Repurposing the Microsoft Kinect for Windows v2 for external head motion tracking for brain PET

P J Noonan¹, J Howard², W A Hallett² and R N Gunn¹,²

¹ Department of Medicine, Imperial College London, Hammersmith, London, UK
² Imanova Imaging Centre, Hammersmith, London, UK

E-mail: p.noonan@imperial.ac.uk

Received 20 May 2015, revised 21 September 2015
Accepted for publication 23 September 2015
Published 3 November 2015

Abstract

Medical imaging systems such as those used in positron emission tomography (PET) are capable of spatial resolutions that enable the imaging of small, functionally important brain structures. However, the quality of data from PET brain studies is often limited by subject motion during acquisition. This is particularly challenging for patients with neurological disorders or with dynamic research studies that can last 90 min or more. Restraining head movement during the scan does not eliminate motion entirely and can be unpleasant for the subject. Head motion can be detected and measured using a variety of techniques that either use the PET data itself or an external tracking system. Advances in computer vision arising from the video gaming industry could offer significant benefits when re-purposed for medical applications. A method for measuring rigid body type head motion using the Microsoft Kinect v2 is described with results presenting \( \leq 0.5 \text{ mm} \) spatial accuracy. Motion data is measured in real-time at 30 Hz using the KinectFusion algorithm. Non-rigid motion is detected using the residual alignment energy data of the KinectFusion algorithm allowing for unreliable motion to be discarded. Motion data is aligned to PET listmode data using injected pulse sequences into the PET/CT gantry allowing for correction of rigid body motion. Pilot data from a clinical dynamic PET/CT examination is shown.

Keywords: positron emission tomography, brain imaging, microsoft kinect v2, motion tracking

(Some figures may appear in colour only in the online journal)
1. Introduction

Subject motion has long been recognised as a limiting factor in medical imaging procedures, and remains a largely unsolved problem, leading to data inaccuracies that impact on costs and effective diagnosis/treatment. This presents a particular challenge for positron emission tomography (PET) brain imaging of patients with neurodegenerative disorders using the latest generation of high resolution PET scanners. Algorithms to correct for motion are well established, yet the lack of effective, affordable, reliable motion tracking hardware has prevented widespread adoption in both research and clinical settings.

Extensive literature exists describing various data driven techniques that aim to derive motion parameters directly from the PET data itself, such as automatic image registration (Woods et al 1992) or mutual information (Collignon et al 1995, Wells et al 1996). Data driven techniques that use multiple acquisition frames (Picard and Thompson 1997) are excellent when the subject motion consists only of short movements separated by long periods of rest since it is possible to reframe the PET data to reduce the effect of inter-frame motion. If subject motion consists of gradual drifts, or rapid and frequent displacements then generally external motion tracking offers a more suitable solution due to the potential for high sampling frequency (>30 Hz) and high spatial sensitivity (<1 mm).

Depth sensing devices such as the 3dMD (3dMD Ltd, London, UK), AlignRT (VisionRT Ltd, London, UK), and Polaris Spectra and Vicra Position Sensors (NDI Ontario, Canada) have been adapted for motion tracking in medical imaging and radiotherapy (Lopresti et al 1999, Schöffel et al 2007, Peng et al 2010). More recently, a number of consumer grade depth sensors have been released that offer a number of advantages in terms of cost, and performance. In principle they eliminate the need to attach markers or tracking tools to the subject that can slip relative to the subject leading to failure of motion tracking. In particular the Microsoft Kinect, a small, low cost, infrared (IR) based depth sensor, has been applied in many medical applications such as gait analysis (Stone and Skubic 2013) or fall detection (Mastorakis and Makris 2014).

Four currently available consumer grade depth sensors are listed in table 1. Of these, the Kinect v1 uses structured light (SL) and the others use time of flight (ToF) to measure depth information. Descriptions of SL and ToF depth sensing technology can be found in Lindner et al (2010) and Khoshelham and Elberink (2012).

In a previous paper, we investigated the Kinect v1 as a markerless based motion tracking system for brain PET (Noonan et al 2012). The Kinect v1 was able to measure the rigid body motion of a polystyrene mannequin phantom to comparable accuracy to the Polaris Vicra Position sensor. Tracking real subjects with the Kinect v1 was unreliable due to the non-rigid parts of the face, such as the mouth and jaw, being included in the tracking algorithm. This issue was confounded by the Kinect’s decrease in depth sampling resolution as a function of distance to the sensor and the 0.5 m minimum operating distance of the Kinect v1.

The Kinect v2 was released in July 2014 and represents a significant improvement over the Kinect v1 sensor. This paper describes modifications to the v2 sensor for subject motion tracking in the routine clinical PET/CT environment at an operating range of 10–50 cm. We describe methods to rigidly position the Kinect v2 in the PET/CT scanner, and synchronise the motion tracking data to the PET listmode event data. To validate the system, experiments were undertaken to demonstrate the accuracy, stability, sensitivity, and robustness of the proposed real-time motion tracking system. We present data demonstrating 0.44 mm and 0.2° root mean square error compared to digital caliper and protractor measurements. We also propose a method for the identification and removal of any unreliable motion data caused by non-rigid facial movements. Finally, we present motion data from a 90 min
clinical PET/CT scan where even the small $\leq 1$ mm motion of the head due to breathing is resolved.

2. Materials

2.1. Kinect V2

Both the Microsoft Kinect v1 and v2 were originally used as video game input devices to measure the user’s body positions. They perform body tracking on the 16 bit depth data which each camera returns at 30 frames per second. To measure depth the Kinect v1 emits a static pseudo-random structured light pattern of speckled dots of IR light. Three-dimensional (3D) IR opaque structures interact with the emitted pattern and shift the reflected dots relative to a calibrated position dependent on the distance of the object to the Kinect. The standard operating range of the Kinect v1 is 0.5–4.0 m with the closest distance limited by the ability of the IR sensor to resolve different speckle points. The Kinect v2 uses three phases of modulated IR light and a TOF principle to measure the distance to surfaces. Similar to the Kinect v1, the Kinect v2 has a standard minimum operating distance of 0.5 m, limited by saturation of the IR sensor by the reflected IR light. A major difference between Kinect v2 compared to the first generation is that a depth measurement is obtained directly for each pixel in the image. For the Kinect v1, depth has to be interpolated between two points of the speckle pattern. Theoretically, this allows the Kinect v2 to have a much larger range of depth than the Kinect v1 as the optics of the IR camera can be changed to sample a specific region of space at a specific distance from the sensor. Section 3.1 describes the modifications performed to enable the Kinect v2 to be used inside a clinical PET/CT scanner where a range of $\leq 200$ mm is required.

2.2. KinectFusion

KinectFusion (Newcombe et al 2011) is an algorithm developed by Microsoft Research Cambridge and is available in the official Software Developer Kit. KinectFusion is a fast iterative closest point (ICP) algorithm that uses the parallel processing power of a general purpose graphics processing unit (GPU) to align sequential depth frames into a single volume. This can be used to build a 3D model or template of an object or scene by moving the Kinect relative to the static object or scene. At the Kinect frame rate of 30 Hz, there is generally not much difference in perspective between sequential frames and the ICP algorithm only has to iterate $\leq 7$ times to converge to the transformation required to register the new frame to the existing template. Using a modern GPU, a frame can be processed and integrated into the volume within the 33 ms before a new Kinect frame is available, resulting in real-time functionality. For these studies, a gaming grade laptop with a 2.7 GHz Intel core i7 3820 QM and a 4 GB Nvidia GTX 680 m GPU was used.
KinectFusion is mainly used for scanning the 3D structure of static objects, however it can also be used to measure rigid body motion since for successful integration of a new depth frame into the volume, KinectFusion requires the knowledge of the relative position of the Kinect to the scene. This method to measure rigid body motion is insensitive to skin tone and lighting conditions, and uses dense ICP, i.e. all the available depth points are used rather than a subset, in the ICP registration. Section 3.4 describes the application of KinectFusion to obtain the rigid body head motion of a subject.

3. Modifications

The following sections describe the hardware and software modifications performed to repurpose the consumer grade Kinect v2 depth sensor in conjunction with KinectFusion for head motion tracking in clinical brain PET.

3.1. Near mode

As mentioned in section 2 the standard configuration of the Kinect v2 has a minimum operating distance of 0.5 m. This prevents Kinect v2 from directly viewing the 3D facial features needed to perform ICP based KinectFusion tracking when the subject is within the PET scanner bore. In our previous work, a mirror was used to reflect the structured light pattern onto the subject’s face (Noonan et al 2012). A front surfaced mirror would be required for Kinect v2 to prevent multiple reflections degrading the ToF depth information.

We obtained a developmental ‘near mode’ firmware upgrade through the Kinect for Windows v2 Developer Preview Program which lowered the intensity of the emitted IR laser light so that closer objects did not saturate the sensor. This also required a specific Windows service executable to allow the Kinect v2 to return valid depth values below 0.5 m, as without this service, these values would be null. As an alternative way to enable near mode without requiring a firmware update, an IR neutral density filter was used to reduce the light output of the IR emitters. Since Kinect v2 is designed to operate over a range of 0.5–8.0 m, the IR lens is out of focus at distances less than 0.4 m. The sensor was refocused for near mode by increasing the distance from the lens to the sensor array. This also required a recalibration to determine the new intrinsic parameters of the modified IR sensor. A checkerboard pattern was imaged at 20 different poses and openCV (Bradski 2000) was used to calculate the camera calibration matrix of the sensor using algorithms based on Zhang (2000). The near mode camera intrinsic parameters are then used in KinectFusion to enable correct depth estimation. Figure 1 shows two KinectFusion scans of an eye before and after refocussing and recalibrating for near mode.

3.2. Scanner mount

In order to mount the Kinect v2 in the scanner environment, a tension ring of 5 mm thick acrylic was fitted inside the scanner bore. The Kinect v2 was attached to the top of the ring using a quick release camera mount adapter. This allowed the Kinect to be held securely in the optimal position for tracking without the Kinect or tension ring encroaching into the PET or CT field of view and without any modification to the scanner, as is shown in figure 2. For the Siemens HiRez and TrueV Biograph PET/CT scanners, the microphone recess was used to feed the Kinect data cable out of the scanner without entering either the PET or CT field of view.
3.3. Temporal alignment triggering

Temporal alignment of the motion tracking data to the PET data is essential for motion correction. Techniques to achieve this have been implemented previously, from comparing file time stamps to injecting trigger gates into the PET listmode (Bloomfield et al 2003). In this work we used an Arduino microcontroller to inject 5V TTL level pulses into the PET/CT gantry gate ports. The Arduino is connected to the computer processing the Kinect data via a USB 2 port. A De Bruijn coded sequence (de Bruijn 1946) using an alphabet of numbers 1–9 was used to create a unique sequence. Every 300 frames of Kinect data a value from the sequence is written to the Kinect data file and the PET listmode via the Arduino and the gantry gate ports. The listmode can be scanned for the gate tags which can be corresponded to the non-repeating De Bruijn sequence in the Kinect data.

3.4. Pre-processing of raw depth data

At high contrast boundaries in the Kinect v2 depth data between foreground and background regions we observed a ‘flying pixel’ noise effect, common to many ToF depth sensors.
This effect was more pronounced when the v2 was modified for near mode, so that the accuracy of the KinectFusion ICP registration reduced as the integrated volume became dominated by noise. This effect can be ameliorated by masking regions in the incoming depth frame that contain boundaries. This is not ideal as it also masks valid depth values and boundaries may enter the field of view with large movements.

In order to remove artefacts from the raw depth data in near mode we implemented an experimental 3D data filter developed by the Microsoft Kinect for Windows team which performs filtering in real time on each new raw depth frame by using a 3D spatial kernel that removes pixels that are more than a set distance from other surfaces. The effect of using a 3D spatial filter on the depth data and a KinectFusion scan is shown in figure 3.

3.5. Global frame of reference

To monitor any motion of the Kinect v2 relative to the scanner during operation, a square marker was attached to the PET/CT gantry and its position was measured using the 1920 × 1080 resolution colour camera and the Perspective-n-Point (PnP), algorithm in the Aruco (Garrido-Jurado et al 2014) and openCV libraries. The PnP problem can be used to estimate the pose of a flat marker of known size using a single camera. Solutions to PnP use point correspondences between the 3D points of the marker corners and their projections onto the image plane of a calibrated camera. In the case of a square marker \( n = 4 \) and the transformation of the marker in 3D can be estimated using an iterative cost function. The marker tracking can be seen in figure 2.

3.6. Spatial alignment calibration

In this work a threshold was applied to the CT data to create a single isosurface mesh representing the subject’s skin surface, which was then rigidly aligned to the KinectFusion generated point cloud surface mesh using ICP. The transformation matrix between KinectFusion space and CT space can then be applied to the Kinect measured transformations to define them in the CT coordinate system.
4. Methods

As described in section 1, our previous work with the Kinect v1 demonstrated that it was capable of measuring the rigid body motion of a rigid head phantom to within 1 mm of the measurements provided by the Polaris Spectra Position Sensor. The specular reflectivity of the polystyrene phantom caused artifacts in the Kinect v2 data so the phantom was replaced with a skull phantom. The skull phantom was manufactured using a powder bed and inkjet head 3D printing process. The printer used gypsum plaster that formed a Lambertian surface which was imaged well by the Kinect v2. In this work we sought to verify that the KinectFusion algorithm applied to data from the Kinect v2 was also capable of at least the same accuracy. However comparing measurements from the Kinect v2 and the Polaris Sensor is difficult to achieve as the near infra-red light from each sensor can confuse the other. It could be possible to measure the discrete positions of the phantom by covering the IR sources of each device in turn. This method would allow for realistic, complex transformations to be applied and measured which contain both translations and rotations.

In Wiles et al (2004) a passive tool tracked by the Polaris contains positional errors of 0.23 mm and rotational errors of 0.38°. With this rotational error, a point at 100 mm distance from the tracking tool will include an uncertainty of 0.66 mm. Therefore it was decided that the Polaris was not a suitable tool to compare the accuracy of another measuring device. Rather we used a linear motion (LM) guide to move the phantom known distances measured with high precision digital callipers. To measure rotational accuracy, the phantom was securely fixed to a milling machine high precision rotating table and a digital protractor was used to measure the angle of the table to within 0.1°. These techniques are able to precisely measure the applied motion however it is acknowledged that the motion is not realistic for head motion as it is constrained to single dimension translations and in-plane rotations.

In the following experiments, KinectFusion is used to generate a template of the object being scanned, either phantom or subject. This process involves moving the object relative to the Kinect so that KinectFusion integrates depth data from multiple view positions to build a model of the object without holes or missing data caused by occlusions from any one single view point. After manual assessment of the quality of the template, integration of new depth data is halted, and the template is saved to disk.

4.1. Comparing calliper and Kinect measured translations

The phantom was securely fastened to a rigid platform attached to the LM guide. Firstly, this was crudely orientated along the optical axis of the Kinect (z) and secondly, transaxial (x, y) at a perpendicular distance of 170 mm, the expected distance between the Kinect and the subject in the PET/CT scanner.

For the axial motion experiment, the template was generated with the phantom in the centre of the depth of focus at a distance of approximately 170 mm. 21 positive and negative displacements from this position along the LM guide were manually applied and measured using the callipers and KinectFusion. For the transaxial motion experiment, a new template was generated and eight positive and negative manual translations were applied over a 55 mm range to cover the transaxial field of view for the phantom at a distance of 170 mm. Single measurements relative to the time the template was finalised were taken with the callipers at each point and only a single time point in the Kinect data was used for each measurement position.
4.2. Comparing protractor and kinect measured rotations

Manual in-plane rotations were applied to the horizontally positioned rotating table over a range of 45° at the same distance of 170 mm used for transaxial linear motions. The Kinect was raised above the height of the table to enable an unobstructed view of the phantom. Similarly to section 4.1 single measurements relative to the template were recorded from the digital protractor and from single time points in the Kinect data corresponding to each angle.

4.3. Static phantom measurements

Measurement stability was assessed by tracking the position of a stationary phantom for 90 min. A template was generated by slightly rotating the phantom relative to the stationary Kinect v2 sensor. The displacement of a point on the surface of the phantom was measured relative to its starting position, prior to the generation of the template. The experiments in Lachat et al (2015) present data showing that the depth data from the Kinect v2 drifts during the initial 40 min from powering on, suggesting the Kinect v2 requires a ‘pre-heating’ time before reliable data is obtained. The stationary phantom was monitored for an additional 90 min directly preceding the first experiment. A new template was generated at the start of the second 90 min scan. All the following experiments in this paper were performed with a Kinect v2 that had been powered on for at least 60 min before data acquisition.

4.4. Using alignment energy for estimating occurrence of non-rigid body motion

Due to the close proximity of the Kinect v2 to the face of the subject, it is both possible and advantageous to only view and measure the motion of the more rigid, upper parts of the face. KinectFusion reports an alignment energy (AE) after every registered frame of depth data, which indicates how successfully the new depth frame has been registered to the template. AE is stated in Newcombe et al (2011) as the global point-plane energy between the vertex points in the current depth frame point cloud and the rigid global model. It is suggested that this metric can be used to indicate the reliability of each estimated pose, since it increases when the skin deforms non-rigidly compared to the rigid global model.

An experiment was performed with two volunteers where each participant was positioned on the PET/CT scanner bed using the normal procedures for securing the head during scanning, using foam padding and a forehead strap. The volunteers were asked to remove their spectacles (if applicable) and their hair was swept away from the forehead. They were asked to try to keep their head in a fixed position throughout the monitoring session.

After an initial period of inactivity for 30 s, the volunteer was prompted to talk normally for 30 s, whilst aiming to keep their head stationary in the head rest. After this period the volunteer was then asked to keep silent and stationary for another 30 s. Following this, the volunteer was asked to frown and grimace to distort the skin around the eyes and forehead for 30 s. Finally, the volunteer was asked to keep still for 30 s.

Throughout the experiment the subject’s head was tracked using the Kinect v2 in head tracking position, with the input depth frame masked so that only a $10 \times 10 \text{cm}^2$ region centred over the right eyebrow was used.

4.5. Clinical motion tracking data

The Kinect v2 was fitted inside a Siemens HiRez Biograph 6 PET/CT scanner using the tension ring. The Arduino controlled pulse generator was attached to the gating signal.
inputs on the PET/CT gantry and to the motion tracking acquisition PC. The Kinect v2 was powered on 90 min before the scan to warm up. A subject undergoing a 90 min dynamic PET scan with the 5-HT2A ligand [11C]-CIMBI-36 (Ettrup et al 2010) was tracked using the Kinect v2 and the tracking data and alignment energy were recorded at 30 frames per second. The motion of a point on the bridge of the nose was calculated using the rigid body motion data.

Finally, a comparison was made between the displacement of the bridge of the nose as calculated by the Kinect v2 and by the PET data driven derived motion parameters from the Mutual Information (MI) image coregistration routine in SPM (The FIL Methods Group 2015). The PET data was reconstructed into 26 frames and the MI routine was used to estimate the transformations between each frame and a reference frame. The $x$, $y$, $z$ position of the bridge of the nose as transformed by each Kinect measurement was compared to the position of the bridge of the nose as transformed by the corresponding frame's MI transformation. The RMSE values for $x$, $y$, and $z$ was calculated for the entire 90 min data set.

5. Results

5.1. Comparing calliper and kinect measured transformations

Figure 4(a) plots the measured position of the phantom using the near mode Kinect v2 compared to accurate measurements with digital callipers, as it was moved 120 mm axially on a LM guide. The region 140–210 mm contains small, sub mm, differences between calliper and Kinect measured translations, however these differences quickly increase outside this region. For the transaxial experiment, where the phantom was moved over 55 mm at 170 mm depth, the RMSE between the calliper and Kinect measured translations was 0.46 mm.
5.2. Comparing protractor and kinect measured rotations

To measure the rotational accuracy of the Kinect v2, 17 in-plane rotations over a range of 40° were manually applied using a precise rotating table and were measured using a digital protractor. The data is shown in figure 4(b) and the RMSE was calculated to be 0.2°.

5.3. Static phantom measuring

The Kinect v2 measured the position of a static phantom for two consecutive 90 min sessions. Plots of the measured position of the phantom are shown in figure 5. Drift occurs in the initial 45 min as the Kinect warms up. It is believed that the activity observed at 19–25 min is caused by the Kinect v2 fan turning on and altering the thermal properties of the Kinect v2. Alignment Energy (AE) also increases as the template that was created at the start of the scan becomes less valid as the depth data converges to a steady state. The second scan immediately follows the first and shows the step decrease in AE relative to the preceding scan, which remains constant for the next 90 min. The standard deviation of the x, y, z measured positions in figure 5(b) was 0.13, 0.14, 0.31 mm respectively.

5.4. Using alignment energy for estimating occurrence of non-rigid body motion

Figure 6 shows the 2.5 minute tracking data from two volunteers alternating between 30 s periods of no motion, talking, and grimacing. Generally, the KinectFusion measured position of the volunteer’s head remains constant during the static and talking sections. Figure 6(a) shows more apparent motion during talking than figure 6(b), and both show large apparent motions during grimacing. During these periods however the Alignment Energy is elevated or spikes exist indicating that the motion data at those time points is unreliable.

5.5. Clinical motion tracking data

The motion plot from a 90 minute [11C]-CIMBI-36 scan showing the displacement of the nose bridge on the surface of the subject’s skin is shown in figure 7. A zoomed-in section of the motion data is shown in figure 7(b) where the high sensitivity of the tracking system is able to observe the sub-mm motion of the head caused by breathing.
Figure 6. For both sessions; 0–0.5 min: stationary, 0.5–1.0 min: talking, 1.0–1.5 min: stationary, 1.5–2.0 min: grimacing, 2.0–2.5 min: stationary. The translation of a point on the face is plotted in $x$, $y$, $z$, additionally the alignment energy is plotted on the right hand axis. (a) Shows that talking affects the Kinect v2 measured motion of the face, whereas reduced motion is measured during talking in (b). Both sets of data show that the tracking data is affected by grimacing. The alignment energy greatly increases from a steady baseline during grimacing in both volunteers and is elevated during talking for (c).

Figure 7. The Kinect (solid line) and MI (dashed line) transformations, of the nose bridge for the 90 min duration of the $[^{11}C]$-CIMBI-36 PET/CT scan is shown in (a) and a short section where a complex series of movements occurred is shown in (b).
The RMSE ± standard deviation between the position of the nose bridge as measured by the Kinect and MI image registration was 1.49 ± 1.43, 2.13 ± 1.38, and 1.62 ± 1.57 mm in x, y, and z, respectively.

6. Discussion

Markerless motion tracking is an active research area for brain PET due to a combination of the lack of clinically suitable solutions offered by marker based techniques, and the need for higher spatio-temporal resolution than is currently possible using data driven methods. Other markerless based tracking systems proposed for use in research PET offer excellent spatial resolution (Olesen et al 2013, Kyme et al 2014). Despite the relative low cost of these systems compared to existing motion tracking equipment they do not match the consumer grade cost and off-the-shelf availability of the Kinect. This ease of access is beneficial as it results in vast amounts of Kinect based code, such as KinectFusion, that is written by Microsoft, academia, and the open source community.

The Kinect v1 could track objects to comparable accuracy and sensitivity as the Polaris Spectra Position Sensor, however we found it was unable to operate sufficiently reliably in a clinical setting. The Kinect v2 has a similar standard operating range and minimum distance to the Kinect v1, however we have shown that it is possible to reduce this range to 0.1–1.0 m by modifying the sensor optics. This allowed the Kinect v2 to be fitted inside the PET/CT gantry enabling direct imaging of the face. Besides improving the resolution, fitting the Kinect v2 inside the gantry reduces the likelihood of it being knocked and misaligned and ensures the line of sight cannot be inadvertently obscured by someone moving between it and the subject.

In near mode, ToF artefacts such as ‘flying pixels’ became more prevalent in the depth data which can degrade the quality of the KinectFusion registration. The ‘flying pixels’ were successfully removed from the depth frame data using a 3D spatial filter.

The Kinect v2 should be turned on at least 60 min before the start of the PET scan to allow for the unit temperature to stabilise. The effect of temperature on depth sensors has been noted before with the Kinect v2 (Lachat et al 2015) and other 3D ToF sensors (Kahlmann and Ingensand 2006). These papers suggest that the process is due to the shape of the IR pulse emitted by the IR LEDs changing due to the temperature of the LED. The Kinect v2 has an active fan and substantial heat sinks and appears to thermally stabilise after 40–60 min of use. The increasing jitter seen in figure 5(a) is a result of the template generated at the start of the scan no longer representing the surface as being observed in new depth frames. In clinical practice this means that the Kinect v2 should be incorporated into the morning daily quality control procedures. This could involve using a phantom of known shape and a previously obtained template when the Kinect was at operating temperature. In this case the AE reported by KinectFusion will begin high and reduce as the Kinect warms.

The use of the KinectFusion algorithm enables the tracking data to be obtained in real-time at 30 frame per second, as the processing and registration of each new depth frame is achieved in under 33 ms. Using KinectFusion with Kinect v2 operating at 140–210 mm from the face, the position and angular pose of the face can be measured to within 0.5 mm and 0.2°. These experiments have shown that at least 3 of the 6 degrees of freedom of the pose estimation of the phantom can be measured using the modified Kinect v2. To fully evaluate the tracking capability of the system a robot arm could be used to accurately and reliably drive the motion of the phantom.

Alignment energy is a potential indicator of the occurrence of non-rigid surface deformation and therefore unreliable tracking data. By thresholding the AE signal for large gradients it
may be possible to obtain a criteria for reliable rigid motion data. Investigations on the talking and grimacing of volunteers, as well as the clinical data, show that the AE is at a constant level until non-rigid surfaces are detected. In figure 7(a) AE remained locally constant even during periods where the subject was moving (0–30 min). This is consistent with AE being the point to plane error metric for new depth data compared to the template. The spikes seen in the AE appear to correlate with non rigid motion definitely occurring in the grimacing sections of figure 6 and when there was a high chance of non rigid motion occurring in figure 7 as the spikes in this clinical data set temporally align well with the acquisition arterial blood samples. More investigations into AE is required with either phantoms with non rigidly deforming surfaces, or by repeating the volunteer experiment with a stereotactic frame.

Initial examination of the motion plots from the clinical subject appear promising, with even the sub-mm motion due to the breathing cycle clearly resolved demonstrating the tracking system’s sensitivity. The agreement with the MI motion parameters is encouraging, and we will proceed to validate on a larger cohort of subjects by correcting the PET data for the Kinect measured motion and assessing the impact on outcome measures of interest.

7. Conclusion

The need to develop accurate and reliable subject head motion tracking and correction is urgent with the increasing use of imaging in research into neurodegenerative diseases using high resolution scanners. The work presented in this paper demonstrates that, with some modifications, the Kinect v2 can be successfully used as a motion tracking device for brain PET.

References

Bloomfield P M, Spinks T J, Reed J, Schnorr L, Westrip A M, Livieratos L, Fulton R and Jones T 2003 The design and implementation of a motion correction scheme for neurological PET Phys. Med. Biol. 48 959–78
Bradski G 2000 The OpenCV Library Dr. Dobb’s J. Softw. Tools 2236121
Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P and Marchal G 1995 Automated multimodality image registration based on information theory Inf. Process. Med. Imaging 5 263–74
de Bruijn N G 1946 A combinatorial problem Proc. K. Ned. Akad. Wet. 49 758–64
Ettrup A et al 2010 Radiosynthesis and evaluation of 11C-CIMBI-5 as a 5-HT2A receptor agonist radioligand for PET J. Nucl. Med. 51 1763–70
Garrido-Jurado S, Muoz-Salinas R, Madrid-Cuevas F and Marr-Jimnez M 2014 Automatic generation and detection of highly reliable fiducial markers under occlusion Pattern Recognit. 47 2280–92
Kahlmann T and Ingensand H 2006 Calibration of the fast range imaging camera swissranger for use in the surveillance of the environment Proc. SPIE 6396 639605
Khoshelham K and Elberink S O 2012 Accuracy and resolution of kinect depth data for indoor mapping applications Sensors 12 1437–54
Kyme A, Se S, Meikle S, Angelis G, Ryder W, Popovic K, Yatigammana D and Fulton R 2014 Markerless motion tracking of awake animals in positron emission tomography IEEE Trans. Med. Imaging 33 2180–90
Lachat E, Macher H, Mittel M, Landes T and Grussenmeyer P 2015 First experiences with kinect v2 sensor for close range 3d modelling ISPRS—Int. Arch. Photogramm. Remote Sens. 1 93–100
Lindner M, Schiller I, Kolb A and Koch R 2010 Time-of-flight sensor calibration for accurate range sensing Comput. Vis. Image Underst. 114 1318–28
Lopresti B J, Russo A, Jones W F, Fisher T, Crouch D G, Altenburger D E and Townsend D W 1999 Implementation and performance of an optical motion tracking system for high resolution brain PET imaging IEEE Trans. Nucl. Sci. 46 2059–67
Mastorakis G and Makris D 2014 Fall detection system using Kinects infrared sensor J. Real-Time Image Process. 9 635–46
Newcombe R A, Davison A J, Izadi S, Kohli P, Hilliges O, Shotton J, Molyneaux D, Hodges S, Kim D and Fitzgibbon A 2011 KinectFusion: real-time dense surface mapping and tracking Proc. of the IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)

Noonan P J, Howard J, Cootes T F, Hallett W A and Hinz R 2012 Realtime markerless rigid body head motion tracking using the Microsoft Kinect Proc. of the Nuclear Science Symp. and Medical Imaging Conf. (NSS/MIC IEEE) pp 2241–6

Olesen O, Sullivan J, Mulnix T, Paulsen R, Hojgaard L, Roed B, Carson R, Morris E and Larsen R 2013 List-mode pet motion correction using markerless head tracking: proof-of-concept with scans of human subject IEEE Trans. Med. Imaging 32 200–9

Peng J L, Kahler D, Li J G, Samant S, Yan G, Amdur R and Liu C 2010 Characterization of a real-time surface image-guided stereotactic positioning system Med. Phys. 37 5421–33

Picard Y and Thompson C J 1997 Motion correction of pet images using multiple acquisition frames IEEE Trans. Med. Imaging 16 137–44

Reynolds M, Dobos J, Peel L, Weyrich T and Brostow G J 2011 Capturing time-of-flight data with confidence Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition pp 945–52

Schöffel P J, Harms W, Sroka-Perez G, Schlegel W and Karger C P 2007 Accuracy of a commercial optical 3D surface imaging system for realignment of patients for radiotherapy of the thorax Phys. Med. Biol. 52 3949

Stone E E and Skubic M 2013 Unobtrusive, continuous, in-home gait measurement using the Microsoft Kinect IEEE Trans. Biomed. Eng. 60 2925–32

The FIL Methods Group 2015 Statistical parametric mapping (www.fil.ion.ucl.ac.uk/spm/)

Wells W M, Viola P, Atsumi H, Nakajima S and Kikinis R 1996 Multi-modal volume registration by maximization of mutual information Med. Image Anal. 1 35–51

Wiles A D, Thompson D G and Frantz D D 2004 Accuracy assessment and interpretation for optical tracking systems Proc. of the Society of Photo-Optical Instrumentation Engineers vol 5367 pp 421–32

Woods R P, Cherry S R and Mazziotta J C 1992 Rapid automated algorithm for aligning and reslicing pet images J. Comput. Assist. Tomogr. 16 620–33

Zhang Z 2000 A flexible new technique for camera calibration IEEE Trans. Pattern Anal. Mach. Intell. 22 1330–4