1-1-2017

Investigation Of The Microsoft Kinect V2 Sensor As A Multi-Purpose Device For A Radiation Oncology Clinic

Evan Asher Silverstein
Wayne State University,

Recommended Citation

Silverstein, Evan Asher, "Investigation Of The Microsoft Kinect V2 Sensor As A Multi-Purpose Device For A Radiation Oncology Clinic" (2017). *Wayne State University Dissertations*. 1877.
http://digitalcommons.wayne.edu/oa_dissertations/1877

INVESTIGATION OF THE MICROSOFT KINECT V2 SENSOR AS A MULTI-PURPOSE DEVICE FOR A RADIATION ONCOLOGY CLINIC

by

EVAN ASHER SILVERSTEIN

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2017

MAJOR: MEDICAL PHYSICS

Approved By:

____

Advisor Date

____

____

____

____

____

DEDICATION

To my parents, for continuing to encourage me and reminding me that I have the ability to persevere.

To the many friends I have made throughout my years at Wayne State, for their support, informed discussions, and allowing themselves to be ~~victims~~ volunteers.

To Michelle, for always being there for me and believing in me; for keeping me sane outside of school; and for never letting me forget that, although we started at the same time, she completed her doctorate before me.

ACKNOWLEDGMENTS

I would like to thank the members of my dissertation committee: Dr. Jay Burmeister, Dr. Michael Snyder, Dr. Joe Rakowski, and Dr. Ning (Winston) Wen. Without your continued support, encouragement, and ideas as I worked my way through this project, I would not have been able to complete it. I would like to specifically thank my advisor, Dr. Michael Snyder, for his unique teaching style and advisory role as I continued through this process. Without his guidance and understanding as I balanced my workload with school, research, and work, I would not have been able to complete this project, obtain a residency, or become the researcher that I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 “INTRODUCTION”

In 1953, just 60 years after Wilhelm Röntgen discovered X-rays and both Marie Curie and Henri Becquerel discovered radioactivity, the first linear accelerator (linac) based radiotherapy treatment of cancer was performed in London. Over the next 60 years, radiotherapy treatments using linacs have advanced dramatically to allow for increased precision while decreasing unwanted dose[1]. Procedures such as Stereotactic Radiosurgery (SRS), Stereotactic Body Radiation Therapy (SBRT), and Volume Modulated Arc Therapy (VMAT) all allow for highly conformal radiotherapy treatments to be created for a wide array of tumors and locations. As improvements upon these procedures progresses, each requires an ever increasing amount of supplemental software and hardware to ensure the treatment to be delivered is done so with extreme accuracy.

The tracking of patients through identity verification procedures and through intra- and inter-fraction motion ensures these exacting and precise procedures are performed to the correct patient and to the correct treatment location. Tracking of the patient and verifying identity as they progress through a radiation oncology clinic becomes exceedingly important to ensure patient safety. Tracking of intra- and inter-fraction motion in real-time helps ensure that these complex and conformal treatments are as precise as possible as well as quantifying said motion to understand how it can affect treatment. Multiple vendors produce a wide variety of devices to fulfill these needs within a radiation oncology clinic, allowing for the clinic to increase patient safety and treatment efficacy using real-time identification and real-time motion tracking.

Patient Verification

In a radiation oncology clinic, verification procedures can generally be grouped into two major categories: Plan Verification and Patient Verification. Both are integral to ensuring patient safety throughout the treatment process and both have recommendations from various governing bodies to ensure clinics meet specific standards for each verification process. Plans created utilizing techniques such as Intensity Modulated Radiation Therapy (IMRT) or Volume Modulated Radiation Therapy (VMAT) are run through various QA procedures to verify the highly specific MLC movements and beam modulation calculated within the treatment planning system by delivering the plan to phantoms or other radiation measuring devices. These pre-treatment QA delivery procedures are compared to the original plan and are required to pass specific criteria before actual patient treatment can proceed[2,3]. For all non-IMRT treatment plans, AAPM TG-40 and TG-114 recommend that monitor units calculated from a treatment planning system have an independent, secondary calculation performed for various sites within treatment plan[4,5]. This ensures that the physicist is not required to solely rely on the ever increasing complexity of the treatment planning system's MU calculations, but rather ensure that two independent systems can calculate the same number of MU's to a specific point on a patient to deliver a specified dose.

In contrast, procedures utilized for patient identify verification as they enter into a radiation oncology clinic do not currently have the same extensive requirements of those applied to plan verification procedures. The 2017 National Patient Safety Goals only require a minimum of two patient identifiers when verifying identity and, typically, these simply involve the patient's name and date of birth[6]. As a minimal improvement on this, most clinics have also incorporated a photograph of the patient to be added to the electronic chart for increased verification accuracy. As technology advances and becomes more readily attainable, many new devices have become commercially available for use to increase the accuracy and reliability of patient identification and verification beyond these simple measures. Disposable RFID bracelets are often utilized in clinics and issued to patients in order to be scanned as a patient enters or exits various rooms throughout the facility [7,8]. Still other clinics have implemented more advanced identification devices such as palm scanners [9], iris scanners[10], or fingerprint scanners[11]. All of these procedures are implemented to help increase accuracy and speed of patient verification and are especially useful for clinics with a large patient load.

Facial recognition has become a novel approach to patient verification and identification. With this process, human errors can be eliminated when a patient answers to the wrong name or when photograph identification is not sufficient. Devices that implement facial recognition can be incorporated into the treatment planning system in order to include patient check in, automatic setup of couch, and patient verification as they enter the treatment vault [12,13]. These additional features to facial recognition processes can make patient verification and identification a quick and easy process that can facilitate movement and treatment of a patient through a radiation oncology clinic.

Patient Motion Tracking

In a typical treatment process, once identity has been confirmed and the patient is setup on the couch prior to treatment, setup images are taken utilizing either kV imaging or Cone Beam CT (CBCT) imaging procedures available onboard the linac. Alignment verification of the patient and internal anatomy is then performed by comparing the current images to those taken during Computed Tomography Simulation (CT-SIM). For any misalignments that cannot be easily adjusted by use of a couch shift, the radiation oncologist may need to be notified to decide if the patient can be treated with the newly shifted alignment. This process can take extended periods of time due to the fact that the oncologist not only needs to be available at the time to make the decision, but to also physically come down to the treatment console in order to confirm if treatment should continue. During this time, the patient is required to remain in the exact same position as for the CBCT acquisition. In order to quantify this extended period of time required by the patient, the latent time between CBCT and treatment was measured for 593 cases at the Karmanos Cancer Institute in Detroit, MI. It was found that this time period took, on average, 4 minutes, with a maximum value of 14 minutes (see Figure 1).

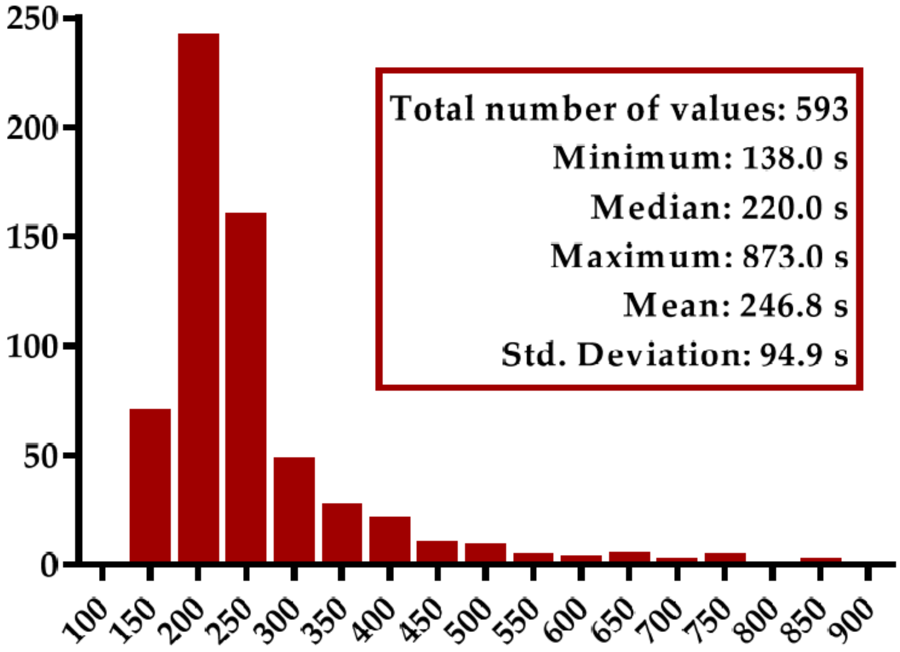


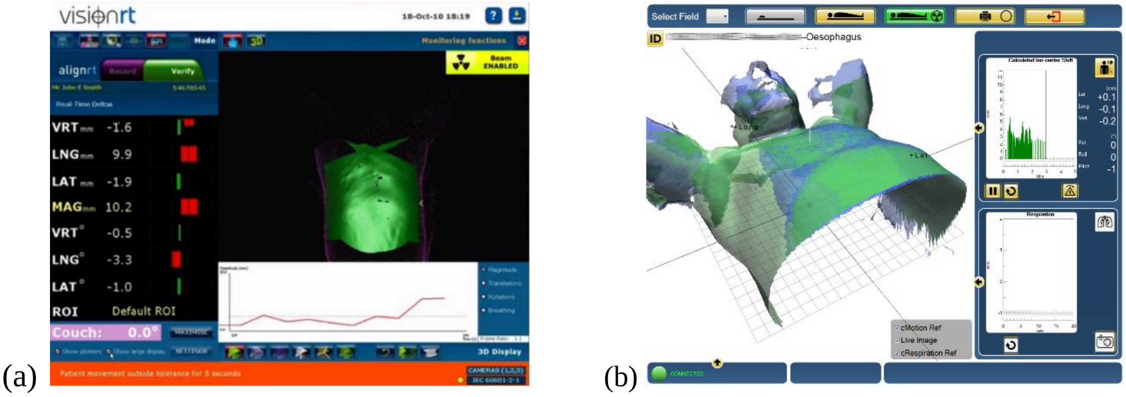
Figure 1: A frequency plot of latent time to treatment post CBCT

Depending on the specific type of radiotherapy treatment being performed, treatment times can range from 5 minutes to 15 minutes[14,15]. When adding together the gap between CBCT and treatment as well as the treatment time itself, this can require the patient to remain in the same position for a timeframe ranging anywhere from 5 minutes to 30 minutes. In addition to the extended period of time required to remain still, some treatments require the patient to remain in an unnatural position. Lung or breast treatments, for example, require the patient to raise one or both arms over their head while in a supine position. When holding this position for extended periods of time, patients may, unconsciously, adjust one arm or simply move an arm down to their side due to strain or discomfort. Doing so can drastically shift positioning of treatment areas compared to setup positioning. Correcting large movements such as this becomes difficult as the therapist must continuously monitor the patient through a CCTV camera and treatment must be paused to ensure proper patient positioning for the rest of the treatment.

While gross movements of major body parts may be visible through monitoring of the CCTV camera, smaller movements can occur that may not be as noticeable. Pain in a patient's hip while in the supine position due to the hard surface of the couch may cause the patient to shift their body positioning or a slight sinking of the patient's body into an alpha cradle can occur after the patient has begun to relax. In each case, these minor movements can cause an unexpected deformation of the body. For treatments with higher dose rates, a minor shift of a few mm due to these factors could cause a drastic dose deposition outside of the intended PTV. As the movements become smaller and more subtle, they becomes difficult, if not impossible, to identify when the only visual queue is coming from the CCTV camera.

The PTV itself is created during the planning process to accommodate the uncertainties of patient motions. The ICRU Report 50 and Report 62 originally defined the PTV to encompass the tumor volume (GTV), microscopic disease (CTV), as well as internal volume movement (ITV) and any movement of the patient that may occur during setup[16,17]. Although the PTV was originally created as a simple margin added to the volumes just described, its application has changed since its inception due to the increased precision and accuracy of radiation therapy procedures. The created PTV may now vary in size for different treatment locations in the body and for different treatment techniques. The expansion can vary from a few mm to over a cm and can be done so with margins set to different sizes in different anatomical directions[18, 19]. Much research and studies have been done to determine optimal PTV expansions based on different treatment sites and treatment modalities. Bell et al. found that, for post-prostatectomy IMRT treatments, the optimal PTV included an expansion of 0.5 cm in all directions except for the AP direction, which required a 1 cm expansion[20]. Chen et al. found that a 0.3 cm expansion in all directions was optimal for the PTV in cases of head and neck IMRT treatments[21]. Burgdorf et al. found that reducing the PTV for whole breast treatments from 0.7 cm to 0.5 cm, doses to the heart and lungs could be reduced while achieving a similar success rate[22]. As radiation prescriptions may incorporate the volume of the created PTV to ensure a specific percent of the volume receives a specific dose, it becomes paramount that extraneous movement be accounted for to ensure that the PTV does not move from the expected position.

To address this issue, various vendors have created devices to continuously monitor the patient in order to quantify any patient movement all while displaying the information in real-time. Vision RT created the AlignRT system which uses multiple IR cameras and sensors mounted around the patient to create a deformable, 3D representation of a user specified area on the patient being monitored[23]. It does so by projecting a known pseudo-random speckle pattern onto the patient and the system analyzes how that pattern is deformed compared to the known pattern. The use of multiple cameras allows the system to calculate and display various parameters of the patient’s current position compared to an initial position, including orthogonal movement and rotational movement (see Figure 2a). C-RAD created the Catalyst system which utilizes near UV wavelengths of light in order to create a similar 3D representation of the patient with real-time displays[24]. As with the AlignRT, the Catalyst utilizes multiple cameras to achieve deformable registration of the body as well as ascertain patient movement in multiple directions (see Figure 2b). With the ability to detect movement in the mm range, devices such as these are integrated into the linac software to not only give a visual indication to the therapist that the patient has moved outside an accepted threshold, but to also ensure that treatment can automatically be paused if those movement thresholds are reached. This visual feedback and ability to create thresholds for movement by the user ensures increased accuracy of treatment without the need to constantly monitor the CCTV cameras.



VRT mm -1.6
LNG mm 9.9
LAT mm -1.9
MAG mm 10.2
VRT° -0.5
LNG° -3.3
LAT° -1.0
ROI Default ROI
Couch: 0.0°

Patient movement outside tolerance for 5 seconds

Lat +0.1
Lng -0.1
Vrt -0.2
Rot 0
Roll 0
Pitch -1

Figure 2: Images of the user interface for (a) AlignRT and (b) Catalyst[23,24]. Both systems allow for threshold values to be set for movements and rotations of the patient.

Respiratory Motion Tracking

While tracking gross motion associated with the patient's body requires a high degree of spatial resolution to ensure accuracy in the millimeter range, motion associated with a patient's respiratory cycle requires a high degree of both spatial and temporal resolution. Due to its involuntary and cyclic nature, respiratory motion should be monitored and accounted for whenever treatments involve the thoracic or abdominal region. Depending on the type of breathing performed (shallow versus deep), studies have shown that organs such as the liver, kidney, or pancreas, move anywhere from 10mm up to 80mm in the superior-inferior (SI) direction during a respiratory cycle[25,26,27]. These large deviations from the expected position for these organs create unwanted uncertainties in the treatment planning process. Additionally, when focused on just tumor movement located within the lungs, positional variation can occur in all three orthogonal directions. Studies have shown that movement in the superior-inferior (SI) direction can vary drastically (from 0mm up to 30mm) with smaller, yet still significant, movements in the anterior-posterior (AP) and left-right (LR) directions (from 0mm up to 10mm)[28,29,30].

In 2006, AAPM Task Group 76 was created to address the difficulties of managing respiratory motion within a radiation oncology setting[31]. Aside from attempting to eliminate respiratory motion during treatments through the use of compression, breath-holds, and shallow breathing techniques, many different respiratory gating techniques are discussed. Tracking and recording of the respiratory cycle allows for a quantification and visualization of the respiration process by use of some external imaging device and is utilized during CT-SIM and, if tumor excursion is excessive, during actual radiotherapy treatments by gating the radiation beam during specific portions of the respiratory cycle. This former process, CT-SIM, is the beginning of every radiotherapy treatment planning as this is where the CT images are acquired for treatment planning and creates a baseline to which the patient is aligned during treatment.

When utilizing respiratory tracking with CT-SIM, a retrospective binning process is used for the images obtained to create a 4D-CT image set. Figure 3 gives an example of this binning process by which images obtained at specific points along the respiratory cycle are placed into corresponding bins. Ideally, the images within each bin will be exactly the same as each image represents the same position of internal structures at the same time in each respiratory cycle. The images for each bin are then averaged and the average for each bin is stitched together to create, essentially, a movie. This movie then represents the internal motion of the tumor and internal organs throughout one respiratory cycle [32]. It is at this point that the determination is made whether or not gaiting may be required during treatment based on the maximum movement of the tumor. Many centers consider the threshold of tumor movement to be 1.0 cm before gaiting is required for treatment and Liu et al. have shown that major component of tumor movement in the lungs was in the cranial-caudal direction[33]. With this threshold, it becomes important that any device used for measuring a respiratory trace have significant spatial and temporal resolution.

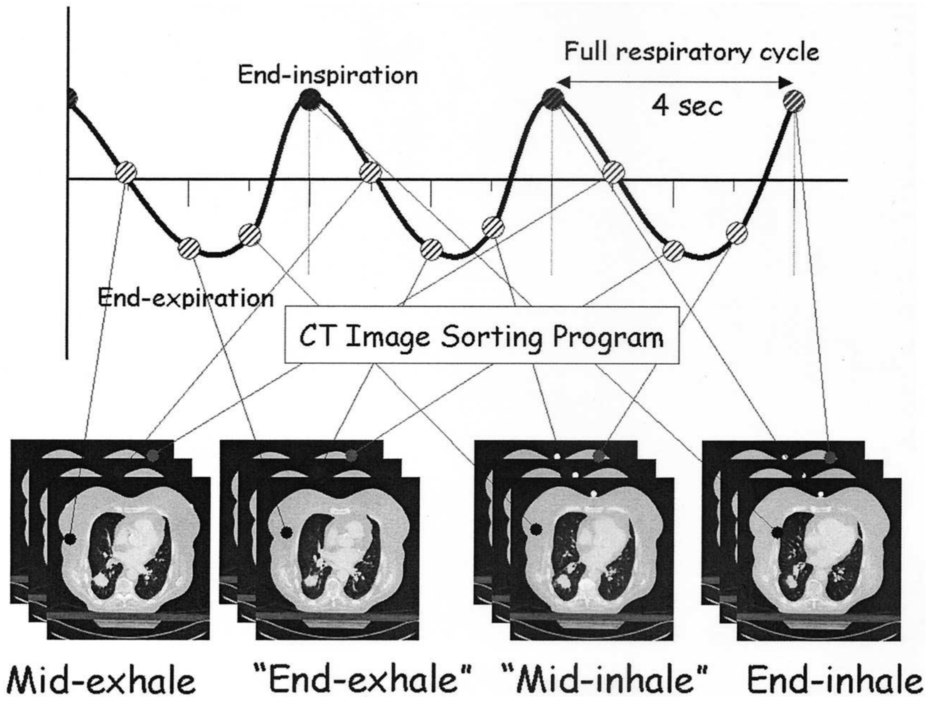


Figure 3: Example of binning CT images during respiration cycle[34]

To visualize and obtain a patient's respiratory cycle, be it for binning of CT images or gating radiation therapy, or both, an external imaging device is typically required. Current devices such as Vision RT's GateCT, Varian's RPM Respiratory Gating System, and the Anzai Gating system all utilize different processes to obtain the respiratory trace[23,35, 36, 37]. While some devices require physical apparatus to be in contact with the patient either by use of an external marker on the patient or pressure sensor belt attached to the patient, newer technologies are allowing the same trace to be obtained using a marker-less process. In either case, a cyclic, respiratory motion trace is created by the device's software and is incorporated into the treatment planning process for purposes of 4D-CT creation and, if warranted by tumor movement beyond the clinic threshold, assisting in gated radiation therapy treatments. To accomplish gating during radiation therapy, the device tracks the same respiratory motion as was done during CT-SIM, but now interacts with the linac in order to turn the beam on/off during specific portions of the patient's respiratory cycle[38]. This allows for radiation treatment to occur when the tumor is in the exact expected position within the breathing cycle based on the 4D-CT images created during CT-SIM.

Microsoft Kinect v2 Camera

It is the purpose of this manuscript to detail the processes by which the Microsoft Kinect v2 Camera can be implemented to solve the need of a radiation oncology clinic for patient verification, gross motion management, and respiratory motion tracking. Outlined in the following chapters are details on the hardware and software involved when working with the Kinect v2 (Chapter 2), as well as specifics on how the Kinect can be utilized for patient verification by way of facial recognition and recall (Chapter 3), to track both gross anatomical patient motion as well as smaller localized motion within a user drawn ROI (Chapter 4), and to generate respiratory traces of patients with accuracy comparable to currently available commercial hardware (Chapter 5). Creation and validation of these processes will help widespread incorporation of the Kinect as a multi-purpose device within a radiotherapy clinic.

CHAPTER 2 “MICROSOFT KINECT V2 SENSOR”

The Kinect v2 is multi-sensor camera/microphone system developed by Microsoft for the purposes of anatomical motion tracking. Developed and produced in 2014, the Kinect v2 is the successor to the original Kinect produced by Microsoft in 2010 (see Figure 4). Many of the specific differences and improvements upon the Kinect are listed in Table 1[39]. The largest improvement upon the Kinect involves the depth camera resolution which is based upon an updated measurement process. The original version of the Kinect implemented a Pattern Projection principle by which a known pseudo-random speckle pattern was projected by the IR projector onto objects within its field of view. The resulting IR pattern was then captured by the IR sensor and the system analyzed any deformation by comparing to the known pattern[40,41]. With a known distance between the IR sensor and IR projector, the disparity between the reference and live speckle pattern allows for a depth value calculation for each pixel within the frame.

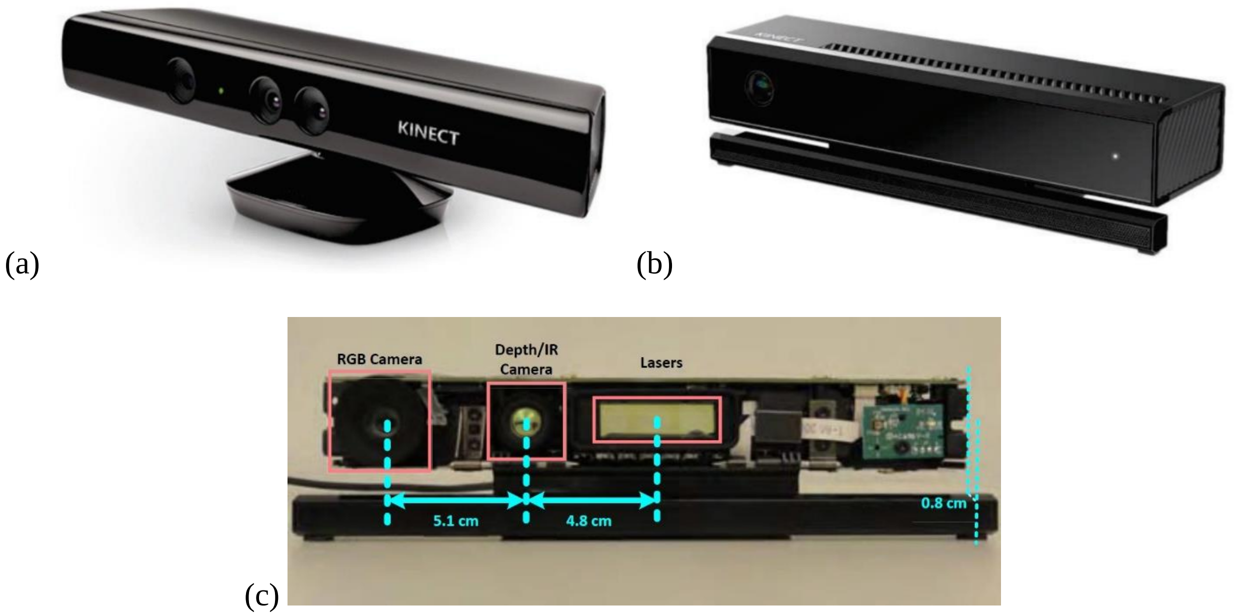


Figure 4: Images of original Kinect (a), Kinect v2 (b), and internal structure of Kinect v2 (c)[42]

Specification	Kinect v1	Kinect v2
Color Camera Resolution	640 x 480	1920 x 1080
Depth Camera Resolution	320 x 240	512 x 424
Depth Measurement Process	Pattern Projection	Time-Of-Flight
Horizontal FOV	57 degrees	70 degrees
Vertical FOV	43 degrees	60 degrees
Skeleton Joints Defined	20 Joints	25 Joints
Simultaneous Bodies Tracked	2 Bodies	6 Bodies

Table 1: Specification differences between Kinect v1 and Kinect v2

To more accurately calculate depth values for each pixel, the Kinect v2 IR sensor was upgraded to include a time-of-flight camera. Many commercial devices currently available for purchase include time-of-flight cameras for purposes of distance measurement due to the increased resolution and accuracy when compared to IR pattern projection processes[43-46]. With typical time-of-flight cameras, the IR projector emits photons of a highly specific IR frequency. As the light is reflected off an object and back to the IR sensor, a phase shift occurs which is then computed by the system and allows for depth values to be calculated for each pixel within the array (see Figure 5). The Kinect improves upon this basic process by emitting three specific IR frequencies (120 MHz, 80 MHz, and 16 MHz)[39]. The use of three different emitted frequencies was added as a tradeoff between depth measurement precision and maximum measurable range. A larger frequency (shorter wavelength) assists in increasing the resolution of the camera due to smaller possible phase shifts that could occur corresponding to smaller possible distances. A smaller frequency (longer wavelength) allows for much larger possible phase shifts to occur which are associated with measurements of larger maximal distances[47]. The combination of three different frequencies produced by the Kinect, allows for a unique middle ground that ensures precision of measurement out to a larger distance[48].

Additionally, the use of three different frequencies allows for a less computationally intensive process to solve the problem of phase wrapping that occurs for time-of-flight recorders. For a photon with a specific frequency/wavelength, the phase is typically given in radians from 0-2π and describes the cyclic nature of the propagating wave. When a photon is reflected, the wavelength of that reflected photon can be offset from the original photon depending on the distance traveled. This offset is the phase shift and can easily be measured by the system when the reflected photon interacts with the detector. However, if the reflected photon is shifted far enough, it can appear in phase again given that phase shift values can only occur between 0-2π. For example, a photon could appear to be $\pi/2$ out of phase, but this calculated shift would be the same if the reflected photon was $5\pi/2$, $9\pi/2$, or $13\pi/2$ out of phase given that each of these phase shifts are simply $\pi/2 + 2\pi * k$ where k is some integer multiple. The system that absorbs and records the reflected photon must have a way to distinguish the total phase shift of the reflected photon in order to determine the total distance traveled. Utilizing 3 different frequencies allows for a phase unwrapping that becomes a simple implementation into the hardware and remains accurate enough to overcome noise (see Figure 6)[49, 50]. This unwrapping occurs when the phase shifts of all three frequencies line up allowing for the actual distance traveled to become apparent. Including this process into the distance measurement procedure allows the Kinect to calculate distance values for each pixel in increments as small as 1mm while measuring out to a distance of 4.5m.

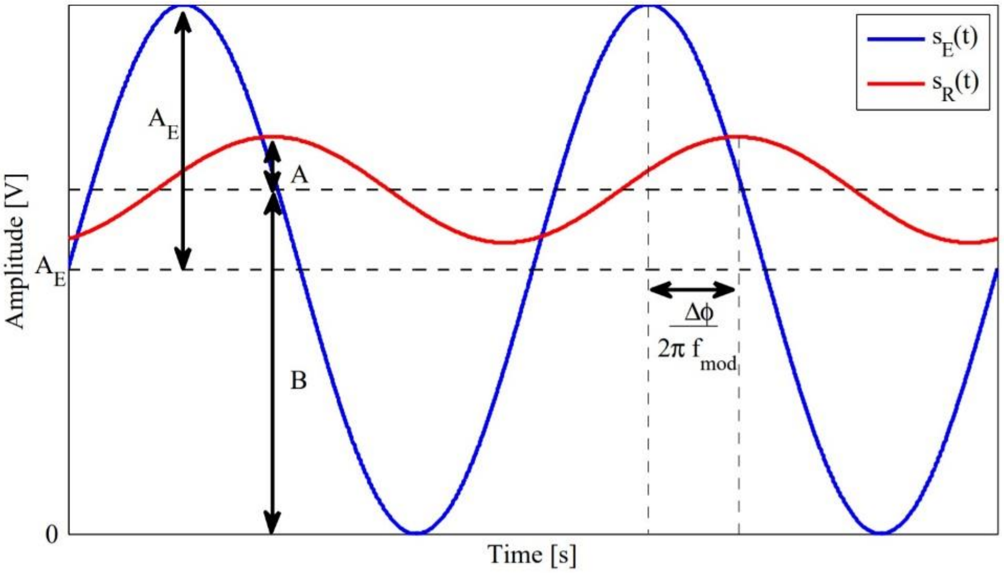


Figure 5: Example of specific IR frequency signal emitted (blue, $s_E(t)$) and phase-shifted IR signal received (red, $s_R(t)$)[51]

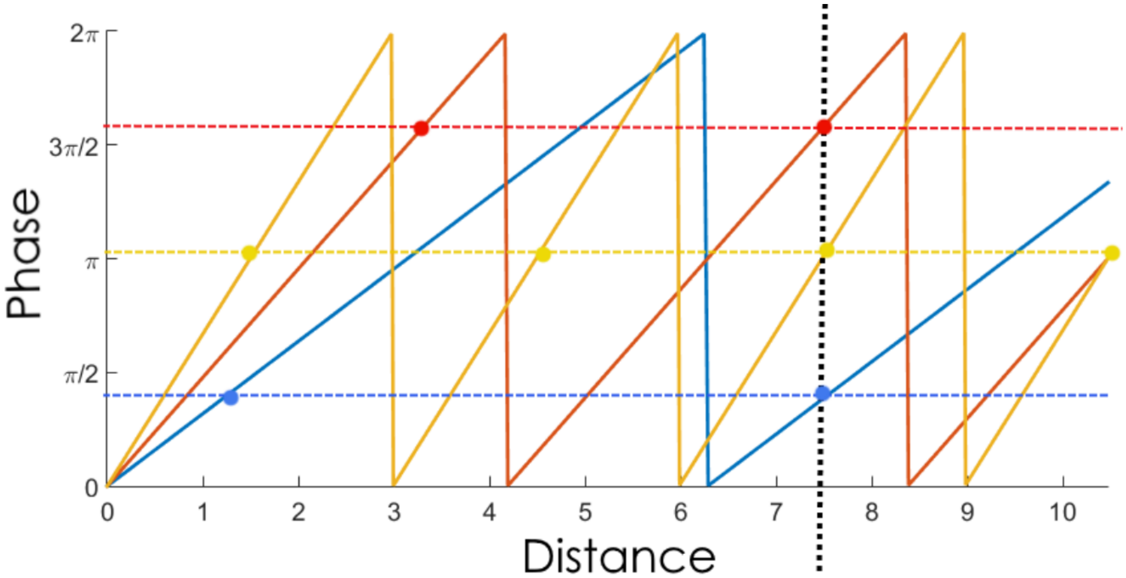


Figure 6: Visualization of phase unwrapping with multiple frequencies. Each colored line represents data from three different frequencies. The dots represent multiple measurements for each frequency. When comparing each frequency to each other, only one distance value, the correct distance, will properly unwrap the phases[50].

The IR sensor in the Kinect v2 is also unique in that it is made up of a differential pixel array. Here, each pixel contains two photodiodes which are timed to be turned on/off with the same clock signal that controls the IR emitter. The actual IR signals emitted are square waveforms and, as such, the clock signals are either in a high or low state. If each pixel has photodiodes A and B, when the clock is in a high state, the photons falling onto the pixel will contribute charge [a] to A. When the clock is in a low state, the photons falling onto the pixel will contribute charge [b] to B. According to Sell et al, this allows for extraction of useful properties from the image acquired[48,52]:

- [a] – [b] indicates phase information that is used to calculate the depth values after an arctangent calculation
- [a] + [b] indicates a "normal" grayscale image which is illuminated by ambient light ("ambient image")
- $\sqrt{\sum([a]-[b])^2}$ indicates the grayscale image that is independent of ambient light ("active image")

In order to easily access the data and imaging information generated, Microsoft has created the Kinect to function on a royalty-free development platform and has distributed a free software development kit (SDK) containing many sample applications to allow quick access to the various processes of the Kinect. These include accessing depth information, color images, body and skeletal joint tracking, as well as basic facial recognition[53, 54]. Interfacing with the Kinect is relatively simple and the hardware supports programming languages of C++, HTML, Java, as well as C#. Coding done throughout this dissertation was accomplished utilizing C# within Visual Studio. The sample applications provided within the SDK created a basis for accessing various features of the Kinect, but the projects mentioned in subsequent chapters expanded upon these samples to accomplish specific goals.

The programming framework utilized by the Kinect groups much of the information into classes for various “frames” Each class has a number of properties and sub-classes associated with it including frame descriptions (width, height, number of pixels in image, etc), time stamp for the frame, as well as the raw frame data acquired by the Kinect. Each of the “frames” created are associated with the different imaging hardware attached to the Kinect. The “DepthFrame” contains information associated with the depth calculations for each pixel within the IR sensor. The “ColorFrame” contains imaging information related to the high resolution color camera. The “BodyFrame” and “BodyIndexFrame” contains information specific to how many bodies are recognized by the Kinect, the tracked joints associated with each body, and which pixels are associated with each body recognized.

Accessing the information associated with these frames requires understanding of exactly how the data is formatted. The information contained within the DepthFrame is utilized extensively throughout this dissertation given its importance in tracking and imaging objects in 3D space. The raw depth data is simply the actual depth value registered to an object in front of the camera per pixel within the 512 x 424 frame. Instead of a 2D array of values, the depth values are contained within a 1D array of 217088 values, the total number of pixels associated with the 512 x 424 IR sensor. In this manner, isolating an area of pixels within the depth frame for a Region of Interest (ROI), requires knowledge of the pixel locations on the 512 x 424 grid. This can be accomplished in a two-step process. First, by utilizing the (X,Y) pixel location of the upper left corner ROI, the starting index within the 1D array of depth values can be located with the following line of code:

$$startIndex = (X + (Y * depthFrame.FrameDescription.Width)) - 1$$

Here, *startIndex* is the variable integer that will be saved as the starting index for the depth values required, and *depthFrame.FrameDescription.Width* is a call to obtain the width of the original DepthFrame.

The second step to parse out the specific depth values within an ROI requires creating a new 1D array of depth values by looping through the original array and incorporating only those pixels within the ROI:

$$for\ (int\ i = 0, i\ < ROIHeight, i++)$$

$$\{$$

$$Array.Copy(depthFrame,$$

$$(startIndex + i * depthFrame.FrameDescription.Width),$$

$$newDepthFrame, i * ROIWidth, ROIWidth);$$

$$\}$$

In this piece of code, the Array.Copy function will populate the array called “newDepthFrame” by looping through all data within the original “depthFrame.” It will do so according to the following algorithm:

1. Start at the index *startIndex* in the original *depthFrame* as i=0
2. Copy a *ROIWidth* number of values into *newDepthFrame* where *ROIWidth* is the pixel width of the ROI
3. Increase i by one

4. Start at the index *startIndex* + i**depthFrame.FrameDescription.Width* in *depthFrame* as this will be the starting point of the next row in the original 512 x 424 frame
5. Copy a *ROIWidth* number of values into *newDepthFrame* starting at the index of i**ROIWidth* of *newDepthFrame* where *ROIWidth* is the pixel width of the ROI
6. Repeat 3-5 until i < ROIHeight, where *ROIHeight* is the pixel height of the ROI.

Through this simple line of code, depth values within the ROI created and be extracted without the need to analyze the entire data set of depth values obtained by the Kinect.

The "BodyIndexFrame" is also a very useful frame to incorporate into this process as this allows for the created ROI to be cross referenced in order to eliminate pixels that are not associated with a specific body. The raw data within this frame is similar to the depth frame in that it consists of a 1D array of 217088 values (the total number of pixels in the 512 x 424 IR sensor). Each value in this array is either a 0 or 255 where a value of 0 indicates there is no body present in that specific pixel and a value of 255 indicates a body is present in that specific pixel. Given the similar structure of the BodyIndexFrame from the DepthFrame, the same algorithm used for the depth frame can be applied to retrieve information only applicable to a created ROI. Combining the two filtered arrays together allows the program to analyze information only applicable to data within the ROI and data only applicable to the body recognized by the Kinect within that same ROI.

The easy to access data generated by the Kinect and the open-sourced, multi-faceted availability facilitated by Microsoft with the available SDK has allowed Kinect developers to create an array of applications for research and commercial purposes:

- Real-time translation of sign language into spoken language and vice versa[55]

- Controlling robotics with hand and body movements[56,57]
- Manipulation of 3D models for biology, engineering, and geography education[58]

Apart from widespread commercial use, the Kinect has been heavily researched for applications within the medical community. Its ability to combine accurate depth data with high resolution color images have enable researchers and programmers to utilize this versatile devices for numerous medical research opportunities and applications:

- Touch-less interaction of computer systems within surgical suites[59]
- Assisting in muscle rehabilitation and stroke recovery[60,61]
- Tracking of head orientation during PET scans[62]
- Assist in early detection of autism in children[63,64]
- Monitor elderly patient's daily movements to analyze gait and potential falls[65]

Because imaging plays such a large role within radiation oncology, researchers within the field have identified many uses for the Kinect and much research has been completed on the Kinect since its inception:

- Automatic couch and positioning setup for radiotherapy sessions[66]
- 3D scanning of surface to allow for bolus creation from 3D printer[67]
- Assist in patient thickness estimations to improve x-ray imaging techniques[68]
- Respiratory motion tracking using physical markers[69,70]
- Real-time monitoring of patients during radiotherapy treatments[71,72]

With many different devices currently available that assist in real-time positioning or patient monitoring that work on similar principles, the Kinect offers a chance for smaller or more independent clinics to incorporate these advancements in technology and improve their radiation treatment techniques. In subsequent chapters, implementations of these various processes are presented and analyzed for their accuracy and feasibility.

CHAPTER 3 “IMPLEMENTATION OF FACIAL RECOGNITION WITH MICROSOFT KINECT V2 SENSOR FOR PATIENT VERIFICATION”

With radiotherapy procedures becoming more precise and complex, ensuring that errors in treatment do not occur becomes exceedingly important. The process of eliminating errors in treatment starts with patient verification. Although the Joint Commission has stated that patient identification within a hospital setting only requires two identifiers[6], typically given as a verbal recall of the patient’s name and date of birth, many clinics continue to improve upon this process by requiring additional identifiers. Some may simply add a patient photograph to the chart for visual identification, while others have installed palm or iris scanners outside of the treatment vault to scan a patient before entering [9,10,73,74]. In either case, additional verification procedures such as these can greatly benefit any radiation oncology clinic to ensure that patient errors are kept to a minimum.

To continue with the advancement of patient verification processes, the Kinect v2 was adapted to create a facial recognition process. The Kinect SDK was used to begin this process as it contains a facial mapping library that can be utilized to collect the coordinates of a vast number of facial feature locations in 3D space. It also includes a sample application for a basic implementation of this library. However, facial mapping is distinct from facial recognition, and methodology was created to use this mapping information in a manner that facilitates facial identification. The information presented in this chapter has been published within *Medical Physics* in 2017 but is also presented here as part of this dissertation[75].

For this process, we have used the facial mapping library to create a straightforward system for performing facial recognition and recall. The matching algorithm is based on the comparison of vector magnitudes between facial fiducial points in a pre-collected reference set to those collected through real-time sensor capture. The overall performance of the system has been benchmarked through a sensitivity and specificity analysis, and potential limitations of the system are analyzed with respect to varying light conditions between reference data collection and real-time matching to collected sensor data.

The Kinect v2 camera was utilized to image and capture facial details. The example application presented in the SDK utilizing the HDFaceMapping library was used as a basis for the data collection process. The flowchart given in Figure 7 displays an overview of the entire process. This begins with the system recognizing the closest body to the camera, and isolating the body on the screen, removing all background. The patient then interacts with the Kinect by way of a hand gesture recognition process to initiate the facial mapping. Then, the data collection process of facial features begins through user interactions, facilitated by on-screen instructions and chimes to indicate when each action is complete. The system then continues through the calculations used to specify the details of a specific face, and through the comparison algorithm used to identify whether or not a facial match has been made. Lastly, it concludes with interaction from the individual, via hand gesture recognition, to verify their identity to that of the match made within the database.

The initial facial contour mapping completed by the HDFaceMapping library does so by capturing 16 specific frames from the Kinect utilizing both the color and depth camera: 4 facing the camera, 4 with the head rotated 45° to the right, 4 with the head rotated 45° to the left, and 4 with the head pitched upward at 45°. Figure 8 displays the 4 specific poses needed by the user in order to complete the data collection process. Rotation, pitch, and yaw of the head are all calculated by the system and are incorporated into this process to determine the extent at which the head is still required to be moved.

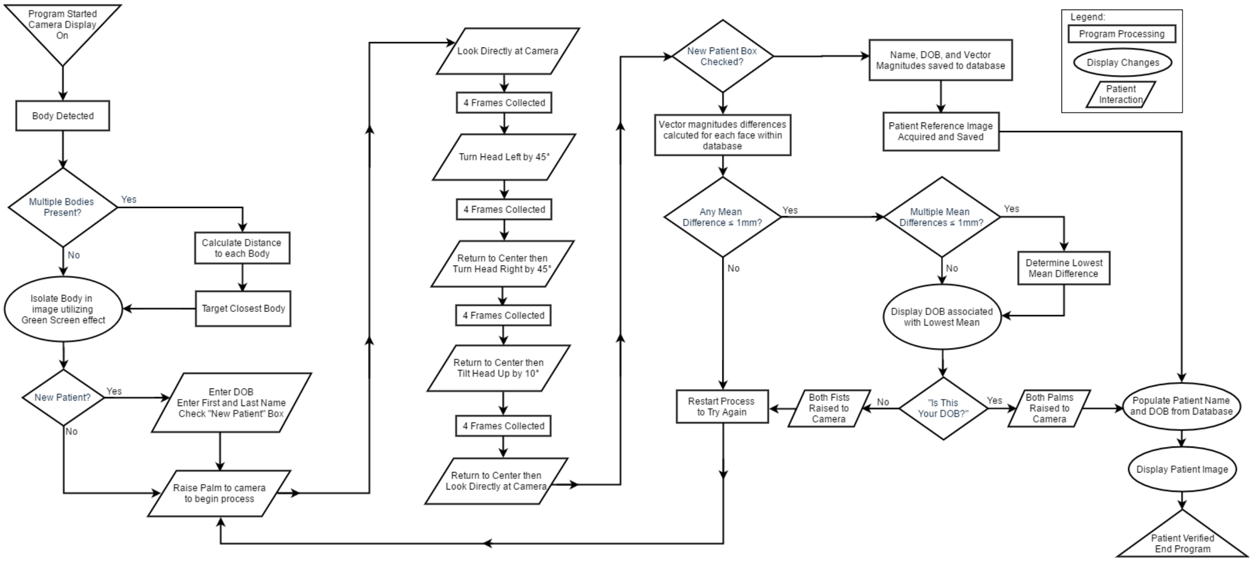


Figure 7: Flowchart illustrating patient interface and program processes from initiation to match confirmation. Body recognition and patient initiation is contained on the left of the flowchart and SDK data acquisition is contained within the middle of the flowchart. Algorithm, calculations, and database references are all part of the code written for this process and are displayed on the right side of the flowchart.

The facial contour created by the SDK from the 16 specific frames consists of 1347 facial points tracked in 3D space to specific locations on the individual's face as it moves in front of the camera. Of the points collected, 35 are specific to various facial landmarks (see Table 2 and Figure 8). Through initial analysis of the data extracted, it was found that 4 points were highly variable when capturing facial data from the same person through subsequent acquisitions. These 4 points were located on the corners of the mouth and center of the lower lip. As such, they were ignored during the analysis leaving a total of 31 points to be used for facial comparisons.

Midline	Bilateral	Removed
Chin	Cheekbone	Mouth – Left Corner
Forehead	Cheek – Center	Mouth – Right Corner
Mouth Upper Lip – Mid Bottom	Eyebrow – Center	Mouth Lower Lip – Mid Bottom
Mouth Upper Lip – Mid Top	Eyebrow – Inner	Mouth Lower Lip – Mid Top
Nose Bottom	Eyebrow – Outer	
Nose Tip	Eye – Inner Corner	
Nose Top	Eye – Mid Bottom	
	Eye – Mid Top	
	Outer Corner	
	Lower Jaw – End	
	Nose Bottom	
	Nose Top	

Table 2: List of facial location for the 35 extracted facial points including 4 points removed from the analysis due to their extreme variability within acquired data. The nomenclature refers to that used in the SDK

For any point, p, the position in 3D space with respect to the color camera lens is extracted from the system as p=(x, y, z). For any two points, p_1 and p_2, the absolute vector magnitude between those points can be calculated using of the following:

$$p_1 = (x_1, y_1, z_1) \quad p_2 = (x_2, y_2, z_2) \tag{1}$$

$$|\vec{v}_{12}| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{2}$$

To ensure these 3D points are being collected from the same reference position for each acquisition, the individual is instructed to look straight into the camera while the 3D positions of the 31 points are collected. From the 31 points extracted, 465 absolute vectors magnitudes are calculated and saved such that all 465 absolute vector magnitudes act as a representation of the original face.

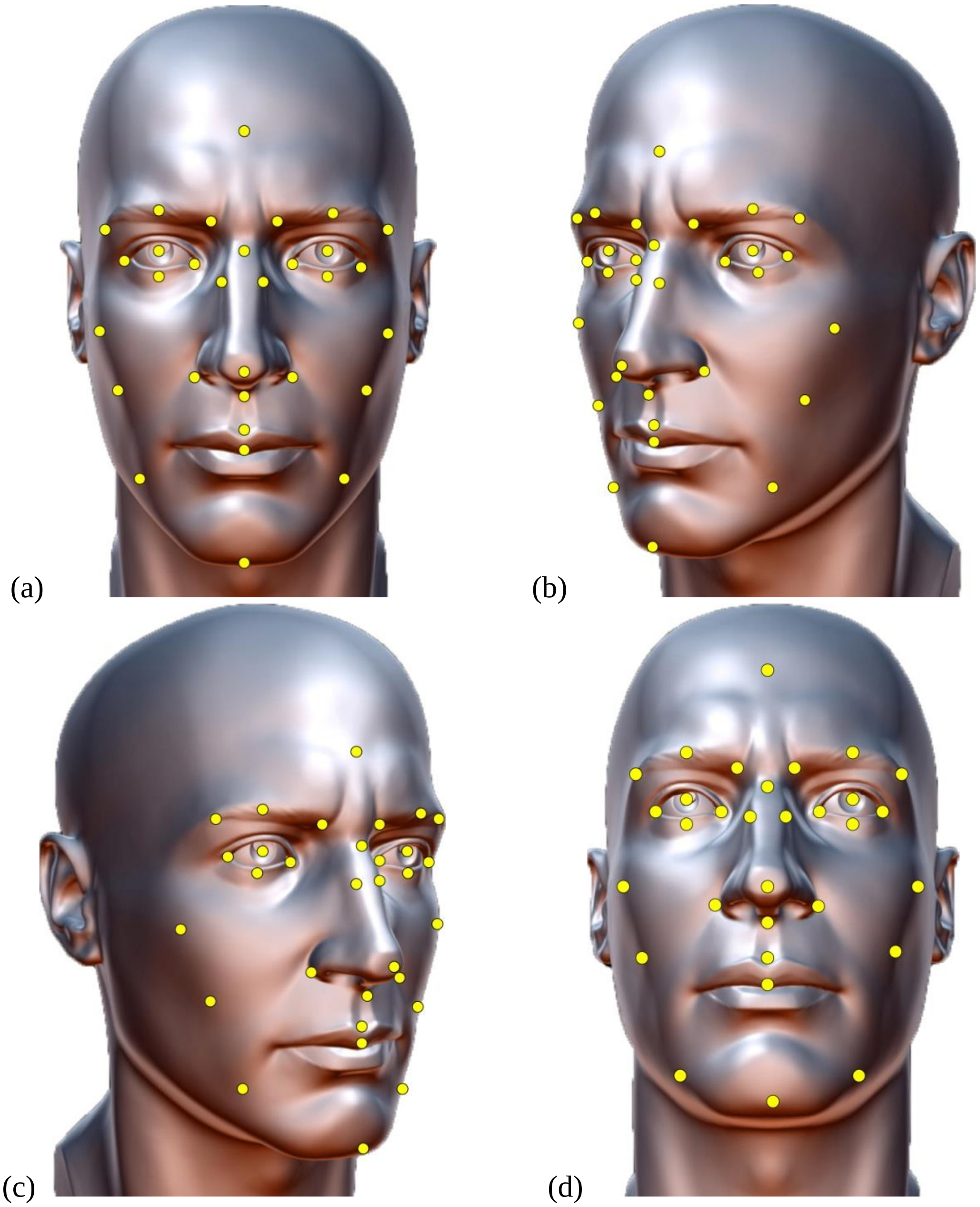


Figure 8: 4 poses required to complete the facial mapping procedure: (a) Straight ahead. (b) 45° rotation to the right. (c) 45° rotation to the left. (d) 45° pitch upwards. 31 Facial Points used in the analysis are also shown. All points remain on specific facial landmarks as the head moves throughout 3D space.

For each acquisition, the Kinect was mounted on a tripod set to a height of 160 cm (see Figure 9). To ensure data collection under identical conditions, optimal camera-to-individual distance was needed and was found to be dictated by two competing processes: body tracking requirements and depth sensor resolution. The Kinect software has the ability to detect when a human body has entered the frame of the camera. The software can then isolate and differentiate said body when compared to a non-human object such as a chair or animal. Once detected, the software will track the body while it is still within the frame. Isolation of the body within the frame can then be utilized to create a “green screen effect” such that all background is removed from the image and only the individual is displayed. However, the body detection and isolation can only occur if a significant portion of the individual’s body (head + torso) is visible within the frame. This “green screen effect” was implemented into the code and dictated the minimal camera-to-individual distance required.

The maximum camera-to-individual distance was dictated by the depth sensor resolution which was a function of the Kinect hardware. Specifications on the depth sensor state that it has the ability to detect distances ranging from 0.5m to 4.5m[76]. However, given that the absolute magnitudes being calculated are within the mm range and area represented by an individual pixel increases with distance, having the individual as close to the camera as possible allows increased accuracy while measuring such minute details. As such, a distance of 1m was chosen to maximize the percentage of a body within the frame while minimizing resolution loss of the depth sensor. This is incorporated into the display and will indicate to the individual how close or far away they are from the required distance of 1m.

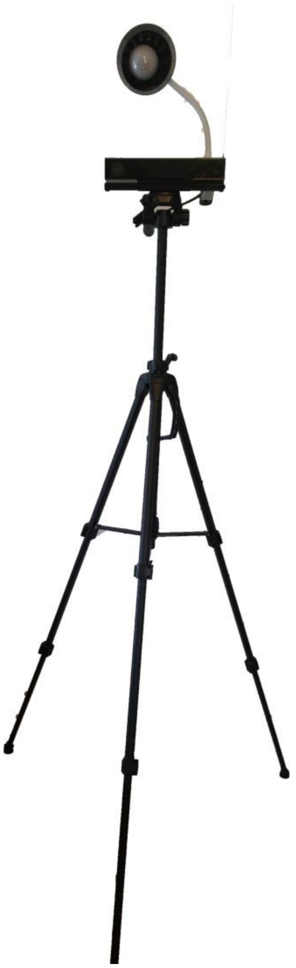


Figure 9: Kinect mounted on tripod set at height of 160 cm. Additional light source was added to the top of the Kinect allowing for consistent lighting during each acquisition.

Additionally, to ensure the process is as user friendly as possible, a hand gesture recognition feature was implemented with the interface. The Microsoft SDK contains sample codes to recognize hand gestures such as an open palm or a closed fist. The code itself is minimal and is easily implemented into any existing code. As such, the hand gesture recognition process was incorporated into the facial recognition process to allow for a tactile-free interface. To initiate facial recognition, the user can simply open their palm to the camera. Once the recognition process has been completed, the user can use an open palm to confirm identity or a closed fist to deny identity. In each case, the user is prompted through visual cues within the display to interact with the system in this manner.

As the facial mapping process is susceptible to varying levels of lighting, an LX1330B Digital Lux Meter was used to measure the ambient light 1m from the camera. Typical values ranged between 110 Lux to 220 Lux and were highly dependent on the room or hallway used during the acquisition. In an attempt to create identical lighting conditions for each location used, an additional light source was added to the tripod, above the Kinect camera, to ensure adequate and identical lighting (see Figure 9).

The recognition algorithm relies upon the calculation of the absolute magnitude difference between vectors acquired at a pre-collected reference session and those acquired at later verification sessions. Figure 10 plots sample data calculated for all 465 absolute magnitude differences between a sample correct facial match and a sample incorrect facial match. Visualizing the data in this manner allowed the identification of the mean, $\bar{x}$, and median, $\tilde{x}$, of these absolute magnitude differences to be tested as similarity scores for match/non-match determination due to the fact that a correct facial match appeared to have much lower absolute magnitude differences for the majority of the 465 vectors generated.

In order to adequately test the designed algorithm and determine specific threshold values for the mean and median that would indicate a facial match, a study was designed to acquire a facial database with as many faces as possible with as many repeated acquisitions as possible. This would ensure that the system could not only recognize the same individual over multiple sessions, but also be able to discriminate against many different individuals within the database. As such, a database of 39 different faces was acquired to begin the testing process. 37 of these individuals were compared to the database multiples times over multiple sessions as well as 12 additional individuals not in the database. This created a total of 115 specific acquisitions and allowed for 5299 individual comparison trials. For each acquisition, the 465 vector magnitudes obtained were stored in a separate database for further analysis.

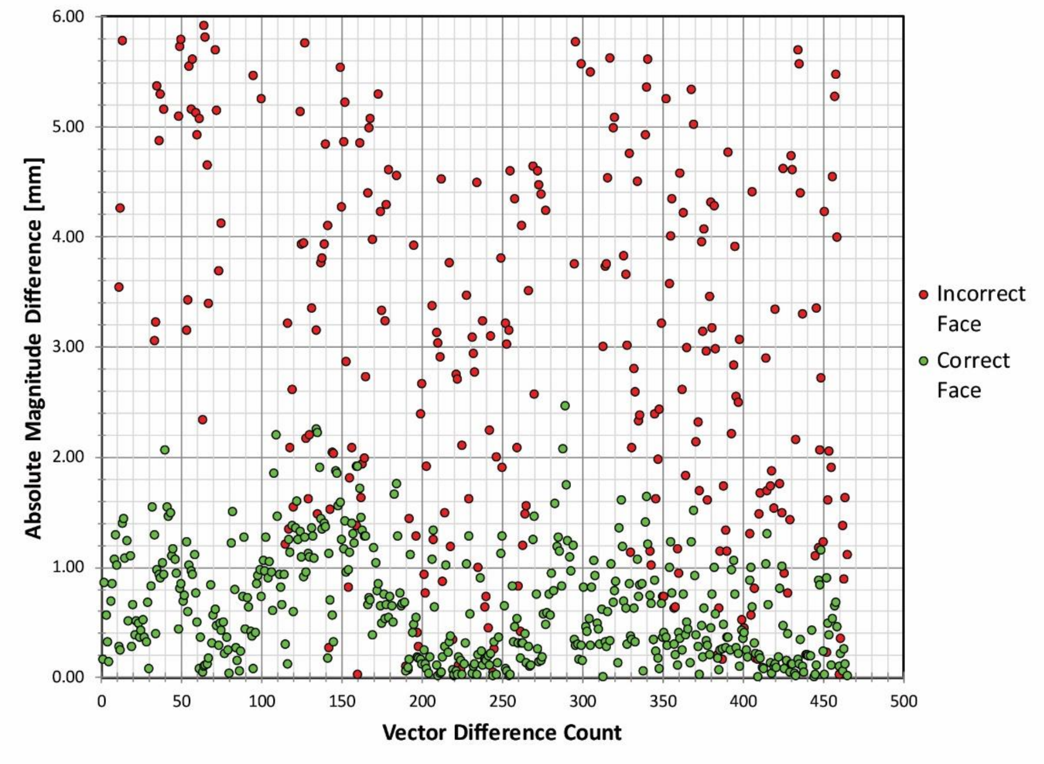


Figure 10: Graph of the Absolute Magnitude Differences calculated for a correct facial match and an incorrect facial match. Values for each absolute magnitude difference for a correct facial match (green) are much lower than that for an incorrect match (red).

Comparison of the data was done through a one-to-one matching process. This is typically a verification of identity in that the face being acquired is simply being checked against the existing face on file for that individual. In this case, the face acquired was individually compared to each face within the database to allow for True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values to be generated. As the mean and median were chosen for similarity scores, each parameter was calculated for the absolute magnitude differences calculated in each comparison. Figure 11 displays the distribution of data for each parameter given a correct match and an incorrect match. As the spread of data for a correct match was smaller and more narrowly peaked compared to the large spread of data for an incorrect match, both statistical parameters were used for match determination and calculation of the optimal threshold for each was then required. To accomplish this, Receiver Operator Characteristic (ROC) curves were constructed.

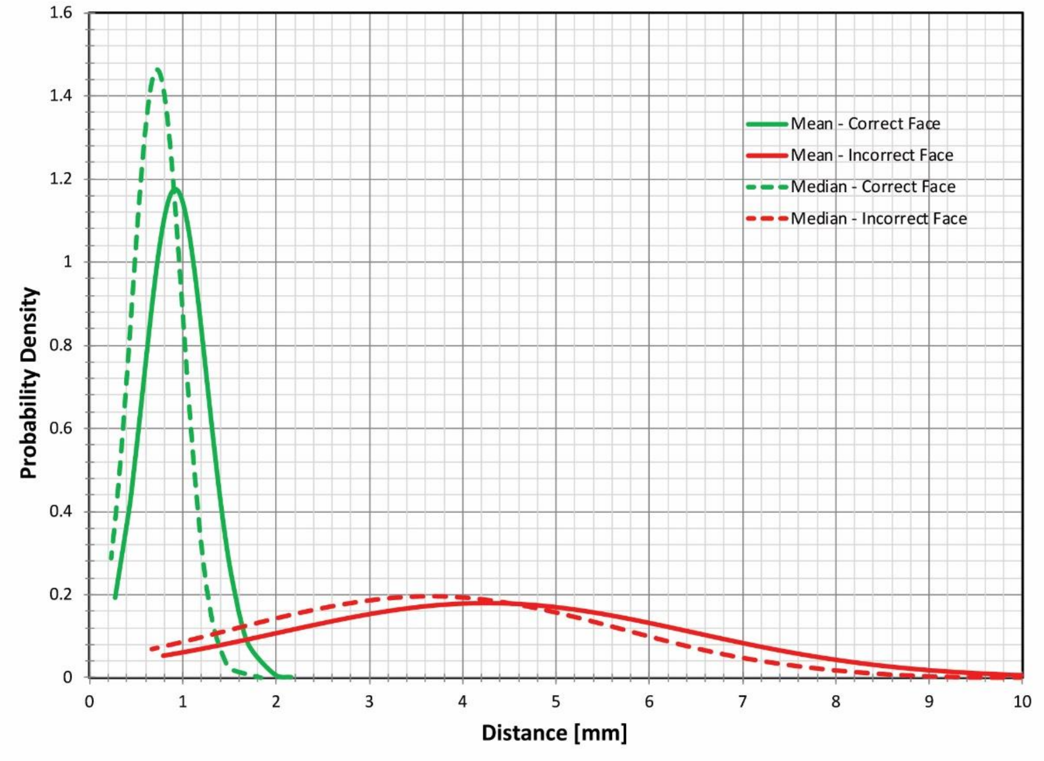


Figure 11: Plot of Similarity Score distributions calculated for both a correct match and incorrect match when mean and median of the absolute magnitude differences are calculated. The correct matches (green) show a highly peaked, lower valued mean and median when compared to an incorrect match (red) which has a broadly peaked, larger valued mean and median.

Although used for many diagnostic testing procedures, ROC curves are also used to determine the validity and accuracy of identity verification processes such as fingerprint identification scanners[77] or, as in this study, facial recognition. Identity verification typically involves layers of pattern matching to ensure valid True Positive and True Negative results. The ROC curves generated are based off of calculated values for Sensitivity and Specificity given by Equations 3 and 4. Both values are important in this study as sensitivity indicates the percentage of comparisons that will correctly identify an individual when comparing a live acquisition to a previous session (True Positive Rate). The specificity calculated indicates the percentage of comparisons correctly identified an incorrect match when comparing a live acquisition to all individuals within the database (True Negative Rate).

The ROC curves in this study were constructed by first calculating the mean and median of the absolute magnitude differences in each of the 5299 comparisons in this study. Once obtained, the threshold value of each parameter, which would ultimately determine whether a comparison was a match, was varied between 0.0007mm and 0.004mm. These threshold limits were chosen simply because moving past these values yielded no difference in the TP, FP, TN, and FN values. As each threshold value was changed in increments of 0.0001mm, the number of TP, FP, TN, and FN were counted with sensitivity and specificity values calculated from those counts as defined by Equations 3 and 4. Each pair of sensitivity and specificity values calculated were then plotted on a semi-log plot in order to generate the ROC curves.

$$Sensitivity = \frac{TP}{TP+FN} \tag{3}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4}$$

ROC curves generated by varying the match threshold for both similarity scores are shown in Figure 12. The area under the curve (AUC) calculated for each curve was ~99% indicating that both the mean and median are excellent parameters to differentiate between correct and incorrect matches.

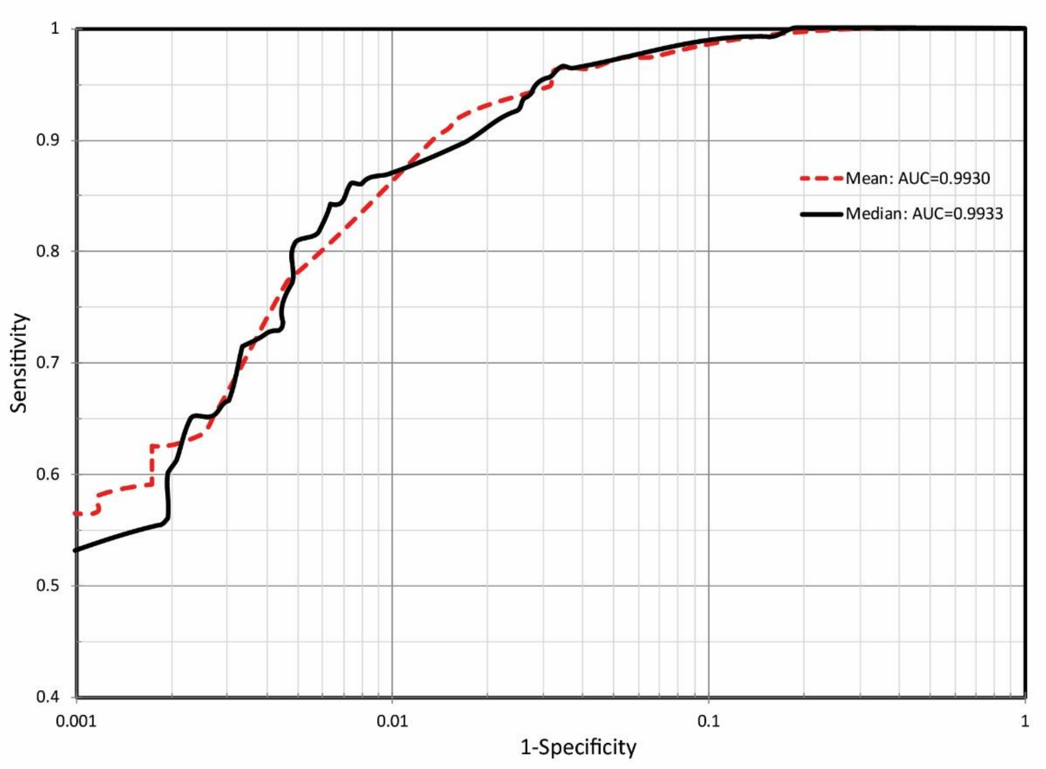


Figure 12: ROC curves generated from facial match determination by varying threshold of Mean or Median value required for a match. Logarithmic scale used to better visualize data.

Threshold optimization utilizing the ROC curves was calculated using both the Youden Index (largest vertical distance from the line of equality) and the smallest distance from the curve to the point (0,1)[78]. Both optimization criteria resulted in the threshold of 1.61mm for the mean and 1.27mm for the median (see Table 3). These optimization criteria are done without any reflection of the type of procedure being performed.

Parameter	Threshold	Sensitivity	Specificity
Mean	1.61 mm	96.52% (111/115)	96.76% (5016/5184)
Median	1.27 mm	96.52% (111/115)	96.68% (5012/5184)

Table 3: Summary of ROC curve analysis and threshold calculations

Although the threshold optimization criteria indicated specific values for the mean and median leading to excellent values for sensitivity and specificity (~96% for both), each of these parameters are the result of a percentage of False Positives or False Negatives and do not account for the actual number of False Positives/Negatives. Given the results from the threshold optimization, 170 false positives would have been encountered which is not acceptable for a real-world biometric verification process. Altering the threshold to more acceptable values can dramatically decrease the false positives while increasing the false negatives.

Lastly, it was found that ambient light plays an important role regarding the consistency of the acquisition process. In order to visualize and quantify this effect, two faces were acquired under varying levels of light, multiple times. The average of the vector magnitudes were taken for each light level and the absolute vector magnitude differences were calculated as done in the initial study. In this case, however, each face was only compared to acquisitions of the same face (i.e. Face 1 was only compared to Face 1 and Face 2 was only compared to Face 2). Mean and median values were determined for these differences and plotted as a function of light level to determine the correlation between decreasing light level and acquisition accuracy.

For this light intensity study, a face was imaged 5 times at 8 specific light levels. These levels varied from 60 lux to 285 lux. The 5 acquisitions taken at the highest light level, 285 lux, were averaged together and saved to a database in order to be used as a reference. Each acquisition taken at all 8 specific light levels was then compared to the reference facial data taken at 285 lux. Comparisons were performed exactly as before but with the average absolute magnitude differences calculated for each of the identical 465 vector magnitudes. The mean and median of those average differences for each light level was then determined. This acquisition study was performed on two different volunteers under identical conditions. The results of this analysis are displayed in Table 4 and Figure 13 and show a decrease in both the mean and median vector magnitude difference as light level is increased. To achieve the most accurate and consistent acquisitions, the ambient light around the face being acquired should be above 200 lux and should be consistent from one acquisition to the next.

	Face 1		**Face 2**	
Light Level	**Mean Magnitude Difference**	**Median Magnitude Difference**	**Mean Magnitude Difference**	**Median Magnitude Difference**
60 lux	0.812 mm	0.710 mm	1.420 mm	1.187 mm
80 lux	0.781 mm	0.657 mm	1.473 mm	1.249 mm
105 lux	0.687 mm	0.581 mm	1.452 mm	1.168 mm
135 lux	0.668 mm	0.578 mm	1.083 mm	0.935 mm
165 lux	0.643 mm	0.508 mm	0.943 mm	0.708 mm
185 lux	0.633 mm	0.566 mm	0.786 mm	0.571 mm
215 lux	0.589 mm	0.536 mm	0.707 mm	0.514 mm
285 lux	0.399 mm	0.348 mm	0.345 mm	0.328 mm

Table 4: Mean and Median values of the Absolute Magnitude Differences obtained and varying light levels for two volunteers.

Some practicality issues were encountered during this process that should be noted. Light glare from a person's glasses could cause the program to misread facial features and, thus, facial points. Scarves around a person's neck could cause some errors during the gesture recognition process and hair covering part or all of the forehead can cause the program to misread facial points around that area. All of these issues were overcome through appropriate visual instructions to the participant before data capture, although the need to issue such instructions does display some of the weakness of the technique.

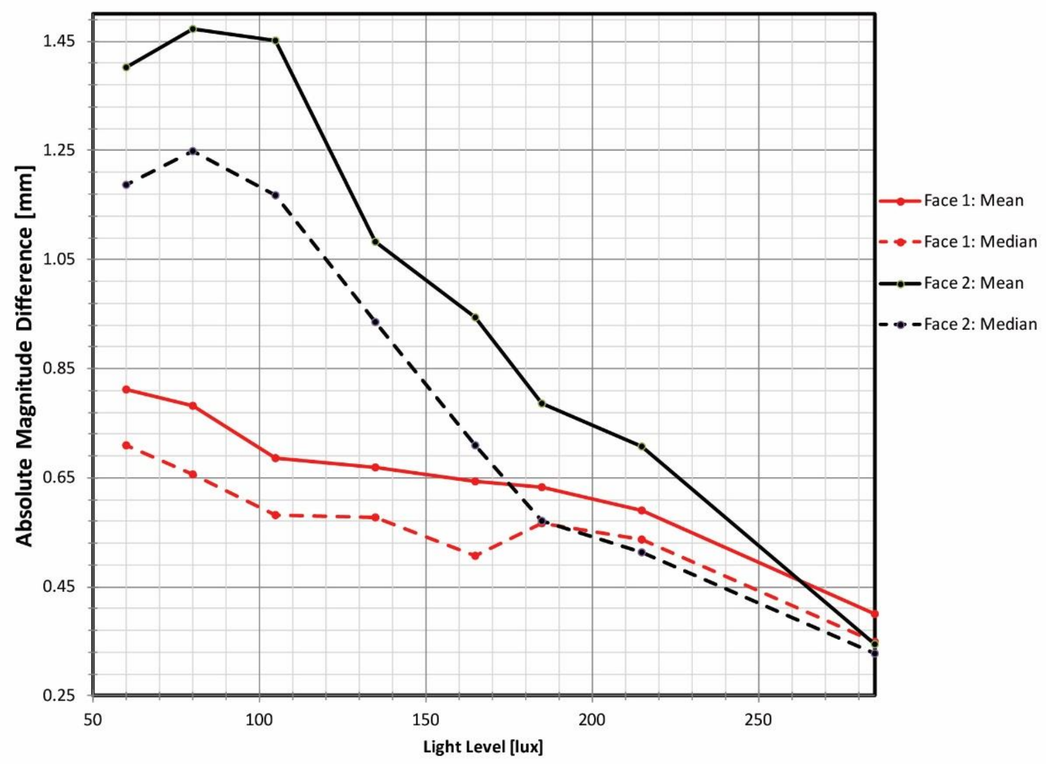


Figure 13: Mean and Median values of the Absolute Magnitude Differences graphed as a function of light level. As ambient light levels are increased, the mean and median of the absolute magnitude differences for the same face decrease indicating a more accurate acquisition at higher light levels.

Even though this facial recognition process has these limitations, the core technique has shown to be functional and has the ability to perform as intended. The unique properties of the Kinect enable a 3D facial points to be extracted and utilized in this process all while allowing for an interface that only requires hand gestures to operate. Forced interaction with a patient for verification may prove to be slightly cumbersome, but further improvements may increase the speed of the verification process allowing a more natural interaction between the patient and program.

CHAPTER 4 “AUTOMATIC MARKER-LESS PATIENT MOTION TRACKING UTILIZING THE MICROSOFT KINECT V2 SENSOR”

Verification of a patient’s identity enhances patient safety within the radiation oncology clinic by ensuring the correct patient is receiving the correct treatment. However, with said treatment’s ever increasing precision and complexity, delivery techniques also become more advanced. Thus, the next step to ensure patient safety and treatment efficacy is to monitor and quantify patient motion before and during radiation treatments.

Although a small amount of patient motion is accounted for within the setup margin of the PTV, gross patient motion can become a significant source of error. Treatment procedures such as SBRT and SRS may deliver high dose fractions or high dose-rate treatments where even small patient movement could result in not only an underdose to the target, but an excessive overdose to surrounding normal tissue[79]. As previously mentioned, for treatments to the lungs or breast, patients may be required to hold their arms over their head which can be an unnatural and uncomfortable position to hold. Patients may shift position or even move one or both hands down to their side, shifting the lungs or breast tissue away from the expected location causing a large portion of the dose to be delivered to the incorrect site.

To ensure little to no patient movement, intra-fraction motion should be monitored both during radiotherapy treatment as well as between CBCT alignment and actual treatment. Commercial devices such as the Align RT[80] and C-Rad Catalyst[24] are currently used for high-precision, real-time surface tracking but their implementation within a radiotherapy clinic may be outside the reach of some locations due to factors such as hardware integration or cost. In this chapter, the capabilities of the Microsoft Kinect v2 sensor are investigated in order to accomplish similar real-time surface tracking to quantify and monitor both gross and local patient motion to ensure that the initial setup of patient positioning remains throughout the entire treatment. The information presented here has been submitted for publication in 2017 within the *Journal of Applied Clinical Medical Physics* but is also presented as a part of this dissertation[81]

The Kinect was adapted to save an initial state of a patient's position and continuously compare the patient's current position to that initial state through the use of both the depth sensor and automatic skeletal tracking provided by the SDK. Previous studies with the Kinect have provided proof of concept for marker-less patient positioning setup and motion tracking of non-human shaped objects[41,66,71]. The research presented here builds upon these studies by employing real-time patient motion tracking and incorporating automatic skeletal tracking capabilities available with the Kinect sensor. With this, the Kinect can identify large anatomical movements as well as smaller, more subtle movements associated with the treatment area.

The data collection processes presented here focused on two separate collection methods. The first involved looking at the entire patient to detect gross motion through the use of 3D joint data collected by the Kinect's automatic skeletal tracking. Tracking the position of these joints in 3D space allows for a broad overview of the patient's movement while simultaneously allowing the user to track and focus on specific joints as needed. The second data collection method involved narrowing movement detection from the entire body or specific joints down to a Region of Interest (ROI) that could be selected by the user. For this second method, depth values of individual pixels within the ROI were tracked and an algorithm was derived to determine movement of the ROI based on pixel depth values within it. This allows for tracking in areas of the body where the treatment area is small and where the accuracy of the body joint tracking is not sufficient for the area of interest.

Both of the data collection methods require the use of the Kinect's depth sensor to obtain depth data of objects in front of the camera. As previously mentioned, the depth sensor resolution is 512 x 424 and has the ability to detect distances ranging from 0.5m to 4.5m. Depth data is returned for each pixel within the 512 x 424 frame in 1mm increments and can image at a rate of 30 frames per second[76]. Acquiring the depth data is relatively simple, requiring only a few lines of code, and the SDK produced by Microsoft has multiple applications and examples to access the information. The skeletal tracking is accomplished by the system generating 25 specific anatomical joints and tracking their relative position on the body. Each joint is tracked in 3D space with X, Y, and Z coordinates given relative to the camera. Specific details regarding each process and testing is listed in subsequent subsections.

As a general overview of the process, after the patient has been placed into the correct position on the couch for treatment and the start button is clicked, filters are used to smooth out the depth and joint data. 30 frames of depth and joint data (1 second worth of frames) are averaged together during this process. The initial segmentation mask, composed of an average of 30 frames of both joint and depth data, is saved within the program to be used as reference for movement. Once the segmentation mask has been compiled, the system will continuously filter the live stream of depth and joint data for comparison. For the depth data, a running average of 30 frames is used and each pixel within the running average is compared to the corresponding pixel within the segmentation mask and a depth difference for each pixel is calculated. The depth difference calculated for each pixel could be positive or negative depending on whether the area represented by the pixel has moved toward or away from the camera. For the joint data, a similar process is done by which the data is averaged over 30 frames and the distance between the initial and current positions of each joint are calculated using the X, Y, and Z joint data generated by the Kinect.

Skeletal Tracking

The incorporation of depth data to begin body and skeletal tracking is done automatically through the BodyBasics and BodyIndexBasics libraries within the SDK. The code itself allows the system to identify and isolate a body that is present in front of the camera and generate an approximation of a simple skeletal structure by the use of 25 specific anatomical joints as listed in Table 5 and displayed in Figure 14. Although approximate in their actual anatomical location on the body, each joint is locked to relative positions on the tracked body itself.

The exact process by which these joints are created is not divulged by Microsoft. However, the body tracking functionality, and thus, skeletal tracking, is known to incorporate body contours and high depth gradients near the edges of the recognized body by the depth sensor[82]. Each joint's position is given in 3D space allowing for the tracking of joints to their relative distance and position from the camera. As the joint recognition process is typically used on a person standing in front of the camera, it was found that, when in the supine position, some joints were not as static as others with regards to their placement on the body. Through this study, it was found that 9 specific joints were more stable than the rest and, as such, were the only joints tracked for their validity (see Table 5, Figure 14, and Figure 15).

Joints Used	Joints Ignored
Spine-Shoulder	Head
Spine-Mid	Neck
Spine-Base	Wrist - Left and Right
Shoulder Left	Hand – Left and Right
Shoulder Right	Hand Tip – Left and Right
Elbow Left	Thumb – Left and Right
Elbow Right	Knee – Left and Right
Hip Left	Ankle – Left and Right
Hip Right	Foot – Left and Right

Table 5: List of the 25 skeletal joint locations created by SDK. The nomenclature refers to that used in the SDK. The second column lists the 16 joints not used in this study due to their inconsistent positioning while a patient is in the supine position.

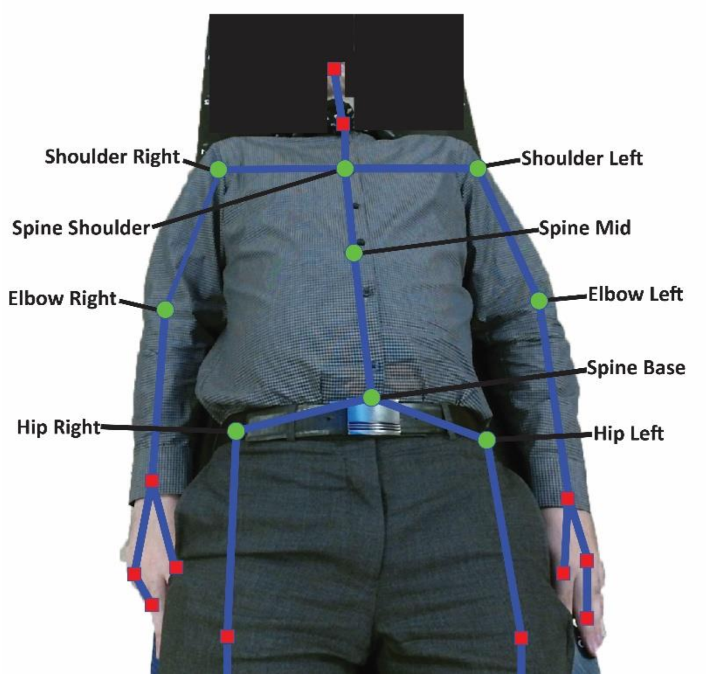

[redacted]

Figure 14: 25 Skeletal Points created by the Kinect when tracking a body. Square joints are the 17 joints not used in this study due to their inconsistent placement. Circle joints consist of the 9 joints chosen to be tracked in this study with the SDK nomenclature identified.

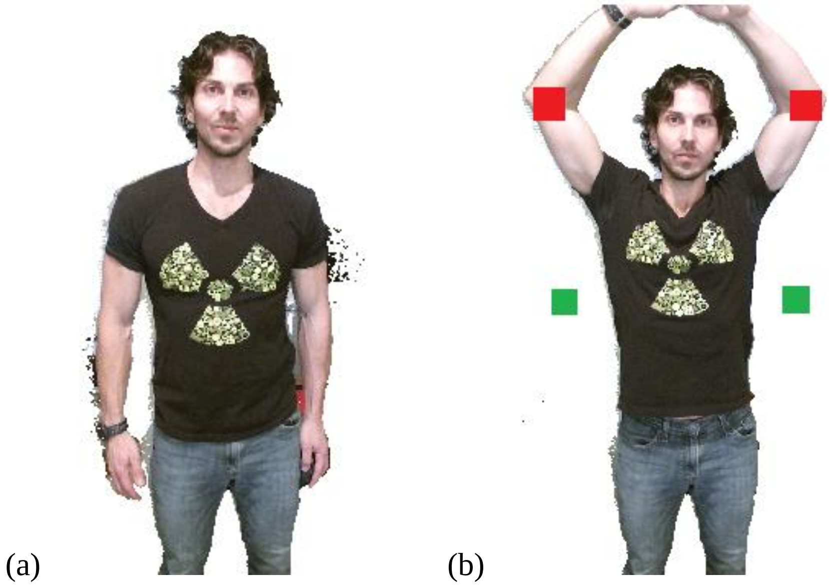


Figure 15: Example of image displayed while tracking joint movement of LeftElbow and RightElbow: (a) initial position and (b) movement of both arms. Green squares idicate original position of both elbows and Red squares indicate current position of tracked joints that have been moved beyond threshold value.

Smoothness and stability of the joint data was accomplished by use of the Holt-Winters Double Exponential Smoothing filter[83]. This specific filter removes small fluctuations in the joint data while the entire skeleton is tracked and calculated by the system [82]. Sample code for this smoothing filter was also provided by Microsoft and is available on their website [84]. The double exponential filter has the ability to incorporate trends present in data as well as configurability of certain parameters in order to better suit specific types of data. This filter is governed by Equations 5 and 6[85]:

$$Trend\colon b_n = \gamma\left(\hat{X}_n - \hat{X}_{n-1}\right) + (1-\gamma)b_{n-1} \qquad 0 < \gamma < 1 \qquad (5)$$

$$Filter\ Output\colon \hat{X}_n = \alpha X_n + (1-\alpha)(\hat{X}_{n-1} + b_{n-1}) \qquad 0 < \alpha < 1 \qquad (6)$$

Here, γ is the trend smoothing factor that allows configurability of the filter to be slower (values closer to 0) or faster (values closer to 1) in correcting towards the raw data. α is the data smoothing factor that allows for a smoother trend with increased latency (values closer to 1) or a less smooth trend with decreased latency (values closer to 0). When tracking and filtering joint data, 0.25 was given as the optimal value for both γ and α. The trend value (b_n) is essentially a smoothed difference of two concurrent filtered values ($\hat{X}_n$and $\hat{X}_{n-1}$) while the current filtered value ($\hat{X}_n$) is an adjustment based on the previous trend value and previous filtered value[17].

Lastly, given that the 3D coordinates measured by the Kinect are utilizing orthogonal directions with respect to the Kinect face, any rotation of the Kinect requires rotation of the joint coordinates measured by the system. With the Kinect mounted directly over the patient, rotation occurs about the X axis (medial-lateral direction). As such, the X coordinate is unaffected but the Y (superior-inferior) and Z (anterior-posterior) coordinates are modified by the standard coordinate rotation equations for rotations about the X axis:

$$y(t)' = y(t)cos\theta - z(t)sin\theta \qquad (7)$$

$$z(t)' = y(t)sin\theta + z(t)cos\theta \qquad (8)$$

Where y(t) and z(t) are the y and z coordinates of the joint at time t, and θ is the angle of the Kinect face to the couch. Modifying the joint coordinates generated by the Kinect with these equations allows the new coordinates to be within same coordinate system as movement in the plane of the couch.

In order to verify the accuracy of the skeletal joint data, a volunteer was imaged while lying on a PerfectPitch 6DoF couch. This couch allows for highly accurate movements, down to 1/10mm, to ensure joint movement picked up by the Kinect can be correlated to exacting movement of the couch. The couch was moved in increments of 2mm along each orthogonal direction (X, Y, and Z) with recordings of the X, Y, and Z joint positions taken at each interval ranging from -10mm to 10mm. Additionally, the couch was moved to radial distances of 5mm, 10mm, 15mm, and 20mm with varying values of X, Y, and Z coordinates. To calculate the radial distance from the 3D coordinates of a joint, the 3D vector magnitude is calculated using the following:

$$p_1 = (x_1, y_1, z_1) \quad p_2 = (x_2, y_2, z_2) \tag{9}$$

$$|\vec{v}_{12}| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{10}$$

Where p_1 is the live 3D position of the joint and p_2 is the initial 3D position of the joint. 3D coordinates and radial distances recorded by the Kinect were then compared to actual couch movements in order to quantitatively assess the accuracy of the joint positions.

A plot of the joint data recorded for the "Elbow Left" joint is shown in Figure 16. Although a general linear correlation can be seen between the X, and Z movement recorded for these joints and the actual couch movement, the linearity was not exactly 1:1 (i.e. 5mm of couch movement was not necessarily recorded as 5mm by the Kinect). The Y coordinate was particularly poor when recording movement values and continuously under reported the movement in that direction.

To ensure that the angle of the Kinect or unforeseen rotations were not interfering with the results, the radial distance moved was then calculated given that the calculation of the 3D vector magnitude is rotation invariant. Table 6 displays the average radial distances calculated for radial movements on the couch of 5mm, 10mm, 15mm and 20mm and indicate that some joints are more sensitive to movement than others. Additionally, Figure 17 displays the Coefficient of Variation (COV) attached to each joint for those same radial couch movements.

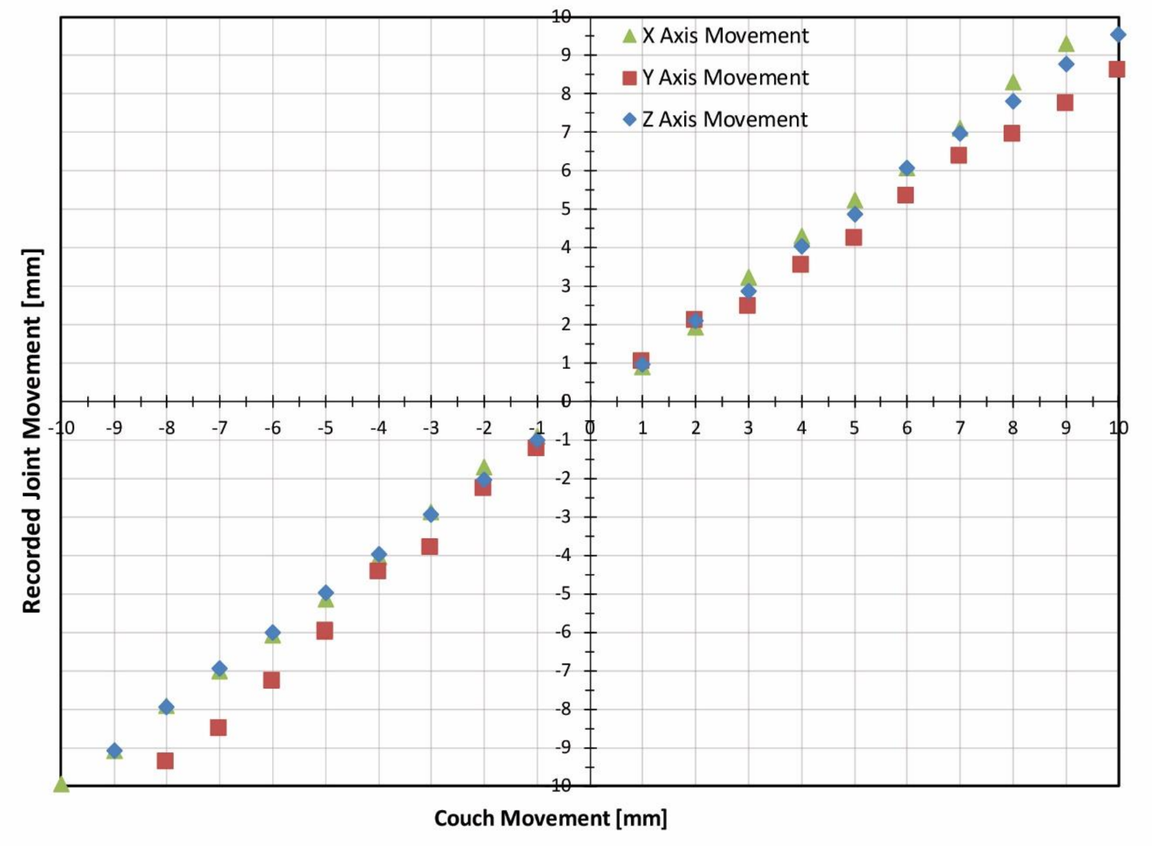


Figure 16: Plots of recorded movement for the ElbowLeft joint. Couch movement performed was done so in three orthogonal directions (X, Y, and Z) from -10mm to +10mm.

Joint	Average Radial Distance Measured By Kinect [mm]			
	R_c=5mm	R_c =10mm	R_c =15mm	R_c =20mm
Spine-Shoulder	3.22 ± 1.07	7.60 ± 2.49	12.14 ± 6.60	18.25 ± 7.58
Spine-Mid	4.12 ± 0.64	6.67 ± 2.30	11.43 ± 4.77	16.84 ± 4.72
Shoulder Left	2.52 ± 0.55	5.37 ± 2.88	9.04 ± 7.82	16.20 ± 6.61
Shoulder Right	3.73 ± 1.08	5.78 ± 2.49	8.96 ± 3.14	18.14 ± 9.28
Spine-Base	5.27 ± 1.35	9.41 ± 2.45	13.45 ± 3.74	20.38 ± 8.71
Elbow Left	5.10 ± 0.82	9.28 ± 1.40	16.43 ± 2.56	17.61 ± 2.10
Elbow Right	4.82 ± 1.06	10.25 ± 2.48	16.33 ± 1.86	18.18 ± 3.92
Hip Left	5.03 ± 1.06	8.60 ± 2.90	12.33 ± 3.20	14.79 ± 3.13
Hip Right	5.89 ± 1.76	10.20 ± 2.20	14.40 ± 4.50	18.26 ± 6.11

Table 6: Average radial distance and standard deviation measured for each joint measured by the Kinect. R_c indicates the radial distance moved by the couch and was done so at distances of 5mm, 10mm, 15mm, and 20mm for varying values of X, Y, and Z.

The Left and Right Elbow joints proved to be the most sensitive with calculated radial distances consistently closer to the couch radial movement and COV values consistently lower than other joints, typically between 15%-20%. These specific joints can prove to be most useful for tracking gross motion during any treatment requiring the patient to hold their arms in an overhead position, as the Kinect would be able to track and identify which elbow and arm may be involved with movement and alert the user to stop treatment.

Additionally, the Left and Right Hip joints, as well as the Spine-Base joint, were also sensitive to movement, similar to that of the elbow joints for 5mm and 10mm. However, COV values increased to 30%-40% for radial movements of 15mm and 20mm. Overall, as couch radial distance was increased, the variability of data obtained did appear to increase with larger standard deviations and COV calculated for many of the joints as movements increased to 20mm. As such, the tracking of joint movement for these joints would be most reliable for radial movements between 5mm and 10mm.

For these 5 joints (Left/Right Elbow, Left/Right Hip, and Spine-Base), threshold values can be created based on their average distance and standard deviation calculated. For 5mm of movement, it would be reasonable to create a threshold of calculated radial distance of 3mm to assume 5mm of radial movement based on an average calculated distance of ~5mm and average standard deviation of ~1mm. Similarly, for 10mm of radial movement, it would be reasonable to create a threshold of calculated radial distance of 7.5mm to assume 10mm of radial movement, based on an average calculated distance of ~10mm and average standard deviation of ~2.5mm.

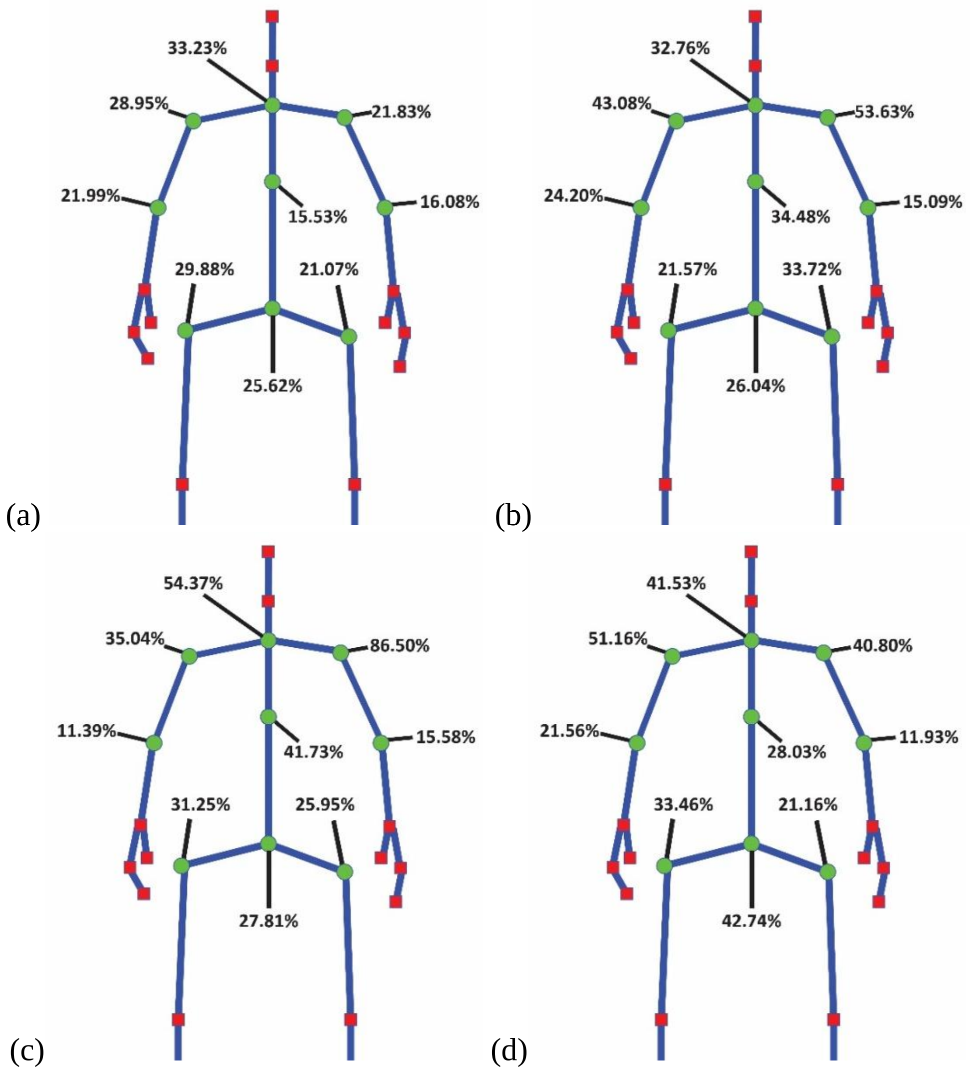


Figure 17: Coefficient of Variation (COV) calculated for each joint when the couch was moved radially (a) 5mm, (b) 10mm, (c) 15mm, and (d) 20mm. COV is defined as the Standard Deviation divided by the Mean (σ/μ).

The Left Shoulder, Right Shoulder, Spine-Mid, and Spine-Shoulder joints had calculated radial distances consistently lower than the couch radial movement and more variable. As such, they may not be viable for tracking movement. Given that joint creation by the Kinect is based off of depth contours and gradients, it may be the case that these joints are not as easily tracked by the system given that higher gradients do not occur in these regions.

When utilizing the skeletal tracking, distance from the camera plays a large role in accurate data collection. The depth sensor is the primary method by which the Kinect generates the skeletal joints through recognition of body contours and edges. When the body is too close to the camera (~ 0.5m – 1m), the skeletal tracking cannot keep a continuous lock onto the patient as intended with joints varying wildly in their position. In this study, it was found that a camera-patient distance is between 1m and 2m was optimal for allowing for appropriate skeletal tracking. Additionally, there are small errors with the depth values generated that should be noted. Yang et al. found that, in the optimal range of 1-3m directly in front of the camera face, average errors were < 2mm[42]. Beyond that range, the errors in depth can become larger, ranging from 2-4mm. As such, ensuring that the camera setup includes an optimal distance to the patient will allow for the depth information to be as accurate as possible.

Depth Data from Region of Interest

Secondary to the skeletal joint tracking, the Kinect v2 sensor's depth camera was utilized to obtain raw depth information for a user specified region of interest (ROI). One of the advantages of incorporating the body tracking capability of the Kinect is its ability to differentiate between pixels associated with a body vs. pixels belonging to the background. Once the body is recognized by the Kinect, the background can be removed from the image displayed allowing for an easy visualization of the patient through the GUI created. Once displayed and isolated on the screen, the user can then select a region of interest by simply drawing a rectangle on a specific area of the body. As different radiotherapy procedures may require tracking of different anatomical locations based on treatment area, the creation of a region of interest enables the user to track important anatomical locations without the need to quantify motion throughout the rest of the patient's body.

In order to reduce noise as much as possible and to ensure stable information for the depth data, a median filtration algorithm was implemented for data obtained within a specific frame. Typical noise from the depth sensor registers as either 0 for the depth or a value much greater than an expected depth. To account for this, a 7 x 7 grid of pixels is created around the pixel containing bad data. Depth values from all 48 surrounding pixels are analyzed and the median of those pixels is used as the correct value for the center pixel. This process enables correction of the bad pixel data without being effected by any outliers within the 7x7 grid. Figure 18 gives a visualization of this process.

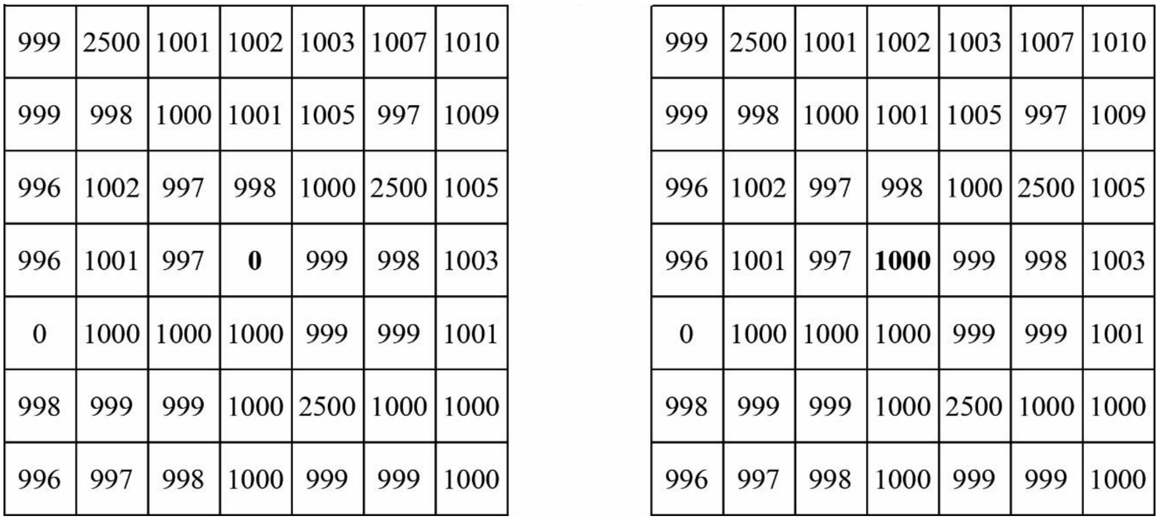
To quantify gross movement within this ROI, a threshold value of depth difference is needed. This threshold is used by the system to make the determination as to whether or not any specific pixel has a depth value that has changed significantly from the corresponding pixel in the segmentation mask. A visualization of this is displayed on the screen (Figure 19) with pixels within the ROI colored green, blue, or red depending on the depth difference compared to the threshold:

$$Threshold > D_i - D_t \qquad \text{Pixel Color} = \text{Green} \tag{11}$$

$$D_i - D_t > Threshold \qquad \text{Pixel Color} = \text{Red} \tag{12}$$

$$D_i - D_t < -Threshold \qquad \text{Pixel Color} = \text{Blue} \tag{13}$$

Where D_i is the depth value of the specific pixel within the segmentation mask and D_t is the depth value of the same pixel in the current frame.



999	2500	1001	1002	1003	1007	1010
999	998	1000	1001	1005	997	1009
996	1002	997	998	1000	2500	1005
996	1001	997	**0**	999	998	1003
0	1000	1000	1000	999	999	1001
998	999	999	1000	2500	1000	1000
996	997	998	1000	999	999	1000

999	2500	1001	1002	1003	1007	1010
999	998	1000	1001	1005	997	1009
996	1002	997	998	1000	2500	1005
996	1001	997	**1000**	999	998	1003
0	1000	1000	1000	999	999	1001
998	999	999	1000	2500	1000	1000
996	997	998	1000	999	999	1000

Figure 18: Visualization of smoothing algorithm for depth data. Left – 7x7 pixel array centered on 0 depth value before smoothing process. Right – same pixel array with center pixel corrected with median depth value of 48 surrounding pixels

Once the determination of whether one pixel has a depth difference above or below the threshold, the system must then determine if the patient has moved. Rather than simply counting the fraction of pixels that were turned red or blue, the area represented by individual pixels allows for a more accurate quantification of the gross patient movement. Microsoft has provided the following formula to calculate the area represented by any specific pixel[82]:

$$Area = (depthValue * inverseFocalLength)^2 \qquad (14)$$

Where the Inverse Focal Length of the Kinect has been calculated by Microsoft to be 0.0027089166 and the Depth Value is the value returned by the depth sensor for that specific pixel in mm. Using this equation, not only can the area represented by pixels that have moved be calculated, but the entire area of the body within the ROI can also be calculated. Thus, a percentage of body area within the ROI that has moved outside the threshold value can be calculated and used as the indicator of movement.

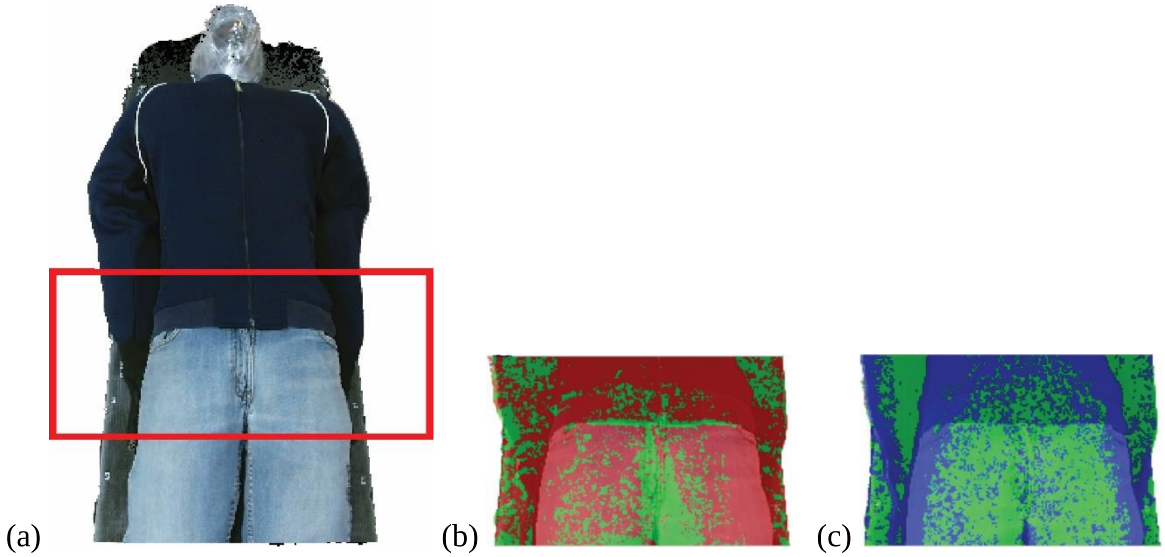


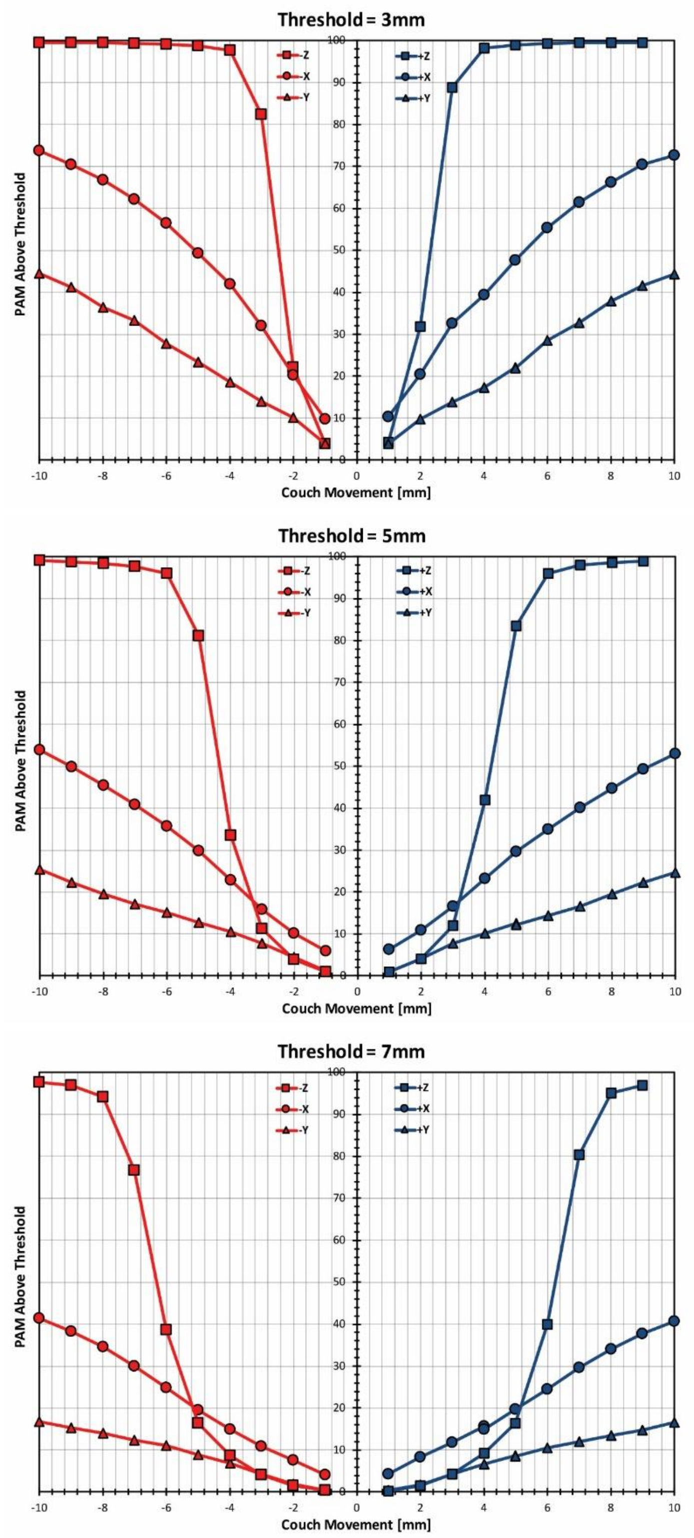
Figure 19: (a) RANDO phantom on couch with background removed and pelvic ROI selected. A threshold of 5mm was used during comparison. Display of ROI with (b) pixels that have moved away from the camera more than 5mm in red and (c) pixels that have moved towards the camera more than 5mm in blue.

To accurately quantify movement for this study, an anthropomorphic RANDO phantom was used to ensure that the patient being imaged remained static. The phantom itself consists of a head, chest, abdomen, and pelvic region. To ensure the system would begin the body tracking process with the phantom, a jacket, gloves, and pants were added to allow for an approximate, humanoid shape that would be recognizable as a body. The phantom was placed on a PerfectPitch 6DoF couch, as was done in the joint data study. The couch was again moved in increments of 1mm along each orthogonal direction (X, Y, and Z) with data recorded at each interval with movement ranging from -10mm to 10mm. Additionally, the couch was moved to radial distances of 3mm, 5mm, 7mm, and 10mm with varying values of X, Y, and Z coordinates. The ROI was set to the pelvic region and the threshold of movement was set to values between 1mm and 10mm in 1mm increments. The percentage of pixels and area outside of each threshold were recorded and compared to actual couch movement with the “Red”, “Blue” and “Green” area recorded separately for analysis. For the remainder of this section, this value will be referred to as Percent Area Movement (PAM).

Figure 20 and Figure 21 display the data obtained while tracking the pelvic ROI through various orthogonal movements. Figure 20 displays plots of PAM with pixel depth differences at or greater than the threshold value listed (3 threshold plots are shown as an example). Figure 21 plots the same calculated PAM but only the values calculated when the couch movement is the same as the threshold value (i.e. when the system should indicate movement has occurred).

As evident in Figure 20, the PAM value for movement at or exceeding a specific threshold is highly dependent on the orthogonal direction of movement, with the Y direction having the least sensitivity and the Z direction having the most sensitivity, particularly when couch movement was at or near threshold. Figure 21 indicates that the PAM obtained when at various thresholds does remain relatively constant in the X and Y direction for movement greater than 2mm.

However, when compounded with movement in more than one axis, the percentage values were found to no longer be reliable indicators of quantitative motion. For example, when moved to the (4mm, 3mm, 0mm) coordinates which represent radial movement of 5mm, the PAM was calculated to be 22.2% when threshold was set to 5mm. When set to the same threshold and same radial distance but moved to the (4mm, 0mm, 3mm) coordinates instead, the resulting PAM was 34.2%. To be a reliable movement indicator, movement at similar radial distances in any direction should have similar PAM values, which was not the case here. Table 7 summarizes some of the values obtained while attempting to validate this process. This trend persisted no matter which radii or coordinates were used and was most likely due to the high sensitivity of this process for motion in the Z direction.



(a)

(b)

(c)

Figure 20: Plots of PAM (Percent Area Movement) within the pelvic ROI with depth differences above the threshold value of (a) 3mm, (b) 5mm, and (c) 7mm. The +/- and X/Y/Z nomenclature within the legend refers to movement in the positive or negative direction of the specified orthogonal axis.

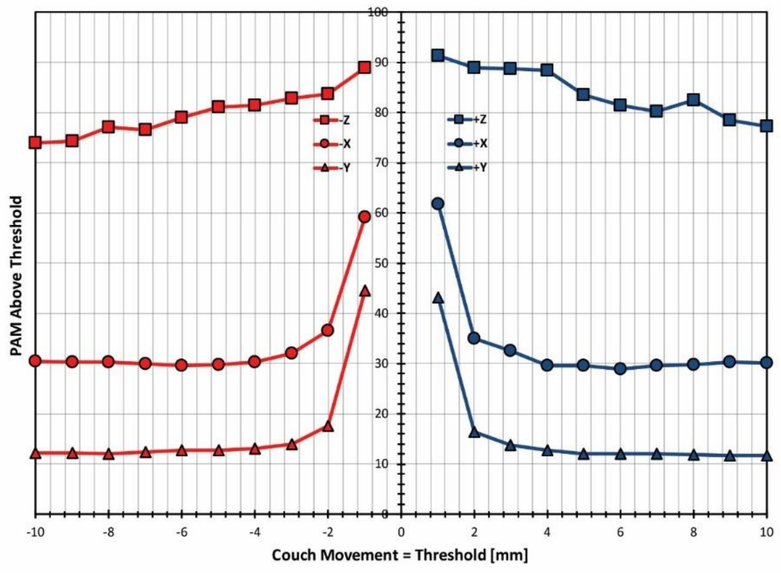


Figure 21: Plot of PAM (Percent Area Movement) within pelvic ROI with depth difference at or above threshold value when the couch had been moved to a distance equal to the threshold. The +/- and X/Y/Z nomenclature within the legend refers to movement in the positive or negative direction of the specified orthogonal axis

During this analysis, it was found that the PAM for movement in the Z direction when at threshold were always large and were larger than values calculated when movement occurred in the X or Y direction. The PAM calculated when movement in the Z direction was at threshold was always greater than 60%. Knowing this, the PAM for an ROI can simply be utilized as an indication of movement within the Z direction. Even when compounded with movement in the X or Y direction, the percentage value calculated would only reach over 60% when movement in the Z direction reached the threshold value (Table 8). As both the red and blue pixels (movement toward and away from the camera, respectively) within the ROI are recorded separately, this indication of movement beyond threshold can even indicate which direction along the Z axis that movement has occurred.

Threshold [mm]	Radius [mm]	Z Movement [mm]	X Movement [mm]	Y Movement [mm]	PAM
3	3	0	0.0	3.0	06.9%
			1.6	2.5	17.9%
			3.0	0.0	27.5%
3	3	1	0.0	2.8	12.4%
			2.0	2.0	22.9%
			2.8	0.0	32.0%
3	3	2	0.0	2.2	25.9%
			1.5	1.6	30.5%
			2.2	0.0	38.6%
5	5	0	0.0	5.0	06.6%
			4.0	3.0	22.2%
			5.0	0.0	27.9%
5	5	1	0.0	4.9	14.2%
			3.5	3.5	23.9%
			4.9	0.0	29.1%
5	5	3	0.0	4.0	17.6%
			2.6	3.0	27.7%
			4.0	0.0	34.2%
7	7	0	0.0	7.0	06.4%
			5.0	4.8	19.3%
			7.0	0.0	28.5%
7	7	1	0.0	6.8	06.7%
			4.0	5.0	16.5%
			6.8	0.0	27.8%
7	7	3	0.0	6.3	10.5%
			4.0	4.9	22.4%
			6.3	0.0	32.8%

Table 7: Comparison of PAM values calculated for movements occurring at radial distances of 3mm, 5mm, and 7mm with the threshold value matching the radial distance. Note the increase in PAM as the X and Z movement increases with no consistency in the actual PAM value obtained when compared to similar radial distances

Use of a tool such as this that can monitor intrafraction motion for a specific ROI in the Z direction can prove quite useful in a clinical setting. Studies regarding intrafraction motion in the anterior-posterior direction for prostrate treatments indicate that movement of a few millimeters can and does occur even before the start of treatment[86,87].

Threshold Value [mm]	Z Movement [mm]	X Movement [mm]	Y Movement [mm]	PAM
3	3	6.3	0.0	65.0%
		4.0	0.0	66.2%
		6.2	1.0	65.4%
		4.0	1.0	62.7%
		5.5	2.0	64.6%
		3.5	2.0	61.7%
		2.6	3.0	64.3%
5	5	4.8	0.0	62.4%
		0.0	0.0	66.9%
		4.9	2.0	64.5%
		8.6	2.0	62.1%
		0.0	4.0	64.6%
		7.7	4.0	62.1%
		6.2	6.0	63.1%
7	7	7.0	1.0	63.6%
		0.0	1.0	66.3%
		6.8	2.0	63.4%
		5.7	4.0	65.3%

Table 8: Comparison of PAM values calculated for movements in the Z direction and various movements in the X and Y direction. Note the consistent PAM values when movement in the Z direction is at threshold.

These two studies have provided two distinct processes of patient motion tracking with the Kinect v2 which can be shown to correlate with actual movement in such a way to create threshold values for an indication movement. For large scale anatomical movement, automatic skeletal tracking and body recognition by the Kinect showed 5 specific joints to be more accurate with gross motion in a radial direction when compared to an initial state: Left Elbow, Right Elbow, Right Hip, Left Hip, and Spine-Base. Their COV calculated for these joints was approximately 20% for the Elbow joints and 25% for the Hip and Spine-Base joints when measured at 5mm and 10mm radial movements. This allowed for thresholds to be set for calculated radial distances of 3mm to indicate 5 mm of actual movement and calculated radial distances of 7.5mm to indicate and 10mm of actual movement.

For smaller areas or areas of specific interest, a ROI can be drawn on a patient to continually monitor a specific portion of the body to indicate movement in the Z direction. The surface area of the pixels representing the body within the ROI can be calculated and if the percentage of that area that has moved beyond a threshold value exceeds 60%, it can be interpreted that the ROI has moved in the Z direction beyond said threshold. The combination of both the joint tracking and ROI with the Kinect allow for a robust implementation of patient motion tracking for both small and large patient movements allowing the user to receive indication that a beam shutoff is required or that re-imaging may be required due to extraneous patient movement from the initial setup position.

CHAPTER 5 "COMPARATIVE ANALYSIS OF RESPIRATORY MOTION TRACKING USING MICROSOFT KINECT V2 SENSOR"

Previous chapters have explored using the Kinect to improve patient safety and treatment efficacy in the radiation oncology clinic by imaging with the assumption of a static patient or area of interest in front of the camera. Although this assumption may be valid for many treatment sites, the simple act of breathing can deform a patient's external contours and induce internal motion. This movement becomes exceedingly important to track and quantify for tumors located within the thorax and abdomen as they are significantly affected by respiratory motion[28,29, 30]. As radiotherapy treatments become increasingly precise, identifying and visualizing tumor movement during treatment becomes exceedingly important.

One specific way to acquire and process this information is through the use of a 4DCT, by which the respiratory motion of the patient is tracked using a respiratory surrogate [31, 88]. As mentioned previously, these processes typically require some manner of physical device attached to the patient by way of a marker placed on the patient's surface or apparatus worn by the patient. However, these may require repositioning and multiple attempts to get an accurate respiratory motion trace due to irregular breathing and can restrict the respiratory motion tracking to one specific area on the patient, typically the lower abdomen. Here, the Microsoft Kinect v2 sensor was adapted to trace and record a patient's breathing cycle by way of a marker-less process, doing away with any requirement for external hardware to be attached to the patient.

Previous research into respiratory motion tracking using the Kinect utilized either the Kinect v1 or required a translational marker to be placed on the patient surface or embedded within clothing worn by the patient, similar to other respiratory tracking systems currently available for purchase [69,70,89]. The latest version of the Kinect contains higher resolution sensors than the previous model, which helps remove the requirement for a translational marker to track respiratory motion. The removal of this requirement allows a simpler process to be employed with less trial-and-error to obtain a useful respiratory trace.

In this chapter, the Microsoft Kinect v2 sensor's ability to trace and record a patient's breathing cycle by way of a marker-less process was investigated. The cost of utilizing a marker-less approach is the inability to guarantee tracking of a specific point on the patient's surface. This is due to the fact that the tracking process is done with respect to pixels in an image frame as opposed to fixed anatomical locations. Motion of the patient's surface during breathing will, in general, cause slightly different anatomical points within some connected surface area to pass through the tracked pixels within the image. This inherent difference between marker-based and marker-less tracking could theoretically lead to differences in recorded breathing traces between the methodologies. As a result, our evaluation of the Kinect v2 sensor as a motion tracking device also includes, by necessity, an overarching evaluation of a general marker-less approach whereby the motion tracking is in some sense decoupled from the motion of singular points on the patient's surface. The information presented here has been submitted for publication in 2017 within the *Journal of Applied Clinical Medical Physics* but is also presented as part of this dissertation[90].

The Kinect respiratory tracking process created was compared against both the Varian RPM Respiratory Gating system (RPM) and the Anzai Gating system (Anzai). For comparison and accuracy measurements, RPM and Anzai were both employed on a subject at the same time with the Kinect mounted above the patient. RPM traces the movement of a propriety marker placed on the subject's abdomen through the use of infrared sensors at a rate of 30 fps [35]. Anzai utilizes a belt strapped around the subject's abdomen which contains a pressure sensor to track the respiratory motion at a rate of 40 fps[36,37]. The Kinect returns depth values, in mm, for every pixel within the depth frame at a rate of 30 fps[82]. All three devices acquired data simultaneously with the RPM marker placed directly on top of the Anzai belt and data were exported from all three for analysis.

Currently available 4DCT procedures employ either a phase based or amplitude based binning process when incorporating respiratory motion[91,92]. As such, the traces recorded for all three devices were analyzed with each process in mind. With a phase based binning process, the period of one cycle is obtained and divided up into 10 phase portions with bins of equal width. With an amplitude based binning process, the bins are divided up into percentages of the maximum and minimum amplitude throughout one cycle, typically calculated as 100%, 80%, 60%, 40%, 20%, and 0%. These percentages correspond to specific physical states of the breathing cycle (mid-inhalation, maximum exhalation, etc.). Figure 22 gives a visualization of the difference between the two processes. Given irregularities that can occur in a patient's breathing pattern which may cause shifts in the phase but not amplitude, many binning procedures are moving away from a phased based process in favor of an amplitude based process[93]. However, in this manuscript, both binning procedures are used to test the validity of data being recorded by the Kinect.

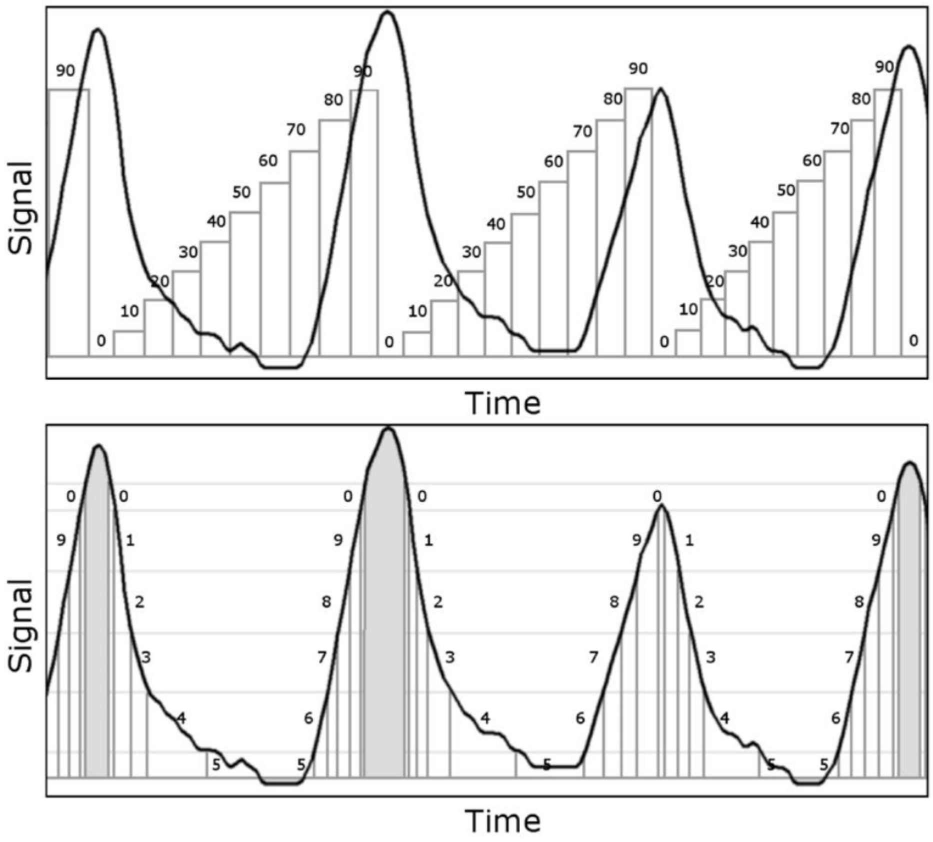


Figure 22: Visualization of a phase based binning process (top) and an amplitude based binning process (bottom). Notice the equal width of bins for the phase based binning process compared to the variable bin widths for the amplitude based binning process[94].

Calculation and identification of the local maximum and minimum for each breathing cycle (100% amplitude, and 0% amplitude, respectively) was implemented through a simple local comparison algorithm. To mitigate possible misidentification of per-cycle maxima and minima due to temporally small, noisy perturbations, each individual data point of the trace was compared to the 10 data points acquired before and after, allowing for 20 comparisons in total. If the data point in question was greater than or equal to the 20 points surrounding it in time, it was considered 100% amplitude for that breathing cycle. If the data point was less than or equal to the 20 points surrounding it in time, it was considered 0% amplitude. Similar to analyses performed in the clinic when acquiring respiratory traces, multiple values of 100% or 0%

amplitude may be identified by the system for the same breathing cycle. As such, manual adjustment was required to remove duplicate local maximum or minimums.

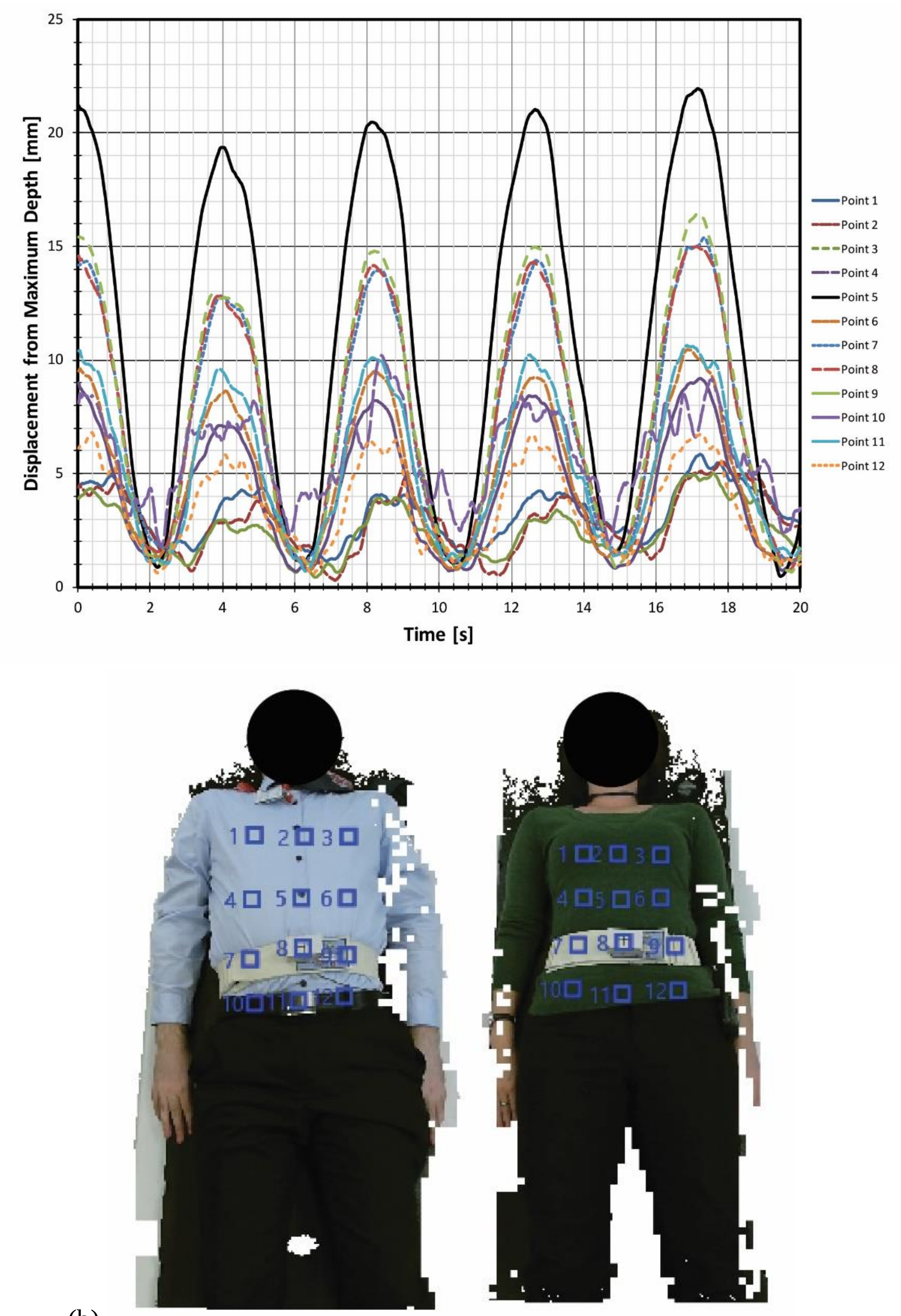
In order to obtain data for the respiratory trace, the Kinect v2's depth camera was utilized. The depth camera returns depth data for each pixel within its 512 x 424 frame in 1mm increments. Rather than track movement associated with a specific location on the body and monitor depth changes as it moves across the frame, as would be done with a physical marker, the system is designed to track specific pixels from the depth image and record the depth values returned over time. Although different from the typical respiratory tracking processes, which track a specific location on the body, this study investigates if both processes can produce the same respiratory trace with congruent results.

To begin the data collection process, the user manually selects 5-12 points anywhere on the patient for respiratory motion tracking. Data collection duration is also selected by the user and the process can be stopped manually if needed. Each point has depth data continuously recorded during the acquisition process, with visual displays of each trace, and the program can choose the most accurate representation of the respiratory motion by calculating the largest difference between the maximum and minimum distances recorded for each point created. Additionally, as all traces during acquisition are saved, the user has the ability to view and select traces from different points to those chosen by the program in order represent respiratory motion if so desired.

To ensure that the data collection process and GUI were as user friendly as possible, the body tracking capabilities of the Kinect were implemented. The Kinect software has the ability to detect when a human body has entered the frame of the camera and can differentiate between pixels associated with a body versus pixels belonging to the background. Once the body is recognized by the Kinect, the background can be removed from the image displayed allowing for an easy visualization of the patient. The advantage to utilizing this process is that the image displayed is aligned, pixel for pixel, exactly to the depth images generated. This allows selection of specific points on the patient to exact depth data generated by the depth sensor.

During each tracking session, the Kinect was mounted directly over the subject pointing down at an angle of roughly 45 degrees and was set at a height of roughly 0.75m. Data acquisitions were performed on both a male and female subject for approximately 120s and each were asked to breathe in a manner typical for the individual with no breath holds.

Figure 23 displays a sample respiratory trace from the Kinect with all 12 points selected by the user as well as images of each subject with all 12 points shown. As previously mentioned, the point selected by the system to represent the respiratory motion is done so by calculating the largest amplitude between the traces created for all points selected. In this example, point 5 (located on the diaphragm) has the largest difference between the maximum and minimum values throughout the trace and, as such, it would be chosen by the system as the representation of the patient’s respiratory motion. For analysis of the trace generated by the Kinect, point 5 from each subject was chosen to represent the respiratory motion. This allowed for analysis and comparison of a trace obtained from a location that was different from those obtained from RPM and Anzai while still containing amplitudes large enough to be compared to both systems.



(a)

(b)

Figure 23: (a) Sample respiratory trace from Kinect with all 12 points selected by the user. Values along the y-axis indicate the difference between the current depth value and the maximum depth value for that user selected point throughout the recorded trace. (b) Subjects 1 and 2 with points 1-12 selected

Initial comparisons of the traces between systems involved implementing typical amplitude and phase based binning methods that would be used for gating purposes within the clinic. With the amplitude binning method, the maximum and minimum displacement values were obtained for each breathing cycle and amplitude values were obtained for 100%, 80%, 60%, 40%, 20%, and 0% of the maximum value. The times at which each percentage occurred within each breathing cycle were then obtained across all three devices. For the phase based binning method, the maximum displacement value was again utilized for each breathing cycle and the period of the cycle was divided into 10 equal bins. The times for each bin were then obtained across all three devices.

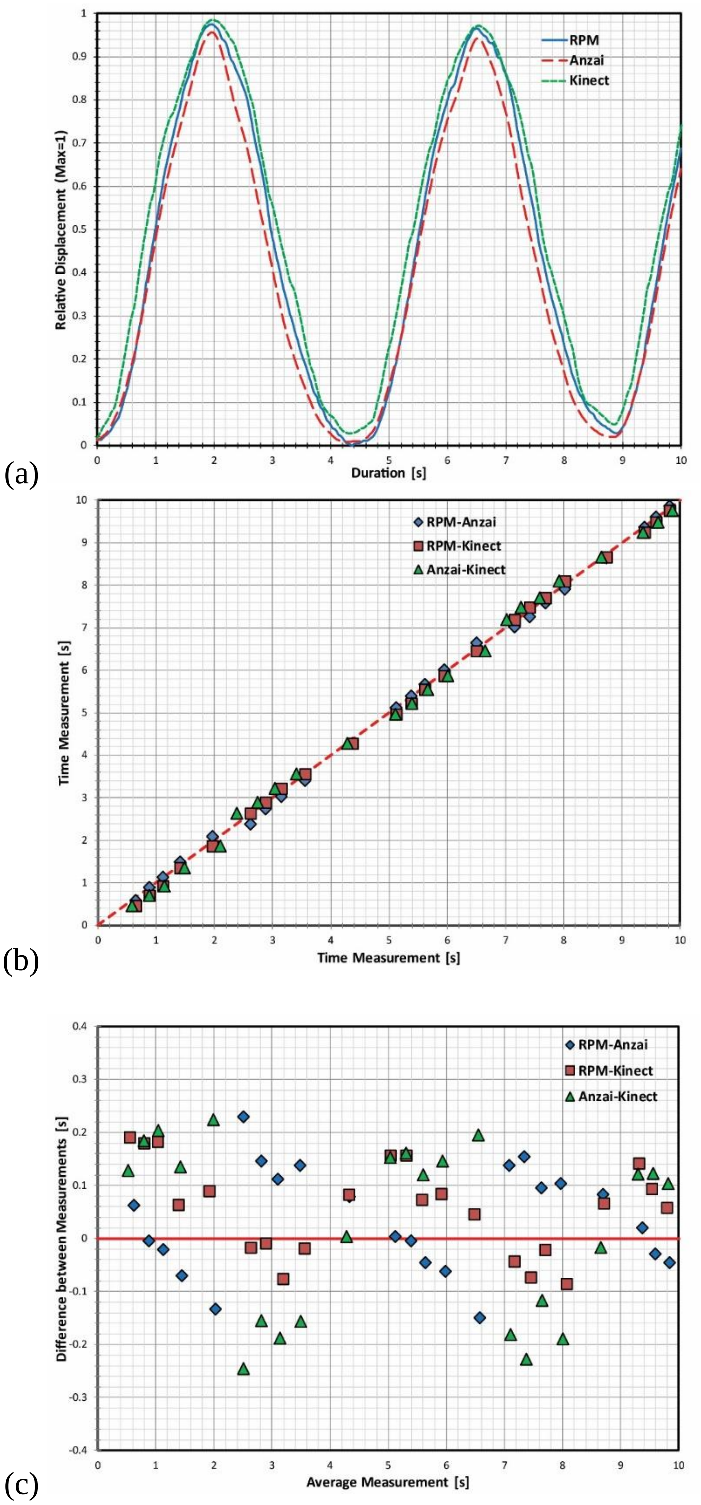
Portions of the data obtained with all three respiratory systems collecting data are displayed in Figure 24a and Figure 25a. To align and overlap the data, the relative displacement was used based on the global maximum displacement during the respiratory tracking. Initial analysis of the times obtained for the amplitude binning process was accomplished using a Bland-Altman approach[95]. First, measurements between two of the devices were plotted along a line of Y=X for simple comparability (see Figure 24b and Figure 25b pertaining to Subjects 1 and 2, respectively) with one device measurement as the X coordinate for a point, and another device measurement as the Y coordinate. The closer each point is to the line of Y=X, the more similar the measurements. Next, all comparisons were analyzed utilizing a Bland-Altman plot to test for agreement as shown in Figure 24c and Figure 25c for Subjects 1 and 2, respectively. Here, the plot contains data comparing two devices with each point on the plot having the X and Y coordinates calculated by the following:

$$(X, Y) = \left(\frac{t_A + t_B}{2}, t_A - t_B\right) \tag{15}$$

The X coordinate of a point, $\frac{t_A + t_B}{2}$, represents the average time measurement for a specific amplitude percentage between two devices (t_A for device A, and t_B for device B). The Y value, $t_A - t_B$, represents the difference between the time measurements from the two devices being compared. In essence, the difference between two time measurements for a specific amplitude percentage (Y value) is plotted against the average of those same two measurements (X value)[95,96]. The data analyzed here with the Bland-Altman approach only represents the data obtained from the amplitude binning process. This was simply done for clarity as analysis for the phase based binning process would yield similar results.

Further analysis utilized the Bland-Altman plots for amplitude time values obtained throughout the 120 seconds of recording. Here, the data is plotted around the line representing the mean for all measurements as well as lines representing the mean ± 1.96*SD (i.e. the 95% Confidence Interval). Figure 26 displays comparisons from all three devices with values obtained for Subject 1 and Subject 2.

With the Bland-Altman plots created in Figure 26, the agreement between two devices producing similar measurements lies with the percentage of values that fall within the span of the mean ±1.96*SD. Typically, two devices can be shown to produce similar measurements if roughly 95% of the data within the plot falls inside this range. Table 9 summarizes the percent of values within the range specified and indicates that all three devices have similar agreement with one another regarding the time values obtained for the amplitude percentages.



(a)

(b)

(c)

Figure 24: Plots created for Subject 1 based on two full respiratory cycles. (a) Plot with all three traces overlapped, (b) plot of time measurement versus time measurement obtained for the amplitude binning process for all three device comparisons, (c) Bland-Altman plot generated using the same measured values plotted in (b).

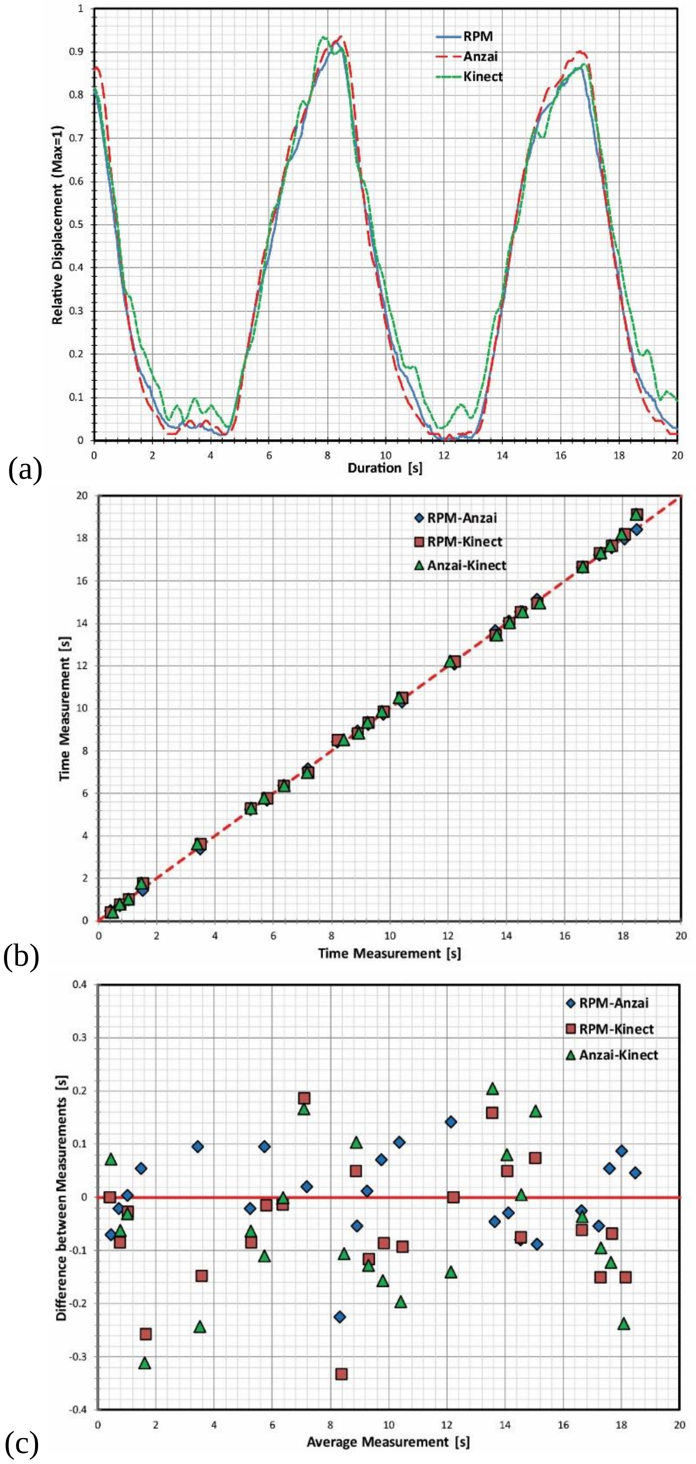
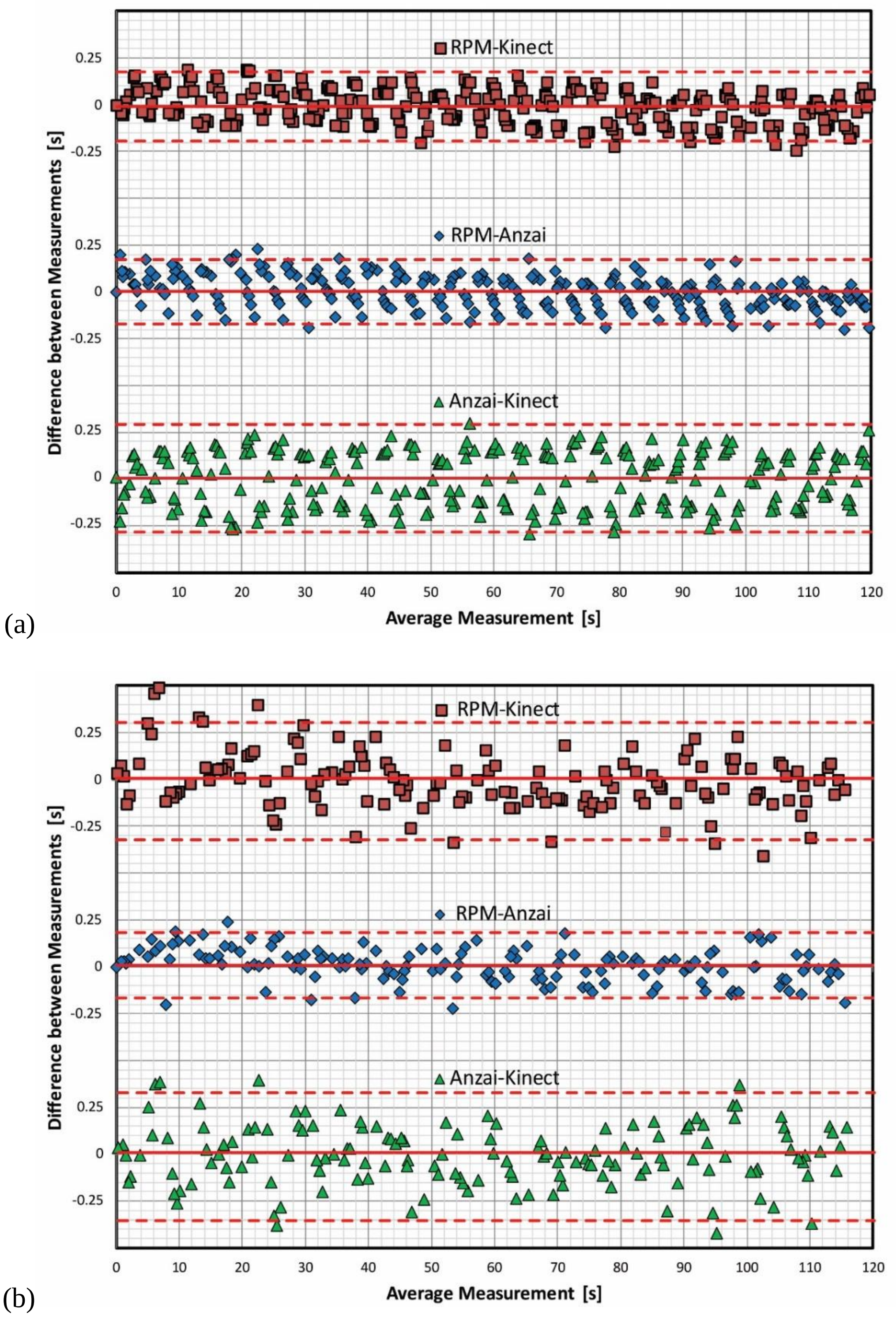


Figure 25: Plots created for Subject 2 based on two full respiratory cycles. (a) Plot with all three traces overlapped, (b) plot of time measurement versus time measurement obtained for the amplitude binning process for all three device comparisons, (c) Bland-Atman plot generated using the same measured values plotted in (b).



(a)

(b)

Figure 26: Bland-Altman plots generated for (a) Subject 1 and (b) Subject 2 based on time values obtained through the amplitude binning process. Comparisons were made between RPM and Kinect (top plot), RPM and Anzai (middle plot), and Anzai and Kinect (bottom plot). The solid line is the mean of all values and the dashed lines represent the Mean ±1.96 SD (i.e. the 95% confidence interval).

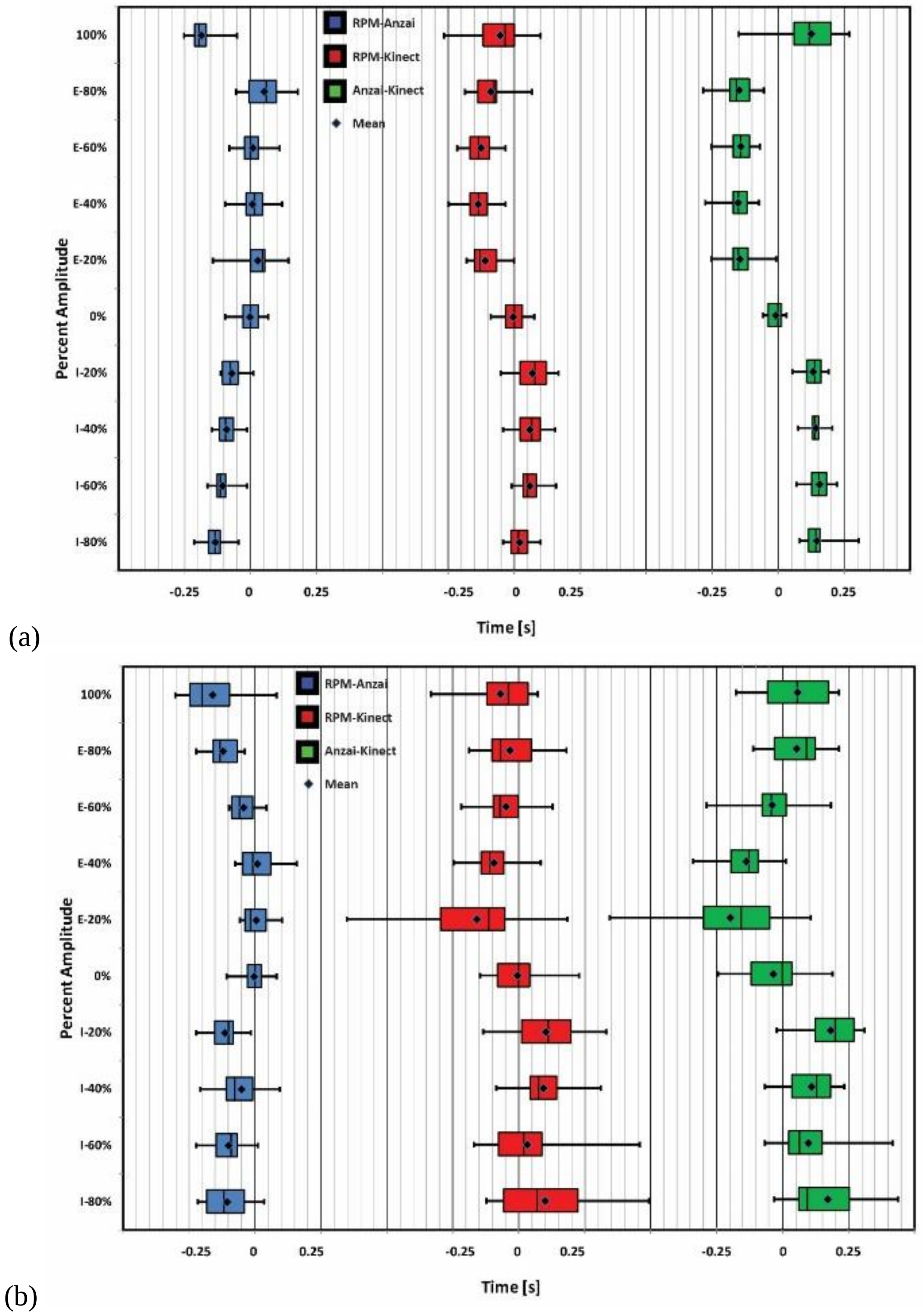
Subject	Percentage of Values Within 95% Confidence Interval		
	RPM-Anzai	**RPM-Kinect**	**Anzai-Kinect**
Subject 1	96.09%	96.44%	98.93%
Subject 2	96.03%	93.38%	94.04%

Table 9: Summary of Bland-Altman values based on the data plotted in Figure 26

Lastly, the difference between the times obtained for each device within the amplitude and phase based binning process was calculated and the average difference across devices for each percentage was calculated. Figure 27 displays the Interquartile Range (IQR) for the amplitude time differences using a Box and Whiskers plot for both subjects. Figure 28 displays the IQR for the phase time differences across each of the calculated bins utilizing similar Box and Whiskers plots for both subjects.

The IQR becomes an important quantifier when analyzing the differences between traces as it indicates a range of time that specific percentages of amplitude and phase differ between devices. A summary of IQR values can be found in Table 10 and Table 11 with Table 10 displaying the average time span within the IQR for each comparison and Table 11 containing the average mean time difference and standard deviation for each comparison.

When analyzing traces with the amplitude based binning process for each breathing cycle, the IQR for the time differences between devices was low overall, typically lower than 0.2s. Subject 1 had much better agreement across devices with the IQR spanning a time frame of ~0.07s, while the IQR for Subject 2 spanned a time frame of ~0.15s.



(a)

(b)

Figure 27: Box and Whiskers plots displaying the IQR of time differences obtained between devices when utilizing an amplitude binning process. (a) Subject 1 and (b) Subject 2 both performed natural breathing patterns over a period of 120 second. "I" and "E" next to the percentage value on the y-axis indicate "Inhalation" and "Exhalation", respectively. The mean for each comparison is indicated with a point within each box.

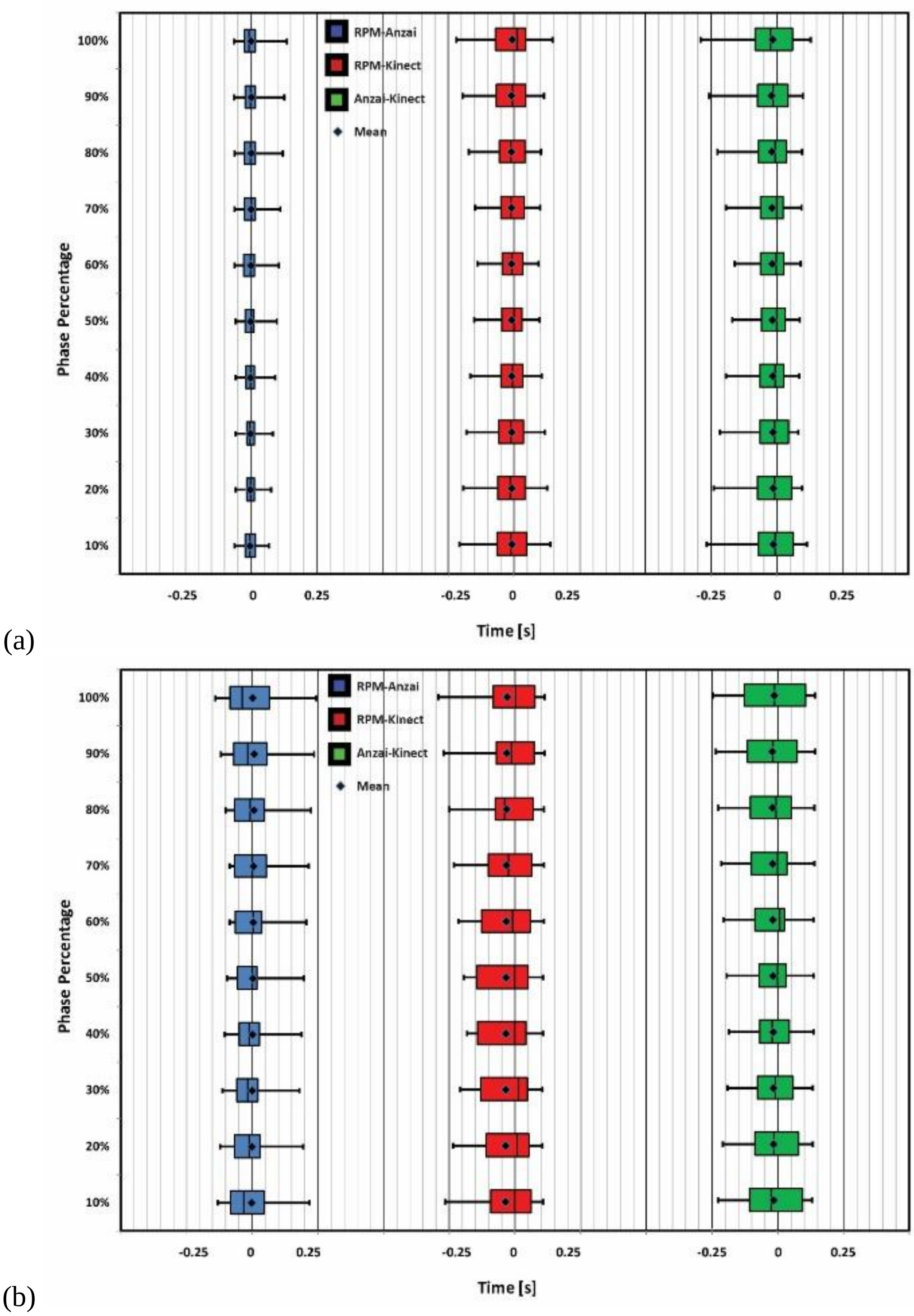


Figure 28: Box and Whiskers plots displaying the IQR of time differences obtained between devices when utilizing a phase based binning process. (a) Subject 1 and (b) Subject 2 both performed natural breathing patterns over a period of 120 second. The mean for each comparison is indicated with a point within each box.

Binning Process	Subject	Average IQR Time Spans [s]		
		RPM-Anzai	RPM-Kinect	Anzai-Kinect
Amplitude	Subject 1	0.056	0.076	0.062
	Subject 2	0.095	0.157	0.160
Phase	Subject 1	0.036	0.096	0.110
	Subject 2	0.106	0.167	0.154

Table 10: Average time spans of the Interquartile Range (Q3-Q1) calculated between each device with 2 subjects. Values were averaged over all 10 amplitude and 10 phase bins per cycle created in the above analysis.

Binning Process	Subject	Average Mean Time Difference Throughout Trace [s]		
		RPM-Anzai	RPM-Kinect	Anzai-Kinect
Amplitude	Subject 1	0.009 ± 0.087	-0.002 ± 0.095	-0.011 ± 0.144
	Subject 2	0.020 ± 0.110	-0.007 ± 0.185	-0.026 ± 0.233
Phase	Subject 1	-0.137 ± 0.034	-0.031 ± 0.067	0.106 ± 0.0742
	Subject 2	-0.082 ± 0.086	-0.072 ± 0.115	0.010 ± 0.0984

Table 11: Average time difference throughout trace between each device with 2 subjects. Values were averaged over all 10 amplitude and 10 phase bins per cycle created in the above analysis.

The largest deviation when comparing all three devices in this manner occurred for Subject 2 during the 100% portion (Max Inhalation) and 20% Exhalation portions of the curve. Here the IQR spanned ~0.25s for both portions when comparing the Kinect to Anzai or RPM. However, when comparing RPM directly to Anzai, the 100% portion had a time span of ~0.12s whereas the 20% Exhalation bin spanned ~0.10s.

When analyzing traces with the phase based binning process, the Kinect values from Subject 1 were, again, in much better agreement with RPM and Anzai belt compared with Subject 2, yet time differences for each bin between the devices were still quite low. For Subject 1, the IQR spanned a time frame of ~0.08s when comparing the Kinect to the RPM or Anzai verses a difference of ~0.07s when RPM was compared to Anzai directly. For Subject 2, the IQR spanned a larger range of ~0.16s when the Kinect was compared to RPM or Anzai but was ~0.12s when RPM was compared directly to Anzai.

Given these ranges of time differences for the IQR, it becomes important to quantify how this would affect a 4DCT being generated by incorporating the couch feed. In our scanning protocols at Karmanos Cancer Institute, a typical 4DCT may include a couch pitch of 0.1, 0.5 gantry rotations/s, and detector configuration of 24 x 1.2mm, giving the effective movement of the couch as 5.76 mm/s. Although the scans are helical in nature, we can estimate reconstruction differences of "effective slices" using this information and the variation between the respiratory traces. Assuming a constant rate of movement and 1.5mm thick slices, it can be said that 3.84 effective slices are acquired every second with deviations of the expected time for slice acquisition creating a slice offset. Table 12 summarizes what minimal impact these IQR values would have during a 4DCT acquisition process.

(a)

Binning Process	**Subject**	**Couch Movement @ 5.76 mm/s [mm]**		
		RPM-Anzai	**RPM-Kinect**	**Anzai-Kinect**
Amplitude	Subject 1	0.35	0.42	0.37
	Subject 2	0.54	0.83	0.88
Phase	Subject 1	0.25	0.48	0.50
	Subject 2	0.67	1.01	1.05

(b)

Binning Process	**Subject**	**Fraction of Slice Offset @ 3.84 slices/s**		
		RPM-Anzai	**RPM-Kinect**	**Anzai-Kinect**
Amplitude	Subject 1	0.17	0.21	0.18
	Subject 2	0.26	0.41	0.43
Phase	Subject 1	0.13	0.24	0.24
	Subject 2	0.33	0.50	0.52

Table 12: Summary of (a) couch movement and (b) fraction of slices that would have occurred during the IQR time spans calculated for each subject.

The analyses presented in this chapter of the respiratory traces obtained with the Kinect v2 respiratory tracking process indicate that the Kinect traces are congruent to those obtained with RPM and Anzai. Visually, when overlapping traces from all three devices, there is minimal difference between them. When analyzing the traces through an amplitude and phase based binning process, time values associated with each amplitude and phase percentage were extracted and compared across each device. Using the Bland-Altman approach, it was shown that between 93%-96% of the time values fell within the 95% confidence interval when comparing the Kinect to RPM and between 94%-99% of the time values fell within the 95% confidence interval when comparing the Kinect to Anzai. These ranges indicate that each of the devices recorded similar measurements to one another. IRQ values were then calculated for comparisons between devices for the amplitude and phase based binning processes. Again, values obtained for comparisons between the Kinect and RPM or Anzai were shown to be similar to those obtained when comparing RPM to Anzai. Deviations that did occur with the IRQ values in these comparisons were shown to have minimal effect on the couch movement or slice offsets that would occur during a 4DCT acquisition process.

One item of note is in regards to the time values obtained from the traces associated with Subject 2. The traces used in the analysis were noticeably more noisy than those used for Subject 1, indicating an overall reduction in the magnitude of the patient surface motion. When analyzing the raw data, it was found that the reduction of magnitude was evident in that the average difference between the maximum and minimum depth values of each respiratory cycle was 19.1mm for Subject 1 but only 7.8mm for Subject 2. As mentioned previously, the analysis performed for the Kinect traces was done so utilizing Point 5 (directly over the diaphragm). This point was chosen as the point of comparison simply to analyze a trace that would be obtained from a different location as the RPM and Anzai trace. Although Point 5 was shown to be accurate and comparable to both RPM and Anzai, increased agreement between devices could be obtained if the system had automatically chosen the point based on the largest amplitude difference. With this criterion in mind, Point 9 (directly to the left of the RPM block) would have been chosen as the representation for respiratory motion. Here, the average difference between the maximum and minimum depth values for each respiratory cycle increased to 9.5mm.

The difference between the two points can be visualized in Figure 29 which displays traces for RPM and Anzai overlapped with traces obtained for both Point 5 and Point 9 from the Kinect for Subject 2. Here, much of the noise present for Point 5 during maximum inhalation and maximum exhalation has dissipated for the trace associated with point 9. Additionally, Table 13 shows the change in average IQR time spans when comparing the traces from Point 5 and Point 9 to RPM and Anzai. It can be seen how the average IQR decreases with the trace from Point 9 to values that are closer to that of the RPM and Anzai comparison. This indicates that further study may be required to determine the effect that selections of different point on the body may have on noise introduced within the Kinect system.

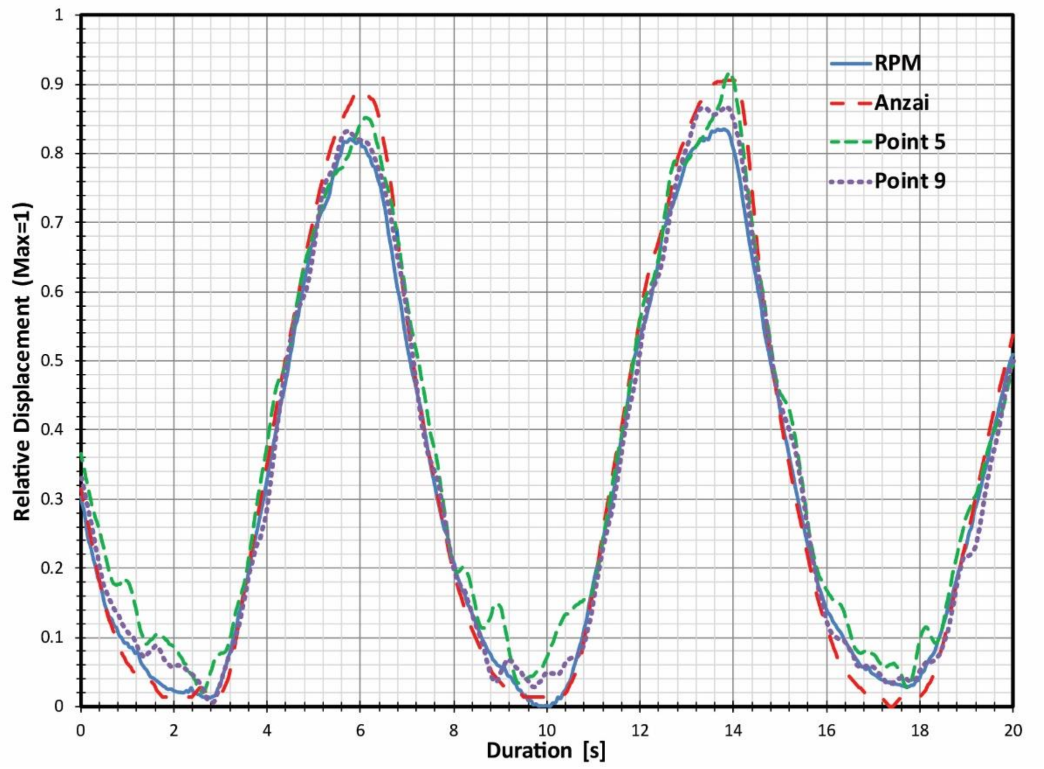


Figure 29: Overlapped respiratory traces obtained for Subject 2 utilizing RPM, Anzai, and both Point 5 (diaphragm) and Point 9 (left of RPM block) from the Kinect. Note the noise generated at maximum inhalation and exhalation for the trace associated with Point 5 has decreased significantly for the trace associated with Point 9.

Binning Process	Point	Subject 2 Average IQR Time Spans [s]		
		RPM-Anzai	RPM-Kinect	Anzai-Kinect
Amplitude	Point 5	0.095	0.157	0.160
	Point 9	0.095	0.098	0.108
Phase	Point 5	0.106	0.167	0.154
	Point 9	0.106	0.137	0.114

Table 13: Comparison of the Average IQR time spans obtained for Subject 2's traces associated with Point 5 (diaphragm) and Point 9 (left of RPM block). Note the decrease in IQR time spans for Point 9 and their similarity to the RPM-Anzai comparison.

Lastly, the previous analyses have quantified the ability of the Kinect to detect standard, cyclic breathing patterns but have not indicated its ability to track irregular breathing patterns. As a patient attempts to remain still and control their breathing during respiratory tracking sessions, the inevitable interruption of the cyclic breathing pattern can occur via hiccough, cough, sneeze, etc. Detection of this interruption should be evident when viewing the respiratory trace being recorded. For each type of interruption, the amplitude of the live trace would increase considerably from any previous local maxima due to the sharp rise in the patient’s abdominal region that would result. When gating radiation therapy or using Deep Inspiration breath Hold (DIBH) during therapy sessions with amplitude based tracking, these irregular patterns can quantitatively be determined by utilizing thresholds for the maximum amplitude, above which the radiation beam can be turned off as the system will have determined irregular breathing has occurred.

Figure 30 displays a sample of irregular breathing patterns obtained from Subject 1 during a subsequent tracking session. Here, the subject was asked to induce coughing midway through the tracking session. Again, the RPM and Anzai systems were tracking respiratory motion concurrently with the Kinect. As evidenced by the traces in Figure 30, the first 10 seconds show 3 cycles of regular breathing patterns. After which, irregular patterns are indicated by all three devices due to the induced coughing. One item of note is the slight differences of said traces between devices. The Kinect and RPM both track respiratory motion utilizing IR imaging (albeit, through difference processes) whereas the Anzai system does so via a pressure sensor within a belt around the patient. Although the irregular breathing pattern is visible from all three devices, it is the Kinect and RPM system that have very similar traces, especially in the regions of extremely high amplitudes, corresponding to fast, abdominal excursions in the anterior direction. It would appear that the IR tracking processes are able to track and record subtleties in the respiratory motion of the subject during these portions that are not tracked by the pressure sensor system of the Anzai belt. This is most likely due to the fact that the IR tracking process records actual distance values moved by the portion being tracked as opposed to pressure values recorded by the Anzai pressure sensor.

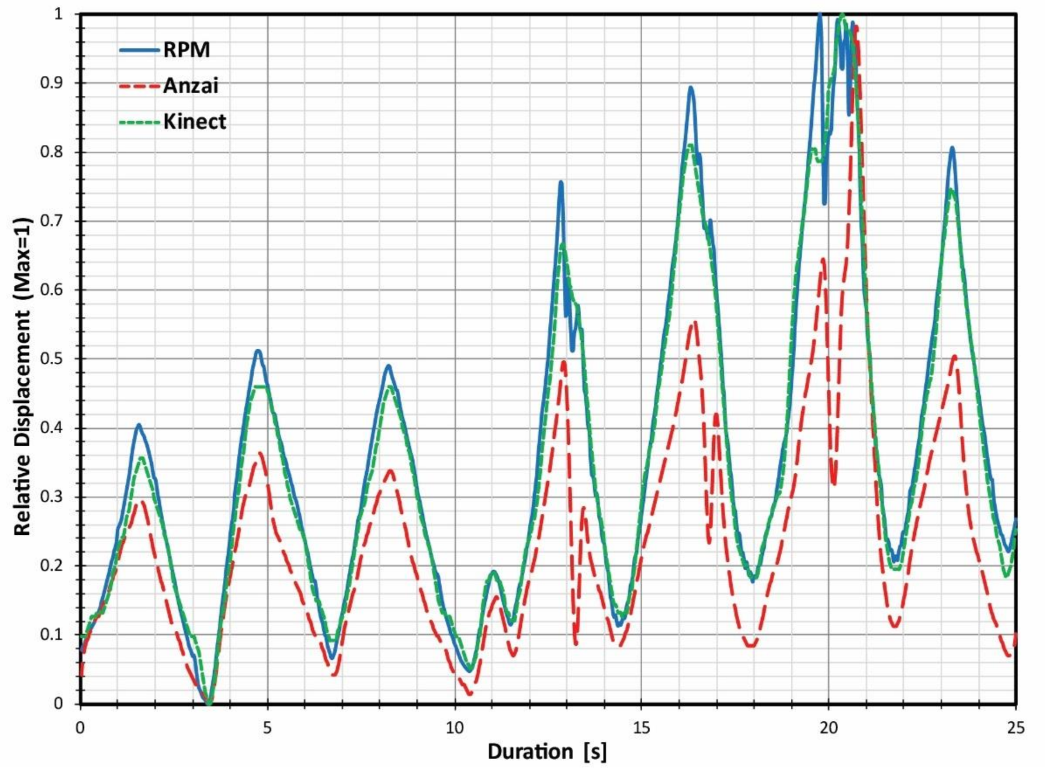


Figure 30: Respiratory traces created from all three devices while subject 1 induced irregular breathing patters through coughing. Notice the irregular patterns begin to develop at approximately 10 seconds into the tracking session.

The various analyses performed in this chapter have shown that recording respiratory motion with the Kinect v2, by way of recording depth values for specific pixels on the depth image, rather than anatomical locations, can be as accurate as the Varian RPM system or Anzai belt and can be easily implemented. The ability to select multiple points on a patient to be used for respiratory tracking through the GUI, allows for a unique and user-friendly setup. Without the need for physical hardware attached to the patient for tracking, points can be selected anywhere on the patient, including the area of the tumor, without interfering with a CT scan or radiation therapy.

CHAPTER 6 “CONCLUSIONS AND FUTURE WORK”

This body of work sought to investigate the Microsoft Kinect v2 sensor in a variety of ways which could useful within the settings of a radiation oncology clinic. As the sensor has already seen many applications within the field of medicine, incorporating it has a multi-purpose device within a radiation oncology clinic would be greatly beneficial given its ease of use and multi-faceted programs.

Firstly, in order to enhance patient safety, a facial recognition and recall process was created using the Kinect for patient verification purposes. In Chapter 3, it was shown that by utilizing the HDFaceMapping library within the SDK provided by Microsoft, a facial recognition process could be created through a specific facial mapping procedure. Here, 31 points are mapped to specific facial landmarks with each point given in 3D space. This allowed 3D vectors to be calculated between each of the 31 points, resulting in 465 vector magnitudes defining a specific face. By creating a database of 39 faces (each represented by 465 vector magnitudes) real-time recognition and recall could be performed by calculating the difference between identical vector magnitudes for a face being acquired in real-time and those within the database. Based on the fact that, under ideal conditions, the average vector magnitude difference for all 465 vector magnitudes between two acquisitions of the same face should be 0, the mean and median of those vector magnitude differences were used as similarity scores.

Analysis showed that when comparing different acquisitions for the same face, both the mean and median of the vector magnitude differences would be very low, whereas, when comparing two different faces, the mean and median had a very large spread of values, typically much higher. ROC curves generated based on varying thresholds for both the mean and median indicated the process was very good at identifying True Positives and True Negatives. Sensitivity and Specificity values were both ~96% with Area Under the Curve (AUC) of approximately 99% for both parameters. Ambient light was also found to play a crucial role in the acquisition process and it was shown that light levels above 200 lux were optimal to ensure consistent and accurate acquisitions.

Although the recognition and identification process is slightly cumbersome for the average patient, future iterations of the process may be able to not only decrease the amount of time required for interface with a patient but to create a simpler process that requires less interaction. Quickly ensuring that the patient in front of the camera is in fact the correct patient will facilitate the verification process ensuring better compliance with the process.

Next, a real-time patient motion tracking system was created utilizing various functions of the Kinect. Chapter 4 showed that such a system was possible by implementing two different, but complimentary processes. First, by utilizing the skeletal and body tracking capabilities of the Kinect, specific joints generated by the system could be tracked, in 3D space for large, anatomical motion tracking. By creating an initial reference state of the patient and their joints, the total distance each joint moved could be tracked in 3D space. By calculating the radial distance moved by each joint, accuracy of movement could be improved by removing any errors caused by unforeseen rotations. 5 specific joints (Left/Right Elbow, Left/Right Hip, and Spine-Base) proved to be stable and consistent enough during tracking to calculate threshold values that would relate to actual movement of the patient. For 5mm of actual movement, a threshold calculated radial distance of 3mm was created and 10mm of radial movement, a threshold calculated radial distance of 7.5mm was created.

For smaller tracking areas, a region of interest (ROI) could be drawn by the user over the patient in order to track more subtle movements. The depth values generated by the Kinect of each pixel within the ROI were tracked and compared to the initial state of the patient. A threshold distance value is set by the user to determine when movement has occurred. If the difference in a pixel's live depth value when compared to the initial state is greater than the threshold value, the pixel is determined to have moved. The area of each pixel can be calculated based on the depth value measured as well as a total area associated with the patient within the ROI. Thus, a percentage of area to have moved (termed PAM in this study) can be determined and used as a quantifier to alert the user that movement has occurred.

Through testing of this process, it was determined that the PAM value calculated was not sensitive to movement in the X (Left/Right) or Y (Superior/Inferior) direction and no correlation could be made between movements in those directions and the PAM value. However, the PAM value was particularly sensitive to movement in the Z (Anterior/Posterior) direction and it was found that, when movement in the Z direction was at or above the threshold distance set, the PAM value would be 60% or greater no matter what distance had been moved in the X or Y direction. This allows for a movement tracking process that can indicate subtle movement in the Z direction (either towards or away from the camera).

Creating two different processes to track and quantify patient motion management with the Kinect allows for a versatile system to be implemented easily into a radiation oncology clinic. With the ability to track large movements in real time, the user would quickly be able to identify if the of the hips or arms have moved beyond some threshold value. Additionally, with the ability to track smaller movements in the Z direction, this process can be particularly useful for patients who may settle into an alpha cradle once relaxed after the initial setup for treatment.

Lastly, the Kinect was utilized to create a marker-less, respiratory motion tracking process that could be implemented for purposes of 4DCT or gaiting during radiotherapy treatments. The unique aspect of this process is that it does away with any physical markers or hardware that are typically required to be attached to the patient, as is the case with many other devices currently available. With a fixed camera position, natural respiratory motion will slightly move a specific anatomical point across the field of view of a camera. Rather than tracking this specific point as it moves across the field of view, this process tracks the depth values returned by the Kinect for specific user selected pixels. In doing so, respiratory traces were obtained and compared to those recorded by the Varian RPM Respiratory Gating system and the Siemens Anzai Gating system.

Overlapping the traces from all three devices showed that they were visually similar to each other. Statistical analysis was performed according to phase and amplitude based binning processes (similar to binning procedures currently performed in radiation oncology clinics while obtaining respiratory traces to be used with 4D-CT). To do so, the times at which each trace reached various amplitude percentage levels and the times at which each trace would be binned for phase percentage values were extracted. Taking the difference of analogous times between one device’s trace to another, the Interquartile Range (IQR) could be calculated for differences throughout the traces. For two subjects, the average IQR between the Kinect and either RPM or Anzai was very low, with values typically less than 0.16s for either binning process. In terms of how this would affect a 4DCT scan, given parameters currently used at Karmanos Cancer Center, during this time frame the couch would only have moved approximately 0.9mm resulting in a fractional slice offset of approximately 0.5.

The analyses performed when comparing the Kinect traces to those traces obtained with RPM and Anzai have shown that Kinect respiratory tracking system can create traces that are comparable to each device. Further improvement upon this process could involve investigation and mitigation of noise introduced for patients who are shallow breathers during maximum exhalation.

The Kinect itself is an easy to use piece of hardware that is easily attainable, and affordable, for any radiation oncology clinic. Creating an interface with the Kinect has been made relatively simple due to the SDK produced by Microsoft. With many different imaging devices currently available to assist in the radiation oncology clinic for patient verification, motion management, and respiratory motion tracking, the ability of the Kinect to accomplish all of these tasks without complex implementations and with accuracy comparable to current commercial hardware, allows it to be a versatile device; one that can be incorporated into a clinic as a multi-purpose device.

BIBLIOGRAPHY

1. Thwaites D, Tuohy J. Back to the future: the history and development of the clinical linear accelerator. *Physics in Medicine and Biology*. 2006;51(13):R343-62.

2. Ezzell G, Galvin J, Low D, Palta J. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT subcommittee of the AAPM radiation therapy committee. *Medical Physics*. 2003;30(8):2089-2115.

3. Low D, Dempsey J. Evaluation of the gamma dose distribution comparison method. *Medical Physics*. 2003;30(9):2455-2464.

4. Kutcher G, Coia L, Gillin M, Hanson W. Comprehensive QA for radiation oncology: Report of AAPM Radiation Therapy Committee Task Group 40. *Medical Physics*. 1994;21(4):581-618.

5. Stern R, Heaton R, Fraser M, Goddu S. Verification of monitor unit calculations for non-IMRT clinical radiotherapy: Report of AAPM Task Group 114. *Medical Physics*. 2011;38(1):504-530.

6. The Joint Commission. Hospital: 2017 National Patient Safety Goals. *The Joint Commission*. 2017. Available at: https://www.jointcommission.org/hap_2017_npsgs. Accessed May 02, 2017.

7. XECAN. User of RFID to enhance the patient experience, increase safety and eliminate treatment errors. *XECAN*. 2017. Available at: http://www.xecan.com/online/docs/Oncology_RFID_WP.pdf. Accessed May 02, 2017.

8. Chowdhury B, Khosla R. RFID-based Hospital Real-time Patient Management System. *6th IEEE/ACIS International Conference on Computer and information Science (ICIS 2007)*. 2007.

9. Basavatia A, Fret J, Lukaj A. Right Care for the Right Patient Each and Every Time. *Cureus*. December 2016.

10. California Healthcare News. The Eyes Have It: Iris Biometrics Safely Identify UCSD Patients for Radiation Oncology Treatment. *California Healthcare News*. 2017. Available at: http://www.cahcnews.com/newsletters/ca-nhaile-0613.pdf. Accessed May 02, 2017.

11. Palmgren JE, Lahtinen T. Fingerprint recognition to assist daily identification of radiotherapy patients. *Journal of Radiotherapy in Practice*. 2009;8(01):17-22.

12. Santhanam A, Dou H, Kurihara A, et al. Three-dimensional Feature Recognition-based Automated Patient Treatment Mismatch Verification System for Radiation Therapy. *International Journal of Radiation Oncology*. 2012;84(3):S742.

13. Xecan. Xecan. *Xecan Smart Facial Recognition*. 2017. Available at: http://www.xecan.com/online/docs/XecanFaceRec_1.pdf. Accessed June 19, 2017.

14. Mayo Clinic. What you can expect. *Test and Procedures: Radiation Therapy*. 2017. Available at: http://www.mayoclinic.org/tests-procedures/radiation-therapy/basics/what-you-can-expect/prc-20014327. Accessed April 29, 2017.

15. American Cancer Society. External Radiation Therapy. *A Guide to Radiation Therapy*. 2017. Available at: https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/radiation/radiation-therapy-guide/external-radiation-therapy.html. Accessed April 29, 2017.

16. Landberg T, Chavaudra J, Dobbs J, et al. ICRU Report 50. *Journal of the International Commission on Radiation Units and Measurements*. 1993;26(1).

17. Landberg T, Chavaudra J, Dobbs J, Gerard J, Hanks G. ICRU Report 62. *Journal of the International Commission on Radiation Units and Measurements*. 1999;32(1).

18. Merlotti A, Alterio D, Vigna-Taglianti R, Muraglia A. Technical guidelines for head and neck

cancer IMRT on behalf of the Italian association of radiation oncology - head and neck working group. *Radiation Oncology*. 2014;9(264).

19. Burnet N, Thomas S, Burton K, Jefferies S. Defining the tumour and target volumes for radiotherapy. *Cancer Imaging*. 2004;4(2):153-161.
20. Bell L, Cox J, Eade T, Rinks M, Herschtal A, Kneebone A. Determining optimal planning target volume and image guidance policy for post-prostatectomy intensity modulated radiotherapy. *Radiation Oncology*. 2015;10:151.
21. Chen A, Farwell D, Luu Q, Donald P, Perks J, Purdy J. Evaluation of the planning target volume in the treatment of head and neck cancer with intensity-modulated radiotherapy: what is the appropriate expansion margin in the setting of daily image guidance? *International Journal of Oncology Biology Physics*. 2011;81(4):943-949.
22. Burgdorf B, Freedman G, Teo B. Evaluation of CTV-To-PTV Expansion for Whole Breast Radiotherapy. *Medical Physics*. 2016;43(6):3437.
23. VisionRT. VisionRT Publications. 2017. Available at: http://www.visionrt.com/education/publications/. Accessed April 27, 2017.
24. C-RAD. Catalyst. 2017. Available at: http://c-rad.se/product/catalyst-2/. Accessed April 27, 2017.
25. Suramo I, Paivansalo M, Myllyla V. Cranio-caudal movements of the liver, pancreas and kidneys in respiration. *Acta Radiol Diagn (Stockh)*. 1984;25(2):129-131.
26. Davies SC, Hill AL, Holmes RB, Malliwell M, Jackson PC. Ultrasound quantitation of respiratory organ motion in the upper abdomen. *British Institute of Radiology*. 1994;67(803):1096-1102.

27. Bryan PJ, Custar S, Haaga JR, Balsara V. Respiratory movement of the pancreas: An ultrasonic study. *Jounal of Ultrasound in Medicine*. 1984;3(7):314-320.

28. Ekberg L, Holmberg O, Wittgren L, Bjelkengren G, Landberg T. What margins should be added to the clinical target volume in radiotherapy treatment planning for lung cancer? *Radiothearpy & Oncology*. 1998;48:71-77.

29. Erridge SC, Seppenwoolde Y, Muller SH, et al. Portal imaging to assess set-up errors, tumor motion and tumor shrinkage during conformal radiotherapy on non-small cell lung cancer. *Radiotherapy & Oncology*. 2003;66(1):75-85.

30. Plathow C, Ley S, Find C, et al. Analysis of intrathoracic tumor mobility during whole breathing cycle by dynamic MRI. *International Journal of Radiation Oncology Biology Physics*. 2004;59(4):952-959.

31. Keall PJ, Mageras GS, Balter JM, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Medical Physics*. 2006;33:3874-3900.

32. Li G, Citrin D, Camphausen K, et al. Advances in 4D Medical Imaging and 4D Radiation Therapy. *Technology in Cancer Research & Treatment*. 2008;7(1):67-81.

33. Liu H, Balter P, Tutt T, Choi B, Zhang J, Wang C. Assessing respiration-induced tumor motion and internal target volume using four-dimensional computed tomography for radiotherapy of lung cancer. *International Journal of Radiation Oncology Biology Physics*. 2007;68:531-540.

34. Kay CS, Kang YN. Curative Radiotherapy in Metastatic Disease: How to Develop the Role of Radiotherapy from Local to Metastases, Frontiers in Radiation Oncology. *Frontiers in Radiation Oncology*; 2013.

35. Varian RPM. Real-time Position Management System. 2007. Available at:

https://www.varian.com/sites/default/files/resource_attachments/RPMSystemProductBrief_RAD5614B_August2007.pdf. Accessed September 01, 2017.

36. Li X, Stepaniak C, Gore E. Technical and dosimetric aspects of respiratory gaiting using a pressure-sensor motion monitoring system. *Medical Physics*. 2006;33(1):145-154.

37. Heinz C, Reiner M, Belka C, Walter F, Sohn M. Technical evaluation of different respiratory monitoring systems used for 4D CT acquisition under free breathing. *Journal of Applied Clinical Medical Physics*. 2015;16(2):4917.

38. Schweikard A, Shiomi H, Adler J. Respiration tracking in radiosurgery. *Medical Physics*. 2004;31(10):2738.

39. Pagliari D, Pinto L. Calibration of Kinect for Xbox One and Comparison between the Two Generations of Microsoft Sensors. *Sensors*. 2015;15(11):27569-27589.

40. Wasenmüller O, Stricker D. Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision. *Computer Vision – ACCV 2016 Workshops Lecture Notes in Computer Science*. 2017:34-45.

41. Khoshelham K, Elberink S. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors*. 2012;12(12):1437-1454.

42. Yang L, Zhang L, Dong H, Alelaiwi A, Saddik A. Evaluating and Improving the Depth Accuracy of Kinect for Windows v2. *IEEE Sensors Journal*. 2015;15(8):4275-4285.

43. PMD Technologies. PMD Technologies. *PMD Technologies*. 2014. Available at: http://www.pmdtec.com/. Accessed September 01, 2017.

44. SoftKinetic. DepthSense Cameras and Modules. *SoftKinectic A Sony Group Company*. 2016.

Available at: https://www.softkinetic.com/products/depthsensecameras. Accessed September 01, 2017.

45. Fotonic. Fotonic. 2017. Available at: http://www.fotonic.com/. Accessed September 01, 2017.

46. Heptagon. Time of Flight (TOF). *Heptagon*. 2017. Available at: http://hptg.com/technology/time-of-flight/. Accessed September 01, 2017.

47. Bamji C, O'Connor P, Elkhatib T, et al. A 0.13 μm CMOS system on-chip for a 512× 424 time-of-flight image sensor with multi-frequency photodemodulation up to 130 MHz and 2 GS/s ADC. *IEEE Journal of Solid-State Circuits*. 2015;50(1):303-319.

48. Valgma L. 3D reconstruction using Kinect v2 camera. *University of Tartu, Faculty of Science and Techonology*. 2016. Available at: https://www.tuit.ut.ee/sites/default/files/tuit/atprog-courses-bakalaureuset55-loti.05.029-lembit-valgma-text-20160520.pdf. Accessed September 01, 2017.

49. Zanuttigh P, Marin G, Mutto C, Dominio F, Minto L, Cortelazzo G. *Time-of-Flight and structured light depth cameras: technology and applications*. Switzerland: Springer; 2016.

50. Crabb R. UC Santa Cruz. *Fast Time-of-Flight Phase Unwrapping and Scene Segmentation Using Data Driven Scene Priors*. 2015. Available at: http://escholarship.org/uc/item/7x91t94. Accessed September 04, 2017.

51. Mutto CD, Zanuttigh P, Cortelazzo GM. Time-of-Flight Cameras and Microsoft Kinect. *pringerBriefs in Electrical and Computer Engineering*. 2012:17-32.

52. Sell J, Oconnor P. The Xbox One System on a Chip and Kinect Sensor. *IEEE Micro*. 2014;34(2):44-53.

53. Microsoft. Buy Kinect for XBox One. *Microsoft Store*. 2017. Available at: https://www.microsoftstore.com/store/msusa/en_US/pdp/Kinect-Sensor-for-Xbox-One/productID.2267482500. Accessed May 25, 2017.

54. Microsoft. Kinect for Windows SDK 2.0. *Microsoft Download Center*. 2017. Available at: https://www.microsoft.com/en-us/download/details.aspx?id=44561. Accessed May 25, 2017.

55. Microsoft. Microsoft Research. *Kinect Sign Language Translator*. 2013. Available at: https://www.microsoft.com/en-us/research/blog/kinect-sign-language-translator-part-2/. Accessed September 01, 2017.

56. ISSEL. ISSEL. *Roki - Robto control using Microsoft Kinect*. 2012. Available at: https://www.youtube.com/watch?v=kECNyr7v0kM. Accessed September 01, 2017.

57. Jackson L. IGN. *NASA Uses Kinect And Oculus Rift To Control A Robotic Arm*. 2013. Available at: http://www.ign.com/articles/2013/12/31/nasa-uses-kinect-and-oculus-rift-to-control-a-robotic-arm. Accessed September 01, 2017.

58. Microsoft. Channel 9. *Kienct 3D = Fusion4D*. 2012. Available at: https://channel9.msdn.com/coding4fun/kinect/Kinect--3D--Fusion4D. Accessed September 03, 2017.

59. Jacob M, Wachs J, Packer R. Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images. *Journal of the American Medical Informatics Association*. 2013;20(e1).

60. Webster D, Celik O. Systematic review of Kinect applications in elderly care and stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation*. 2014;11(108).

61. Reflexion Health. Reflexion Health. *Virtual Exercise Rehabilitation Assistant*. 2017. Available

at: http://reflexionhealth.com/. Accessed September 01, 2017.

62. Noonan P, Howard J, Hallett W, Gunn R. Repurposing the Microsoft Kinect for Windows v2 for external head motion tracking for brain PET. *Physics in Medicine and Biology*. 2015;60(22):8753-8766.

63. Postawka A, Sliwinski P. A Kinet-Based Support System for Children with Autism Sprecturm Disorder. *International Conference on Artifical Intelligence and Soft Computing*. 2016;9693:189-199.

64. Uzuegbunam N, Wong WH, Cheung Sc, Ruble L. MEBook: Kinect-based self-modeling intervention for children with autism. *2015 IEEE International Conference on Multimedia and Expo*. 2015:1-6.

65. Varsanik J, Kimmel Z, de Moor C, Gabel W, Phillips G. Validation of an ambient measurement system (AMS) for walking speed. *Journal of Medical Engineering and Technology*. 2017;41(5):362-374.

66. Bauer S, Wasza J, Haase S, Marosi N, Hornegger J. Multi-modal Surface Registration for Markerless Initial Patient Setup in Radiation Therapy using Microsoft's Kinect Sensor. *IEEE International Conference on Computer Vision Workshops*. 2011:1175-1181.

67. Fisher T, Bligh M, Laurel T, Rasmussen K. 3D Extrusion Based Printing of Custom Bolus Using a Non-Invasive and Low Cost Method. *Medical Physics*. 2013;40(6Part15):280.

68. RSNA. RSNA Press Release. *Researchers Use Gaming Technology to Create Better X-Rays*. 2015. Available at: https://press.rsna.org/timssnet/media/pressreleases/PDF/pressreleasePDF.cfm?ID=1852. Accessed September 01, 2017.

69. Ernst F, Saß P. Respiratory motion tracking using Microsoft's Kinect v2 camera. *Current Directions in Biomedical Engineering*. 2015;1(1).

70. Xia J, Siochi RA. A real-time respiratory motion monitoring system using KINECT: Proof of concept. *Medical Physics*. 2012;39(5):2682.

71. Tahavori F, Alnowami M, Jones J, Elangovan P, Donovan E, Wells K. Assessment of Microsoft Kinect technology (Kinect for Xbox and Kinect for windows) for patient monitoring durign external beam radiotherapy. *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*. 2013.

72. Edmunds D, Bashforth S, Tahavori F, Wells K, Donovan E. The feasibility of using Microsoft Kienct v2 sensors during radiotherapy delivery. *Journal of Applied Clinical Medical Physics*. 2016;17(6):446-453.

73. Hendee W, Herman M. Improving patient safety in radiation oncology. *Practical Radiation Oncology*. 2011;1(1):16-21.

74. Chera B, Mazur L, Buchanan I. Improving Patient Safety in Clinical Oncology. *JAMA Oncology*. 2015;1(7):958.

75. Silverstein E, Snyder M. Implementation of facial recognition with Microsoft Kinect v2 sensor for patient verification. *Medical Physics*. 2017;44(6):2391-2399.

76. Microsoft. Kinect Hardware. *Windows Development Center*. 2014. Available at: https://developer.microsoft.com/en-us/windows/kinect/hardware. Accessed January 11, 2017.

77. Tico M, Kuosmanen P. Fingerprint matching using an orientation-based minutia descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2003;25(8):1009-1014.

78. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evalutation. *Caspian Journal of Internal Medicine*. 2013;4(2):627-635.

79. Solberg T, Balter J, Benedict S. Quality and safety considerations in stereotactic radiosurgery and stereotactic body radiation therapy: Executive summary. *Practical Radiation Oncology*. 2012;2(1):2-9.

80. Alight RT. Atfiscia. *Summary of Key Journal Papers: Align RT*. 2012. Available at: http://www.atfisica.com/site/upload/ficheros/key_journal_papers.pdf. Accessed August 01, 2017.

81. Silverstein E, Snyder M. Automatic Marker-Less Patient Motion Tracking Utilizing the Microsoft Kinect v2 Sensor. *Submitted to Journal of Applied Clinical Medical Physics*. 2017.

82. Microsoft. Channel 9. *Programming Kinect for Windows v2: (07) Advance Topics: Skeletal Tracking and Depth Filtering*. 2014. Available at: https://channel9.msdn.com/series/Programming-Kinect-for-Windows-v2/07. Accessed August 01, 2017.

83. Kalekar P, Rekhi K. Semantic Scholor. *Time series Forecasting using Holt-Winters Exponential Smoothing*. 2014. Available at: https://www.semanticscholar.org/paper/Time-series-Forecasting-using-Holt-Winters-Exponen-Kalekar-Rekhi/1eae285bd8d62eaf9dc37e8d2351845ba2903664. Accessed August 01, 2017.

84. Microsoft. Microsoft Developer Network. *Joint Smoothing code for C#*. 2014. Available at: https://social.msdn.microsoft.com/Forums/en-US/850b61ce-a1f4-4e05-a0c9-b0c208276bec/joint-smoothing-code-for-c?forum=kinectv2sdk. Accessed August 01, 2017.

85. Microsoft. Microsoft Developer Network. *Skeletal Joint Smoothing White Paper*. 2012.

Available at: https://msdn.microsoft.com/en-us/library/jj131429.aspx. Accessed August 01, 2017.

86. Budiharto T, Slagmolen P, Haustermans K. Intrafractional prostate motion during online image guided intensity-modulated radiotherapy for prostate cancer. *Radiotherapy and Oncology*. 2011;98(2):181-186.

87. White P, Yee C, Shan L, Chung L, Man N, Cheung Y. A comparison of two systems of patient immobilization for prostate radiotherapy. *Radiation Oncology*. 2014;9(1):29.

88. Mageras G, Pevsner A, Yorke E. Measurement of lung tumor motion using respiration-correlated CT. *International Journal of Radiation OncologyBiologyPhysics*. 2004;60(3):933-941.

89. Lim S, Golkar E, Rahni A. Respiratory Motion Tracking using the Kienct Camera. *IEEE Conference on Biomedical Engineering and Sciences*. 2014:8-1.

90. Silverstein E, Snyder M. Comparative Analysis of Respiratory Motion Tracking Using Microsoft Kinect v2 Sensor. *Submitted to Journal of Applied Clinical Medical Physics*. 2017.

91. Vedam S, Keall P, Kini V, Mostafavi H, Skukla H, Mohan R. Acquiring a four-dimensional computed tomography dataset using an external respiratory signal. *Physics in Medicine and Biology*. 2002;48(1):45-62.

92. Wink N, Chao M, Antony J, Xing L. Individualized gating windows based on four-dimensional CT information for respiration-gated radiotherapy. *Physics in Medicine and Biology*. 2007;53(1):165-175.

93. Li H, Noel C, Garcia-Ramirez J. Clinical evaluations of an amplitude-based binning algorithm for 4DCT reconstruction in radiation therapy. *Medical Physics*. 2012;39(2):922-932.

94. Kruis M, Kamer J, Belderbos J, Sonke JJ, Herk M. 4D CT amplitude binning for the generation fo a time-averaged 3D mid-position CT scan. *Physics in Medicine and Biology*. 2014;59(18):5517-5529.

95. Altman D, Bland J. Measurement in Medicine: The Analysis of Method Comparison Studies. *The Statistician*. 1983;32(3):307.

96. Giavarina D. Understanding Bland Altman analysis. *Biochemia Medica*. 2015;25(2):141-151.

ABSTRACT

INVESTIGATION OF THE MICROSOFT KINECT V2 SENSOR AS A MULTI-PURPOSE DEVICE FOR A RADIATION ONCOLOGY CLINIC

by

EVAN ASHER SILVERSTEIN

December 2017

Advisor: Dr. Michael G. Snyder

Major: Medical Physics

Degree: Doctor of Philosophy

For a radiation oncology clinic, the number of devices available to assist in the workflow for radiotherapy treatments are quite numerous. Processes such as patient verification, motion management, or respiratory motion tracking can all be improved upon by devices currently on the market. These three specific processes can directly impact patient safety and treatment efficacy and, as such, are important to track and quantify. Most products available will only provide a solution for one of these processes and may be outside the reach of a typical radiation oncology clinic due to time or cost of implementation and incorporation with already existing hardware. This manuscript investigates the use of the Microsoft Kinect v2 sensor to provide solutions for all three processes all while maintaining a relatively simple and easy to use implementation.

To assist with patient verification, the Kinect system was programmed to create a facial recognition and recall process. The basis of the facial recognition algorithm was created by utilizing a facial mapping library distributed by Microsoft within the Software Developers Toolkit (SDK). Here, the system extracts 31 fiducial points representing various facial

landmarks. 3D vectors are created between each of the 31 points and the magnitude of each vector is calculated by the system. This allows for a face to be defined as a collection of 465 specific vector magnitudes. The 465 vector magnitudes defining a face are then used in both the creation of a facial reference data set and subsequent evaluations of real-time sensor data in the matching algorithm. To test the algorithm, a database of 39 faces was created, each with 465 vectors derived from the fiducial points, and a one-to-one matching procedure was performed to obtain sensitivity and specificity data of the facial identification system.

In total, 5299 trials were performed and threshold parameters were created for match determination. Optimization of said parameters in the matching algorithm by way of ROC curves indicated the sensitivity of the system was 96.5% and the specificity was 96.7%. These results indicate a fairly robust methodology for verifying, in real-time, a specific face through comparison from a pre-collected reference data set. In its current implementation, the process of data collection for each face and subsequent matching session averaged approximately 30 seconds, which may be too onerous to provide a realistic supplement to patient identification in a clinical setting. Despite the time commitment, the data collection process was well tolerated by all participants. It was found that ambient light played a crucial role in the accuracy and reproducibility of the facial recognition system. Testing with various light levels found that ambient light greater than 200 lux produced the most accurate results. As such, the acquisition process should be setup in such a way to ensure consistent ambient light conditions across both the reference recording session and subsequent real-time identification sessions.

In developing a motion management process with the Kinect, two separate, but complimentary processes were created. First, to track large scale anatomical movements, the automatic skeletal tracking capabilities of the Kinect were utilized. 25 specific body joints (head, elbow, knee, etc) make up the skeletal frame and are locked to relative positions on the body. Using code written in C#, these joints are tracked, in 3D space, and compared to an initial state of the patient allowing for an indication of anatomical motion. Additionally, to track smaller, more subtle movements on a specific area of the body, a user drawn ROI can be created. Here, the depth values of all pixels associated with the body in the ROI are compared to the initial state. The system counts the number of live pixels with a depth difference greater than a specified threshold compared to the initial state and the area of each of those pixels is calculated based on their depth. The percentage of area moved (PAM) compared to the ROI area then becomes an indication of gross movement within the ROI.

In this study, 9 specific joints proved to be stable during data acquisition. When moved in orthogonal directions, each coordinate recorded had a relatively linear trend of movement but not the expected 1:1 relationship to couch movement. Instead, calculation of the vector magnitude between the initial and current position proved a better indicator of movement. 5 of the 9 joints (Left/Right Elbow, Left/Right Hip, and Spine-Base) showed relatively consistent values for radial movements of 5mm and 10mm, achieving 20% - 25% coefficient of variation. For these 5 joints, this allowed for threshold values for calculated radial distances of 3mm and 7.5 mm to be set for 5mm and 10mm of actual movement, respectively. When monitoring a drawn ROI, it was found that the depth sensor had very little sensitivity of movement in the X (Left/Right) or Y (Superior/Inferior) direction, but exceptional sensitivity in the Z (Anterior/Posterior) direction. As such, the PAM values could only be coordinated with motion in the Z direction. PAM values of over 60% were shown to be indicative of movement in the Z direction equal to that of the threshold value set for movement as small as 3mm.

Lastly, the Kinect was utilized to create a marker-less, respiratory motion tracking system. Code was written to access the Kinect's depth sensor and create a process to track the respiratory motion of a subject by recording the depth (distance) values obtained at several user selected points on the subject, with each point representing one pixel on the depth image. As a patient breathes, a specific anatomical point on the chest/abdomen will move slightly within the depth image across a number of pixels. By tracking how depth values change for a specific pixel, instead of how the anatomical point moves throughout the image, a respiratory trace can be obtained based on changing depth values of the selected pixel. Tracking of these values can then be implemented via marker-less setup. Varian's RPM system and the Anzai belt system were used in tandem with the Kinect in order to compare respiratory traces obtained by each using two different subjects.

Analysis of the depth information from the Kinect for purposes of phase based and amplitude based binning proved to be correlated well with the RPM and Anzai systems. IQR values were obtained which compared times correlated with specific amplitude and phase percentage values against each product. The IQR spans of time indicated the Kinect would measure a specific percentage value within 0.077 s for Subject 1 and 0.164s for Subject 2 when compared to values obtained with RPM or Anzai. For 4D-CT scans, these times correlate to less than 1mm of couch movement and would create an offset of one half an acquired slice. These minimal deviations between the traces created by the Kinect and RPM or Anzai indicate that by tracking the depth values of user selected pixels within the depth image, rather than tracking specific anatomical locations, respiratory motion can be tracked and visualized utilizing the Kinect with results comparable to that of commercially available products.

AUTOBIOGRAPHICAL STATEMENT
EVAN ASHER SILVERSTEIN

EDUCATION

2012-2017	Ph.D. Medical Physics	Wayne State University
2009-2010	M.S. Applied Physics	CSU Long Beach
2007-2008	B.S. Astrophysics	University of California, Irvine

CERTIFICATIONS

2013	American Board of Radiology: Part 1

PROFESSIONAL APPOINTMENTS

2017-2019	Medical Physics Resident	Mayo Clinic, AZ
2016	Assistant Lecturer	Wayne State University
2013-2016	Graduate Research Assistant	Wayne State University

ACADEMIC AND PROFESSIONAL MEMBERSHIPS

2012-Present	American Association of Physicists in Medicine
2013-Present	AAPM – Great Lakes Chapter
2015-Present	AAPM – Working Group on Students and Trainee Research
2010-Present	American Physical Society

PUBLICATIONS

1. **Silverstein, E**, Snyder, M. Implementation of facial recognition with Microsoft Kinect v2 sensor for patient verification. *Medical Physics*. 2017; 44(6):2391-2399.
2. **Silverstein, E**, Snyder, M. Automatic Marker-Less Patient Motion Tracking Utilizing the Microsoft Kinect v2 Sensor. **Pending Approval:** *Journal of Applied Medical Physics.*
3. **Silverstein, E**, Snyder, M. Comparative Analysis of Respiratory Motion Tracking Using Microsoft Kinect v2 Sensor. **Pending Approval:** *Journal of Applied Medical Physics.*
4. **Silverstein, E**, Burmeister, J, Fullerton, G. SDAMPP Student Guide to a Medical Physics Career. *SDAMPP Reports*. 2016
5. Gredig T, **Silverstein E,** Byrne M. Height-Height Correlation Function to Determine Grain Size in Iron Phthalocyanine Thin Films. *Journal of Physics: Conference Series*. 2013; 417:012069
6. Gredig T, Werber M, Guerra JL, **Silverstein EA**, Byrne MP, Cacha BG. Coercivity Control of Variable-Length Iron Chains in Phthalocyanine Thin Films. *Journal of Superconductivity and Novel Magnetism.* 2012;25(7):2199-2203