

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Automation of Region Specific Scanning for Real Time Medical Systems

Denis Kow Son Wong

A dissertation submitted to the Department of Electrical Engineering,
University of Cape Town, in fulfilment of the requirements
for the degree of Master of Science in Engineering.

Cape Town, May 2012

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the degree of Master of Science in Engineering in the University of Cape Town. It has not been submitted before for any degree or examination in any other university.

Signature of Author

Cape Town

15th July 2011

Acknowledgements

I would like to thank my supervisor, Dr Fred Nicolls, for his ongoing insight, support and honesty towards my thesis. I would also like to thank both Professor de Jager and Mayuresh Kulkarni for their advice and suggestions and to the LODOX research group for their weekly meetings, which helped me stay on track.

I would also like to acknowledge my girlfriend, Joanna Szewczyk, who remained by my side throughout my academic studies.

University of Cape Town

Abstract

X-rays have played a vital role in both the medical and security sectors. However, there is a limit to the amount of radiation a body can receive before it becomes a health risk. Modern low dose x-ray devices operate using a c-arm which moves across the entire human body. This research reduces the radiation applied to the human body by isolating the region that needs exposure. The medical scanner that this work is based on is still under development and therefore a prototype of the scanner is developed for running simulations. A camera is attached onto the prototype and used to point out the regions that are required to be scanned. This is both faster and more accurate than the traditional method of manually specifying the areas. Using a camera to locate a region also gives way for inevitable minor movements from the patient. An analysis is performed on the automation process as there are many variables such as speed, accuracy and searching thresholds that need to be catered for in the experiment. It is found that the correct region of interest can be located with the use of reliable feature points and that certain regions of the body are easier to locate than others. Currently, partial scans are done manually and this is a step forward towards automating the process completely.

Contents

Declaration	i
Acknowledgements	ii
Abstract	iii
List of Figures	viii
List of Tables	xi
1 Introduction	1
1.1 Lodox	1
1.1.1 Background	1
1.1.2 Statscan	2
1.1.3 Versasean	4
1.2 Objectives	6
1.3 Overview of Research	7
1.3.1 Problem	7
1.3.2 Proposed Solution	7
1.3.3 Approach	8
1.4 Outline of report	13

CONTENTS

2	Literature Review	14
2.1	Template Matching	14
2.1.1	Cross Correlation	15
2.1.2	Normalized Cross Correlation	16
2.1.3	Other Methods used in Template Matching	16
2.2	Feature-Based Matching	17
2.2.1	Edges	17
2.2.2	Corners	19
2.2.3	Scale Invariant Feature Transform	20
2.3	Summary	22
3	Proposed Versascan Setup	23
3.1	Versascan	23
3.2	Camera	25
3.2.1	Objective	25
3.2.2	Configuration	26
3.2.3	Model	26
3.2.4	Calibration	27
3.2.5	Field of View	27
4	Radiographer Workflow	30
4.1	Patient in Required Position	30
4.2	Initial Video Capture	31
4.3	Marking the Region of Interest	31
4.4	Second Video Capture	31
4.5	Comparison	32

CONTENTS

5	Workflow Processes	33
5.1	Reference Image	33
5.1.1	Capturing Video Data	34
5.1.2	Stitching Video Data	34
5.1.3	Video Data Loss	37
5.2	Finding Feature Points	38
5.3	Determine Reliable Points	39
5.4	Locating the Region of Interest	40
5.4.1	Searching Ahead	40
5.4.2	Online Searching Method	40
6	Experiments, Results and Findings	45
6.1	Ground Truth	45
6.2	Thresholds	46
6.2.1	Number of Matches	48
6.2.2	Searching Threshold	49
6.2.3	Point Reliability Threshold	50
6.2.4	Summary	52
6.3	Performance of Different Body Regions	53
6.4	Other Findings	58
6.4.1	Horizontal versus Vertical Accuracy	58
6.4.2	Failed Tests	60
7	Conclusions	61
8	Further Research	63
8.1	Searching Algorithms	63

CONTENTS

8.2	Camera	63
8.3	3-Dimensional Reference Image	63
A	Video Datasets	64
B	Experimentation Results	65
C	Code	66
	Bibliography	67

University of Cape Town

List of Figures

1.1	X-ray image of diamond ring	2
1.2	Statscan medical imaging device.	3
1.3	Important components of the Statscan.	3
1.4	Examples of x-ray images captured from Statscan.	5
1.5	Concept drawing of the forthcoming Versasecan medical imaging device.	6
1.6	Flowchart of performing a partial scan automatically.	8
1.7	Garage motor track used to imitate c-arm's movement.	9
1.8	Examples of reference images taken from different scanning systems. (a) Captured from c-arm on Statscan. (b) Captured from garage motor track.	10
2.1	Matching process for template matching.	15
2.2	Array of pixel intensity values shown as a function. (a) Function $f(t)$. (b) Derivative $f'(t)$. (c) Second derivative $f''(t)$	18
2.3	Results of using the Harris Corner Detector.	20
2.4	Example of using SIFT matching.	21
2.5	Converting image gradients and orientation to keypoint descriptors. (a) Image gradients and orientation. (b) Image descriptors.	22
3.1	Versasecan environment mimic setup. (a) Versasecan booth. (b) Versasecan c-arm.	24
3.2	Distance between camera and x-ray tube.	27
3.3	Field of view with focal length of 2.8mm	28

LIST OF FIGURES

3.4	Field of view of camera attached to c-arm where points a, b and c show the camera's perpendicular view, the physical offset and the maximum 'look-ahead' distance respectively.	29
4.1	Radiographer workflow to perform a partial scan.	30
4.2	Examples of reference images with different regions of the body selected. (a) Abdomen. (b) Chest. (c) Head.	32
5.1	Detailed workflow process to perform partial scan.	33
5.2	Examples of reference images. (a) and (b) Anteroposterior position. (c) and (d) Lateral position.	35
5.3	Important parts of the reference image.	36
5.4	Process of image stitching from video dataset where r represents the rows considered for stitching and d_r is the distance between each frame captured.	37
5.5	Obtaining reliable feature points from an abdomen region. (a) Reference image with feature points. (b) Feature points only within region of interest. (c) Reliable points only within region of interest.	39
5.6	Illustration of numbering reliable points and searching method.	42
5.7	Result of online search after finding two corresponding reliable points. The yellow and green lines indicate the position of the c-arm and the camera's viewpoint respectively.	43
5.8	Result of online search indicating the location of the marked region on scanned image. The yellow and green lines indicate the position of the c-arm and the camera's viewpoint respectively. The red box is the estimated location of the marked region.	44
6.1	Entire scanned image with an estimated location of the region of interest and ground truth indicated by red and green respectively.	47
6.2	Results of tests passed with varying number of matches required.	48
6.3	Results of passed tests that are within 2% and 3% accuracy when varying the number of matches required.	49
6.4	Results of tests passed with varying the search threshold.	50

LIST OF FIGURES

6.5	Results of passed tests within 2% and 3% accuracy when varying the search threshold.	51
6.6	Reference image after abdomen marked as region of interest and reliable points determined. Point reliability threshold set to (a) 10, (b) 20 and (c) 50.	53
6.7	Results of tests passed with varying the point reliability threshold.	54
6.8	Results of passed tests within 2% and 3% accuracy when varying the point reliability threshold.	55
6.9	Results of tests passed for each body region.	57
6.10	Results of each body region and accuracies within 2% and 3% shown as red and blue respectively.	57
6.11	Results of the experiment using all and only specific regions where the blue, red and green indicates the tests passed and accuracies within 2% and 3% respectively.	58
6.12	Histogram of horizontal and vertical pixel distances between ground truth and estimated marked region shown as red and blue respectively.	59
6.13	Results of passed tests within their respective accuracies where red and blue show the horizontal and vertical directions respectively.	59

List of Tables

3.1	Specifications of the medical scanner.	25
3.2	Selected specifications of the imaging device.	26
4.1	X-ray scanning routine views for different body regions. Regions marked with a * are not considered for the experiments.	31
6.1	Table of results for locating the region of interest.	56

University of Cape Town

Chapter 1

Introduction

Digital image processing has become a field of growing interest, especially in industries such as the medical sector to aid in the analysis of the human body. Computer algorithms that perform image processing on digital images [32] can be used to aid the analysis of the human body. Image processing can be in the form of analysis or manipulation of digital images.

This report is based on one of the medical x-ray scanner systems from Lodox Systems, so a brief discussion of their scanners is made. The aim of this research is to see whether it is possible to perform a partial scan of a moving human body using image processing techniques such as feature-based matching and template-based matching.

1.1 Lodox

1.1.1 Background

Lodox Systems (Pty) Ltd. is a South African company that is responsible for worldwide sales, distribution and customer service support of its medical imaging products. Lodox Systems designed and developed its medical x-ray scanners which originated from the Scannex [5], an x-ray security scanner developed by De Beers to prevent diamond theft within the mining industry [1].

Diamonds are very noticeable on an x-ray image [11] as shown in figure 1.1, which shows an example of a diamond ring. Other uses of the x-ray security scanner include the detection of firearms and other dangerous weapons to reduce security risks. X-ray machines are also found in hospitals to analyze a patient's bone structure, in areas like the arm, chest, head and leg.

The Lodox scanners are unique in that they can produce a full body x-ray image after a single 13

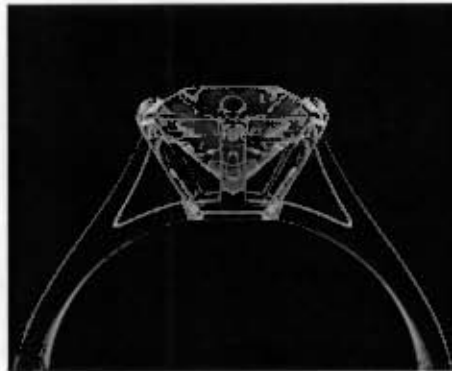


Figure 1.1: X-ray image of diamond ring

second scan. Lodox Systems currently has a medical scanner in the market, called the Statscan, and is in the process of developing their latest medical device, the Versascan. Considering that this report is based on the Lodox scanners, a detailed discussion is made on both the Statscan and Versascan.

1.1.2 Statscan

Figure 1.2 shows an example of the Statscan, which is a full body digital x-ray scanner developed and manufactured by Lodox. This particular Statscan scanner is found in Red Cross hospital in Cape Town. The Statscan's objective is to generate high quality, full body x-ray images of trauma patients with injuries in emergency rooms [36].

There are four main advantages in using the Statscan to obtain an x-ray image:

1. The Statscan takes a maximum of 13 seconds, to travel 2m, to do a full body scan.
2. With x-ray dosage being a sensitive matter, the Statscan only uses approximately 25% of the x-ray dose of a single conventional chest x-ray to produce a full body image.
3. Lodox scanners have minimal x-ray scatter. Therefore, unlike with conventional x-ray equipment, there is no need for personnel to operate the machine from a distance protected by lead screens and so the Statscan can be placed in any room.
4. Operating the Statscan is simple and can be done by trauma personnel, therefore no special training is needed to use the machine.

Figure 1.3 shows an image of the Statscan where the relevant components are labelled. The local positioning console is an important component: it is used to maneuver the c-arm, whether it be moving it across the platform, along the z-axis, or rotating it about the z-axis.



Figure 1.2: Statiscan medical imaging device.



Figure 1.3: Important components of the Statiscan.

The c-arm, x-ray source side, contains two important components for this project, a laser and the beam width limiter. The laser shines directly onto the platform, as a thick line, to indicate the exposure area above the patient. The beam width limiter is a collimator used to increase or decrease the width of the exposed area.

A limitation with the Statscan is seen when wanting to perform a partial scan of the human body. Lodox scanners are designed to perform full body scans but partial scans are not impossible. If a partial scan is required, the start and end position of the c-arm would have to be specified and the width of the collimators would have to be defined.

In practice, in order to perform a partial scan, the operator uses the local positioning console to move the c-arm to the starting point, using the laser as a guide. The beam width limiter is adjusted as required, using the length of the laser lines as a guide. Adjusting the beam width limiter and the position of the c-arm is done with the local positioning console.

Once the operator is satisfied with the start position, the positioning console is used to send a signal to the workstation of the starting point. The stopping point is specified in a similar manner. The workstation now has the information required to perform a partial scan.

The operator can adjust the starting and stopping points on the workstation if necessary. Operating the x-ray source to expose the person on the platform can only be done from the workstation and not with the local positioning console.

Examples of x-ray images of different body regions taken with the Statscan is shown in figure 1.4.

1.1.3 Versascan

The Versascan is a medical x-ray scanner currently being developed by Lodox. A concept drawing of the forthcoming Versascan is shown in figure 1.5. It is designed to be a multi-purpose, self-contained and transportable digital radiography system for general and orthopedic radiography.

Both medical scanners, Statscan and Versascan, are designed by Lodox so that the advantages of the Statscan apply to the Versascan. The main difference between the two is that the Versascan is a vertically-orientated scanner as opposed to the Statscan's horizontal orientation, so the patient is vertically positioned in a booth instead of lying on a platform. The Versascan's orientation therefore gives it the capability to capture x-ray images of patients in more varied positions and poses.

The Versascan can perform partial scans of the human body, but at this stage it would have to be done manually, like with the Statscan. Performing partial scans manually leads to human errors when having to specify the starting and stopping points, using the laser as a guide, and requires that

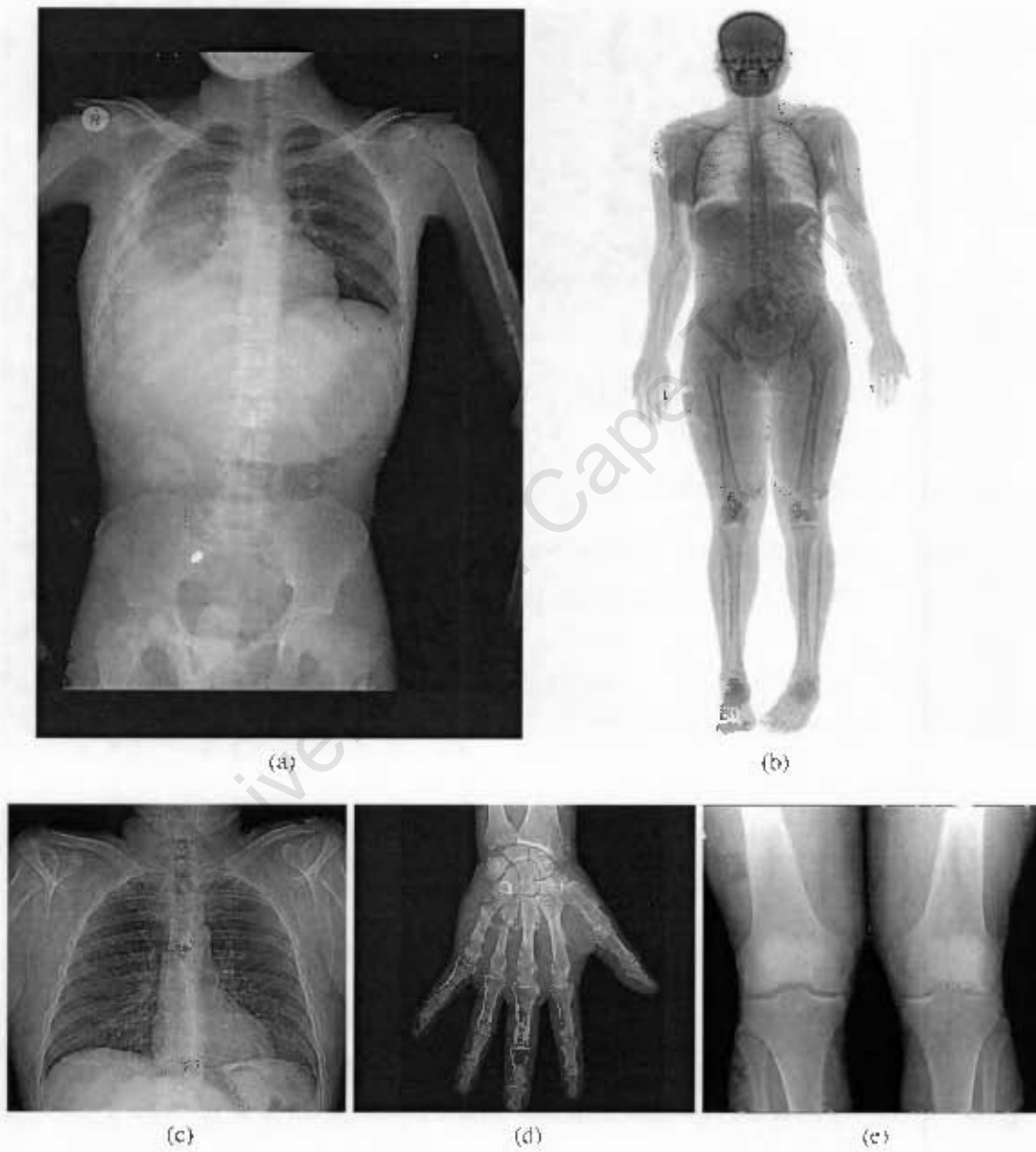


Figure 1.4: Examples of x-ray images captured from Statscan.



Figure 1.5: Concept drawing of the forthcoming Versascan medical imaging device.

the patient stand very still during the scan.

This report aims to provide the information needed in order to perform a partial scan of the human body. This proposed setup uses a camera on the c-arm to capture a full body image of the patient during a pre-scan where x-rays are disabled. The operator uses the full body image to mark the region that requires scanning. Using a camera to locate the region of interest, in real-time, provides an allowance for the patient to move slightly.

1.2 Objectives

There is an incentive to lower the x-ray radiation dose because there is a limit to how much radiation a body can absorb before it becomes a health risk. This project aims to reduce the total radiation exposed to the body by isolating the region of interest so that only the necessary region is scanned.

Therefore the main aim of this project is to find whether it is possible to perform a partial scan, on the Versascan, with the aid of a camera. It is proposed that the camera be attached to the c-arm, above the x-ray source.

The approach taken is to begin with video data being captured with the camera while the c-arm performs its first pass where it moves to the other end of the booth. The video data is used to produce a single image of the full body, referred to as the reference image, which the operator uses to mark the region of interest that is to be scanned. The marked region of interest provides two

essential pieces of information to perform a partial scan: the width of the beam limiter collimator and the vertical starting and stopping points for the x-ray exposure.

After a region is marked, the actual x-ray scan occurs as the c-arm performs its second pass and returns to its original position. As this is happening, the camera is capturing information that is used to locate the marked region when the region of interest is detected. Once the location of the marked region is found, an analysis is performed on accuracy, robustness and speed of the searching method.

The objectives of this project can be broken up into five major parts:

1. Create an apparatus to mimic the action of the Versascan, to obtain data. This step is necessary as the Versascan has not yet been manufactured.
2. Control the c-arm, with a camera attached, and move to one end of the booth and obtain a full body image, stitched from video data (pass 1). The full body image, referred to as the reference image, is used by the operator to mark the region of interest.
3. Return the c-arm to its original position and simultaneously locate the marked region for the x-ray exposure (pass 2). Both feature-based and template-based matching methods are explored for locating the marked region.
4. Once the region of interest is successfully located, investigate and illustrate the factors that affect the results.
5. Determine whether a particular region is easier to locate than others.

1.3 Overview of Research

1.3.1 Problem

The main problem, and the reason for this research, is that it is difficult to perform partial scans using the Versascan. This research looks at the possibility of using a camera and image processing techniques to identify the region of interest to perform a partial scan automatically.

1.3.2 Proposed Solution

A detailed overview is given of the Versascan as this research is dependent on the physical parameters of the device. A proposal is made for a single modification to the design of the Versascan prototype to enable it to perform partial scans and locate the region of interest automatically.

There are two important practical aspects in this research, namely the workflow for the radiographer and the processes that occur to locate the marked region. Figure 1.6 is a flowchart containing both aspects for performing a partial scan automatically. In figure 1.6, the elliptical elements show the steps taken by the radiographer and the rectangular elements indicate what processes are performed, at each step, in order to locate the region of interest automatically.

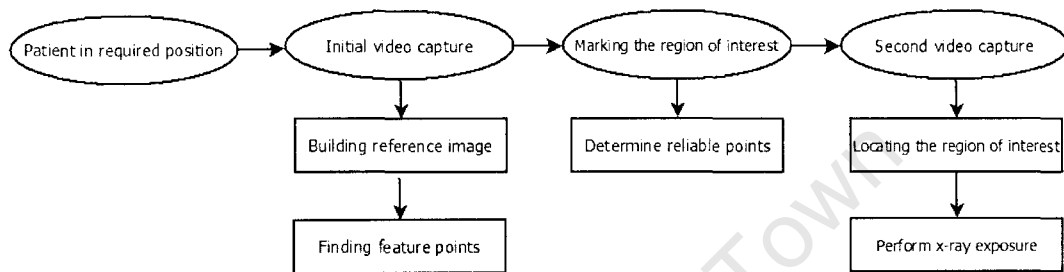


Figure 1.6: Flowchart of performing a partial scan automatically.

The workflow for the radiographer is important if the proposed modification is to be accepted for the Versascan design. The reason is because the radiographer needs to know what the procedures are in order to perform a partial scan and locate the region of interest automatically.

The remainder of this section uses the flowchart shown in figure 1.6 to give an overview of how this report proposes to perform a partial scan and locate the region of interest automatically. More detail of the radiographer’s procedures and the processes that follow are discussed in chapters 4 and 5.

1.3.3 Approach

The Versascan is in its design phase and there is currently no prototype of the physical medical scanner. The apparatus used to capture data needs to mimic that of the Versascan as closely as possible.

A simple motorised garage door opener, vertically orientated, is used to capture simulated visual data of patients. An image of the vertically orientated garage motor track with the camera attached is shown in figure 1.7.

A comparison is made of the video data captured between the movement of the garage motor track and the c-arm on the Statscan. Figure 1.8 shows examples of already stitched reference images from the two different scanning systems. It is seen that the video data is comparable with the differences being the patient’s orientation. Therefore data captured from the garage motor track is assumed to

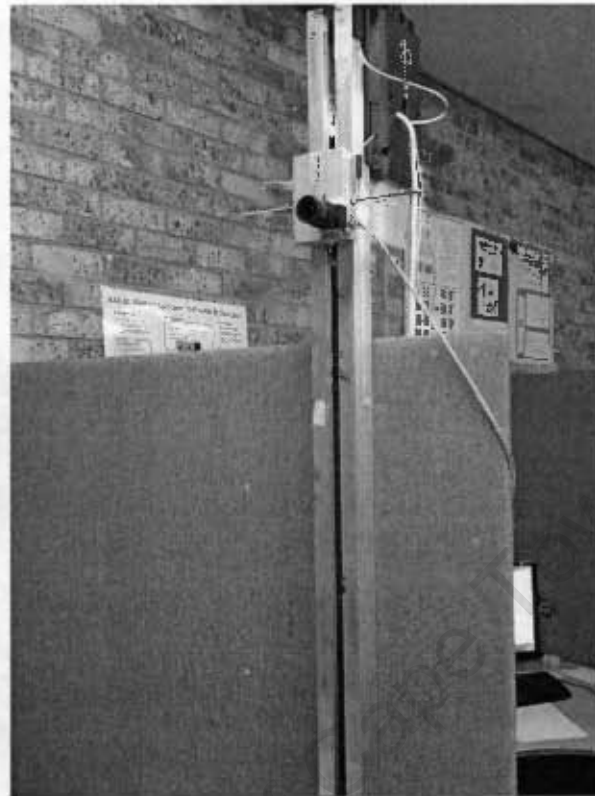


Figure 1.7: Garage motor track used to imitate c-arm's movement.

be similar to that of the Versascan. A detailed discussion is presented on the method used to stitch these reference images, but what is important is that the images are similar.

For the remainder of the report, the garage motor track is referred to as the c-arm unless otherwise specified. The reason why the c-arm is being mentioned is because this report is based on how the proposed system would work as if implemented on the Versascan. This section continues by giving a brief overview of the entire workflow process as shown in figure 1.6.

Patient in required position

The radiographer begins by placing the patient in the required position where the region of interest is visible to the x-ray source. Random poses and regions are not used as it is important to cater for realistic cases when performing a partial scan on the Versascan. Therefore, standard poses, for instance the anteroposterior and lateral positions, are considered when testing to do partial scans with common regions, such as the abdomen, arm and chest.



(a)



(b)

Figure 1.8: Examples of reference images taken from different scanning systems. (a) Captured from c-arm on Statiscan. (b) Captured from garage motor track.

Initial video capture

Once the patient is in the required position, the radiographer initiates a vertical scan of the c-arm. The initial video capture of the patient occurs on the c-arm's first pass. The video data is stitched together to produce a single full body image, referred to as the reference image. A corner detector is then used to find feature points on the reference image and it is these points that will be used for subsequent detection of the region of interest. A broader discussion on the stitching method is discussed in later chapters.

Marking of region of interest

After the feature points are found, the radiographer receives the reference image. The feature points on the reference image are not visible to the radiographer to avoid confusion. Using standard regions as a guide, the region of interest is marked on the reference image.

A test is performed on the feature points found within the marked region to determine how reliable they are. Determining the reliability of feature points is necessary to eliminate corners with similar surroundings. Reliability is determined by using a template-based matching method on each feature point as its centre coordinate and calculating if there is another position nearby that closely matches it.

Increasing the reliability of the feature points is seen to increase the accuracy of the results, but consequently decreases the number of feature points available. Thus, there is a trade-off between the reliability and the number of feature points available. This trade-off is evaluated by varying the reliability of the feature points and is shown later in this report.

Second video capture

Once the radiographer has marked the region of interest, there is enough information acquired to perform a partial scan automatically. The information used to perform a partial scan are the reliable feature points found, the marked region and the reference image. The radiographer's final task is control the c-arm to perform its second pass.

As the c-arm performs its second pass, the search for the region of interest begins. The camera captures its second set of video data of the patient, referred to as the scanned image, but with one small difference: the stitching method is adjusted so that the camera is effectively looking ahead to where the c-arm is approaching. This is required because it is necessary to detect the region of interest before the c-arm reaches it to signal the x-ray tube to turn on and off at the appropriate

positions.

Under the assumption that the patient has not moved too much, the marked region is used as a reference of location. The search for the marked region only begins fifty frames, approximately 65mm, before the expected region of interest is reached and the search continues until the region of interest is found. A more detailed discussion of converting pixels to distances is made in section 5.1.3.

The searching method is composed of both feature-based and template-based matching methods. A small template around each reliable point is used in the search for the marked region. Each small template is used to search, using normalized cross correlation as a measure, around its expected position to find where it best fits. A threshold value is set to measure the quality of the best fit and if the expected best fit is below that threshold then the match is ignored. The matching threshold is varied in the experimental stage to see how it affects the results.

Once a set number of reliable matches has been found on the scanned image, the location of the estimated region of interest is calculated. The number of reliable points found is also varied in the experimental stage to see how it affects the result. Locating the corresponding reliable points on the scanned image is discussed in more detail in chapter 6.

An accuracy measure needs to be in place to determine how accurate the calculated region of interest is. A calculated ground truth is taken to compare against the region found. The ground truth is calculated in a similar manner when finding the expected region of interest. Instead of using a set number of reliable points, all the reliable points on the reference image are used to find a best fit on the scanned image. This is assumed where the ground truth is calculated to be and accuracy is measured as a comparison of the two regions found.

The accuracy is important to such an extent that a maximum error of 2% source to image detector distance (SID) is allowed for the Versascan to be approved by the Food and Drug Administration (FDA). A patient is never completely still and so tests are performed on bodies that have moved between scans.

The physical distance between the source to the image detector is approximately 1000mm. Therefore, in order to achieve an accuracy to within 2% SID, a maximum error of 20mm is permitted. For experimental purposes, results with accuracies within 2%, 3% and 5% are accepted as passed tests as they have successfully located the region of interest.

Failed experiments are cases where the region of interest is not found, found to be incorrect, or where the c-arm passes the region of interest without turning on the x-ray tube. A more detailed discussion on different failed experiments and reasons why they failed is made in section 6.4.2.

Datasets

For the experimental stage, two pairs of video data are required in order to perform the relevant tests, one for the reference image and one for the scanned image to search for the region of interest. The pairs of video data are of patients in standard poses with slight adjustments to their stances so as to mimic minor movements in a realistic situation.

Considering that this project is for a specific application, a large dataset is required in order to make a reliable recommendation. Seven people were selected to take part in this experiment where five video pairs were captured and six or eight regions, depending on the patient's pose, were selected per pair. Therefore, there are approximately two hundred and sixty tests which are used in the experiment. Video data of the seven people taking part in the experiment included two people in the lateral position and the remainder in the anteroposterior. The experiment is to indicate whether using a combination of feature-based and template-based matching methods, and the aid of a camera, can locate a region of interest from a moving c-arm.

1.4 Outline of report

Chapter 1 introduces the dissertation and proposes that Lodox Systems should include a camera in the new Versascan design to perform a partial scan automatically. Chapter 2 looks at relevant academic research from various papers, articles, websites and books in the image processing field. This chapter is helpful in solving the problem and investigating the advantages and disadvantages of some of the methods that could be used to aid the searching process.

The physical details of the Versascan setup and the implementation of the camera follows in Chapter 3, which describes the proposed modification for the Versascan.

Chapters 4 and 5 discuss the entire workflow for the radiographer and the processes to perform a partial scan automatically. All assumptions and experiments performed are evaluated in chapter 6 where results are shown and recommendations made.

Finally, chapters 7 and 8 present a summary of the entire project with its findings, followed by a discussion of ways in which this research could be taken further.

Chapter 2

Literature Review

This chapter describes the various methods that are available to aid in the search for the region of interest on the medical scanner. In order to achieve this, both an overview and a look into relevant literature of some of the different methods available are made. For the purpose of this project, two main categories of searching methods are discussed in detail, namely: feature-based and template-based matching methods.

Feature-based matching finds feature points, such as edges and corners, and uses them to locate objects. Template-based matching uses a portion of the original image to compare and locate on another image. Both methods have been used in various applications such as identifying or tracking objects.

2.1 Template Matching

Template matching is a technique commonly used for object recognition and stereo-matching. Template matching techniques compare portions of images against one another. Figure 2.1 shows the template matching process where the template, $t(x, y)$, on the input image moves to all possible positions, generally limited to a search window. Correlation values are calculated in the various positions that indicate how well the template matches the image. The matches are done on a pixel-to-pixel basis.

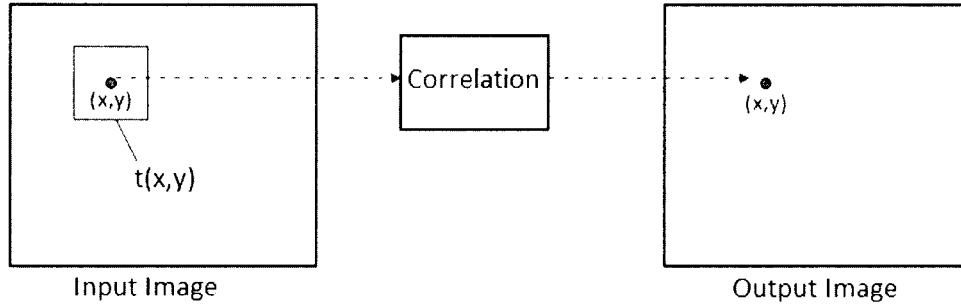


Figure 2.1: Matching process for template matching.

2.1.1 Cross Correlation

Correlation is a measure of the degree to which two variables agree, not necessarily in actual value but in general behaviour. A common way to calculate the position (u, v) of the object in the search window is to evaluate the cross correlation coefficient value, c , at each point (u, v) for function f and the template t . The positions (u, v) represent the shift of u in the x direction and by v in the y direction.

Cross correlation is motivated by the Euclidean distance, which is a measure of similarity and is shown as

$$d(u, v) = \sqrt{\sum_{x,y} [f(x, y) - t(x - u, y - v)]^2}. \quad (2.1)$$

Euclidean distance is only appropriate for data measured on the same scale as no adjustments are made for differences in scale. Equation 2.2 is obtained by expanding d as

$$d^2(u, v) = \sum_{x,y} [f^2(x, y) - 2f(x, y)t(x - u, y - v) + t^2(x - u, y - v)]. \quad (2.2)$$

In equation 2.2, if $f(x, y)$ and $t(x - u, y - v)$ are standardized, the sums are both equal to a constant value n . Therefore, $\sum f(x, y)t(x - u, y - v)$ is the only non-constant term just as it is in the reduced formula for the correlation coefficient:

$$c(u, v) = \frac{n \sum_{x,y} f(x, y)t(x - u, y - v) - (\sum_{x,y} f(x, y))(\sum_{x,y} t(x - u, y - v))}{\sqrt{(n \sum_{x,y} f^2(x, y) - (\sum_{x,y} f(x, y))^2)(n \sum_{x,y} t^2(x - u, y - v) - (\sum_{x,y} t(x - u, y - v))^2)}}. \quad (2.3)$$

Lewis states that there are a few disadvantages to using the cross correlation coefficient for a measure of similarity [21]. Some of the disadvantages mentioned are that the range of the correlation

coefficient value is dependent on the size of the feature and that it is not invariant to changes in scale and lighting conditions.

2.1.2 Normalized Cross Correlation

Lewis states that the difficulties with the cross correlation can be overcome by normalizing the image, with search window size of 168×86 , to unit length [21]. The normalized cross correlation is shown as

$$\gamma(u, v) = \frac{\sum_{x,y} [f(x, y) - \bar{f}_{u,v}] [t(x - u, y - v) - \bar{t}]}{\sqrt{(\sum_{x,y} (f(x, y) - \bar{f}_{u,v})^2) (\sum_{x,y} (t(x - u, y - v) - \bar{t})^2)}} \quad (2.4)$$

where \bar{t} is the mean of the feature and $\bar{f}_{u,v}$ is the mean of $f(x, y)$. Normalized cross correlation is a popular measure of similarity as its easy hardware implementation makes it useful for real-time applications.

Work has been done on increasing the performance of normalized cross correlation with the use of basis functions. Briechle and Hanebeck proposed using rectangular basis functions where the number of calculations depend linearly on the number of basis functions used [6]. The specific example used in [6] has an outcome that results in a computational reduction of 47 times using basis functions.

There have been some image matching methods performed based on normalized cross correlation [14, 41, 31]. However, these methods do not perform well as normalized cross correlation is not invariant to rotation. Zhao, et. al. propose a hybrid method, consisting of both feature points and templates, to improve the results of normalized cross correlation [13]. The hybrid method consists of using feature points on the two images to determine the rotation and scale changes according to the characteristic scale and dominant direction of the points. The invariant normalized cross correlation is then applied at the corresponding feature points.

The main difference between the works of [13] and [6] is that the one potentially eliminates the measure's variance and the other increases its performance by reducing its computation time.

2.1.3 Other Methods used in Template Matching

Another application for template matching is to not only classify an object but also to track it. Object tracking is usually categorized into two classes. One is where tracking takes place while the camera

is stationary and the other is when the camera is moving. The most important characteristic is that to make a real-time system, the image captured by the camera must be processed before the next frame is digitized.

Pal and Biswas propose an automated correlation based tracking approach using edge strength and Hausdroff Distance Transform (HDT) technique for tracking moving targets [37]. The approach produces a complete real-time video tracking system for both detecting and tracking moving targets from optical image sequences.

Other methods used for detecting objects can be seen in [19] and [29], with differences being that the object's shape is known. Cole, et. al describe how a 2D model can be used [19] and Gupta, et. al. show how detecting objects can be done with a 3D model [29].

2.2 Feature-Based Matching

The feature-based matching approach is the easiest method for finding image displacements. The method finds features in an image, such as edges and corners, and calculates the change in distances of the position of the feature points from the original image to another.

When video data is considered, the displacement is calculated from frame to frame. This is basically a two-step approach. Firstly, feature extraction is performed on two or more consecutive frames, to both reduce the amount of information to be processed and to obtain a higher level of understanding of the image scene. Secondly, these feature points are matched between frames to find any change in the positions of the points. Generally, changes in feature point positions between frames usually mean a movement of some object or background.

Feature-based matching is usually preferable when an image has strong features, such as sharp corners, in it. A feature-based approach is generally faster than a template-based approach because it does not consider the entire image but rather only the feature points found. This section investigates two different methods of feature-based matching, namely corners and edges as well as the Scale Invariant Feature Transform (SIFT).

2.2.1 Edges

Edges indicate boundaries in an image, which makes them important for image processing. Edges in an image usually appear as intensity changes in pixels situated next to each other. There are many different methods of edge detection but they can be grouped into two categories, gradient-based and

Laplacian-based. Mlsna and Rodriguez show that the difference between the two categories is that the gradient methods consider maximum values in the first derivative of an image and Laplacian methods look for zero crossings in the second derivative of an image [27].

Mathematically, a change in the intensity values of pixels in the 1D continuous case indicates an edge. Suppose figure 2.2a shows the intensity values of a certain line of pixels. Figure 2.2b shows the derivative of the intensity values which indicates an edge as the gradient decreases from the vertical axis. An edge can also be found by using the second derivative and finding the zero crossing because all maximum and minimum values will be zero, as shown in figure 2.2c.

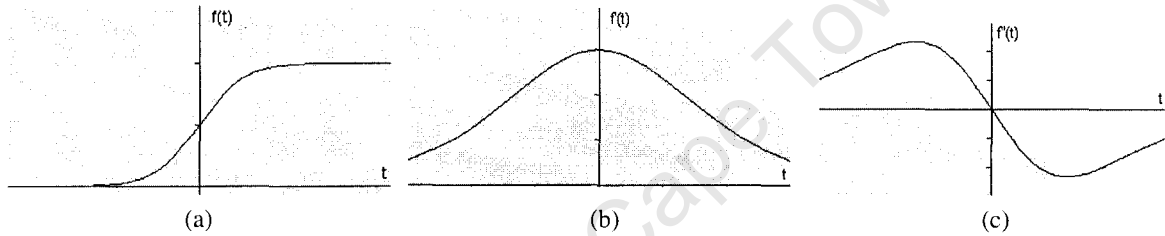


Figure 2.2: Array of pixel intensity values shown as a function. (a) Function $f(t)$. (b) Derivative $f'(t)$. (c) Second derivative $f''(t)$

It is shown in [9] that, in practice, a zero crossing filter is created by performing Gaussian smoothing followed by Laplacian filtering. The Laplacian-of-Gaussian (LoG), which is the convolution mask of the zero crossing operator, can be obtained from using various orders of linear filters and the rotational symmetry of Gaussian filter:

$$w_z(r) = N \left(\frac{r^2}{\sigma^2} - 1 \right) e^{-\frac{r^2}{2\sigma^2}}. \quad (2.5)$$

Equation 2.5 shows the Laplacian-to-Gaussian convolution mask where N is the normalisation constant, $r^2 = x^2 + y^2$ is the square distance from the centre of the mask, and σ is the scaling parameter of the mask. The Difference of Gaussian filters (DoG) is another implementation of the zero crossing filter and is more efficient but less accurate than the LoG.

A few examples of other edge detectors are Sobel, the Canny, the Local Threshold and Boolean Function Based edge detectors [2] and color edge detection using euclidean distance and vector angle [39]. Nadernejad, et. al. have performed a greater analysis of various edge detectors, including the ones previously mentioned [12].

2.2.2 Corners

Corners are the intersections of two edges of sufficiently different orientations. Therefore corners contain two dimensional features and can potentially represent object shapes. The ability to represent object shapes play important roles in matching and pattern recognition.

There are many different corner detectors that exist such as the Principal Curvature-Based Region (PCBR) detector [16] and the Harris operator [10]. Corner detectors have many applications in motion tracking, stereo matching and image database retrieval. Mokhtarian and Suomela modify the corner detector to make it more robust, based on the curvature scale-space (CSS) representation [28]. The quality of a corner detector is determined by its ability to detect the same corner in multiple images of the same scene but under different conditions, like lighting, translation, rotation and other transformations.

The Harris corner detector is a good method to use to detect corners as it provides good quality corners under varying rotation and illumination and may detect interest points other than corners. The Harris corner detector is popular for its simplicity and effectiveness and is based on the following matrix:

$$C = \begin{pmatrix} (dI/dx)^2 & (dI/dx)(dI/dy) \\ (dI/dx)(dI/dy) & (dI/dy)^2 \end{pmatrix} \quad (2.6)$$

In practice, the Harris corner detector uses a small window size and sums each of the derivatives over their given window region. Therefore, the change of intensity for the shift (u, v) is shown as

$$E(u, v) = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2 \quad (2.7)$$

where $w(x, y)$ is the window function and $I(x, y)$ is the calculated intensity. Some examples are provided in figure 2.3 to show the results of using the Harris corner detector [34, 33]. For the purpose of this thesis the Harris corner detector is considered due to its strong invariance to rotation, scale, illumination variation and image noise [7].

Finally there has been work done in combining the edge and the corner detector. Harris and Stephens propose combining the corner and edge detector based on the local auto-correlation function to obtain feature points for tracking algorithms [17]. Weijer, et. al. propose combining the two detectors by photometric quasi-invariants [18] and Ando by gradient covariance [4].

Parks and Gravel provide a detailed comparison of over 10 various corner detectors including ones

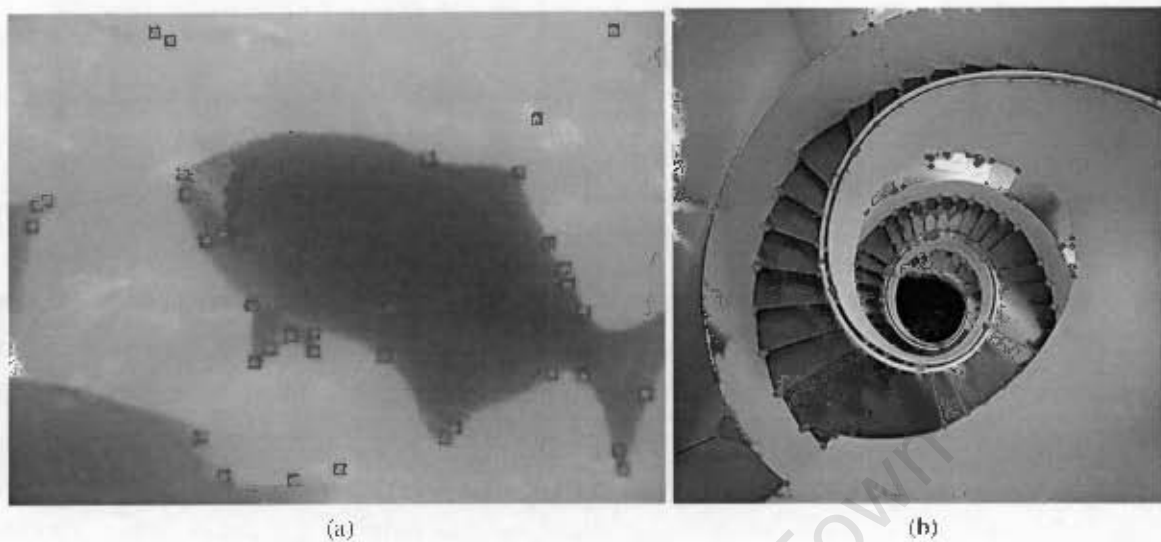


Figure 2.3: Results of using the Harris Corner Detector.

mentioned previously [30].

2.2.3 Scale Invariant Feature Transform

Lowe developed and published the algorithm called Scale Invariant Feature Transform (SIFT) to detect and describe local features in images [24]. The University of British Columbia has patented this algorithm but it is available to the public for research purposes only and there are papers available by Lowe that give a better understanding of the SIFT keypoint detector method [22, 23, 24]. An example of the SIFT keypoint detector is shown in figure 2.4.

The SIFT algorithm is robust because, as the name suggests, it is able to handle image transformations like scale, rotation and deformation. There are four steps that SIFT goes through to transform image data into scale invariant coordinates relative to local features [24].

The first step is the scale-space extrema detection. This first stage of computation searches over all scales and image locations. It is implemented efficiently using a Difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.

Unfortunately, the scale-space extrema detection produces too many keypoints, some of which are unstable. Therefore the second step is the keypoint localization, which takes each candidate location and fits a detailed model to determine location, scale and ratio of principal curvatures. Keypoints are selected based on the measures of their stability, so points that are sensitive to noise are rejected.

With fewer keypoints, the next step is the orientation assignment. One or more orientations are



Figure 2.4: Example of using SIFT matching.

assigned to each keypoint location based on local image gradient directions. This step is important in order to achieve invariance to rotation, but the disadvantage is that it limits the descriptors because it does not use all the information it has.

The previous steps have found keypoint locations at certain scales and assigned orientations to them which ensures invariance to image location, scale and rotation. The final step is to compute a descriptor vector for each keypoint so that it is partially invariant to other factors, such as changes in illumination.

The steps which produce keypoint locations at particular scales that have orientations assigned to them are shown in figure 2.5a. At each sample point, shown in figure 2.5a, a weight is applied by a Gaussian window shown by the circle. Figure 2.5b shows a 2×2 descriptor array computed from an 8×8 set of samples, where each descriptor was calculated with the length of each arrow corresponding to the sum of the gradient magnitudes in the 4×4 subregions.

Aly has shown that SIFT can be used to find feature points in a face to identify a person for surveillance and access control [3]. There are various other applications that use SIFT such as image stitching [40], video tracking [8] and 3D modeling [20, 38].

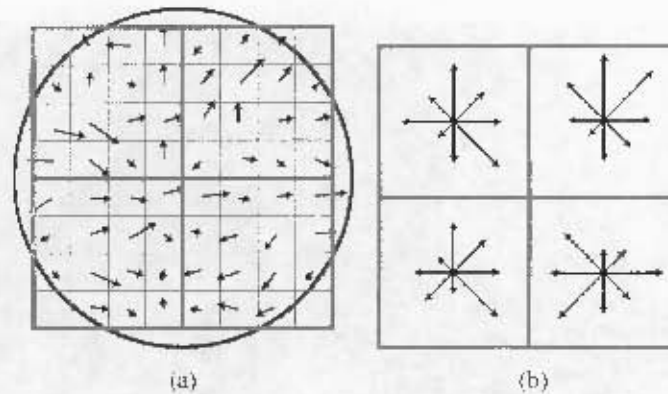


Figure 2.5: Converting image gradients and orientation to keypoint descriptors. (a) Image gradients and orientation. (b) Image descriptors.

2.3 Summary

The various academic research which is mentioned in this chapter are those of template and feature based matching methods. In this report, these two methods of matching are explored in great detail to aid in performing an online search for a region of the body.

The advantages and disadvantages of the two matching methods are mentioned in this chapter in order to get a better understanding of them. A good understanding of the matching methods is necessary in order to create an accurate and reliable online search to locate a region of the body on the c-arm's second pass.

Chapter 3

Proposed Versascan Setup

A detailed discussion is made on the Versascan, namely the physical dimensions, and the characteristics of the camera. This chapter provides a better understanding of how the proposed modification for the Versascan would look like and operate.

3.1 Versascan

Lodox Systems is responsible for the support of its medical imaging products. This project is based on the Lodox Versascan. It is a new design which Lodox Systems is hoping to release in the near future. The Versascan is similar to the Statscan with the main difference being that it is vertically-orientated and is potentially more versatile.

The datasets required to develop the methods required in this work cannot be captured from the Versascan. However, specifications have been obtained so as to create, as closely as possible, the datasets as if taken from the Versascan. Although specifications of the Versascan are not available, the dimensions of the device are similar to those of the Lodox Statscan. Therefore all dimensions referred to and used in this project are either provided by a technical officer from Lodox or the Statscan product specifications and physical dimensions datasheet [25].

The Versascan is potentially more versatile than the Statscan as it can scan a larger number of regions of the body. This is due to its vertical orientation, which means that the only requirement needed is that the patient be in a pose where the region to be scanned is visible to the c-arm. Unfortunately, this means that the patient is essentially standing which means that minor movements can occur. These minor movements are catered for in the experiments.

Figure 3.1 shows images of the setup used to mimic data captured from the Versascan. The setup is

made up of an office panel and a garage motor track. The office panel mimics the detector for the medical scanner. A white sheet is used to cover the panel to resemble the Versascan's booth, with no landmarks, as shown in figure 3.1a. The camera used for detecting the region of interest is attached to the linear slide of the garage opener, as shown in figure 3.1b. The slide travels to both ends of the garage motor track. Coincidentally, the garage door opener takes approximately 13s to travel along its 2m track, so datasets captured using this setup closely resemble those that would be taken from the Versascan.



Figure 3.1: Versascan environment mimic setup. (a) Versascan booth. (b) Versascan c-arm.

The speed of the c-arm can actually be adjusted to half or a quarter of its speed. This project assumes that the c-arm operates at its maximum speed, travelling a distance of 2m in 13s. Only the maximum speed for the c-arm is considered as slower speeds mean that there is more time to capture data. Therefore, proving that this proposed modification can work on the maximum c-arm speed is an indication that it should also work at slower speeds.

Table 3.1 shows some important parameters of the Versascan medical scanner. The distance between the camera and the patient is important because the closer the patient the less information there is available due to the camera's field of view. The thickness of patient specification was provided by an employee from Lodox Systems. Therefore, the worst case distance, as shown in table 3.1, is the smallest distance between the camera and the patient. This worst case distance is used in the experimentation.

Specification	Lodox medical scanner
Distance from camera to detector	900mm
Thickness of patient	100mm - 300mm
Worst case distance	600mm

Table 3.1: Specifications of the medical scanner.

3.2 Camera

In this section a discussion is provided of the camera that is proposed to be attached to the c-arm to aid in performing a partial scan automatically. Different configurations of how the camera can be attached to the c-arm are compared and a recommendation is made. Finally, other characteristics of the camera that are important, such as the field of view and calibration, are mentioned. These are used when performing tests in later chapters.

3.2.1 Objective

Currently the only way to perform a partial scan is for the radiographer to do it manually. In order to perform a partial scan, the radiographer begins by using the local positioning console to move the c-arm to the required position. The radiographer then uses a laser, located within the c-arm on the x-ray source side, as a guide to indicate the required region to be scanned. The points indicated are sent to the workstation where the radiographer will finally be able to perform the partial scan. Manually indicating a region can lead to possible minor human errors.

The aim of the camera is to identify the region that requires scanning and to eliminate the possibility of minor human errors. Another advantage with using a camera to indicate and identify the region to be scanned is that it caters for the patient moving slightly.

Therefore, using a camera to indicate the region of interest is more robust than doing it manually. Manual scans are static points sent to the workstation to indicate which area requires exposure, whereas using a camera is more dynamic because it searches for the region of interest. Searching for the region of interest is more robust as the area that requires exposure changes as the patient moves.

Finally, the proposed modification aims to provide the ability to perform a partial scan automatically without needing any special training for the radiographer.

3.2.2 Configuration

Various configurations for how a camera is used to identify the region of interest are possible. The main objective is to locate the region of interest for the x-ray tube to expose that specific area, so the camera is attached on the c-arm on x-ray source side. It is advantageous having the camera attached to the moving c-arm because the location of the x-ray source is always known while the search for the region occurs. It is proposed that the camera be attached to the top of the arm so that there is no obstruction when the c-arm reaches the bottom.

Increasing the number of cameras attached to the c-arm could be used to provide more information, such as depth via stereo reconstruction. However, the main aim is to locate the region of interest and this can be done with a single camera. Depth information is nevertheless available with a single camera in this configuration as frames are captured from different viewpoints as the c-arm moves.

Finally, a patient is less likely to be intimidated with a single camera than multiple cameras in front of them. This is an important factor if patients are to remain still but more importantly calm through the scanning process. Therefore, the proposed configuration is to attach a single camera on the c-arm.

3.2.3 Model

The camera that is used for this project is the Point Grey Firefly MV FFMV-03M2C. The specifications of the camera are important as the experimental results are obtained using this device. The important details of the camera, including the image sensor, that is used for testing purposes can be found in table 3.2 [15, 26]:

Specification	FFMV-03M2C
Image Sensor Type	1/3" progressive scan CMOS
Image Sensor Model	Micron MY9V022
Image Sensor Dimensions	4.51mm(H) × 2.88mm(V) 5.35mm Diagonal
Maximum Resolution	752 × 480
Pixel Size	6.0μm × 6.0μm
Transfer Rate	400Mb/s
Maximum Frame Rate	752 × 480 at 60fps

Table 3.2: Selected specifications of the imaging device.

3.2.4 Calibration

The configuration proposed is to attach a camera on the c-arm, above the x-ray source. The ideal position for the camera is as close to the x-ray tube as possible. The proposed configuration is the closest distance to the x-ray tube as it is contained within the c-arm, x-ray source side.

Calibration is necessary, due to the proposed configuration, and figure 3.2 shows the distance between the camera and the c-arm containing the x-ray tube. The distance from the lens to the top of the c-arm was measured, with a ruler, to be 18mm and 110mm from the top of the c-arm to the x-ray tube. The distance d_{off} is now referred to as the offset distance in the scanning direction and is measured to be 128mm.

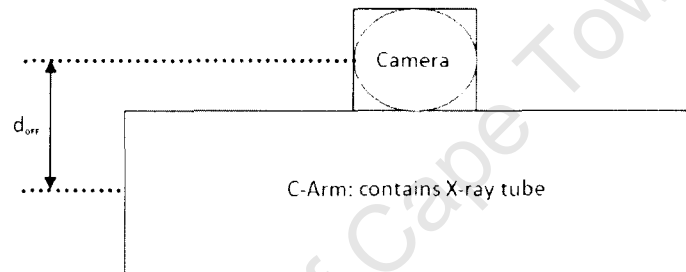


Figure 3.2: Distance between camera and x-ray tube.

3.2.5 Field of View

The field of view of the camera is important because it provides different degrees of information depending on the focal length of the lens. The smaller the focal length the wider the field of view and thus the more information that is captured in every frame. Similarly with a larger focal length this results in a more narrow field of view.

Figure 3.3 shows the field of view with the 2.8 - 12mm varifocal lens used for this project. The focal length, f , is set to 2.8mm, the minimum length, so that the largest area of data is obtained at every frame. The worst case distance, D , is 600mm from camera to patient.

Using simple geometry we can calculate that

$$\theta = \arctan \left(\frac{f}{\frac{1}{2} \text{Image sensor height}} \right) = 62.78^\circ. \quad (3.1)$$

Therefore, the length of the patient captured at every frame is

$$L = 2 \left(\frac{D - 2.8}{\tan \theta} \right) = 614.26 \text{mm}. \quad (3.2)$$

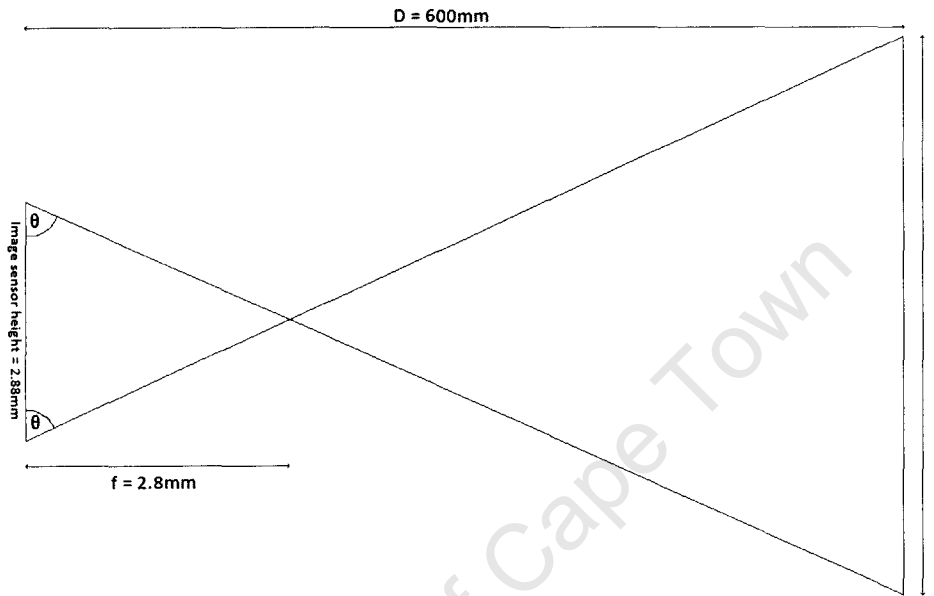


Figure 3.3: Field of view with focal length of 2.8mm

Figure 3.4 shows the different interest points from the camera, assuming it is looking perfectly horizontally, while the c-arm performs its second pass downwards. The distance between a and b is the measured distance d_{off} . The distance between a and c is half the length of the patient captured at every frame, $\frac{L}{2}$. The distance between b and c , which is referred to as the 'maximum look-ahead' distance, is important because as the c-arm moves, the camera needs to identify whether the region of interest is present in this portion of every frame. Both feature-based and template-based matching methods are used to locate the region of interest.

If the region of interest is found within the look-ahead distance, a signal is sent to the Versascan to turn on and off at the respective points. If the region of interest passes the look-ahead distance without being identified by the camera, then the test fails. The maximum look-ahead distance is calculated as

$$\text{Maximum look-ahead distance}_{Down} = \frac{L}{2} - d_{off} = 179.13 \text{mm}. \quad (3.3)$$

However, if the c-arm is to be moving upwards for its second pass, the look-ahead distance would be

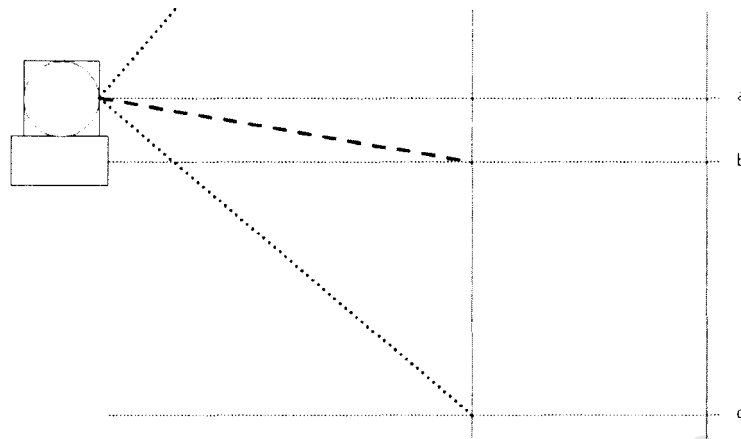


Figure 3.4: Field of view of camera attached to c-arm where points a, b and c show the camera's perpendicular view, the physical offset and the maximum 'look-ahead' distance respectively.

$$\text{Maximum look-ahead distance}_{Up} = \frac{L}{2} + d_{off} = 435.13\text{mm}. \quad (3.4)$$

Therefore, due to the larger maximum look-ahead distance, a recommendation is made that the c-arm's first pass start from the top moving downwards and from the bottom moving upwards for its second pass. The following chapters of this report uses this recommendation for the c-arm's two passes.

Chapter 4

Radiographer Workflow



Figure 4.1: Radiographer workflow to perform a partial scan.

This chapter describes the entire workflow process for the radiographer on the proposed modification for the Versascan. Figure 4.1 shows the procedure that the radiographer follows in order to perform a partial scan automatically. Finally, the last section gives a brief outline of the major differences in performing a partial scan between the original and modified Versascan.

4.1 Patient in Required Position

X-rays are taken of different regions of the body depending on the area that needs to be analysed. To be able to set up a good experimental environment, a list of regions of the body that are commonly exposed by x-rays are shown in table 4.1 [35]. Routine views are generally in the anteroposterior and lateral positions. Therefore, video data is captured of the patients in these two views as they cater for a wide range of different regions. Only a subset of all the body regions are used and those that are not being considered are marked with an asterisk, *, following their region names in the table.

	Anteroposterior	Lateral	Internal Oblique	External Oblique	Tunnel	Tangential
Abdomen	X					
Ankle*	X	X	X			
Chest	X	X				
Elbow*	X	X	X	X		
Foot*	X	X	X			
Hand	X	X		X		
Head	X	X				
Hip	X	X				
Knee	X	X			X	X
Ribs	X					
Shoulder	X					

Table 4.1: X-ray scanning routine views for different body regions. Regions marked with a * are not considered for the experiments.

4.2 Initial Video Capture

Once the patient is in the required pose, the radiographer uses the workstation to control the c-arm to perform its first pass. The result of the initial video capture is a single image of the entire patient's body, the reference image, which is displayed on the workstation where the radiographer then marks the region that requires exposure.

4.3 Marking the Region of Interest

After obtaining a suitable reference image, the radiographer marks the desired region of interest for the x-ray scan. Figure 4.2 shows three reference images with different hypothetical regions of interest marked. These are just a few examples, from table 4.1, of the different regions of the body that are being considered.

The area marked by the radiographer is important as it is used to locate the region of interest to perform the partial scan.

4.4 Second Video Capture

Finally, the radiographer uses the workstation to control the c-arm to perform its second pass. During the second pass, the video data is used not only to stitch a second image, referred to as the scanned image, but also to search for the marked region and signal the x-ray tube when to turn on and off.

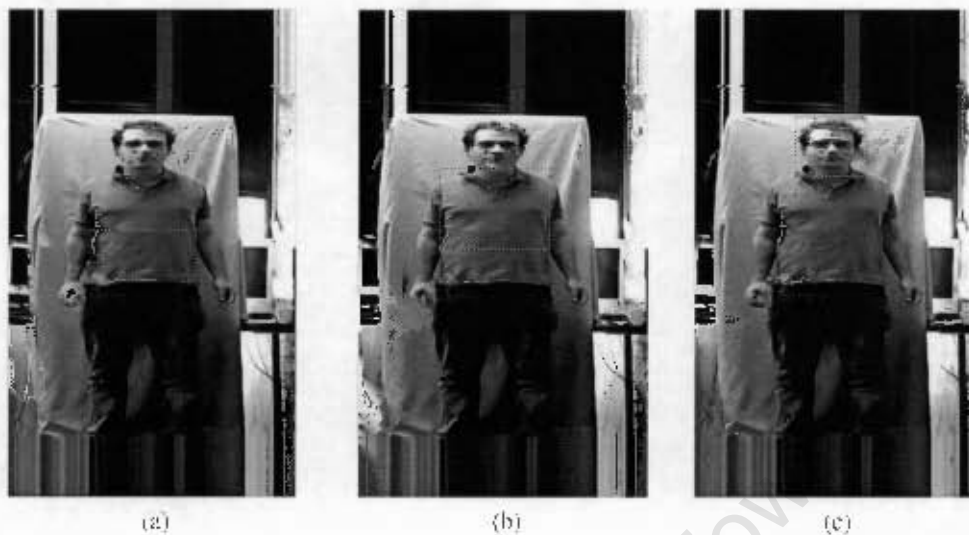


Figure 4.2: Examples of reference images with different regions of the body selected. (a) Abdomen. (b) Chest. (c) Head.

4.5 Comparison

There are a few differences in performing a partial scan between the original and modified Versascan. The main difference is that the modified version uses a camera to locate the region of interest as opposed to the radiographer specifying it with the use of a projected laser line. Using a camera to locate the region of interest provides more flexibility as it searches for the area as opposed to scanning at static points specified manually. Additionally, searching for the marked region caters for the patient moving slightly.

An advantage with the modified version is that the radiographer need only use the workstation to perform both a full body or partial scan. Special training for the radiographer is not required for the modified Versascan as they simply operate the c-arm at the workstation and mark the region of interest using a peripheral, such as a mouse. The workflow process mentioned in this chapter is used and followed to perform the experiment mentioned in the next two chapters.

Chapter 5

Workflow Processes

This chapter discusses the methods used, following the order of the radiographer’s workflow, in order to perform a partial scan. To achieve a better understanding of the entire proposed system, the methods used are broken up into separate processes. Figure 5.1 shows the radiographer’s workflow where the respective processes are performed at each stage.

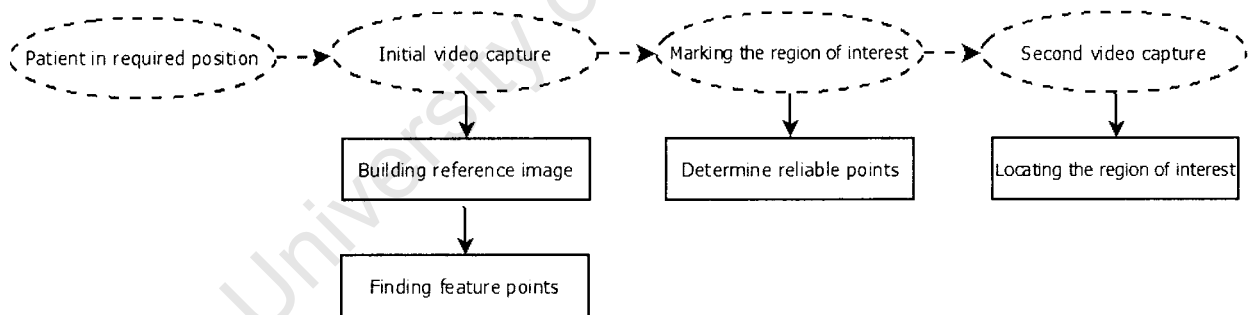


Figure 5.1: Detailed workflow process to perform partial scan.

5.1 Reference Image

The reference image is the first image displayed on the workstation and this is where the radiographer marks the region of interest. The reference image is obtained during the first pass by stitching the video captured from the camera at 60fps. Figure 5.2 shows examples of stitched reference images.

As mentioned previously, routine views are generally in the anteroposterior and lateral positions. Figures 5.2a and 5.2b show examples of the anteroposterior position and figures 5.2c and 5.2d show

the lateral position. The process of stitching is discussed later, along with a discussion of how to overcome the data loss experienced with the stitching method.

5.1.1 Capturing Video Data

Video data was not captured on the Versascan as this particular medical scanner is still under development. This limitation was overcome by creating a similar setup, as shown in figure 3.1, to mimic the data as if it had been obtained from the Versascan. Reference images produced from using this setup are shown in figure 5.2.

The video captured during the first pass consists of 780 frames using a camera at $60fps$ for the full-body scan taking 13s. The reference images shown in this report are constructed from video data consisting of 1080 frames. The reason for the additional 300 frames, or 5s, is due to the delay between operating the camera and the garage opener. It was found that the time taken for the workstation to configure the camera and begin capturing was inconsistent. To overcome this, a delay was implemented, such that after 3s of capturing a signal was sent to start the garage opener. The remaining 2s is used to cater for the approximate time taken for the slide to reach the other end.

Therefore, there are 3 important aspects in the reference images in this report, as shown in figure 5.3. There are the two delays, at the beginning and the end, and when the slide starts moving. When the garage opener is signaled to operate the entire setup shakes for a brief moment. This causes the horizontal misalignment artifacts visible near the top of the images, and indicated in figure 5.3. It is unnecessary to accommodate for the brief shaking as the regions which are important for this project are below this point.

In the proposed modified Versascan, the camera captures video data consisting of 780 frames while the c-arm moves. Therefore, both delays and the initial brief shake will not be present on the actual Versascan.

5.1.2 Stitching Video Data

Image stitching is the process in which multiple images are aligned by various registration algorithms and blended together in a seamless manner [32]. As mentioned previously, the video datasets were captured by a camera mounted on a vertically-orientated garage motor track. The reference image was created by taking a number of rows, r , at each frame captured, as shown in figure 5.4. The rows that are used for stitching the reference image are at the centre of every frame.

The method of stitching performed uses a number of rows at each frame and stacks them underneath



(a)



(b)



(c)



(d)

Figure 5.2: Examples of reference images. (a) and (b) Anteroposterior position. (c) and (d) Lateral position.



Figure 5.3: Important parts of the reference image.

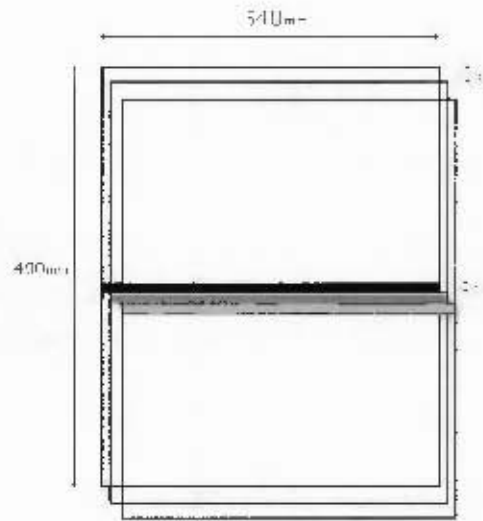


Figure 5.4: Process of image stitching from video dataset where r represents the rows considered for stitching and d_r is the distance between each frame captured.

or above each other, depending on the scan direction, i.e. pass 1 or pass 2. The result of taking into account the centre rows of the video data produces the reference image. The only problem with this method of stitching is the possibility of data loss, and whether such loss impacts on the objective of locating the region of interest.

Using the centre rows of every frame is effectively providing the information that is directly in front of the camera. However, for the second pass, the camera would have to 'look-ahead' to identify the area before the c-arm reaches it by considering the rows that are higher than the centre, considering the c-arm travels upwards for its second pass.

5.1.3 Video Data Loss

The method of image stitching mentioned previously causes some concern for data loss as it simply takes a number of rows at the centre in each frame and constructs the reference image from that. In figure 5.4, with a camera capturing at $60fps$, the distance d_r between each frame can be calculated as follows:

$$d_r = \frac{2000mm}{13s} \cdot \frac{1}{60fps} = 2.564mm. \quad (5.1)$$

From equation 3.2 it is seen that the vertical extent of the patient captured at every frame, L , is $614.26mm$ which is calculated using the worst case distance from the camera to the patient. The

rows that are considered for stitching are at the centre so the height, d_{row} , of a single row in any frame is

$$d_{row} = I/480 = 614.26/480 = 1.280mm. \quad (5.2)$$

Therefore, if only 1 row is considered there is a significant amount of data loss. However, if 2 rows are considered at every frame, the data loss becomes insignificant and it can be assumed that the reference image is a close representation of the patient. The percentage of data loss with 2 rows considered is

$$Data\ loss = 2.564 - (2 \times 1.280) = 0.004mm = 0.16\%. \quad (5.3)$$

Therefore, if the patient's thickness is smaller than 300mm, overlap occurs in the stitching method but all the information remains in the reference image.

5.2 Finding Feature Points

Feature points are important as they are used to locate the region of interest on the scanned image. The Harris corner detector has been selected to find feature points on the reference image [7] for reasons mentioned in section 2.2.2. However, one problem was found with this approach: when there is ambiguity or patterns present around the region of interest, corners are sometimes found at other locations that are visually similar.

Ambiguity is found when, for example, corners are found along the contours of a rough mountain. These corners are similar, provided that the contours have a relatively similar gradient and the background is comparable. Patterns that exhibit a repetition of shapes can lead to determining corners being found around these areas and are very similar but are positioned at different locations.

Therefore, to remove corners that are either ambiguous or found in patterns, a need for reliable feature points is necessary. The process of identifying reliable points is done after finding all corners on the reference image. The next section discusses the test that is performed on each individual feature point to determine whether it is reliable or ambiguous.

5.3 Determine Reliable Points

The process of determining reliable feature points occurs after the radiographer has marked the region of interest, as only the feature points which fall within the marked region are considered and the rest are ignored. Only feature points within the marked region are considered as these are the ones used to locate the region of interest on the scanned image. Therefore it is not necessary to determine if all the feature points are reliable or ambiguous.

The approach for determining whether a feature point is reliable or ambiguous consists of looking at the neighbourhood of each point. A small window of size 15×15 , centred at the detected corner, is considered. Template matching is then performed over the surrounding area of size 75×75 to see whether there are other locations which have similar appearances. Normalized cross correlation is selected as the measure of how similar the feature point is to the background. Various thresholds have an impact on the resulting correlation value which determines how reliable each corner is and this is discussed in more detail in section 6.2.

Figure 5.5 shows three images and demonstrates the process of obtaining the reliable points. Note that this is a zoomed in reference image. First all feature points are found, then points outside of the region of interest are removed, and finally the reliable points are calculated.



Figure 5.5: Obtaining reliable feature points from an abdomen region. (a) Reference image with feature points. (b) Feature points only within region of interest. (c) Reliable points only within region of interest.

5.4 Locating the Region of Interest

After the region of interest has been marked, the radiographer controls the c-arm to perform the second pass. The second set of video data is not only being stitched together, but also being used to locate the region of interest.

This section discusses how reliable feature points are used to locate the region of interest and signal the x-ray source at those respective points. The assumption has been made that the signal to trigger the source is instantaneous, and this assumption has been confirmed to be reasonable by a Lodox Systems employee.

5.4.1 Searching Ahead

The stitching method used is to take two rows at the centre of every frame and sequentially stack them until the final frame is captured. In order to find the region of interest and provide the location to the x-ray source before it passes it, a search is required to take place ahead of the c-arm. Searching ahead of the c-arm is done by adjusting the stitching method during the second pass.

Instead of using two rows at the centre of every frame, higher rows are considered. The rows to use are effectively determined by the desired look-ahead distance discussed in section 3.2.5. The look-ahead distance is not set to a constant value but is varied depending on the height of the marked region. In the case where the height of the marked region is greater than the maximum look-ahead distance, the maximum look-ahead distance is considered.

The height of the marked region is measured by the number of pixels along a single column. Therefore, the varying look-ahead distance is calculated by converting the number of pixels to a distance, with the use of equation 5.1, at 2.56mm per pixel.

5.4.2 Online Searching Method

The online searching method is necessary in order to locate the region of interest in the scanned image efficiently and accurately. An online searching method means that a search is constantly being performed while the c-arm performs its second pass. The time taken to locate the region of interest is important as it is necessary to identify the marked area before the c-arm reaches it.

The time available for the online search is catered for with the varying look-ahead distance. As mentioned previously, in order to achieve accuracy within 2% source to image detector distance (SID), the actual region scanned as a result of the search needs to fall within a distance of 20mm

from the region marked by the radiographer. The measure of accuracy is discussed later in this report.

The approach is to use reliable points found within the region of interest, on the reference image, in order to identify the respective area in the scanned image. A template of size 15×15 pixels centered on each reliable point is used for the search in the scanned image. Each template is referred to as $T_n(x_n, y_n)$ where n is the number of the reliable point and (x_n, y_n) is the pixel coordinate of the n^{th} reliable point. Normalized cross correlation is used as a distance measure on the scanned image for each reliable point to find where the best match is situated.

In an ideal situation, if the patient had remained perfectly still then each template would find a best match at the same location on the scanned image, but with a vertical offset of the look-ahead distance. However, since minor movements are permitted an online search is necessary. To cater for these minor movements, a template of size 15×15 pixels and a search window of size 51×51 pixels is used to provide an allowance of minor movements of 25 pixels in any direction. One factor that determines the size of the search window is the accuracy as a distance error of more than 20 pixels is greater than 5%, which is regarded as a failed test.

Figure 5.6a shows an illustration of a number of reliable points using the pixel coordinate system. Figure 5.6b shows the same number of search windows, centered at the original reliable points, to where the corresponding match is expected to be.

The search for the matches for the reliable points on the scanned image yields normalized correlation coefficient values at each point within the search window. The point with the highest match value is regarded as the best match. If the highest match value is greater than some search threshold, then that point is considered a reliable match. If it is below the threshold then the corresponding match is determined to have not been found and the match pair is thus ignored. Once a specified number of corresponding reliable points are found, an estimate of the marked region on the scanned image can be calculated. An evaluation of this strategy is made later in this report where both the number of corresponding points and the searching threshold are varied in the experiments to determine their effect on the results.

The estimated location of the marked region on the scanned image is calculated with the use of the pixel coordinates of both the original and the corresponding match points found. First, the pixel distances are measured, both horizontally and vertically, between each reliable point and the marked region on the reference image. These distances are then transferred to the corresponding match points and are used to calculate the location of the marked region on the scanned image. In principle the marked region can be found by using the pixel distances measured on any single reliable match as they all indicate the location of the marked area.

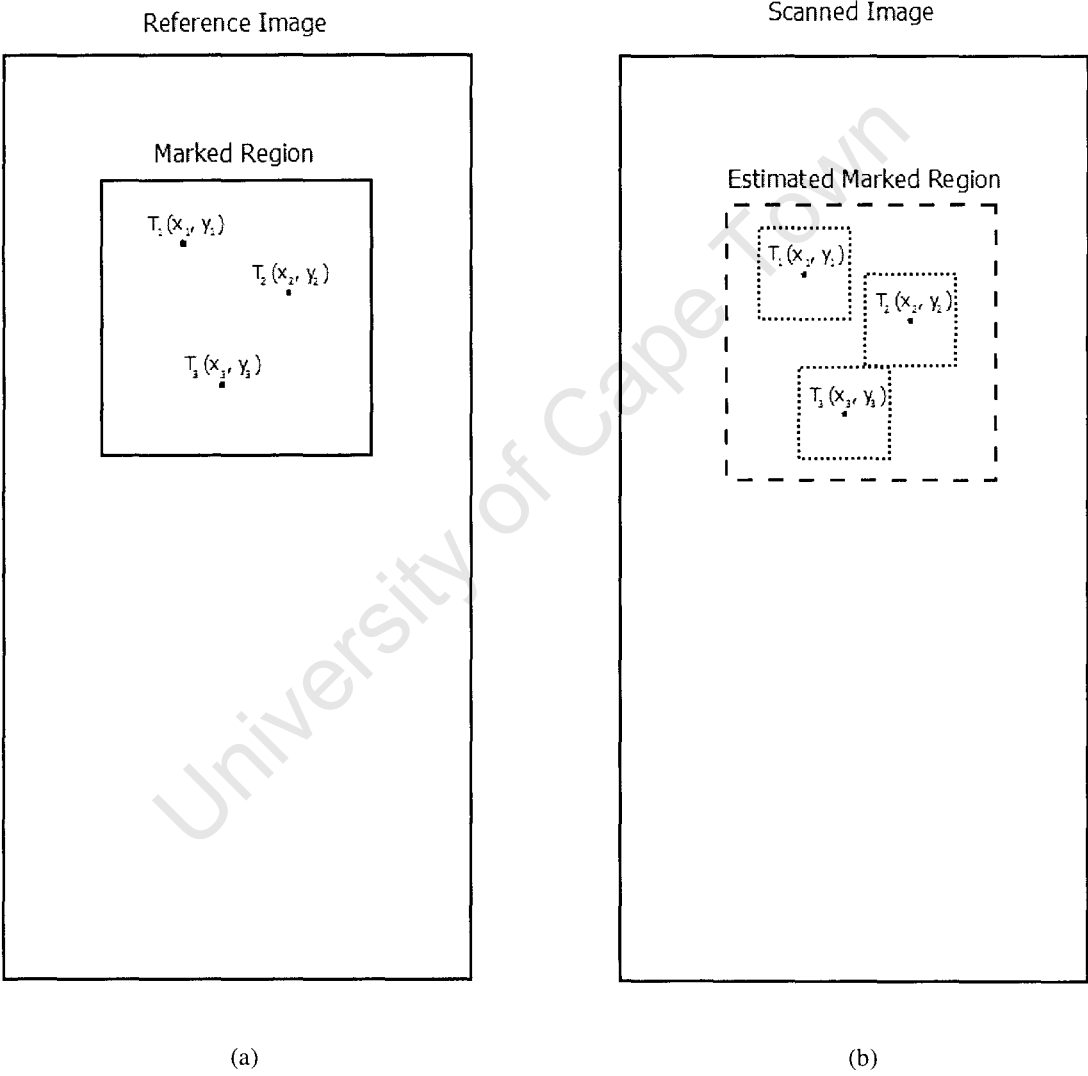


Figure 5.6: Illustration of numbering reliable points and searching method.

In an ideal situation, with no movement of the patient, the offset of each match would indicate the same marked region on the scanned image. However, in the presence of minor movements each matching pair of points can potentially indicate a different location to where the marked region is. The average location predicted by all matching pairs of points is taken to be the estimated region of interest.

The method of locating the estimated marked region is dependent on the accuracy of the radiographer's ability to mark the region of interest on the reference image. If the radiographer has marked an area which does not contain the entire body region, then this method would potentially only scan that portion of the body. Therefore it is the radiographer's responsibility to ensure that the marked region is accurate and correct.

A recommendation has been made to search for the marked region as the c-arm scans in the upward direction due to the greater look-ahead distance, shown in equation 3.4. Figure 5.7 shows the reference image, on the left, and the scanned image, on the right, as the c-arm moves downwards and performs an online search. In this case, the number of corresponding reliable points is 2. The yellow line indicates the position of the c-arm and the green line indicates the camera's viewpoint, which is also the last row that has been stitched.



Figure 5.7: Result of online search after finding two corresponding reliable points. The yellow and green lines indicate the position of the c-arm and the camera's viewpoint respectively.

A closer look at the scanned image is needed to see where the estimated location of the marked

region is. Figure 5.8 shows the same scanned image from figure 5.7 with the addition of the red box which indicates the estimated location of the marked region.



Figure 5.8: Result of online search indicating the location of the marked region on scanned image. The yellow and green lines indicate the position of the c-arm and the camera's viewpoint respectively. The red box is the estimated location of the marked region.

The workflow process mentioned in this chapter is used for the online search on the large dataset acquired and the results of these experiments are discussed in more detail in the next chapter.

Chapter 6

Experiments, Results and Findings

This chapter provides a detailed analysis of the experiments performed on the datasets acquired for this project. The first section creates a comparison for the estimated marked region found on the scanned image in order to get an understanding of how accurate it is. Various thresholds have been used in the experiments and this chapter evaluates each individual threshold and demonstrates how they influence the results. Finally, the performance of each individual region is calculated to determine whether certain regions perform better than others.

6.1 Ground Truth

A measure of accuracy needs to be defined for the estimated location of the region of interest on the scanned image. A maximum error of 2% of the SID is allowed in order for the proposed modification of attaching a camera onto the c-arm to be accepted for the Versascan. The ground truth is only used as a measure of accuracy to see how well the search method performs.

Visual inspection can be used to see whether the estimated marked area has captured the required body region, but this doesn't provide a quantified accuracy measure. Therefore, once the second video has been captured, another search is performed. However, in this case instead of doing a progressive search, all of the reliable points on the reference image are used. The region found using all the reliable points, referred to as the ground truth, is then compared against the estimated marked area for an accuracy measure. The ground truth is assumed to be the closest location to the original marked region. In addition to the ground truth, a visual test is also made to determine whether the correct marked region is found on the scanned image.

The distance between the ground truth and the estimated region is the distance error used to de-

termine accuracy in pixels. This can be converted to world coordinates using the pixel to distance conversion from equation 5.1, namely 2.56mm per pixel.

Using the previous example, where figure 5.8 shows the estimated location of the marked region on the scanned image, the ground truth is determined as shown in figure 6.1. Figure 6.1 shows the entire scanned image where the red and green boxes represent the estimated locations of the marked region and the ground truth respectively.

Using figure 6.1 as an example, the error is found to be 7 pixels horizontally and 4 pixels vertically, which is equivalent to 17.92mm and 10.24mm. Therefore, the example illustrates a successful test as it resulted in locating the region of interest correctly within 2% accuracy. The results of each test in the experiment are analyzed using the ground truth to determine whether the correct region has been found and what the accuracy is.

6.2 Thresholds

Various thresholds were mentioned in the workflow processes, all of which influence the results in some way. This section evaluates the different thresholds mentioned in this report and a range of optimum values is identified that give a suitable result. The three thresholds evaluated are the number of reliable points found on the scanned image, the correlation coefficient value when searching for the reliable points, and the reliability measure of the feature points used. These three thresholds are evaluated in more detail in this section and are referred to as: number of matches, searching threshold and point reliability threshold accordingly. Finally, a discussion is made on how combinations of the thresholds influence the results.

The experiment consists of two hundred and sixty tests where different regions of the body were marked and searched for. The results in this section have been captured by repeating the experiment and changing the various thresholds accordingly. For experimental purposes, tests which have achieved an error within 2%, 3% and 5% are recorded as passed tests as the marked area identified on the scanned image contains the body region.

Each threshold is varied and a recommendation is made based on two results, the percentage of tests passed and the percentage of those tests passed that are within 2% and 3% error. The outcome of combining the two results is a percentage of tests passed within a certain accuracy. Therefore, a recommendation is made for each threshold based on the combination of the two results. For all the accuracy illustrations, the red and blue points show passed tests within 2% and 3% respectively.



Figure 6.1: Entire scanned image with an estimated location of the region of interest and ground truth indicated by red and green respectively.

6.2.1 Number of Matches

The number of matches required is an important parameter in the matching process. If this parameter is set to be one, the estimated location of the region of interest would be obtained from the relative position of a match of one point. This makes the process of finding the estimated region fast but potentially inaccurate. On the other hand, if the parameter is set too high then the estimated region might not be found because the number of matches required within the marked region might never be obtained.

Therefore, the varying number of matches required used for the experiment are 2, 5, 10, 20 and 50. This particular experiment used the search and point reliability threshold of 0.9 and 10 respectively.

It is found that the greater the number of matches required, the more tests that fail. Figure 6.2 shows the results of the range of the number of required matches considered.

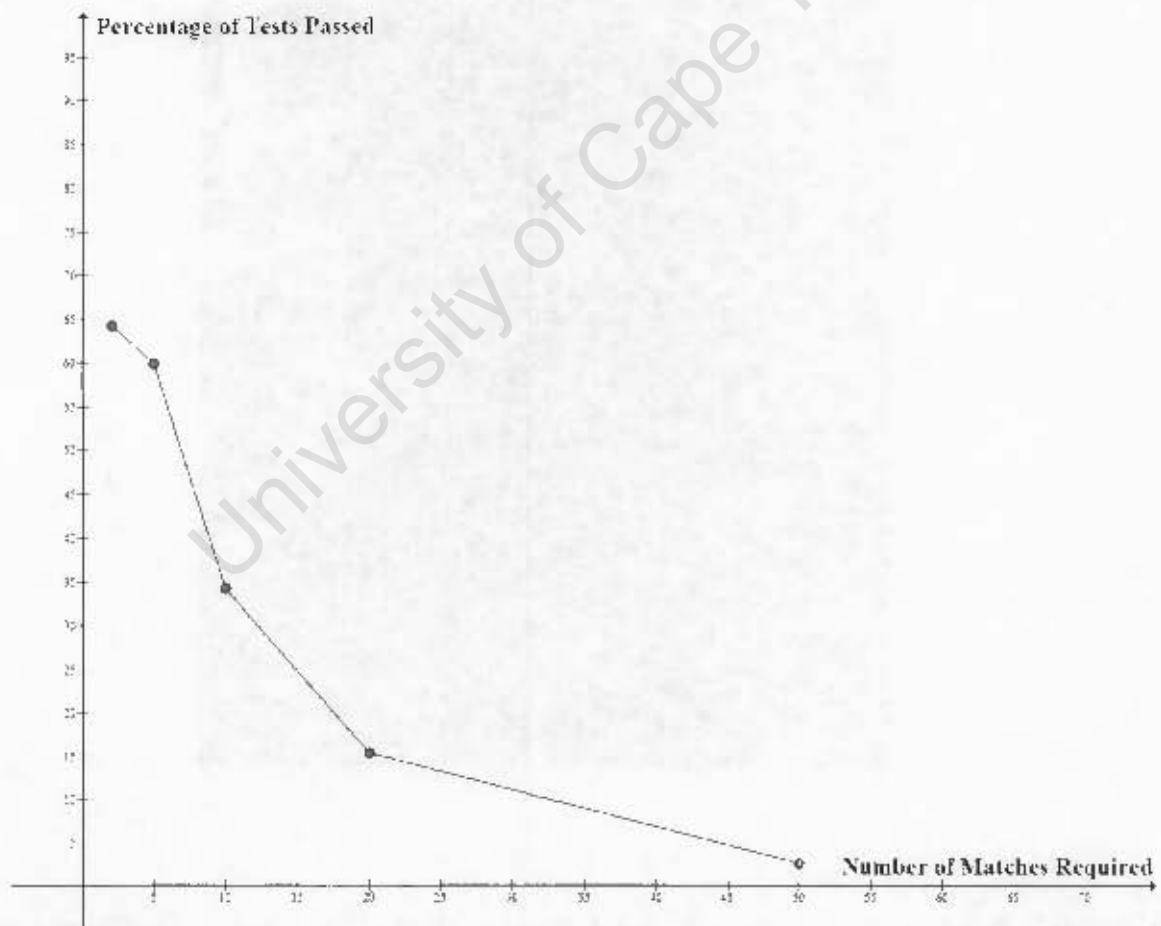


Figure 6.2: Results of tests passed with varying number of matches required.

The results suggest that one should use a low number of matches for the search. However, figure 6.3 shows an increase in obtaining more accurate results as the number of required matches increases to

approximately 20.

The product of combining the two results are, in order: 0.225, 0.273, 0.168, 0.092 and 0.004. Therefore a recommendation is made to set the number of matches required to 5 to cater for both correct and accurate results.

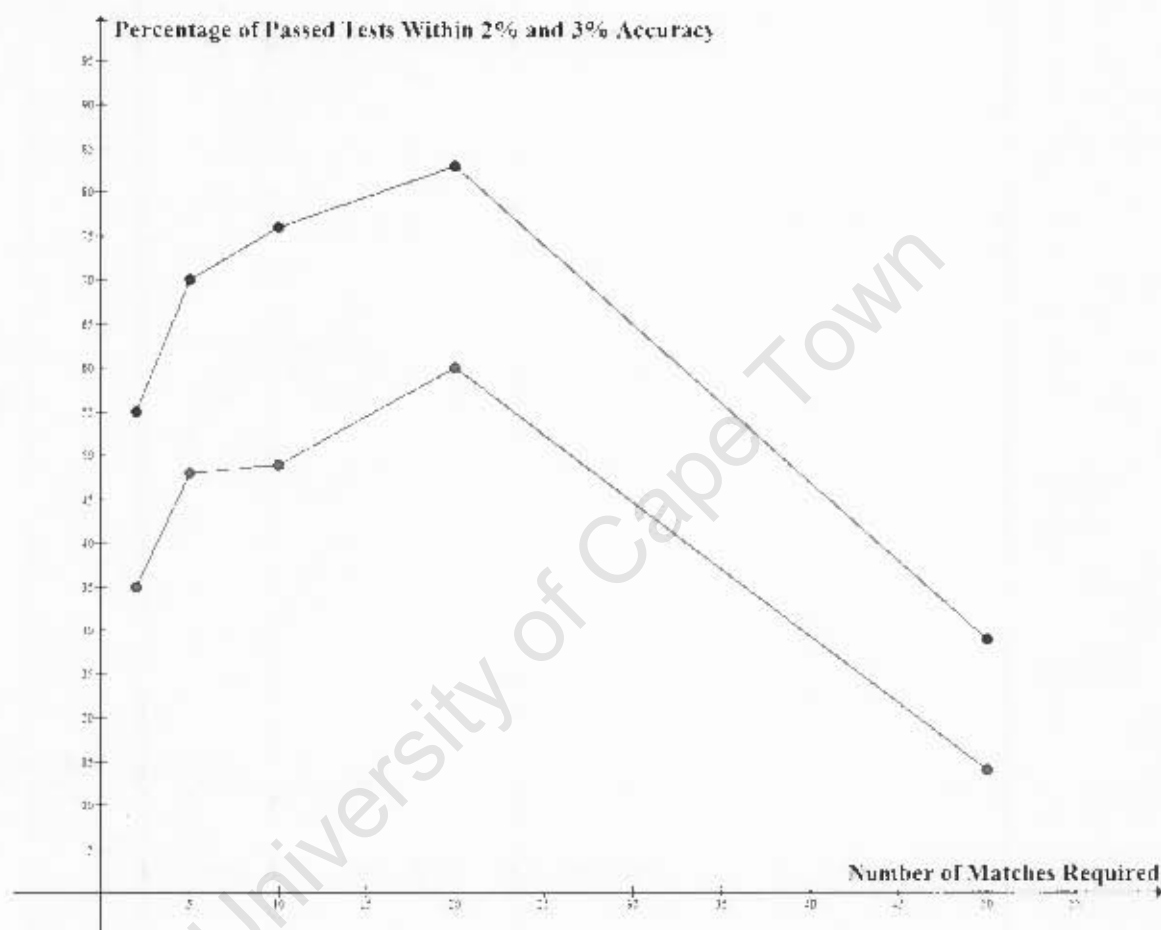


Figure 6.3: Results of passed tests that are within 2% and 3% accuracy when varying the number of matches required.

6.2.2 Searching Threshold

The correlation coefficient values from the search must be greater than the search threshold for a match to be declared. The highest correlation coefficient value around the search window is then used as the best match location to where the matching point is. The search threshold is therefore an indication of how good a match has to be for it to be considered reliable.

The thresholds used for this experiment are 0.8, 0.85, 0.9, 0.95 and 0.98. This experiment used a required number of matches and a point reliability threshold of 5 and 10 respectively.

Figure 6.4 shows the results of the experiment where an increase in the search threshold results in a decrease in the percentage of tests passed.

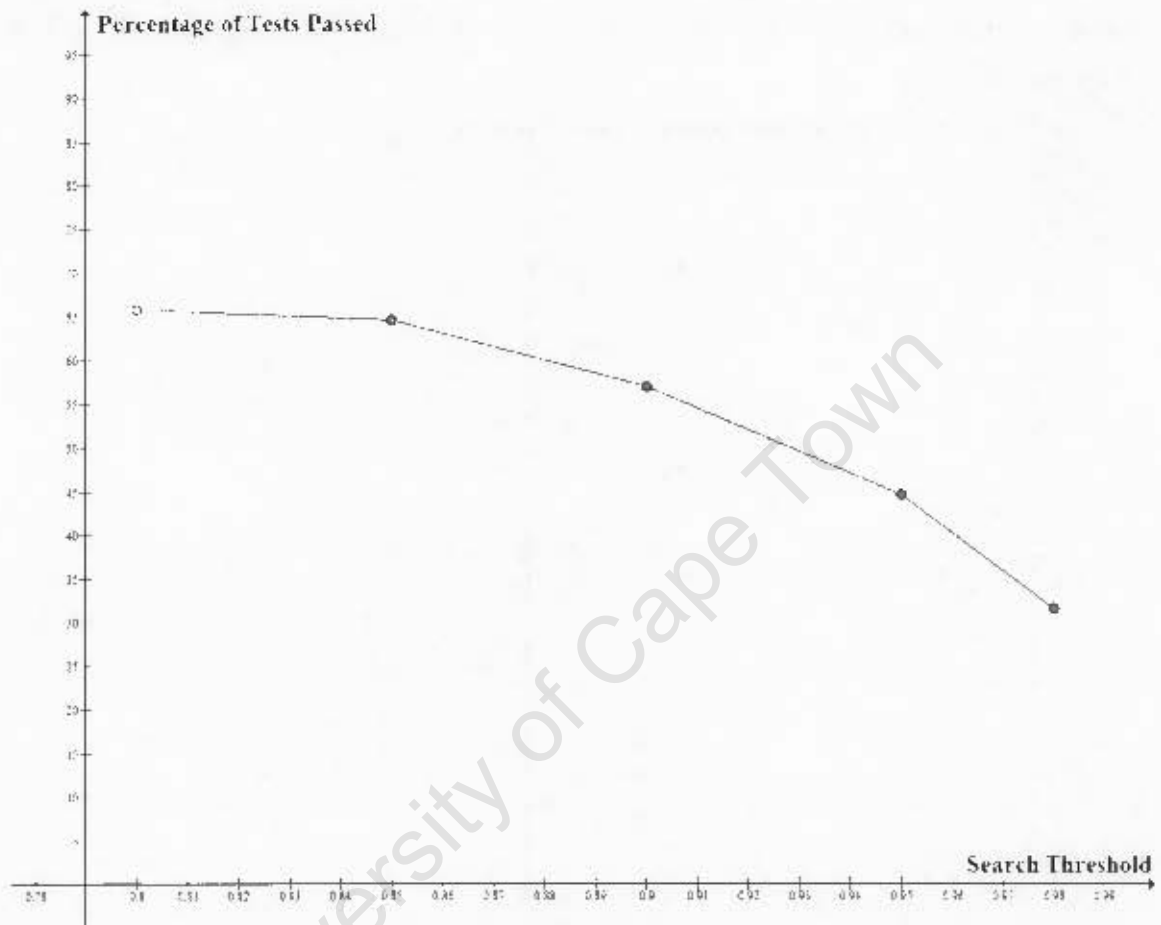


Figure 6.4: Results of tests passed with varying the search threshold.

The accuracy of the tests passed is not drastically affected by the varying search threshold, as shown in figure 6.5. The product of combining both results are, in order: 0.257, 0.271, 0.273, 0.210 and 0.136. Therefore, a recommendation is made to use a search threshold of 0.9 to cater for both correct and accurate results.

6.2.3 Point Reliability Threshold

A process of determining whether a feature point is reliable or ambiguous was proposed to achieve more accurate results. Determining whether a feature point is reliable occurs after the radiographer has marked the region of interest. The reason for determining the reliability of the feature point after the region has been marked is because only feature points within the marked region are considered. Determining the reliability of each feature point within the marked region caters for the minor

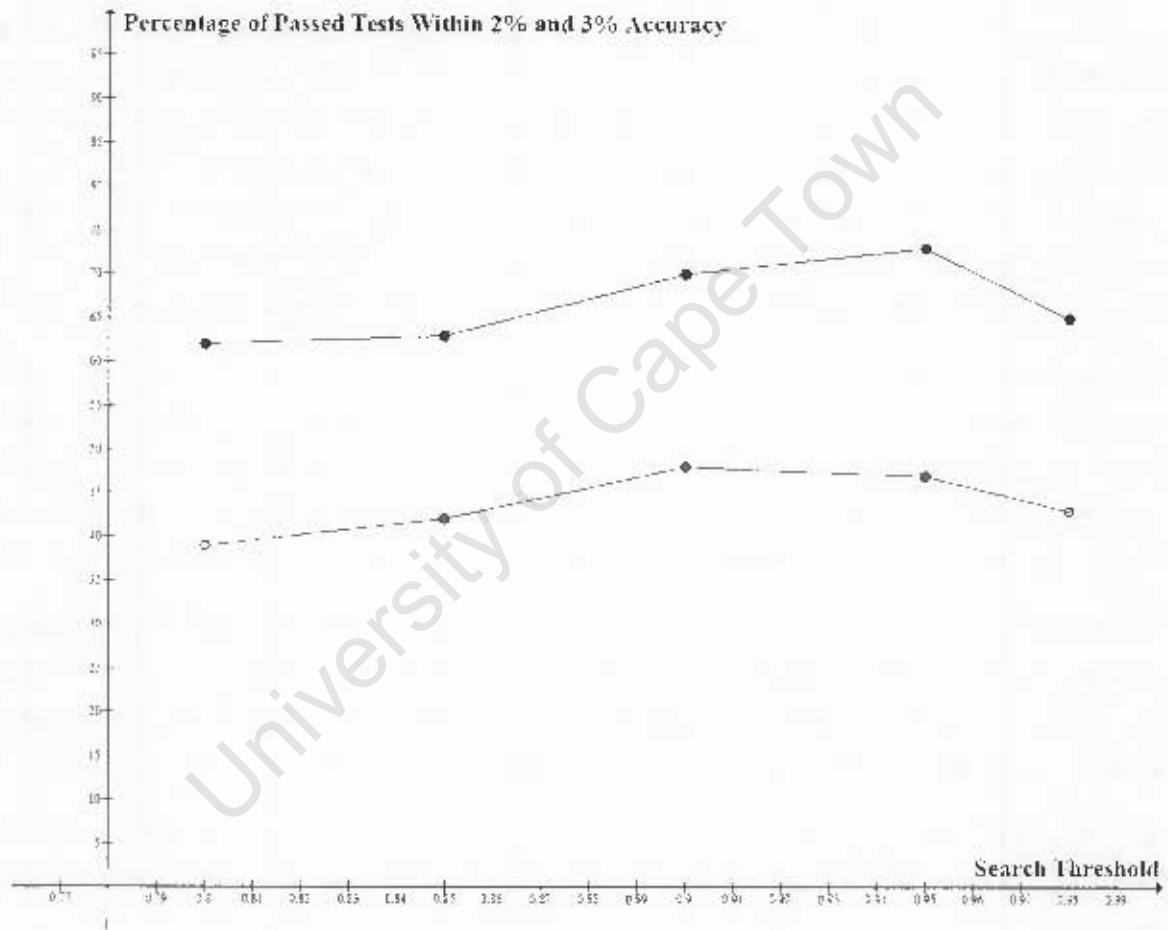


Figure 6.5: Results of passed tests within 2% and 3% accuracy when varying the search threshold.

movements made by the patient.

A test is performed on each feature point individually to specify whether it is reliable or ambiguous. This test uses template matching of size 15×15 and a search window of size 75×75 with each feature point as its centre, to determine whether there is a similar point nearby. Normalized cross correlation is used as the template matches around the search window.

All the correlation coefficient values are evaluated and accumulated as a weighting to how reliable the point is. If the correlation coefficient is 1, this is generally the case where the template is in its original position and therefore ignored. If the correlation coefficient is greater than 0.95, it is assumed that there is a similar point in the search window and 5 is added to the accumulated weighting. If it is greater than 0.9, then only 1 is added as it is not strongly similar. If the highest correlation coefficient value in the search window is less than 0.9, it is assumed that there are no points similar and ignored. Therefore, feature points with a smaller accumulated value are more reliable than ones with a larger accumulated value. Reliable points are necessary in order to locate the marked area accurately as the patient could have moved slightly and the surrounding area is possibly similar.

The accumulated weighting value is then compared to the point reliability threshold. If the weighting is smaller than the threshold then it is identified as a reliable point. If the weighting is greater than the point reliability threshold it is identified as an ambiguous point and is ignored when performing a search for the marked region. In the experiment, the point reliability threshold values considered are 10, 20 and 50, and the different outcomes on the reference images are shown in figure 6.6.

To see the effects of the feature point reliability threshold in the experiment, the required number of matches and the search threshold have been set to 5 and 0.9 respectively. It is shown in figure 6.7 that a higher point reliability threshold results in an increase towards the percentage of tests passed. However, reliable points achieve better accuracy than non-reliable points, as shown in figure 6.8.

The product of combining the two results are, in order: 0.290, 0.250 and 0.226. Therefore, a recommendation is made to use a point reliability threshold of 10 to cater for both correct and accurate results.

6.2.4 Summary

The thresholds used for the experiments are important as they impact on whether a region is found correctly and accurately.

In the experiment, it was found that the different thresholds influenced each other in various ways. The point reliability threshold was found to be dominant as it determined not only the reliability of the feature point but also the number of points available to search for.

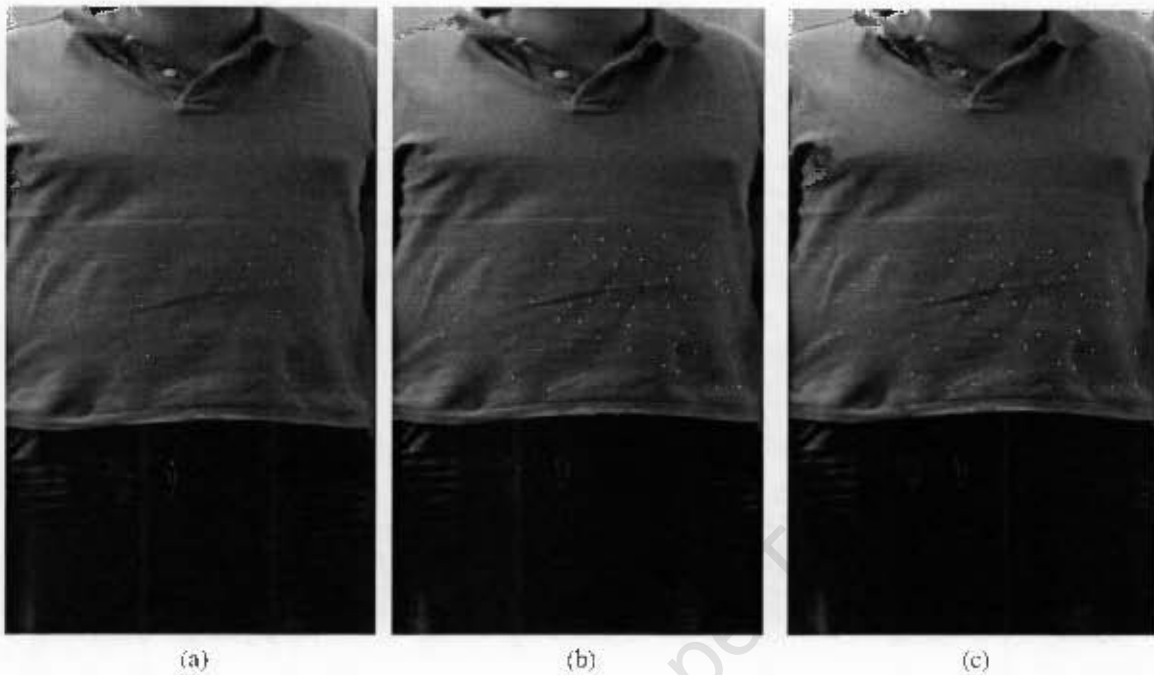


Figure 6.6: Reference image after abdomen marked as region of interest and reliable points determined. Point reliability threshold set to (a) 10, (b) 20 and (c) 50.

If the number of matches required is greater than the number of feature points available within the marked area, the search would fail. However, if the feature point is reliable and not at all ambiguous, a high search threshold is not necessary and therefore increases the chance of finding a reliable match.

After considering every relation of each threshold it is recommended that the number of matches required, the search threshold, and the point reliability threshold be set to 5, 0.9 and 10 respectively.

6.3 Performance of Different Body Regions

This section evaluates each body region individually to see if some regions are found more easily than others. Recommended values for the different thresholds mentioned previously are used in determining the performance of different body regions.

The previous section discussed the effect of varying the different thresholds in the experiments. Using the recommended values for the thresholds, an analysis is done on each body region and all the body regions together.

The format of the results is as shown in table 6.1. Due to the large number of tests done, only a small

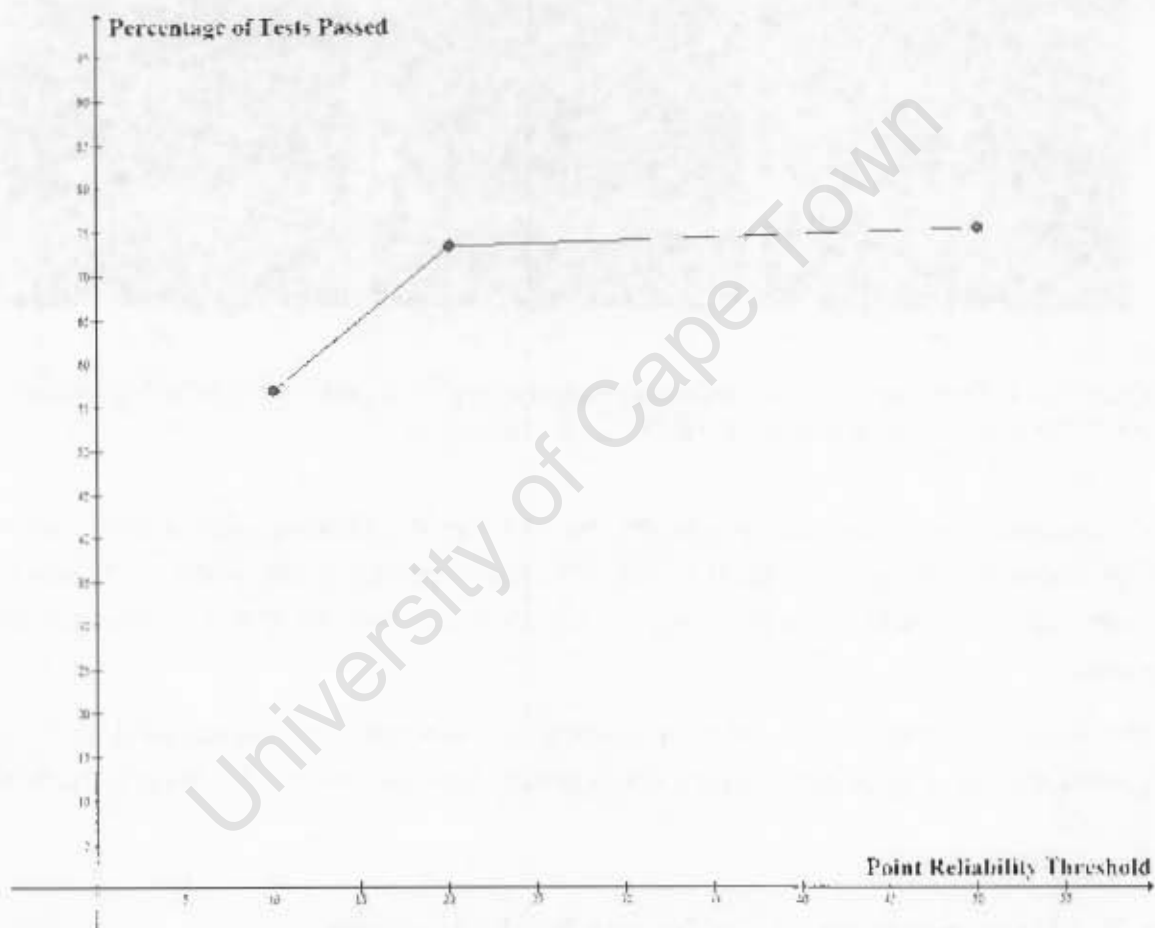


Figure 6.7: Results of tests passed with varying the point reliability threshold.

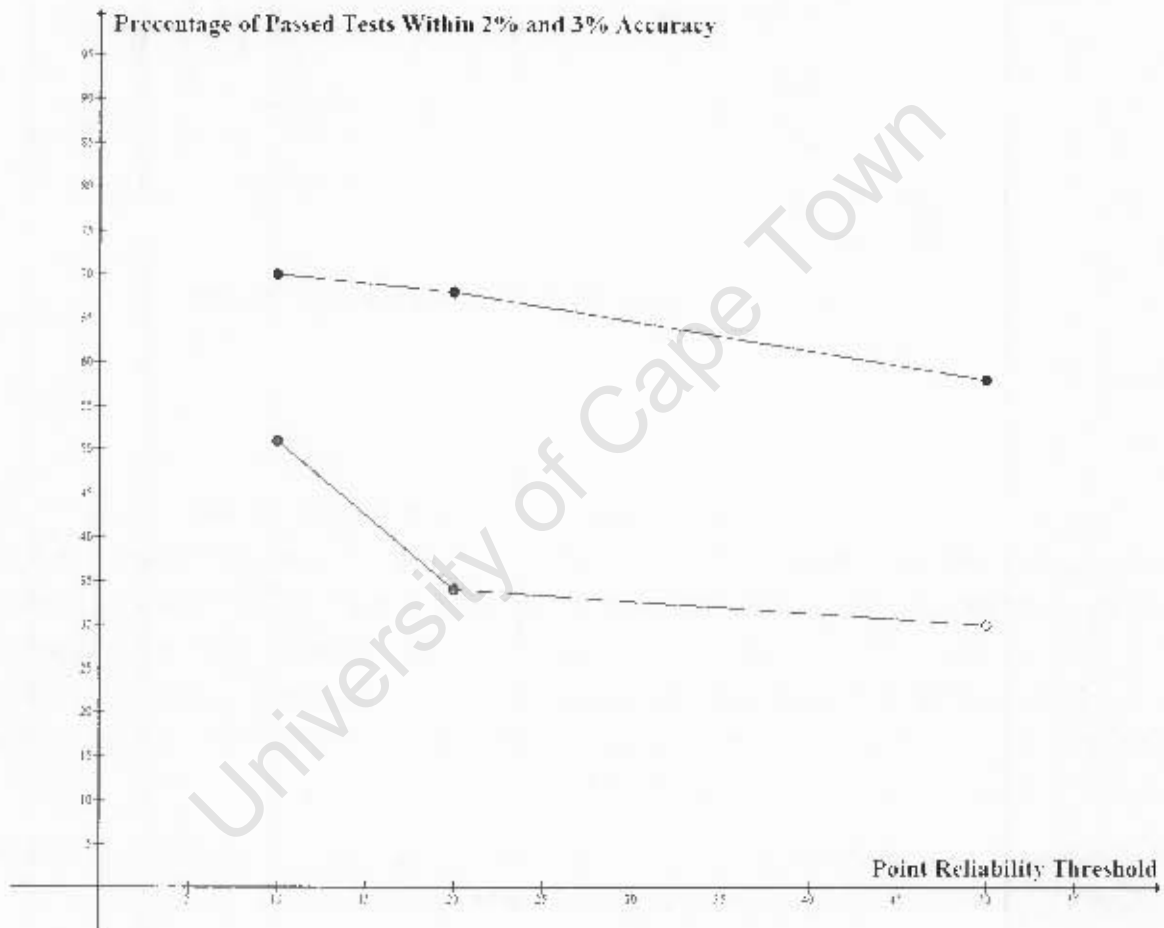


Figure 6.8: Results of passed tests within 2% and 3% accuracy when varying the point reliability threshold.

section of the tests are shown. The spreadsheet containing all the results mentioned in this chapter can be found in Appendix B. For the sake of anonymity, names of subjects are not revealed and thus they are instead referenced by numbers.

Patient	Data	Region	V.Acc	H.Acc	ROI Found	FDA	Signal	Notes
1	1-2	Abdomen	30	4	Y	N	Y	1*
		Chest	24	14	Y	N	Y	1*
		Head	7	3	Y	Y	Y	
		Knee			N	N	N	2*
...

1* - Error greater than 5%

2* - Region not found

Table 6.1: Table of results for locating the region of interest.

The first five columns of table 6.1 are self-explanatory: patient number, the dataset used, the region selected, and both the vertical and horizontal pixel errors. The next column determines whether the region of interest was found at the end of the experiment and the FDA column is whether the accuracies are within 2%. It is necessary to know whether the region of interest is found before the c-arm has passed the first signal point and this is indicated by the signal column. Finally, the last column is for any additional notes that are seen in the results, such as the region not being found or indicating what the accuracy is if greater than 2%.

An analysis is performed on each body region to see whether some regions perform better than others. Figure 6.9 shows the percentage of tests passed for each region.

The accuracies of each body region of the respective passed tests are shown in figure 6.10, where red indicates an accuracy within 2% and blue within 3%.

The poorer performing regions, namely the abdomen, the head and the ribs, have been excluded to observe how it affects the overall performance. Figure 6.11 shows the performance using all the body regions and the other excluding the poorer performing regions. The blue, red and green indicates the tests passed and accuracies within 2% and 3% respectively. Removing the poorer performing regions resulted in an increase in the number of tests passed without having an impact on accuracy, as shown in figure 6.11. This shows that certain regions of the body are easier to locate than others using the proposed online search.

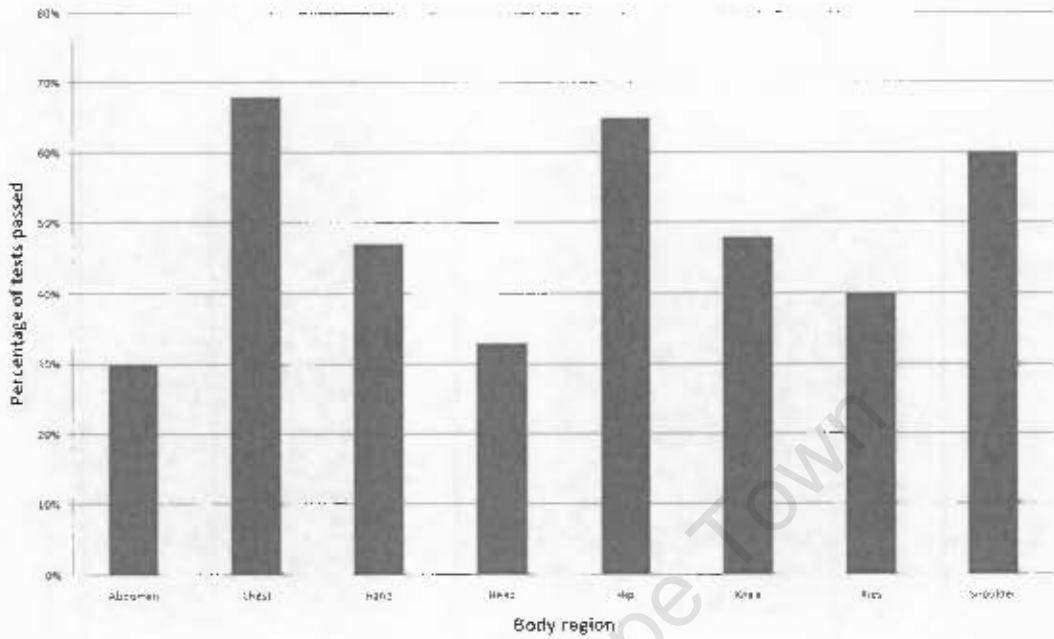


Figure 6.9: Results of tests passed for each body region.

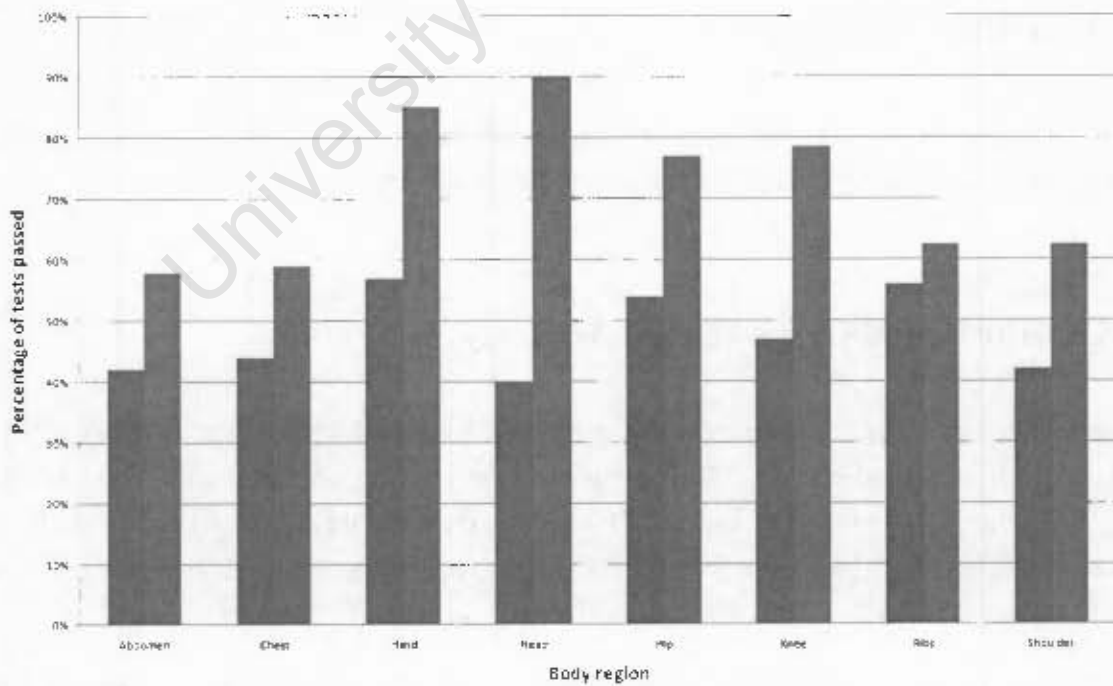


Figure 6.10: Results of each body region and accuracies within 2% and 3% shown as red and blue respectively.

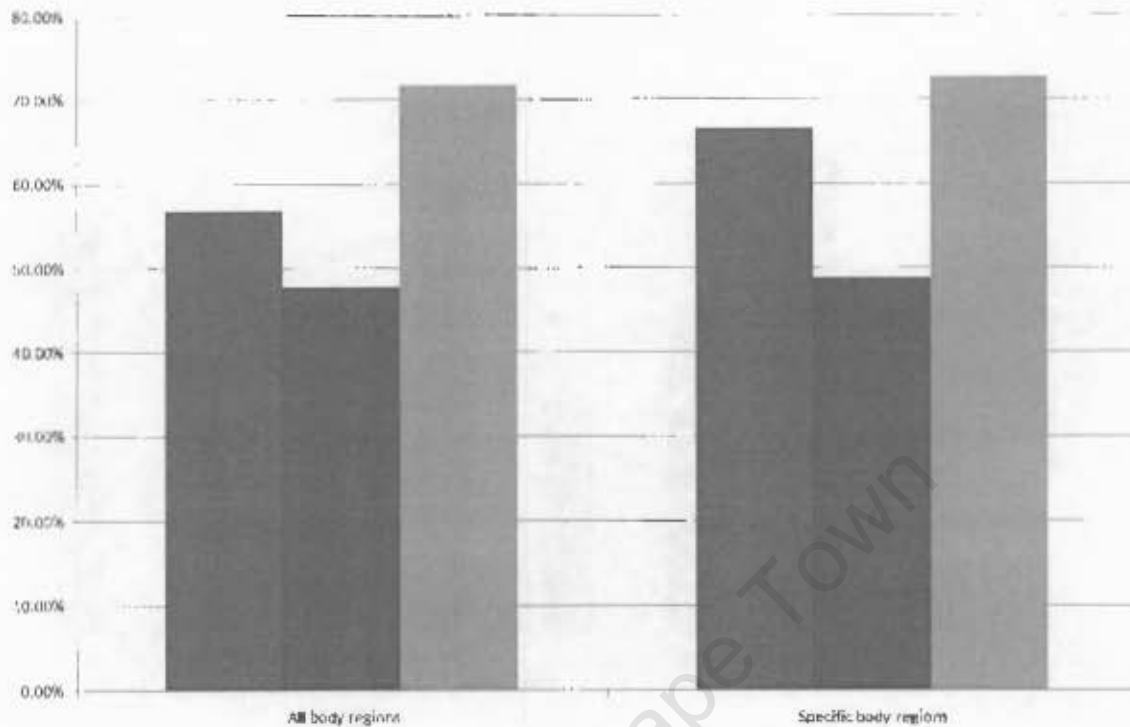


Figure 6.11: Results of the experiment using all and only specific regions where the blue, red and green indicates the tests passed and accuracies within 2% and 3% respectively.

6.4 Other Findings

This section provides additional information to the accuracies of the results and reasons for the failed tests. The recommended values for the different thresholds are used in this experiment.

6.4.1 Horizontal versus Vertical Accuracy

This section calculates accuracy by considering both horizontal and vertical distances between the ground truth and the estimated marked region when determining whether it falls within a certain criterion to achieve the required accuracies. This section separates the horizontal and vertical distances to determine which of these has a greater effect on the results. The illustrations shown in this section are represented as red and blue indicating horizontal and vertical directions respectively.

In this experiment, 177 tests have found an estimated region of interest successfully, and the horizontal and vertical distance errors are captured accordingly. Figure 6.12 is a histogram representing the pixel distances between the ground truth and the estimated marked region. Figure 6.13 groups the pixel distances to their respective accuracy percentages.

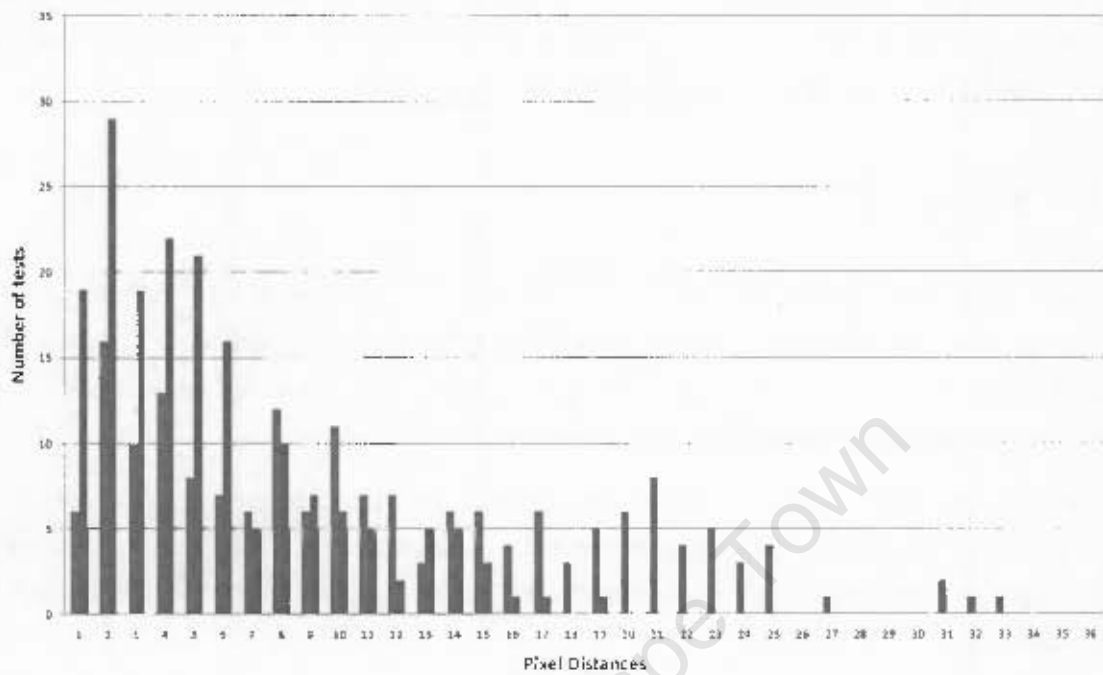


Figure 6.12: Histogram of horizontal and vertical pixel distances between ground truth and estimated marked region shown as red and blue respectively.

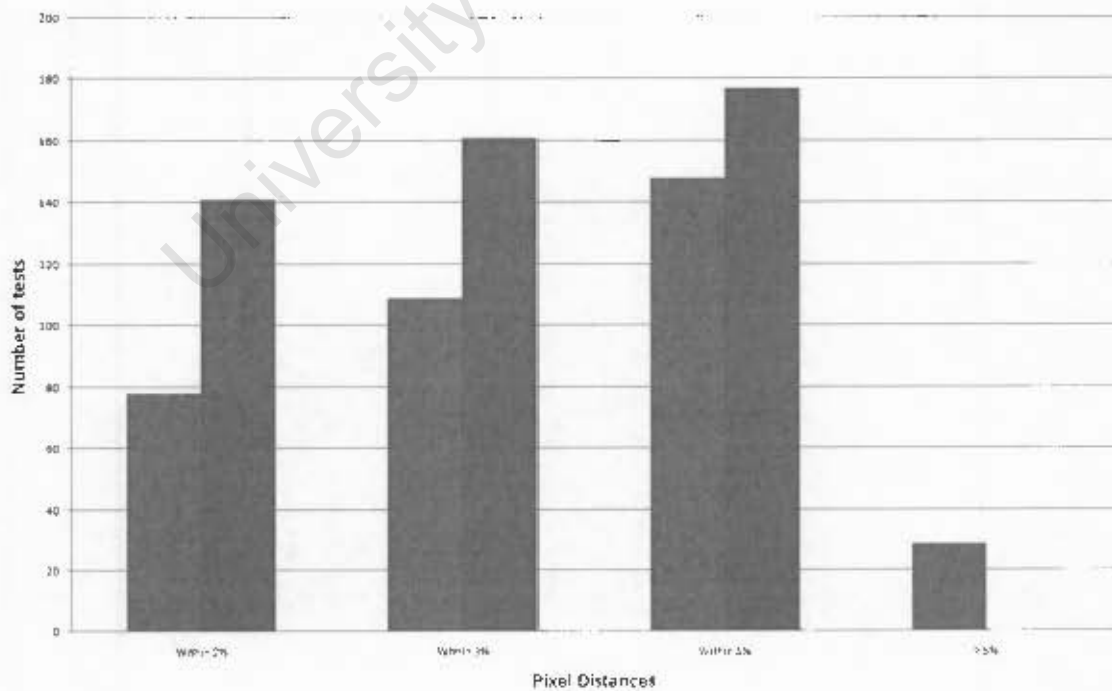


Figure 6.13: Results of passed tests within their respective accuracies where red and blue show the horizontal and vertical directions respectively.

The results shown in figure 6.12 and 6.13 show that horizontal location estimates are more accurate than vertical ones. Additionally, no horizontal distance errors were greater than 5%. Therefore, the online search performs better horizontally than it does vertically.

6.4.2 Failed Tests

An evaluation of the experiment showed that there are two main reasons why the tests fail. The first is that the region of interest could not be found, due to the insufficient number of matches. The other reason is that the distance between the ground truth and the estimated marked region is greater than 5%, 50mm, which is then regarded as the incorrect region found.

The same experiment is considered with the recommended thresholds except now only the failed tests are evaluated. An evaluation of this experiment shows that the failed tests are broken up into two categories. Approximately 40% resulted in the incorrect region found and 60% did not find the marked region.

Chapter 7

Conclusions

The most significant result is that it is possible to automate the search for a region of interest on a real-time medical scanner. After performing an experiment consisting of 260 tests, it has been found that it is possible to locate a region on the body, marked by the radiographer, with the aid of a camera attached to a c-arm.

Due to the nature of the proposed setup, a look-ahead distance of 435.13mm is available when the c-arm returns upwards, whereas only 179.13mm is available if the c-arm returns downwards. Therefore, to cater for larger regions of the body being marked a recommendation is made that the c-arm's first pass start from the top moving downwards and from the bottom moving upwards for its second pass.

The main factors that influence the search are the thresholds placed on the number of matches required, the search threshold, and the reliability of the feature points. An evaluation on the various thresholds, which consisted of varying the threshold values, was performed in order to see the impact on the results and a recommended value was provided for each threshold. Taking the recommended threshold values into consideration, the results are found to have an overall performance of 57%, of which 48% and 72% were within 2% and 3% accuracy.

It was also found that certain regions of the body were easier to locate than others. When ignoring the regions that were harder to locate, namely the abdomen, the chest and the head, the overall performance increased to 67%, of which 49% and 73% were within 2% and 3% accuracy respectively. Therefore, by ignoring regions that are harder to locate, the number of regions found successfully increases but the accuracy remains similar. This shows that certain regions of the body are easier to locate than others and that the proposed online searching method is approximately 49% and 73% within 2% and 3% accuracy respectively.

A more detailed experiment was performed to determine whether horizontal or vertical matches had a greater impact on the results. It was found that the distances between the ground truth and the estimated location of the marked region were closer in the horizontal direction than they were in the vertical.

Finally, an analysis was done on the tests that failed and it was found that approximately 60% of the failed tests were due to the region not being found and 40% due to a region found with an error greater than 5%.

Actual data could not be acquired as the Lodox Versascan is still under development. This was overcome by mimicking the Versascan environment as closely as possible and obtaining datasets accordingly. Once the Versascan is operational, it would be of interest to acquire datasets from the actual device and compare them to the results found in this report. As this experiment has shown, the method used to perform an online search to locate the region of interest is moderately successful.

University of Cape Town

Chapter 8

Further Research

8.1 Searching Algorithms

The searching method used for this project was one of many combinations of feature-based and template-based matching methods. Other combinations could be used to potentially achieve similar results.

8.2 Camera

The camera that was used for the project was the Point Grey Firefly MV FFMV-03M2C. The camera properties that had the greatest influence were the frame rate and lens field of view. Any other camera may be used within similar research, but in order to obtain comparable results, these two properties are important.

8.3 3-Dimensional Reference Image

Both the reference image and the template searching were performed on two-dimensional images. This was adequate for locating the region of interest and signaling the medical scanner when to turn on and off. However, there is scope to increase the number of dimensions to three by attaching an additional camera.

Appendix A

Video Datasets

Video datasets were captured from two different ways. The first was using the horizontally orientated medical device, Lodox Statscan, and the other was using a vertically mounted garage door opener. The datasets from these two devices can be found on the attached DVD. Only a portion of all the datasets captured are included in the DVD due to its size. All video data captured can be found on the UCT DIP server.

Appendix B

Experimentation Results

The experiments mentioned in this project were tabulated in a spreadsheet format. Considering the large number of tests, the spreadsheet can be found in the DVD. Note that there is only one spreadsheet that contains all the results of the experiment mentioned in this thesis.

University of Cape Town

Appendix C

Code

The final folder contained in the DVD is the code that was used in this project.

University of Cape Town

Bibliography

- [1] HERCA Working Group 2. Facts and Figures Concerning the use of Full Body Scanners using X-rays for Security Reason. *Oslo HERCA Plenary Meeting*, June 2010.
- [2] M.B. Ahmad and T.S. Choi. Local Threshold and Boolean Function Based Edge Detection. *IEEE Transactions on Consumer Electronics*, 45(3):674–679, August 1999.
- [3] M. Aly. Face Recognition using SIFT Features. Computer Science Department, California Institute of Technology, 2006.
- [4] S. Ando. Image Field Categorization and Edge/Corner Detection from Gradient Covariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2):179–190, February 2000.
- [5] De Beers. SCANNEX X-ray Body Scanner, 2011. URL http://www.debeersgroup.com/ImageVault/Images/id_1893/scope_0/ImageVaultHandler.aspx.
- [6] K. Briechle and U.D. Hanebeck. Template matching using fast normalized cross correlation. In *Proc. SPIE 4387*, pages 95–102, 2001.
- [7] C. Schmid, R. Mohr and C. Bauckhage. Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, 37(2):151–172, June 2000.
- [8] Z. Chaoyang. Video Object Tracking using SIFT and Mean Shift. Master's thesis, Chalmers University of Technology, Sweden, 2011.
- [9] D. Csetverikov. Basic Algorithms for Digital Image Analysis. Institute of Informatics, Budapest, 2009.
- [10] K.G. Derpanis. The Harris Corner Detector. Technical report, York University, October 2004.
- [11] Diamond. Diamond Light Source Ltd, 2010. URL <http://www.diamond.ac.uk/>.
- [12] E. Nadernejad, S. Sharifzadeh and H. Hassanpour. Edge Detection Techniques: Evaluations and Comparisons. *Applied Mathematical Sciences*, 2(31):1507–1520, 2008.

BIBLIOGRAPHY

- [13] F. Zhao, Q. Huang and W. Gao. Image Matching by Normalized Cross-Correlation. *In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP2006*, 2: 14–19, 2006.
- [14] W. Forstner. A Feature-Based Correspondence Algorithm for Image Matching. *International Archives of Photogrammetry and Remote Sensing*, 26(3):150–166, 1986.
- [15] Point Grey. Firefly MV Data Sheet. May 2009. URL <http://vision.lusterinc.com/UploadFile/assistant/Camera/PointGrey/DS-FMVU.pdf>.
- [16] H. Deng, W. Zhang, E. Mortensen, T. Dietterich and L. Shapiro. Principal Curvature-Based Region Detector for Object Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, July 2007.
- [17] C. Harris and M. Stephens. A Combined Corner and Edge Detector. *In Alvey Vision Conference*, pages 147–151, 1988.
- [18] J. van der Weijer, T. Gevers and J.M. Geusebroek. Edge and Corner Detection by Photometric Quasi-Invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4): 625–630, April 2005.
- [19] L. Cole, D. Austin and L. Cole. Visual Object Recognition using Template Matching. *Australian Conference on Robotics and Automation*, 2004.
- [20] L. Lei. Three Dimensional Shape Retrieval using Scale Invariant Feature Transform and Spatial Restrictions. Technical report, National Institute of Standards and Technology, August 2009.
- [21] J.P. Lewis. Fast Normalized Cross-Correlation. *Vision Interface, Canadian Image Processing and Pattern Recognition Society*, pages 120–123, 1995.
- [22] D.G. Lowe. Object Recognition from Local Scale-Invariant Features. *International Journal of Computer Vision*, 2:1150–1157, September 1999.
- [23] D.G. Lowe. Local Feature View Clustering for 3D Object Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 1:682–688, December 2001.
- [24] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, January 2004.
- [25] Lodox Systems Ltd. Statscan Product Specifications and Physical Dimensions 2011. http://www.lodox.com/images/statscan_product_specs.pdf.
- [26] Micron Technology, Inc. 1/3-Inch, Wide-VGA CMOS Digital Image Sensor Data Sheet. 2006.

BIBLIOGRAPHY

- [27] P.A. Mlsna and J.J. Rodriguez. Gradient and Laplacian Edge Detection. *The Essential Guide to Image Processing (Second Edition)*, pages 495–524, 2009.
- [28] F. Mokhtarian and R. Suomela. Robust Image Corner Detection Through Curvature Scale Space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1376–1381, December 1998.
- [29] N. Gupta, R. Gupta, A. Singh and M Wytock. Object Recognition using Template Matching, 2008. URL <http://www.stanford.edu/class/cs229/proj2008>.
- [30] D. Parks and J.P. Gravel. Corner Detection. URL <http://www.cim.mcgill.ca/~dparks/CornerDetector/harris.ht>.
- [31] M. Pilu. A Direct Method for Stereo Correspondence Based on Singular Value Decomposition. *Computer Vision and Pattern Recognition*, pages 261–266, 1997.
- [32] I. Pitas. *Digital Image Processing Algorithms and Applications*. John Wiley and Sons, Inc., 2000.
- [33] The Glowing Python. Corner Detection with OpenCV, 2011. URL <http://glowingpython.blogspot.com/2011/10/corner-detection-with-opencv.html>.
- [34] RoboRealm. Harris Corners, 2011. URL <http://www.roborealm.com/help/Harris.php>.
- [35] RTstudents. Radiographic Routine Procedures Guide, 2011. URL <http://www.rtstudents.com/radiology-positioning.htm>.
- [36] S. Beningfield, H. Potgieter, A. Nicol, S. Van As, G. Bowie, E. Hering and E. Latti. Report on a New Type of Trauma Full-Body Digital X-ray Machine. *Emergency Radiology*, 10(1):23–29, April 2003.
- [37] S. Pal and P.K. Biswas. Modified Hausdorff Distance Transform Technique for Video Tracking. *Indian Conference on Computer Vision, Graphics and Image Processing*, 2000.
- [38] S. Se and P. Jasiobedzki. Stereo-Vision Based 3D Modeling and Localization for Unmanned Vehicles. *International Journal of Intelligent Control and Systems*, 13(1):46–57, March 2008.
- [39] S. Wesolkowski and E. Jernigan. Color Edge Detection in RGB Using Jointly Euclidean Distance and Vector Angle. *In Proc. of the IAPR Vision Interface Conference*, pages 19–21, May 1999.
- [40] Z. Hua, Y. Li and J. Li. Image Stitch Algorithm Based on SIFT and MVSC. *Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, pages 2628–2632, August 2010.

BIBLIOGRAPHY

- [41] Z. Zhang, R. Deriche, O. Faugeras and Q.T. Luong. A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. *Artificial Intelligence Journal*, 78:87–119, 1995.

University of Cape Town