Markerless Motion Tracking for Motion-Compensated Clinical Imaging

Andre Kyme, *Member*, *IEEE*, Stephen Se, *Senior Member*, *IEEE*, Steven Meikle, *Senior Member*, *IEEE*, and Roger Fulton, *Senior Member*, *IEEE*

Abstract– Motion-compensated brain imaging can dramatically reduce image artifacts and degradation associated with head motion. However, it has been slow to enter routine clinical use, and remained largely a research tool. One possible reason for this is the lack of a practical motion tracking method for the clinical setting. Here we present the initial validation of a highly convenient markerless motion tracking method, previously developed for motion-compensated PET imaging of rats, for human head motion estimation in a mock imaging scenario. Sixteen reclining volunteers followed the projected path of a robotically controlled laser beam during a 2 min experiment in which their resulting head motion was simultaneously tracked at 30 Hz from the anterior and posterior using the markerless system and a marker-based system, respectively. Motion estimates were compared in a common coordinate system and generally showed excellent agreement, especially in early frames (RMSE of 1-3 mm). Larger discrepancies were associated with the unwanted detection of non-rigid features which could be prevented in future studies. These results are a promising first step in translating the markerless tracking system to clinical use.

I. Introduction

Motion-compensated brain imaging has been shown to dramatically reduce the artifacts and quantitative degradation associated with voluntary and involuntary subject head motion during scanning [1, 2]. However, motion compensated brain imaging protocols have not found their way into routine clinical use, remaining largely research tools. One possible reason for this is the lack of a practical motion tracking method for the clinical setting.

Nearly all optical motion tracking systems available require the rigid attachment of markers to the patient's head, a procedure which can lead to errors due to slippage of the markers, and which can be time-consuming to set up. Although many attachment methods have been used with some success (e.g. goggles, neoprene caps and headbands) [e.g. 3, 4], a tracking method capable of providing accurate pose estimates over a large range of motion, but without the need to attach markers, would be advantageous. There is current interest in such tracking approaches for clinical applications [5].

Previously we developed and reported a markerless optical tracking system for motion-compensated PET imaging of rats [6]. Here we describe the first stage in translating this system to clinical use, validating the markerless tracking approach for human head motion in a mock imaging scenario.

II. Methods

A. System Overview

The markerless motion tracking system used in this work has been described in detail previously [6]. Briefly, the system consists of four synchronized and calibrated CCD cameras (Point Grey Research, Canada) arranged front-on to the subject and in two pairs viewing different sides of the subject's head. We detect native features [7] on the face/head from the multiple 2D views in order to reconstruct 3D landmarks on the face/head surface. A 3D-2D registration between accumulated landmarks and image features in the newly acquired camera images is then used to estimate the changing head pose [8].

B. Experimental validation

A mock imaging scenario was set up in the lab for tracking the head motion of human volunteers (n=16) as shown in Fig. 1. The markerless tracking system was positioned on an optical bench in front of the volunteer who reclined in a chair to simulate an oblique viewing angle typical when out-of-bore tracking systems are used. A marker-based optical motion tracking system (MicronTracker, Claron Tech. Inc., Toronto, Canada) was positioned on a tripod behind the volunteer and used as a reference. A marker attached to a neoprene swimming cap (Fig. 2) or head band worn by the volunteer was used for the marker-based tracking. The marker was not visible to the markerless tracking system and therefore did not act as a source of additional (artificial) features. The markerless tracking system was used to trigger the marker-based system so that synchronized pose measurements could be obtained. The two systems were also cross-calibrated so that the pose estimates from each could be compared.

To achieve some degree of uniformity in the head motion performed by each volunteer, a 6-axis robot (Epson C3-A601S, SEIKO Corp., Japan) with a laser pointer on the end-effector executed several pre-programmed motion sequences during each experiment, lasting approximately 2 min. This resulted in zigzag, circular and square motion paths of the laser spot on the wall in front of the volunteers, which they were instructed to follow using head, rather than eye, motion. The motion was deliberately large to thoroughly test the

This work was supported by Australian Research Council Grant No. DP 0988166.

André Kyme is with the Faculty of Health Sciences and the Brain & Mind Research Institute, University of Sydney, NSW 2006, Sydney, Australia (e-mail: andre.kyme@sydney.edu.au).

Stephen Se is with MDA Systems Ltd., 13800 Commerce Parkway, Richmond, B.C. V6V 2J3 Canada.

Steven Meikle and Roger Fulton are with the Faculty of Health Sciences and the Brain & Mind Research Institute, University of Sydney, NSW 2006, Sydney. Roger Fulton is also with the School of Physics, University of Sydney, NSW 2006, Sydney, Australia, and with the Department of Medical Physics, Westmead Hospital, NSW 2145, Sydney, Australia.

system. During the motion sequence both tracking systems acquired data at 30 Hz. These data were processed offline to obtain pose estimates in a common coordinate frame for comparison.

III. RESULTS AND DISCUSSION

Figure 3 shows an example of the detected features (red) and matches (white) across camera views. Table 1 shows the range and rate of motion for each degree-of freedom. Range varied from approximately 17-100 deg. and 50-280 mm. Maximum rotational and translational speeds were approximately 150 deg/s and 350 mm/s, and mean speeds were approximately 10 deg/s and 25 mm/s. These data confirm that the head motion was more extreme than what would typically be expected from a compliant patient during neuroimaging.

Figure 4 shows the *x*-rotation component of head motion for one of the volunteers as measured by the marker-based (black) and markerless (red) tracking systems. These data show movements in MicronTracker coordinates, relative to the initial pose. The estimated movements were very well matched.

Figure 5 shows a comparison of the *y*-rotation estimate for a short segment (500 frames) of the data so that the jitter is more obvious. These data suggest that the markerless pose estimates may be inherently less noisy than those from the marker-based system, at least for early frames.

Figure 6 shows, in red, the cloud of landmarks (in the initial head pose) that were obtained using the markerless pose estimates for one of the volunteer studies. Superimposed on this, in black, is the landmark cloud obtained using pose estimates from our reference system, the MicronTracker. The zoomed region shows clearly a discrepancy (see arrow) between the landmark locations obtained using the two systems. We quantified this discrepancy in terms of the root mean square error (RMSE), computed over all landmarks. Table 2 shows the RMSE results for all 16 volunteers as more poses were included: the first 1000 poses (first column), 2000 poses (second column), 3000 poses (third column) and all 4000 poses (fourth column). It was not unusual for the RMSE to be between 1-2 mm early on, and even to remain this low for the whole sequence. Given that the cross calibration error between the markerless tracking system and MicronTracker contributed somewhere between 1-2 mm, this suggests the systems were performing very similarly. It was also apparent that the RMSE tended to increase over the course of the sequence. We attributed this to non-rigid features on the face (caused by changing facial expression), clothing and background resulting in drift error in the pose estimates obtained using the markerless tracking system. There is clearly more work to be done to limit features to the face, and to deal appropriately with non-rigid features, either by excluding them or minimizing their contribution to pose estimation.

IV. CONCLUSION

We have presented results from the first stage of translating a markerless motion tracking approach, originally developed for motion-compensated PET imaging of rats, to the clinical setting. The approach appears well suited to tracking human faces and preliminary results indicate that accuracy comparable to state-of-the-art marker-based systems is feasible. The markerless tracking system seems to provide reduced measurement jitter, which may be due to the large number of well-spaced features used for pose estimation. Challenges for this approach include non-rigid features resulting from changing facial expressions and background features. Future work will focus on addressing these issues, in particular by incorporating error modelling to reduce the impact of drift and uncertainty in database landmarks.

REFERENCES

[1] R. Fulton et al., "Correction for head movements in positron emission tomography using an optical motion-tracking system," *IEEE Trans. Nucl. Sci.*, vol. 49, pp. 116–123, 2002.

[2] A. Montgomery et al., "Correction of head movement on PET studies: comparison of methods," *J. Nucl. Med.*, vol. 47, pp. 1936–1944, 2006.

[3] P. Bloomfield et al., "The design and implementation of a motion correction scheme for neurological PET," *Phys. Med. Biol.*, vol. 48, pp. 959-978, 2003.

[4] H. Herzog et al., "Motion artifact reduction on parametric PET images of neuroreceptor binding," *J. Nucl. Med.*, vol. 46, pp. 1059-65, 2005.

[5] O. Olesen et al., "Motion tracking for medical imaging: a nonvisible structured light tracking approach," *IEEE Trans. Med. Imaging*, vol. 31, pp. 79-87, 2012.

[6] A. Kyme et al., "Novel SLAM-Based Markerless Motion Tracking of Conscious Unrestrained Rodents in PET," *Proc. 2011 IEEE Nucl. Sci. Symp. Med. Imaging. Conf.*, pp. 3554-3557, 2011.

[7] D. Lowe, "Object recognition from local scale invariant features," *Proc. Int. Conf. Comp. Vis.*, pp. 1150-1157, 1999.

[8] S. Se et al., "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *Int. J. Robot. Res.*, vol. 21, pp. 735-758, 2002.

ACKNOWLEDGEMENT

This project was supported by the Australian Research Council.

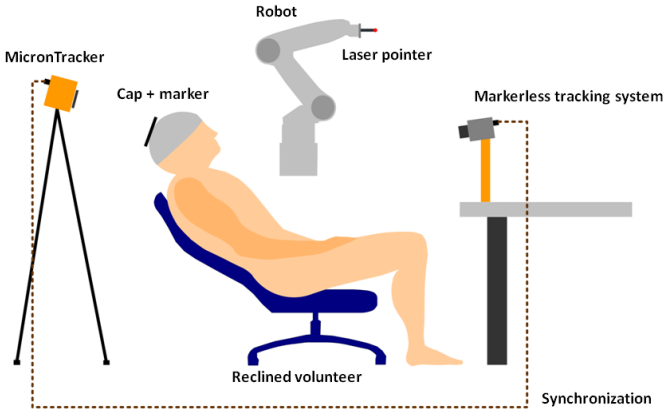

Fig. 1. Experimental setup. A reclined volunteer followed a light source projected onto the wall using a robot-controlled laser pointer undergoing controlled motion. Head motion was tracked using the markerless and marker-based (MicronTracker) tracking systems simultaneously.

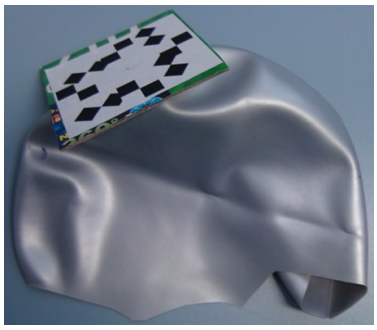

Fig. 2. Swimming cap with a marker attached for use with the marker-based motion tracking system.

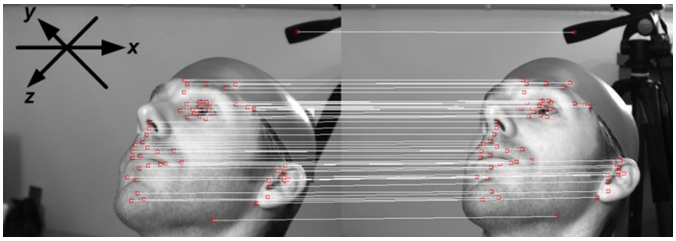

Fig. 3. Feature detection and matching for a human face. Features are shown in red and feature matches as white lines. Details of the feature detection and matching can be found in [6]. An outlier match can be seen on the tripod; this was removed using outlier rejection methods prior to pose estimation.

Table 1. Range and rate of volunteer head motion*.

DoF	Range (deg or mm)	Rate (deg.s^{-1} or mm.s^{-1})	
		Max	Mean
***x*-rot**	29 – 65	103	11
***y*-rot**	17 – 66	92	9
***z*-rot**	51 – 106	146	15
***x*-trans**	69 – 279	316	24
***y*-trans**	95 – 270	354	36
***z*-trans**	56 – 135	182	17

* See Fig. 3 for the coordinate system.

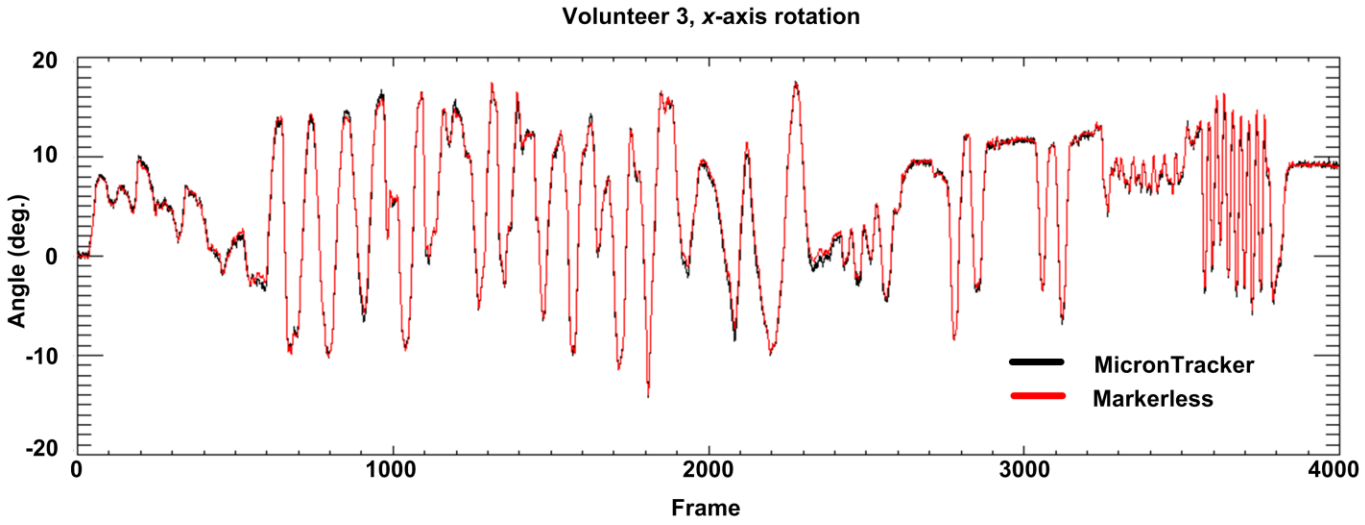

Fig. 4. Pose estimate comparison. Estimated *x*-axis head rotation of volunteer 3 determined using the marker-based (black) and markerless (red) tracking systems for one of the volunteers.

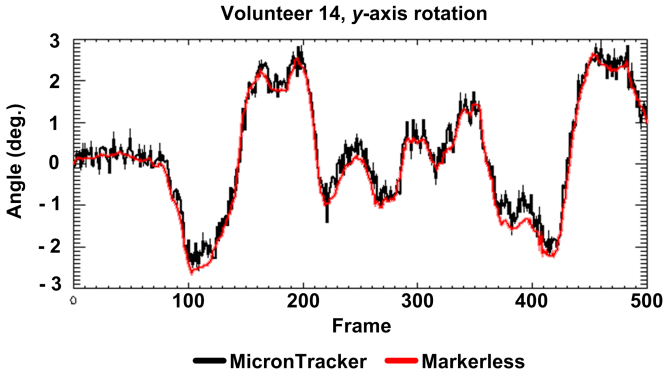

Fig. 5. Noise comparison. A 500-pose segment showing the estimated *y*-axis head rotation for volunteer 14 determined using the marker-based (black) and markerless (red) tracking systems.

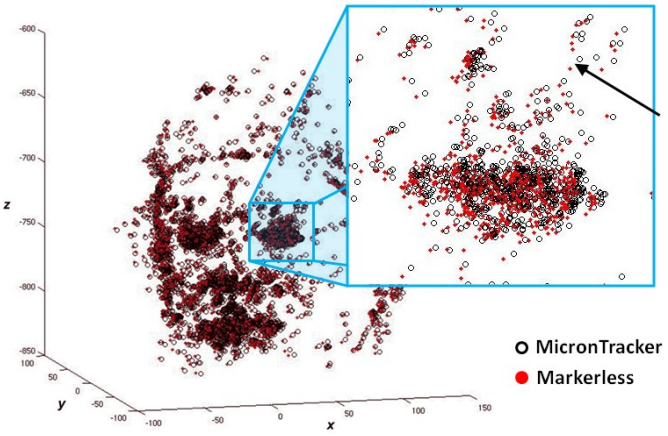

Fig. 6. Head landmarks estimated using the markerless (red) and marker-based (black) pose estimates (shown superimposed). The zoomed view shows the discrepancy more clearly (e.g. see arrow). This discrepancy was quantified in terms of the root mean square distance (Euclidean) computed over all landmarks (see Table 2).

Table 2. RMSE between the two tracking systems for each volunteer.*

Volunteer	RMSE (mm)			
	0-1k	0-2k	0-3k	0-4k
1	1.8	1.8	1.9	1.9
2	1.5	1.6	1.8	1.9
3	1.0	1.2	1.3	1.4
4	1.3	1.4	1.5	1.6
5	1.0	1.2	1.5	2.0
6	1.5	2.1	2.3	2.4
7	1.5	1.9	2.0	2.2
8	2.6	3.2	3.5	3.9
9	2.3	2.8	3.3	3.9
10	1.9	2.1	2.1	2.3
11	1.6	2.0	2.1	2.8
12	1.2	1.5	1.9	2.3
13	3.4	4.4	4.9	4.9
14	1.8	2.1	2.5	3.1
15	-	-	-	-
16	1.9	2.2	2.2	2.1

* Note that no data are shown for volunteer 15 due to a loss of synchronization between the two tracking systems during the acquisition.