


(e)



(f)

(g)



(h)

(i)

Figure A.2 Segmentation and region of interest extraction results on multispectral data collected on real normal babies (subjects 8-10) in a clinical environment, for three representative cases: (a) original RGB-D images; (b) background removal results; (c) after applying skin color filter; (d) recognized anatomical joints; (e) potential abdominal area restricted by blue quadrilaterals; (f) initial estimation of abdominal region; (g) refined extracted region of abdomen with skin filter segment; (h) original IR images; and (i) final region of interest extracted in IR images.

Normal11

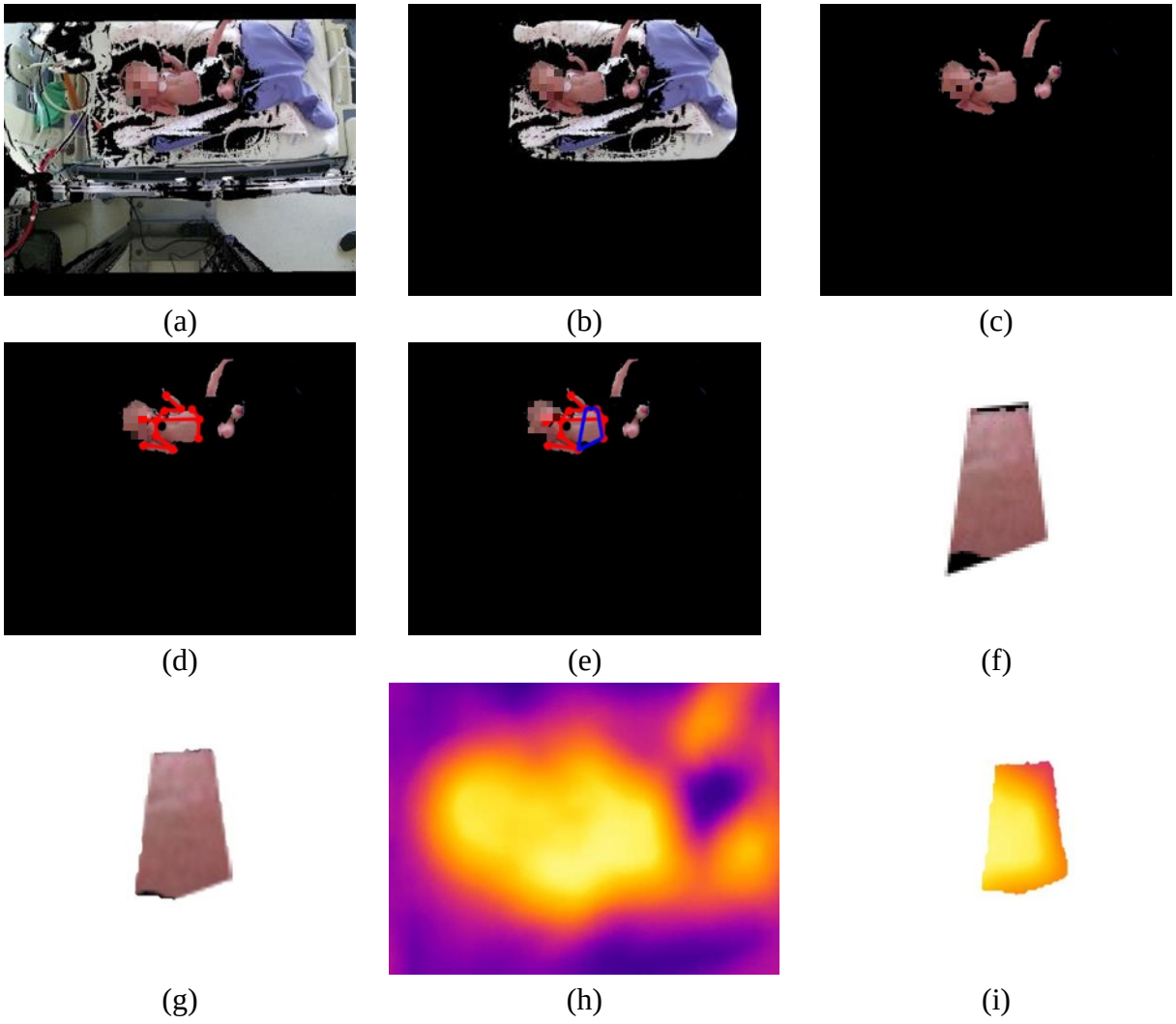


Figure A.3 Segmentation and region of interest extraction results on multispectral data collected on a real normal baby (11) in a clinical environment: (a) original RGB-D image; (b) background removal results; (c) after applying skin color filter; (d) recognized anatomical joints; (e) potential abdominal area restricted by blue quadrilaterals; (f) initial estimation of abdominal region; (g) refined extracted region of abdomen with skin filter segment; (h) original IR images; and (i) final region of interest extracted in IR images.