Marker-free optical stereo motion tracking for in-bore MRI and PET-MRI application

Andre Z. Kyme[a)]
School of Biomedical Engineering, Faculty of Engineering and Computer Science, University of Sydney, Sydney, Australia
The Brain & Mind Centre, University of Sydney, Sydney, Australia

Murat Aksoy
Department of Radiology, Stanford University, USA

David L. Henry
The Brain & Mind Centre, University of Sydney, Sydney, Australia
Faculty of Health Sciences, University of Sydney, Sydney, Australia

Roland Bammer and Julian Maclaren
Department of Radiology, Stanford University, USA

(Received 28 November 2019; revised 12 April 2020; accepted for publication 15 April 2020; published xx xxxx xxxx)

Purpose: Prospective motion correction is arguably the "silver bullet" solution for magnetic resonance imaging (MRI) studies impacted by motion, applicable to almost any pulse sequence and immune from the spin history artifacts introduced by a moving object. In prospective motion correction, the magnetic field gradients and radio frequency waveforms are adjusted in real time in response to measured head motion so as to maintain the head in a stationary reference frame relative to the scanner. Vital for this approach are accurate and rapidly sampled head pose measurements, which may be obtained optically using cameras. However, most optical methods are limited by the need to attach physical markers to the skin, which leads to decoupling of head and marker motion and reduces the effectiveness of correction. In this work we investigate the feasibility and initial performance of a stereo-optical motion tracking method which does not require any attached markers.
Methods: The method relies on detecting distinctive natural features or amplified features (using skin stamps) directly on the forehead in multiple camera views, and then deriving pose estimates via a 3D-2D registration between the skin features and a database of forehead landmarks. To demonstrate the feasibility and potential accuracy of the marker-free method for discrete (step-wise) head motion, we performed out-of-bore and in-bore experiments using robotically and manually controlled phantoms in addition to in-bore testing on human volunteers. We also developed a convenient out-of-bore test bed to benchmark and optimize the motion tracking performance.
Results: For out-of-bore phantom tests, the pose estimation accuracy (compared to robotic ground truth) was 0.14 mm and 0.23 degrees for incremental translation and rotation, respectively. For arbitrary motion, the pose accuracy obtained using the smallest forehead feature patch was equivalent to 0.21 ± 0.11 mm positional accuracy in the striatum. For in-bore phantom experiments, the accuracy of rigid-body motion parameters (compared to wireless MR-sensitive markers) was 0.08–0.41 ± 0.18 mm/0.05–0.3 ± 0.12 deg and 0.14–0.16 ± 0.12 mm/0.08-0.17 ± 0.08 deg for the small and large feature patches, respectively. In vivo results in human volunteers indicated sub-millimeter and sub-degree pose accuracy for all rotations and translations except the depth direction (max error 1.8 mm) when compared to a registration-based approach.
Conclusions: In both bench-top and in vivo experiments we demonstrate the feasibility of using very small feature patches directly on the skin to obtain high accuracy head pose measurements needed for motion-correction in MRI brain studies. The optical technique uses in-bore cameras and is consistent with the limited visibility of the forehead afforded by head coils used in brain imaging. Future work will focus on optimization of the technique and demonstration in prospective motion correction.
 [https://doi.org/10.1002/mp.14199]

Key words: motion tracking, magnetic resonance imaging, prospective motion correction

1. INTRODUCTION

Head motion can limit the spatial and quantitative accuracy of MR measurements even more so than factors such as scanner design and magnetic field strength.[1] Motion of 1 mm amplitude may introduce blurring and ghosting artifacts in anatomical imaging, false activations in functional MRI (especially when head motion is correlated with the task), bias in the functional parameters derived from diffusion-weighted MRI, and reduced spectral resolution and reproducibility in MR spectroscopy.[2–6] Noninvasive head immobilization and support devices can reduce the amplitude of motion but do not fully prevent it.[7] It is not surprising, therefore, that current MRI research includes a strong focus on motion correction methods.

Many motion correction methods have been proposed for MRI, and several are in widespread use clinically. Motion correction methods can be broadly classified into retrospective and prospective approaches, which aim to estimate a consistent set of *k*-space data offline and online, respectively. Widely used retrospective methods include PROPELLER[8] and SNAILS,[9] both of which rely on redundant sampling of the center of *k*-space during each sequence repetition time. While such methods are useful, they cannot fix spin history effects related to motion, and oversampling lowers scan efficiency and can lead to colored noise effects. They also require that you have time in the pulse sequence and do not perturb signal evolution, which is difficult with spoiled gradient recalled (SPGR) or fast spin echo (FSE) readout.[10]

Prospective motion correction is a general approach to compensate for head motion in MRI. It involves real-time adjustment of the magnetic field gradients and radio frequency (RF) waveforms during scanning to maintain a fixed spatial relationship between the head and imaging volume.[1] Unlike retrospective motion correction methods, prospective motion correction automatically compensates for the spin history effects introduced by a moving object.

Accurate and rapidly sampled head pose (position and orientation) estimates are vital for real-time gradient and RF waveform control in prospective motion correction. One way to obtain motion estimates is via navigator sequences interleaved with the main sequence. Navigator-derived motion estimates can either be obtained directly from *k*-space subsampling based on well-known mathematical relationships between translations/rotations in the spatial domain and phase shifts/rotations in the Fourier domain,[11,12] or can be based on image registration in the spatial domain. Examples of the latter include regular sampling and alignment of orthogonal 2D slices[13] or whole volumes.[14]

Although navigator approaches for motion estimation in prospective motion correction have the benefit of being MRI-based, and so do not require additional hardware, they necessarily use up some of the available time within the sequence cycle and, therefore, are not generalizable to all sequences. This has motivated the use of independent motion tracking to obtain motion estimates. Of the available motion tracking methods, optical approaches have often been preferred due to the potential for very high accuracy.[15] In MRI, however, there is always the challenge of line-of-sight for optical systems due to the very tight space constraints imposed by the bore and the presence of the head coil. Thus, there has been a related effort in recent years to improve line-of-sight by developing fully MR-compatible in-bore optical tracking systems which can operate close to the subject's face.[16–20]

Despite the technological advances in prospective MC based on in-bore camera-based motion estimates, three major challenges remain for practical implementations. Firstly, nearly all current optical tracking approaches rely on the noninvasive fixation of physical markers to the head. To fully correct for head motion, perfectly rigid fixation of these markers is necessary. In practice, however, decoupling of head and marker motion easily and often occurs. Indeed, although the accuracy of marker-based tracking may be a few tens of microns in bench-top experiments,[16] it may be several millimeters in the clinic due to the difficulty of attaching markers rigidly.[1] This error is often larger than the ~ 1-mm intrinsic spatial resolution typical of clinical MRI scanners. Numerous marker fixation methods have been used in MRI and other imaging modalities to reduce the decoupling problem (e.g. attachment to caps, dental molds, bandages, goggles and skin[21–24]), but no approach has found widespread use in any imaging modality. Secondly, marker-based experiments can be challenging to set up if they require a high degree of cooperation from the subject. And thirdly, the presence of a head coil creates an enclosed environment which limits visibility of the head. Therefore, although the development of in-bore MR-compatible cameras has helped with line-of-sight issues and also increased the potential accuracy of motion estimates by reducing the working distance, it has also presented a new challenge: how to leverage this potential accuracy when only small patches of the face are visible through a head coil.

To overcome these issues and to facilitate motion-compensated imaging that is highly accurate, practical and reliable, a reassessment of conventional motion tracking methodology is needed. Our goal is to address the challenges through the development of a marker-free head tracking approach in which pose is accurately determined from skin features directly on the skin and confined to very small patches of the forehead. We adapt a feature-based tracking method previously developed for rats[25] and investigate its feasibility for brain MRI using out-of-bore and in-bore phantom and volunteer experiments with ground truth motion. In the remainder of the paper we describe the feature-based tracking method (Section 2A, 2B); an out-of-bore robotic test bed and validation (Sections 2C, 3B); an in-bore phantom and in vivo validation (Sections 2D, 2E, Sections 3C, 3D); and the implications of our results and future plans (Section 4).

2. METHODS

2.A. Overview of feature-based motion tracking and pose estimation

The feature-based motion tracking system is a two-view adaptation of the four-view system developed for estimating the head pose of freely moving rats during positron emission tomography.[25] The two spatially calibrated cameras in the two-view system acquire synchronized frames in which we detect highly distinctive native features.[26] SIFT features are small image patches characterized by extreme gradients across multiple resolutions. These features are mapped to descriptors (128-element vectors) based on radial sampling of the image gradients in a patch, and in a way that ensures scale, shift and rotation invariance. This makes SIFT descriptors well-suited for robust matching across disparate views from multi-camera systems. Since the cameras are spatially calibrated, any corresponding features that are matched across the two views can be triangulated, resulting in a 3D landmark.[27] As more stereo frames are acquired, we amass a growing database of such landmarks, along with their associated descriptors. At any given time, this landmark database represents a sparse 3D model of the object. Finally, by matching SIFT features in a stereo frame to the landmark database, the changing object pose can be estimated using a 3D-to-2D registration in which the reprojection error is minimized.[28] Namely, if a new camera frame shares feature descriptors with the database, we use a Gauss-Newton approach to find the rigid-body transformation (translation and orientation) resulting in the best alignment between the database descriptors and frame descriptors. The method is summarized in Fig. 1 and further details can be found in Kyme et al.[25] Incorrect feature matches, referred to as "outliers", generally contribute noise to the pose estimation. We use several outlier rejection methods to detect and ignore such spurious matches: (a) a multiple-match filter, which immediately rejects any features that are matched to more than one other feature; (b) a geometric filter, which ensures the plane connecting the candidate feature match and the triangulated landmark satisfies epipolar geometry[27]; and (c) a statistical filter, which recognizes and rejects matches that contribute the most error during pose estimation.

2.B. Camera and coil setup

The feature-based system comprises two MR-compatible USB2 CMOS cameras (HobbitView Inc., 640x360 or 1284x724 resolution) fitted with wide (80°) field-of-view (FoV) lenses. The sensors were securely mounted to a standard 8-channel head coil (3T GE Healthcare) using a 3D-printed bracket [Fig. 2(a)]. Both sensors had line-of-sight through the same channel opening of the head coil. The sensors had a nominal physical separation (stereo baseline) of 50 mm and subtended an angle of approximately 30 deg. Lens focus was adjusted for a working distance of 8–10 cm. Stereo camera calibration was performed by programming a 6-axis robot to move a checkerboard pattern of 4 mm squares to 30 arbitrary poses within the FoV of both cameras [Fig. 2(b), 2(c)]. Intrinsic and extrinsic camera parameters were then computed using the Matlab Calibration Toolbox.[29,30]

2.C. Out-of-bore robotic test platform

The out-of-bore test platform [Fig. 3(a)] was based around a 6-axis robotic arm (C3-A601ST, Epson America Inc.) capable of highly repeatable (20 μm) arbitrary rigid-body motions. The robot was fixed securely in front of the camera-mounted head coil and used to apply discrete arbitrary motion in six degrees of freedom to a polystyrene head phantom inside the coil. An image from Subject #5 of the face recognition database[31] was color-printed on paper and glued to the forehead of the phantom to simulate a realistic skin surface [Fig. 3(b)]. The phantom was tested with and without a small (25 mm × 6.4 mm) and large (50.8 mm × 19.1 mm) feature patch, consisting of arbitrary text and symbols, above the left eyebrow to simulate a concentrated source of features on the skin [Fig. 3(c)]. Feature patches were applied using a custom-made stamp with skin-safe black ink. This approach avoided any physical attachment of features, which is characteristic of traditional marker-based tracking methods.

Two motion experiments were performed using the robotic test platform. First, the robot was programmed to execute an

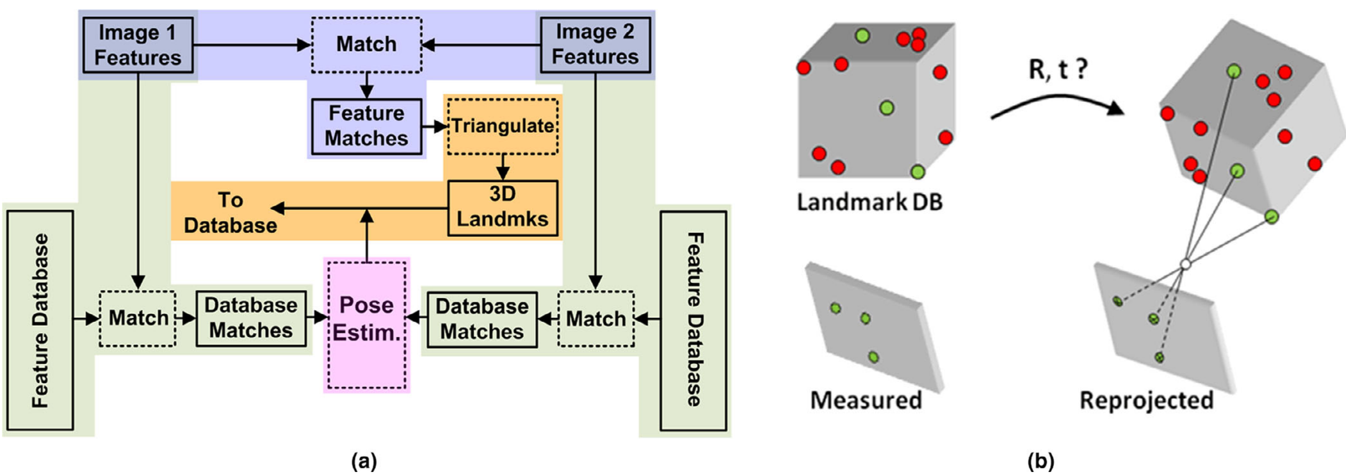

FIG. 1. Principle of feature-based motion estimation. (a) Overview: in a given stereo frame comprising two synchronized and calibrated camera views, distinctive features are detected and matched (blue) and then triangulated to produce 3D landmarks (orange) which are stored in a database along with their corresponding feature descriptors. Collectively, the database landmarks form a sparse representation of the tracked object. For each stereo frame, features are also matched to the database (green) to enable pose estimation (pink). Note that dashed boxes refer to processes and non-dashed boxes to data (inputs/outputs). (b) Pose estimation: after identifying feature matches (green) between all possible features (red) in the landmark database and those in the image, pose is estimated via a 3D-2D registration. Specifically, we determine the rotation, **R**, and translation, **t**, minimizing the reprojection error between measured and estimated feature locations. Further details can be found in Kyme et al[25].

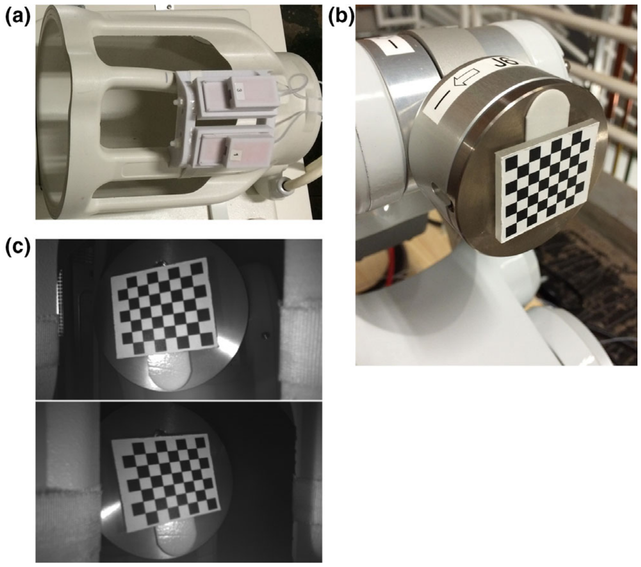

FIG. 2. Camera setup and calibration. (a) Cameras mounted to the head coil with line-of-sight through a channel opening. (b) Robotic manipulation of a checkboard marker for the extrinsic (stereo) camera calibration. (c) View of the first calibration pose from the two cameras. A total of 30 calibration poses were used for the extrinsic calibration.

incremental one-dimensional translation along the head-feet (z) axis from 1–10 mm in 1 mm steps and then an incremental rotation around the head-feet axis (head "shaking") from 1–10 deg in 1 deg steps. Second, we applied 19 arbitrary rigid-body poses to the phantom simulating arbitrary head motion. In each experiment, synchronized stereo camera images were collected at successive poses for offline motion estimation. For the incremental motion test, estimated poses were compared directly to robotic poses by using the pure rotation and translation magnitudes, eliminating the need for a formal cross-calibration. For the arbitrary motion test, a cross-calibration was used to convert estimated poses to the robot frame for comparison with the ground truth. The cross-calibration was obtained using Horn's method[32] with 30 paired pose measurements of a calibration marker, the center of which coincided with the robot tool center (Fig. 4, left).

2.D. In-bore phantom experiment

To demonstrate that the motion tracking system was functioning correctly in a MR environment, we applied discrete movements to a phantom inside a MR750 3T scanner (GE Healthcare). It was not possible to use the robot in this case since it was not MRI compatible. Therefore, a custom-designed spherical plastic phantom (Fig. 5) was fitted with three wireless active markers[23] to track motion directly in the MR frame and serve as ground truth. To cross-calibrate the MR and motion tracker frames, we rigidly attached an 8-point optical marker to the phantom (Fig. 4, right) and collected simultaneous MR and optical measurements at approximately 50 discrete poses, which were applied manually to the phantom inside the camera-mounted head coil. A detailed description of the calibration tool, wireless markers and tracking sequence is available in Maclaren et al.[19] These pose data were input to a hand-eye cross-calibration to determine a closed-form solution relating the MR and tracker frames.[33]

We then replaced the optical calibration marker with the same ink stamps used in the out-of-bore robotic test platform (Section 2C) and manually moved the phantom to approximately 15 discrete poses during a single acquisition. At each pose, a synchronized stereo image frame from the motion tracking system was acquired together with continuous measurements from the wireless MR markers. The wireless marker data were processed offline to identify the step-wise transitions and compute an average pose per step. Pose estimates from the feature-based tracking system were also computed offline and compared with the ground truth MR-based pose measurements after applying the cross-calibration.

2.E. In-bore volunteer experiment

Two volunteers were scanned on a GE MR750 3T MR system and instructed to move their head to six different poses in turn. The volunteers had the same ink stamps on their forehead as were used in both the robotic test platform and in-bore phantom experiments. At each of the six poses, the volunteer was imaged for 45 s using a 3D FSPGR navigator

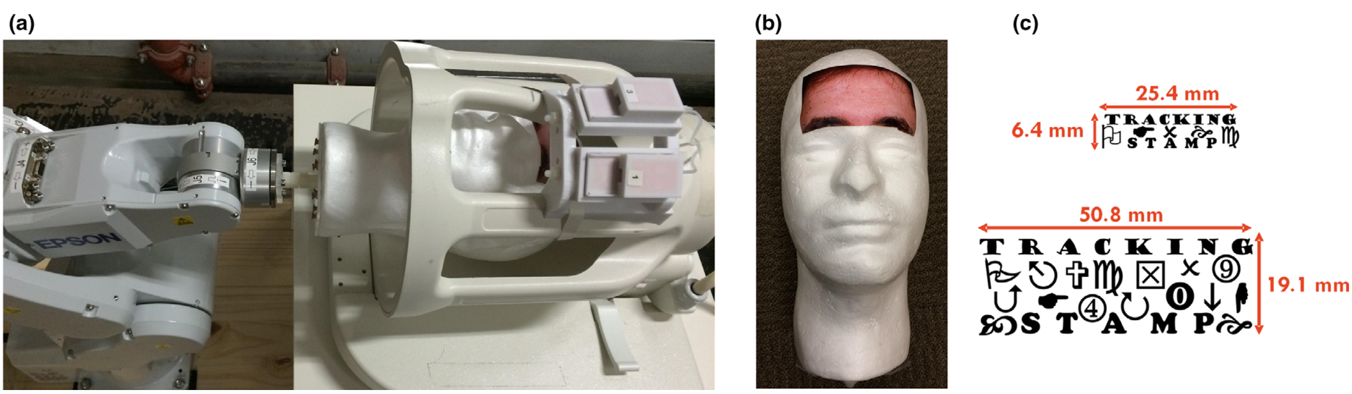

FIG. 3. Out-of-bore test platform. (a) Epson C3 6-axis robot to control a styrofoam mannequin head inside the head coil. (b) Close-up of the mannequin with a photograph of a forehead skin patch affixed, allowing the setup to be tested using realistic geometry and texture. (c) The small and large feature stamps applied to the skin patch to amplify the number of features derived from the forehead.

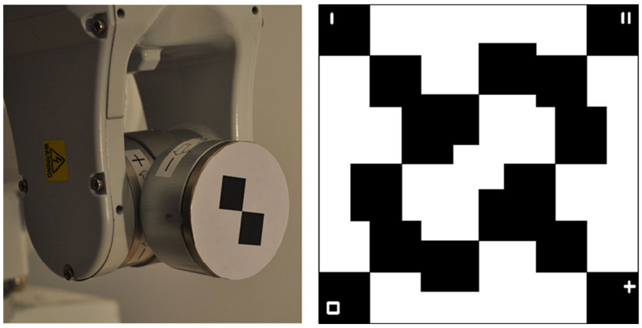

FIG. 4. Cross-calibration marker for the out-of-bore (left) and in-bore (right) experiments.

(TE: 1.5 s, TR: 4.9 s). Reconstructed image volumes from the six scans were aligned retrospectively to the first volume using Statistical Parametric Mapping (SPM) image registration and directly compared with movements estimated from the feature-based motion tracking system by applying the cross-calibration from Section 2D. Finally, the feature-based estimates were used for retrospective alignment of the images for motion correction and the resulting mean images were compared to the SPM-based motion correction results. Human volunteer experiments were performed in compliance with IRB protocols.

3. RESULTS

In the following results, we use the designations x, y and z to specify axis directions. These directions map to the standard brain imaging coordinate system designations left/right (L/R), superior/inferior (S/I) and anterior/posterior (A/P), respectively.

3.A. Camera calibration

The principal point, focal lengths and distortion coefficients estimated from the intrinsic camera calibration resulted in a mean reprojection error of 0.02 ± 0.02 mm at the working distance of the forehead. This error is 1–2 orders of magnitude below our minimum motion tracking accuracy requirement ($\leq$1 mm) and therefore can be considered negligible. The extrinsic (stereo) calibration indicated a stereo baseline of 50.0 ± 0.1 mm and an angle of 37.0 ± 0.16 deg subtended by the two cameras. Figure 6 shows the stereo calibration output. The coordinate system for tracking was arbitrarily chosen to align with the left camera frame.

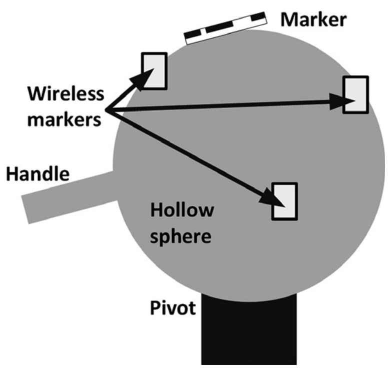

FIG. 5. Phantom rig for the in-bore experiments. The spherical phantom had three wireless active markers and one feature-based marker affixed for simultaneous tracking in the MRI and tracker frames, respectively.

3.B. Out-of-bore robotic test bed

The mean camera/robot cross-calibration error was 0.09 mm (max 0.18 mm), 0.07 mm (max 0.13 mm) and 0.10 mm (max 0.25 mm) in x, y and z, respectively. Figure 7(a) shows synchronized camera images of the mannequin head with a bare (non-stamped) forehead for a single frame and [Fig. 7(b)] shows the detected and matched features after background masking.

Figure 8 shows a comparison of the estimated incremental translation and rotation versus the ground truth robotic motion for the case of a bare forehead. The mean absolute error was 0.14 mm for translation and 0.23 deg for rotation. For arbitrary motion, it can be more intuitive to measure accuracy in terms of the measured and expected location of specific test points relevant to the application. Thus, Fig. 9 shows the discrepancy (in millimeters) between the ground truth and estimated positions of two test points, one located on the forehead surface and one in the striatum. The physical displacement of each test point from pose to pose is also shown (right axis) for comparison. For the forehead test point, the average error was 0.19 ± 0.09 mm, 0.22 ± 0.11 mm and 0.19 ± 0.11 mm for the bare forehead, feature stamp 1 (small) and feature stamp 2 (large), respectively. The three regimes therefore performed very similarly for a test point located within the region sampled by the stereo calibration, suggesting that the cross-calibration error was the limiting factor in this case. However, for the striatum test point, differences between the regimes were more obvious: average error was 0.41 ± 0.10 mm, 0.21 ± 0.11 mm and 0.19 ± 0.09 mm for the bare forehead, feature stamp 1 (small) and feature stamp 2 (large), respectively. In this case, having more features led to improved accuracy.

3.C. In-bore phantom experiment

Figure 10 shows the six degree-of-freedom motion tracking data from the in-bore phantom experiment using the smaller and larger feature stamps and Table 1 compares the mean absolute error ($\pm$1 SD) between the feature-based (estimated) and wireless marker-based (ground truth) estimates. For the smaller feature stamp the mean absolute error was $\leq$0.3 deg for rotations and $\leq$0.41 mm for translations; for the larger stamp the mean absolute error was $\leq$0.17 deg for rotations and $\leq$0.16 mm for translations. Although the motion tested had a larger amplitude than would typically be expected in practice, our aim in these experiments was to demonstrate tracking performance and capability across a full motion range and for all degrees of freedom.

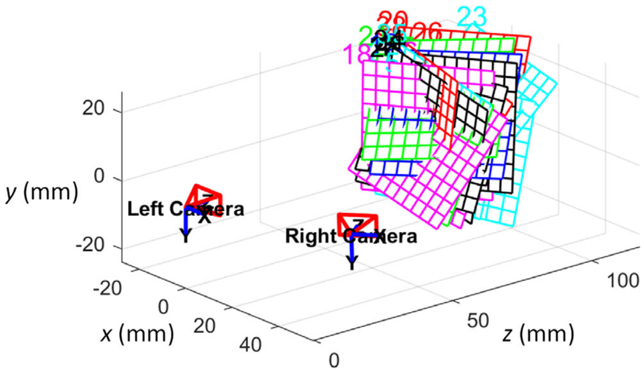

FIG. 6. Extrinsic (stereo) camera calibration. The relative position and orientation of the two cameras is shown with respect to the 30 checkboard orientations used for calibration. The tracker frame was arbitrarily chosen to coincide with that of the left camera.

3.D. Human volunteers

Figure 11 shows each of the volunteers with the small feature stamp imprinted above the left eyebrow and an example of feature detection and matching within a single stereo frame. Figure 12 shows a comparison of the feature-based motion estimates (using the smaller stamp) and SPM-based motion estimates for the first volunteer, indicating sub-millimeter and sub-degree agreement in all cases except the z-translation (depth) component for some of the poses (maximum error was 1.8 mm). The poorer agreement for the z-translation component was again likely due to error propagating from the cross-calibration. We also note that because the registration-based poses used here for reference were almost certainly impacted by intra-scan motion, we do not expect an exact match to the feature-based poses. Figure 13 shows the motion correction result obtained by using the motion estimates to register the navigator volumes for each volunteer. It is clear that blurring was reduced and contrast was increased after motion correction using both methods, and that the feature-based approach gave qualitatively comparable results.

4. DISCUSSION

The aim of this work was to investigate the feasibility of accurately estimating head motion using a close-range binocular head tracking approach that relies on skin features confined to very small patches of the forehead. The approach was tested out-of-bore using controlled robotic step-wise motion in conjunction with a realistic test phantom. It was also tested in-bore on phantoms and human volunteers undergoing discrete motion, where ground truth was provided via wireless active markers or image registration. Our results demonstrate that even feature-rich patches with an area $<2\ \mathrm{cm}^2$ are sufficient for sub-millimeter and sub-degree rigid-body head pose measurement using this tracking method. Specifically, for out-of-bore phantom tests, the pose estimation accuracy (compared to robotic ground truth) was 0.14 mm and 0.23 deg for incremental translation and rotation, respectively. For arbitrary motion, the pose accuracy obtained using the smallest forehead feature patch was equivalent to 0.21 ± 0.11 mm positional accuracy in the striatum. For in-bore phantom experiments, the accuracy of rigid-body motion parameters (compared to wireless MR-sensitive markers) was 0.08–0.41 $\pm$ 0.18 mm/0.05–0.3 $\pm$ 0.12 deg and 0.14–0.16 $\pm$ 0.12 mm/0.08–0.17 $\pm$ 0.08 deg for the small and large feature patches, respectively. And the in vivo results in human volunteers indicated sub-millimeter and sub-degree pose accuracy for all rotations and translations except the depth direction (max error 1.8 mm) when compared to a registration-based approach.

In this pilot study, we used a stamp on the skin to ensure that sufficient features were available for tracking. A key advantage of our approach over marker-based tracking methods is that patches can be printed directly onto the skin. This removes the potential for decoupling of head and marker

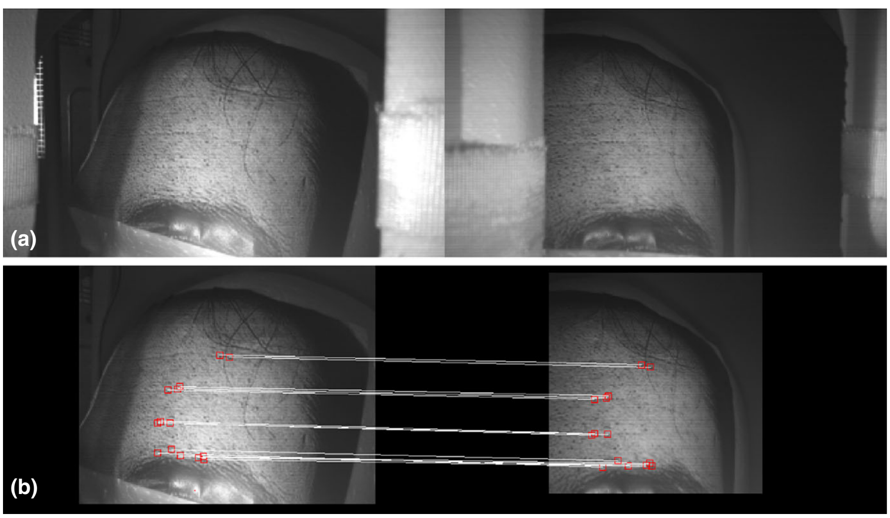

FIG. 7. Out-of-bore robotic phantom feature tracking and matching. (a) Raw synchronized images of the head phantom from Camera 1 (left) and Camera 2 (right). (b) Images after background masking showing detected and matched features.

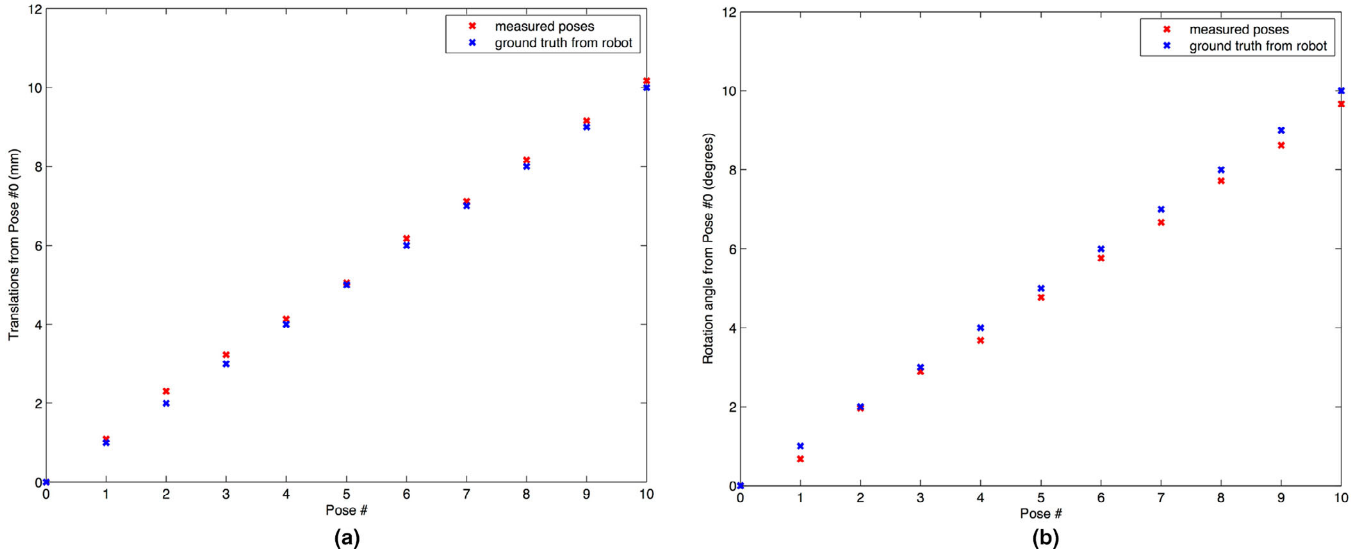

FIG. 8. Out-of-bore feature-based estimation of incremental translational (a) and rotational (b) robotic motion of the head phantom.

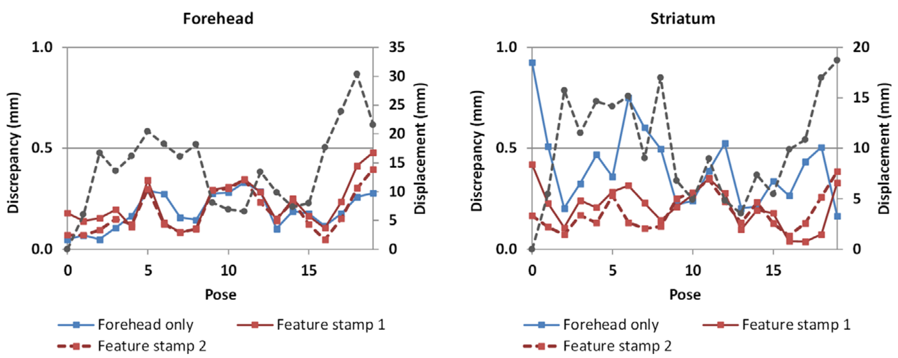

FIG. 9. Discrepancy between the measured and expected location of a test point on the forehead surface (left) and in the center of the brain (right) undergoing arbitrary motion. The physical displacement of the test point for each movement is plotted on the right axis for reference (gray dashed line).

motion that readily occurs when markers are attached to goggles and caps or with adhesive to the skin. Another advantage of our method is that the entire feature patch need not be visible to the cameras at any given time – features can move in and out of the FoV of the cameras and still be useful. For this reason, larger feature patches may be more beneficial even though the FoV remains very focused due to both the proximity of the cameras to the forehead and the line-of-sight restrictions imposed by the head coil.

We are aware of at least one other marker-free tracking approach, based on structured light, which has been reported for use in PET and MRI.[34,35] Because structured light methods rely on the registration of surfaces to infer pose changes, robustness is usually contingent on sampling areas of high topological change, such as the bridge of the nose and around the eyes. This requirement potentially limits the flexibility of camera positioning and compatibility with the variety of enclosed head coils used in MRI. By contrast, our sparse feature-based method is not restricted by surface topology and works well anywhere on the forehead, making it very flexible in the space-constrained environment. Computation of pose from a sparse set of features is also computationally much less demanding than dense surface registration which, for real-time rates, would usually necessitate specialized hardware.[35]

There are several limitations with the present implementation of our motion tracking system and the initial benchmarking experiments reported here:

(i) The camera boards used in this study operated entirely independently, without any synchronization or triggering capability. This hardware limitation restricted testing to discrete, step-wise motion. However, each camera is capable of acquiring at >30 Hz and, indeed, the authors have demonstrated high frame-rate monocular performance with these cameras previously.[16] Thus, there is no "in-principle" temporal limitation due to the cameras: with hardware synchronization, the system would be capable of more than sufficient temporal resolution for human head tracking. We are currently implementing a custom stereo board for synchronized tracking at up to 100 Hz.

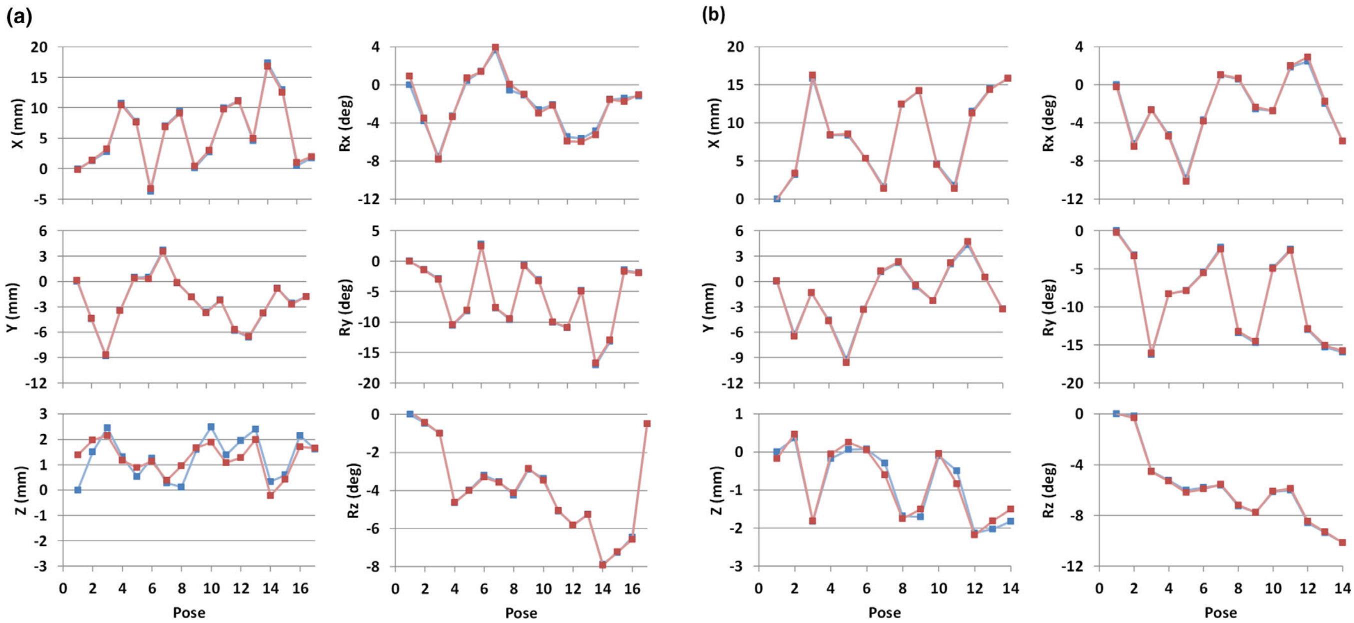

FIG. 10. Motion estimation for the in-bore phantom. Feature-based (blue) and wireless active marker-based (red) motion parameters estimated for the arbitrary motion of the in-bore phantom using (a) the smaller feature stamp and (b) the larger feature stamp.

TABLE 1. Mean absolute error (±1 SD) between feature-based and wireless marker-based motion estimates for in-bore phantom motion.

	R_x (deg)	R_y (deg)	R_z (deg)	X (mm)	Y (mm)	Z (mm)
Small stamp	0.30 (0.23)	0.14 (0.08)	0.05 (0.05)	0.29 (0.14)	0.08 (0.06)	0.41 (0.34)
Large stamp	0.17 (0.13)	0.15 (0.07)	0.08 (0.05)	0.14 (0.13)	0.15 (0.12)	0.16 (0.11)

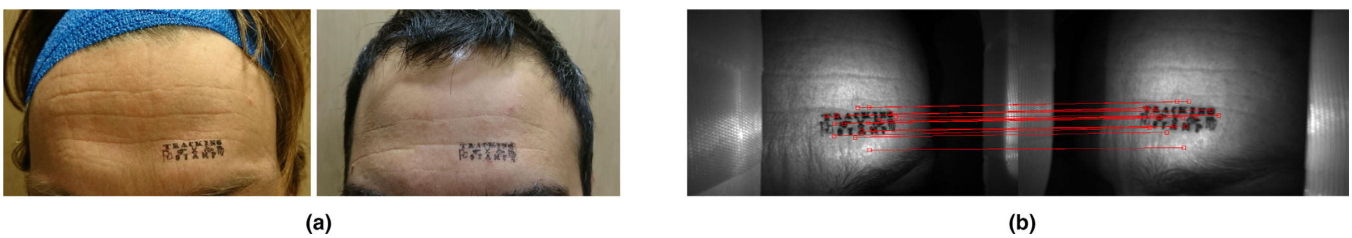

FIG. 11. In vivo volunteer tests. (a) Photo showing each volunteer with the small feature stamp imprinted on the forehead; (b) example of feature detection and matching for a single stereo frame for one of the volunteers.

(ii) The restriction to discrete motion using the current camera hardware meant that in vivo tests lacked a suitable ground truth. Choosing 45 s navigator scans was a compromise: short enough to minimize head motion but long enough to provide adequate volume images for robust registration. Nevertheless, it was not possible to avoid head motion during the scans and the quality of the individual navigator volumes was still rather poor; thus we would not expect highly reliable registration parameters and motion correction. The problem of intra-scan motion was mitigated to some extent by using an average pose from the tracking system for each step-wise motion, however the comparison is still inherently limited.

(iii) Poses are currently computed offline, and in order for the method to be applied in prospective motion correction, the computational efficiency must be improved to allow real-time pose processing with low latency. No optimization of the computational efficiency was attempted in this work. We note, however, that offline motion correction is normal in PET and thus the current approach would be suitable for the PET component of a hybrid PET-MRI acquisition.

(iv) Although the two stamps used in this study did amplify the number of native forehead features, they still had a rather low feature density compared to a typical real-world scene. This was a limitation of the rubber stamp

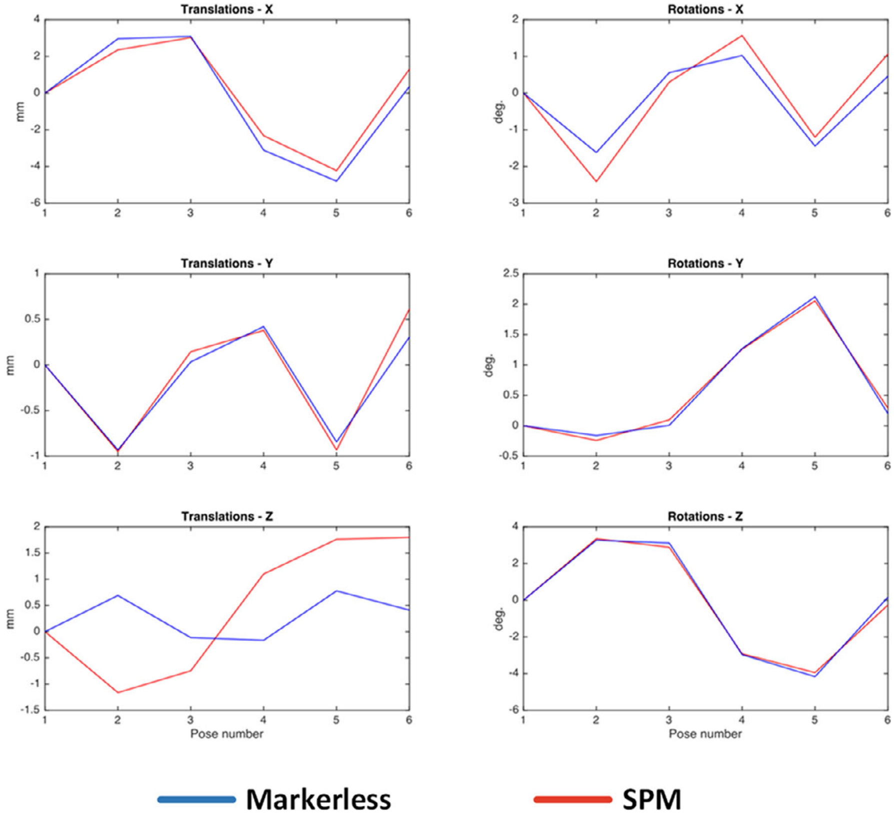

FIG. 12. In vivo motion estimation results for volunteer 1 using the smaller feature stamp. All six degrees of freedom are plotted for the feature-based (blue) and SPM registration-based (red) pose estimates relative to the initial pose.

fabrication. Thus, there is scope to adapt our approach to alternative skin patch-based methods which leverage a higher feature density for improved accuracy. This will be investigated in future work.

(v) Currently we make no attempt to handle non-rigid deformation of the forehead and instead assume that the feature patches move rigidly. In reality, facial movement can and will violate this assumption, leading to erroneous pose estimation. We are therefore working on developing methods to decouple rigid motion from "mixed" (rigid + non-rigid) motion of the forehead based on traditional computer vision approaches[36,37] and more recent methods involving deep neural networks.[38] However, it is also important to point out that the use of attached markers is not immune from non-rigid motion of the forehead. Indeed, non-rigid motion of the forehead can cause a rigid marker to move very differently to the brain, and it is impossible to know this from the marker measurements alone. Thus, an important advantage of our method is that by directly tracking the skin we can detect and quantify the deformation. This in turn means that our approach may be used to switch prospective motion tracking on and off based on the absence/presence of intermittent non-rigid deformation of the forehead.

(vi) Finally, notwithstanding the benefits of imprinting features directly onto the skin, avoiding the need to apply anything to the skin is still preferable since the clinical goal is seamless motion correction that requires no input from the technologist or cooperation from the patient. Future work should, therefore, explore the use of different feature detectors or lighting schemes (e.g. UV, IR) to best highlight exclusively native features for tracking.

Despite these limitations, it is clear from our results that an accuracy of 0.2 mm and 0.2 deg for translation and rotation, respectively, is certainly feasible with our approach. Moreover, two factors suggest that the motion tracking performance we report here for our system is somewhat conservative. Firstly, the residual error in the motion estimates for the out-of-bore phantom experiments is in part due to cross-calibration error resulting from a mismatch between the cross-calibration volume (the immediate vicinity of the robot tool) and the volume from which features were derived for pose estimation (the mannequin forehead, approximately 250 mm from the cross-calibration volume). We did not quantify the cross-calibration error, but it is well known that this error increases rapidly away from the cross-calibrated volume.[39] Secondly, no geometric and/or algorithmic (e.g. feature detection, matching) optimization was performed in this work. There is, therefore, scope to refine all of these aspects of the method to improve performance. This is the focus of ongoing work.

In summary, the close-range stereo-optical motion tracking approach reported here, which relies on native skin features to estimate rigid-body motion in highly space-

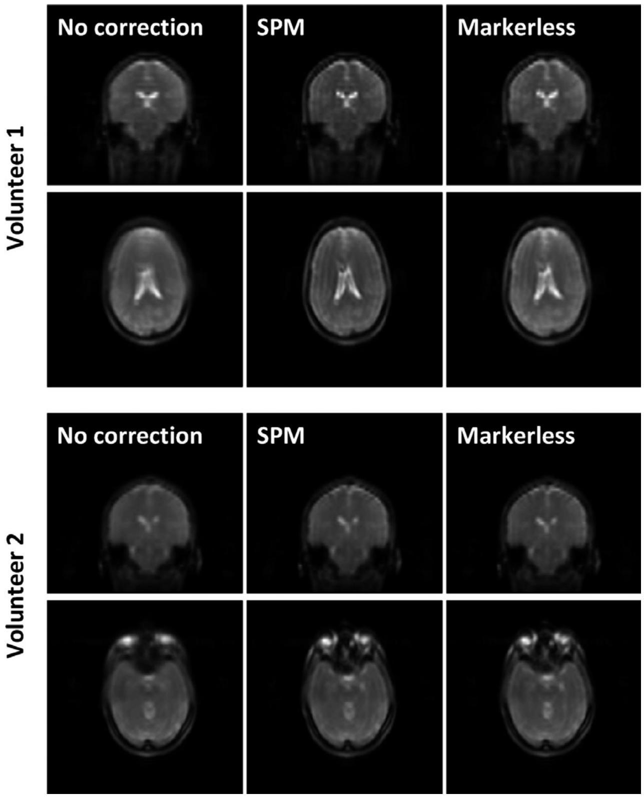

FIG. 13. Motion correction of MR-based navigators for the in vivo studies. Coronal and axial reformats of the 3D navigator data from volunteer 1 and 2 are shown for the mean unregistered (i.e. non motion-corrected) image from all six poses (left), the mean image data after registration using SPM (middle), and the mean image data after registration using pose estimates derived from the feature-based tracking method (right).

constrained conditions, appears capable of delivering sub-millimeter and sub-degree accuracy, at least for discrete motion. It remains to be tested for continuous motion with fully synchronized cameras. The approach therefore appears to be a promising motion tracking candidate to support prospective motion correction within the tight geometry and limited line-of-sight conditions afforded by modern multi-channel head coils inside MRI and PET-MRI scanners.

5. CONCLUSIONS

We report efforts towards highly accurate and convenient optical tracking of head motion within the extremely tight space constraints of an MRI scanner and head coil. This is vital for the clinical viability and optimal performance of prospective motion correction in standalone MRI and hybrid MRI-PET systems. Using very small feature patches imprinted directly on the skin for marker-free optical pose estimation, we avoid head-marker decoupling and demonstrate the potential for sub-degree and sub-millimeter pose estimates in phantom and volunteer experiments. The approach is currently limited to tracking discrete motion due to hardware limitations, however our ongoing work focuses on developing the system to handle continuous motion, non-rigid motion, and eliminating the need for imprinting the skin.

ACKNOWLEDGMENTS

A. Kyme was supported by a Cassen Fellowship, Education and Research Foundation, SNMMI, USA, during part of this research, and also an ECR Research Innovation Award, Faculty of Health Sciences, University of Sydney. Support for the project was also obtained by the NIH (2R01 EB002711, 5R01 EB008706, 5R01 EB011654) and the Center of Advanced MR Technology at Stanford (P41 RR009784), Lucas Foundation.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

[a)]Author to whom correspondence should be addressed. Electronic mail: andre.kyme@sydney.edu.au; Telephone: +612 93512660

REFERENCES

1. Maclaren J, Herbst M, Speck O, Zaitsev M. Prospective motion correction in brain imaging: a review. *Mag Reson Med*. 2013;69:621–636.
2. Schulz J, Siegert T, Bazin PL, et al. Prospective slice-by-slice motion correction reduces false positive activations in fMRI with task-correlated motion. *NeuroImage*. 2014;84:124–132.
3. Brown TT, Kuperman JM, Erhart M, et al. Prospective motion correction of high-resolution magnetic resonance imaging data in children. *NeuroImage*. 2010;53:139–145.
4. Hajnal JV, Myers R, Oatridge A, Schwieso JE, Young IR, Bydder GM. Artifacts due to stimulus correlated motion in functional imaging of the brain. *Magn Reson Med*. 1994;31:283–291.
5. Ling J, Merideth F, Caprihan A, Pena A, Teshiba T, Mayer AR. Head injury or head motion? Assessment and quantification of motion artifacts in diffusion tensor imaging studies. *Hum Brain Mapp*. 2012;33(1):50–62.
6. Keating B, Deng W, Roddey JC, et al. Prospective motion correction for single-voxel 1H MR spectroscopy. *Magn Reson Med*. 2010;64:672–679.
7. Mawlawi O, Weiss R, Shinn AK, Pidcock J, Slifstein M, Laruelle M. Performance characteristics of a head immobilization device for PET imaging. *J Nucl Med*. 1999;40:281.
8. Pipe JG. Motion correction with PROPELLER MRI: application to head motion and free-breathing cardiac imaging. *Magn Reson Med*. 1999;42 (5):963–969.
9. Liu C, Bammer R, Kim DH, Moseley ME. Self-navigated interleaved spiral (SNAILS): application to high-resolution diffusion tensor imaging. *Magn Reson Med*. 2004;52:1388–1396.
10. Herbst M, Maclaren J, Weigel M, Korvink J, Hennig J, Zaitsev M. Prospective motion correction with continuous gradient updates in diffusion weighted imaging. *Magn Reson Med*. 2012;67:326–338.
11. Welch EB, Manduca A, Grimm RC, Ward HA, Jack CR. Spherical navigator echoes for full 3D rigid body motion measurement in MRI. *Magn Reson Med*. 2002;47:32–41.
12. van der Kouwe AJW, Benner T, Dale AM. Real-time rigid body motion correction and shimming using cloverleaf navigators. *Magn Reson Med*. 2006;56:1019–1032.
13. White N, Roddey C, Shankaranarayanan A, et al. PROMO: real-time prospective motion correction in MRI using image-based tracking. *Magn Reson Med*. 2010;63:91–105.
14. Tisdall D, Hess A, Reuter M, Meintjes E, Fischl B, van der Kouwe A. Volumetric navigators (vNavs) for prospective motion correction and selective reacquisition in neuroanatomical MRI. *Magn Reson Med*. 2012;68:389–399.
15. Kyme A.*PhD dissertation*. University of Sydney; 2012.
16. Maclaren J, Armstrong BSR, Barrows RT, et al. Measurement and correction of microscopic head motion during magnetic resonance imaging of the brain. *PLoS ONE*. 2012;7:e48088.

17. Schulz J, Siegert T, Reimer E, et al. An embedded optical tracking system for motion-corrected magnetic resonance imaging at 7T. *Magn Reson Mater Physics Biol Med*. 2012;25:443–453.
18. Qin L, van Gelderen P, Derbyshire JA, et al. Prospective head-movement correction for high-resolution MRI using an in-bore optical tracking system. *Magn Reson Med*. 2009;62:924–934.
19. Maclaren J, Aksoy M, Ooi MB, Zahneisen B, Bammer R. Prospective motion correction using coil-mounted cameras: cross-calibration considerations. *Magn Reson Med*. 2018;79:1911–1921.
20. Spangler-Bickell MG, Khalighi MM, Hoo C, et al. Rigid motion correction for brain PET/MR imaging using optical tracking. *IEEE Trans Radiat Plasma Med Sci*. 2019;3:498–503.
21. Zaitsev M, Dold C, Sakas G, Hennig J, Speck O. Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system. *NeuroImage*. 2006;31:1038–1050.
22. Bloomfield PM, Spinks T, Reed J, et al. The design and implementation of a motion correction scheme for neurological PET. *Phys Med Biol*. 2003;48:959–978.
23. Ooi MB, Maclaren J, Aksoy M, Watkins RD, Bammer R. Prospective motion correction using inductively coupled wireless RF coils. *Magn Reson Med*. 2013;70:639–647.
24. Kim J, Nuyts J, Kyme A, Kuncic Z, Fulton R. A rigid motion correction method for helical computed tomography (CT). *Phys Med Biol*. 2015;60:2047–2073.
25. Kyme A, Se S, Meikle S, et al. Markerless motion tracking of awake animals in positron emission tomography. *IEEE Trans Med Imaging*. 2014;33:2180–2190.
26. Lowe D. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis*. 2004;60:91–110.
27. Hartley R, Zisserman A. *Multiple View Geometry in Computer Vision*, 2nd edn. UK: Cambridge University Press; 2004.
28. Se S, Lowe D, Little J. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *Int J Rob Res*. 2002;21:735–758.
29. Bouguet Complete Camera Calibration Toolbox for Matlab http://www.vision.caltech.edu/bouguetj/calib_doc/. Accessed Jan 1 2019.
30. Zhang Z. A flexible new technique for camera calibration. Technical Report MSR-TR-98-71. (Microsoft Research). 2000;1–22.
31. Weyrauch B, Heisele B, Huang J, Blanz V. Component-based face recognition with 3D morphable models. *Proc 2004 Conf Comput Vis Pat Recog*. 2004;5.
32. Horn BKP. Closed-form solution of absolute orientation using unit quaternions. *J Opt Soc Am*. 1987;4:629–642.
33. Daniilidis K. Hand-eye calibration using dual quaternions. *Int J Rob Res*. 1999;18:286–298.
34. Olesen OV, Sullivan JM, Mulnix T, et al. List-mode PET motion correction using markerless head tracking: proof-of-concept with scans of human subject. *IEEE Trans Med Imaging*. 2013;32:200–209.
35. Slipsager JM, Ellegaard AH, Glimberg SL, Paulsen RR, Tisdall MD, Wighton P. Markerless motion tracking and correction for PET, MRI, and simultaneous PET/MRI. *PLoS ONE*. 2019;14:e0215524.
36. Jensen SH, Del Blue A, Doest ME, Aanæs H. A benchmark and evaluation of non-rigid structure from motion. *ArXiv*. 2018;abs/1801.08388.
37. Lebeda K, Hadfield S, Bowden R. Direct-from-Video: Unsupervised NRSfM. *ECCV Workshops*. 2016.
38. Kong C, Lucey S. Deep non-rigid structure from motion. *IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019;2019:1558–1567.
39. Fitzpatrick JM, West JB, Maurer CR. Predicting error in rigid-body point-based registration. *IEEE Trans Med Imag*. 1998;17:694–702.