

Motion Tracking in Narrow Spaces: A Structured Light Approach

Oline Vinter Olesen^{1,2,3}, Rasmus R. Paulsen¹, Liselotte Højgaard²,
Bjarne Roed³, and Rasmus Larsen¹

¹ Informatics and Mathematical Modelling, Technical University of Denmark
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark
ovol@imm.dtu.dk

<http://www.imm.dtu.dk/>

² Department of Clinical Physiology, Nuclear Medicine & PET, Rigshospitalet,
Copenhagen University Hospital, University of Copenhagen

³ Siemens Healthcare, Siemens A/S, Denmark

Abstract. We present a novel tracking system for patient head motion inside 3D medical scanners. Currently, the system is targeted at the Siemens High Resolution Research Tomograph (HRRT) PET scanner. Partial face surfaces are reconstructed using a miniaturized structured light system. The reconstructed 3D point clouds are matched to a reference surface using a robust iterative closest point algorithm. A main challenge is the narrow geometry requiring a compact structured light system and an oblique angle of observation. The system is validated using a mannequin head mounted on a rotary stage. We compare the system to a standard optical motion tracker based on a rigid tracking tool. Our system achieves an angular RMSE of 0.11° demonstrating its relevance for motion compensated 3D scan image reconstructions as well as its competitiveness against the standard optical system with an RMSE of 0.08° . Finally, we demonstrate qualitative result on real face motion estimation.

1 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 this time will result in a lower image quality or even render the examination useless, cf. [1] for PET imaging. The probability of patient motion occurring grows with increasing acquisition time. For structural or anatomical imagery, patient motion can sometimes be estimated and compensated directly from the scan recordings, e.g. in cardiac MRI [2] and lung CT [3]. For functional 3D scans such as PET and fMRI lower contrast and more spatially sparse events hampers the direct estimation of motion from the recordings themselves. Our focus is on the Siemens High Resolution Research Tomograph (HRRT) PET brain scanner, which is a

brain dedicated scanner with a spatial resolution down to 1.4 mm [4]. Movement induced image degradation increases with increasing scanner resolution and thus head motions end up counteracting the technological advances of high resolution scanners. Due to the low count rate and resulting low contrast information the patient motion is assumed known for most of the suggested motion correction methods in PET imaging [5, 6, 7, 8]. An optical real-time motion tracking system has been preferred (Polaris System, Northern Digital Inc.) [9]. This system registers a rigid tracking tool with 3-6 infrared reflecting markers. The tracking tool is fixed to the patient's head using different types of band-aid, helmets, wet-caps, or goggles. However, it has been reported that, using fixation, these methods can cause artifacts on the PET images [10]. We have previously described a structured light (SL) based system for 3D face surface reconstruction that: (1) does not need any markers; (2) fits to the narrow geometry of the Siemens HRRT PET scanner; and (3) can potentially be built into future PET scanners [11].

In this paper we will show how 3D point clouds captured using this system by use of resistant and robust iterative closest point (ICP) registration to a template surface can be used to estimate rigid body motion of the head inside the scanner patient tunnel. In the method proposed, the pose estimation is based on the rigid alignment of scans to either a pre-computed reference scan or the previous scan. The alignment is a special case of the 3D point cloud registration. Point correspondence cannot be assumed and since the scanner output are patches, there will only be partial overlap between reference/previous scans and the current scan. Our method is based on the classical ICP algorithm [12] that aligns two point clouds with no prior correspondence. Several efficient variants of the ICP algorithm have been published [13]. The possible enhancements of the normal ICP includes matching based on differential properties of point sets and rejection of invalid point matches. Due to the partial overlap of our scans, we employ a point rejection approach, where points falling on the border of the target are rejected. It is therefore necessary to detect borders of the 3D scan. This is non-trivial to do with raw points clouds. To overcome this, we represent our target as a triangulated surface that has been computed using a state-of-the-art surface reconstruction algorithm [14]. Compared, to the popular Poisson surface reconstruction algorithm [15], the Markov Random Field surface reconstruction algorithm [14] deals particularly well with human body scans. We provide a quantitative evaluation of the performance of the structured light system for pose estimation.

2 Experiments and Methods

The SL system consists of a DLP projector (DLP Pico Projector, Texas Instruments) with HVGA resolution (480×320) and two grayscale CCD cameras (Chameleon, Point Grey Research) with a resolution of 1280×960 as described in [11]. The SL system was mounted on the gantry of the HRRT PET scanner as shown on Fig. 1 just above the patient tunnel. The performance of the new tracking approach is 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 a patient would be. It was mounted onto a nano rotary motor stage from Thorlabs. The stage made it possible to perform highly controllable rotation of the head. The stage was programmed to rotate in steps of 5 degrees from -15 to 15 degrees and the movements were repeated four times. At each stationary position a set of images were captured with the SL system.

The Polaris tracking tool was fixed to the forehead of the mannequin head using a band-aid as used for patients to track the head motions during the PET acquisition. While the Thorlabs stage provides baseline rotation data, the Polaris system recorded the motions of the head simultaneously with the image capturing of the SL system. Fig. 1 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. 1(a) shows the head in the reference position at 0 degrees where the region of interest (ROI) is seen as the bright region around the bridge of the nose. This ROI is chosen due to the limited facial movements of the bridge of the nose. At each of the seven positions of the four runs of experiments 3D point clouds are reconstructed using phase-shifting interferometry (PSI) [16]. PSI is used to determine the correspondence between the two image planes; the projector image plane and the image plane of one of the cameras. From a series of three captured interferograms (2D images) the wavefront phase is computed and converted to line positions on the projector image plane. Thus a given phase of the cosine patterns on the captured images correspond to a position on the projector image plane after phase unwrapping. 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 [17] perform well with our data.

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

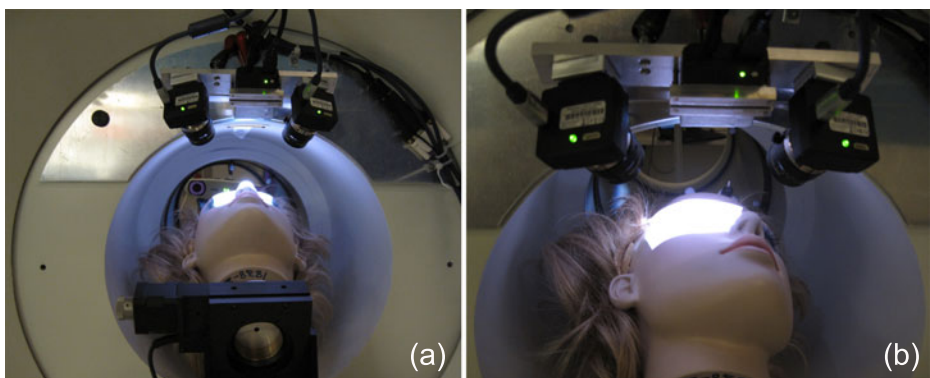


Fig. 1. Photographs of the mannequin head inside the HRRT PET scanner with the SL system in the front mounted to the HRRT PET gantry. (a) The motor stage is seen in the bottom and the Polaris sensor in the back. (b) Mannequin head rotated to the right (the tracking tool can just be discerned above the forehead).

the calibrations parameters for all three components are known. The details of system calibration and 3D coordinate computations can be found in [11].

We use the scan acquired at 0° as the reference and in the following, the pose of the head is estimated relatively to this reference in the following. We want 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 specialised version of the ICP [12] algorithm. Initially, a surface is created based on the reference scan. Both cameras produce a point cloud representation of the part the head in its field of view. These two point clouds are aligned and merged to create a reference scan that covers the field of view of both cameras. The surface is created using the novel Markov Random Field surface reconstruction algorithm [14]. 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 post-processing 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. d is estimated as the average neighbour distance in the input point cloud. The result is a polygonised surface patch, where the edge vertices are defined by having only one adjacent triangle. For each point in the current scan, the closest point on the triangulated surface is found using a k D-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 [18]. Using this method the transformation bringing the current scan into alignment with the reference surface is computed. The transformation consists of an estimated 3×3 rotation matrix \mathbf{R} and a translation vector \mathbf{t} .

To be able to compare the rotation estimate from our method and the Polaris system with the baseline rotation provided by the Thorlabs stage, the rotation angle θ , direction of rotation axis \mathbf{v} , and a point on the line \mathbf{c} is determined as [18]:

$$\begin{aligned}\theta &= \arccos((\text{trace}(\mathbf{R}) - 1)/2) \\ \mathbf{v} &= 1/(2 * \sin(\theta)) * [R_{32} - R_{23} \ R_{13} - R_{31} \ R_{21} - R_{12}]^T \\ \mathbf{c} &= (\mathbf{I} - \mathbf{R})^{-1}\mathbf{t}\end{aligned}$$

3 Data and Results

Fig. 2 shows the 3D point clouds at the different positions for one of the four experiment runs with the right (red) and left (blue) camera respectively. The image to the left represent the position at -15 degrees and the image to the right at a position of 15 degrees. These are aligned into the reference surface seen in the center. As seen, the point clouds are highly detailed with little noise and outliers, supporting a high spatial resolution of the system. Two results of the ICP alignment are shown in Fig. 3 at ± 10 degrees from the right camera. The

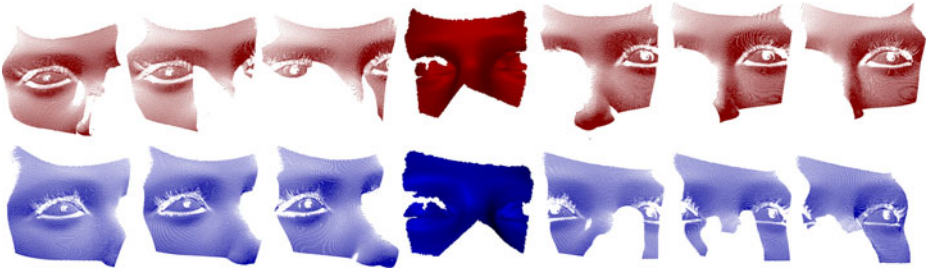


Fig. 2. 3D point clouds of the mannequin head at the seven positions. Left to right from -15 degrees to 15 degrees. Blue images represent the left camera and red images represent the right camera. The center images at 0 degrees are the reference surfaces.

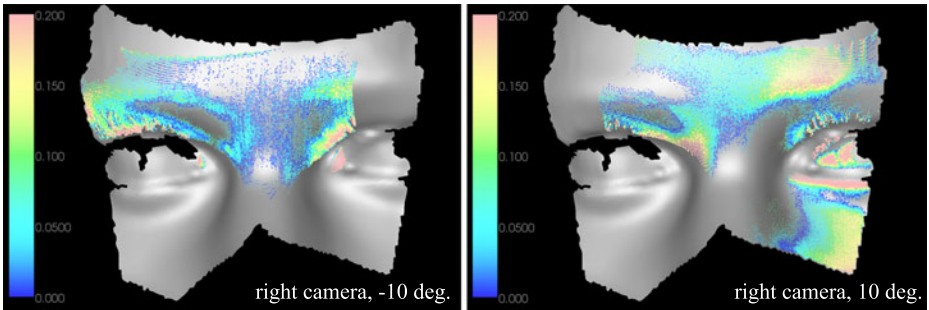


Fig. 3. Results of the ICP alignment of the right camera at two positions ± 10 degrees. Alignments into the reference position are shown on top of the reference surfaces. The colors represent the errors as the distance to target in mm.

errors between the target and the aligned points are in the order of 0-0.2mm with the largest errors around the eyes. 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. 4. The red and blue points represent the right and left camera respectively and the black points are the results from the Polaris system. The errors of the SL system are less than 0.2 degrees 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 0.11 degrees. This is a similar result as the Polaris system, which has a RMS error of 0.08 degrees. There are two main reasons why the results are less accurate when the head is rotated toward the camera; the overlap with the target around the bridge of the nose becomes less and the overlap around the eye becomes larger. The eye is a non-robust region due to edges and hair. This is an even larger problem on humans. In the future, this region will be excluded prior to alignment.

The system has been tested on a human test subject in the HRRT setup. The purpose is to demonstrate the clinical usability of the system. Currently,

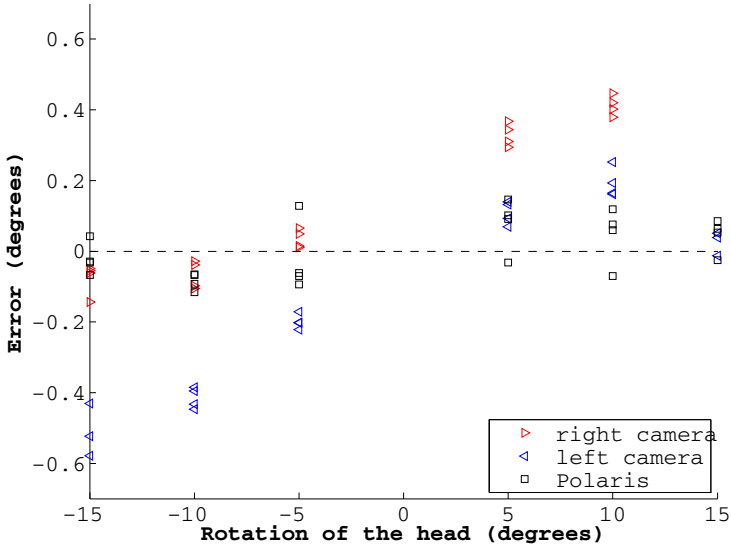


Fig. 4. Comparison of the SL system and the Polaris system; differences between the estimated rotations and the performed ones as a function of the performed rotations.

we are only aiming for a quantitative evaluation and therefore no ground truth data was recorded and no PET acquisition was done. The subject moved 1-3 cm between each scan. The results can be seen in Fig. 5, where the reference scan is represented as a grey reconstructed surface. Two following scans are aligned to the reference scans and the aligned point clouds are seen. It is clearly seen, that the reference scans are less complete than for the mannequin. This is due to problems with shadows around the nose. We are currently optimising the hardware configuration of the system with respect to real humans in the clinical setting. In Fig. 5, the colour coding of the aligned scans represents the individual

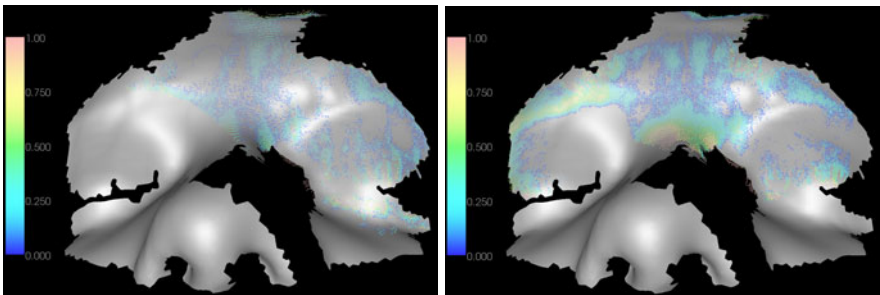


Fig. 5. Test with real human scanings. Point clouds aligned to the reconstructed surface of the reference scan (grey surface) and the point errors are colourcoded. The medians of the point errors are 0.13 and 0.17 mm.

per-point alignment error. It is computed as the distance from the point to the closest point on the reference scan. The median values of these point errors are 0.13 and 0.17 mm for the two scans in Fig. 5. These results are representative for the experiments we have performed on human volunteers. While not optimal, the magnitude of this error indicates that the system will be able to accurately determine the pose changes for real humans in the clinical environment.

4 Summary and Conclusions

We have presented a structured light system adapted for motion compensation in high-resolution PET scanners. While the systems accuracy is comparable with the current state-of-the-art optical trackers, it is more flexible and easier to adapt to the narrow patient tunnels of PET scanners. Furthermore, the system is fully automatic and does not rely on markers that are notoriously difficult to use in a clinical setting. The system was tested on a setup, where a mannequin head mounted on a robot system created baseline data. The results show, that the proposed framework was able to estimate the head rotation with an accuracy better than 0.2° for a head movement between -15° and 15° . Furthermore, preliminary experiments were performed on human test subjects. Quantitative analysis shows that the system is able to robustly estimate the pose changes for human subjects, strongly indicating that the method will be useable in the clinical practise.

Acknowledgements. The authors would like to thank the staff at The Mechanical Workshop at Rigshospitalet, Copenhagen, Denmark who helped us produce the mechanical devices. Ms. Olesen also thanks the Danish Industrial PhD programme for supporting this research.

References

1. Green, M.V., Seidel, J., Steina, S.D., Tedder, T.E., Kempner, K.M., Kertzman, C.: Head movement in normal subjects during simulated PET brain imaging with and without head restraint. *J. Nuclear Medicine* 35(9), 1538–1546 (1994)
2. Stegmann, M.B., Larsson, H.B.W.: Motion-compensation of cardiac perfusion MRI using a statistical texture ensemble. In: Magnin, I.E., Montagnat, J., Clarysse, P., Nenonen, J., Katila, T. (eds.) *FIMH 2003*. LNCS, vol. 2674. Springer, Heidelberg (2003)
3. Ehrhardt, J., Werner, R., Frenzel, T., Säring, D., Lu, W., Low, D., Handels, H.: Reconstruction of 4D-CT data sets acquired during free breathing for the analysis of respiratory motion. In: *Proc. SPIE Medical Imaging*, vol. 6144, pp. 614414–1–8 (2006)
4. Olesen, O.V., Sibomana, M., Keller, S.H., Andersen, F., Jensen, J.A., Holm, S., Svarer, C., Højgaard, L.: Spatial resolution of the HRRT PET scanner using 3D-OSEM PSF reconstruction. In: *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, vol. M13-225, pp. 3789–3790 (2009)
5. Picard, Y., Thompson, C.J.: Motion correction of PET images using multiple acquisition frames. *IEEE Trans. on Medical Imaging* 16(2), 137–144 (1997)

6. Woo, S.K., Watabe, H., Choi, Y., Kim, K.M., Park, C.C., Bloomfield, P.M.: Sinogram-based motion correction of PET images using optical motion tracking system and list-mode data acquisition. *IEEE Trans. on Nuclear Science* 51(3), 782–788 (2004)
7. Rahmim, A., Dinelle, K., Cheng, J.C., Shilov, M.A., Segars, W.P., Lidstone, S.C., Blinder, S., Rousset, O.G., Vajihollahi, H., Tsui, B., Wong, D.F., Sossi, V.: Accurate event-driven motion compensation in high-resolution PET incorporating scattered and random events. *IEEE Trans. on Medical Imaging* 27(8), 1018–1033 (2008)
8. Raghunath, N., Faber, T.L., Suryanarayanan, S., Votaw, J.R.: Motion correction of PET brain images through deconvolution: II. Practical implementation and algorithm optimization. *Physics in Medicine and Biology* 54(3), 813 (2009)
9. Lopresti, B.J., Russo, A., Jones, W.F., Fisher, T., Crouch, D.G., Altenburger, D.E.: Implementation and performance of an optical motion tracking system for high resolution brain PET imaging. *IEEE Trans. on Nuclear Science* 46(6), 2059–2067 (1999)
10. Herzog, H., Tellman, L., Fulton, R., Pietrzyk, U.: Motion correction in PET brain studies. In: *IEEE Proc. The Fourth International Workshop on Multidimensional Systems*, pp. 178–181 (2005)
11. Olesen, O.V., Jørgensen, M.R., Paulsen, R.R., Højgaard, L., Roed, B., Larsen, R.: Structured light 3D tracking system for measuring motions in PET brain imaging. In: *Proc. SPIE Medical Imaging*, vol. 7625, p. 76250X (2010)
12. Besl, P.J., McKay, N.: A method of registration of 3D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14(2), 239–256 (1992)
13. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: *Proc. Int. Conf. 3-D Digital Imaging and Modeling*, pp. 145–152 (2001)
14. Paulsen, R., Bærentzen, J., Larsen, R.: Markov Random Field Surface Reconstruction. *IEEE Transactions on Visualization and Computer Graphics* (2009)
15. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson Surface Reconstruction. In: *Proc. Symposium on Geometry Processing*, pp. 61–70 (2006)
16. Huang, P., Hu, Q., Jin, F., Chiang, F.: Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring. *Optical Engineering* 38, 1065 (1999)
17. Herráez, M., Burton, D., Lalor, M., Gdeisat, M.: Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Appl. Opt.* 41, 7437–7444 (2002)
18. Horn, B.K.P.: Closed form solution of absolute orientation using unit quaternions. *Journal of the Optical Society A* 4(4), 629–642 (1987)