AN INVESTIGATION INTO MOTION CORRECTION
SCHEMES FOR HIGH RESOLUTION 3D PET AND PET/CT

A THESIS
SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY (PhD)
IN THE FACULTY OF MEDICAL AND HUMAN SCIENCES

PHILIP JOHN NOONAN

SCHOOL OF MEDICINE
2014
## CONTENTS

Abstract ................................................................................................................. 13

Declaration .............................................................................................................. 15

Copyright .................................................................................................................. 17

Acknowledgements ................................................................................................. 19

1 Thesis Introduction and Overview ......................................................................... 21
   1.1 General Thesis Introduction ............................................................................. 21
   1.2 General Thesis Overview .................................................................................. 22

2 PET Physics ............................................................................................................ 25
   2.1 Introduction ...................................................................................................... 25
   2.2 Emission and Detection .................................................................................... 26
   2.3 PET Scanners .................................................................................................. 30
   2.4 PET Image Reconstruction .............................................................................. 30
   2.5 Conclusion ....................................................................................................... 33

3 Motion Tracking and Correction ............................................................................ 35
   3.1 Introduction ...................................................................................................... 35
   3.2 Motion Tracking for Brain PET ........................................................................ 37
      3.2.1 Data Driven Motion Measuring ................................................................. 37
      3.2.2 Marker Based Motion Tracking ............................................................... 39
      3.2.3 The Polaris Position Sensor ..................................................................... 40
      3.2.4 Alternative Marker Based Motion Tracking .......................................... 46
      3.2.5 Markerless Motion Tracking .................................................................. 49
      3.2.6 Spatial Registration Using Iterative Closest Point .................................. 60
   3.3 Motion Tracking for Whole Body PET ............................................................ 61
      3.3.1 Data Driven Gating ................................................................................... 61
      3.3.2 Marker Based Motion Tracking ............................................................... 62
      3.3.3 Markerless Surface Measuring ................................................................ 64
6.3.2 Co-Registering Single Kinect Frames and a Kinectfusion Generated Surface ........................................ 111
6.3.3 Realtime Registration using the Kinect ...................... 111
6.4 Discussion ......................................................... 118
6.5 Conclusions ...................................................... 121

7 Clinical Motion Tracking Using Kinect .......................... 123
7.1 Introduction ...................................................... 123
7.2 Methods .......................................................... 125
  7.2.1 Software Advancements .................................... 125
  7.2.2 Hardware Changes .......................................... 125
  7.2.3 Scanning Protocol .......................................... 127
7.3 Results ............................................................. 128
7.4 Discussion ........................................................ 129
7.5 Conclusion ........................................................ 132

8 Multiple Target Marker Tracking ................................. 133
8.1 Introduction ...................................................... 133
8.2 Methods .......................................................... 135
  8.2.1 Software .................................................... 135
  8.2.2 Hardware ................................................... 136
  8.2.3 Marker Tracking Accuracy ................................. 136
  8.2.4 Marker to PET Calibration ................................ 137
  8.2.5 Stability Assessment ...................................... 138
  8.2.6 Volunteer Tracking ........................................ 140
8.3 Results ............................................................. 140
  8.3.1 Marker Tracking Accuracy ................................ 141
  8.3.2 Marker to PET Calibration ................................ 143
  8.3.3 Stability Assessment ...................................... 145
  8.3.4 Volunteer Tracking ........................................ 147
8.4 Discussion ........................................................ 147
8.5 Conclusion ........................................................ 152

9 Respiratory Tracking ................................................. 153
9.1 Introduction ...................................................... 153
9.2 Methods .......................................................... 155
  9.2.1 Single Point Motion Tracking ............................. 156
  9.2.2 Validating Respiratory Motion Tracking ................. 157
  9.2.3 Augmented Respiratory Monitoring ..................... 158
  9.2.4 Oncology Patient Data ..................................... 159
9.3 Results ............................................................. 160
## CONTENTS

9.3.1 Respiratory Gating .............................................. 160
9.3.2 Augmented Respiratory Monitoring ............................. 162
9.3.3 Oncology Patient Data ........................................... 164
9.4 Discussion ............................................................ 165
9.5 Conclusion ............................................................. 167

### 10 Future Prospects .................................................. 169

10.1 Thesis Summary ....................................................... 169
    10.1.1 Limitations of Work ........................................ 171
10.2 Future Kinect Projects in Medical Imaging ....................... 171
    10.2.1 Internal organ motion correction ........................... 172
    10.2.2 3D model for electrical impedance tomography ........... 172
    10.2.3 3D model for x-ray mammography .......................... 173
10.3 Research Outcome .................................................... 174
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Positron in a cloud chamber</td>
<td>26</td>
</tr>
<tr>
<td>2.2</td>
<td>The types of events that can occur during PET imaging</td>
<td>28</td>
</tr>
<tr>
<td>2.3</td>
<td>A simple example showing filtered back projection image reconstruction</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>Motion of the head due to breathing</td>
<td>36</td>
</tr>
<tr>
<td>3.2</td>
<td>Centroid planogram data</td>
<td>38</td>
</tr>
<tr>
<td>3.3</td>
<td>Randoms rate and centroid data from h00634</td>
<td>38</td>
</tr>
<tr>
<td>3.4</td>
<td>Photograph of the Polaris position sensor</td>
<td>41</td>
</tr>
<tr>
<td>3.5</td>
<td>The pinhole camera model</td>
<td>42</td>
</tr>
<tr>
<td>3.6</td>
<td>Diagram illustrating epipolar geometry</td>
<td>43</td>
</tr>
<tr>
<td>3.7</td>
<td>Polaris vicra images with epipolar lines</td>
<td>44</td>
</tr>
<tr>
<td>3.8</td>
<td>Stereo calibration of two cameras can be achieved using multiple images of a chessboard pattern</td>
<td>44</td>
</tr>
<tr>
<td>3.9</td>
<td>The Polaris Vicra tracking tool as used at the WMIC</td>
<td>45</td>
</tr>
<tr>
<td>3.10</td>
<td>Marker patterns being tracked with a stereo-optical camera</td>
<td>47</td>
</tr>
<tr>
<td>3.11</td>
<td>Marker patterns tracked in 3D using a single, calibrated digital camera</td>
<td>47</td>
</tr>
<tr>
<td>3.12</td>
<td>The pose and orientation of a target marker estimated using a calibrated camera and prior knowledge of the size of the marker</td>
<td>48</td>
</tr>
<tr>
<td>3.13</td>
<td>Motion tracking of a 5 mm$^2$ target marker</td>
<td>49</td>
</tr>
<tr>
<td>3.14</td>
<td>Detecting features on a taxidermied rat head</td>
<td>51</td>
</tr>
<tr>
<td>3.15</td>
<td>Inking patterns on mice heads allow for more features to be detected on rigid regions of the head</td>
<td>51</td>
</tr>
<tr>
<td>3.16</td>
<td>Playstation4 stereo camera with 2 high definition cameras.</td>
<td>52</td>
</tr>
<tr>
<td>3.17</td>
<td>The PMD Camboard Nano is a small, TOF depth sensor</td>
<td>52</td>
</tr>
<tr>
<td>3.18</td>
<td>Three colour photometric stereo</td>
<td>53</td>
</tr>
<tr>
<td>3.19</td>
<td>Simple Swept Line 3D Scanner</td>
<td>55</td>
</tr>
<tr>
<td>3.20</td>
<td>Temporal Encoding Structured Light Patterns</td>
<td>56</td>
</tr>
<tr>
<td>3.21</td>
<td>Structured light using a series of 3 striped line patterns</td>
<td>56</td>
</tr>
<tr>
<td>3.22</td>
<td>Kinect patent illustrations</td>
<td>57</td>
</tr>
<tr>
<td>3.23</td>
<td>Illustration showing depth from disparity</td>
<td>58</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.24</td>
<td>Diagram showing Kinect speckle pattern, depth map, and coloured point cloud.</td>
<td>60</td>
</tr>
<tr>
<td>3.25</td>
<td>Varian RPM tracking system</td>
<td>63</td>
</tr>
<tr>
<td>3.26</td>
<td>Anzai pressure belt</td>
<td>64</td>
</tr>
<tr>
<td>3.27</td>
<td>Diagram highlighting the need to extrapolate a motion corrected LOR</td>
<td>68</td>
</tr>
<tr>
<td>3.28</td>
<td>Diagram showing the possible effect of motion correcting an oblique LOR close to the edge of the FOV</td>
<td>68</td>
</tr>
<tr>
<td>3.29</td>
<td>Comparison of various motion correction schemes on brain PET data</td>
<td>70</td>
</tr>
<tr>
<td>4.1</td>
<td>Prototype new tracking tool attached to goggles</td>
<td>77</td>
</tr>
<tr>
<td>4.2</td>
<td>Polaris frame rate counter</td>
<td>80</td>
</tr>
<tr>
<td>4.3</td>
<td>Meshlab ICP tool alignment</td>
<td>81</td>
</tr>
<tr>
<td>4.4</td>
<td>Graph showing the Polaris clock drift</td>
<td>82</td>
</tr>
<tr>
<td>4.5</td>
<td>Example of a frame by frame motion corrected phantom with motion tracking using the Polaris Vicra</td>
<td>83</td>
</tr>
<tr>
<td>5.1</td>
<td>Motion tracking of subject h00262_01</td>
<td>90</td>
</tr>
<tr>
<td>5.2</td>
<td>Motion tracking of subject h00228_01</td>
<td>91</td>
</tr>
<tr>
<td>5.3</td>
<td>Motion tracking of subject h00207_01</td>
<td>92</td>
</tr>
<tr>
<td>5.4</td>
<td>TAC analysis of h00207_01</td>
<td>93</td>
</tr>
<tr>
<td>5.5</td>
<td>Motion tracking of subject GSK111013_000107</td>
<td>94</td>
</tr>
<tr>
<td>5.6</td>
<td>Contour analysis of subject GSK111013_000107</td>
<td>94</td>
</tr>
<tr>
<td>5.7</td>
<td>Motion tracking of subject GSK111013_000102</td>
<td>95</td>
</tr>
<tr>
<td>5.8</td>
<td>Contour analysis of subject GSK111013_000102</td>
<td>95</td>
</tr>
<tr>
<td>5.9</td>
<td>Motion tracking of subject GSK111013_000106</td>
<td>96</td>
</tr>
<tr>
<td>5.10</td>
<td>Contour analysis of subject GSK111013_000106</td>
<td>96</td>
</tr>
<tr>
<td>5.11</td>
<td>Comparison of FBP and AW-OSEM contour plots</td>
<td>99</td>
</tr>
<tr>
<td>6.1</td>
<td>Comparison of Meshlab and Point Cloud Library implementations of ICP</td>
<td>103</td>
</tr>
<tr>
<td>6.2</td>
<td>Overview of KinectFusion algorithm</td>
<td>105</td>
</tr>
<tr>
<td>6.3</td>
<td>Illustration Of KinectFusion global volume</td>
<td>106</td>
</tr>
<tr>
<td>6.4</td>
<td>Photograph of head phantom</td>
<td>107</td>
</tr>
<tr>
<td>6.5</td>
<td>Positioning the Kinect to acquire head motion tracking requires a mirror to be positioned on the PET/CT scanner.</td>
<td>108</td>
</tr>
<tr>
<td>6.6</td>
<td>KinectFusion scan of a flat square test object</td>
<td>111</td>
</tr>
<tr>
<td>6.7</td>
<td>Single frames of Kinect data being registered to a KinectFusion generated template.</td>
<td>112</td>
</tr>
<tr>
<td>6.8</td>
<td>Motion tracking parameters showing drifting without the use of the mirror</td>
<td>114</td>
</tr>
<tr>
<td>6.9</td>
<td>Motion tracking using the Kinect with a mirror</td>
<td>115</td>
</tr>
<tr>
<td>6.10</td>
<td>Simultaneous motion tracking using the Polaris Vicra and the Kinect</td>
<td>116</td>
</tr>
<tr>
<td>6.11</td>
<td>Simultaneous motion tracking using the Polaris Vicra and the Kinect in each axis</td>
<td>117</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>6.12</td>
<td>Kinect tracking of a static phantom shows no drift.</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>Application interface of KinectFusion tracking</td>
<td></td>
</tr>
<tr>
<td>7.2</td>
<td>Mirror fixation devices to the PET/CT gantry</td>
<td></td>
</tr>
<tr>
<td>7.3</td>
<td>KinectFusion meshes of first clinical PET scans</td>
<td></td>
</tr>
<tr>
<td>7.4</td>
<td>KinectFusion motion tracking of first clinical PET scans</td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td>Initial 1.8 minutes of KinectFusion tracking of subject 0605a</td>
<td></td>
</tr>
<tr>
<td>7.6</td>
<td>Displacement plot showing head motion due to breathing</td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>Diagram showing HRRT profile with the line of sight of Kinect and webcam</td>
<td></td>
</tr>
<tr>
<td>8.2</td>
<td>Tracking application window</td>
<td></td>
</tr>
<tr>
<td>8.3</td>
<td>Point source target marker</td>
<td></td>
</tr>
<tr>
<td>8.4</td>
<td>Illustration of the perspective view of flat and curved markers</td>
<td></td>
</tr>
<tr>
<td>8.5</td>
<td>Position of three target markers on hairline of volunteer</td>
<td></td>
</tr>
<tr>
<td>8.6</td>
<td>Camera resolution test image and profile</td>
<td></td>
</tr>
<tr>
<td>8.7</td>
<td>Separation of markers in fixed array</td>
<td></td>
</tr>
<tr>
<td>8.8</td>
<td>Rotation of markers in fixed array</td>
<td></td>
</tr>
<tr>
<td>8.9</td>
<td>Plot of marker positions during marker to PET space calibration</td>
<td></td>
</tr>
<tr>
<td>8.10</td>
<td>Tracking of three static target markers</td>
<td></td>
</tr>
<tr>
<td>8.11</td>
<td>Affect of varying lighting on tracking</td>
<td></td>
</tr>
<tr>
<td>8.12</td>
<td>Motion and separation of 3 target markers worn by volunteer without facial changes</td>
<td></td>
</tr>
<tr>
<td>8.13</td>
<td>Motion and separation of 3 target markers worn by volunteer with facial changes</td>
<td></td>
</tr>
<tr>
<td>9.1</td>
<td>KinectFusion image of a volunteer inside a PET/CT scanner.</td>
<td></td>
</tr>
<tr>
<td>9.2</td>
<td>The Kinect point cloud of the torso is used to monitor respiratory motion.</td>
<td></td>
</tr>
<tr>
<td>9.3</td>
<td>Multiple target markers are used to monitor the motion of various regions on the body.</td>
<td></td>
</tr>
<tr>
<td>9.4</td>
<td>Simultaneous motion tracking using the Kinect and the Varian RPM.</td>
<td></td>
</tr>
<tr>
<td>9.5</td>
<td>Simultaneous motion tracking of a phantom using the Kinect and the Anzai pressure belt.</td>
<td></td>
</tr>
<tr>
<td>9.6</td>
<td>Simultaneous motion tracking of a volunteer using the Kinect and the Anzai pressure belt.</td>
<td></td>
</tr>
<tr>
<td>9.7</td>
<td>The spatial sensitivity of the Kinect allows for the abdominal aortic pulse to be observed.</td>
<td></td>
</tr>
<tr>
<td>9.8</td>
<td>Simultaneous tracking of the Kinect and the target marker</td>
<td></td>
</tr>
<tr>
<td>9.9</td>
<td>Multiple target marker tracking of 6 points on the chest</td>
<td></td>
</tr>
<tr>
<td>9.10</td>
<td>Kinect data can be used to produce a range of quality of images depending on application need.</td>
<td></td>
</tr>
<tr>
<td>9.11</td>
<td>Anzai tracking of Patient A07062013</td>
<td></td>
</tr>
<tr>
<td>9.12</td>
<td>Single frame reconstruction from 10 minute listmode</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF TABLES

2.1 PET Radioisotopes .................................................. 27
2.2 PET scanner features ............................................. 30

5.1 Summarised motion tracking results from WMIC scans .... 93
5.2 Summarised contour analysis of motion tracking results from CIC scans . 93
5.3 Voxel variation in time analysis ................................ 98

6.1 ICP residuals for head phantom ................................. 113
6.2 RMS Hausdorff distance between frame pairs ............... 113
6.3 Standard Deviation of Hausdorff distance .................. 113
6.4 Summary of results from Kinect motion tracking without the use of the mirror. ........................................ 115
6.5 Summary of Kinect motion tracking using a mirror .......... 116
6.6 Range and standard deviation of long term kinect tracking under ideal conditions ........................................ 119

8.1 Mean and Standard Deviation of the Separation of the 4 Markers .... 141
8.2 Rotational Root Mean Squared Difference from the 4 Markers .... 144
8.3 Root Mean Squared Difference Between Transformed Target Marker Position and the Measured Point Source Voxel Location .......... 144
8.4 Standard deviation of markers on a flat and curved surface .... 147
8.5 Standard deviation of rotation parameters for different sized target markers 147
Although motion correction in medical imaging is well established and has attracted much interest and research funding, a gap still exists in that there is a lack of reliable, low-cost hardware to enable such techniques to be widely adopted in healthcare. Motion correction of brain Positron Emission Tomography (PET) data for instance is an important step in realising the potential offered by modern high resolution PET scanners. Since it is not likely that subjects can remain stationary throughout the PET scan, which can last 60 minutes or more, accurate and reliable motion tracking is needed to correct the PET data for any observed motion. A commercially available marker based motion tracking system was evaluated and found to produce unreliable data. This was due to the possibility of the tracking tool slipping from the subject. This thesis describes the investigations into alternative and novel tracking techniques for use in PET. These included a markerless tracking system using the Microsoft Kinect (a low cost depth sensor) as well as a multiple target marker tracking system. The performance characteristics of both systems (low cost, high spatial and temporal accuracy, and real-time operation) were evaluated using phantom and clinical experiments. Investigations into using these two tracking techniques in whole body PET, specifically measuring the respiratory rate during lung imaging, were developed and compared against current commercially available solutions.
The University of Manchester
PhD by published work Candidate Declaration

Candidate Name: Philip John Noonan

Faculty: Medical and Human Sciences

Thesis Title: An Investigation Into Motion Correction Schemes for High Resolution 3D PET and PET/CT

Declaration to be completed by the candidate:

I declare that no portion of this work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning. In Chapter 5 the centroid analysis implementation was provided by Dr. Shail Segobin, and the automatic image registration was provided the Vinci software platform and performed by Jose Anton. In Chapter 6, the Kinect drivers was provided by the by the OpenNI initiative and the software was developed from the opensource Point Cloud Library. Engineering support was provided by Dr. Jon Howard and was essential in developing and constructing the prototype glasses, pulse counting circuit, and mirror mount. In Chapter 7, the Kinect drivers were provided by Microsoft and the software was also developed from sample Microsoft code. In Chapter 8, the software was developed from the opensource OpenCV library using examples from GoblinXNA and ALVAR. Chpater 9 contains examples from all previously mentioned driver and software sources.

Signed:                                  Date: 28th April 2014
The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property” and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproduction”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of Theses.
It is a truth universally acknowledged, that a young researcher in possession of an inquisitive mind must be in want of a PhD. However, it is also true that without the immense support and encouragement from many close friends and colleagues this thesis would have remained a distant target. The boundary between my friends and my colleagues is very blurred, indeed in a thesis concerned with reducing blur in medical images, it may appear defeatist to begin with such a blurred view on who to thank. Regardless, being a Physicist, I concluded that the simplest method to thank everyone who has helped me get to this point was an alphabetised list.


My thanks go out to everyone mentioned here. Without your discussions (be they work related or about the local sports teams) I doubt I would be as well informed as I am now.
ACKNOWLEDGEMENTS

In particular, my thanks go out to my sister, MaryAnn. You started me on my academic path and have been invaluable by proof reading my work, setting down benchmarks, and generally always being there as a mentor. To my wife and lifelong best friend, Laura, I could not have completed this part of my life without you and your continual support; here’s to our future! To my parents and family, thank you for raising me in a happy home and encouraging me to grow with a curious mind. I would finally like to express my sincere thanks to my supervisory team; Rainer, Will, Peter, and Tim. I thank Rainer for allowing me the investigative freedom to explore strange new technologies and I thank Will who successfully focused my mind and efforts on bringing my ideas into clinical use. I fondly remember winning the best team prize with Peter at the Turku PET symposium in 2011. I feel very fortunate to have been supervised by Tim, indeed without his Matlab lectures I doubt I would have developed a passion for programming and computer vision.

I wish the best of luck to my very close friends Jose, Sophie, and Joe who are currently working towards their PhD. Your friendship made the last few years a pleasure.

I look back with many fond memories of my time at the WMIC, CIC, and Imanova, and I look forward to an exciting future.
GENERAL THESIS INTRODUCTION AND OVERVIEW

1.1 General Thesis Introduction

This thesis tackles the challenge of accurately measuring patient movement during medical imaging in order to allow correction for motion induced artifacts. The work presented in this thesis spans multiple scientific fields including medical imaging, optical physics, and computer vision. The research outcomes that were developed were only possible by having a working knowledge of how these interdisciplinary fields can interact. The work in this thesis was focussed on motion tracking for Positron Emission Tomography, PET, a nuclear medicine imaging modality. Since PET scans can last upwards of an hour, it is vitally important to have an accurate and reliable motion correction scheme in place to correct for the movements that are guaranteed to occur, such as breathing and subject repositioning.

In brain PET, the motion of the head can be said to be of a rigid body, i.e. the brain does not deform during movements. In whole body PET, the internal organs such as the heart and lungs undergo large deformations, and as such motion correction for whole body PET is considerably more challenging than for brain PET. In brain PET, head restraints are often used to minimise head motion, however the more they restrain the head the more likely the restraints are to cause discomfort to the subject. For whole body PET, the cardiac and respiratory cycles are unavoidable and so a motion correction scheme has to be used to reduce the effect of blurring.

There has been much research interest in different motion correction algorithms, however there has been relatively few attempts at developing a novel motion tracking technique. Some motion correction algorithms generate motion tracking information from the PET data itself, however care has to be taken to ensure that the motion that is observed in the PET data is actually due to the subject moving, rather than the radiolabelled tracer redistributing during the scan. The inherently noisy property of PET data also prevents high temporal resolution motion correction in data driven techniques. An alternative to determining the motion from the PET data is to use an external tracking device. External tracking devices have potentially higher spatial and temporal resolution motion tracking than data
driven techniques. Since they are tracer independent, the quality of motion tracking will not be affected if the PET signal is low. Accurate motion correction is essential for brain PET as modern PET scanners can achieve spatial resolutions of $\sim 2$ mm$^3$.

The motivation for the work in this thesis came about from the lack of a reliable solution to track and correct for head motion in brain PET. Commercial motion tracking systems were available but many PET research centres worldwide have struggled to optimise the tracking system’s integration into a routine PET scanning protocol. The Polaris Position Sensor is a commonly used motion tracking device that monitors a tool that needs to be worn by the subject. Two major problems existed with the commercially available system; describing the motion tracking data in the PET image frame of reference, and the difficulty in ensuring that the tracking data is reliable by preventing the tracking tool from slipping during use. These issues have prevented the motion data from external tracking systems being successfully used in the state-of-the-art motion correction algorithms. The main research aim in this thesis was to provide an accurate and reliable motion tracking solution for PET.

To meet this research objective, the problem was first approached by evaluating the currently used commercial tracking systems. A process was developed to spatially and temporally register the tracking data to the PET data and effort was put into investigating a robust method for attaching the tracking tool to the subject’s head. The results from these investigations into the commercial solution prompted the development of two novel tracking techniques for use in brain PET. The first system used the Microsoft Kinect, a low cost depth sensor, and a real time registration algorithm, KinectFusion, to track rigid body motion without using tracking tools or markers. The second system used multiple, small, printed target markers fixed to the hairline and tracked using a webcam and open-source computer vision algorithms in order to robustly track head motion. These systems are both novel, operate in real time, and are low cost, using consumer grade hardware and opensource software. The investigations into motion tracking were extended to include whole body motion tracking. The knowledge gained from developing head tracking using the Kinect and multiple target markers allowed for tracking systems to be designed for monitoring respiratory motion. These systems were evaluated against commercially available solutions for tracking respiratory motion.

1.2 General Thesis Overview

This thesis generally follows the chronological progression of work over three years of research starting with an introduction of Positron Emission Tomography and motion (Chapters 2 and 3). These two chapters aim to provide the understanding of the rationale of this work by describing the physics of PET imaging and the crucial role that motion tracking and motion correction has in producing highly quantifiable and qualitative image data. Chapter 3, in particular, attempts to describe the current field of motion correction
1.2. GENERAL THESIS OVERVIEW

in PET through reviews of the relevant literature from the last decade since PET scanners achieved high enough spatial resolution that motion became one of the most significant barriers to high resolution imaging. Where appropriate, specific devices and techniques are discussed to ensure that the research from Chapter 4 onwards can be understood. Crucially for the direction of these investigations, Chapter 3 highlights the current gap in the motion correction pipeline; accurate and reliable motion tracking. Many algorithms have been developed that elegantly correct for motion, however the majority require the tracked motion to be perfectly representative of the motion which occurred during the scan.

Chapter 4 is the first of six research chapters and begins with a description of the work done to integrate the Polaris position sensor into an automated frame by frame realignment motion correction scheme. Chapter 4 highlights the importance of the design of the tracking tool fixation device, as this is the main factor that reduces the reliability of the Polaris tracking technique. In Chapter 5, results from clinical brain PET studies that had been tracked using the Polaris system are presented and discussed. These data sets were obtained from the Wolfson Molecular Imaging Centre, University of Manchester, Manchester, UK, and the Clinical Imaging Centre, GlaxoSmithKline, Hammersmith, London, UK. This chapter presents solutions to the difficult task of designing a tool fixation modality that does not allow for tool slippage, yet is sufficiently comfortable for the subject to wear for the duration of the scan. Chapter 6 represents the beginning of the novel investigations into developing a markerless motion tracking system. The launch of the Microsoft Kinect in 2011 prompted investigations which sought to determine whether this new hardware would be appropriate to the needs of a markerless motion tracking system. Chapter 6 proceeds by validating the use of the Kinect as a motion tracking device by first registering single Kinect frames using an offline registration algorithm, then by using a real time registration algorithm which uses a graphics processing unit. Chapter 7 documents the first use of the real time registration algorithm in a clinical PET study and describes the steps taken to ensure that motion tracking is achieved without interference to the subject. Chapter 8 details the investigations into an advanced marker based approach for use in scanners such as the High Resolution Research Tomograph (HRRT), whereby the geometry of the bore prevents the use of the Kinect. Additional work was undertaken using the Kinect for respiratory monitoring for whole body PET and is described in Chapter 9, the final research chapter. The thesis is concluded in Chapter 10 with an overview of the research outcomes and a look forward as to how the work presented here could be used in the wider field of medical imaging.
AN OVERVIEW OF THE PHYSICS OF POSITRON EMISSION TOMOGRAPHY

2.1 Introduction

Positron Emission Tomography (PET) is a medical imaging technique which produces 3D volumetric images based on detecting the distribution of a positron emitting radionuclide within a volume or body. The radionuclide can form part of a drug or compound to create radiopharmaceuticals which can be injected or inhaled by the patient and, depending on its biological function, it is distributed by the body enabling different regions or systems within the body to be imaged. A PET event is triggered when the radioisotope undergoes $\beta^+\,$ decay and emits a positron which, after travelling $\sim 1\,$ mm, collides and annihilates with an electron from the surrounding material. This produces two annihilation photons which are emitted at nearly $180^\circ\,$ separation. A PET scanner generally consists of rings of detectors that record the location of the co-incident pairs of annihilation photons that have escaped the body and then are absorbed by the detectors. The line that connects the detection points is known as a line of response (LOR) on which, or very close to, the original event occurred. Over the course of a PET scan, $10^6$ to $10^{10}$ events will be detected and passed to a 3D reconstruction algorithm. State of the art reconstruction algorithms can correct for many image degrading factors to produce images where the signal intensity inside each voxel (3D analogy of a pixel) directly correlates to the activity of the radiopharmaceutical at that point in the body. A strategic benefit of PET scanning is the versatility of the radiopharmaceutical used, allowing different areas and systems within the body to be investigated. PET is a significant imaging modality due to its ability in obtaining functional data in a qualitative and quantitative manner. During a clinical scan, blood tests are performed regularly to measure the concentration of the tracer so that the physiological processes which uptake the tracer in the body can be induced. With knowledge of the injected activity, the tracer distribution can be quantifiably determined from the PET images.

The remainder of this chapter aims to describe the basic physical properties of PET scanning to secure a working knowledge of the challenges of producing high quality med-
ical images.

2.2 Emission and Detection

Figure 2.1: Photograph of a cloud chamber experiment performed by Carl D. Anderson in 1933 showing the path of a positively charged particle but with the mass to charge ratio of an electron, figure taken from [Anderson 1933]

In 1928, Dirac proposed that electrons could have either negative or positive charge [Dirac 1928]. The search for a positively charged electron was concluded in 1933 by Anderson in his Nobel Prize winning cloud chamber experiments [Anderson 1933]. Figure 2.1 shows the curved path of a 63 MeV positively charged particle moving from the bottom half of the chamber, through 6 mm of lead and continuing its curved path to a distance at least 10 times greater than expected if the particle was a proton. The positron was the first antimatter particle to be discovered and is produced by the following 3-body decay equation,

\[ ^{A}ZX \rightarrow ^{A-1}Z^{-1}X + \beta^{+} + \nu_{e} \] (2.1)

The daughter nucleus, \( \beta^{+} \), and \( \nu_{e} \) have a range of possible energies after the decay event. The energy of the positrons produced by Anderson were much higher than those used for PET (63 MeV compared to \( \sim 1 \) MeV) as the energy of a positron determines how far it travels before annihilation.

The initial kinetic energy of the positron controls the distance between the initial decay site and the point of annihilation, as the positron needs to lose energy through inelastic scattering before it has sufficiently low enough energy to collide and bind with an electron to form positronium, the bound system of a positron and electron. The distance travelled by the positron before annihilation is known as the positron range, and creates the first finite limit on the spatial resolution available for PET imaging. In the absence of a strong magnetic field, the positron is emitted from the parent nucleus at a random angle and continues in a random walk until it forms positronium. The full width half maximum
(FWHM) of the positron range in water is also listed in Table 2.1

Table 2.1: Radioisotopes commonly used in PET imaging, their half lives, maximum energy and range in water of the emitted positron, adapted from [Levin et al. 1999]

<table>
<thead>
<tr>
<th>Radioisotope</th>
<th>Half Life</th>
<th>$E_{\text{max}}$</th>
<th>Range$_{\text{FWHM}}$</th>
<th>Range$_{\text{FWTM}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{18}$F</td>
<td>109.8 min</td>
<td>0.63 MeV</td>
<td>0.102 mm</td>
<td>1.03 mm</td>
</tr>
<tr>
<td>$^{11}$C</td>
<td>20.4 min</td>
<td>0.96 MeV</td>
<td>0.188 mm</td>
<td>1.86 mm</td>
</tr>
<tr>
<td>$^{15}$O</td>
<td>122 s</td>
<td>1.73 MeV</td>
<td>0.501 mm</td>
<td>4.14 mm</td>
</tr>
</tbody>
</table>

Positronium is an unstable system existing for $10^{-10}$ s [Colombino et al. 1965] [Karshenboim 2004] before matter and anti-matter annihilate converting the rest mass of the 511 KeV particles into two 511 KeV annihilation photons. Annihilation events where three daughter photons are produced occur at a lower rate than two photon events [Ore et al. 1949] and are not considered in conventional PET. In the simplest case whereby the system has no net momentum, i.e. the centre of mass frame of reference when the electron and positron are briefly at rest before they annihilate, the two photons will be produced at a $180^\circ$ separation. In the lab frame of reference, the positronium may still contain kinetic energy which results in an angular distribution which can be represented (in water) as a Gaussian with a mean of $180^\circ$ and FWHM of $\sim 0.5^\circ$ [Colombino et al. 1965][DeBenedetti et al. 1950][Berko et al. 1956]. This non-collinearity provides the second limiting factor in PET spatial resolution the effect of which increases with the distance separating the detectors. Before the photons can escape the body to be detected, there is a high probability of one or both of the photons interacting with the surrounding tissue. In the relatively low density case of human tissue, the scattering process is predominately the inelastic scatter of the photon off a free or loosely bound electron via Compton Scattering. Figure 2.2 lists the possible types of events that can occur during PET data acquisition. The 511 KeV photons are detected in coincidence by the PET detectors, and in the perfect scenario due to the well defined geometric relationship between the two photons, the location of the original event can be located to within the line of response. In reality due to factors such as those mentioned above, this location is at best very close to the line of response.

The majority of modern PET detectors use scintillating crystals to detect the annihilation photons. For 511 KeV photons the scintillating material needs to be optically thick so that the photons are likely to interact with the scintillator via the photoelectric effect. The incident high energy photon interacts with the dense scintillator, depositing its energy to the crystal lattice. The excited electrons occupy elevated states then decay back towards the ground state emitting optical photons at a characteristic spectrum. The scintillator Lutetium Oxyorthosilicate (LSO) is a common choice for modern PET scanners as it has high attenuation for 511 KeV photons, high light output per incident 511

---

Footnote: The full width tenth maximum (FWTM) is given alongside the FWHM as the positron range is non-Gaussian with a strong central peak with relatively wide tails.
CHAPTER 2. PET PHYSICS

Figure 2.2: Multiple types of events are detected. Scattered events are photons originating from the same event, but one or both of the photons have undergone Compton scattering. An energy loss is associated with the scattering, so highly scattered events can be screened by an energy window. Random events add significant noise to the data. They occur when one photon from 2 different events is lost and the resulting two photons are recorded as one single event by the coincidence controller. Other types of events include Single and Multiple events. A Single is where only one photon is detected and a multiple event is where more than 2 photons arrive within the time window. These are instantly discarded as there is no possibility for the co-incidence board to distinguish the true pair.

KeV photon, and a fast decay time allowing for high count rates to be used. The cascade of visible photons from the scintillator is guided down light channels cut into the crystal that controls the lateral spread of the visible photons. The visible light is amplified by photomultiplier tubes coupled to the scintillator block and the location of the annihilation photon point of interaction, as well as the energy of the photon, can be calculated from the differing voltage outputs of the photomultiplier tubes.

Events are processed on the coincidence controller board to determine event site pairs, or a coincident event, that have been triggered within a short, 6-9 ns time window. To measure the rate of random events, a parallel delayed time window is triggered from the pulse from one of the two detectors, therefore any coincident events that are measured in this delayed window can only be from random events. The rate of random events can be subtracted from the original (prompt + random) coincidence rate however this can lead to an increase in the statistical uncertainty of the true coincident signal, especially in situations with relatively high random events.

PET data can be recorded in histogram mode or listmode, the former being where the
events are binned on the fly into sinograms and the latter is when the events are sequentially recorded to file in a binary data stream. A listmode file consists of binary data that encodes the information for a coincident event such as the detector index, energy window, depth of interaction, time of flight bin, etc. Listmode data also contains tag words, such as time markers, block singles, gating markers, which are used in the image reconstruction process.

Before image reconstruction can take place, the raw PET data needs to be corrected for normalisation and attenuation. In emission tomography, data quality is affected by the tissue that the photons need to pass through before being detected outside of the body. The attenuation coefficient along a LOR is the linear sum of the individual attenuation coefficients of each emitted photon since as a pair the photons have had to traverse the same amount of tissue irrespective of the event location along the LOR. The Beer-Lambert law that describes the attenuation of electromagnetic radiation through material can be modified to allow for materials with regions of differing attenuation coefficients, as written in Eq. 2.2 as a probability of the photon escaping along a path of length \( L \) and attenuation \( \mu(x) \).

\[
p = \frac{N}{N_0} = e^{-\int_0^L \mu(x) dx}
\] (2.2)

For annihilation photon pairs originating at some point \( x \) along the LOR of length \( L \) the total attenuation of the pair is simply the product of the individual attenuation that each photon undergoes.

\[
p_{1+2} = e^{-\int_0^x \mu(x) dx} \cdot e^{-\int_x^L \mu(x) dx} = e^{-\mu L}
\] (2.3)

Therefore to correct for attenuation it is necessary to know the tissue density as it varies throughout the volume being imaged. In PET this is achieved by either using an external source of radiation (transmission scan) or by using a Computed Tomography (CT) scan. In a transmission scan, an external source is rotated around the object by rastering a collimated external source around the object being imaged. The PET camera is then set to acquire in single incidence mode where the known position of the transmission source gives the LOR to the detection site. Due to the fact that the activity of the transmission source is well known, a map of the attenuation (\( \mu \) map) can be generated and used to correct the emission scan. The external source provides transmission data which describes the total attenuation along each LOR and when compared to a similar external source scan without any attenuating object in the field of view (a blank scan) an attenuation map of the object is calculated. A CT scan provides high resolution anatomical information which can be interpolated into an attenuation map for the higher energy annihilation photons.
2.3 PET Scanners

Ideally a PET scanner would have detectors covering the entire $4\pi$ sr volume surrounding the object being scanned. The cost and practicality implications of building such a machine currently limit the design of a PET scanner to typically consist of a series of rings that each fully surround one axial plane. Four PET scanners were used during the investigations in this thesis; the CTI/Siemens High Resolution Research Tomograph\(^1\) (HRRT), and three types of Siemens Biograph, the HiRez\(^2\), TrueV\(^1,2\), and the mCT\(^3\). The HRRT is a brain PET only scanner whereas the Biographs are multimodal, whole body capable, PET/CT scanners. The HRRT consists of 8 planar detector heads arranged in an octagonal array, whereas the Biographs consists of rings of detector buckets. In the case of the HiRez the rings are in a barrel arrangement whereas for the TrueV and mCT the rings are in the shape of a cylinder. The mCT is the newest of the scanners that were used and can take advantage of Time of Flight information to improve PET data quality. With very fast detector electronics, TOF capable scanners, such as the mCT, are able to define the location of an event to one of thirteen separate sinogram bins that correspond to thirteen divisions along an LOR. This is achieved by comparing the arrival times of each event in a co-incident pair. Table 2.2 summarises the four different scanners.

Table 2.2: Cross scanner comparison between the four PET systems used in the investigations performed in this thesis. Spatial resolution data taken from [Jong et al. 2010][Brambilla et al. 2005][Jakoby et al. 2009][Jakoby et al. 2011].

<table>
<thead>
<tr>
<th></th>
<th>HRRT</th>
<th>HiRez</th>
<th>TrueV</th>
<th>mCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scintillating crystal</td>
<td>LSO/LSO</td>
<td>LSO</td>
<td>LSO</td>
<td>LSO</td>
</tr>
<tr>
<td>Spatial Resolution / mm</td>
<td>2.3</td>
<td>4.6</td>
<td>4.1</td>
<td>4.4</td>
</tr>
<tr>
<td>Axial FOV / mm</td>
<td>251</td>
<td>162</td>
<td>216</td>
<td>218</td>
</tr>
<tr>
<td>Transaxial FOV / mm</td>
<td>312</td>
<td>585</td>
<td>605</td>
<td>700</td>
</tr>
<tr>
<td>Number of detector elements</td>
<td>119808</td>
<td>24336</td>
<td>32448</td>
<td>32448</td>
</tr>
</tbody>
</table>

2.4 PET Image Reconstruction

This section aims to give an overview of the main methods used to solve the inverse problem of producing a PET image from a set of measured projection data from the PET scanner. This projection data describes the total number of counts detected along every LOR over the duration of the scan and under an ideal scenario is proportional to the line integral of the activity distribution that the LOR intersects. For further reading, excellent

\(^1\)Installed at the Wolfson Molecular Imaging Centre, University of Manchester, Manchester, UK
\(^2\)Installed at Imanova Centre for Imaging Sciences, Hammersmith, London, UK, formerly GlaxoSmithKline Clinical Imaging Centre
\(^3\)Installed at the Nuclear Medicine Centre, Central Manchester University Hospitals, Manchester, UK
2.4. PET IMAGE RECONSTRUCTION

overviews of the mathematics of PET image reconstruction can be found in the literature such as [Leahy et al. 2000] and [Qi et al. 2006].

The two main classes of image reconstruction algorithms are analytical and iterative. Backprojection is a simple example of an analytical method in which LORs are superimposed onto an image matrix. Since most LORs will not be aligned with this grid, the LORs are weighted by a factor determined by the path length of the LOR inside each voxel. The image matrix is then updated over all detector pairs which results in an image showing an approximation of the activity distribution, however the image is blurred, as activity has been assigned equally along the LOR.

Filtered backprojection (FBP) utilises the central section theorem that states that a 1D Fourier transform (FT) of a projection at an angle $\theta$ is equivalent to the 2D-FT of the image in the radial direction at angle $\theta$. This allows various filters to be applied in the frequency domain to the set of 1D-FT sinograms to suppress undesirable features of backprojection such as the $1/r$ blurring and rapid noise fluctuations. A simple example of a 2D FBP implementation is shown in Figure 2.3

![Figure 2.3](image)

Figure 2.3: A worked example of analytical reconstruction techniques is shown using projection data in the form of a 2D sinogram (a) generated from a greyscale image (d) using the Matlab Radon transform function. Each row of the (a) is Fourier transformed to form (b). Each row is then multiplied by the filter function to produce (c). The inverse Fourier transform is performed at this stage, followed by the backprojection into Cartesian space. The difference between standard backprojection (e) and filtered backprojection (f) is easily apparent.

There are many methods to extend 2D FBP into 3D which mainly operate by first rebinning the 3D projection data into a set of parallel sinograms with an azimuthal angle
of 0°. Single-slice rebinning places events into sinograms corresponding to their mean axial location. This technique is suitable for events located on the axis but fails for off-axial events. More advanced rebinning techniques exist such as Fourier rebinning (FORE) which aims to solve the resolution degradation for off axis events [Defrise et al. 1997] yet remain computationally inexpensive. Analytical reconstruction techniques are undeniably more efficient and are simpler than iterative methods, however since analytical algorithms cannot model the complexities of the emission and detection process their popularity has declined with the rise in affordable and powerful computers available for image reconstruction.

Iterative image reconstruction has arguably become the standard reconstruction technique as the computationally expensive demands of iterative algorithms have been addressed with modern computing hardware. There are many different flavours of iterative algorithms differing by factors such the generation of the system matrix and the choice of the cost function which is used to compare iterations. In brief summary, an iterative algorithm would begin by multiplying an initial estimate of the activity distribution $f$ by a projection matrix $M$ into sinogram space. This projected data set would be compared to the measured PET sinogram data $g$ and the resulting differences are used to modify the estimate of the activity distribution. This relationship between emitted and detected activity is shown in Equation 2.4 where matrix $M_{i,j}$ describes how the sinogram element $g_j$ is the sum over all image voxels of the probability of events being detected in that sinogram bin.

$$g_j = \sum_i M_{i,j} f_i$$ (2.4)

This process continues for a certain number of iterations or until a convergence criteria has been reached. Iterative algorithms can include corrections for various emission and detection features, such as the previously discussed positron range and photon non-collinearity, in addition to modelling the scanner geometry and detector characteristics. Collectively these resolution limiting effects are known as the point spread function (PSF) i.e. the systems blurred response to imaging a point source. By modelling the PSF in the system matrix $M$, iterative algorithms can go beyond the limitations of using the line-integral model and recover resolution potential.

The PSF can be simulated or directly measured by imaging a point source at all points inside the field of view of the scanner. Imaging a single point source over the entire field of view would be impractical due to time constraints and would need a robot arm to accurately and reliably move the point source through the field of view. A simpler method to measure the PSF by using arrays of point sources was developed by using a modified inkjet printer and a mixture of ink and radiotracer\(^1\). This allows for a spatially variant PSF to be measured as in [Kotasidis et al. 2011] for relatively little effort. With

\(^1\)Original idea and initial development was performed by D. Gorman and P. Noonan as part of a Masters project in 2008 at the WMIC
PSF modelling it is possible to achieve spatial resolutions of the order of 1-2 mm in the HRRT and 2-3 mm in PET/CT scanners, however this resolution is rarely met in practice as subject motion will unavoidably occur.

2.5 Conclusion

This chapter is designed to provide a general background on the PET process, from positron emission to image formation. This is by no means a comprehensive guide to every aspect of PET, rather particular attention has been given to features that have specific relevance to the work presented later in this thesis. For a recent review of PET, its past achievements, current state and future directions the reader is directed to [Jones et al. 2012]. In general, PET is only made possible by the combined efforts from many cross disciplinary scientists and clinicians, from the radiochemists who manufacture the radiotracers to the analytical modellers who analyse the PET data to obtain the quantitative information that makes PET such a valuable imaging tool. The work in this thesis is focussed on improving the quality of PET data by monitoring subject motion to enable accurate motion correction. Motion correction in PET is an essential step in producing useful quantitative data however it has often been difficult to succinctly integrate into routine clinical use as will be described in following chapter.
3.1 Introduction

Sophisticated imaging techniques are now widely used in healthcare. For example, 1.52 million PET or PET/CT scans were performed in the USA in 2008 [Hricak et al. 2010], yet their utility is undoubtedly limited by patient movement. As a challenging example, current PET scanners are capable of 2mm resolution but this is seldom achieved due to respiratory or involuntary movement during the scan. Correction for patient movement would lead to more accurate diagnosis, and, in oncology for example, better targeted radiotherapy treatment, with the aim of greater survival and fewer complications. Movement has long been a concern for PET studies in neurological research, and many research centres have developed motion tracking. Typically this involves attaching markers to the head and tracking these using range-finding equipment. However this is uncomfortable and distracting for the subject, and the markers often slip, making the motion data unusable. PET is also widely used clinically, for instance as an essential imaging modality to routinely characterise solitary pulmonary nodules in patients with lung cancer as recommended by the National Institute of Clinical Excellence (NICE) [White et al. 2011]. The spatial resolution of clinically used PET/CT scanners is not achieved in practice due to patient motion. This is a particular problem near the diaphragm where respiratory motion is typically around 2 cm, making accurate diagnosis or staging of tumours more difficult and can lead to additional Computed Tomography (CT) scans incurring costs and higher radiation doses. Patient motion measuring devices are now available that rely on patient contact to obtain respiratory gated images. However these were initially developed for radiotherapy patients who are first coached to produce a reliable respiratory signal. In clinical PET, time is seldom available for coaching, patients are frequently too unwell to comply, and it can lead to increased staff radiation doses.

This problem has long been identified as a confounding effect in medical diagnosis for all imaging modalities (MR, CT, PET, SPECT and so on) and, to some extent also, treatment delivery (external beam radiotherapy) and has consequently attracted much research
effort, particularly in the development of mathematical algorithms. Nevertheless this has
had little direct impact on clinical practice, where monitoring of and correction for patient
motion is hardly used. For example, PET plays a key role in tumour detection, staging and
treatment planning, yet because a typical clinical scan takes minutes to acquire, detection
and accurate assessment of lung tumours is limited by the effects of respiratory motion.
Current commercially available motion sensing systems only measure movement of the
chest in a limited area, and rely on physical contact with the patient. These systems are
less than ideal for use in patients, who may have difficulty in remaining still or breathing
regularly. Well established as a key research tool in neurology, PET also promises to have
increasing clinical impact with the development of commercial tracers for neurodegener-
ative conditions such as Alzheimer’s disease. However clinical evaluation of patients will
be far less effective if head movement during the scan is not measured and corrected for.

Since a typical dynamic PET scan can last upwards of 60 minutes it is unreasonable
to expect that the subject can remain perfectly stationary during PET data acquisition.
For brain PET imaging head restraints may be used however they can only reduce rather
than eliminate motion. Motion of the head can be caused by a number of factors such as
swallowing, falling asleep, through repositioning of the body which changes the balance
of weight on the neck, or simply by breathing. Figure 3.1 shows the periodic motion of
the head due to the subject breathing. The motion tracking was performed with the Polaris
position sensor (see §3.2) and the data was smoothed to better observe the cyclic motion.

![Figure 3.1: Head motion can be caused by the subject breathing. This data was ob-
tained using the Polaris position sensor and a breathing rate of $\sim 15$ respiratory cycles per
minute. The tracking data shown represents the displacement of the tracking tool in the
axial direction of the PET/CT and corresponds to the head tilting forwards and backwards
as the subject respires. Data processed from WMIC subject h00264 with permission from
Dr. P. Talbot.](image)

These different events cause the head to either undergo a rapid, relatively high amp-
3.2 Motion Tracking for Brain PET

3.2.1 Data Driven Motion Measuring

It is possible to deduce the occurrence of motion directly from the PET data itself. Such methods have other significant advantages over external monitoring such as requiring no additional operating procedures during an already busy scanner schedule as the algorithms can be performed in post-processing. Disadvantages generally arise from the relatively low quality of the PET data, especially at low count rates or during short time frames as there is simply not enough data to accurately deduce motion parameters. In brain PET, some data driven methods are only able to detect when motion has occurred yet some data driven methods aim to also quantify the amount of motion that took place. A simple way to monitor patient motion, without external tracking devices, is to dynamically calculate the centre of gravity (centroid) of detected lines of response. A recent version of this method developed in the PhD thesis [Segobin 2009] has been implemented on the octagonal head design HRRT installed at the Wolfson Molecular Imaging Centre, University of Manchester. The LORs between two pairs of opposing detector heads are used to form ‘planograms’ such as can be seen in Figure 3.2.

Centroid analysis offers a technique that is able to provide general information about motion and is relatively simple to compute. However it can only detect translations and not rotations. This method fails to accommodate for the physiological and metabolic processes affecting the distribution and transport of the tracer. This is most apparent at the beginning of scans where centroid data shows a large \( z \) drift, whereas this is simply due to the tracer arriving in the brain. Therefore a centroid type method is capable of detecting sharp movements but it would be difficult to distinguish patient drifting and tracer kinetics. An additional method which can be used to detect when motion has occurred uses the head curve data. By processing the head curve, i.e. an array of singles, prompts, and randoms over time, it is possible to detect motion as the event rates varies with motion. The prompts rate will vary if the any activity inside the head or neck moves in or out of the field of view and the randoms will vary if external activity is brought close to the field.
CHAPTER 3. MOTION TRACKING AND CORRECTION

Figure 3.2: (a) The planogram between HRRT head 0 and 4 is used to calculate the centre of mass of the LORs in \( z \) and \( y \). (b) The planogram between HRRT head 2 and 6 is used to calculate the centre of mass of the LORs in \( z \) and \( x \). This results in two measurements of \( z \) motion and a single measurement of \( x \) and \( y \). The gaps in the detector head, as can be seen in (b), are one of the factors that prevent centroid data from reliably quantifying motion.

... of view. For example if the subject checks their watch, the activity in the hand and arm will add to the randoms rate and it is likely that the action of moving the arm will affect the position of the head. Figure 3.3 shows the randoms rate and centroid of a subject who persisted in checking their watch.

Figure 3.3: (a) Shows the randoms rate plotted on a \( \log_{10} \) plot. (b) Shows the same subject motion detected using a centre of mass, or centroid approach. Data processed from WMIC subject h0636 with permission from Dr Julian Matthews.

The two previously described data driven methods are able to determine when motion has occurred, to an extent. They provide no useful or quantifiable information about the amount or direction of motion. Data driven methods are available, such as Automatic Image Registration (AIR) [Woods et al. 1992] or Mutual Information [Collignon et al. 1995] [Wells et al. 1996], that allow for the PET frames to be coregistered into a single
PET, CT, or MRI space, thereby tracking and correcting for motion. An excellent review can be found at [Hill et al. 2001]. Data driven methods have been shown to be effective in correcting for between frame motion, however the need for good statistics for good coregistration limits the minimum frame length. Results from a comparison of the motion parameters as measured and calculated from using the Polaris Vicra and image registration performed in the Vinci medical image viewer can be found in §5.3.

3.2.2 Marker Based Motion Tracking

Head motion is considered a rigid body type motion as no deformation occurs during normal movements, therefore it is possible to determine the displacement of any point inside the head by measuring the motion of the head as a whole. As such, the problem of accurately measuring the motion of the head is well suited to external measuring devices and the following section reviews the use of different types of external measuring devices. Early motion tracking devices [Daube-Witherspoon et al. 1994] [Goldstein et al. 1997] [Picard et al. 1997] used in PET often consisted of an array of light emitting sources arranged in a known position and attached to the patient’s head. Video surveillance would then be used to compute the positions of these sources and measure their movements. These methods reported adequate accuracy of measuring the position of the light sources compared to the spatial resolution of their respective PET systems. However the use of active markers has not been widely adopted and much of the later work used passive, infrared (IR) retro-reflecting markers. The advantages of using passive rather than active markers is that the tracking tool does not need to be electrically powered, resulting in a lighter and less expensive tool. In [Atkins et al. 1996] the investigation used stereoscopic IR video cameras and measured the position of a set of retro-reflecting markers placed on a tracking tool. The tool was attached to a cast that was uniquely constructed before the scan to grip the teeth of the patient’s upper jaw. Since IR light is invisible to the naked eye it does not interfere with the rest state of the patient and it is largely insensitive to most room lighting conditions.

The use of dental casts is arguably the best way to ensure that a tracking device best represents the rigid motion of the brain, but construction of the grip is time consuming and overall the device is uncomfortable for the patient. A similar tracking tool has been used in a MRI motion correction scheme where the authors noted that due to the high precision of the dental cast it was possible in follow up studies to reposition the mouthpiece and therefore the patient’s head in the same position inside the field of view as in the initial examination [Zaitsev et al. 2006].

Various commercially available optical tracking devices that apply the principle of stereovision are currently in use in many nuclear medicine centres worldwide [Fulton et al. 2002] [Beach et al. 2004] [Buhler et al. 2004] [Beach et al. 2006]. Many of these systems operate by bathing their field of view in IR light and detecting the light from retro-reflecting spheres positioned on a tool. The position and rotation of the spheres can then
be computed with simple geometry, as detailed in §4.2. This tool can be attached to the
patients head so that theoretically any movement of the tool directly represents the rigid
motion of the brain as in [Lopresti et al. 1999] where the authors were the first to present
motion tracking using the now widely adopted Polaris position sensor. However concerns
have been raised about the system as it is difficult for the operator and uncomfortable for
the patient to securely attach the tool to the head so that the true correspondence between
tool position and brain position can be ascertained. The success of IR marker based
tracking as a motion tracking device is now mainly dependent on the method of attaching
the tracking tool to the patient, with different centres adopting different methods. These
include neoprene or wetsuit caps, modified glasses and the aforementioned dental casts.
In §5.3 results are presented that compare the varying efficacy of motion tracking using
the Polaris system where the tracking tool is fixed by either a wetsuit cap or by using a
set of glasses.

3.2.3 The Polaris Position Sensor

The Polaris Position Sensor, a commercially available IR tracking system from North-
ern Digital Inc., ON, Canada, consists of a pair of IR sensitive digital video cameras which
are surrounded by a ring of IR light emitting diodes, LEDs. The LEDs pulse at a rate of 20
Hz on the smaller Polaris Vicra or 20/40/60 Hz on the larger Polaris Spectra. The tracking
tool used by the Polaris system uses IR retro-reflecting spheres to back reflect the emitted
IR light to the cameras. Without highly mirrored objects in the field of view, this ensures
that the spheres are the brightest objects in each image which aids their detection. The
3D location of each sphere can then be calculated by using a technique similar to those
described in §3.2.3, and once the points are located the prior knowledge of the geometry
of the tracking tool can be used to calculate the position and orientation of the tracking
tool. The Polaris Spectra and Vicra offer high precision tracking of IR reflective spheres
that can be attached to a tool and tracked with up to 6 degrees of freedom with positional
errors of 0.23 mm for a 3 sphere tool [Wiles et al. 2004]. The Polaris user guide does not
specify the computer vision technique used to calculate the 3D position of the tracking
tool, however the following subsection describes a routine that is likely used.

Epipolar Geometry

Computer vision is routinely used in most motion tracking systems as it enables the 3D
world to be represented and measured in software. Therefore an investigation into motion
correction needs to begin with the understanding of the computer vision techniques used.
Since most external tracking systems employ the use of a camera, it is useful to describe
the simplest camera model, the pinhole camera. The pinhole camera is an ancient optical
device however it remains a useful description of an ideal camera. Light rays coming off
an object in the cameras field of view travel through the pinhole, or centre, of the camera
3.2. MOTION TRACKING FOR BRAIN PET

Figure 3.4: The position sensor consists of a pair of IR sensitive cameras and two rings of IR light emitting diodes that illuminate the field of view. The spheres reflect the light back to the cameras and the position sensor calculates the position and orientation of the entire tool using the predefined physical definition of the spheres on the tool from the appearance of the tool in the two images.

and form an inverted image on a surface placed behind the pinhole. In Figure 3.5 the image plane is placed in front of the camera centre, O, for simplicity.

The process of projecting 3D points onto a 2D plane through the centre of the camera is known as central projection. The mapping of a 3D point $P$ from 3D space to 2D image plane can also been seen in Figure 3.5. The image plane coordinate $y_1$ is shown to be given by the similar triangle relationship $y_1/f = x_1/x_3$. A similar relationship is found for the other image plane coordinate $y_2$ to give

$$\begin{pmatrix} y_1 \\ y_2 \\ 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix} \sim C \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix} \tag{3.1}$$

This describes the relationship in Cartesian geometry however extending Equation 3.1 using homogeneous coordinates allows the Camera Matrix to be defined as in Equation 3.2,

$$\begin{pmatrix} y_1 \\ y_2 \\ 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix} \sim C \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix} \tag{3.2}$$

where $\sim$ denotes that the relationship is scalable by a non-zero number. A single image cannot typically be used to reconstruct 3D information since Equation 3.2 is rank deficient. When stereo cameras (like the Polaris position sensor) view a 3D scene, corresponding points in each 2D image acquired from the stereo rig can be used to calculate the 3D position of the point. Using a pinhole camera model, epipolar geometry describes the relationship between multiple camera views of the same scene. The pinhole model says that a ray from $Q$ passes through the centre of the camera and intersects the image plane at a single point. In real cameras the image plane is behind the camera centre, $O$, and is mirrored, however for simplicity Figure 3.6 shows the virtual and un-mirrored image plane which is in front of the camera centre. The labelled points $y$ and $y'$ in each image
Figure 3.5: A pinhole camera describes how rays of light from a 3D object pass through the centre of the camera to form an image on a plane at a distance of the focal length of the camera, \( f \). (a) A point at \( Q \) appears on the image plane at point \( P \) at pixel location \((y_1, y_2)\). (b) shows the similar triangles that exist in the pinhole camera model.

The epipole, for each image, is the point on the line \( O - O' \) that intersects the image plane. Epipolar lines \( l \) and \( l' \) are the lines on the image plane on which lies the epipoles \( e \) and \( e' \) and the image ray intersection points \( y \) and \( y' \). Using homogeneous coordinates, \( l' = Fy \) where \( F \) is the Fundamental matrix. The Fundamental matrix is an algebraic representation of epipolar geometry which states that

\[
y'^\top F y = 0
\]  

for all point correspondences, \( y \leftrightarrow y' \). The Fundamental matrix can be estimated geometrically from multiple corresponding points or algebraically if the camera matrices are known [Xu et al. 1996]. Many solutions that solve for the Fundamental matrix are freely available online, such as the functions `StereoCalibrate` and `FindFundamentalMat`
3.2. MOTION TRACKING FOR BRAIN PET

Figure 3.6: When viewed from two cameras, an epipolar plane is formed between a point commonly seen in all images and the centres of each camera. The epipole is the point along the rays that join the camera centres and each image plane. An epipolar line is the line that joins the epipole and the ray from each camera centre and the commonly viewed point in space $Q$.

in OpenCV\(^1\) ($\geq 2.0$) and the open source Camera Calibration Toolbox for Matlab\(^2\). Once the fundamental matrix is known and a method is in place to match corresponding points along their epipolar lines, the task of calculating the 3D position of $Q$ can be solved using linear algebra.

To demonstrate the process of epipolar geometry, IR images were obtained from the Polaris Vicra using the Polaris tracking software. In these images, 8 points were identified and used to calculate the Fundamental matrix. The epipoles were then calculated by taking the singular value decomposition of the Fundamental matrix (using the 8 point algorithm [Longuet-Higgins 1981], [Hartley 1997], and [Hartley et al. 2004]) and are marked on the two images shown in Figure 3.7.

The 3D points of the spheres can now be calculated if the points are perfectly located in each image and if the images have no lens distortion. Since this is unlikely it is common for cameras to be calibrated, i.e. the intrinsic and extrinsic camera parameters are calculated by imaging a known pattern over a number of positions in the field of view. A chessboard pattern is commonly used as the high contrast edges and corners can be easily and quickly detected in images. Since it is relatively simple to calibrate a modern digital camera using solutions like those mentioned above, it is often preferred to simply using the Fundamental matrix to calculate the depth of a point.

**Polaris Tracking Data**

The Polaris position sensor reports the position of origin of the tracking tool in mm and the orientation of the tool in quaternions. Quaternions represent rotation as a complex

\(^1\)OpenCV (Open Source Computer Vision) is a library of programming functions for computer vision. OpenCV is released under the BSD license and is free for both academic and commercial use.

\(^2\)The Matlab function estimateFundamentalMatrix computes the Fundamental matrix from corresponding points in images and is available in the proprietary Image Processing toolbox (v2010b) or Computer Vision Systems ($\geq v2011a$)
Figure 3.7: The fundamental matrix of the Polaris Vicra can be calculated from images taken from the Vicras IR cameras if 8 or more points can be commonly seen in both images. The epipole is the point of intersection of the different coloured epipolar lines where each colour corresponds to the similar epipolar line in the other image.

Figure 3.8: A stereo camera rig can be calibrated using multiple images of a chessboard pattern which is imaged at many points in the cameras field of view. Once calibrated, a stereo camera rig can calculate the 3D location of point correspondences using stereo triangulation by finding the crossing point of the back projected image rays. Image taken from the open source Camera Calibration Toolbox for Matlab.
number with 3 different imaginary parts as shown in Equation 3.4 and can be rewritten as a $4 \times 4$ orthonormal matrix which contains a $3 \times 3$ submatrix representing the rotation and a $1 \times 4$ submatrix denoting the Cartesian translation, as shown in Equation 3.5.

\[ q = q_w + iq_x + jq_y + kq_z \]  
\((3.4)\)

\[ A = RT \]
\[
\begin{bmatrix}
q_w^2 + q_z^2 - q_y^2 - q_x^2 & 2(q_xq_y - q_wq_z) & 2(q_zq_w + q_yq_y) & 0 \\
2(q_xq_y + q_wq_z) & q_y^2 - q_x^2 + q_z^2 - q_y^2 & 2(q_yq_z - q_wq_x) & 0 \\
2(q_xq_z - q_wq_y) & 2(q_yq_x + q_wq_y) & q_z^2 - q_x^2 - q_y^2 + q_z^2 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & T_x \\
0 & 1 & 0 & T_y \\
0 & 0 & 1 & T_z \\
0 & 0 & 0 & 1
\end{bmatrix}
\]  
\((3.5)\)

The rotation submatrix expresses a pure rotation if it has a determinant of +1 and its transpose is equal to its inverse. Transformation matrices are noncommutative i.e. $[A][B] \neq [B][A]$. To apply a series of transformations $T_1$ then $T_2$ then $T_3$, the total transformation is $T = T_3 \times T_2 \times T_1$.

The tracking tool used at the WMIC consists of 4 retro-reflective spheres (diameter 11 mm and supplied by the manufacturer) arranged in precise positions. The layout of the spheres can be changed, if certain geometric constraints are met, allowing for different designs of tracking tool to be constructed. At the WMIC the tracking tool is attached to the subject through the use of a wetsuit headcap with Velcro strips which fix the tracking tool in place, as can be seen in Figure 3.9.

![Figure 3.9: Wetsuit cap with tracking tool attached with Velcro. Tracking tool is shown in more detail showing the metallic mounting posts that attach the spheres to the tool body.](image)

To calculate the spatial transformation between the Polaris Position Sensor space and PET space, it is necessary to locate 4 (or more) points that exist in both coordinate frames. The transformation can then be calculated by using an absolute orientation approach as described in [Horn 1987]. Since the tracking tool is constructed from an attenuating material it is observable in $\mu$-map generated from either a transmission or CT scan for the
HRRT or PET/CT systems respectively. As mentioned in §2.2 these scans are necessary for the reconstruction process and are acquired in each scanning session. Attempts have been made to locate the tracking tool in emission space by placing point sources of activity on certain parts of the tracking tool, as in [Fulton et al. 2002], would be too time consuming for routine clinical use. In [Olesen et al. 2010] the tracking tool was set up so that the centre of the spheres were all in one transaxial PET image plane. A high count rate HRRT transmission scan of the tracking tool was then obtained. After $\mu$-map generation, a threshold is applied to each transaxial plane to detect the higher attenuating metal screws that fix the spheres to the body of the tool. The location of the tool can then be determined when 4 blobs are detected in an image slice. This technique is suitable if the position of the Polaris position sensor and the reference tool is fixed between scans. This calibration procedure can be time consuming as it requires careful positioning of the tracking tool and the use of a long, high count rate transmission scan.

### 3.2.4 Alternative Marker Based Motion Tracking

In addition to the Polaris position sensor there are many other methods which have been proposed to track head motion in brain imaging. Some of the more recent techniques are described below. These subsections aim to review the current rigid body motion tracking methods and gives an introduction to some of the techniques that are research later on in this thesis.

**Stereo Target Tracking**

In [Kyme et al. 2011a] the authors used a stereo-optical motion tracker (MicronTracker Sx60 Claron technology Inc., Toronto Canada) which uses a similar method to track markers employed by the Polaris however instead of IR reflective markers, the MicronTracker tracks the position of X-points on target markers as can be seen in Figure 3.10. The authors report adequate motion tracking accuracy for pre-clinical PET and that the 8 X-point tool outperforms the 3 point tool. However, the following methods detail how target marker tracking can be performed with a single digital camera and a small (5 mm$^2$) target marker.

**Monocular Target Tracking**

With modern digital cameras and advanced computer vision algorithms it is possible to accurately track in 3D a marker using a single monocular camera. The attractiveness of this type of system is that it is very scalable to different tasks using simple off-the-shelf hardware.

A tracking marker consists of a binary pattern on a 5 $\times$ 5 grid surrounded by a single ring of black squares as can be seen in Figure 3.11a. The marker detection and location uses the ALVAR library which uses OpenCV to provide simple access to this type of
3.2. MOTION TRACKING FOR BRAIN PET

Figure 3.10: (a) Example of a three point target marker and an 8 point target marker. A ruler is provided for guidance on relative size. (b) The MicronTracker (yellow box) is a stereo-optical camera that tracks X-point on target markers. Four reference target markers can be seen on the PET system. Figures taken from [Kyme et al. 2011a].

Figure 3.11: An array of 24 target markers (a) out of a possible 1024 unique markers can be created and a single marker (b) with a red grid overlaid to show how the pattern is created out of a $7 \times 7$ grid.

marker tracking. The algorithm which is used first applies an adaptive threshold to the colour video to obtain the borders that are present in the scene. Borders that have too few pixels to represent a tracking marker are filtered out and borders that form a concave contour with exactly four corners are labelled as possible tracking markers. The rectangles are transformed using homography to obtain a frontal view as to remove the projection perspective and this allows the identification of the unique internal code to uniquely label the marker. A description of how the algorithm estimates the pose of the marker from measuring the corners of the markers can be found in §8.2. A simple diagram is shown in Figure 3.12 that illustrates that with a calibrated camera the rays that connect the camera centre, the marker corners in the image plane, and the marker corners in real space can
be drawn. Pose estimation iteratively calculates the optimal pose of the marker using \textit{a priori} knowledge of the marker size.

Figure 3.12: A target marker is imaged allowing for the 4 corners of marker to be detected in the image. If the true distance between the corners is known, then there can be only one position along the back projected rays where a marker of that size would produce the image. To obtain the correct pose of the marker, the internal layout of the $5 \times 5$ grid is used to determine the unique identity of each corner. The marker will always have a negative surface normal (y axis) since the marker has to face the camera in order to be seen.

Since the algorithm is fairly computationally inexpensive this marker tracking can be performed in real time using a regular desktop computer. This type of monocular, marker based tracking could be used for both clinical and preclinical applications. The high quality macro capabilities of modern digital cameras allow for a small marker of $2.5 \text{ mm}^2$ to be tracked at a distance of 5 cm which makes this system applicable for PET scanners that have restrictions due to bore size, such as the HRRT or the Siemens Inveon, a preclinical micro PET scanner.

This method allows for very small targets to be used in head motion tracking, rather than the large and heavy tracking tool used by the Polaris position sensor. Using smaller targets has two advantages; the fixing method does not need to be as strong, and multiple targets can be used on the head. As discussed previously, tool slippage degrades the reliability of any marker based tracking. However by using multiple tracking targets, which can only realistically be achieved by using very small targets, the tracking system has increased robustness against the target moving non-rigidly with the head. Multiple targets can be fixed to the subject’s head at points that are less likely to move non rigidly during the scan, such as the bridge of the nose, the temples and across the hair line, and if all are simultaneously tracked then if a facial twitch occurs only some of the tracking markers should move non-rigidly. So long as a majority of tracking markers are moving in
3.2. MOTION TRACKING FOR BRAIN PET

Figure 3.13: A tracking target with a 5 mm² printed square pattern is fixed to a head phantom and tracked using a Logitech c910 webcam. (Left) shows the raw colour video used by the tracking algorithm with the thresholded video being shown in (right). A cube with the local coordinates of the target is digitally drawn over the original video.

agreements with each other, then their motion should be representative of the true motion of the head. This technique is further expanded on in Chapter 8, as it forms the basis for the experiments performed therein.

3.2.5 Markerless Motion Tracking

In addition to these more experimental motion monitoring methods, recent improvements in computer vision algorithms and the lowering cost of high quality digital cameras have allowed for the potential of markerless motion tracking in brain PET to be explored. As this is a relatively recent development in the field of motion tracking in PET there are currently no commercially available or routinely clinically used devices that employ markerless motion tracking. The attraction of not having a tracking tool that may slip, rendering the motion data unreliable, or be uncomfortable for the patient, potentially inducing additional motion, has encouraged many attempts to solve this currently unmet clinical need. In [Ma et al. 2009] the naturally occurring facial landmarks, such as the corners of the eyes, nostrils, mouth etc were tracked using a stereo camera rig. Other more recent attempts at markerless motion tracking are discussed later in this Chapter however to give an overview, any successful markerless tracking technique needs to have the following features

1. Practicality: Can the system be used in a PET scanner without affecting normal scanning operations?

2. Accuracy: Can it be used in high resolution medical imaging?

3. Robustness: Is the tracking sensitive to non-rigid facial deformations?

4. Stability: Does the tracking drift over time due to compound errors?

To address all of these points, it would be sensible for a markerless tracking system to operate in IR so to be insensitive to skin tone and to be invisible to the subject. The
components of skin colour, melanin and haemoglobin have little absorption in the near-IR wavelengths [Graaff et al. 1993] which allows for high contrast IR images to be obtained with all skin tones. Any technique that aims to be routinely used in clinical environments needs to be applicable to all skin tones. The use of IR is also essential for Fluorodeoxy-glucose, FDG, activation studies that need to be in complete darkness to avoid the ocular regions of the brain becoming activated as described in [Schlindwein et al. 2008]. High resolution, IR sensitive digital cameras are now readily available as off-the-shelf components, allowing for highly accurate computer vision algorithms to be deployed on consumer grade hardware, however the overall success of the technique depends on how sensitive it is to facial feature changes such as swallowing, talking, or blinking.

Stereo Cameras

An extension to the epipolar geometry and stereo triangulation method discussed in Section 3.2.3 where the corresponding points are automatically determined in each image pair can be used to calculate the 3D position of the points and therefore the object being tracked. This can be considered to be similar to the Polaris technique, however the tracking tool is made up of interesting features in the image, such as the corners of the mouth, eyes, nostrils etc. These feature points can be tracked through the video sequence to improve the reliability of the tracking as in the study [Kyme et al. 2011b]. The main advantage of this type of system is that it has the potential to be very accurate, as the overall accuracy of the tracking is determined by the resolution of the pair of cameras used. On rigidly moving phantoms [Kyme et al. 2011b] presented sub-mm accuracy in position tracking.

The limitations of this technique stem from the ability to reliably locate the features in each image. The scale-invariant feature transform (SIFT) feature detection algorithm was used in [Kyme et al. 2011b] to produce a database of approximately 1500 3D landmarks in 5000 frames of video data. These landmarks form a sparse model of the object being tracked. In practice it is unlikely that more than 20 landmarks would be used to calculate the 3D pose of the object in any one frame, as most of the 1500 landmarks are hidden from the camera at any one point and only become visible if the object moves. It also becomes more computationally expensive if more corresponding landmarks need to be detected so a trade-off is needed to be made. In [Kyme et al. 2011b] a figure is shown with 11 matched features for a taxidermied rat head.

As can be seen in Figure 3.14 many of the potential feature points are around the edge of the head on the high contrast boundary between the white fur and the black background and many of the point joined by epipolar lines connect to features that are expected to move and twitch on a awake animal, such as the whiskers, nose and eyelids.

Sparsely populated models are generally sufficient for rigidly moving objects as all matched points agree with calculated registration. If an object has regions of non rigid deformations, such as regions around the mouth, the calculation of the pose can be greatly
3.2. MOTION TRACKING FOR BRAIN PET

Figure 3.14: Potential features are marked as red squares and successfully matched features have been connected between the two images. Image taken from [Kyme et al. 2011b] affected. In the case of the 11 feature pairs in [Kyme et al. 2011b], 6 of the pairs are from regions that would be expected to move non-rigidly with the skull such as the nose, whiskers and ears. It can also be dangerous to rely on features being present in colour images as the surface may be poorly textured, have low contrast or be under poor lighting conditions in which case the model will be very sparsely filled and the pose estimation may fail.

Figure 3.15: Ink creates high contrast features on regions of the head which are unlikely to move non-rigidly during awake imaging. Image taken from [Kyme et al. 2012].

To aid the feature detection the authors adapted their methods in [Kyme et al. 2012] proposing the use of indelible ink to mark features on the forehead of the animal, as in Figure 3.15. The authors acknowledge that the technique is no longer truly markerless however marking the animals forehead greatly increases the chance that the features detected will be stable and the process of applying the markers is fast and simple. However since markers have been applied it may be simpler to rely fully on markers, using a technique such as the one discussed in §3.2.4. Applying indelible ink markers requires the colour of the fur to have good contrast against the colour of the ink and even though markers have been applied it is still necessary to run the feature detection algorithm which is very time consuming.

The electronic entertainment industry has introduced the world to cutting edge hardware at consumer grade prices. This industry is highly competitive and rapidly changing
and introducing new devices like the Microsoft Kinect. A recent announcement from Sony Computer Entertainment detailed a high definition stereo camera (see Figure 3.16) device for use with the Playstation 4 home video game console. Early specifications detail the cameras will each have $1200 \times 800$ resolution with an $85^\circ$ field of view.

**Time of Flight Camera**

Time of flight cameras operate by measuring the different times that pulses of light arrive on a sensor after being reflected off a surface. Each pixel on the sensor is opened for a short window collecting the reflected light. The peak in intensity against time corresponds to the amount of time taken for the light to be reflected off the surface and return to the TOF sensor. This gives the $z$ coordinate of the surface to which the light was reflected off. The $x,y$ real world coordinate can be found by projecting the $i,j$ pixel coordinate ray out using the cameras intrinsic parameters as in Eq. 3.2. Since no collimator is used, light scattered off other parts of the surface can interfere with the signal recorded at each pixel which distorts the depth measurement. The spatial accuracy of TOF cameras depend on the speed of the electronics used which results in high costs for top end, high resolution, cameras.

**Figure 3.17:** The PMD CamBoard nano, is a small profile, time-of-flight depth sensor. It has a spatial resolution of $160 \times 120$ with a field of view of $90^\circ \times 68^\circ$. This means that the sensor needs to be closer than 80 mm for 1 image pixel to equal 1 mm$^2$. The sensor can record data at frame rates of up to 90 Hz which may allow for temporal noise smoothing. Image taken from the PMD user information guide.
There has been much commercial interest in devices that allow for natural user interface, which has pushed down the cost of the commercially available devices. The Camboard nano from PMD is a TOF depth sensor that can be purchased for £420 GBP, a significant price drop compared to the available technology just a few years ago.

**Photometric Stereo**

Photometric stereo, or shape from shadows, aims to reconstruct the shape of a surface from the different intensities of light on the surface after being illuminated by multiple light sources from different directions. In a simple example a single camera is used with three non-colinear white light sources that illuminate the scene in turn, so that in camera frame 1 the scene is only illuminated from light source 1 and so on. This method requires multiple video frames that are synchronised to the alternating light sources, however an extension presented in [Vogiatzis et al. 2012] which uses different coloured lights (red, green, and blue) that allows for the three images needed for photometric stereo to be obtained by splitting the red, green, and blue channels from a single video frame, as can be seen in Figure 3.18.

![Photometric Stereo Example](image)

(a) Figure 3.18: Three non-collinear, different coloured light sources are shone onto the face. The individual red, green and blue (RGB) components of the digital image can be separated into 3 images (a). The photometric stereo algorithm can then be used to reconstruct a depth map (b) from these simultaneously obtained images. Figure adapted from [Vogiatzis et al. 2012]
In three colour photometric stereo there exists a relationship between the camera pixel-wise response, \( c \), and the normals \( n \) of the surface being imaged.

\[
c = V \cdot L \rho n
\]  

(3.6)

Where \( V \) is a calibration matrix, \( L \) is a matrix of vectors describing the direction of the three light sources, and \( \rho \) is the surface albedo (the reflection coefficient or the fraction of light which is reflected). Once \( V \cdot L \) has been estimated by calibrating the system using an object of known reflectance and geometry, \( n \) can be easily calculated.

This method is capable of producing very high resolution depth images from a single video camera, however the need to illuminate the scene with brightly coloured lights would have been uncomfortable for the subject. It may be possible to use three wavelengths of IR light, as it would be invisible to the subject and skin is more lambertian to IR wavelengths (i.e. reflected light is more diffuse and less specular), however a bespoke tri-band bayer filter would have to be designed.

3D Scanning With Swept Planes

In the description of Epipolar geometry and stereo cameras in §3.2.3 the 3D location of a co-imaged point can be found if the baseline or extrinsic calibration between the two cameras is known. If one of the cameras are replaced by a laser pointer, then the 3D position of where the laser point is incident on a surface, can similarly be calculated. The advantage of projecting a point is that it is not necessary to locate and determine point pairs in the two camera images, however the disadvantage is that only a single point is being sampled at any time. To address this, it is possible to use multiple points or by sweeping the single point across the field of view. In systems that use multiple points (such as the Microsoft Kinect) care needs to be taken to ensure that each measured point can be uniquely identified to determine its projection line. Rather than sweeping a single point through the scene it is common to use a projected line which forms a plane of light, under the pinhole camera model. The position and orientation of the plane can be estimated by placing planar surfaces behind the object being scanned as in Figure 3.19.

After camera calibration the 3D surface can be reconstructed through triangulation by equating the intersection of each pixel’s optical ray and the plane and sweeping the line across the scene. The spatial resolution of such a system is dependant on the spatial resolution of the digital camera used and the line spread function of the projected line, therefore high resolution surfaces can be created with very low cost hardware. However, the temporal resolution is limited as the line source needs to be swept throughout the scene in order to fully sample to object and any motion of the object will create artifacts in the reconstructed surface. Additionally the requirement of a background planar object limits this methods applicability in space limited scenarios.
3.2. MOTION TRACKING FOR BRAIN PET

Figure 3.19: (a) Shows the set up demonstrated in [Lanman et al. 2009] where a wooden dowel is used to create shadow lines from a standard incandescent light bulb. (a) Shows the planar surfaces used in the background to estimate the equation of the plane for each image. The markers placed on the boards are used to calculate its position using techniques such as those described in §3.2.4.

Structured Light

Structured light encode the identity of the individual planes spatially, temporally, or a combination of the two by using a digital projector. The use of a digital projector removes the need to mechanically sweep the projected line by encoding the identity of the plane in a single projected image frame (spatial encoding) or across multiple frames (temporal encoding). Spatial encoding is more suited to dynamically moving scenes whereas temporal encoding can benefit from redundancy, improving the identification of lines to reduce image artifacts. Spatial encoding can be performed using coloured lines [Boyer et al. 1987] however methods that reply on colour to identify projected lines assume that the reflected light is not modified by the scene. There exists many temporal encoding approaches, for example binary patterns [Posdamer et al. 1982] or gray codes [Bitner et al. 1976]. Figure 3.20 shows the sequence of structured light using a binary and grey code pattern. The number of subdivisions (or number of frames per sequence) needed to identify which projected line each image pixel belongs to is given by the \( \log_2 \) of the image width, i.e. for a \( 1024 \times 768 \) image, \( \log_2(1024) = 10 \) frames are needed.

To reduce the number of frames needed, three phase-shifting can be used where each frame consists of a \( 120^\circ \) shifted sine wave intensity pattern. The phase at each pixel can be calculated from the ratio of the measured pixel intensities over the three images (as in [Zhang et al. 2006]), and this can be used to calculate surface depth using a phase-to-height (for example [Legarda-Sa et al. 2004]) calculation from a flat reference frame.

Additional reviews and assessments on structured light scanning and the various encoding strategies can be found in [Lanman et al. 2009] [Geng 2011] [Salvi et al. 2004] [Salvi et al. 2010].

This method means that the effect temporal resolution is \( 1/3 \) of the resolution of the...
CHAPTER 3. MOTION TRACKING AND CORRECTION

Figure 3.20: The first 4 projected images used in the binary pattern (a) and gray code (b) structured light techniques for temporal encoding. The sequence of the binary pattern (c) and gray code (d) where each row represents an image pattern bit profile illustrates how the gray code demonstrates improved robustness for boundary pixels. Image taken from [Lanman et al. 2009]

Figure 3.21: (a)(b)(c) Show three images taken where the striped pattern is shifted by $2\pi/3$. The images can be reconstructed into a point cloud as in (d). Image taken from the open source Structured-Light project [McDonald 2013]
projector/camera pair and any rapid motion that occurs between the 3 frames will greatly distort the surface. A structured light approach has been adapted for use in the HRRT in [Olesen et al. 2012] using a bespoke IR projector and camera array, however suffered from a slow refresh rate of the camera which resulted in an effectual temporal resolution of $\sim 5$ Hz. At this rate the effects of tremors can cause motion artifacts due to between video frame motion. However, with the use of a high speed IR camera (120Hz), intra-frame motion is unlikely to be an issue. Using higher frame rates results in a higher processing load, potentially resulting in non-real time operation. This is unlikely to remain a problem with the ever increasing processing power of CPUs and GPUs.

The Microsoft Kinect

The Microsoft Kinect is a $150, consumer grade, hand held 3D depth sensor comprising of an 830 nm infra-red (IR) laser projector and an IR sensitive camera. The projector emits a static, pseudo-random speckle pattern of dots which is distorted by 3D objects. The Kinect calculates the distance to the surfaces using the disparity of the expected IR dot positions and their detected locations in the image plane of the IR camera. The operation of the Kinect is not directly in the public domain, however information can be gathered from patents filed by Primesense Ltd. the company which Microsoft licensed the Kinect technology from.

US Patent application 2010/0290698 shows an illustration demonstrating depth from focus, the principle that object depth can be estimated from determining how blurred the object is. Specifically in this patent application an astigmatic lens is used where the focal length is different in $x$ and $y$. As illustrated in Figure 3.22a a projected point becomes an ellipse whose orientation depends on the distance to the projector.

![Figure 3.22: (a) US Patent illustration showing the principle of depth of focus which is possibly used in the Kinect to determine the approximate distance to the object. Figure taken from US Patent 2010/0290698. (b) US patent illustration showing three regions for unique reference images to be targeted to reflected points at specific depth regions. Figure taken from US Patent 2010/0177164.](image)

Other patents (for example US patent 2010/0118123) from the Kinect developers ex-
ist which describe other potential methods for sensing depth, and as such it is difficult to determine exactly how the Kinect operates. The following analysis is based on the published calibration experiments of the Kinect sensor performed in [Khoshelham et al. 2012].

The depth of a reflected point can be calculated from its observed disparity from its position in a previously obtained reference image. Such a reference speckle pattern can be obtained by projecting onto a flat surface at known distance. This reference is stored in memory and is used to compare new IR images against. A speckle point on a surface not at the reference distance will be shifted in the IR image plane in the direction of the baseline connecting the projector and IR camera. This shift, or disparity, is calculated for each speckle point observed in IR image. Patent 2010/0177164 describes the use of three regions with a specific reference image for each region which could be used in combination with three superimposed speckle patterns projected by the Kinect. Each of the three speckle patterns could be separated (for example by the shape, size, or brightness of each point) and compared to their relevant reference image giving a higher resolution near region and lower resolution middle and far regions. Figure 3.22b shows a illustration of the three regions as described in the patent.

![Diagram showing the principle of depth from disparity](image)

Figure 3.23: Diagram showing the principle of depth from disparity, where the distance to an observed point can be calculated by measuring the pixel distance to its position in a reference image. Figure taken from [Khoshelham et al. 2012]

In Figure 3.23, the reference pattern is obtained at a distance $Z_0$ and the object is at distance $Z_k$. This results in point $k$ being displaced by a distance of $D$. On the image plane at focal length $f$, the disparity, $d$, is the pixel shift between the expected and measured speckle point location. The disparity, $d$, is in direction of the baseline $b$, which connects the camera centre, $C$ and the laser projector, $L$.

Using similar triangles the following is true:

$$\frac{D}{b} = \frac{Z_0 - Z_k}{Z_0}$$

(3.7)
\[
\frac{d}{f} = \frac{D}{Z_k}
\]  

(3.8)

By substituting \( D \) in Equations 3.7 and 3.8, \( Z_k \) can be expressed by the following:

\[
Z_k = \frac{Z_0}{1 + \frac{Z_0}{f_b}d}
\]  

(3.9)

Once \( Z_0, f, \) and \( b \) are calculated from calibration, depth can be calculated for each observed speckle point. Once depth is know the 3D position of the speckle point can be calculated using the camera model as described in Equation 3.1.

It must be noted that this technique does not provide a continuum of depth values across a surface as depth is only calculated at the speckle points. Interpolation is required to fill in gaps between speckle points and this results in surfaces further from the sensor being less sampled and therefore less accurately measured.

In the software provided by Primesense and Microsoft, sensor calibration has been performed at the factory. The raw Kinect depth data is in the form of an 11-bit video which is sent to the host PC via a USB 2.0 cable. Since 1 bit is reserved for null depth data, 1024 depth levels remain to describe the disparity measured for each pixel. [Khoshelham et al. 2012] show that the minimum difference in depth that can be measured, i.e. the depth resolution \( \delta_Z \), is proportional to the squared distance to the sensor, \( Z \). In Equation 3.10 they use a factor of \( 2.85e^{-6} \) which was determined by measuring the slope of the linear relationship between the normalised measured disparity and the inverse depth of a planar surface at different measured distances from the Kinect.

\[
\delta_Z = 2.85e^{-6} \times Z^2
\]  

(3.10)

This relationship gives a depth resolution of 1.8 mm at a distance of 800 mm, rising to 71.3 mm at a distance of 5000 mm. The resolution in the lateral XY plane is dependant on the spacing of the speckle points, as it is only possible to sample depth where a point reflects off a surface. The point density, or number of points per unit area, and therefore the lateral resolution is inversely proportional to the squared distance to the Kinect. The lateral resolution will also have step differences between the three depth regions if the three unique patterns have different point spacing.

Each video frame can be either stored to disk or used in a real-time process. In these experiments the depth video data is projected into 3D space in the form of a point cloud, an array of points which represent the underlying surface vertex point in Cartesian space. The optimal operating range of the Kinect is 0.5-1.0m and no data can be obtained closer than 0.4m as the IR laser pattern is too bright to be resolved. The Kinect has a field of view of 57° horizontally and 43° vertically which at the native resolution of the Kinect depth data of 640 × 480 results in a head sized object of 200 × 150mm at a distance of 0.8m consisting of over 14,000 pixels of depth data. Compared to the total number of pixels in the depth image of 307200 pixels, the relatively few depth pixels that represent
Figure 3.24: Brief overview of Kinect depth data acquisition. A speckle pattern of IR dots is distorted due to 3D structures. The distortion is used to calculate the depth of the surface away from the Kinect sensor and this depth information is recorded as a depth video. The video can be projected into 3D space using the intrinsic camera properties of the Kinect to form a point cloud of the scene.

the surface of a face exemplifies the challenge faced in using the Kinect to image and track the bulk motion of small objects especially as the face will contain non-rigidly moving regions. The low cost of the Kinect is in part due to the fact that it was originally intended for use for whole body skeletal tracking for use in a video games console. Raw Kinect data is inherently noisy and holes can be created from shadows and some surfaces. The Kinect is used extensively throughout this thesis, and more detailed information about how the Kinect is used as part of a markerless motion tracking system is provided in the relevant research chapters.

3.2.6 Spatial Registration Using Iterative Closest Point

To close this section on depth sensors, it is useful to describe how the point cloud depth information can be used to track motion. The four previously described methods can all produce point cloud data, i.e. an array of points that represent the underlying surface of the object being imaged. To monitor and measure the movements of the object it is possible to register subsequent point clouds or a point cloud to a single reference cloud. Iterative closest point (ICP) is an accepted algorithm to register multiple point clouds and the transformation returned by the algorithm will describe the rigid body motion of the surface between these two time points [Chen et al. 1992][Besl et al. 1992]. ICP attempts to iteratively minimise the difference between two sets of 3D points by using a cost function.

ICP is a very powerful method of aligning multiple point clouds however it needs good initial conditions, i.e. for the clouds to be almost registered, otherwise the registra-
tion can fail. Other confounding effects on the success of ICP include if the surface is deforming non-rigidly in time, or if the two point clouds contain high levels of noise. Non-rigid sections of point clouds may be filtered out by methods such as \textsc{Random SA}mple Consensus, \textsc{RANSAC} [Fischler et al. 1981], or by simply setting the distance to nearest neighbour in the target point cloud to be small, e.g. \( \leq 5\text{mm} \). As long as there is a sufficient number of points that behave rigidly and with low noise, ICP should converge to a good estimation of the transformation between the two point clouds. Consideration to the implementation and parameters used for ICP have to be made to trade off between number of points used and time per registration. In [Izadi et al. 2011] and [Newcombe et al. 2011] Microsoft Research developed an ICP implementation on a general purpose graphics processing unit (GPU) which drastically sped up computational time and allowed for real time ICP registration using data from the Kinect. Further information regarding KinectFusion can be found in §6.2.2.

In this thesis, figures are presented of point clouds registered using various implementations of ICP. Where applicable, these point clouds have also been analysed by measuring for each point in one of the samples the distance to the nearest point in the other cloud. This is known as the Hausdorff distance.

### 3.3 Motion Tracking for Whole Body PET

In whole body medical imaging of the thorax the predominate source of motion is the respiratory cycle. For modalities with short acquisition time, like CT, a breath hold technique can be used to reduce the blurring, however for PET when acquisition times can be many minutes this approach is clearly infeasible. Similarly to head tracking there are systems that aim to track motion by either using external monitoring or by using the PET data itself, however unlike brain imaging the motion of the internal organ in the thorax is decidedly non-rigid in nature.

#### 3.3.1 Data Driven Gating

An \textit{a posteriori} gating method is introduced in [Visvikis et al. 2003] that assumes that due to the deformable nature of tissue, different pixels from the same organ will experience different motion amplitudes, however the frequency of the periodic motion is the same. The paper presents a method of Fourier transforming time activity curves from different small regions of interest to determine the frequency of the motions enabling the regions of interest to be placed on high contrast boundaries. The authors used simulated data and reported that the calculated frequency was within 2\% of the simulated frequency. This showed that it was possible to obtain gating frequencies without using devices that directly measure chest expansion such as pressure sensing chest belts [Kokki et al. 2007].

A similar centre of mass method has been investigated in [Buther et al. 2009] where data was acquired on a Biograph Sensation 16 scanner. The process requires oblique LOR
to be rebinned via the single-slice rebinning algorithm into 47 transverse sinograms. The true coincidence rate per 50 ms frame in each transverse histogram was determined from the listmode data to form an axial histogram $T(i, t)$ (where $t$ and $i$ denote time frame and transaxial slice number respectively) of measured coincidence events within time frame $t$. The axial centre of mass as a function of time can then be computed for each histogram to infer a respiration and cardiac rate.

The geometry of 3D PET scanners result in an inhomogeneous sensitivity profile, as oblique LOR cannot be measured at the extremities of the field of view. This feature can be used to estimate the motion of a moving source of activity, since when the source moves through the inhomogeneous sensitivity profile the number of detected events will vary. This method has been investigated in both simulated data sets [He et al. 2007] and patient data [Buther et al. 2009]. In the latter study, list mode data from the scanner was analysed by recording the number of prompt and delayed events in 50 ms frames, and taking into account the natural decrease in total events due to the exponential decay curve, the true coincidence rate can be estimated. This is generally a noisy signal, so can be Fourier transformed to give the underlying harmonics which are identified as a low frequency contribution caused by respiratory motion and a higher frequency component caused by cardiac motion. The inverse Fourier transformation can then be used on these identified frequencies to give gating curves for whole body motion correction.

### 3.3.2 Marker Based Motion Tracking

There are many proposed solutions to measuring the respiratory or cardiac rate to enable gating such as using ECG [Klein et al. 1998] for measuring the electrical activity of the heart, thermistors [Boucher et al. 2004] that detects the exhaled air from the patients, and video surveillance of the motion of objects placed on the patient’s chest [Nehmeh et al. 2002] [Beach et al. 2004] [Beach et al. 2006] [Kokki et al. 2007] [Mukherjee et al. 2009]. Commercially available solutions are available such as the Varian RPM and Anzai pressure belt. With ECG it is possible to accurately monitor the cardiac cycle so that gates can be placed at precise points during each cycle. However, since organ motion is non-rigid, these tracking systems can generally only be used to gate the PET data into motion reduced frames. Efforts to correlate external markers with internal organ motion aim to allow for precise and real time organ tracking for targeted radiotherapy [Seppenwoolde et al. 2002] These radiotherapy applications demonstrate the clinical need for accurate knowledge of internal organ location during treatment, however the benefits for imaging are also numerous. There have been many studies evaluating the correlation between internal organ motion and external placed fiducial markers [Hoisak et al. 2004][Gierga et al. 2005][Beddar et al. 2007] which state that there is, to a degree, a linear relationship between organ motion and external markers. More sophisticated non-linear models of internal organ motion are being developed [Tsoumpas et al. 2010] [Guérin et al. 2011] [King et al. 2012] and may become more widely used with the arrival of simultaneous
3.3. MOTION TRACKING FOR WHOLE BODY PET

PET/MR scanners.

**Varian Real-time Position Management**

The Varian RPM is similar to the Polaris systems in that it uses IR light to detect IR reflective markers on a tracking tool which is worn on the chest. The simplest tracking tool consists of a lightweight block with two reflective discs spaced 3 cm vertically apart. The topmost disc is tracked in the video sequence and the apparent distance in pixels to the bottom disc is used to calibrate the amplitude of the motion. The respiratory phase is then calculated from the translation of the tracking tool and gating triggers are injected into the PET listmode in real time. Figure 3.25 shows the software interface of the system tracking a block placed on a phantom torso.

![Figure 3.25: The tracking block consists of a an array of IR reflective markers placed on a flat surface. This image shows the tracking of a target block placed onto a phantom torso. Image taken from the Varian user manual.](image)

**Anzai pressure belt**

The Anzai (Anzai Medical, Tokyo, Japan) pressure belt consists of a elasticated belt which contains a pressure sensing pad. During respiration the force of the chest wall against the pressure pad changes and this varying pressure signal is used to determine the respiratory phase. Anzai data is discretised into pressure levels and two sensors are included with the device, a high sensitivity and a low sensitivity sensor. These are used selectively depending on the expected dynamic range of the subjects breathing. Figure 3.26 shows a subject being fitted with the belt.
3.3.3 Markerless Surface Measuring

The Microsoft Kinect

Recently, a use of the Kinect to track a respiratory signal has been demonstrated [Xia et al. 2012]. This system tracks the displacement of a board placed on the subject’s chest. This allows for the Kinect to be placed at a shallow angle relative to the subject’s chest. Without the board, any creases in the clothing of the subject would result in very noisy Kinect data as the 3D data from the Kinect is poorly defined at edges. The Kinect’s poor performance at boundary regions is due to the problem of the increasing speckle point separation with distance to the sensor (see earlier discussion on the Kinect in §3.2.5). The Kinect depth data is only sampled at the points of reflected speckle light, and surfaces are interpolated between sample points. At boundary regions, the points can be distorted by a rapid fall off in Z (blurring the speckle point) or points can be lost in background regions.

Additionally the shallow angle of the Kinect relative to the board and patients chest would result in the depth measurement of the board extend across a large range of depth values and therefore have increasing noise with the decreasing depth and lateral resolution. [Alnowami et al. 2012], in a system to model the surface of the chest during respiration, used a Kinect directly in front of the subject and showed that the Kinect reports linear depth over a distance range applicable for Medical imaging. In PET/CT imaging, the scanner gantry prevents the positioning of the Kinect directly in front of the chest, due to the Kinect’s minimum operating distance.

Surface Measurement In Associated Medical Disciplines

Patient motion affects many medical disciplines, not solely in medical imaging. For example in image guided radiotherapy (IGRT), any difference between patient position
3.3. MOTION TRACKING FOR WHOLE BODY PET

and the static CT scan used to plan treatment results in radiation being delivered to the incorrect location. This can lead to reduced effectiveness of treatment and can cause damage to healthy tissue. IGRT can be coupled with a cone beam CT such as in [Jaffray et al. 2002] to provide a CT map of patient on the radiotherapy bed. Currently this cannot be performed in real time as the image reconstruction is performed requires off-line processing. This can be done to produce a patient specific motion model to then be used with an external trigger as in [Li et al. 2006], however this assumes regular and predictable breathing patterns. The use of 4D CT to continuously track motion would also lead to an undesirable increase in dose to the patient. The use of internal structural markers in CT and MRI to estimate the deformation of internal organs during respiration is discussed further in §3.4.2.

Non-ionising commercially available surface measurement systems that have been used in medical applications exist, such as the 3dMD (3dMD Ltd, London, UK), and AlignRT (VisionRT Ltd, London, UK). These systems utilise variants of the previously described methods for obtaining depth from stereo and projected patterns. In [Schöffel et al. 2007] the AlignRT showed rigid body translational error of $\sim$1 mm when compared to a stereo-optical passive IR reflective marker tracking system. However the temporal resolution of the AlignRT has been observed at 1-1.7 frames/s [Peng et al. 2010]. This low frame rate may hinder the operation of the AlignRT on quickly deforming surfaces as rapid motion may be missed due to the low refresh rate.

More recent developments for IGRT have aimed to address the need for real time operation by specifying that the depth measurements should be obtained at rates exceeding 20 Hz. Fourier transform filtering has been widely used and is based on the work performed in [Takeda et al. 1982] which showed that Fourier transforming a fringe pattern transformed onto a surface gives the underlying topology of the surface. In [Price et al. 2012] the authors present a surface measurement system utilising Fourier transform profilometry which can operate at rates above 20 Hz with a 2 mm conformance of the system measured surface to a cone beam CT generated surface for over $90\% \pm 3\%$ of the surface. This system directly measures the patient’s skin at $512 \times 512$ resolution and operates by projecting a modulating fringe pattern onto the skin. After removing the background, the image of the fringe pattern is Fourier transformed to obtain the orthogonol spatial frequencies of the pattern which has been distorted by the surface relative to a reference frame. The Fourier domain image is filtered to isolated the first term of the Fourier transformed fringe pattern. The inverse Fourier transform of the filtered Fourier space image gives the spatial phase distribution, initially wrapping around every $2\pi$ however this can be unwrapped to give a continus distribution.

The calibration procedure developed by the authors is beyond the scope of this review however it must be noted that they gave particular consideration to making the system practically usable for IGRT, where factors such as relative projector to patient position are fixed due to the requirements of radiotherapy procedure.
3.4 Motion Correction Schemes

3.4.1 Brain Imaging Motion Correction

The goal of any investigation into patient motion is to be able to correct the PET data for movements that occurred during the scan. There are generally two groups of techniques to accomplish motion correction; image based and event based. In image based motion correction the PET data is framed into discrete time frames which are then co-registered, either using motion parameters measured by an external device or by using the PET data itself. Event based motion correction is sometimes referred as LOR rebinning as the aim of such techniques is that individual LORs are transformed to the position that they would have been had no motion occurred. As such, event based techniques have the potential to produce motion free PET images whereas image based methods will always retain intra-frame motion and so can only be considered as motion reduction.

Image Based Motion Reduction

Image based schemes are generally less computationally demanding than event based methods. For example an average dynamic PET data set with 26 frames would only need 26 registration operations to be performed for an image based scheme however, all $\sim 10^9$ LORs would have to be rebinned for an event based method. Image based schemes have historically proved more popular, due to the vast difference in computational cost, and effort has been made in attempting to reduce the intra-frame motion, that increasingly effect longer frames. A method that has been shown to correct for sharp distinct movements involves the use of Multiple Acquisition Frames (MAF) [Picard et al. 1997] [Fulton et al. 2002] [Montgomery et al. 2006]. The data stream from the scanner is saved into new frames if a displacement larger than a predetermined amount is measured. These frames are individually reconstructed then coregistered to compensate for the motion. Coregistration can be performed by manual realignment of the individually reconstructed frames or by automated methods for example those developed in [Minoshima et al. 1992] and [Woods et al. 1992]. If the threshold of acceptable movement is very small, then many frames will be reconstructed. There will be little intra-frame motion but the individually reconstructed frames will have poor statistics. Conversely if the threshold is too large, then only a few frames will be reconstructed. These frames will have good statistics but will be blurred due to high levels of intra-frame motion. MAF type schemes are a relatively effective method for producing motion reduced images without the need for complex, time consuming computing.

Deconvolution of image blur is well understood, however it has not been widely adopted for medical imaging purposes because it requires accurate knowledge of the patient motion. A Fourier-based deconvolution method was presented in [Menke et al. 1996] that applied a spatially invariant blurring function to the reconstructed image which was derived from the motion tracking data acquired by two IR sensitive digital cameras mon-
monitoring the position of reflective tape fixed to a tracking tool however this method could not include rotations and also amplified noise.

Event Based Motion Correction

Event based methods apply motion correction prior or during the reconstruction process and as such they have the potential to be more accurate than image based methods as they operate on every LOR in the whole PET data set. An event based method for motion correction involves the geometric transformation of individual LOR to the coordinates that they would have occupied had motion not occurred. This method was implemented in [Menke et al. 1996] but has since been improved by taking into account the difference in the normalisation factors between the recorded LOR and the motion corrected LOR, which has been shown to cause artifacts [Buhler et al. 2004]. Such event based methods require scanners capable of either on-the-fly normalisation or the capability to record data in list-mode so that the correction of the LORs can be performed later on.

A study was performed in [Fulton et al. 2004] that compared the performance of LOR rebinning against a MAF technique. The investigation was performed using the Hoffman brain phantom with a Siemens/CTI Ecat Exact HR+ scanner. The phantom was stationary for a 30-minute emission scan followed by another 30-minute emission scan in which frequent arbitrary movements were applied to the phantom. The LOR rebinning was performed by deriving a transformation from the inverse of the motion recorded by the external tracking device. In general the coordinates of the transformed LOR do not correspond with the surface of the detector rings therefore the LOR needs to be extrapolated and the points of intersection from the extrapolated LOR and the detector surface should be used, as described in [Buhler et al. 2004][Woo et al. 2004] and illustrated in Figure 3.27.

After transformation, events that missed the detectors or that exceeded the maximum ring difference were discarded. Therefore, to address the loss of counts, the reconstruction was multiplied by a global factor of $\frac{1}{1-(1-f)}$ where $f$ is the fraction of discarded events; reported as 0.08 in [Fulton et al. 2004]. To assess the accuracy of the MAF and the LOR rebinning methods, the summed squared differences between both the reconstructions and a motion free image were compared to a reconstruction that received no motion correction. The MAF and LOR rebinning reduced the summed square differences by factors of about 7 and 10, respectively. However it was noted that the external tracking device was not accurately synchronised resulting in errors of about 1 or 2 seconds between the two data sets.

A method to overcome the shortcoming of basic LOR rebinning while taking the advantages offered by the latest PET hardware (such as high resolution, high sensitivity and the option for iterative reconstruction) motivated [Carson et al. 2003] to design a framework for Motion-compensation OSEM List-mode Algorithm for Resolution recovery Reconstruction, or MOLAR, for use with the HRRT. Using motion data from an external
CHAPTER 3. MOTION TRACKING AND CORRECTION

Figure 3.27: Detected LOR \( i' \) is transformed to LOR \( i \) however the corresponding transformed points of detection \( Q1' \) and \( Q2' \) no longer lie on the detector surface and extrapolation needs to be performed to obtain \( P1' \) and \( P2' \). Figure taken from [Buhler et al. 2004].

Due to patient motion, events can exit the FOV that would have otherwise been detected, resulting in a loss of data. Conversely, events that are not normally detected can move into the FOV and following motion correction these detected events do not correspond to actual detector pairs as in Figure 3.28.

Figure 3.28: (Left) Illustration of (1) a LOR that would have been detected but due to motion, no longer corresponding to a detector pair, and (2) a detected LOR that when motion corrected, does not correspond to a detector pair. (Right) Detected LOR \( i' \) is transformed to its position had no motion occurred by the operator \( \mathcal{L}(i) \). Image voxel \( j' \) is similarly transformed with operator \( \mathcal{M}(i) \). Figures taken from [Rahmim et al. 2008a].
These two issues have been shown to decrease signal to noise and produce artifacts, respectively [Buhler et al. 2004] [Rahmim et al. 2004] [Thielemans et al. 2004] [Qi et al. 2002] proposed to model motion into the system matrix of the expectation maximisation algorithm. The system matrix determines the probability of an event originating in voxel \( j \) being detected along LOR \( i \). This approach incorporated both of the above issues and was implemented in [Rahmim et al. 2008a] on a simulated brain phantom. The following summarises the histogram-mode algorithm for motion compensated reconstruction. A scan, of duration \( T \), is divided into \( Q \) motion-intervals \((t = 1, \ldots, Q)\) each with time duration \( \Delta T_t \), such that within each interval movements are less than a small threshold. The authors argue that this can be small without increasing the computational load excessively and use a motion threshold of a tenth of the scanner resolution.

Motion corrected sinograms are obtained by transforming detected LOR \( i' \), to the positions they would have occupied if no motion had occurred, \( i \), during time interval \( t \). The invertible operator \( L_t() \) can be obtained from an external motion tracking device and acts such that \( i' = L_t^{-1}(i) \). The probability of an event \( j \) being detected along LOR \( i' \) during time interval \( t \) (i.e., prior to motion correction and therefore will be binned along \( i \) following motion correction) is given by Eq 3.11.

\[
P_{i'j} = g_{ij} A_i N_i' \delta_{i'}
\]

Where the geometric \( g_{ij} \), attenuation \( A_i \), and normalization \( N_i' \) factors are time-varying due to motion and \( \delta_i \) corrects for LOR that fall outside of the FOV due to motion by

\[
\delta_i = \begin{cases} 
1, & \text{if LOR corresponds to a detector pair} \\
0, & \text{otherwise}
\end{cases}
\]

Following motion correction, the overall probability \( p_{ij} \) of detected an event originating in \( j \) and being detected along LOR \( i \) any time during the scan \((t = 1, \ldots, T)\) is written as

\[
p_{ij} = \sum_{t=1}^{Q} p_{i'j} \frac{\Delta T_t}{T} \text{ with } i' = L_t(i)
\]

Combining Eq 3.11 and Eq 3.13, the overall system matrix is given by

\[
p_{ij} = g_{ij} A_i \sum_{t=1}^{Q} N_i' \delta_{i'} \frac{\Delta T_t}{T} \text{ with } i' = L_t(i)
\]

The authors propose a method that avoids the computationally intense expression \( \sum_{t=1}^{Q} N_i' \delta_{i'} \frac{\Delta T_t}{T} \) by performing the sensitivity term calculation in image space. This is possible if the sinogram bins are pre-corrected for attenuation. This derivation can be found in [Rahmim et al. 2004]. To summarise; this method can result in a factor of \( \sim 30 \) increase in speed in the calculation of the sensitivity factors. Unlike MOLAR, which
only uses a semi-random selection of points to determine a global sensitivity factor, pre-correcting for attenuation allows for the accurate calculation of the sensitivity factor at all points.

This motion correction scheme has been further elaborated to include a method that incorporates the presence of scattered and random events which can be applied to various motion correction schemes. A full derivation can be found in [Rahmim et al. 2008b]. This algorithm has been applied to simulated and patient data. The simulated data sets showed fewer qualitative artifacts and better noise/bias trade-off than either no motion correction or a purely LOR driven approach as can be seen in Figure 3.29.

![Figure 3.29: Reconstructed images from simulated data showing the motion free case, non corrected, LOR rebinning only, and the proposed method where the data is pre-corrected for attenuation. Figure taken from [Rahmim et al. 2008b].](image)

Compared to the more traditional image based motion correction, event based motion correction has the following benefits;

1. prevents the PET data from suffering from within frame motion
2. allows for motion correction of early or low count frames
3. reframing the data, which results in frames with lower statistics, is no longer necessary
4. prevents the need to transform (and therefore introduce interpolation effects to) the reconstructed PET images

The limitations of event based motion correction, such as the difficulty in calculating the differential sensitivities of the transformed LORs and the loss of sensitivity due to LORs either moving in or out of the FOV are beyond the scope of this work. It must be noted that one of the aims of this thesis is the ability to obtain accurate and reliable motion tracking, which arguably is more crucial to the success of event based rather than image based motion correction.
3.4. Whole Body Motion Correction

Both CT and MRI have relatively higher spatial resolution than PET. Cone Beam (CB)CT is based on ability to reconstruct 3D objects from cone beam projections [Nalcioglu et al. 1978]. CBCT has been widely used in various healthcare fields, including radiotherapy, where it has been employed to obtain structural information of the patient immediately prior to radiotherapy [Mosleh-Shirazi et al. 1998]. [Jaffray et al. 2002] demonstrates the use of a CBCT in image guided radiotherapy setting however without consideration and correction for respiratory motion, the utility of CBCT for image guided radiotherapy will be diminished. 4D respiratory correlated CBCT [Sonke et al. 2005] uses a externally measured respiratory signal to sort corresponding projections of CB data prior to reconstruction. Motion Compensated (MC) CBCT [Li et al. 2006] uses a motion deformation field, specific to the patient and obtained from a previously acquired 4D CBCT, to correction for respiratory deformation for single CBCT images. A comparison study of respiratory motion correction using 4D CBCT and MC CBCT can be found in [Rit et al. 2011], in which they showed that MC CBCT improved image quality while reducing acquisition time and therefore radiation dose to the patient.

The approach of calculating a patient specific deformation field from a set number of respiratory cycles has also been used in simulated PET/MR data [Tsoumpas et al. 2011]. This technique can potentially be used to enable MRI derived motion fields on emission data obtained on PET/CT or PET only scanners. Problems arise using previously derived motion fields if the subject’s breathing pattern changes after the motion field is calculated.

As indicated in §3.3 the problem of motion correcting whole body PET data is extremely complex and attracts much research interest due to its importance in allowing for small regions of interest to be resolved such as lung lesions in oncology. This is currently an exciting field of motion correction as simultaneous PET/MR has allowed for the possibility of MR sequences to be used to obtained alongside PET data acquisition. However for existing PET or PET/CT systems which cannot use MR navigator sequences, methods need to be used to attempt to reduce the effect of motion in whole body PET.

Gated PET

Gating PET involves splitting the PET data into segments that correspond to specific points along the respiratory cycle. In phase gating a new gate is triggered once the respiratory cycle has reached a specific phase for example every eighth of a respiratory cycle. In amplitude gating a new gate is triggered after motion of a certain amount is observed for example when an external marker placed on the abdomen is moved by 5 mm. In real cases, the observed respiratory cycle is not sinusoidal or regular and in oncology studies the amplitude of motion of the chest surface can be small as some patients do not breath deeply.
Dynamic Reconstruction

The principles of event based motion correction that were introduced in §3.4.1 can be extended to non-rigid applications as in [Gravier et al. 2005] [Mingwu et al. 2007], however care must be taken when considering the varying tracer distribution due to the cyclic motion and the longer drifts that occur during a scan due to either internal physiological processes (peristalsis, digestion, etc) or patient movements. An extension to dynamic reconstruction is presented in [Jin et al. 2006] [Niu et al. 2009] wherein the authors reconstruct and display full 3D volumetric images of a gated cardiac Single Photon Emission Tomography (SPECT) image on two different time axes, one describing the changes over the time scale of a single cardiac cycle whereas the other shows the change that occur gradually over the period of the study. This allows for the intensity variations due to the heart beating to be differentiated from the intensity variations due to the tracer kinetics or organ motion. The authors implemented this scheme on simulated data and noted that images were obtained that showed both time-varying tracer distribution and cardiac motion. Improved SNR was also obtained as compared to reconstruction with only spatial smoothing. For rigid body motion applications, the two time dimensions collapse into the standard 4D reconstruction.

Deformable Mesh Models

The respiratory motion of organs in the thoracic region, such as the heart, have been corrected for by using a rigid body transformation or a centre of gravity approach [Bruyant et al. 2002] similar to as described above. However for oncology it is of interest to study multiple organs which increases the complexity of motion correction. With gated SPECT or PET studies, each gate suffers from a decrease in statistics caused by splitting the dataset into short time frames [Visvikis et al. 2004]. There has been effort into constructing a deformable mesh model for the heart in cardiac gated SPECT since such studies suffer from large movements. To attempt to resolve this [Brankov et al. 2002] [Brankov et al. 2004] have replaced the uniform voxel grid with a deformable mesh model. Since the mesh elements are allowed to deform over time with the motion of the object, it is possible to determine the motion field vector $d_{k\rightarrow l}(x)$ that maps a mesh element $x$ from the current frame $k$ to frame $l$. This improves the SNR of the reconstruction of a frame since it applies temporal smoothing derived from statistics from other frames in the sequence. Other studies have shown similar success [Marin et al. 2009].

Other inter-frame registration methods include an affine transformation of organs as described in [Lamare et al. 2007b] whereby a simulated phantom was considered in three distinct parts; the lungs, the heart and three organs located under the diaphragm. The data was reconstructed into eight frames, one reference and seven other frames that were registered to it. The affine transformations were performed for each of the three parts considered. The optimal transformation is then applied to the original individual LOR prior to secondary reconstruction, as described in [Livieratos et al. 2005]. The authors report
up to 60% improvement in determining the translational position of the simulated lung lesions before and after registration. The authors also determined that it was insufficient to only apply the transformation determined for the lungs onto the other organs, indicating that, as expected, organs such as the liver, stomach and spleen have transformations similar to, but not directly related to, the lungs. This is a limitation of the method as the affine transformation has to be applied to all of the LOR, not just segments.

Further studies in [Lamare et al. 2007a] have extended the list-mode based motion compensation by using non-rigid body transformations. This required b-spline deformations to be derived from 4D CT data and incorporated into the system matrix. More care has to be taken when considering non-rigid transformations of the system matrix than the previously discussed rigid motion system matrix modelling, since post transformation; a voxel may overlap several other voxels of the grid. The authors compared the reconstructions with non-corrected images and with the results from [Lamare et al. 2007b] and showed improved lesion contrast, size and positional information with the non-rigid transformations than the other methods. Other studies with b-spline deformations [Chen et al. 2009], but which only used the PET images themselves for registration, report similar improvements in reconstructed images. However they note that registration errors are introduced due to the high levels of noise in the PET images.

§3.5. Conclusion

Motion has been a topic of great interest since the beginning of PET. With lower spatial resolution scanners the majority of motion induced blurring was hidden within the other more prevailing image degrading factors, such as detector sensitivity, crystal element size etc. However, now that modern PET scanners and reconstruction algorithms have increased the maximum spatial resolution of PET imaging, motion has become the main factor that limits the resolution of PET data. The problem of motion in PET can be subdivided into two parts; motion tracking and motion correction. Motion correction has been researched and investigated with many solutions that can reduce or even eliminate motion from PET data, however they generally require accurate and reliable motion tracking, especially in brain PET. Motion tracking solutions are less numerous and less sophisticated than correction schemes and often consist of simply fixing an object to the patient and then tracking that object. It is the goal of this thesis to develop a system that provides accurate and reliable motion tracking information to be then used in the advanced motion correction schemes.

§3.2.5 introduces a set of criteria that a markerless motion tracking system should achieve for it to be successfully used in a PET environment to track head motion. Indeed, these criteria apply equally to marker based tracking systems, and consideration to them was made when considering what systems should be investigated as part of this thesis. These criteria are necessarily challenging as failure to fulfil even one will reduce the
applicability of the tracking system, even if the system excels in another criteria. The two systems that were used for tracking in Chp. 6 and Chp. 8, the Kinect and the monocular target marker tracking, respectively, were chosen for a range of reasons.

As described earlier the Kinect has relatively low depth resolution of $\sim 2$ mm at a distance of 80 cm and has a minimum operating distance of 40 cm. These affect the potential spatial accuracy and practicality of the Kinect for use in a motion tracking system. In Chp. 6 investigations are done on using the Kinect to track the bulk motion of the head which reduces the effect of the uncertainty on individual points due to the number of sampled points over the head. Additional investigations are performed which use a mirror to enable optimal positioning of the Kinect inside the confined gantry of a PET/CT scanner.

Monocular target marker tracking presents different challenges as it was shown earlier that it has potentially very high accuracy tracking of a single point and can be positioned inside a PET scanner without interference to the subject or scanning procedure. However since this system only tracks the motion of a single point which has been added to the head rather than tracking the motion of the head itself, the tracking will be unreliable if the marker point is able to move non-rigidly to the head. In Chp. 8 investigations are made into the use of multiple small markers which are attached to the head where the separation of the markers can be monitored to determine if the markers are moving non-rigidly to each other and therefore are likely to be non-rigidly moving relative to the head.

This chapter has described many different ways in which motion can be monitored, inferred, or measured, however it should not be considered an exhaustive list. These techniques demonstrate that there are multiple definitions to what consists of a "motion tracking system". In whole body imaging, often it is sufficient to simply obtain the respiratory phase of the subject, so a motion tracking system here may only consist of a method to obtain some surrogate single parameter of surface motion, for example monitoring the pressure changes of a pressure sensitive belt worn by the subject. In brain imaging, the head is often considered to be a rigid body, so the motion of the brain can be measured by monitoring the centroid displacement or any single point added to the head in the form of a tracking marker. Such systems do not measure the surface of the object since it is assumed that it will not deform or move non-rigidly. This assumption does not always hold for medical imaging for example in brain imaging, any change in facial expression will distort the skin surface which may move any single point being independently monitored. In these situations where the surface may be unpredictably deforming, it is necessary to sufficiently densely measure the surface so that the non-rigidly moving regions can be distinguished from the rigid bulk motion of the object of interest.
THE USE OF THE POLARIS POSITION SENSOR FOR MOTION CORRECTION

Abstract

Head motion in PET reduces the quantifiability of PET data by introducing image blur. The motion experienced by the brain is generally considered to be rigid body in nature and as such is well suited to external device motion tracking. The Polaris position sensor is a widely used motion measuring device for brain PET. However, before the Polaris data can be used to compensate the PET data in a motion correction scheme, the Polaris data needs to be spatially and temporally aligned to PET space and scanner time. This chapter describes the procedures developed to enable automatic spatial and temporal alignment for each individual scanning session. The design of the tracking tool and the mode of fixation to the subject’s head is explored with the aim of reducing tracking tool slippage.

4.1 Introduction

As discussed in Chapter 3, the potential benefits of a fully operational and reliable motion correction scheme using external tracking make for a compelling case to use devices such as the Polaris Position Sensor to track subject head movements. This chapter aims to give a technical background into the marker based motion tracking systems that were used at the WMIC and the CIC and the developmental processes that went into optimising these systems for routine clinical use. It must be stressed that without careful set up and detailed knowledge of how to apply movement information to the PET data, the recorded motion data can only be used as a rough indication of whether motion has occurred.

With improvements in the spatial resolution of PET cameras, the detrimental effect of subject motion on image quality increases. Of the approaches to correct emission tomography data for motion, those using external motion tracking are independent of tracer distribution and are therefore preferred in terms of robustness and general applicability. A proposed motion tracking scheme adopted by many PET centres uses an infrared (IR)
reflective tracking tool, consisting of multiple reflective markers in a known arrangement, attached to the subject’s head. Due to the use of multiple imaging devices (PET, CT, IR camera), data from these modalities are acquired in different coordinate systems. Spatial transformations between these coordinate systems need to be established and accurate spatial calibration is required. In addition to spatial calibration it is important to ensure that the data from the different imaging systems can be temporally aligned. This is less important for a frame by frame realignment scheme but is crucial for an event based motion correction approach.

The mode of attachment of the tracking tool to the subject’s head is the most important factor to ensure reliable motion tracking data. [Dinelle et al. 2011] compares motion tracking from a Polaris Vicra to an automatic image registration (AIR) implementation for frame by frame motion correction\(^1\). The investigation was to determine if AIR could be used as a validation tool to detect Polaris scans whereby tool slippage has occurred. This recent paper highlights the difficulty the wider PET community has had in successfully implementing the Polaris as a head tracking system, especially considering that the Polaris was first introduced to PET as a head tracking tool in 1999 [Lopresti et al. 1999]. Two modes of tool fixation were investigated in this thesis, and are discussed in the following sections.

4.2 Methods

Prior to the commencement of these Ph.D. investigations, the Polaris Vicra was set up at the WMIC in 2007 and motion tracking was obtained on various studies. As discussed in §3.2.3 the tracking tool consisted of 4 spheres mounted on metallic screw posts embedded into a 8 mm thick acrylic body. The tracking tool was attached to a wetsuit cap which was worn by the subject. A new tracking tool was designed and built for use in a PET/CT scanner. This method of tool attachment was in use prior to the start of these investigations and has obvious limitations in that the wetsuit neoprene cap has a higher coefficient of friction against the foam headrest than the head has inside the felt inner surface of the cap. The means that if the head moves, it is possible that not all motion will be transferred to the cap and tracking tool, resulting in an under-measurement of the true head motion.

The new tool was fixed to a pair of safety glasses that were worn by the subject, as can be seen in Figure 4.1a. By design, glasses have few points of contact on the head. the points of contact (the bridge of the nose and behind the ears) are not areas that generally move non-rigidly. For this reason, and for the increased comfort to the subject compared to having to wear a wetsuit cap, the glasses based approach was adopted for the studies in the CIC.

The tracking tool was designed to be low attenuating and lightweight. The first models

\(^1\)This paper is further discussed in §5.4 in reference to similar work performed in that Chapter.
of the new tool were manufactured using a 3D printing technique that used a powder resin to form the body of the tool. This produced very lightweight and low attenuating tools, however they were discovered to be too brittle for routine use, therefore a new tool was cut from a single sheet of 8 mm thick acrylic, and can be seen in Figure 4.1b. The good rigidity of the 8 mm acrylic material allowed for much of the body of the tool to be cut away to leave behind a skeleton tool, with a low attenuation profile. The metallic screws used in the older tool were replaced by nylon screws to eliminate metal artifacts that would be produced in a CT scan which could confound the tool detection process.

Figure 4.1: The updated tracking tool which was used for PET/CT brain PET had to have no metallic components. The prototype tool (a) was manufactured using a 3D printer whereas the tool used in clinical studies (b) was cut from a sheet of acrylic.

The method of mounting the tracking tool on a pair of glasses restrained the likely position that the tool will occupy in different scan data. In other words, the tracking tool is worn above the eye line and so will be roughly in the axial centre of the PET field of view and at top of the vertical axis.

4.2.1 Spatial Calibration: Detection of Tracking Tool

To calibrate PET to Polaris space, 4 or more points need to be co-detected in each space. Since the tracking tool consists of 4 spheres it is sensible to attempt to locate the position of these in PET space to then register to Polaris space. In [Olesen et al. 2010] a method for registering PET and Polaris space is presented that performed a single spatial calibration using a high count transmission scan. This is where the transmission source is rastered at a slower rate than normal around the field of view, resulting in a more detailed \( \mu \)-map. This method is described in §3.2.3 however for use at the WMIC, the assumptions of a static position sensor or reference tool could not be made. Hence it was decided to use the routinely acquired normal count rate transmission scan obtained for each subject as the starting point for spatial calibration.

The method developed to detect the tracking tool was performed by implementing
an rigid body Iterative Closest Point (ICP) registration algorithm on a pre-labelled point cloud of the tracking tool (the template) and a point cloud of the subject’s μ-map (the target). ICP is an accepted computer vision technique for registering point clouds that attempts to minimise the difference in distance between two sets of points by iteratively applying spatial transformations [Besl et al. 1992]. Many flavours of ICP exist, differentiating mainly through cost function, however to be used in detecting the tracking tool it was necessary to use an ICP algorithm which contains a robust initialisation technique. Generally, ICP needs good initialisation to help avoid the iterations converging to a local minima. In the open source PointClouds library, PCL, registration between point clouds is achieved by first detecting keypoints that best describe the underlying geometry in the scene. Feature descriptors are created for each keypoint which describes how the keypoints are co-related. This then allows for correspondences to be found in the different point clouds by pairing the most similar feature descriptor histograms. ICP can run on the subset of keypoints rather than the full data set, which reduces computational load and allows for outlying keypoints to be detected and removed from the registration. Once the keypoints have been registered, a second ICP algorithm with finer convergence criteria can be applied to the whole data set that has been roughly pre-registered by the initial use of ICP. The open source 3D mesh viewer, Meshlab, has a implementation of ICP which requires the user to pre-align the point clouds manually. Further discussion on the two implementations of ICP can be found in §6.2.1.

To generate point clouds from μ-map data, an isosurface was applied to the transmission image at a threshold of 0.6 cm⁻¹. The template point cloud was generated from a transmission scan of the tracking tool which had been cropped so that only points from the tracking tool were visible in the template. The coordinates of the centre of the spheres were calculated by taking the centre of mass of the points representing the surface of the spheres. The target point cloud, generated from each transmission scan was axially cropped by 7 cm so that only the tracking tool and the top of the skull was visible. Once the template and the target point clouds have been registered, the pre-labelled sphere coordinates in template space were multiplied by the template-to-target transformation to obtain the spheres coordinates of the target point cloud in PET space. The position of the spheres in Polaris space can now be aligned to the calculated positions in PET space using absolute orientation [Horn 1987].

4.2.2 Temporal Calibration: Data Stream Alignment

Temporal alignment is an essential step to allow for accurate PET data motion correction when using an external motion measuring device, as it is imperative to apply the measured motion parameters to the same time points in the PET data. This procedure is complicated in the case of the Polaris as the hardware lacks the ability to output a time stamp associated with each motion frame, and as a result the frame rate has to be used to infer the elapsed time. The Polaris acquisition software allows for event logging, which
prints the communications between the position sensor and the host computer to a text file. When the acquisition software starts, the local time of the host computer is written to the log file and by determining the time offset between the host Polaris computer and the PET acquisition computer (ACS) it is possible to align the Polaris and PET data streams.

This method assumes the following:

1. The Polaris frame rate is 20 Hz.
2. The Polaris frame rate does not drift over the period of a PET scan.
3. The Polaris host computer local clock does not drift over the period of a PET scan.
4. The PET ACS local clock is reliable.

An experiment was designed using the Polaris Vicra to determine whether the 20 Hz refresh rate was both accurate and reliable. The experiment involved taking photographs of the computer monitor which was running the Polaris acquisition software with a previously calibrated stopwatch also in view. 6 sets of 10 photographs were taken at intervals of 100, 120, 150, 170, 240 and 1315 minutes. Multiple photographs were taken to increase reliability of the measurements by accounting for the exposure time of the photograph and the refresh rate of the computer monitor.

To develop a system that automatically aligns the Polaris and PET space, with consideration to the features listed above, it was decided to inject tags into the PET listmode that corresponded to specific frame numbers of the Polaris data. This removed the need for any computer clock to be accurate or pre-aligned, as the Polaris frame number can be described in PET listmode time by interpolating the tag position as it appears between listmode time tags. To achieve automatic temporal calibration, a circuit was designed and built that directly measures the light output from the position sensor. The light sensor was clipped to the Polaris unit above the LED rings so that it could be triggered from the light output without blocking the LED ring from illuminating the scene. The light output, or Polaris frame rate, is digitally counted using a series of decade counters. Four, 4017B ICs were connected in series with the tenth sequential output of the first three counters being connected to the clock input of the next counter. The tenth sequential output from the first and fourth counters, which correspond to multiples of the Polaris frame 10 and $10^3$ and when a certain number of frames have been measured a gating tag is recorded in the PET listmode. A simplified circuit diagram of the overall layout of the counting circuit is shown in Figure 4.2.

Frame 10000 relates to 8 minutes 20 seconds after the start of Polaris data acquisition. This time delay should ensure that the first occasion of the slow pulse will be seen in the PET listmode data rather than the transmission listmode since the transmission scan takes 6 minutes 6 seconds to complete. The frame $\times 10$ fast tags are used to ensure any slight drifting or variations in Polaris frame rate are noted in the PET listmode.
Figure 4.2: A series array of decade counters is used to digitally count the Polaris frame rate by detecting the light output using a photoresistor. The fast gate is triggered every time the first decade counter receives 10 input pulses. Additionally at every 10 pulses, the second decade counter is triggered and so on until the final decade counter in the chain counts 10 pulses which corresponds to the 10000 Polaris frame and the slow gate is triggered.

4.2.3 Applying Motion Tracking Data to Image Based Motion Correction

Once the motion tracking data had been spatially and temporally aligned with the recorded PET data, a frame by frame motion correction scheme was constructed. Prior to emission image reconstruction, the \( \mu \)-map is transformed using the Polaris measured transformation between \( \mu \)-map space and emission frame space. The PET data is then reconstructed with each frame having unique attenuation data. The reconstructed PET frames are then transformed back into a common reference frame, which was selected to be \( \mu \)-map space. To perform these transformations using the Polaris tracking data, the transformation calculated in the auto-calibration procedure, \( T_{\mu2pol} \), is used to first transform the \( \mu \)-map into Polaris space. Then the \( \mu \)-map is transformed by the mean transformation between reference space and each frame, \( i \), \( T_{ref2i} \), as measured by the Polaris. The \( \mu \)-map is then returned into PET space using the transformation \( inv(T_{\mu2pol}) \) resulting in a total transformation for each frame given by Eq. 4.1.

\[
T_i = T_{\mu2pol} \times T_{ref2i} \times inv(T_{\mu2pol})
\]  

(4.1)

4.3 Results

4.3.1 Tool Detection

Registration of PET to Polaris space requires at least 4 points to be co-imaged in each space. A registered target and template point cloud from the ICP alignment routine is shown in Figure 4.3. ICP converges to the correct solution even if part of the target is
missing due to being outside of the FOV, as can be seen at the top of the glasses based point cloud in Figure 4.3. To give a metric of success, the ICP algorithm returns the mean distance between points in the template and to the nearest point in the target point cloud. For the scans shown in Figure 4.3, the wetsuit cap ICP algorithm reported a mean error between nearest points in the two clouds as 0.8 mm and the glasses based tool had a mean error of 1.0 mm. Using all the points in the templates the RMS distance ± standard deviation between nearest neighbouring points was calculated to be 1.2 ± 0.5 mm for both sets of point clouds. The colour scale used in Figure 4.3 shows the distance between nearest neighbour points between the registered point clouds (the Hausdorff distance). The reliability of the motion tracking depends on the accurate alignment of PET and Polaris space, hence the need for accurate determination of the position of the tracking tool. Even though the RMS distance between points in the target and template point cloud is 1.2 mm, the error on the position of the tool (its centre of mass) decreases with the number of points in the cloud. For reference, the glasses based tool shown in Figure 4.3 contains over 2000 points.

Figure 4.3: Meshlab ICP alignment of the tracking tool to a prepared tool template for the wetsuit cap tool (left) and the glasses based tool (right).
4.3.2 Temporal Alignment

After spatial registration, it is essential to temporally align the motion tracking data to the PET listmode. For event based motion correction schemes it is crucially important to not miscorrect the LORs for motion that occurred at a different time. Initial studies into the temporal accuracy of Polaris data investigated the sampling rate of the Polaris unit. Since no time stamp is given to Polaris frames it was important to validate the frame rate of the sensor.

Figure 4.4: A regular drift of 4.1 seconds per hour of the Polaris Vicra compared to a stopwatch. Each apparent data point is actually a cluster of data points, due to the multiple photographs that had been taken. $8 \times 10^4$ seconds equates 22.22 hours.

The photographs taken to measure the Polaris frame rate indicated that the Polaris Vicra unit tested did not operate at 20 Hz and gains 4.1 seconds per hour compared to the stopwatch. The Polaris unit acquires a single measurement of the pose of the marker for every image frame, as described in §3.2.3. This is the Polaris frame rate which was assumed to be fixed at 20 Hz. The results showed that this was a constant factor and did not appear to vary. The counting circuit was used so that the PET elapsed time can be assigned to each Polaris Vicra frame. The PET elapsed time is interpolated over the Polaris Vicra data using the inserted fast tags that correspond to every 10 Polaris frames. Data interpolation was also needed due to occasional skipped frames in the Polaris data. This was observed when inspecting the Polaris Vicra data and some data entries were missing (∼100 frames per hour). While this was a relatively small number of missing frames, if not accounted for, it could affect a large proportion of the Polaris data.

4.3.3 Motion Correction Using Spatial and Temporal Alignment

Figure 4.5 shows the summed image of the reconstructed frames of a uniform $^{18}$F phantom before and after motion correction. The data set was reconstructed without issue from the additionally inserted tags. On inspection of the emission and transmission list mode, there was no effect on the emission data but in the transmission data some 32 bit
words were missing. This did not affect the attenuation data and indeed this loss of words has been observed in high count rate sessions. The extra two 64 bit words that the scanner has to process per second is believed to be an insignificant load compared to the amount of events processed during a transmission scan.

Figure 4.5: A summed image before motion correction where different colours represent the different frames is shown on the left and the right hand side shows the summed frames after frame by frame motion correction using Polaris Vicra motion tracking.

4.4 Discussion

Since the Polaris was first examined for use in brain PET [Lopresti et al. 1999], it was hoped it would become the standard for motion tracking using an external measuring device. From inception it was known that secure fixation of the tracking tool would be the crucial factor for using the Polaris for brain PET. Other barriers to overcome before the Polaris data is used for motion correction, such as Polaris to PET alignment (spatial and temporal), have been solved in a variety of methods. In [Olesen et al. 2010] a high count rate transmission scan was obtained of the tracking tool in which it was assumed that the tool was aligned in a single transaxial plane and that the centre of mass of the spheres corresponded to the tip of the metallic screws. If this calibration is set up perfectly then this routine would provide an accurate measurement of the position of the tracking tool. However it was the complexity of setting up a calibration routine such as this which prompted adapting this technique to a more general case. An iterative closest point algorithm was implemented to register a tracking tool template to the $\mu$-map. By using the entire tracking tool, rather than just the spheres, the object that is being located becomes larger and less weighting is given to each point. In other words, in the likely event that the tracking tool moves during the transmission scan, the blurring effect would be more pronounced for individual tracking spheres and has less of a significant effect on
the shape of the tracking tool as a whole. This approach allowed for automated Polaris to PET space calibration to be performed on every scan including legacy data.

A desire for succinctness and automation led the approach to temporally align the two data streams. Experience with the Polaris Vicra at the WMIC had shown that using local time stamps to align data streams became inaccurate as the clocks began to drift relative to each other. Additionally, Polaris data does not contain timing information with each frame. Rather it assigns a frame number and it assumes a constant frame rate of 20 Hz. This frame rate was measured using calibrated stopwatches and was shown to be not the case. Over an hour, the WMIC Polaris Vicra suffered a 4.1 second drift, which may seem inconsequential if using frame by frame realignment, however for event based motion correction any temporal discrepancies would greatly affect the quality of motion correction. It was decided that the best approach short of injecting the PET listmode with the actual Polaris measured quaternion transformations, would be to inject timestamps into the listmode that corresponded to the Polaris frame rate. This was performed using an independent hardware solution directly measuring the light output and therefore frame rate of the Polaris position sensor.

An additional barrier to motion correction being routinely used is the effort needed to apply the various motion correction techniques to the PET data. Therefore effort was made to create a simple and robust environment in which to apply motion tracking to PET listmode data for all PET scanners used in these investigations. Automated routines were created that accurately and quickly aligned PET space and Polaris space by detecting the tracking tool in the transmission image. This step is a necessary alternative to simply calculating this spatial registration once and using it over multiple scanning sessions. This is because if the reference tool moves or is knocked between scanning sessions, the Polaris data can no longer be accurately described in PET space. Temporal alignment was solved by using the IR pulse detection and gate pulse injection circuit. This simple method injects gate tags into the listmode to allow for the Polaris data to be accurately aligned with the PET events, without having to account for any local computer clocks that may drift relative to each other. To improve upon the counting technique it is planned that a fifth decade counter will be added and the slow gate tag will be replaced by multiple slow gates that are uniquely triggered by using the first, third, and fifth sequential outputs on the new decade counter. For example, HRRT gate 0 could be used for the fast gate with HRRT gates 1,2,3 used for the slower pulses from frames $10^3$, $3 \times 10^3$, $5 \times 10^3$, or 8 minutes 20 seconds, 25 minutes 00 seconds, and 41 minutes 40 seconds after Polaris data acquisition start. This would allow for temporal alignment to be achievable even if there is a longer than usual pause in between starting the Polaris system tracking and the start of the emission scan.
4.5 Conclusion

A method has been developed that enables motion correction of PET data using motion tracking data from the Polaris position sensor. The tracking information provided by the Polaris exists in an external space and time frame to the PET data therefore steps are needed to temporally and spatially align the Polaris and PET data before motion correction can be performed. These steps are automatically implemented in a combination of Matlab and C++ scripts. A new tracking tool and method for inputting gating tags into the PET listmode were designed and constructed to be used on clinical PET systems. The need for an improved tracking tool came from the observed failings of the wetsuit cap based design. For the patient, the wetsuit cap is uncomfortable. The nature of the cap also results in points of friction between the headrest and cap. This can result in the head moving within the cap and not transferring motion to the tracking tool and the Polaris is therefore not able to detect all patient motion. Velcro is also not suited to rigidly secure two surfaces together as the microfibres and hooks can flex under strain.
CHAPTER
FIVE

ANALYSIS OF CLINICAL DATA RECONSTRUCTED WITH MOTION DATA COLLECTED BY THE POLARIS POSITION SENSOR

Abstract

The Polaris position sensor is a widely used device for measuring rigid body head motion that requires the subject to wear a tracking tool that is monitored using a pair of IR cameras. Clinical PET data was obtained with motion tracking data from the Polaris position sensor. The PET data was reconstructed without motion correction and with Polaris enabled frame-by-frame motion correction. To assess the quality of the Polaris motion tracking data, the motion parameters are compared to automatic image registration. Results are presented of cases where motion correction has appeared to have successfully been applied and in cases where the Polaris motion tracking data has failed to reliably track head motion.

5.1 Introduction

Since the Polaris position sensor was first investigated as a device for tracking subject head motion for brain PET [Lopresti et al. 1999], it has become a standard device for brain PET research. Despite being 14 years since its first use in brain PET, the Polaris remains a complex system and requires many processing steps before the motion tracking data can be used in a motion correction scheme. These spatial and temporal calibration procedures can be performed in a variety of methods however they are often time consuming [Olesen et al. 2010] and so a fully automated method for spatially and temporally aligning Polaris and PET data was devised in §4.2. Automation was an important factor when designing the calibration techniques as one of the disadvantages of independent source motion tracking is the extra work needed to do in order to correct the PET data for motion. Image based coregistration, conversely, requires no extra procedure during scanning and the registration jobs can be queued up and left overnight.
This chapter contains the results of clinical data obtained at the Wolfson Molecular Imaging Centre, Manchester, and the Clinical Imaging Centre, Hammersmith, London. The data originating from the WMIC was taken on the PET only HRRT using the Polaris Vicra and had already been obtained prior to the start of these investigations. This caused issues with handling the motion tracking data as the quality of this data varied from scan to scan due to an unoptimised and non-standardised acquisition process. The improvements listed in §4.2 were directly influenced by handling the legacy data. The initial set up of the Polaris Spectra at the CIC on a TrueV and HiRez Biograph PET/CT included these improvement and as a consequence the motion tracking data from the CIC was of a generally higher quality. Image based analysis to assess the quality of motion correction on three scans from each centre is presented showing examples of both successful and unsuccessful motion correction. For the WMIC scans, the Polaris motion tracking data was used in a frame by frame motion correction technique and compared against an automatic registration algorithm using Mutual Information (MI) using the same frame definitions and the control case where no motion correction had been applied.

5.2 Methods

5.2.1 Motion Correction of HRRT Data Using the Polaris Vicra

Motion data from legacy HRRT brain studies is presented from the Polaris and Centroid analysis. The presented Polaris data is the x,y,z position of the tracking tool during the period of motion acquisition. The Centroid data displays the x,y,z1,z2 measured parameters (see §3.2). Additionally, data is presented showing the displacement of the central image voxel as it has been transformed in each image frame by the mean measured Polaris transformation and the calculated Mutual Information (MI) transformation.

The 13th PET emission frame was chosen to be the reference frame for the MI coregistration. MI coregistration was performed using the Multi-Modality Matching (MMM) plug-in in the Vinci medical image viewer platform. All other frames were reconstructed without attenuation correction and then were registered to the reference frame. The transmission image was also registered to the reference frame, and then was registered to all other PET emission frames using the inverse of the other frames transformations to reference space.

Polaris to PET space calibration was not performed on the original \(\mu\)-map, rather spatial calibration was performed on the \(\mu\)-map which had been previously transformed into the 13th emission frame space using the MMM plug-in. This ensured that the Polaris tracking data was in the same reference space as the MI corrected images. This method removes any possibility of errors in the Polaris measurement affected the spatial alignment.

On a scan where there was a large difference in the mean measured Polaris and the calculated MI frame transformation, a time activity plot, showing the activity in a volume
of interest on a region of high contrast, was performed to investigate if the discrepancy affects the quality of motion correction.

5.2.2 Motion Correction of TrueV Data Using the Polaris Spectra

The data presented in this section was obtained as part of a $^{11}$C-PHNO PET/CT study with permission from Dr. Eugenii Rabiner. The PET/CT data was reconstructed using the offline e7_recon tools provided from Siemens Healthcare. These tools are identical to the programs used on the reconstruction PC attached to the scanner. Specific thanks go to Dr. Mike Casey and Dr. Judson Jones for enabling this.

To analyse the PE/CT data obtained with Polaris motion tracking data, a threshold was applied to specific slices in each reconstructed frame to segment the outline of the head. The Matlab function `bwboundaries` was used to obtain the pixel list of the perimeter of the segmented binary image. The boundary from each frame was weighted by the length of the frame and then summed over all 26 frames. For a static object, the perimeter should remain stationary and the final summed image should be a contour of the object 1 pixel thick with each pixel value in the contour being 5400 (the length of the scan in seconds). In reality, the early frames show poor statistics and the boundary function often includes FBP artifacts, however since these are less weighted than later frames where the boundary is better defined, the method can still be used as a guide to determine if motion correction has been successful in the later frames. The results are given as a percentage of the mean contour length of the final 10 frames.

5.3 Results

WMIC data

In this subsection, motion tracking data from the WMIC is presented from centroid analysis, MI coregistration, and the Polaris Vicra. The following three data sets are from subjects who undertook a 120 min PET acquisition following injection of the 5-HT2A receptor selective tracer $^{11}$C-MDL100907 where the subject was tracked using the Polaris Vicra with the wetsuit cap attachment. For the MI and Polaris Vicra comparisons, the centre of the PET volume was used as a test point and was transformed according to the motion parameters calculated with the two methods. The displacement in x,y,z and the total displacement is then compared.

The mean differences in displacement of the test point when transformed with Vinci MMM and the Polaris Vicra of these data sets are summarised in Table 5.1. Further analysis of subject h00207_01 was performed by creating Time Activity Curves (TAC) for a Volume Of Interest (VOI) over a region of high contrast. In the TAC graph, the Vinci image based coregistration method produces a smoother curve indicating a more successful motion correction than the Vicra. In Figure 5.4 two regions where the TAC
Figure 5.1: Subject h00262_01. (a) Shows the Polaris measured displacement of the tracking tool. (b) Shows the temporally aligned centroid tracking data of the displacement of the centre of mass of activity. (c) Shows the displacement of the central voxel in PET space using Polaris and MI motion tracking. (d), (e), (f) show the displacement of the central voxel in axis x, y, z, respectively. Subject h00262_01 displays motion drift until 40 minutes after which the motion is predominantly short jumps followed by periods of low motion. The data driven methods correlate well with the external Vicra monitoring system, suggesting that the tracking tool did not slip during this scan.
5.3. RESULTS

Figure 5.2: Subject h00228_01. (a) Shows the Polaris measured displacement of the tracking tool. (b) Shows the temporally aligned centroid tracking data of the displacement of the centre of mass of activity. (c) Shows the displacement of the central voxel in PET space using Polaris and MI motion tracking. (d), (e), (f) show the displacement of the central voxel in axis x,y,z, respectively. The motion of the tracking tool in its local coordinate system shows that the tracking tool undergoes periods of rapid twitching and has a long drift in Tx and Tz over the 120 minutes. Centroid analysis of this scan has difficulty in resolving the rapid twitches and the long drift.
Figure 5.3: Subject h00207_01. (a) Shows the Polaris measured displacement of the tracking tool. (b) Shows the temporally aligned centroid tracking data of the displacement of the centre of mass of activity. (c) Shows the displacement of the central voxel in PET space using Polaris and MI motion tracking. (d), (e), (f) show the displacement of the central voxel in axis x, y, z, respectively. Subject h00207_01 initially shows a jump in the Vicra data between the TX position (1st minute) and the start of the PET emission (8th minute). Following this the motion of the head is relatively stable until a large motion is observed in the Polaris data around the 60 minute mark. Peaks are also observed in the centroid and Vinci MMM data however they are much smaller in magnitude.
Table 5.1: Root Mean Square Difference ± Standard Deviation and Maximum Range in the differences from calculated from transforming the centre of PET space using Vinci MMM co-registration and the Polaris Vicra. Data is given in the format Mean ± StD (maximum difference).

<table>
<thead>
<tr>
<th>WMIC ID</th>
<th>D / mm</th>
<th>Dx / mm</th>
<th>Dy / mm</th>
<th>Dz / mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>h00262_01</td>
<td>0.1 ± 1.0 (3.2)</td>
<td>0.4 ± 0.6 (2.1)</td>
<td>3.9e-2 ± 0.4 (1.0)</td>
<td>0.7 ± 0.9 (3.7)</td>
</tr>
<tr>
<td>h00228_01</td>
<td>1.0 ± 1.6 (4.1)</td>
<td>0.3 ± 1.1 (3.6)</td>
<td>0.2 ± 0.9 (1.6)</td>
<td>0.5 ± 2.1 (4.5)</td>
</tr>
<tr>
<td>h00207_01</td>
<td>4.2 ± 3.5 (14.5)</td>
<td>0.4 ± 2.2 (5.6)</td>
<td>0.4 ± 1.8 (3.5)</td>
<td>1.5 ± 5.5 (13.6)</td>
</tr>
</tbody>
</table>

differs by over 5% are highlighted.

Figure 5.4: TAC analysis of a VOI of a high contrast region in h00207 show the Vicra motion correction inducing image errors by incorrectly correcting for falsely observed motion. The transaxial slice outlines where the VOI (shown in red) were created.

CIC data

The following three scans were taken as part of a $^{11}$C-PHNO PET/CT study. The Polaris motion corrected reconstructed images are compared to uncorrected reconstructed images using a novel contour analysis technique.

Table 5.2: Percentage of pixels with intensity greater than 3000 compared to the mean number of pixels in the final 10 frames. Results are compared between the three studies without motion correction (non) and with motion correction (MC).

<table>
<thead>
<tr>
<th></th>
<th>107 non</th>
<th>107 MC</th>
<th>102 non</th>
<th>102 MC</th>
<th>106 non</th>
<th>106 MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transaxial</td>
<td>14.3 %</td>
<td>26.9 %</td>
<td>1.8 %</td>
<td>20.6 %</td>
<td>0.9 %</td>
<td>5.8 %</td>
</tr>
<tr>
<td>Sagittal</td>
<td>32.8 %</td>
<td>38.9 %</td>
<td>19.2 %</td>
<td>28.3 %</td>
<td>1.5 %</td>
<td>2.6 %</td>
</tr>
<tr>
<td>Coronal</td>
<td>43.1 %</td>
<td>52.5 %</td>
<td>16.6 %</td>
<td>33.2 %</td>
<td>7.2 %</td>
<td>20.3 %</td>
</tr>
</tbody>
</table>
CHAPTER 5. ANALYSIS OF POLARIS MOTION TRACKING

Figure 5.5: Subject GSK111013_000107. (a) shows the displacement of the tracking tool in mm. Motion has mainly occurred in the Tz and Ty axis, i.e. the head has tilted back in the head rest. There is minimal motion in the Tx axis which indicates that no side to side movements occurred.

Figure 5.6: Subject GSK111013_000107. Transaxial, coronal, and sagittal summed contour images at slice number 75,110,44 respectively of the uncorrected data (a) and Polaris motion corrected data (b). In the sagittal plane an interesting feature can be seen at the top ridge of the eye socket. This corresponds to the tracking tool position and is likely an artifact caused by scattered LORs.
Figure 5.7: Subject GSK111013_000102. (a) shows the displacement of the tracking tool in mm. Motion drift and short discrete movements are visible in each axis. Motion in Tx corresponds to the side to side rotation of the head. This motion is the least desirable as the arms of the glasses are likely to pinch the headrest if the head undergoes this type of motion.

Figure 5.8: Subject GSK111013_000102. Transaxial, coronal, and sagittal summed contour images at slice number 75, 100, 35 respectively of the uncorrected data (a) and Polaris motion corrected data (b).
Figure 5.9: Subject GSK111013_000106. This subject presented severe motion during the scanning session. Motion is mainly in the Ty axis which corresponds the subject lowering into the headrest.

Figure 5.10: Subject GSK111013_000106. Transaxial, coronal, and sagittal summed contour images at slice number 75,100,35 respectively of the uncorrected data (a) and Polaris motion corrected data (b).
5.4 Discussion

The results in this section do not seek to compare different motion correction techniques. Frame by frame realignment was used so that a comparison can be made between the Polaris motion tracking and Mutual Information image registration, which uses the standard predefined frames. Discussions on the benefits of multiple acquisition frames have been made [Fulton et al. 2004] [Picard et al. 1997] however the creation of shorter than standard frames to correct for within frame motion can result in decreased signal to noise for iteratively reconstructed data [Montgomery et al. 2006].

WMIC Data

Motion analysis has been performed on HRRT brain PET data using various methodologies. Polaris Vicra motion tracking data has been compared with centroid analysis and MI coregistration, two examples of data driven techniques. The motion tracking data shown here presents a typical cross section of scanning sessions at the WMIC. Subject drift occurred in most scans and was mainly seen in the Tz (axial PET) plane. Theoretically, the Polaris should be the best at detecting and measuring subject drift, since centroid analysis, and to a lesser amount automatic image registration, are affected by tracer redistribution [Hill et al. 2001]. Centroid data generally always begins with a rise in the Tz plane as the activity arrives from the neck to spread across the brain, which manifests as an apparent positive translation in Tz. Another feature common to centroid data is an increase in noise with time, as the count rate drops. This noise increase can make it difficult to observe motion in the centroid data. Structure can be detected in late centroid data by applying a temporal smoothing filter however due to the initial sampling rate of 1 second, rapid motions will not be able to be detected, especially when smoothed.

The TAC analysis shown in Figure 5.4 shows an extreme situation where correcting the PET data using Polaris motion tracking has severely and obviously degraded image quality. Due to the ill posed nature of kinetic analysis used to extract and fit kinetic parameters to PET data and TACs, small errors in TAC data can propagate and result in large errors in kinetic parameter estimation. In this case it is obvious that the Polaris data is inaccurate and should be discarded, however it is not always this clear.

CIC Data

Movement in the Tx direction corresponds to lateral motion of the head as the subject tilts their head in the headrest. It became apparent when inspecting the displacement data on scans where the frame by frame motion correction has obviously failed, that large discretisations appear in the Tx data. On scans where the motion correction appears to have been successful, the motion in Tx is smooth, such as in Figure 5.5a. These jumps in the Tx data are most likely due to the arms of the glasses becoming pinched on the head rest causing the tracking tool to start moving non-rigidly with the rest of the head.
The contour analysis was performed in an attempt to provide a simple metric quantifying whether motion correction has been successful. It is generally insensitive to early frame motion as the threshold used to detect the contour often produces internal contours or detects FBP artifacts in low count frames. It is possible that a dynamic threshold based of the maximum pixel intensity in each frame could reduce internal contours from appearing in the early frames.

The following equation presents a method developed in [Hafezian et al. 2012] where motion correction efficacy can be evaluated by comparing the mean of the sum of the difference between voxels in sequential time frames, using the equation

$$R = \frac{1}{T-1} \sum_{t=2}^{T} |V_t - V_{t-1}|$$

(5.1)

where $T$ is the maximum number of time frames $t$ of image data $V$. The range of the following percentage differences agree with the authors analysis on their own data between non motion corrected and motion corrected images.

Table 5.3: The mean of the sum of the difference between voxels in sequential frames is shown below for the three data sets. All data sets show a negative percentage change between non motion corrected and motion corrected, indicating that there is less variation in sequential voxels.

<table>
<thead>
<tr>
<th></th>
<th>non MC</th>
<th>MC</th>
<th>% diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>107</td>
<td>$2.865 \times 10^7$</td>
<td>$2.561 \times 10^7$</td>
<td>-10.6%</td>
</tr>
<tr>
<td>102</td>
<td>$2.772 \times 10^7$</td>
<td>$2.589 \times 10^7$</td>
<td>-6.6%</td>
</tr>
<tr>
<td>106</td>
<td>$2.905 \times 10^7$</td>
<td>$2.517 \times 10^7$</td>
<td>-13.4%</td>
</tr>
</tbody>
</table>

At this point, it becomes difficult to directly compare the results from Table 5.3 and Table 5.2, as subject 106 has the best percentage decrease and lowest mean voxel value of the three studies, according to Table 5.3. Yet according to Table 5.2, 106 performs the worst, consistently having fewer contour pixels above the 3000 second threshold than the other two scans. On visual inspection of movies of the image slices, 106 appears to have been least well motion corrected, with 107 appearing to have been motion corrected the best. Although this visual approach is at best subjective, it does appear to correlate more with the results from the contour analysis. These simple comparisons between motion correction quantification schemes, and the difficulty in comprehending the data highlights the difficulty in reliably validating one motion correction scheme against another on clinical, non-phantom, data.

The contour analysis can be improved if the data is reconstructed with an iterative algorithm such as OSEM, if only a few iterations are selected as the reconstructed images will have less noise and better definition at high contrast regions. This can be seen in Figure 5.11 where the images show the same slice from the same subject for frames 2:26 reconstructed with FBP and AW-OSEM.
5.4. DISCUSSION

Figure 5.11: (a) Cumulative sum of transaxial slices of frames 2:26 of subject GSK111013_00107 reconstructed using FBP. (b) Cumulative sum of transaxial slices of frames 2:26 of subject GSK111013_00107 reconstructed using AW-OSEM with 2 iterations.

Reconstructing the data using OSEM should allow for the contour analysis to be able to describe the effect of motion on earlier frames than the FBP images. Finally, the use of attenuation correction in the reconstruction algorithm can potentially affect the contour analysis results. Misaligned \( \mu \)-map and emission data will affect the intensity of the regions in which there is measured activity, but due to the misalignment, no attenuating matter. In these results, the \( \mu \)-map was transformed to emission space, however it is incorrect to assume that this transformation is accurate.

[Dinelle et al. 2011] present a paper similar in principle to the work in this chapter. In it, the authors seek to use automatic image registration to validate the Polaris Vicra on a scan by scan basis so that the Vicra tracking data does not have to be manually inspected for reliability. Similar to the results presented in this chapter the authors show data sets with TACs that are negatively affected by incorrect Polaris tracking data. The authors report that 11\% of their protocol-accepted images were determined to be falsely accepted, with discrepancies of these incorrectly accepted images ranging from 1.2 to 3.6 mm. Their proposed method cuts down on manual inspection time, but does not completely remove the need of a trained user to evaluate images when they are flagged on comparison with an AIR corrected image.

In agreement with the results in this Chapter, [Dinelle et al. 2011] found that care needed to be taken to ensure that the Polaris data was reliable, as correcting for falsely measured motion will degrade the quality of the PET data. Whereas the authors developed a semi automated check for quality assurance, it was the decision for the future work in this thesis to investigate and develop a new method of motion tracking, which has reliable
tracking.

5.5 Conclusion

This chapter presents results from historical WMIC PET data that was obtained using motion tracking information from the Polaris Vicra and the wetsuit cap and from PET data from the CIC obtained using the glasses based attachment. Validating the success of a motion tracking system such as the Polaris is complicated. It has been shown to be successful in phantom work however it has been demonstrated that motion tracking of clinical patients using the Polaris is inherently unreliable as tool slippage can occur and it is difficult to accurately determine when and by how much. By comparing the external motion tracking to a data driven technique, one can validate the reliability of the external tracking device down to the spatial and temporal resolution of the data driven measurements. However, this reduces the impact of the external tracking system if data driven motion tracking has to be performed for every scan. A combination method could be used so that external tracking is preferred until the agreement between the external and data driven tracking falls below a certain threshold. Past this point, data driven tracking is preferred, or the external tracking system could be reset. Changes to the tool design or mode of attachment can improve the reliability of motion tracking however if the possibility of tool slippage is at all present, then the motion tracking cannot be relied upon. The Polaris position sensor provides excellent motion tracking of the tracking tool however its practicality in a clinical setting is diminished since the tool cannot be comfortably and rigidly fixed to the patient.

The design of the tracking tool and how it is attached to the patient is the limiting factor when using a marker based tracking system, such as the Polaris position sensor. Therefore future work was focused on developing a marker-less tracking system and systems which use smaller markers. Recent advancements in consumer grade IR digital projectors have opened the possibility of an inexpensive coded/structured light 3D scanner. These options were explored with the aim to produce an accurate motion tracking system without the need for attaching markers to the patient or using visible light which could distract the patient. The following chapters set out novel alternatives to motion tracking and their applicability to medical imaging.
CHAPTER SIX

THE DEVELOPMENT, APPLICATION, AND ASSESSMENT OF THE KINECT FOR TRACKING RIGID BODY HEAD MOTION

Abstract

The motion tracking methods described previously rely on the assumption that the tracking tool remains fixed in place during PET data acquisition. Since this is not always the case, and it impossible to determine if the motion of the tool is truly representative of the head motion, novel methods for motion tracking were explored. Advances in the field of computer vision have provided potential solutions to the problem of tracking head motion and some of these approaches were considered and are discussed below. A critical drawback of a Polaris type system is that due to the size of the tool it is only feasible to attach one marker to the subject’s head at any one time and as such it is impossible to determine if any tool slippage has occurred. This chapter aims to investigate if markerless tracking techniques can measure head motion more reliably than the Polaris system, as by their nature there is no tool that could slip.

6.1 Introduction

The past decade has seen various attempts to solve the problem of tracking the rigid body motion that occurs during brain PET (see §3.2). In general, the more popular techniques employ a tracking tool that is fixed to the subjects head and monitored by an external sensor. The Polaris position sensor is such a marker-based system and its spatial resolution is dependent on the shape of the tool and the quality of the cameras used in the sensor, although tool tracking with sub millimetre resolution is reported [Wiles et al. 2004]. For the Polaris, this high spatial resolution, coupled with a frame rate of 20 Hz, encouraged much research interest. Initial findings were promising [Lopresti et al. 1999]. However, it became apparent that tool slippage adversely affected the PET data after motion mis-correction. Effort began in developing markerless methods that would remove
the need of a tracking tool thereby sidestepping the problem of securing the tracking tool to the head.

Early attempts at markerless motion tracking systems, such as [Ma et al. 2009] that focussed on detecting features in video sequences, such as the corners of the eyes, nose, mouth etc. However these techniques were not sophisticated enough to address the main problem with markerless tracking techniques; non-rigid facial feature changes. Without local paralysis, deformations of the face are inevitable and will result in poor tracking. Especially for early attempts at markerless motion tracking which used relatively low resolution digital video cameras which coupled with less computational power, resulted in system that could only detect a small number of features in each video frame. If even a small number of these move non-rigidly, then the tracking accuracy would suffer. An example of the current state of the art in feature detection markerless motion tracking in PET is the work in [Kyme et al. 2012] and the benefits and limitations are described in §3.2. This system specifically tracks features that tend to persist in the video sequence building a library of points that have appeared to move rigidly previously in the sequence. While feature detection tracking can provide very accurate tracking it remains problematic to base motion correction methods on features that may move non-rigidly. A solution to this problem is to detect more features, then use a selection process to remove the non-rigid outliers, such as RANdom SAMple Consensus, RANSAC [Fischler et al. 1981]. However more features equates to more computing time in detecting these points. The logical extension is to detect the entire surface of the face rather than trying to detect a subset of interesting points. The entire surface of the face will inevitably contain regions of points that move non-rigidly but the sheer number of points that move rigidly should, under most conditions, allow for accurate rigid body tracking. Before the release of the Microsoft Kinect, depth sensors were generally large and expensive devices, but with its advent, the Kinect allows for depth measurements to be acquired from a hand-held, consumer grade device and opened the possibility of obtaining surface maps of the head for use in a motion tracking scheme. The Kinect was initially devised for use as a video game peripheral to track the whole body skeletal motion of multiple people, and therefore was not immediately applicable to a rigid body head motion tracking system. This chapter sets out the work that was done to translate the Kinect depth sensing technology into a clinically relevant motion tracking system.

6.2 Methods

6.2.1 Iterative Closest Point

As described in §3.2.6 the Iterative Closest Point, ICP, algorithm [Chen et al. 1992][Besl et al. 1992] is a accepted technique to align multiple point clouds. ICP operates iteratively over the following three steps until the point clouds are converged.
1. For each point in the first point cloud, find its nearest neighbour in the second point cloud.

2. Estimate the rotation and translation using a cost function that best aligns the points matched in the 1.

3. Transform the first point cloud by the transformation calculated in 2.

Convergence can be defined as when the nearest neighbours are separated less than a certain distance. It is important for the point clouds to be roughly aligned prior to ICP to prevent the algorithm converging incorrectly. This pre-alignment can be performed in numerous ways. Meshlab uses a manual multiple point alignment tool which requires the user to specify 4 or more paired points on the two surfaces. Point Cloud Library includes functions to detect a subset of interesting geometric features on the two point clouds. Feature histograms are created which describe the neighbourhood of each feature point and these are then matched to similar histograms in the target point cloud.

![Figure 6.1: Comparison of Meshlab and Point Cloud Library implementations of ICP of the coregistration of two KinectFusion generated point clouds. The KinectFusion normal map rendered surface meshes are shown at the bottom as inserts. The KinectFusion mesh shown on the left has an occluded region behind the nose. This causes a hole in the point cloud which displays edge artifacts when displaying the distance between nearest neighbour points in the co-registered point clouds.](image)

Figure 6.1 compares the result of applying the Meshlab ICP implementation with manual pre-alignment to the PCL ICP implementation with feature point detection as
pre-alignment on two point clouds of a head phantom. The colour map of the point cloud displays the Hausdorff distance from a point to its nearest neighbour in the other point cloud. The importance of this experiment is not to determine which ICP produces the more accurate alignment, rather it demonstrates that there are multiple methods to perform ICP, which can be used according to preference. The feature point detection algorithm requires careful parametrizing for different types of geometrical models however using Meshlab requires manual input and cannot be used in a fully automated pipeline.

Many ICP implementations will specify a maximum distance that is used to locate a nearest neighbour. This value should be used sensibly to ensure that points that come from a region of the surface that only exists in one of the point clouds do not get incorrectly paired.

6.2.2 KinectFusion

Microsoft Research Cambridge developed an algorithm called KinectFusion that implements dense iterative closest point calculations, where every 3D point is used, in real time on Kinect depth data by utilising the parallel processing power of a GPU.

KinectFusion can be summarised in the following process, see also Figure 6.2

1. A block of GPU memory is created that represents the active target volume. A 32 bit $512 \times 512 \times 512$ voxel grid would take up 512 MB of graphics memory.

2. The depth map is converted into a point cloud (a surface vertex and associated normal map).

3. ICP is used to rigidly transform the current point cloud to the previous frame. The transforms are incrementally applied for each subsequent frame to obtain continuous tracking of the Kinect position relative to the scene.

4. The surface of the volume is represented in GPU memory as the average distance, per voxel, to the predicted surface. This is know as the Truncated Signed Distance Function, TSDF. The TSDF contains the running average of the distance for each voxel to the surface as more Kinect frames are integrated.

5. The implicit surface is rendered to the user using raycasting. Over multiple integrations, this raycasted volume becomes less noisy than the individual Kinect depth frames and can be used as the global reference frame for the ICP step, rather than the previous depth frame.

Even though the integration of new Kinect frames into the weighted average volume occurs in real time (i.e. at the 30 Hz frame rate of the Kinect sensor) the generation of a KinectFusion surface cannot be considered real time, since multiple frames need to be integrated before the weighted averaging smooths out the noise, errors, and holes present in individual Kinect frames.
6.2. METHODS

The fine detail of the KinectFusion method is beyond the scope of this thesis, however a full description can be found in [Newcombe et al. 2011] and [Izadi et al. 2011]. The original implementation of KinectFusion uses ICP to track the location of the Kinect depth sensor as it is moved through a static scene. As the Kinect is moved, subsequent Kinect frames are registered and as new views of the scene become available to the Kinect, the holes and the noise on the surfaces gets filled in and smoothed, as shown in the 3D model in Figure 6.3. A global Kinect volume is initiated as a volume of fixed size which is locked to the static structure of objects that fall within this volume. If the Kinect is moved within this scene, new viewpoints offer new depth perspectives of the scene. Each new depth frame is registered to its previous by estimating a 6 Degree Of Freedom (DOF) transformation. These transformations are incrementally applied together with every new depth frame to give the global transform of the Kinect to its starting position in global Kinect volume.

In this case the global Kinect volume is fixed to the scene and the Kinect is moved within this space. However to implement KinectFusion for head tracking, the Kinect sensor is fixed in real space with the head moving freely. As the head moves, KinectFusion calculates the global transformation as before, yet in this case the transformation describes how the global Kinect volume, which is locked to the head, moves around the stationary Kinect.

An open source implementation of the KinectFusion algorithm, KinFu, is provided by the Point Cloud Library [Rusu et al. 2011], PCL, open source project. The GPU code is written in for the CUDA environment which requires the use of an Nvidia GPU. A range of consumer grade computers were used to evaluate the KinFu code. Specifically the GPUs used were an Nvidia GeForce GTX 1.5GB 580, an Nvidia GeForce 3GB 580, and an Nvidia GeForce GTX 4GB 680m (a laptop GPU). Experiments were performed on Windows 7 64bit systems running Nvidia driver 306.23 and CUDA toolkit 4.2. The Kinect for Windows Toolkit (versions ≥1.7) provides an encapsulated version of the KinectFusion algorithm, which claims to run at interactive rates for DirectX 11 compatible
CHAPTER 6. PHANTOM MOTION TRACKING USING KINECT

Figure 6.3: A global volume of $0.65m^3$ is used with $512^3$ voxels and is defined in space by $T_x$ (blue), $T_y$ (green), and $T_z$ (red). The number of voxels is limited by graphics memory on the GPU with $512^3$ 32 bit voxels requiring 512 MB of memory.

GPUs such as the NVIDIA GTX560 or AMD Radeon 6950 (i.e. not only CUDA compatible cards).

Real time processing is a very important factor for motion tracking as most offline markerless based tracking systems take many days to calculate the tracking parameters from the recorded video data. The Kinect system does not store the video data of the patient from the scanning session. The following criteria set out the main areas that any motion tracking scheme needs to satisfy to be considered for use in high resolution brain PET motion correction.

6.2.3 Resolution Test Object

A basic resolution test object consisting of a flat $80 \times 80 \text{ mm}^2$ square piece of paper was placed on an optically transparent sheet of Perspex. KinectFusion was used with 768 voxels per metre resulting in a voxel size of $1.30 \text{ mm}^3$. The Kinect sensor was moved relative to the test object and frame integration was halted when the model did not visibly improve, approximately 2 seconds of data acquisition. The dimensions and topology of the mesh created by KinectFusion was analysed inside Meshlab, by measuring the interedge distance and calculating the distance of each point to a flat surface.

6.2.4 Co-Registering Single Kinect Frames and a KinectFusion Generated Surface

In [Newcombe et al. 2011] an experiment was performed with a fixed Kinect and objects placed on a rotating tabletop. The authors compared the drift characteristics of frame-to-frame tracking to frame-to-model (KinectFusion), where the global volume is...
6.2. METHODS

updated. the frame-to-frame case quickly accrued tracking errors resulting in a poor recon-
truction of the scene, however KinectFusion was able to position the Kinect even after
multiple whole rotations of the tabletop. In this experiment, the roughly 1 m³ tabletop
contained items which provided a lot of depth information, essential for a tracking sys-
tem. For head motion tracking, it is important to determine if there is sufficient depth
information on the face for the Kinect to detect and track.

To investigate registering Kinect frames to a smooth model rather than co-registering
noisy Kinect frames an experiment was performed using an offline ICP implementation.
A head phantom was imaged using the Kinect at 6 different positions within 1 m of the
Kinect sensor. KinectFusion was used to generate a surface map of the head phantom,
using $1.3 \times 1.3 \times 1.3$ mm³ voxels. KinectFusion was used until the model looked visu-
ally smooth. ICP was used to register the Kinect frames and the KinectFusion model
to a single reference position, using the Meshlab Align ICP function. The Meshlab ICP
algorithm was set to use 10000 sample points, and had a target convergence distance of
1 mm. A photo of the head phantom used throughout these investigations is shown in
Figure 6.4. The white, polystyrene head phantom had a reflective surface such that at
distances less than $\sim 80$ cm the Kinect failed to measure depth values reliably. Course
medical tape was applied as an additional reflective surface, however it still lacked the
lambertian properties of real skin.

![Figure 6.4: A photograph of the head phantom used in these investigations. The area where tape was applied to provide an additional and less specular reflective surface can be seen over the eyes.](image)

6.2.5 Practicality

It is essential that any technique used to track head motion can be applied under PET
scanning conditions without interfering with the PET data acquisition. As mentioned
previously the Kinect has a minimal operating distance of 0.4 m. This distance prevents
the Kinect from being optimally positioned face on to the subject inside the confine of the
PET/CT scanner bore. A consequence of the Kinect being placed outside the bore with
a shallow view of the subject is that the surface area of interest extends away from the sensor, which as shown in [Khoshelham et al. 2012] has an increasing effect on the error on the depth measurement. Figure 6.5a illustrates this using a mesh of a subject inside a PET/CT scanner. To limit the effect of increasing noise, the target surface area of interest should be contained within a narrow depth range. An experiment was performed tracking a static head phantom with the Kinect fixed to the outside of the gantry directly viewing the top of the head phantom. If tracking was successful then KinectFusion would have measured no transformations, however it was expected that KinectFusion would have difficulty as the head phantom had little 3D structure from the perspective of the Kinect.

Positioning the Kinect face on provides the best depth data for the ICP registration and minimises the shadows caused by facial features such as the nose. As an initial test to determine if the Kinect could be placed outside of the scanner bore.

![Figure 6.5: (a) The depth resolution as derived from the calibration performed in [Khoshelham et al. 2012] is plotted alongside an image of a subject in a PET scanner. The difference in error between the feet at a distance of 0.4 m and the head at 2.2 m is 13.3 mm. At a distance of 80 cm the depth resolution, i.e. the minimum bound on the accuracy of the depth measurement for each point, is 1.8 mm. (b) The Kinect can be positioned optimally if a mirror is used to reflect the speckle pattern onto the head. Using a mirror allows for the head to appear at a distance of 0.8 m from the depth sensor.](image)

To allow for face on positioning of the Kinect, a mirror was used to reflect the speckle pattern onto the subjects face as shown in Figure 6.5b. The mirror used was a thin (∼2mm) sheet of back surface mirrored acrylic. Front surface mirrors were considered too prone to scratches and more care has to be taken with their cleaning. The acrylic was transparent to the IR wavelength used by the Kinect (830 nm) and no major distortions of the dots were seen on inspection of the speckle pattern using a separate IR camera. In these initial experiments the mirror was arranged flat against the top side of the scanner bore for rigidity. This necessity results in the Kinect and mirror set up shown, where it can be seen that only half of the Kinect data is actually reflected towards the subject. A
static head phantom was used in this set up to determine if tracking could be achieved.

6.2.6 Spatial Accuracy

After determining if the Kinect can be used in real scanning situations, it was necessary to compare the KinectFusion tracking with a current method for external tracking. A tracking tool was rigidly fixed to the back of the head phantom and was monitored using the Polaris Vicra. Simultaneously, the Kinect was placed in front of the head phantom to track the face. Movements such as tilting forward and side to side rotations were applied by hand to the head phantom to simulate typical movements that occur during head scanning. A range of amplitudes and frequencies were applied for 1 minute.

A CT image of the head phantom was obtained to register to the Kinect global model and also to detect the 4 spheres of the Polaris tracking tool so that both the KinFu and the Vicra tracking could be translated into the common CT space.

Previous work with this particular Polaris Vicra [Noonan et al. 2010] showed that the sample rate of the unit was 20.02 Hz. The Kinect tracking data was obtained at a rate of 20 Hz, as measured on the local PC clock and was linearly interpolated to match the Vicra data sample rate. Without interpolation the two tracking streams drift by 1.2 frames per minute.

6.2.7 Robustness and Stability

Markerless motion tracking schemes can be sensitive to occlusions and deformations. Events such as the subject scratching their nose, coughing or talking occlude or distort the surface of the face and are expected to occur at some point during routine clinical use. Evaluating a tracking system’s response to these types of events is difficult as it is not simple to design a phantom or test object to have realistic distortions.

To provide robustness to the surface of the face distorting, or being occluded, the tracking software was written so that the integration of new frames into the KinectFusion global model is paused after a certain period of time. Pausing integration provides robustness against facial deformations as the global model is not incorrectly updated with surfaces from non rigidly moving sections of the face. Additionally, a library of point clouds is constructed of previous camera positions. If tracking is lost, for example if the object is occluded or moves out of the field of view, this library is repeatedly searched until the new depth frame matches a library point cloud. For example, the subject is asked to remain expressionless while the Kinect builds up a detailed point cloud of the head. Integration is then paused and the PET scan is performed.

6.2.8 Temporal Accuracy

To enable the use of Kinect based motion tracking in a PET motion correction scheme it is necessary to temporally align the motion tracking data with the PET listmode. This is
especially important for event by event based motion correction schemes which requires high temporal accuracy. In previous work on temporally aligning Polaris motion tracking data to PET data §4.2.2, a method of directly sampling the clock cycle and therefore the frame rate of the external tracking system was constructed. Attempts to do this with the Kinect were made more complicated as the light output from the Kinect is static. On probing into the electronics inside the Kinect the clock frequency of the components could be isolated and sampled, however it proved impossible to locate a trigger that corresponded to any information of the rate of data transfer to the host PC, i.e. a raw measure of the Kinect’s frame rate on the hardware level. As such temporal alignment was crudely performed using time stamps from the host PC as the Kinect frame is processed by the KinectFusion algorithm, aligned to listmode time using the file creation timestamp in the listmode header.

6.3 Results

6.3.1 Resolution Test Object

Figure 6.6 shows the KinectFusion mesh of the square test object. Artifacts appear at the edge of the square surface and the corners are rounded. Manually measuring the distance between vertices in the corners gives a mean edge length of 78.3 mm. The edge distance between the mid point of the edges was measured at 80.0 mm and 79.3 mm for the orthogonal axis. The diagonal length, which should be 113.1 mm, was measured to be 111.2 mm and 109.1 mm. These values are lower than expected due to the incorrectly reconstructed corner regions. The corners should be right angles however KinectFusion has been unsuccessful at determining the structure and smooths out the corner points, as can be seen in Figure 6.6. This is likely due to interpolation and integration errors, where for small structures or boundary regions, a speckle point may be reflected off the surface in one frame however with motion the speckle point may be moved past the surface. As the Kinect is moved relative to the test object, the weighted average for these regions in the integrated volume is reduced and the corners would appear to be rounded, rather than straight edged. In addition to the rounded corners and the ridged boundary, the centre of the test object shows non-flat structure, which corresponds to the peak in the histogram around 0.45 mm. This small difference may be due to the physical surface being non-flat or due to KinectFusion integration errors.

The resolution of a KinectFusion volume is set at the construction of the KinectFusion application, however the 3D voxel resolution of this volume is not the effective resolution of the generated surface. Resolution, in the case of a KinectFusion model, is dependant on a number of factors such as the distance to the sensor (which affects the lateral separation of adjacent speckle points and the depth resolution of individual points), the number of frames integrated into the volume, and if the motion of the Kinect relative to the surface resulted in the speckle pattern being swept across the surface, allowing for more of the
6.3. RESULTS

Figure 6.6: This figure shows the digital square point cloud (left) the KinectFusion generated point cloud of the square test object (middle) and the KinectFusion point cloud coloured by the Hausdorff distance from each point in in the measured cloud to its nearest neighbour in the digital cloud (right). This histogram, with range from 0 - 1.0 mm, shows the distribution of distances across the measured surface.

surface to be sampled by a speckle point.

6.3.2 Co-Registering Single Kinect Frames and a Kinectfusion Generated Surface

Table 6.1 shows the reported ICP residual error between the transformed subset of points and the nearest point in the target frame. 10000 points were used by the Meshlab ICP algorithm. Table 6.2 shows the RMS distance between all points in the first mesh and the nearest neighbour (within 5 mm) of the second mesh, as calculated by Meshlab’s Hausdorff Distance function. This explains the higher RMS when KF is the second mesh as opposed to the first, as the regions of the larger KinectFusion mesh where there are no corresponding points in the single Kinect frame return larger Hausdorff distances as they search for a nearest neighbour. Meshlab also calculates the standard deviation of the Hausdorff distance and these are reported in 6.3. Frames 1 and 4 show similar results to when using the KinectFusion mesh as the second mesh. As can be seen in Figure 6.7 these frames correspond to when the phantom was face on to the Kinect covering a large area without many holes. Figure 6.7 graphically displays the results of the 6 single frames measured against the KinectFusion point cloud which is displayed as white points. The colourmap shows the Hausdorff on each point in the first mesh compared to KinectFusion.

6.3.3 Realtime Registration using the Kinect

Unless stated otherwise, the figures in this section show the position and orientation of the voxel at the centre of the phantom after being transformed by the motion tracking
Figure 6.7: The six single frames of Kinect depth data were obtained from a wide range of angles. This results in holes in the point cloud where structures such as the nose created occlusions, for example frame 3. The single frame point clouds are coloured depending on the distance to their nearest neighbour in the KinectFusion template (white point cloud). A histogram, with range 0-5.0 mm has been plotted showing the distribution of the distances for each registered frame.
6.3. RESULTS

Table 6.1: Reported ICP residual error after coregistration between 6 single Kinect frame meshes and a KinectFusion generated mesh of the head phantom

<table>
<thead>
<tr>
<th>Frame</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>KF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>0.8mm</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>0.8mm</td>
<td>1.0mm</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>0.8mm</td>
<td>1.1mm</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>1.2mm</td>
<td>0.9mm</td>
<td>1.3mm</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.8mm</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>1.2mm</td>
<td>0.9mm</td>
<td>1.0mm</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.9mm</td>
<td>0.8mm</td>
<td>0.8mm</td>
<td>0.9mm</td>
<td>0.9mm</td>
<td>1.0mm</td>
<td></td>
</tr>
<tr>
<td>KF</td>
<td>0.9mm</td>
<td>1.0mm</td>
<td>1.1mm</td>
<td>1.3mm</td>
<td>1.0mm</td>
<td>1.0mm</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2: Root Mean Square Hausdorff distance between frame pairs after ICP registration.

<table>
<thead>
<tr>
<th>Frame</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>KF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5mm</td>
<td>1.4mm</td>
<td>1.2mm</td>
<td>1.3mm</td>
<td>1.7mm</td>
<td>1.1mm</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.4mm</td>
<td>1.7mm</td>
<td>1.3mm</td>
<td>1.5mm</td>
<td>1.4mm</td>
<td>1.3mm</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.3mm</td>
<td>1.7mm</td>
<td>1.3mm</td>
<td>1.4mm</td>
<td>1.7mm</td>
<td>1.2mm</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.3mm</td>
<td>1.6mm</td>
<td>1.4mm</td>
<td>1.5mm</td>
<td>1.6mm</td>
<td>1.4mm</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.2mm</td>
<td>1.5mm</td>
<td>1.5mm</td>
<td>1.2mm</td>
<td>1.6mm</td>
<td>1.3mm</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.5mm</td>
<td>1.4mm</td>
<td>1.8mm</td>
<td>1.4mm</td>
<td>1.5mm</td>
<td>1.6mm</td>
<td></td>
</tr>
<tr>
<td>KF</td>
<td>1.4mm</td>
<td>1.6mm</td>
<td>1.5mm</td>
<td>1.5mm</td>
<td>1.6mm</td>
<td>1.7mm</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3: Standard Deviation of Hausdorff distance between frame pairs after ICP registration.

<table>
<thead>
<tr>
<th>Frame</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>KF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.04mm</td>
<td>0.97mm</td>
<td>0.75mm</td>
<td>0.90mm</td>
<td>1.08mm</td>
<td>0.73mm</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.93mm</td>
<td>1.12mm</td>
<td>0.87mm</td>
<td>1.00mm</td>
<td>0.93mm</td>
<td>0.83mm</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.88mm</td>
<td>1.14mm</td>
<td>0.88mm</td>
<td>0.94mm</td>
<td>1.10mm</td>
<td>0.79mm</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.84mm</td>
<td>1.04mm</td>
<td>0.97mm</td>
<td>0.96mm</td>
<td>1.05mm</td>
<td>0.86mm</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.81mm</td>
<td>1.00mm</td>
<td>0.99mm</td>
<td>0.74mm</td>
<td>1.05mm</td>
<td>0.81mm</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.99mm</td>
<td>0.96mm</td>
<td>1.15mm</td>
<td>0.87mm</td>
<td>1.01mm</td>
<td>0.97mm</td>
<td></td>
</tr>
<tr>
<td>KF</td>
<td>1.00mm</td>
<td>1.11mm</td>
<td>1.07mm</td>
<td>0.98mm</td>
<td>1.03mm</td>
<td>1.10mm</td>
<td></td>
</tr>
</tbody>
</table>
relative to its initial state. The motion tracking parameters follow the convention set out in Figure 6.3. The tables in this section show the range and standard deviation of the tracking measurements as these metrics indicate how stable the tracking is. The table columns show two time frames to give an indication of the stability over both long and short time frames.

The computational cost depends on the complexity of the scene and the resolution selected for the Kinect global volume. However to give an indication of the performance of the various GPUs used, the average registration computing time per frame for the same head phantom scanned was 20ms for the GTX 680m and 10 ms for the two GTX 580 GPUs. The other segments of the KinFu algorithm, such as drawing the model to screen and updating the model with new views, added on average another 20ms per frame. This results in the desktop GTX 580 GPUs being the preferred option as they can operate at the 30 Hz sample rate of the Kinect whereas to enable real time processing on the mobile GTX 680m the sample rate was limited to 20Hz.

Practicality

![Graph showing motion tracking](image)

Figure 6.8: Tracking drift and high levels of jitter are present in the motion parameters as measured from the Kinect when placed outside the PET/CT scanner gantry. Between 50 and 60 minutes, tracking in Tz is particularly poorly performed as the Tz parameter undergoes numerous and rapid jumps of over 1.5mm.

Figure 6.8 shows the motion tracking of the head phantom with the Kinect placed outside the PET/CT scanner gantry. The effect of the lack of 3D structures is apparent with the drift and noise in both the translation and rotations. Table 6.4 shows large ranges of up to 4.98 mm in Tz and 5.28° in Ry in the position and orientation which represents
Table 6.4: Range and Standard Deviation of Kinect Tracking Taken From Outside PET/CT Gantry

<table>
<thead>
<tr>
<th></th>
<th>Range</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Data</td>
<td>5 Seconds</td>
</tr>
<tr>
<td>Tx</td>
<td>4.79mm</td>
<td>1.64mm</td>
</tr>
<tr>
<td>Ty</td>
<td>2.80mm</td>
<td>0.27mm</td>
</tr>
<tr>
<td>Tz</td>
<td>4.98mm</td>
<td>0.50mm</td>
</tr>
<tr>
<td>Rx</td>
<td>2.52°</td>
<td>0.61°</td>
</tr>
<tr>
<td>Ry</td>
<td>5.82°</td>
<td>2.49°</td>
</tr>
<tr>
<td>Rz</td>
<td>5.28°</td>
<td>1.23°</td>
</tr>
</tbody>
</table>

that drifting has occurred. In Figure 6.9 a mirror has been used to reflect the speckle pattern onto the face. Drift is no longer present and the noise levels are also reduced compared to when the Kinect is placed outside the gantry as reflected in Table 6.5.

Figure 6.9: The motion tracking parameters show better stability when using a mirror to reflect the speckle pattern onto the face as there are sufficient 3D features for the tracking to converge to an accurate global transform.

**Spatial Accuracy**

Figure 6.10 shows the 2 minutes of simultaneous Kinect and Vicra tracking. The largest motion occurs in the Ty direction which corresponds to the rotations in the sagittal plane with the least amount of motion occurring in the Tx axis which corresponds to
Table 6.5: Range and Standard Deviation of Kinect Tracking Using A Mirror Inside a PET/CT Gantry

<table>
<thead>
<tr>
<th></th>
<th>Range Total Data</th>
<th>Range 5 Seconds</th>
<th>Standard Deviation Total Data</th>
<th>Standard Deviation 5 Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rx</td>
<td>1.33°</td>
<td>0.38°</td>
<td>0.17°</td>
<td>0.12°</td>
</tr>
<tr>
<td>Ry</td>
<td>0.89°</td>
<td>0.27°</td>
<td>0.13°</td>
<td>0.04°</td>
</tr>
<tr>
<td>Rz</td>
<td>1.51°</td>
<td>0.43°</td>
<td>0.17°</td>
<td>0.14°</td>
</tr>
</tbody>
</table>

Figure 6.10: Plots showing the Polaris and Kinect based motion tracking of a head phantom. The phantom is initially at rest with the motion being applied to the head phantom after 24 seconds. A range of movements were applied by hand which included tilting the head forwards and rotating it side to side. These types of movements commonly occur during brain imaging as the subject relaxes into the head rest.
Figure 6.11: (a) The least amount of motion of the head phantom occurred in Tx. The Polaris tracking shows higher levels of noise than the Kinect data and the Kinect data shows finer movements not visible in the Polaris data. (b) Good agreement exists between the two tracking methods in the Ty axis. Fast and small amplitude movements are visible in both data sets. (c) The Kinect based tracking shows good agreement with the Polaris tracking during periods of rest and of movements. Similar to Tx and Ty, Tz shows there is a slight spatial offset between the two tracking methods.
the rotations in the coronal plane. Figure 6.10 shows good agreement between the two tracking methods during times when the phantom was stationary and when motion was applied.

By zooming into each plane (Figure 6.11a - Figure 6.11c) it is possible to see how the two tracking methods compare to each other. The difference between the Kinect and the Polaris tracking should be zero if the Kinect tracking perfectly matches the Polaris measurements. The root mean square difference ± standard deviation for each plane is 0.41±0.57mm, 0.42±0.21mm, 0.36±0.27mm in Tx, Ty, Tz respectively.

Stability

The head phantom was tracked using the Kinect for over 12 hours under non-scanning conditions. Figure 6.12 shows a clinically relevant length subset of that data and shows no significant drift and low noise.

![Graph showing motion tracking data](image)

Figure 6.12: An 80 minute segment of the 12 hour motion tracking of a static head phantom is shown. This experiment was performed under ideal conditions where the head phantom was placed in the centre of the Kinect field of view. This allowed more of the surface of the phantom to be imaged and used in the registration which increased the stability of the tracking.

6.4 Discussion

Markerless motion tracking aims to detect the surface and track its rigid body motion. This differs from marker based tracking in that the surface topology and rigidity is very
Table 6.6: The range and standard deviation of the Kinect tracking of a static phantom over 12 hours and a clinically relevant time frame of 80 minutes.

<table>
<thead>
<tr>
<th></th>
<th>Range</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Data</td>
<td>80 Min</td>
</tr>
<tr>
<td>Tx</td>
<td>0.70mm</td>
<td>0.51mm</td>
</tr>
<tr>
<td>Ty</td>
<td>0.69mm</td>
<td>0.46mm</td>
</tr>
<tr>
<td>Tz</td>
<td>0.59mm</td>
<td>0.45mm</td>
</tr>
<tr>
<td>Rx</td>
<td>0.41°</td>
<td>0.34°</td>
</tr>
<tr>
<td>Ry</td>
<td>0.69°</td>
<td>0.52°</td>
</tr>
<tr>
<td>Rz</td>
<td>0.78°</td>
<td>0.64°</td>
</tr>
</tbody>
</table>

important to a markerless scheme. Whereas a marker is added to a scene and for every frame of data the position and orientation of the marker can be found, markerless motion tracking requires the generation of a surface or subset of surface points that can be temporally registered in sequential frames to track the rigid body motion of the points. The Kinect offers a low cost and consumer grade solution to measuring surfaces, however the limitations of the device cannot be overlooked. As shown in [Khoshelham et al. 2012] the error of depth measurement depends on distance to sensor. Additionally, in §3.2.5 it was discussed how the surface is only measured at each speckle point, resulting in the need to interpolate between points.

This chapter has aimed to validate the use of the Kinect and KinectFusion to measure head motion in a PET environment. Calibration is generally a prerequisite to validate a device for motion tracking, specifically an optical device, however calibration of the Kinect is confounded due to the degree of control available to a Kinect user over its functions. Attempts to calibrate the Kinect have focussed on calibrating the individual componants, i.e. the IR camera, the colour camera, the depth map etc [Khoshelham et al. 2012]. Patents filed by Primesense give an indication to the underlying technology behind the Kinect, however they cannot be used to precisely determine how the Kinect operates. Due to these limitations and the black-box nature of the device, it was important to attempt to constrain the situations where the Kinect is used. For example all tracking was performed in the near range of the sensor, between 0.8-1.5 m, to minimise the effect of varying depth accuracy with distance.

In [Newcombe et al. 2011] the authors demonstrate KinectFusion producing a metrically consistent reconstruction. This can seem in conflict with the previous discussion on varying depth accuracy, however it is the feature of KinectFusion to update a global model with new depth information which enables a form of self-calibration. As the kinect is moved through a scene, the TSDF volume is updated with the running average of the distance to the predicted surface with new Kinect frames. This improves the KinectFusion model by reducing the noise seen on individual Kinect frames. This can be seen in the paper square experiment in Figure 6.6.
For the head phantom, the difference in frame-to-frame and frame-to-model is visible in the mean and standard deviation of the difference in nearest neighbour points in the two point clouds, as measured as the Hausdorff distance. This result is unsurprising, as the KinectFusion model has less noise as it a smoothed collection of registered Kinect frames. An extension of this experiment would be to register the frames to a surface generated from a different depth scanner. By setting an isosurface threshold, a point cloud can be generated from a CT volume, as performed in the Spatial Accuracy experiment. This can result in errors as the isosurface may not represent the optically reflective surface of the phantom. Indeed the apparent drift seen in Figure 6.11a may be due to the co-registration between CT space and Kinect space being incorrectly estimated.

An ideal experiment would be to compare the surface generated from KinectFusion with the continuous type surface as measured by a Fourier transform filtering [Price et al. 2012] as described in §3.3.3. In contrast to a surface generated from a 3D CT where the CT isosurface chosen may not be the optical surface of the skin, the Fourier transform filtering system measures the surface of the skin, ready to be compared to the KinectFusion data. It would be informative to determine how many frames of Kinect data need to be integrated inside KinectFusion until a surface of comparable quality to the continuous surface of the Fourier transform filtering system can be produced. The need for additional viewpoints for KinectFusion would require the Kinect sensor to be moved relative to the patient which would decrease the effective temporal resolution.

Various phantom experiments were performed to validate the use of the Kinect and KinectFusion as a rigid body motion tracking tool for use in a PET/CT scanner. Before further work was performed using the Kinect it was decided to determine if the Kinect could actually be used in a scanning environment, with the limitations that the geometry of the PET/CT scanner enforces. For optimal Kinect positioning the sensor should be placed directly in front of the face at a distance of 0.5-1.0 metres. This obviously could not be implemented as the Kinect would have to be placed inside the PET detectors, or at least in the field of view. If the Kinect was placed outside the PET/CT gantry the view of the face did not have sufficient structure for stable registrations so a mirror was installed that reflected the structured light onto the subject’s face. To reduce risk of injury to the subject and to minimise the amount of attenuating material, an acrylic mirror was designed and rigidly fit in place using strong adhesive tape. Various other methods were considered such as placing the Kinect in the recess between the CT and PET sections of the scanner or by using lenses. However the minimum operating distance of the Kinect and the distortion that badly fitted lenses can cause [Alnowami et al. 2012] aided the conclusion that the mirror would be the simplest and best solution. The KinFu tracking of the head phantom using the mirror showed minimal drift as the mean range and the standard deviation of the Kinect translation parameters (Tx, Ty, and Tz) were 1.55, 1.15, and 1.49 mm respectively over 70 minutes.

The most encouraging result presented in this chapter is the Kinect ↔ Polaris head
phantom tracking comparison in Figures 6.10 and 6.11. This experiment most clearly shows that the Kinect has the potential to track a head sized object to a comparable accuracy to the Polaris, as the root mean square difference and standard deviation between the two tracking modalities is $0.41 \pm 0.57$ mm, $0.42 \pm 0.21$ mm, and $0.36 \pm 0.27$ mm in $Tx$, $Ty$, and $Tz$ respectively. This sub-mm similarity between the Kinect and the Polaris strengthens the case for further clinical use of the Kinect as a rigid body motion tracking system. It was assumed that the Polaris system would provide more superior tracking, but as can be seen in Figure 6.11a the Polaris tracking has higher levels of noise than the Kinect. As an improvement to this experiment would use a robot arm that could move the head phantom while being tracked by the Kinect, similar to the work done with a taxidermied rat head in [Kyme et al. 2011b]. Figure 6.11a to 6.11c also indicate that the tool slips at some point during the movements despite efforts to rigidly fix it to the head phantom. This can be seen at the stationary periods before and after the applied movements. This uncertainty spatial accuracy should not be present when the ground truth motion is known such as when using a robot arm. This method is preferred to the analysis performed in [Olesen et al. 2012] which compared the corrected and uncorrected PET images after MOLAR reconstruction (see [Carson et al. 2003]) as the quality of motion tracking has not been evaluated, only that the PET images have been improved using MOLAR motion correction. However neither of these analysis techniques can be used on real clinical data where the both the true tracer distribution and the original motion are unknown.

The other experiments performed to evaluate the sensitivity and robustness of the Kinect to events such as non rigid facial deformation or partial occlusions showed that the occluding large portions of the tracking surface did not generally affect the quality of tracking. This is due to the large number of points still available for ICP. This is unlike the Polaris system which has reduced accuracy if a single sphere is occluded and would fail to track the tool if more than 2 spheres were blocked. Of course, in the clinical environment the subject should be reminded to not touch their head and try to remain as expressionless as possible, yet for certain studies (for example, pain or dementia) such events may still occur involuntarily, and the motion tracking system should be unaffected by these.

Temporal alignment remains a significant challenge to solve to allow for KinectFusion to be used routinely in event based motion correction schemes. Future work should investigate the use of the host PC’s digital input/output ports to trigger a gating tag into the PET listmode gantry corresponding to when the tracking application measures a set amount of elapsed time.

### 6.5 Conclusions

A GPU based ICP registration algorithm, KinectFusion, has been used in real time rates on Kinect depth data at 30 frames per second to track the bulk motion of a rigid
head phantom. Experiments to validate the practicality, accuracy, robustness and stability have been performed and these results indicate that the Kinect can be used in a motion correction scheme for brain PET/CT. The millimetre-order spatial accuracy of the bulk tracking is consistent with the depth resolution of the Kinect sensor at a distance of $\sim 80$ cm. As although the depth resolution on a single depth measurement is $\sim 2$ mm, the large number of rigidly moving points sampled on the head phantom allow for the effective determination of the rigid displacement of the head phantom. Further optimisation is required for the method to be employed in routine clinical use, where non-rigidly moving facial regions are expected to occur. The first motion tracking performed on clinical PET subjects is shown in Chapter 7.
ANALYSIS OF MOTION DATA COLLECTED BY THE KINECT USING KINECTFUSION DURING CLINICAL PET SCANNING

Abstract

In the previous chapter, the investigations into real time bulk displacement measurement with the Kinect were performed using KinFu, an opensource implementation of the KinectFusion algorithm. This was due to the official Microsoft version not yet being included in the Kinect for Windows software developer kit, SDK. KinectFusion was included in the 1.7 version of the SDK, released 18 March 2013, and effort was made into porting across the modifications that were made to KinFu. It was anticipated that KinectFusion would offer greater code stability than KinFu, an essential feature for the motion tracking application to be used in routine clinical operation. The changes and improvements facilitated by using the SDK are described in the following section. Motion tracking from three subjects show how adaptations to the code and acquisition protocol improve motion tracking data quality. These data sets prompt the discussion into how to optimally integrate the Kinect as a head motion tracking device in PET/CT.

7.1 Introduction

Markerless motion tracking aims to provide accurate and reliable monitoring of the subject movement. Current markerless motion tracking systems rely on detecting feature points in stereo video sequences [Kyme et al. 2012] or by producing a series of point clouds (arrays of 3D points that describe surfaces) from structured light [Olesen et al. 2012] or time of flight depth sensors [Placht et al. 2012]. These point clouds can be registered with an offline iterative closest point algorithm (ICP) to track motion. ICP is computationally expensive for large point clouds and is prone to fail without good initialisation. ICP can be processed in real-time if the computations are performed on a graphics processing unit, GPU, as the structure of the GPU is highly parallelised over
multiple processing cores and is well suited for such tasks. In [Newcombe et al. 2011] and [Izadi et al. 2011] Microsoft Research Cambridge announced a real-time and GPU based ICP implementation using the Kinect as an input device. They demonstrate the ability of KinectFusion to estimate the pose of a moving Kinect sensor in real-time if moved within a detailed yet static scene. Conversely, with a static Kinect sensor, KinectFusion can calculate the pose of a well detailed rigidly moving object. However, for in-vivo studies where body tissue or facial regions are likely to undergo non-rigid deformations, the performance of KinectFusion in tracking the bulk rigid body motion is likely to worsen.

A detailed description of the investigations into validating the Kinect tracking with rigidly moving head phantoms can be found in Chapter 6. In reference to the criteria set out in §3.2.5 investigations were performed to evaluate the use of the Kinect as a depth sensor in a motion tracking system. To enable the Kinect to be used inside the PET/CT gantry, an acrylic mirror had to be installed at the top of the scanner bore to reflect the Kinect structured light onto the face. The mirror allows for the Kinect to view the surface of the face almost straight to improve registration stability. This confined the structure of the face within a narrow depth range which reduced the effect of varying depth resolution with distance to sensor. The bulk, rigid-body type motion of the head phantom was measured using the Kinect and KinectFusion and showed similar spatial accuracy to that of the Polaris Vicra, a marker based motion tracking system, demonstrating that sufficiently good spatial accuracy tracking could be achieved with the Kinect if the object being tracked did not have non-rigidly moving regions. The time constraints of the PhD prevented the design and construction of an anthropomorphic face phantom, which would allow for the motion of the head to be known and to have a non-rigidly moving facial surface.

This chapter, however, expands the validation work performed in the previous chapter by designing a set up to allow the Kinect to be used for clinical PET scans, where the subject’s would indeed have facial regions that deform non-rigidly. To change from a proof of principle design to a clinically feasible device, modifications to the motion tracking software and the mirror attachment device had to be made. These modifications were designed to improve stability and to allow for the hardware to be safe to use while scanning subjects and easy to remove from the PET/CT scanner when not required. Instead of working with the opensource Kinect drivers and opensource KinFu registration algorithm, the official Microsoft software developer kit (SDK) version 1.7 and the official KinectFusion implementation as described in [Newcombe et al. 2011] and [Izadi et al. 2011] were adopted. This chapter presents tracking data from the first clinical subjects using KinectFusion and describes the improvements to the tracking code and hardware which enabled their collection.
7.2 Methods

7.2.1 Software Advancements

The SDK contains many specific features that the opensource drivers do not have, such as

1. Control of the Kinect motor
2. Cross platform GPU support (ATI and NVIDIA)
3. Access to the raw video data (RGB, depth, and IR)
4. Access to the Kinect audio capture

There is currently no stable solution for using the opensource, PCL code with the Kinect SDK drivers. The SDK 1.7 release included libraries to perform KinectFusion and included an example KinectFusion application. This application was modified to allow the Kinect motor to be used and to record at higher resolutions and larger volumes. Certain parameters of the ICP algorithm appear to be hard-coded into the hidden KinectFusion function calls as opposed to KinFu where all parameters were adjustable. Specifically the distance that ICP was allowed to search for a nearby point was restricted so that KinFu was less sensitive to regions that deform non-rigidly. In Chapter 6 this parameter was set to be 5 mm, hence if a point was beyond this distance to the global model it was discarded as an outlier. At 30 Hz data acquisition rate it was assumed that any point that is more than 5 mm away from the global model would belong to a non-rigidly deforming region. The option to set this distance is not included in KinectFusion however the documentation claims that a robust outlier detection routine is used to filter outlier points. The KinectFusion application program interface (API) also allows for the integration of new Kinect frames to be paused thereby preventing tracking drift from occurring, which is potentially particularly useful for in clinical PET as some faces may have less 3D structure and be more prone to drifting affects.

7.2.2 Hardware Changes

The Kinect for Windows, a new model of the Kinect sensor was used for the following experiments. The main difference between the Kinect for Windows and the Kinect for Xbox360 is the ‘near mode’ of the Kinect for Windows. The near mode allows for objects closer than 1 metre to be imaged with better sensitivity, as the dynamic range of the IR sensor is changed to compensate for the fact that objects closer than 1 metre are brighter and the points can fail to be resolved with the Kinect for Xbox360. The near-mode option can be emulated using a partial IR filter to reduce the amount of IR light that the sensor receives. A piece of non-exposed yet developed camera film provides a good solution for this. The advantage of filtering the IR video, either by using a physical filter or by
CHAPTER 7. CLINICAL MOTION TRACKING USING KINECT

Figure 7.1: This application is written in C# and shows some of the features that have been streamlined from the opensource implementations. These include dynamically altering the depth threshold and resolution of the target scene, features that used to be hard coded and required recompilation to alter. The top right window shows the raw Kinect depth video in grey scale. The bottom right window shows the delta depth map, i.e. the distance that the each pixel in the current Kinect frame is from the global model. Distance is denoted as blue is below the surface and red is above the surface and white is where the current frame matches the surface of the global model. The global model is shown in false colour on the left where the false colour describes the position of each voxel inside the global volume.

using the Near-Mode is that the system is less sensitive to background illumination and the reflected and scattered IR light off the inner surface of the PET/CT gantry. This is essential as the subjects are always within 0.8 m of the Kinect. If the subjects are further away from the Kinect then the apparent size of their head as viewed in the Kinect depth data is too small for accurate ICP registration.

A significant challenge was posed when considering how to best fix the mirror to the PET gantry. An adhesive solution was trialled and performed well for the phantom studies shown in the previous chapter. However, as part of the design brief to use the Kinect in clinical PET studies it was important to be able to remove the mirror from the gantry so to not interfere with scans which did not require the Kinect. An adhesive solution proved incapable of providing the rigidity and ease of removal required, and so two modes of fixation was designed. Firstly, a technique using anchor points provided by the scanner gantry, and secondly using a removable ring of 5 mm thick PVC. The first technique utilised features present in the HiRez PET/CT in Imanova which consisted of a screw used for fixing the gantry cover to the frame of the PET segment and a grill found in the recess between the PET and CT segments. The TrueV PET/CT scanner lacked the screw point and so the PVC ring method was devised. These are illustrated in the diagrams in Figure 7.2.

The mirror was mounted on a hinged platform that allows for the mirrors tilt to be
7.2. METHODS

Figure 7.2: (a) The acrylic mirror is attached to the gantry using features on the gantry itself. A metallic screw and a recessed grill provide anchor points to hang an articulated hinged arm which enables the mirror angle to be modified. Positioning the mirror allows for the entire structure to be outside the PET FOV thereby not causing attenuation. Scattered photons may be scattered back into the PET detectors however these should be discarded due to their lower energy. (b) For scanners without easily accessible features such as the screw point, a 5 mm thick sheet of PVC was bent into a ring and fitted inside the gantry. The tension on the PVC holds the ring securely in place and the mirror and hinge mechanism is hung from the top side of the ring.

adjusted and fixed in place for the scan. This allows for better control during set up to take into account varying sizes of heads and their anticipated position during PET scanning.

7.2.3 Scanning Protocol

Three subjects undertook 90 minute PET scan on the HiRez PET/CT scanner at Imanova. The subjects were from a $^{11}$C-PBR28 study ran by Dr. Eugenii Rabiner. As with KinFu in Chapter 6 the tracking parameters returned from KinectFusion describe the motion of the Kinect relative to a static scene. To convert this data into PET/CT space a mesh of the subject’s face was obtained after the first few seconds of KinectFusion tracking. After the mesh has been saved the integration of new Kinect frames into the global model was paused to prevent tracking drift. Due to time constraints, the first subject was not scanned for a sufficient period of time before the mesh was saved and integration was paused. The second subject was scanned for longer and the third for longer still. The effect of varying the quality of the global model template is shown and discussed, later in this chapter. Ideally, a good, smooth, and hole-free global model should be used however since the subject’s head is only visible to the Kinect just prior to injection, there was insufficient time to adequately scan in to the template before injection. After the scan, the centre of mass of the template was used as a test point to describe the motion of the head and this data is shown in the following section.
7.3 Results

Figure 7.3 shows the meshes from the three subjects that were used as the template for KinectFusion. The quality of the meshes vary as the procedure for obtaining them varied as the motion tracking protocol was adapted. Figure 7.3a was obtained with insufficient time to produce a smooth template before the mesh was saved and the integration was paused. Effort was made with the scans shown in Figure 7.3b and more so in Figure 7.3c to ensure that the template face has higher spatial resolution and has been smoothed from multiple views. The objects that appear in the meshes at the side of the face are the foam padding used in the head rest. While most non-skull objects were filtered from the Kinect it is apparent that this was not entirely successful for some subjects. The effect on motion tracking that these additional objects would have is unknown and ideally the only object in the field of view would be the face. For these tracking scans it is likely that they would confound the ICP algorithm and be classed as registration outliers, as they are substantially smaller than the main surface of the face.

![Figure 7.3: The first three clinical PET/CT subjects that were tracked using the Kinect and KinectFusion. These meshes represent the template that was used for KinectFusion. The quality of the meshes increases from (a) to (b) and again to (c). This is due to the amount of time spent at the start of the scan moving the Kinect relative to the head to provide new views to KinectFusion to include into the template. The additional structures visible are part of the headrest that was not fully filtered out of the active tracking volume.](image)

The centre of mass of the meshes was calculated and the KinectFusion measured transformations were applied to that voxel location. Figure 7.4 show the motion of this point during the 90 minuets PET scan time. The narrow peaks in the tracking data that can be seen most in Figure 7.4a and less so in Figure 7.4b are likely due to ICP failing to accurately register the current frame to the template. In these three scans the rate of these sharp peaks appears to decrease as the quality of the template increases, however more scans are required to evaluate this theory.

Figure 7.5 shows the first 1.8 minutes of Figure 7.4c. Due to the time constraints it was not possible to perform the Kinect sensor sweeping prior to tracer inject and the start of the PET scan. As a consequence the first minute of tracking data has been tarnished by the motion of the Kinect sensor relative to the face. Also from this figure it is clear that the mesh saving routine temporarily suspends the KinectFusion algorithm. This is due to
Figure 7.4: The centre of mass of the meshes in Figure 7.3 was transformed by the KinectFusion tracking parameters for each frame of tracking data. All three scans presented some motion, predominately slow drift. In (a) sharp peaks can be seen in the tracking. This is likely due to the ICP algorithm temporarily failing due to the relative low quality of the template. Fewer such peaks are seen in (b) and almost no sharp peaks are visible in (c) which had the best template.

the need to copy the point cloud array from the GPU memory into CPU memory. For this scan a laptop was used to remotely log into the KinectFusion acquisition PC which again caused the tracking to be temporarily paused.

It is very likely that the higher levels of noise in these clinical data sets compared to the static phantom studies is a result of subject motion, such as breathing. Figure 7.6 shows a region between 48 and 50 minutes where the motion of the head due to breathing can be seen.

### 7.4 Discussion

The experiments presented above demonstrate the feasibility of implementing the Kinect in the real clinical setting. Within this context it is clear from the data that continued development of the Kinect motion tracking system is required. The critical factor limiting the success of the motion tracking system is the integration of the device within the PET
Figure 7.5: Due to time constraints of the PET scan, the effort to make a better template resulted in having to move the Kinect after the PET scan had begun. The motion of the Kinect relative to the head causes apparent motion that can be seen in the tracking data. Also visible in this figure is the effect of saving the mesh to disk. Tracking is temporarily suspended as the point cloud data is copied from the GPU memory. An additional break in the tracking data is caused by remotely logging onto the KinectFusion acquisition PC.

Figure 7.6: A breathing rate of 16-17 cycles per minute can be observed in the Ty motion of the head as measured by the Kinect.

scanner environment. The Kinect was never intended to scan small objects, instead its anticipated application was for use in volumes up to 3 m³. However under ideal conditions the bulk motion monitoring of a rigid head phantom is comparable to the motion data obtained using the Polaris position sensor, and at 30 times per second in real-time. Currently no other markerless motion tracking system has demonstrated this in PET. Of the other markerless motion tracking systems that have been used in clinical or preclinical PET, such as [Olesen et al. 2012] or [Kyme et al. 2012], non have been able to produce real-time at 30 Hz motion tracking, especially at the low cost of the system used in this chapter.

Even in the confines of a PET/CT scanner where a mirror had to be used to reflect the Kinect structured light onto the face, motion tracking was successfully performed on phantom studies. In these studies the position of the Kinect and mirror were carefully optimised so that the head phantom could be accurately tracked. In a clinical environment,
subject's have a range of body shapes and sizes. As a result of this, the initial position and orientation of the head varies between scans and consequently the motion tracking system needs to allow for this.

Markerless bulk motion monitoring of the head using the Kinect was performed on three subjects in real-time using the KinectFusion algorithm. To enable this, it was essential to ensure that any modification to the scanner environment, such as installation of the mirror, would be achieved without interference to the subject or the PET procedure. A removable solution was developed that also allowed for the angle of the mirror to be adjusted using a hinge joint. Once set in place, the mirror remained fixed and allows for the Kinect structure light to be better projected onto the face. Currently, once the mirror is fixed and the subject is in the scanner, the only modification that can be made is the position of the Kinect sensor. This can be performed by hand, but is less than ideal for the scanner operators. For future studies, a recommended modification of fixing an extended arm to the mirror would allow the mirror angle to be adjusted to suit individual scans. However in the near future, it may be possible to forego the need of a mirror altogether with the use of short range depth sensors. These sensors are discussed more in §10.1.1. Hence, although the current use of mirror and manual adjustment of the Kinect has its limitations and shortfalls, future depth sensors will be more suitable for the confine of a PET/CT system.

The displacement plots shown in Figures 7.3a, 7.3b, and 7.3c appear to have higher levels of noise than the phantom experiments in the previous chapter. There are three reasons why noise levels are higher. Firstly the additional structures that can be seen in the meshes that were not rigidly fixed to the head would confuse the ICP algorithm and potentially increase noise as ICP searches for the optimal transformation. Secondly, since the mirror has been placed inside the PET/CT gantry rather than at the edge as in the phantom experiments the distance to the face has increased from 50 cm to 80 cm. This results in a smaller mesh and less points for ICP to operate upon, also resulting in higher uncertainty in registration. Finally, the apparent noise may contain actual small motions of the subject’s head, such as the effect of breathing. This can be seen in Figure 7.6.

To address these particular points and to improve overall tracking quality a series of modification are recommended for future iterations of the tracking software system set up. A filter should be implemented that uses a face detection algorithm on the RGB colour or IR video data so that only the regions of the depth map that lie within the boundary of the face are used in the KinectFusion algorithm. The current filter simply cuts off at a arbitrary X and Y level and the Z distance is scalable. A face detection filter would provide a much more robust and user friendly method for segmenting the surface to be tracked from the head rest. An additional filter could be implemented that splits up the global volume into a foreground and a background volume which can be independently tracked. This allows for the background to be used as a static frame of reference in which the foreground object (the head) moves rigidly. Therefore the Kinect can be moved to
obtain new views of the face to build up a good template without the foreground object appearing to move.

7.5 Conclusion

This Chapter presents the first three kinect motion tracking data sets from clinical PET/CT scans. The challenge for bringing a proof of concept prototype which was used in the previous chapter for the phantom validation studies was considerable. Even though the motion monitoring system is still being optimised, markerless bulk motion monitoring of clinical PET brain scans was performed in real-time using inexpensive, consumer grade hardware.
CHAPTER EIGHT

THE ASSESSMENT OF A LOW COST, MULTIPLE TARGET MARKERS TECHNIQUE FOR TRACKING HEAD MOTION IN THE HRRT

Abstract

The Kinect motion tracking tool is not an appropriate system for the HRRT. It is not possible to position the Kinect in from of the subject as the bore of the HRRT scanner is too narrow. While a mirror system can be used to guide the Kinect IR structured light onto the subject’s face in PET/CT systems, the geometry of the HRRT prevents such a method. As shown in §5.3 motion tracking on the HRRT is insufficiently addressed by the Polaris position sensor. Since a replacement motion tracking system is needed for the HRRT, an alternative, advanced, marker based tracking system was investigated. Of the systems studied, the monocular target marker tracking presented in §3.2.4 offered the most attractive solution due to its low cost (computational, physical, and financial) yet high accuracy tracking. The aim of this chapter was to investigate and evaluate the use of a multiple target marker based motion tracking system for use in brain PET.

8.1 Introduction

The geometry of the HRRT presents a significant challenge to install the Kinect as a motion tracking device. Figure 8.1 shows a profile cross section of the HRRT and illustrates the difficulties in positioning the Kinect for optimal face tracking. It is clear even attempts to implement a mirrored reflecting system would not work in this situation. The minimum operating distance of the Kinect prevents the sensor from being positioned behind the head as used for the PET/CT technique. Furthermore the front edge of the bore prevents the upper regions of the face being imaged when the Kinect sensor is placed in front of the scanner as the nose and the jaw dominate the field of view.

The relative merits and weaknesses of the Polaris position sensor have been addressed previously in §5.3, where results are presented that show the effects of the Polaris track-
The HRRT is a brain and small animal PET only scanner. This means that the bore size is significantly smaller than the whole body PET/CT scanner. The narrow bore prevents the Kinect from being able to view the head from the optimal face-on position, even if a mirror could be used. A webcam was installed in the rear section of the gantry outside of the PET field of view to enable motion tracking using target markers.

In preclinical PET, a stereo-optical motion tracker has been used [Kyme et al. 2011a] which tracks target markers consisting of flat boards (20 × 20 to 60 × 60 mm²) of high-contrast visible patterns. This system has a minimum operating distance of 200 mm and therefore complications may arise when viewing the markers inside the narrow bore of the High Resolution Research Tomograph (HRRT).

Monocular target marker tracking, as described in §3.2 is a marker based motion
tracking technique that can track a printed target marker using a single digital camera in real-time. This chapter aims to investigate the use of the multiple target marker tracking (MTMT) on small ($5 \times 5 \text{ mm}^2$) printed target markers by placing the digital camera inside the scanner bore of the HRRT, as can be seen in Figure 8.1. In [Qin et al. 2009] the authors placed digital cameras inside an MR scanner bore to enable viewing of the patient. The system reported high positional accuracy of the tracking marker ($< 1 \text{ mm std in x,y,z}$) however their system used a stereo camera pair and only used a single marker which may be prone to slippage. The following section follows on from this idea but using the target markers as described in §3.2.4.

8.2 Methods

8.2.1 Software

The Perspective-$n$-Point (PnP) problem can be used to estimate the pose of a flat marker of known size using a single camera [Lowe 1991]. Solutions to PnP use point correspondences between the 3D points of the marker corners and their projections onto the image plane of a calibrated camera. For details on the camera model, see §3.2.3. In the case of a square marker, $n = 4$ and transformation of the marker in 3D can be estimated using an iterative cost function. The implementation used in these studies is provided in the Opensource Computer Vision library, OpenCV, and performs using the following behaviour.

1. A flat and square marker with a unique pattern is imaged by a digital camera.

2. The corner location of the marker in its own coordinate system are known by recognising the identity of the marker and by using a priori knowledge of the marker geometry. These are the objectPoints, $x$.

3. In the image of a calibrated camera, the corners of the marker are detected and are labelled as the imagePoints, $y$.

4. The OpenCV function solvePnP estimates the transformation (rotation, $R$, and translation, $t$) of a marker given its objectPoints and the corresponding imagePoints and the camera matrix, $C$, of the calibrated camera, using the a cost function Equation 8.1.

5. The function iterates to minimise the reprojection error, $e$, between the image plane mapping of the transformed objectPoints and the observed imagePoints.

$$e^2(R, t) = \frac{1}{n} \sum_{i=1}^{n} ||y_i - C(Rx_i + t)||^2$$ (8.1)
Many augmented reality applications have used solutions to the PnP problem to monitor the position and orientation of a marker. In this work, the ALVAR software library, from the VTT Technical Research Centre of Finland, and released under a GNU Lesser General Public License, was used which wraps OpenCV and other computer vision functions.

### 8.2.2 Hardware

For these experiments a Logitech c910 digital camera was used. Despite being an off-the-shelf component, the camera can operate at a resolution of $1600 \times 1200$ at 30 Hz and at a variable focal length. A resolution test sheet was printed and imaged with the camera. The pdf test sheet image was obtained from [Westin 2013] and was based off the ISO12233 chart. An image at $1600 \times 1200$ was obtained of the test sheet to investigate the resolution of the Logitech camera at distance similar to those used for the motion tracking.

It was desirable to focus the camera at a shorter distance so the markers appear larger in the field of view. The focal length of the camera was fixed and camera calibration was performed using the Camera Calibration Toolbox for Matlab. The calibration procedure involved taking 15 images at different positions of a $9 \times 15$ checkerboard pattern, with square dimensions $6.2 \times 6.2$ mm.

The algorithm is CPU based and runs at 30 Hz on an Intel i7 3820QM, a laptop CPU at 2.70 GHz. The greater the camera resolution, the more intensive the processing therefore a compromise was made to have the highest resolution possible but to remain at 30 Hz processing.

### 8.2.3 Marker Tracking Accuracy

Camera calibration is an essential step as it measures the lens distortion and other factors that affect the projection of a 3D point onto the 2D image plane of the camera. As the Logitech Camera has a variable focus mode, it is crucial to fix the focal length after calibration as the calibration only remains valid for the focal length used at the time of calibration.

To validate the success of the camera calibration and the spatial accuracy of the target marker tracking an experiment was performed to measure the distance between four printed target markers on a flat surface. The four $10 \times 10$ mm$^2$ markers were arranged in a rectangle with dimensions $85 \times 54$ mm and with a diagonal separation of 100.7 mm. A program was written using Microsoft XNA framework and ALVAR to simultaneously track the four markers and write each marker’s world-to-marker transformation to file. A screen grab of the application tracking the 4 markers can be seen in Figure 8.2. For each frame of tracking the host PC’s local time was grabbed using the c# function DateTime.Now rather than the XNA specific GameTime.totalGameTime as GameTime uses
METHODS

Figure 8.2: The tracking application shows an augmented reality view of the markers with cubes transposed on the markers and a virtual reality viewport showing the cubes position relative to the camera (yellow pyramid) and original marker position (grey cube).

a method of tick counting to calculate time which is sensitive to the application small variations in frame rate. During the 70 seconds tracking, the camera was moved relative to the marker array; this is analogous to the markers moving with a stationary camera, however it was essential that the array remain flat and undistorted during this experiment. The purpose of the experiment was to measure the separation of the markers as measured by the camera at different positions. If the separation of the markers is independent of the relative camera position, then the camera calibration is accurate and the tracking system can accurately estimate the position of markers.

8.2.4 Marker to PET Calibration

To enable motion correction, the motion tracking parameters need to be transformed into PET space. To enable this, the transformation between Marker space and PET space needs to be determined. A single $10 \times 10$ mm$^2$ marker was prepared by positioning a small drop of $^{18}$F in the centre of the marker to create a point source of $\sim$3 MBq activity that corresponds to the origin of the target marker. The excess was removed using the corner of an absorbent cloth to prevent the activity running or spreading out by absorption into the paper.

The point source marker was placed inside the field of view of the HRRT and was
Figure 8.3: (a) The point source marker was prepared using a syringe and needle to carefully position a small drop of $^{18}$F in the centre of the marker. An absorbent cloth was used to remove the excess. (b) The marker was taped to a 4 mm thick Perspex sheet, to provide an annihilation medium. The front surface was not covered as to not distort the marker’s appearance through refraction.

imaged at 12 different, non-coplanar positions. Tracking was performed to measure the position of the marker, and the resultant PET listmode data was split so that each of the 12 frames only contain a single point source. The point source position was calculated by summing the planes of the image to get a profile view in each axis, where the peak can be seen and measured. The first four point correspondences were used to calculate the transformation between marker and PET space using absolute orientation [Horn 1987]. The following 6 pairs of points were used to validate the transformation by transforming the measured marker position into PET space and calculating the root mean square difference to the measure point source voxel location.

8.2.5 Stability Assessment

Drift and Noise

In this experiment three static target markers were tracked on a static phantom for 40 minutes to determine the stability of the system by measuring drift occurrence. Drift is more likely to occur if the camera mounting is unstable, if auto-focus attempts to alter the raw video feed, or if the marker is not securely fixed to the skin of the subject. Monitoring the position of a target marker for an extended period of time allows for the rigidity of the camera mounting to be analysed. If the adhesive is not able to securely fix the camera to the scanner bore, then the target marker would appear to drift. This experiment also aimed to observe the tracking noise level. This noise is due to small errors in determining the location of the corners of the target marker in the image. This then leads to uncertainty in the iterative process used to calculate the position and orientation of the target marker. To increase signal to noise, the target marker needs to be imaged over more pixels. This can either be achieved by using a larger target marker, or more preferably, by using a higher
resolution camera. In the latter case, the camera can be moved closer to the target area (or an optical zoom could be utilised to enlarge the target area) but this will limit the active target tracking area which limits the amount of motion that the system can track before the marker moves out of view.

**Lighting**

As this tracking system operates in visible light, it is important to investigate how a change in ambient light levels affects tracking. A single $5 \times 5 \text{ mm}^2$ marker was tracked under normal office lighting. During tracking a single LED torch was shone on the marker, before being turned off, to return the system to normal lighting conditions. This experiment was performed first with the camera’s automatic gain, contrast, and brightness control turned on and was repeated with these features disabled.

**Marker Curvature**

![Figure 8.4: A flat square is imaged in a calibrated camera image plane as a square. If rotated relative to the camera, a flat marker appears as a trapezoid. The corner points of a curved marker describe a rectangle, which is not expected by the PnP algorithm.](image-url)

If a marker is bent or distorted it no longer represents a true square and the **object-points** do not correspond to the positions of the marker’s corners. To investigate this effect, a $20 \times 20 \text{ mm}^2$, a $10 \times 10 \text{ mm}^2$, and a $5 \times 5 \text{ mm}^2$ marker were taped to the surface of a cylinder of radius 41.7 mm. Using trigonometry, this results in the horizontal opposite corners of the $20 \times 20 \text{ mm}^2$ marker being separated by 19.75 mm with the centre of the marker being 1.2 mm raised compared to the edge. The mean and standard deviation of the distance to the markers was $137 \pm 15$ mm. Each marker was tracked and compared to identical markers fixed to a flat surface. In the image plane of the calibrated camera, a square marker appears to have curved edges that are shorter than the vertical straight edges. These effects confound the marker detection algorithm as firstly the square marker may be disregarded as not have flat edges, and secondly if the square is detected the shorter horizontal edges would suggest to the PnP algorithm that the marker is at an angle relative to the camera, however the pair of vertical would not agree with this estimation of the pose of the marker. Figure 8.4 illustrates how the perspective view of the
camera images a flat square as a trapezoid if the marker is rotated about its vertical axis. A curved flat marker would show shortened horizontal edges as the rotated flat marker, yet it would also show equal length vertical edges.

Marker Size

As a continuation of the previous experiment into the effect marker curvature has on tracking stability, the marker’s size and its effect on tracking stability was investigated. To do this 4 markers of sizes $20 \times 20$, $10 \times 10$, $5 \times 5$, and $3 \times 3 \, \text{mm}^2$ were fixed to a horizontal flat surface with the camera positioned above and perpendicular to the surface. The camera was translated across the array of markers so that each marker was in the centre of the image while being tracked. This was to ensure a fair test if the calibration camera matrix has uncertainties at different points in the image. For this experiment, the distance from marker to camera was more tightly controlled, than the investigation into marker curvature. The mean distance to the markers was 108.95 mm with a standard deviation of 0.45 mm. At this distance the $1600 \times 1200$ resolution equates to approximately 13 pixels per mm. In descending order the 4 markers occupy 67.6k, 16.9k, 4.2k, and 1.5k pixels out of a total of 1.92M pixels. The assumed camera resolution of 0.25 mm as determined in Figure 8.6 can be represented as a fraction of the marker lengths; in descending order these are $1/80$, $1/40$, $1/20$, and $1/12$.

8.2.6 Volunteer Tracking

To determine the systems robustness to non-rigid facial deformations, a volunteer was tracked using three $5 \times 5 \, \text{mm}^2$ markers attached to the skin below the hairline. The markers were printed onto adhesive paper that were fixed to microporous tape that was in turn attached to the subject’s hairline and tracked for 15 minutes.

Following this, the volunteer was asked to talk, swallow, and frown during tracking. By measuring the relative separation of the markers, a confidence metric can be constructed that the markers are reliably tracking the rigid body motion of the brain.

8.3 Results

The resolution test image indicates that the resolution of the camera at the distances being used in later studies is roughly 0.25 mm. This is based upon the line profile dropping below 75% of its maximum after pixel 650. This corresponds to the 600 lines per 145 mm, or 0.24 mm mm per line. The experiment is not a conclusive proof of resolution, as the test image was not printed on a high resolution printer to professional standards, was not perfectly flat, nor had perfectly even illumination.

The Camera Calibration Toolbox for Matlab calculated the camera matrix with the following errors. The errors are reported by the toolbox as approximately three times the
8.3. RESULTS

Figure 8.5: (a) Target markers are positioned at the hairline over the temples and the centre of the face. Increasing the number of markers increases the tracking systems robustness to non-rigid deformation and tool slippage. (b) The markers are numbered and the Euclidean separation is labelled between each marker pair, d13, d12, and d23.

standard deviation.

Focal Length: [1370.8 1379.1] ± [5.6 5.7]
Principle point: [811.9 606.3] ± [3.8 2.6]
Distortion: [-8.3×10^{-3} -4.0×10^{-2} -1.9×10^{-4} -1.1×10^{-4} 0.0] ± [6.2×10^{-3} 2.3×10^{-2} 5.3×10^{-4} 8.6×10^{-4} 0.0000]
Pixel error: [0.5 0.5]

8.3.1 Marker Tracking Accuracy

Table 8.1: Mean and Standard Deviation of the Separation of the 4 Markers

<table>
<thead>
<tr>
<th></th>
<th>d12</th>
<th>d13</th>
<th>d14</th>
<th>d24</th>
<th>d23</th>
<th>d34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>85.1</td>
<td>54.2</td>
<td>100.9</td>
<td>54.2</td>
<td>101.0</td>
<td>85.1</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.1</td>
</tr>
</tbody>
</table>

After camera calibration the Marker Tracking Accuracy experiment was performed on the flat array of 4 markers. Figure 8.7 plots the separation of each marker pair minus the mean of that separation. For reference, the graph also shows the $x$ displacement of the 4 markers which indicates the amount of motion the camera undertook. Figure 8.7 also highlights the difference between tracking static and moving markers. After the 30 second mark, when the camera is in motion, there is a higher level of tracking noise and spikes appear. This may be due to the limitations of the consumer grade digital camera used in these investigations, as its image compression may introduce blurring for moving
objects and image tearing may occur if its framebuffer is being overwritten by a newly acquired image before the previous image frame is passed to the host PC. These issues could be resolved by a camera which is able to send uncompressed image data at a fixed frame rate to the host PC over a highspeed data connection (USB3, Gigabit Ethernet, etc). However, even with the consumer grade webcam, the mean separation of the markers is 0.2 mm from their true separation. Table 8.1 shows the mean and standard deviation of the separation of the markers from Figure 8.7. For reference, the marker separation was $85 \times 54$ mm with a diagonal separation of 100.7 mm.

Figure 8.8 shows the measurement of the error in the rotational accuracy of the marker tracking. Spikes in the tracking data, such as the observed point at 11 seconds, are due to the algorithm being unsure of the shape of the marker and incorrectly calculating its rotation. These events can happen when the marker is moving quickly and the frame tears,
8.3. RESULTS

Figure 8.7: The separation of the markers was calculated from the tracking of the individual markers as the camera was moved around the array. The mean separation of the markers is 0.2 mm from their true separation.

if the marker is out of focus, or if the marker is not flat.

Figure 8.8: The rotation of each marker from initial template marker space to observed marker 3D position should be the same, as they were arranged on a flat surface in identical orientations. The spike seen at 11 seconds was due to one marker being incorrectly positioned as it was close to the edge of the field of view of the camera.

The root mean square difference from the 4 markers was calculated and is shown in Table 8.2. It is crucial for the error in rotation to be small as the effect on distant voxels increases. For example, at a distance of 200 mm the 0.6° error in Marker 4 corresponds to an uncertainty of 2.1 mm.

8.3.2 Marker to PET Calibration

Figure 8.9 shows the x,y,z position of the target marker during PET data acquisition. The mean error on the calculation of the transformation between the first 4 paired points was 0.3 mm with a maximum error between transformed position and measured
Table 8.2: Rotational Root Mean Squared Difference from the 4 Markers

<table>
<thead>
<tr>
<th></th>
<th>Rx (deg)</th>
<th>Ry (deg)</th>
<th>Rz (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker 1</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Marker 2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Marker 3</td>
<td>0.3</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>Marker 4</td>
<td>0.4</td>
<td>0.6</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Figure 8.9: Tracking of the point source marker as measured by marker tracking. The maximum range in displacement in each axis is 122.3, 61.4, 85.3 mm for Tx, Ty, and Tz respectively. The large spikes shown in the tracking are due to the marker being manually moved between positions.

Table 8.3: Root Mean Squared Difference Between Transformed Target Marker Position and the Measured Point Source Voxel Location

<table>
<thead>
<tr>
<th>Frame</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD (mm)</td>
<td>0.1</td>
<td>0.8</td>
<td>0.9</td>
<td>0.7</td>
<td>0.8</td>
<td>1.2</td>
<td>0.5</td>
<td>0.7</td>
</tr>
</tbody>
</table>
point source position of 0.7 mm. Table 8.3 shows, for each frame, the root mean square difference between the measured marker position transformed into PET space using the calculated transformation and the measured point source position in PET space.

8.3.3 Stability Assessment

Drift

Figure 8.10: MTMT was performed on three static target markers for 40 minutes. (a), (b), and (c) show each of the three target marker’s magnitude of motion respectively. (a) shows a gradual drift of 0.1 mm over the 40 minutes. (b) shows no drift however (c) displays a degree of oscillation.

The stability experiment which tracked three static target markers showed very small, sub millimetre drift and noise. The magnitude of motion, measured as the absolute difference between the initial position of the sensor and its position at any frame, of all three markers are shown in Figure 8.10. The mean range and standard deviation of the 40 minutes of data was 0.49 mm and 0.07 mm respectively. Over a shorter time period of 5 seconds the mean range and standard deviation was 0.03 mm and 0.02 mm respectively. The noise in this tracking system is assumed to inhomogeneous and random as it is caused
by the errors in locating the edges of the target markers in the pixelated image.

**Lighting**

Figure 8.11: The marker was initially under normal lighting conditions. After a torch was shone on the marker, the marker appears to move by 3 mm away from the camera. (b) Shows the effect of disabling automatic gain/contrast/brightness control.

Figure 8.11 shows the effect of drastically changing the lighting levels on the apparent position of the marker. The marker appears to move away from the camera as the regions around the marker boundary appear brighter and larger. This would make the black region of the marker appear smaller which is interpreted by the algorithm as being further away from the camera. Disabling the automatic gain, contrast, and brightness causes a greater apparent shift, as can be seen in Figure 8.11b. The overshoot seen when the torch is turned on in Figure 8.11a is likely due to the automatic control attempting to compensate for the increased brightness. Without this compensation the marker would appear to have moved further, which is the case in Figure 8.11b.

**Marker Curvature**

The standard deviation of the rotation parameters for the flat and curved markers, Table 8.4, show that it is essential for the marker to be flat. This experiment shows that the apparent size of the marker in the image has a greater effect than the curvature of the marker as the curved $20 \times 20\text{mm}^2$ marker shows less variation than the flat $5 \times 5\text{mm}^2$. All markers show more variation if they are on a curved surface, which reinforces the need for a flat and rigid marker.

**Marker Size**

Table 8.5 shows the tracking uncertainty increase as the marker size decreases. The differences between these values and those presented in Table 8.4 could be due to the
8.4. DISCUSSION

Table 8.4: Standard deviation of markers on a flat and curved surface.

<table>
<thead>
<tr>
<th></th>
<th>20×20mm²</th>
<th>10×10mm²</th>
<th>5×5mm²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Flat</td>
<td>Curved</td>
<td>Flat</td>
</tr>
<tr>
<td>Rx</td>
<td>0.05°</td>
<td>0.12°</td>
<td>0.10°</td>
</tr>
<tr>
<td>Ry</td>
<td>0.06°</td>
<td>0.20°</td>
<td>0.13°</td>
</tr>
<tr>
<td>Rz</td>
<td>0.01°</td>
<td>0.03°</td>
<td>0.02°</td>
</tr>
</tbody>
</table>

variation in distance from camera to marker, as this was not as tightly controlled in the Marker Curvature experiment. At a distance of ~10 cm, the 10×10mm² marker appears to have a suitable combination of small size yet low standard deviation in its measured pose.

Table 8.5: Standard deviation of rotation parameters for different sized target markers

<table>
<thead>
<tr>
<th></th>
<th>20×20mm²</th>
<th>10×10mm²</th>
<th>5×5mm²</th>
<th>3×3mm²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rx</td>
<td>0.06°</td>
<td>0.09°</td>
<td>0.22°</td>
<td>0.45°</td>
</tr>
<tr>
<td>Ry</td>
<td>0.07°</td>
<td>0.11°</td>
<td>0.33°</td>
<td>0.88°</td>
</tr>
<tr>
<td>Rz</td>
<td>0.01°</td>
<td>0.02°</td>
<td>0.07°</td>
<td>0.06°</td>
</tr>
</tbody>
</table>

8.3.4 Volunteer Tracking

Figure 8.12 show the motion and the separation of the three markers worn by the volunteer during a 15 minute period where the volunteer was asked to remain motionless.

After 15 minutes, the volunteer was prompted to undergo facial expressional changes to move the markers non-rigidly. Figure 8.13 shows the x-translation and the separation of the markers during this time. Peaks can be seen in the separation data which indicate that a marker has moved non-rigidly to the others and that the tracking is less reliable at this point.

8.4 Discussion

With the progression made in markerless motion tracking it is possible to overlook potential advanced marker based tracking schemes. Markerless motion tracking schemes such as the previous work done with the Kinect in [Noonan et al. 2012a], the work done in [Olesen et al. 2012] using structured light scanners, and the feature detection and tracking in [Kyme et al. 2011a] operate by construction a surface or set of points on the surface of the head. In every frame of tracking data, these points are registered to track the underlying motion of the surface. Therefore motion tracking is only reliable if it possible to filter out the points of the surface that are moving non-rigidly. If the number of points being tracked is small, tracking is more likely to suffer from facial deformations.
Figure 8.12: (b) The target marker tracking of multiple markers shows good correlation between the tracking in each axis. The high spatial accuracy of the motion tracking allows for the head motion due to breathing to be easily seen in the y-axis. (d) The Euclidean separation between the three markers are shown for markers $1 \leftrightarrow 3$, markers $1 \leftrightarrow 2$, and markers $2 \leftrightarrow 3$. The distances remain constant over the 15 minutes, apart from a peak at 2.5 minutes.
Figure 8.13: (a) The displacement of the three target markers in the x-axis over a period of 30 seconds. (b) Shows the Euclidean separation $d_{13}$, $d_{12}$, and $d_{23}$ between markers $1 \leftrightarrow 3$, markers $1 \leftrightarrow 2$, and markers $2 \leftrightarrow 3$ respectively. The peak at 1.6 seconds is present in $d_{12}$ and $d_{13}$. Since $d_{23}$ does not peak then the marker that has moved non-rigidly is marker 1, i.e. the marker that is common to $d_{12}$ and $d_{13}$ but not $d_{23}$.

Marker based tracking operates differently as the system is directly measuring the position of the marker and, it is assumed, the head. Marker tracking does not need to calculating the motion from multiple frames of co-registered point clouds. This reduces the complexity of the algorithm needed to ensure reliable tracking, as non-rigidly moving markers can be identified by measuring the separation of marker pairs, as demonstrated in MTMT.

Marker based tracking system such as the Polaris Position Sensor can not take advantage of multiple tracking tools for use in brain PET as the system has a requirement for the tracking tool to have a minimum separation of spheres of 50 mm. This prevents the use of multiple markers on a subject’s head, as there is simply no method of securing tools of that size to the head securely and comfortably. Therefore the Polaris is only used with a single tracking tool attached to the head and it is difficult to determine if and when tool slippage has occurred.

The experiments performed in this work showed that consumer grade hardware and opensource software could be utilised to provide accurate motion tracking for use in a motion correction scheme. The camera calibration was performed to sub-pixel accuracy which allowed for accurate 3D tracking to be performed. The spatial separation of the
markers in Figure 8.7 agreed with the true values to within 1 standard deviation, with a mean difference of 0.2 mm to the expected values. The rotational similarity of the 4 markers was determined by calculating the root mean square difference between the 4 measurements. The maximum RMSD was 0.6° which would result in a translational shift of 2.1 mm at a distance of 200 mm. This calculation indicates how much the error in the rotation of the marker would affect a voxel’s position at the base of the brain (or a point 200 mm from the marker). The value of 2.1 mm is of the order of a HRRT image voxel and voxels closer to the marker would be less affected. To reduce this effect, more markers could be placed at lower points on the head (for example, each temple) so that the markers are closer to the brain volume. This would require stereo-calibrated, multiple digital cameras to be used to cover the front and side surfaces of the head.

These experiments used a standard optical camera which required visible light to view the target markers. It is recommended that an IR sensitive camera be used for future investigations. The small peaks and slight drift which can be seen in Figure 8.10 could be due to a variation in the lighting levels of the room. indicates that room illumination could effect tracking as changing ambient lighting could affect the contrast and appearance of the corners of the markers. This would cause the marker to appear to move. An IR camera would require an array of IR light emitting diodes that can provide a constant source of illumination for the markers. This would allow the target marker tracking to be performed in darkness for the subject, which is needed for certain scans such as FDG brain activation studies. Figure 8.11 highlights the need for consideration on the position and level of IR illumination to ensure that the position of the marker relative to the light source does not affect the marker’s apparent size in the camera image. A constant and homogeneous level of lighting could be created using multiple IR LEDs reflecting off a diffuse surface.

Tracking stability was investigated to determine which factors have the greatest effect. The marker size and marker curvature experiments demonstrate that it is crucial for the marker to be imaged with sufficient resolution to allow for accurate positioning. At a distance of ~100 mm, the 10 × 10 mm² marker appears to have a good trade off between small size yet stable tracking, as for the flat surface tracking the standard deviation of the rotation parameters were ~0.1°. It must be noted however that when curved, the 10 mm marker stability decreased to be similar to that of the 5 mm marker on a flat surface. To have smaller markers yet to retain stable tracking a higher resolution, or physically closer camera would need to be used. The PnP algorithm is iterative and can generally quickly converge when using 4 points, however to retain real-time rates (30 Hz) a maximum number of iterations is set. A residual error level is set to break out of the iterative process however for curved markers it is possible that the algorithm iterates to its maximum limit. This is due to the corners of a curved marker appearing as a rectangle, as shown in Figure 8.4. The iterative PnP algorithm may introduce bias when reporting the pose and position of the marker. This could manifest as the marker being measured as being at a greater distance from the camera or at an angle relative to the camera. To
investigate the behaviour of the tracking system further it would be useful to plot the reported position of both flat and curved markers as a function of number of iterations.

The choice of camera and lens essentially defines the resolution of the tracking system, as the more focussed on a small area of the head the better resolved a single target marker will be. For using more than one target marker the field of view of the camera needs to be wide enough to view all the markers simultaneously. This reduces the spatial accuracy of the tracking, however increases the robustness against non-rigid facial deformations. A multicamera array would be a possible solution to allow for multiple target markers to be tracked at a high resolution. This array would have to be calibrated using standard multicamera calibration techniques so that each target marker will be tracked in the common multicamera array space.

To spatially calibrate marker space to PET space, a point source marker was produced. By placing a drop of $^{18}$F on the centre of a marker, a point correspondence could be found in both marker space and PET space. It was decided to move the point source marker through the field of view, rather than create an array of point source markers. This was due to the requirement of the point correspondences to be non-coplanar, to reduce contusion of pairing up the points, and to limit the variability of the position of the $^{18}$F on the marker. The 12 positions allowed for the initial 4 to be used to calculate the transformation matrix to then validate on the following 6 frames. This validation showed that the transformation between marker space and PET space could be calculated using this method to a mean error of 0.7 mm.

Temporal calibration remains largely as future work. The current implementation is to start tracking manually at the start of the scan and to use file creation times to align marker tracking data to PET listmode data. This method is user-intensive and ill suited to a event based motion correction scheme where precise alignment is essential. Future work aims to develop of system to input gate tags to the PET listmode through the scanner gantry that correspond to specific time frames as measured by the MTMT computer.

The volunteer tracking was the first attempt to evaluate if the separation of the markers could be used to determine if the tracking is reliable. By comparing the separation plots of the low-motion scenario Figure 8.12d to the case where the face underwent changing expressions Figure 8.13b it is clear that in the latter case the separation of the markers does change. Future work will evaluate if the system can be reliably used to assign a confidence level to each of the marker’s tracking data.

Additional future work exists in the method of fixing the markers to the subject’s head. The method of using a layer of microporous tape before attaching the printed target marker on adhesive label paper allowed the target marker to be securely fixed to the skin with a lowered risk of an allergic reactions. The low cost and simple production process of the markers allow for the markers to be disposed of after use. However it is recommend that a more rigid material than paper is used for future studies to ensure the marker remains flat. Also if the adhesive material was hypoallergenic it would remove the need
of the intermediate layer of microporous tape. This would reduce the surface area of skin in contact to the marker reducing the chance of the marker being stretched or compressed when the skin moves.

8.5 Conclusion

A motion tracking system is developed and evaluated for use in the HRRT. The rational for this work was to provide an accurate and reliable motion tracking system for scanners whose geometry prevents the use of markerless motion tracking systems. An advanced marker based tracking system using a single digital webcam is evaluated and by using multiple, small, printed target markers fixed to the subject’s forehead it is possible to track rigid body head motion while being robust to skin deformations. This robustness is provided by the use of multiple markers whose relative position to each other will noticeable change if the skin distorted non-rigidly to the skull.
CHAPTER NINE

APPLYING ADVANCED MOTION TRACKING TECHNIQUES TO MEASURING RESPIRATORY FUNCTION

Abstract

Clinical FDG oncology PET whole body imaging is acquired over many respiratory cycles which can result in spatial blurring, especially in the thorax. This effect may lead to errors such as a reduction in measured uptake, incorrect volume delineation and misalignment with anatomical imaging. Current commercially available systems for externally monitoring respiration were originally designed for radiotherapy patients who are coached on how to breathe in a regular rhythm while using the tracking system. For PET studies, this coaching would expose the radiographer to increased radiation dose from the patient and could affect throughput in busy clinical environments. These commercial solutions have been shown to perform poorly in PET studies. In this chapter the focus of attention is switched from tracking the head in brain PET, which acts as a rigid body, to measuring the motion of the torso during whole body PET/CT, which undergoes many different types of motion. The experiments detailed in this chapter were performed to assess the efficacy of novel tracking techniques against current commercially available respiratory motion measuring devices.

9.1 Introduction

The main source of motion in whole body PET scans is due to the subject breathing which causes the internal organs to move non-rigidly. This motion is unlike the slow drifting or rapid twitching seen in rigid body head movements. In addition to the cyclic motion caused by breathing, movements caused by the subject repositioning, talking and swallowing occur during whole body imaging. Together these disparate sources and types of motion combine to confound the task of motion correction in whole body PET imaging. In listmode data acquisition it is possible to re-frame the PET data in post processing...
which allows for the data to be gated, i.e. divided into short time frames that correspond to specific time points along the respiratory cycle. By doing this each time frame, or gate, should contain a reduced amount of intra gate motion [Nehmeh et al. 2002]. Respiratory, and cardiac in specific studies, gating has become a standard method of reducing image blur in emission tomography and many methods have been proposed that provide the respiratory signal that can trigger the gating process. These can generally be divided into two classes, data driven and external device driven. Data driven methods such as [Visvikis et al. 2003] and [Buther et al. 2009] are advantageous as they require no additional hardware or additional tasks for the radiographer or scanner technician to perform during scanning. However as with all data driven motion techniques in PET care has to be taken as low count statistics may hinder accurate motion tracking. Externally driven gating uses specific hardware such as pressure belts or tracking tools, such as the Anzai belt and Varian RPM (see §3.3.2) which aim to measure the respiratory rate by monitoring the expansion or surface motion of the torso. In practise these devices often have difficulty in tracking motion accurately as they can slip or are badly positioned on the torso. For clinical oncology FDG PET the patient is generally injected with activity prior to being positioned on the scanner and as such any increase in time that the radiographer has in fixing motion tracking devices to the patient increases the dose to the radiographer. Often the patient has to be trained in how to breathe with the tracking devices which can slow down throughput in busy clinical suites.

Figure 9.1: Kinect surface created using the Point Cloud Library opensource implementation, KinFu. This image demonstrates the availability of surfaces for the Kinect to monitor for whole body PET imaging.

Due to the success of applying the Kinect and monocular target marker tracking tech-
9.2 METHODS

In this chapter, two methods are presented that provide a respiratory trace which use either printed target markers or the Kinect point cloud data. These methods are compared to the tracking data provided by the Varian RPM and Anzai pressure belt (see §3.3.2) on phantom and human volunteer data. In addition to presenting comparative results of respiratory traces, two methods are described where multiple points on the chest can be simultaneously tracked. Finally, a case is presented in which an oncology patient was scanned for 10 minutes while being monitored using the Kinect centre of mass and the Anzai pressure belt. The patient was recruited to a study at Imperial College Healthcare, Radiological Sciences Unit, Charing Cross Hospital, London.

niques in brain PET, it was decided to investigate their implementations into whole body PET imaging, specifically to measure the displacement of the chest for respiratory gating. As yet, there is a distinct lack of markerless respiratory monitoring devices available for whole body PET. Recently, [Xia et al. 2012] used the Kinect to track the displacement of a board that was placed on the surface of the chest. While novel and relevant to the work in this chapter, with their goal of tracking the displacement of the chest using the Kinect sensor, this study failed to exploit the full potential of the Kinect. Their reasoning for using the board was to attempt to avoid occluding regions of folded clothing since the Kinect had been placed at a shallow angle by the feet of the subject. However, there notably exists simpler methods to track flat surfaces such as the previously described monocular target tracking and the stereo target tracking in [Kyme et al. 2011a].

An ideal markerless respiratory gating device would combine the benefits of a high spatial and temporal resolution of an external device with the benefits of a data driven method that has no interference to the patient. The aims of the investigations in this chapter were two-fold. Firstly, replicate the operation of current commercially available respiratory monitoring devices using off-the-shelf equipment and validate the accuracy of these new techniques and secondly to improve upon the currently available solutions by designing methods using the new equipment and techniques to be able to capture more detailed motion to better describe the internal organ movement. To achieve these aims experiments were performed to validate the modified tracking software against currently used commercially available tracking systems. Both phantom and human validation was performed and the Kinect based tracking system was used for a clinical oncology PET/CT scan.

9.2 Methods
9.2.1 Single Point Motion Tracking

The Varian RPM and the ANZAI Pressure Belt

The Varian RPM and Anzai pressure belt systems can only measure one point on the surface of the chest, which may not correspond to the internal tissue or organ of interest. Indeed, if there were multiple regions of interest within the investigation the belts would ideally have to be repositioned on the patient during the scanning session, thereby increasing the time duration of the scan. Due to the contactless nature of the Kinect based tracking system and its large field of view, this would not be necessary as many regions of interest can be tracked simultaneously.

Both commercial systems need to be carefully prepared by the radiographer to ensure that tracking tool or belt is placed over the section of the chest which experiences most movement. Since this varies from person to person this setup has to be done while the subject is on the scanning bed, thus increasing the dose to the radiographer. It is also possible for the tracking tool or belt to slip from its starting position which can degrade the quality of the respiratory trace. Finally, another issue with only using one marker is that it not possible to accurately determine if the subject shifts and repositions mid-scan.

Kinect Respiratory Tracking

As described in previous chapters, the Kinect provides a point cloud representation of surfaces. To acquire a respiratory trace from the point cloud data, a region of interest is drawn that corresponds to a area of the chest and the centre of mass of the point cloud is calculated by taking the mean of all the points in the region of interest. The region of interest was selected as a rectangle $100 \times 150$ depth pixels positioned in the centre of the field of view of the Kinect. The Kinect should be placed as face on as possible to avoid occlusions and the effect of varying depth accuracy with distance to sensor, which would affect the accuracy of the 1D respiratory trace.

![Image of a truncated point cloud of a volunteer's abdomen](image)

Figure 9.2: A truncated point cloud of a volunteer's abdomen is shown as viewed from 3 views. This subsection of the whole Kinect point cloud data is used to monitor the displacement of the chest wall during respiration by calculating the centre of mass, or centroid, of the set of points. This point cloud consists of an array of $100 \times 150$ (15000) points.

In the experiments prior to the oncology patient the region of interest was centred on
the subjects thorax. Under non-scanning conditions, this was found to be possible on
volunteers by placing the Kinect on a 1.8 m tripod positioned at the rear of the PET/CT
gantry with the Kinect facing downwards. The in-house code first determines the mean
orientation of the surface plane over 10 Kinect frames and then measures the component
of motion of the surface along the normal vector of the surface. This ensures that the
maximum amplitude is always measured.

Single Target Marker Tracking

The work undertaken in Section 3.2.4 on tracking a square patterned marker was
scaled up to determine whether it could also be used to track the respiratory cycle. A
50 mm² marker was worn on a subject’s chest and tracked using a Logitech c910 webcam.

Comparing Kinect and Target Marker Tracking

An experiment was performed to compare the respiratory tracking data provided by
the target marker and Kinect centroid approaches using a human volunteer. A target
marker was worn by a volunteer while simultaneously being tracked by the Kinect centroid
method. The purpose of this experiment was to evaluate if the target marker approach
could be used to accurately measure the respiratory rate for situations when the Kinect
could not be used.

9.2.2 Validating Respiratory Motion Tracking

The following experiments were performed to validate the use of the Kinect centroid
tracking method against the commercially available solutions\(^1\).

1. Phantom: Varian RPM ↔ Kinect centroid
2. Phantom: ANZAI pressure belt ↔ Kinect centroid
3. Human: ANZAI pressure belt ↔ Kinect centroid

(1) For the Phantom: Varian RPM ↔ Kinect centroid experiment, a NEMA image
quality phantom was fixed to a surface that moved sinusoidally at a constant rate and
magnitude. A Varian RPM tracking tool was rigidly attached to the surface of the phantom
and was tracked for 45 cycles. The front surface of the NEMA phantom was imaged
with the Kinect to provide simultaneous tracking. This experiment was performed at the
Nuclear Medicine Centre, Central Manchester University Hospitals, Manchester, UK.

(2) For the Phantom: ANZAI pressure belt ↔ Kinect centroid experiment, the ANZAI
pressure sensor was placed inside the ANZAI respiratory motion phantom, which consists
of a mass connected to a linear actuator which is coupled to the pressure sensor. The user

\(^1\)Simultaneous Varian ↔ ANZAI ↔ Kinect experiments were not possible as no clinical suite had both
tracking systems available.
manual for the Anzai phantom states that the stroke length is 20 mm at a rate of 10 strokes per minute. The front surface of the mass was imaged by the Kinect to provide simultaneous tracking. This experiment was performed at Imanova Centre for Imaging Sciences, Hammersmith, London, UK.

(3) Finally, during the Human: ANZAI pressure belt ↔ Kinect centroid experiment, a volunteer was monitored with the high sensitivity ANZAI pressure sensor and the chest pressure belt. The Kinect region of interest corresponded with the region of the chest that the belt was placed upon. The volunteer was asked to undergo periods of normal breathing inter-spaced with breath holds, small amplitude rapid breathing, and deep breathing.

9.2.3 Augmented Respiratory Monitoring

The following three experiments investigate novel implementations for the systems developed in this thesis, assessing how they can provide additional information beyond respiratory tracking.

Contactless heart rate measuring with the Kinect

An experiment was performed to investigate whether the Kinect could resolve the abdominal aortic pulse, which can be seen as a small amplitude tremor of the surface of the abdomen. To observe the abdominal pulse, the volunteer was asked to undergo a breath hold followed by normal breathing. Once the abdominal pulse rate was measured over the breath hold, the period of normal breathing was Fourier transformed to separate the abdominal pulse rate from the respiratory signal.

Respiratory MTMT

To measure the displacement of various points on the torso, multiple target markers tracking (MTMT) can be used. In this experiment, six 50 mm$^2$ markers were placed over the torso and tracked using a single digital camera and the same code developed in §8.2. The camera was placed 50 cm from the markers, so that the field of view of the camera was filled by the array of markers. The resolution of the camera was set to 640 × 480 at 30 Hz. The markers were printed onto card and attached to the subjects clothing using adhesive tape.

Kinect surface meshes

The Kinect centroid method is a simple application of the Kinect to produce a respiratory trace but does not take full advantage of the depth information offered by the Kinect. Using KinectFusion, the Kinect is capable of producing a high resolution 3D surface of the chest. This quality of depth information is useful to a wide range of medical imaging applications such as mammography and electrical impedance tomography (see §10.2), where knowledge of the shape of the object being scanned can improve the quantifiability.
9.2. METHODS

Figure 9.3: An array of 6 target markers were placed on the torso as shown. In this experiment only the direction of motion normal to the chest surface, i.e. the Y-axis of the marker, was needed. Motion in any other axis would correspond to the markers or clothing slipping during scanning.

of the data. Surface meshes of torsos were taken from a KinectFusion scan and from a single frame of Kinect data to observe the range in surface quality.

9.2.4 Oncology Patient Data

Finally, an experiment was conducted to assess the translation of the Kinect tests to the clinic and specifically monitor the respiration of an oncology patient. A patient, as part of a larger study which aimed to scan 30 patients with known tumours in the lower lobes of the lungs of dimensions 10-30 mm, was referred for a PET/CT scan at Charing Cross Hospital. For this patient, an additional 10 minute listmode was acquired at the end of the normal scan over the lower lungs. During this acquisition, the Anzai and Kinect monitored the respiration of the patient to enable gated reconstruction. For this patient, the Anzai system used the low sensitivity sensor and was placed over the patients thorax at the point which was visibly moving with highest amplitude with breathing. The Kinect was placed on a 1.8 m tall tripod at the read of the gantry looking towards the patients chest. The point cloud was able to view the top of the patients chest.

As will be discussed later, during scanning the Anzai pressure belt behaved temperamenteally. Although the pressure belt was measuring and recording a respiratory signal, the gating trigger was not consistently firing. This resulted in only 32 triggers being sent to the PET listmode. Whereas on analysis of the Anzai raw data file, 119 peaks were observed. A Matlab script was written to determine the peaks of the Anzai data from the raw data and insert these as gate tags into the listmode after the previous 32 had been scrubbed from the original listmode. The listmode was then reconstructed using the off-line Siemens e7 tools using 8 phase gates, 4 iterations and 14 subsets OP-OSEM with PSF modelling. The reconstructed images were smoothed by a 5 mm Gaussian filter. Ideally this gated data set would be processed through a non-rigid coregistration algorithm to obtain a single motion reduced image frame, however the purpose of this experiment was to
develop a technique for inserting alternative gate tags into the listmode so that a range of registration techniques could be used.

9.3 Results

9.3.1 Respiratory Gating

(1) The end points of the cyclic motion of the phantom were determined in both Kinect and Varian data sets. The Kinect data matched the Varian system in measuring the rate of the cyclic motion. This indicates that the Kinect system can be used as a replacement for the Varian RPM system for phase based respiratory gating motion correction. Figure 9.4 also shows that the Kinect is much more accurate in determining the amplitude of movements. For this experiment, the end points of the cyclic motion should be constant, a constraint due to the design of the phantom, however the Varian RPM system shows greater variability with this measurement. Over 45 oscillations of the phantom the standard deviation of the measured end points using the Varian tracking was 0.56 mm whereas the Kinect had a standard deviation of 0.08 mm.

Figure 9.4: A sinusoidally moving phantom was tracked using the Kinect centre of mass and the Varian RPM system. The end points of the cyclic motion were calculated and the Kinect is almost an order of magnitude more accurate in measuring the end point positions than the Varian RPM. Figure taken from [Noonan et al. 2012b].

(2) Figure 9.5 shows the simultaneous tracking of the Kinect centroid and the high sensitivity Anzai pressure sensor using the Anzai respiratory motion phantom. Unlike the pure sine wave motion of the respiratory phantom used in Figure 9.4, the Anzai phantom moves in a more anthropomorphic manner. The Anzai data and Kinect tracking data showed very good agreement, after normalisation between displacement and pressure level. The mean stroke length as measured by the Kinect was 20.4 mm at a rate of 6.01 ± 0.04 seconds per stroke. The Anzai high sensitivity sensor does not measure amplitude yet measured 5.94 ± 0.47 seconds per stroke. The order of magnitude greater standard deviation found in the Anzai timing is likely due to the nature of the sensor to saturate at the points of inflection. Smoothing could be used to estimate the peaks hidden in the regions of saturated signal.
Figure 9.5: (a) The Anzai respiratory phantom was used with the high sensitivity pressure sensor, which the resulting respiratory trace shown above. The full range of motion of the actuating arm of the phantom is 20 mm which corresponded to 110 integer levels of pressure. (b) The Kinect was used to simultaneously measure the position of the actuating arm of the Anzai respiratory phantom. The Kinect sensor was placed in line with the direction of motion.

Figure 9.6: (a) The high sensitivity Anzai pressure sensor was worn by a healthy volunteer. The breathing included periods of breath holdings (from 130s-134s, 162s-170s, 230s-235s, 329s-344s) and a period of small amplitude, rapid breathing (from 282s-309s). Similar to the phantom experiment, a Savitzky-Golay smoothing filter was applied. (b) The Kinect was positioned directly facing the chest so that the region of interest corresponded to the position of the pressure sensor. The large translation at 146s was due to the subject repositioning in the supports and is not visible in the Anzai data. (c) The last period of breath hold is shown enlarged for ease of viewing. The discrete levels of pressure can be seen, showing the high sensitivity Anzai pressure sensor is capable of detecting the abdominal aortic pulse. The maxima and minima are calculated from the smoothed trace. (d) The mean peak to peak displacement for this breath hold period is 0.25mm as measured by the Kinect.
(3) Figure 9.6 shows the respiratory trace from a human volunteer while being tracked by the Anzai pressure belt and the Kinect centroid technique. Both tracking methods demonstrate highly accurate tracking of the respiratory motion, however since the Anzai belt can only measure changes in pressure levels, it is unable to determine if the subject has repositioned (for example at 150 seconds). The abdominal pulse is visible in both tracking methods, however a smoothing filter had to be applied to both data sets due to the discretisation of the pressure levels and the noise of the Kinect centroid data.

9.3.2 Augmented Respiratory Monitoring

Figure 9.7: (a) Due to the high accuracy of the Kinect, 5 pulses can be easily observed during a breath hold, beginning at 48.1 and ending at 52.5 seconds resulting in 68.6 beats per minute. (b) Fourier Transforming the whole data set shows that both the heart rate and the breathing rate can be obtained from Kinect data. Figure taken from [Noonan et al. 2012b].

Figure 9.7 shows Kinect based tracking on a human volunteer under normal breathing conditions after a short 5 seconds breath hold. During the breath hold period, the abdominal aortic pulse can be seen as small vibrations of the chest surface. This contactless measurement of the pulse could be used as a trigger for studies requiring cardiac gating and demonstrates a use for the Kinect other than providing a respiratory trace.

Comparing the motion as measured with the target marker and the Kinect on a human subject showed good correlation as can be seen in Figure 9.8. The 6 target markers as shown in Figure 9.3 were tracked using respiratory MTMT. Figure 9.9 shows the normal motion of each marker demonstrating the variation of motion amplitude over the torso due to respiration.

Figure 9.10 shows two Kinect generated surfaces of a volunteer’s torso without (Figure 9.10a) and with (Figure 9.10b) clothing. KinectFusion produces surface meshes with better spatial resolution at the trade off of reduced temporal resolution.
Figure 9.8: The centre of mass of the Kinect measured region of interest varies as the subject breathes. Similarly, the position of the target marker also varies as the chest wall expands. The two tracking systems produce comparable respiratory traces of a subject undergoing normal breathing and periods of breath holds. The small tremors caused by the abdominal aortic pulse are more visible in the Kinect data as it is sampling a larger number of points than the target marker system which is only tracking a single point on the abdomen.

Figure 9.9: The 6 plots shown represent the motion of the chest as measured by the multiple target markers in their Y-axis. Each subplot shows the same 50 seconds tracking and has a y axis scale that spans 10 mm. For this volunteer the majority of motion occurred in the middle and lower regions of the torso. There appears to be no contralateral differences between the motion of the marker pairs.
9.3.3 Oncology Patient Data

The combined monitoring of a lung cancer patient using the Anzai pressure belt and the Kinect resulted in interesting and useful data for future investigations. Initially the Kinect was able to monitor the change in the displacement of the chest, however after 10 seconds of scanning the patient changed their breathing pattern from the top of the chest to their lower abdomen. This resulted in the motion of the chest surface to be outside of the field of view of the Kinect and so Kinect respiratory tracking failed. The Kinect observed a large (>1 cm) displacement however it was not possible to confirm that the patient has repositioned from the Anzai data. The baseline pressure level measured by the Anzai lowered after 4 minutes, indicating a repositioning event had occurred. Figure 9.11 shows the Anzai pressure trace from the low sensitivity pressure sensor. The red bars denote the gates that were injected into the listmode by the Anzai laptop during the scan. It is clear that many peaks in the respiratory trace were not injected into the listmode, hence the need to scrub the tags that had been inserted to then reinsert new tag words into the listmode. These tags were inserted at time points calculated using a Matlab peak detection routine.

Figure 9.12 shows the 10 minute reconstructed data, without gating. A region of high activity can be seen which corresponds to the position of the tumour, however it appears elongated along its major axis. It is likely that a non-rigid registration algorithm applied to the gated image frames would produce a better defined image.
9.4 Discussion

The Kinect based respiratory tracking can monitor a respiratory trace to higher degrees of precision than the commercially available Varian RPM and allowed for repositioning to be measured unlike the Anzai pressure belt. This was shown on human and phantom data taken outside of the PET/CT scanner. Under scanning conditions, it was discovered that the Kinect could not be reliably positioned in such a way as to accommodate for a wide range of body shapes and sizes. As a result of the failure of the Kinect to adapt to the patient changing their breathing pattern, the following recommendations are made for future scans. Firstly the mirror insert should be used such that the mirror directs the Kinects structured light onto the surface of the patients chest. This will enable the Kinect to appear to be directly in front of the chest surface for optimal image acquisition. The size of the mirror and its proximity to the surface of the chest place constraints on the field of view available for respiratory monitoring. To be robust against changing in breathing styles, the field of view needs to cover as much of the chest as possible.

Figure 9.11: The raw Anzai data is shown from the 10 minute listmode scan. At 4 minutes the patient changed breathing styles which appeared to trigger the Anzai gating pulses to the PET/CT scanner. The peaks from the raw Anzai data were calculated using a Matlab script.

Figure 9.12: (Left to Right) Transaxial, coronal, and sagital slice showing the single frame reconstruction of 10 minute listmode acquisition of a lung cancer patient. Crosshairs have been added to the images to show the tumour in the lower lung.
The system shown in [Xia et al. 2012] did not attempt to optimise the position of the Kinect relative to the surface of the chest, instead they required the use of a flat board placed on the chest. By using a mirror the Kinect should be able to accurately measure the displacement of the chest surface without any contact to the subject.

The contactless nature of the Kinect allows for;

1. Less contact time between the radiographer and radioactive subject, reducing the dose received by the radiographer,

2. Less pre-scan preparation time to train the patient how to use the motion tracking device allowing for potentially more scans to be performed in a busy scanning suite.

3. The region of interest can be set for different regions on the subject for the different bed positions so that only the motion of the region that is currently being imaged is monitored,

4. Routine aseptic and clean use.

The target tracking method is more able to be scaled to fit scanners where geometrical constraints may prevent a Kinect type depth sensor from operating. For when geometry is not a problem, however, the Kinect is preferred as it is able to measure the entire front facing surface of the body.

The high sensitivity of the Kinect allows for the abdominal aortic pulse to be measured. Other contactless methods for measuring the pulse have been demonstrated such as measuring the skin colour change when blood is pulsed through in [Wu et al. 2012] however they are sensitive to lighting conditions, skin tone and they need a direct view at the surface of the skin. The Kinect is able to measure the abdominal pulse through items of clothing and is insensitive to both lighting conditions and skin tone. This is unlike other methods that aim to measure respiratory function such as the recently developed Respiratory Volume Monitor, RVM (Respiratory Motion, Inc., Waltham, MA). The RVM consists of 2 sensor pads, each consisting of 3 electrodes, placed at the sternum and the right midaxillary line, in line with the xiphoid. As the subject breathes the electric signal between the two sensor pads will change depending on the volume of the lungs. In [Voscopoulos et al. 2013] the system is evaluated with volunteers undergoing simultaneous monitoring with the RVM and a spirometer. The system demonstrated an average error in the accuracy of measuring the minute volume, tidal volume, and respiratory rate of less than 10%. Unlike the Kinect, the RVM requires the placement of multiple sensor pads on the subject’s skin and accuracy is likely to suffer if these pads slip during use or are poorly attached.

The target tracking method was able to simultaneously track the motion of different region of the abdomen by increasing the numbers of markers placed onto the chest. This demonstrated the scalability of the technique to provide augmented tracking information about the varying movements of the chest surface. Marker tracking would be ideally used
if the markers could be stuck directly to the skin, rather than an intermediary layer of clothing. Rather than tracking a single point, or perhaps 6 points when using the marker targets, on the surface of the chest the Kinect is theoretically able to track many thousands of points. The large amount of raw Kinect data taken over a scanning session makes storage an important consideration as the $640 \times 480$ 11 bit depth video at 30 frames per second for 1 hour would take up 42.5 GB of disk space (if uncompressed).

A real time solution to monitoring the motion over the entire chest surface would be to use a feature of KinectFusion which measures the point to plane error for each new depth image pixel. In such a system, KinectFusion can be used to generate a high resolution surface map at the time of CT acquisition. During the PET listmode acquisition, the integration of new Kinect frames can be paused, then the point to plane metric of each new Kinect depth image pixel can be used to monitor the motion of all points of the surface of the chest.

The natural extension to the investigations set out in this chapter would be to firstly compare the Kinect and target based tracking methods against a range of advanced data driven methods on an anthropomorphic body phantom and human subjects. Following this, an investigation into using the Kinect generated chest surface information as constraints to a non linear deformation and registration algorithm should be a very promising and interesting route of work.

9.5 Conclusion

The initial aim of this chapter was to replicate the currently available external tracking systems in producing a respiratory trace using the Kinect, with the intention of investigating if these new methods could be used to obtain more detailed and useful information such as the shape and position of the body. Respiratory monitoring was successfully performed with the marker tracking and the Kinect centroid techniques. These low cost techniques both offer increased sensitivity as compared the current commercially available solutions for a significantly lower cost. The Kinect technique has the added advantage that it requires no contact with the subject. By reducing the contact time between the subject and the radiographer the dose to the radiographer will be reduced and patient throughput will be improved. The methodology introduced for measuring the entire surface of the torso has potential benefits beyond simple respiratory gating, as this data can be used to restrain non-rigid deformation fields used for internal organ motion correction.
10.1 Thesis Summary

If not accounted for, motion degrades medical imaging data. In PET, motion induces image blurring that reduces the quality of the image data which affects the quantifiability and clinical usefulness of the imaging technique. Motion correction of PET data relies on the ability to measure the motion that has occurred during PET data acquisition. Without accurate and reliable motion tracking data motion correction techniques may miscorrect the data for motion that did not occur.

This thesis on motion correction schemes for PET and PET/CT begins by describing PET as an imaging modality with emphasis on factors which affect image resolution. One of these factors, motion, is the theme of this thesis and is described in Chapter 3 with specific detail given to the methods used to track motion. Motion correction schemes are also discussed however Chapter 3 concludes with the position that there are many advanced motion correction algorithms yet they generally require perfect motion tracking and therein lies the current research gap. The main research objective of this work was to provide an accurate and robust motion tracking technique for brain PET.

The work done in Chapter 4 was to set up an automated motion tracking and frame by frame motion correction scheme using the Polaris position sensor on the PET only High Resolution Research Tomograph and the PET/CT TrueV. Spatial calibration was performed by automatically registering a template of the Polaris tracking tool to the \( \mu \)-map image generated from the transmission or the CT scan for the HRRT and TrueV respectively. Temporal alignment is performed automatically using a device that detects and counts the pulsed light output from the Polaris and injects gating tags into the PET listmodes so that Polaris frame rate can be registered to the elapsed time on the listmode. Chapter 4 concludes by stating the importance of the method of attaching the tracking tool to the subject.

In Chapter 5 analysis is performed on two sets of clinical PET data with Polaris motion tracking and frame by frame motion correction showing examples of apparently success-
ful, questionably successful, and definitely unsuccessful motion tracking. This range in the efficacy of Polaris motion tracking is due to the unreliable rigidity of the tracking tool’s fixation to the subject’s head. For example in a frame of HRRT data the Polaris motion tracking transformed the central voxel of PET space by 14.5 mm away from the position determined by the data driven image co-registration. This is the critical limiting factor which prevents the Polaris from being successfully implemented in a motion tracking scheme as it is essentially impossible to determine from the Polaris data if the tracking tool has slipped and if motion correction is applied to erroneous motion tracking then the PET data will be miscorrected.

A markerless motion tracking scheme is developed in Chapter 6 using the Microsoft Kinect. Markerless motion tracking aims to remove the uncertainty in the tracking data caused by tool slippage by removing the need of a tracking tool. The Kinect provides depth sensing capabilities with a consumer grade device, however it is ill suited to tracking small objects as it was originally designed for use in large spaces. An initial, proof of principle, experiment was performed with offline registration of Kinect point clouds being registered to a CT of a head phantom. The experiment gave encouraging results however the offline registration algorithm took too long (~ 10 seconds) for such a system to be used in a busy clinical environment. The latter part of the chapter describes the method development to validate a realtime implementation of the registration algorithm, Kinect-Fusion, which performs the registration using a graphics processing unit. To validate the Kinect and KinectFusion in a PET/CT scanner a mirror had to be installed to reflect the Kinect’s structured light onto the face of the subject. In phantom studies of a static head phantom the mean range and the standard deviation of the Kinect tracking data was 1.40 mm and 0.21 mm respectively over 70 minutes. In comparisons between Polaris tracking and Kinect tracking, root mean square difference and standard deviation of the two tracking systems was 0.40 ± 0.35 mm. This good correlation with the Polaris tracking combined with the realtime and markerless aspect of the Kinect based system encourages the further investigation of the Kinect for use in clinical brain PET imaging. In Chapter 7 the KinectFusion algorithm and Kinect motion tracking procedure is further optimised on clinical subjects. The software was transferred to the Microsoft implementation of KinectFusion and an adjustable mirror mount was developed. In Chapter 8 an alternative motion tracking system is developed and evaluated for use in the HRRT. The rational for this work was to provide an accurate and reliable motion tracking system for scanners whose geometry prevents the use of markerless motion tracking systems such as the Kinect which require a minimum operating distance which the HRRT cannot provide. An advanced marker based tracking system using a single digital webcam is evaluated and by using multiple, small, printed target markers fixed to the subjects for head it is possible to track rigid body head motion while being robust to skin deformations. This robustness is provided by the use of multiple markers whose relative position to each other will noticeable change if the skin distorted non-rigidly to the skull. Results from these investigations
indicate that the open source tracking system potentially surpasses the spatial accuracy of the Polaris position sensor at much improved comfort to the subject. Motion tracking of a stationary phantom over 40 minutes having a mean range and standard deviation of 0.49 ± 0.07 mm and Over a shorter time period of 5 seconds the mean range and standard deviation was 0.03 ± 0.02 mm. The root mean square difference between target marker tracking and the Polaris position sensor was <0.5 mm.

In Chapter 9, experiments in monitoring respiratory motion using the Kinect and target markers are presented. These experiments compared the respiratory tracking offered by the Kinect and the commercially available Anzai pressure belt and Varian RPM tracking system. A phantom experiment was performed using a sinusoidally oscillating phantom and comparisons of the spatial positioning of the antinode (end) points show the Varian having a standard deviation of 0.56 mm over 45 oscillations whereas the Kinect had a standard deviation of 0.08 mm. Due to the high sensitivity of the Kinect based respiratory tracking it is also possible to observe the abdominal aortic pulse by monitoring the fine tremors of the abdomen. An additional phantom experiment was performed using the Anzai pressure belt and the Kinect which showed comparable results, however on human data the Anzai was not able to determine if the subject has repositioned, unlike the Kinect which is able to calculate the mean displacement of the region of interest.

10.1.1 Limitations of Work

It is clear that work is continuing on obtaining and analysing clinical data. Effort was put into validating the novel tracking methods on phantom data and it is the intention to continue this validation on clinical PET data in the near future. Indeed a research collaboration has been set up between Imanova and Siemens Healthcare to evaluate their developmental motion correction algorithms with the motion tracking data obtained using the techniques developed in this thesis. The phantom validation has shown comparable and improved accuracy in head tracking for the Kinect and target markers respectively and much consideration has been put into ensuring that the tracking techniques are conceptually mature and gives stable, sensible, and reliable data before their introduction to clinical use. For the phantom experiments performed the novel methods had improved accuracy, better sensitivity or required little or minimal contact.

10.2 Future Kinect Projects in Medical Imaging

Through the investigations into head motion tracking in PET, discussions were had with researchers from the wider medical imaging community into how a device like the Kinect could be used to benefit their particular research interests. The investigations in respiratory monitoring presented in this thesis was prompted in part by conversations with the nuclear imaging department at Central Manchester University Hospitals. The
following sections describe some of the interesting methods by which the Kinect has been proposed to be used.

10.2.1 Using the Kinect to constrain non-rigid deformation fields for internal organ motion correction

The Kinect based respiratory measuring in Chapter 9 demonstrated that the Kinect can be used to replace the commercially available solutions for monitoring subject breathing, as the accuracy of the Kinect respiratory tracking outperforms even the marker based motion tracking systems. However, this simple use of the Kinect does not compliment the depth data that the Kinect can obtain. Rather than 1D tracking of a point on the chest, the Kinect is capable of providing depth maps over the whole surface of the torso, allowing for the possibility of using this data in an advanced whole body motion correction scheme. In whole body the problem of internal organ motion is extremely ill conditioned and to date the attempts at defining a motion field have been poorly constrained by the noisy PET data or by insufficient numbers of surface markers. There is both precedence and enthusiasm for non rigid motion tracking of internal organs in radiotherapy applications where precise knowledge of organ motion is crucial for accurate and optimal dose delivery. As a simple example, the Kinect could be used to provide an accurate 3D model of the patient as they progress along multiple treatment deliveries. Body weight can change, which is an indication that the initial treatment plan is no longer valid and the patient would benefit from a new CT to re-plan treatment.

10.2.2 Using the Kinect to provide a real time 3D model for chest electrical impedance tomography

Electrical Impedance Tomography is a low cost, non-invasive, and non-ionizing medical imaging modality which determines spatial distribution of the electrical conductivity of body tissue using electrical measurements from multiple skin electrodes. In the earliest medically applied EIT device [Barber et al. 1983] 16 electrodes were placed in a ring around the forearm. Current was applied across an electrode pair with measurements being taken from all other valid electrode pairs. There are many image reconstruction techniques to solve the non-linear, inverse, and ill-posed problem of producing tomographic images from surface electrode measurements, however a major difficulty in interpreting the images stems from the assumptions made about the bounding shape of the surface where the electrodes are placed. In [Grychtol et al. 2012] it was shown that if a CT of the subject is obtained, the EIT reconstructions are qualitatively superior to simply assuming the bounding shape of the torso is a circle. However it may not always be feasible to obtain CT and the shape of the body is likely to change from CT to EIT acquisition, if they are not obtained simultaneously. In cases such as this, the Kinect could potentially be used to provide real time 3D measurements of surface, providing an accurate and reliable...
10.2. FUTURE KINECT PROJECTS IN MEDICAL IMAGING

10.2.3 Using the Kinect to provide a high resolution model of breast tissue deformation during x-ray mammography

This section was adapted from the co-authored poster ‘Measurement of Compressed Breast Thickness for Breast Density Assessment Using a Games Console Input Device’ at the 6th International Breast Densitometry and Breast Cancer Risk Assessment Workshop 2013, San Francisco, USA.

Breast cancer is expected to account for 29% of new cancer cases in women in 2013 [Siegel et al. 2013]. Early and reliable diagnosis is essential for best treatment and X-ray mammography remains the current gold standard for breast cancer screening. Breast density, the ratio of radio-opaque fibroglandular to fatty tissue in the breast, is one of the strongest risk factors for breast cancer after age. Much of the early work relating density to risk was based on subjective visual assessment by radiologists who estimated the percentage of the breast area occupied by dense tissue. However, density decreases with an increase in BMI and with age [Assi et al. 2011], and BMI in particular is not usually known by the radiologist at the time they review the images. Area-based methods are also subject to variation when the breast is recompressed. Subjective area-based assessments are therefore of limited usefulness when longitudinal measures are required, for example when evaluating response to treatment.

Clearly, precisely measuring mammographic density is of paramount importance when determining breast cancer risk. 2D models based on the percentage area of density in the mammogram do not provide an accurate method of determining the quantity of dense tissue in the breast. As a result, several methods for creating a 3D volumetric measure of breast density have been developed. A calibration object, such as an aluminium step-wedge [Berks et al. 2010], with known attenuation, can be positioned in the X-ray FOV allowing for the attenuation properties of the breast, and therefore breast density, to be calculated. All 3D models of breast density require an accurate estimate of compressed breast thickness. Typically, during mammography, the breast is compressed by a perspex plate to improve the quality of the image, reduce motion blurring and x-ray dose. A machine readout of the compressed thickness is given, but it does not take into account tilt and deformations of the compression plate. By accurately modelling these deformations, an improved model of breast density can be formed, leading to more accurate density estimation.

The specific advantage of using the Kinect is that the IR structured light can pass through the perspex compression plate to directly image the breast tissue. It would then be possible for a KinectFusion type algorithm to create a high resolution 3D model of the breast, allowing for a very accurate and precise calculation of the breast density to be calculated. This would be a low cost upgrade that could be retro-fitted to existing imaging equipment.
10.3 Research Outcome

The main research outcome of these investigations is the demonstration of a real time, markerless, and low cost solution to provide motion tracking information for medical imaging using the Microsoft Kinect. Without the arrival of the Kinect in 2011, and the tremendous open source software support it receives, this significant research outcome would not have been achievable in the time frame of a PhD project. However, by being ideally placed in the field of motion correction in medical imaging, it was possible to quickly realise that the Kinect could be used in the currently unmet clinical need of an accurate and contactless motion tracking device for head and whole body motion. This allowed for the world’s first use of the Kinect for head tracking in brain PET to be demonstrated at the 2012 IEEE Medical Imaging Conference, using the methods and results from §6.2.2.
REFERENCES


REFERENCES


Kyme, A., Se, S., Meikle, S., Ryder, W., Popovic, K. and Fulton, R. (2012). ‘Markerless Motion Tracking Enabling Motion-Compensated PET in Awake Rats’. *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC IEEE)*.


REFERENCES


REFERENCES


LIST OF ABSTRACTS AND PUBLICATIONS


