Abstract—We present a system for head motion tracking in 3D brain imaging. The system is based on facial surface reconstruction and tracking using a structured light (SL) scanning principle. The system is designed to fit into narrow 3D medical scanner geometries limiting the field of view. It is tested in a clinical setting on the high resolution research tomograph (HRRT), Siemens PET scanner with a head phantom and volunteers. The SL system is compared to a commercial optical tracking system, the Polaris Vicra system, from NDI based on translatory and rotary ground truth motions of the head phantom. The accuracy of the systems was similar, with root mean square (rms) errors of 0.09° for ±20° axial rotations, and rms errors of 0.24 mm for ±25 mm translations. Tests were made using 1) a light emitting diode (LED) based miniaturized video projector, the Pico projector from Texas Instruments, and 2) a customized version of this projector replacing a visible light LED with a 850 nm near infrared LED. The latter system does not provide additional discomfort by visible light projection into the patient’s eyes. The main advantage over existing head motion tracking devices, including the Polaris Vicra system, is that it is not necessary to place markers on the patient. This provides a simpler workflow and eliminates uncertainties related to marker attachment and stability. We show proof of concept of a marker less tracking system especially designed for clinical use with promising results.

Index Terms—Motion estimation, positron emission tomography, stereo image processing, stereo vision, structured light system.

I. INTRODUCTION

The tomographic reconstruction of 3D and time varying 3D medical images from a series of scanning modalities including X-ray, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) requires sequential data recording over time. Patient motion during acquisition time will result in a lower image quality or even render the examination useless for PET imaging [1]. Our focus is on the Siemens High Resolution Research Tomograph (HRRT) PET scanner, which is a brain scanner with an isotropic spatial resolution of 1.4 mm [2]. Motion induced image degradation increases with increasing scanner resolution and thus head motion counteracts the technological advances of high resolution scanners. The probability of patient motion occurring increases with acquisition time. For structural or analytical imagery, patient motion can sometimes be estimated and compensated directly from the scan recordings, e.g., in cardiac MRI [3] and lung CT [4]. For functional 3D scans such as PET and fMRI low contrast and spatially sparse events hampers the direct estimation of motion from the recordings themselves, and external motion detection and tracking is preferred [5]–[8]. An optical real time motion tracking system (Polaris System, Northern Digital Inc.) has been preferred for human studies [9]. An alternative optical tracker demonstrated on animals includes [10], [11]. These systems register 3–6 infrared retro reflecting markers fixed in a position relative to each other. They are either glued onto the animal directly or mounted onto a tracking tool and then fixed to the subject. For human head tracking different types of band-aid, helmets, wet-caps, or goggles have been implemented. In a clinical setting attaching markers is an additional element in a busy work flow, additional discomfort to the patient, and with some of the fixation solution listed above the markers are likely to move independently of the patient’s head with head motion giving rise to erroneous motion estimates. Experience also shows that a marker based tracking system has problems registering the markers in the narrow scanner geometry and the accuracy of the system affects the motion correction results.

The purpose of our research is to develop a new 3D head tracking system that 1) does not need any markers, 2) fits to the narrow geometry of the Siemens HRRT PET scanner, 3) is comfortable for the patients, and 4) can potentially be built into future PET scanners. Instead of tracking a geometrical object attached to the patient’s head we propose to track the face itself. The human face has a rich collection of texture and color variations including common features, e.g., eyes, eyebrows, mouth, lips, and person specific features, e.g., moles. However, as we want to infer brain motion we concentrate on geometrical features that are less variant to facial expression, e.g., nose tip, the bridge of the nose, and cheekbones. These features are almost featureless in color and texture space and are characterized by...
ICP algorithm [14] registers and aligns two point clouds with no body motion of the head inside the scanner tunnel. The classical tial precomputed template surface can be used to estimate rigid with a volunteer outside and inside the scanner. For the purpose totype system is shown mounted on the Siemens HRRT scanner based on projection of visible light patterns. In Fig. 1, this pro-

II. SYSTEM REQUIREMENTS

The tracking system must satisfy a number of technical and clinical requirements. 1) The registration of the position must be estimated simultaneously so that a detected PET event known as a line of response (LOR) can be repositioned before the PET image reconstruction. 2) The tracking volume must cover the range of the possible head motion in the HRRT scanner. 3) The system must fit the narrow geometry of the PET scanner. 4) The accuracy of the tracking system has to be better than the spatial resolution of the PET scanner, otherwise the motion correction will increase the blurring instead of reducing it. 5) The system must not interfere with the PET acquisition. 6) The sample frequency has to be at least twice as high as the frequency of head motion to avoid aliasing, according to the Nyquist criterion. However, due to the relatively low count rate in PET, a tracking frequency of 5–10 Hz is adequate [9] or even less if applying the frame repositioning motion correction method [5]. The clinical requirements are at least as important as the technical require-
ments. To be a part of clinical routines the tracking system must be as follows. 1) Simple to use with a preference for a fully automated system. 2) The tracking system must have an easy interface with the PET scanner. 3) It must be robust and have a
flexible design to be a part of the daily routine. 4) The system must be comfortable for the patients, since an uncomfortable patient will introduce motion which is counterproductive for both the patient’s well being and the image quality. 5) Finally, the hygiene requirements of hospital use have to be met.

Commercial surface scanners are available. However, they are not compatible with high resolution PET brain scanners for the following reasons. They do not fit the narrow scanner geometry, are not fast enough as a tracking system, and do not use an invisible light source. A visible light source is uncomfortable for the patients and might introduce motion or at worst may interrupt the scanning.

III. MATERIALS AND METHODS

We intend to integrate a 3D surface scanner into the HRRT PET brain scanner and use it to track patient motion during the PET scans. At each tracking frame a partial 3D point cloud of the patient’s head is processed in the tracker coordinate system and aligned to an initial template surface reconstruction referred to as the reference target.

A. Structured Light Tracking System

The SL system consists of a digital light processing (DLP) projector (DLP Pico Projector, Texas Instruments) with HVGA resolution (480 × 320) and two gray scale charge coupled device (CCD) cameras (Point Grey Research) with a resolution of 1288 × 964. The SL system is designed to match the narrow scanner geometry of the HRRT PET scanner with a flexible design that is easy to mount on the scanner gantry. The SL system mounted on the gantry of the HRRT PET scanner is shown in Fig. 2 just above the patient tunnel. The image plane of the DLP projector consists of micro mirrors that are switched on and off to control every pixel of the projected image. This recent technology has improved the quality of the image projection and made it possible to achieve submillimeter accuracy of surface measurements [13]. The size of this digital micro mirror device (DMD) is 2.42 mm × 3.63 mm and the size of the CCD chip is 3.60 mm × 4.80 mm. The cameras and projector are connected to a computer and synchronized through a custom software setup. Patterns are projected onto the patient’s face and captured as images by the CCDs. The region of interest (ROI) is around the bridge of the nose, seen as the bright region in Fig. 1 bottom. This ROI is chosen due to limited facial motions at the bridge of the nose and because of the high surface curvatures. The system was optimized to have the camera positioned 10–20 cm from the subject and the distance between the projector lens and the camera lens was 8.5 cm resulting in an angle of 30° between the image axis of the camera and the projector.

The SL system was modified with a NIR light emitting diode (LED) in order to meet the clinical requirements for such a system. The system has to be comfortable for the patients and shifting patterns of visible light projecting into the eyes of the patients is not acceptable during long acquisitions. A prototype of the NIR-SL system is seen in Fig. 3 with a specially designed DLP projector (DLP Pico Projector, modified). One of the original three LEDs was bypassed by a NIR LED connected to a separate power supply allowing for adjustment of the light intensity. The inserted LED was cooled by an external fan fixed to the system.

B. Tracking Experiments

1) Phantom Study: The performance of the tracking approach was evaluated by a set of experiments on the HRRT PET scanner with simultaneous tracking using the Polaris Vicra system. A mannequin head was placed inside the patient tunnel as seen in Fig. 2. It was mounted onto a nanorotary motor stage from Thorlabs. The stage made it possible to perform highly controllable rotations of the head. The stage was programmed to rotate in steps of 5° from −20° to 20°. The movements were repeated nine times. At each stationary position a set of four phase-shift images were captured with the SL systems. Furthermore, the phantom was translated in the axial direction in steps of 10 mm across six positions repeated four times. Axial translation is often seen when patients are moving their legs or relaxing neck and shoulders. The performed translation steps were measured with a sliding canvas with an estimated accuracy of ±0.01 mm. At each experiment a 3D point cloud was reconstructed using PSI as described in Section III-D. This study extends the experiment described in [16].

We used the 0° position as the reference position and the pose of the head is estimated relative to this position. The alignment method has also been improved to handle greater motion compared to [16]. This was done by increasing the partial overlap between the reference target and the 3D point clouds at each tracking frame. The reference target was based on 3D point clouds captured outside the HRRT PET scanner just before the subject entered the scanner. Thus a larger region of the face surface was represented in the reference target compared to [16] where point clouds from the initial position inside the HRRT PET scanner were reconstructed and used as the reference target. The reference target was moved into the position at zero degrees before the pose estimation of the 3D point clouds. The Polaris Vicra tracking tool was fixed to the forehead of
the mannequin head as for patients to track the head motions during the PET acquisition [19]. While the Thorlabs stage provides baseline rotation data, the Polaris Vicra system recorded the motions of the head simultaneously with the image capturing of the SL system. Fig. 2 shows the set up of the experiments where the SL system is seen in the front and the Polaris sensor is seen in the back behind the patient tunnel. Fig. 2(a) shows the head in the reference position at 0° and Fig. 2(b) shows the most extreme rotation of the head to the left at −20°. The Polaris Vicra system directly provides a 3 × 3 rotation matrix \( \mathbf{R} \) (with elements \( R_{ij} \)) and a translation vector \( \mathbf{t} \) with respect to a reference position. The SL systems uses a software package Sumatra [20] for the ICP alignment returning \( \mathbf{R} \) and \( \mathbf{t} \). To be able to compare the rigid motion estimates from the SL systems and the Polaris Vicra system with the baseline translations and baseline rotations provided by the Thorlabs stage, the rotation angle \( \theta \), direction of rotation axis \( \mathbf{v} \), and a point on the line \( \mathbf{c} \) is determined from \( \mathbf{R} \) and \( \mathbf{t} \) [21]

\[
\begin{align*}
\theta &= \arccos \left( \frac{\text{trace}(\mathbf{R}) - 1}{2} \right) \\
\mathbf{v} &= \frac{1}{2 \sin(\theta)} \left[ R_{32} - R_{23} \right] R_{13} - R_{51} R_{21} - R_{15} \right]^T \\
\mathbf{c} &= (\mathbf{I} - \mathbf{R})^{-1} \mathbf{t}
\end{align*}
\]

where \( \mathbf{I} \) is the identity matrix.

2) Human Study: We tested the system on a volunteer inside the HRRT PET scanner combined with the Polaris Vicra tracking. The volunteer was positioned as a patient would be and fixed using a vacuum bag as is normally done in the clinical routine at Rigshospitalet, Copenhagen. A reference target was obtained outside the HRRT PET scanner just before the subject entered the PET scanner. The reference target was constructed by aligning 3D point clouds recorded in four different SL system positions and reconstructing a triangulated surface representation. In this way the reference target has full coverage of the face. Inside the HRRT PET scanner the volunteer was asked to move the head into 14 different positions corresponding to movements often observed in PET brain imaging: 1) sidewise rotation, 2) upward rotation, and 3) axial translation. The registered motions are evaluated by comparing the angle of rotation relative to a mean position for each of the systems using (1). The mean rotation \( \mathbf{R} \) is determined as [22]

\[
\mathbf{R} = \arg \min \mathbf{R} \sum_i s_i^2 \left( \mathbf{R}^{-1} \mathbf{R}_i \right)
\]

\[
\approx \arg \min \mathbf{R} \sum_i 3 - \text{trace}(\mathbf{R}^{-1} \mathbf{R}_i)
\]

\[
= \arg \max \mathbf{R} \sum_i \text{trace}(\mathbf{R}^{-1} \mathbf{R}_i)
\]

\[
= \arg \max \text{trace} \left( \mathbf{R}^{-1} \sum_i \mathbf{R}_i \right).
\]

The solution to (2) is found by singular value decomposition (SVD) \( \sum_i \mathbf{R}_i = \mathbf{U} \mathbf{D} \mathbf{V} \). Introducing the matrix \( \mathbf{S} = \text{diag}(1, 1, \det(\mathbf{U} \mathbf{V})) \) we have the mean rotation given as

\[
\mathbf{R} = \mathbf{USV}.
\]

C. Pose Estimation

We wanted to estimate the rigid body transformation from the current 3D scan to the reference scan. The scans are unstructured point clouds where approximate estimates of the point normals exist. We are using a specialized version of the ICP algorithm [14]. Initially, two surfaces for each camera respectively are created based on 2–4 scan positions. Both cameras produce a 3D point cloud representation of the part of the head in its FOV. Scans for each camera are aligned and merged to create a reference target that covers the FOV of each camera using the method described in [20]. In this method the surfaces are created using the Markov random field surface reconstruction algorithm [17]. It is based on an implicit description of the surface combined with a regularization step that makes it well suited for human body scans. Since the surface reconstruction algorithm by default computes surfaces that extend beyond the point cloud, a postprocessing step is needed where the surface is cropped to fit the point cloud. This is done by removing parts of the surface that are not supported by reliable input points. Support is defined as being within a distance \( d \) of an input point, where \( d \) is estimated as the average neighbor distance in the input point cloud. The result is a polygonised surface patch, where the edge vertices are defined as having only one adjacent triangle. For each point in the current scan, the closest point on the triangulated surface is found using a kD-tree based approach. If the point falls on an edge vertex, the point match is discarded. The remaining point matches are used to compute the rigid body transformation using the solution found in [21]. Using this method the transformation bringing the current scan into alignment with the reference surface is computed. Prior to the alignment noisy points, non connected points, and small isolated clusters of points were excluded from the point cloud, following the approach from [17]. The alignment of the partial face surfaces into the reference target is computed twice. In the first round, the point clouds are aligned to a target representing most of the face to generate a robust prealignment. In the second alignment round the target is reduced to include the stable part of the face reconstruction just around the nose bridge.

D. 3D Point Cloud Generation

We use PSI to determine the correspondence between the two image planes; the projector image plane \((u_p, v_p)\) and the image plane of one of the cameras \((u_c, v_c)\) (see Fig. 4). From a series of three captured interferograms (2D images) the wave front phase is computed and converted to line positions on the projector image plane [23]. Thus, a given phase of cosine patterns \( I_k(u_p, v_p) \) on the captured images \( c_k(u_c, v_c) \) correspond to a position on the projector image plane after phase unwrapping. The cosine patterns are generated by

\[
I_k(u_p, v_p) = a \left( 1 + \cos \left( \frac{2\pi}{p} u_p + s_k \right) \right) + b
\]

\[
s_k = \frac{2 \pi}{3} (k-2) \quad \text{and} \quad k = 1, 2, 3
\]

82 IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 31, NO. 1, JANUARY 2012
where $a$ is the amplitude, $b$ is the bias, $s_k$ is the shift, and $k$ is the pattern number of the cosine function. This results in the three captured interferograms

\[ cI_k(u_c, v_c) = I_{av} + I_{mod} \cos (\phi(u_c, v_c) + s_k) \]

with the three unknowns; the phase $\phi(u_c, v_c)$, the average of the intensity $I_{av}$, and the modulation of the intensity $I_{mod}$. Solving the above equation gives the phase

\[ \phi(u_c, v_c) = \arctan \left( \frac{cI_1(u_c, v_c) - cI_3(u_c, v_c)}{2cI_2(u_c, v_c) - cI_1(u_c, v_c) - cI_3(u_c, v_c)} \right). \]

Since the phase is periodic, the phase has to be unwrapped to achieve a continuous phase image. Several methods to perform phase unwrapping exist. Experiments showed that the method described in [24] performs well with our data. This method is a 2D path-independent algorithm where the image is divided into regions based on the $2\pi$ phase jumps. The points on the image planes are converted into 3D coordinates using a simple pinhole model for both the cameras and the projector and assuming the calibrations parameters for all three components are known. The calibration matrices of the cameras $\mathbf{P}_c$ and $\mathbf{P}_p$ and the projector $\mathbf{P}_s$ are $3 \times 4$ matrices and from the perspective camera model we have for one camera

\[ q_c = \mathbf{P}_c Q \]

or

\[ s[u_c \ v_c \ 1]^T = \mathbf{P}_c[X \ Y \ Z \ 1]^T. \]

This can be combined into

\[ u_c = \frac{\mathbf{P}_c(1)Q}{\mathbf{P}_c(3)Q} \quad \text{and} \quad v_c = \frac{\mathbf{P}_c(2)Q}{\mathbf{P}_c(3)Q} \]

where the number of the calibration matrix represents a row e.g., $\mathbf{P}_c(3)$ is the third row of $\mathbf{P}_c$. Similar equations are valid for the projector. A new set of equations can be set up and solved with respect to coordinates in the tracker coordinate system. The coordinates in the CCD image plane and the vertical coordinate of the DMD

\[ u_c \cdot \mathbf{P}_c(3) - \mathbf{P}_c(1) \cdot Q = 0 \]
\[ v_c \cdot \mathbf{P}_c(3) - \mathbf{P}_c(2) \cdot Q = 0 \]
\[ u_p \cdot \mathbf{P}_p(3) - \mathbf{P}_p(1) \cdot Q = 0. \]

The new set of linear equations yields a 3D point in the tracker coordinate system

\[ s[X \ Y \ Z]^T - \mathbf{A}^{-1} \mathbf{b} \]

where $\mathbf{A}$ is a matrix and $\mathbf{b}$ is a vector consisting of calibration parameters. Further details of the system calibration and 3D coordinate computations can be found in [25].

IV. EXPERIMENTAL RESULTS

Fig. 5 (top) shows the 3D point clouds at the nine different positions from $-20^\circ$ to $20^\circ$ (left to right) for one of the nine experiments with the mannequin head. The red and the blue point clouds represent the right and left camera, respectively. As seen, the point clouds are highly detailed with little noise and outliers, demonstrating the high spatial resolution of the system.

The right camera has a more favorable angular position with respect to the surface for the negative rotations, and the left camera for the positive rotations. Thus, we have used point clouds for the camera with the largest angle between the image axis and the surface as shown on Fig. 5. The point clouds are aligned into the reference target and two results of the ICP alignment are shown in Fig. 5 (bottom) at $\pm 10^\circ$ from the right and left camera, respectively. In Fig. 5 (bottom), the color coding of the aligned scans represents the individual per-point alignment error. It is computed as the distance from the point to the closest point on the reference target seen in the back. The errors between the reference target and the aligned points are in the order of 0–0.2 mm with the largest errors around the eyes. The medians of the point errors in Fig. 5 (bottom) are 0.10 mm and 0.09 mm.

As previously mentioned, the motion of the Thorlabs stage is considered the ground truth motion. The errors of the estimated motion are plotted as a function of the ground truth motion in
Fig. 6. Comparison of the SL system, NIR-SL system, and the Polaris Vicra system; Rotation study: (top) differences between the estimated and the performed rotations as a function of the performed rotation. Polaris Vicra rotation experiment 3 has errors $>0.5^\circ$ and not represented in the plot. Translation study: (middle) differences between the estimated and the performed translations as a function of the performed translation; (bottom) the estimated absolute rotations that ideally should be $0^\circ$ for the translation study.

The errors of the SL system are less than $0.2^\circ$ from the performed rotation when using the right camera (red) for negative rotations and the left camera (blue) for the positive rotation with a rms error of $0.089^\circ$. This is a similar result as the Polaris Vicra system, which has a rms error of $0.086^\circ$. One of the Polaris Vicra tracking (experiment no. 3) is treated as an outlier and not included in the rms error. We can not operate the NIR-SL system simultaneously with the Polaris Vicra system since both systems used light at 850 nm and would influence each others estimates.

The NIR-SL systems has been used for the same experiment in a second run and the resulting errors are plotted in light blue. The NIR-SL system has a rms of $0.061^\circ$.

In Fig. 6 (middle) the results from varying the translation is shown. Here we compare the NIR-SL system to the Polaris Vicra system. Again the experiment is done in two runs, one for the NIR-SL system, and one for the Polaris Vicra system in order to avoid interference. The light blue and magenta points represent estimates from the NIR-SL systems left and right camera, respectively. The black points are the results from the Polaris Vicra system. The standard deviation (SD) between the differences of the systems are: 1) $SD(\text{Pol}−\text{SL}_{\text{left}})=0.85^\circ$, 2) $SD(\text{Pol}−\text{SL}_{\text{right}})=0.97^\circ$, and 3) $SD(\text{SL}_{\text{left}}−\text{SL}_{\text{right}})=0.41^\circ$. It is seen that the difference between the systems increases with the angle of rotation from the centre, cf. Fig. 7.

Fig. 8 shows the 3D point clouds before (left) and after (right) the ICP alignment to the reference target for two head poses. These are pose 7 and pose 10 which are close to the head poses shown in Fig. 1. These poses are moved approximately 8 mm and $3^\circ$ for pose 7 and 35 mm and $19^\circ$ for pose 10 compared to the reference position. As seen in Fig. 8, parts of the unaligned point clouds (left) are missing from the aligned point clouds (right). These parts are excluded either due to detected errors or simply due to occlusion and shadowing. The mean of the median values of these point errors is $0.19$ mm ($\pm 0.02$ mm) for the two poses shown in Fig. 8(right). It is noted that the per point errors for the two poses are similar even though pose 10 is one of the outlying poses. Finally, we demonstrate that the NIR illumination has the same performance for surface registration as visual light illumination. Fig. 9 shows results of the NIR-SL system for a human volunteer. This experiment was conducted outside the HRRT PET scanner with the volunteer sitting upright in a chair with support for the back and not for the head. The centroids of the reconstructed 3D point clouds from the four positions were $31−46$ mm from the centroid of the
reference target. The main motions were axial/downward translation and sidewise rotation which are typical motions seen in PET brain imaging. The captured images are nonblurred and have high contrast (utilizing approximately 80% of the 8 bit colors). This supports the high quality of the captured images using the NIR LED. The reconstructed 3D point clouds correspond to the regions of the scanned surface. The errors of the ICP alignment are less than 0.5 mm for the majority of regions. The mean of the median values of these point errors are 0.29 mm (±0.09 mm) for the visual light illumination). Small areas especially at the borders and the nose have errors >1 mm seen as the red areas (Fig. 9).

V. DISCUSSION

A structured light system developed to register 3D head motion was applied to a clinical setting and shown to work given narrow scanner geometries such as the HRRT PET scanner. The design is miniaturized, flexible, and does not need any markers. These qualities make the system usable and valuable in a clinical setting. A markerless system is also timesaving and hygienic in terms of hospital use. Another important advance of not using markers is the elimination of the major source of error when attaching markers onto the head. When using the Polaris Vicra system there is potential risk that the tracking tool moves relative to the skull either because of the attachment or movements not related to the skull such as facial movements. Facial movements can obviously also introduce tracking errors with the SL system. However, this system relies on thousands of points thereby improving the robustness of the transformation estimation.

The structured light system was realized in both visible and NIR versions, SL and NIR-SL, respectively. The two systems were compared to the Polaris Vicra system on a phantom set up in order to get ground truth motions. The accuracy of the SL system was equal to the Polaris Vicra system with a rms error of 0.09° for axial rotation from -20° to 20° and a rms error of 0.26 mm for translation over a range of 50 mm (Fig. 6).

This result is supported by the accuracy found in [16] for a similar experiment. Thus, on a rigid object where the tool is fixed the accuracy of both systems are in the order of a few tenths of a millimeter for the brain region, e.g., if a head is rotating around the point of contact in the back head the displacement of the frontal brain lobe 10 cm away for a 0.09° rotation is \( \tan(0.09°) \cdot 10 \text{ cm} = 0.16 \text{ mm} \).

We believe this accuracy is representative of the achievable accuracy on humans since some aspects can be improved such as facial movements, time of image capturing, and combining the left and right systems as explained below. The accuracy are an order of magnitude less than the current resolution of the HRRT PET scanner.

A human volunteer study was performed inside the HRRT PET scanner to demonstrate the SL and NIR-SL systems usability for the HRRT PET scanner on humans. We have shown detailed 3D point clouds of the face surfaces which are aligned to the target with a median of the per point error of around 0.2 mm (Fig. 8). These errors are due to reflectance of the light, noise, and motion. The output transformations from the SL system are based on thousands of points and thus the accuracy of the rigid body transformation is expected to be \( \ll 1 \text{ mm} \). From the comparison of the systems in 14 head poses we find twice as good agreement between the two SL systems \( \text{SD} (\text{SL}_{\text{left}} - \text{SL}_{\text{right}}) = 0.41° \) compared to the agreement between the Polaris Vicra system and one of the SL systems \( \text{SD} (\text{Pol} - \text{SL}_{\text{left}}) = 0.85° \) and \( \text{SD} (\text{Pol} - \text{SL}_{\text{right}}) = 0.97° \).
Some poses have significant larger difference with the Polaris Vicra system (Fig. 7) for which we have no unique explanation. It could be due to the calibration of the SL system or motion of the Polaris tracking tool. We observed that the per point alignment error (Fig. 8) was not related to the size of the performed motion. The difference between the left and right SL system is at the same level as for the phantom experiment if we do not only consider the system with the best angular position relative to the surface. Therefore, the accuracy of a total SL system is improved by combining information from the left and right system taking the angular distortion into account or simply excluding point clouds with normals perpendicular to the image axis.

We have presented a miniaturized NIR-SL system for 3D head tracking. In this system, the visible LED of the original Texas Instruments Pico DLP was replaced with a NIR LED. We demonstrated the system on a human volunteer for four different head poses. The 3D point clouds are highly detailed as for the visible system. The median of the aligned per point errors is 50% larger than for the human volunteer study with the visible SL system. This could be explained by the differences between the experiments. In the NIR-SL case, the volunteer was sitting in a chair without head support, whereas for the SL case the volunteer was lying in the HRRT PET scanner with head support. This set up might have introduced some motion during the image capturing of a set of images of 3–4 s. The rate of the image capturing has not been optimized for these prototypes of the SL system.

Based on the detailed point clouds, the results of the aligned point clouds, and comparing the quality of the captured images we expect to obtain the same accuracy of this system compared to the visible SL system. The results with the NIR-SL system are representative for the experiments where we performed inside the HRRT PET scanner. This error indicates that the system will be able to accurately determine the pose changes for real humans in a clinical environment with simultaneous PET acquisition. In order to realize motion compensated HRRT PET imaging based on the NIR-SL system time synchronization between the HRRT PET scanner and the NIR-SL must be established, the HRRT PET and NIR-SL coordinate system must be aligned, the NIR-SL must be configured to run continuously, and an algorithm for motion compensated reconstruction must be in place. Basically, two approaches can be taken for motion compensation depending on the nature of the motions of the study subjects. The multiple acquisition frames method assumes that the subject lies still for longer periods of time interrupted by short periods of motion [5]. The motion tracking device is used to identify these periods of no motion. For each such period a PET image (frame) is reconstructed using standard algorithm. The frames are then combined to a reconstructed PET image in a second step. An alternative method does not make an assumption of the motion pattern but requires continuous estimates of pose [26]. In this case each PET event (LOR) is repositioned before using a standard reconstruction method.

VI. CONCLUSION

We present a compact vision system based on a customized Texas Instruments Pico DLP projector fitted with a NIR LED. The system is adapted for motion correction in high resolution PET brain imaging. While the system's accuracy is comparable with the current state of the art optical trackers, it is more flexible and the system is fully automatic and does not rely on markers. Furthermore, the NIR LED ensures a more comfortable experience for the patients. This is a step toward a fully automatic tracking system designed for the HRRT PET brain scanner, but with potential use in other scanners and imaging modalities where an external tracking system is currently needed.

ACKNOWLEDGMENT

The authors would also like to thank Rigshospitalet for making clinical scanner facilities available. They would like to thank The John and Birth Meyer Foundation for the generous donation of the HRRT PET scanner. Finally, they would like to thank the staff at The Mechanical Workshop at Rigshospitalet who helped them produce the mechanical devices.

REFERENCES


