Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

RECEIVED
10 January 2018

REVISED
26 March 2018

ACCEPTED FOR PUBLICATION
11 April 2018

PUBLISHED

PAPER

Markerless motion estimation for motion-compensated clinical brain imaging

Andre Z Kyme[1,2], Stephen Se[3], Steven R Meikle[2] and Roger R Fulton[2,4]

[1] Faculty of Engineering and IT, University of Sydney, Sydney, Australia
[2] Faculty of Health Sciences and Brain and Mind Centre, University of Sydney, Australia
[3] FLIR Systems Inc., Richmond, BC V6W 1K7, Canada
[4] Department of Medical Physics, Westmead Hospital, Sydney, Australia

E-mail: andre.kyme@sydney.edu.au (A Kyme)

Keywords: markerless motion estimation, feature tracking, head motion, motion-compensated imaging, PET, SPECT, CT

Supplementary material for this article is available online

Abstract

Motion-compensated brain imaging can dramatically reduce the artifacts and quantitative degradation associated with voluntary and involuntary subject head motion during positron emission tomography (PET), single photon emission computed tomography (SPECT) and computed tomography (CT). However, motion-compensated imaging protocols are not in widespread clinical use for these modalities. A key reason for this seems to be the lack of a practical motion tracking technology that allows for smooth and reliable integration of motion-compensated imaging protocols in the clinical setting. We seek to address this problem by investigating the feasibility of a highly versatile optical motion tracking method for PET, SPECT and CT geometries. The method requires no attached markers, relying exclusively on the detection and matching of distinctive facial features. We studied the accuracy of this method in 16 volunteers in a mock imaging scenario by comparing the estimated motion with an accurate marker-based method used in applications such as image guided surgery. A range of techniques to optimize performance of the method were also studied. Our results show that the markerless motion tracking method is highly accurate (<2 mm discrepancy against a benchmarking system) on an ethnically diverse range of subjects and, moreover, exhibits lower jitter and estimation of motion over a greater range than some marker-based methods. Our optimization tests indicate that the basic pose estimation algorithm is very robust but generally benefits from rudimentary background masking. Further marginal gains in accuracy can be achieved by accounting for non-rigid motion of features. Efficiency gains can be achieved by capping the number of features used for pose estimation provided that these features adequately sample the range of head motion encountered in the study. These proof-of-principle data suggest that markerless motion tracking is amenable to motion-compensated brain imaging and holds good promise for a practical implementation in clinical PET, SPECT and CT systems.

1. Introduction

Motion-compensated brain imaging can dramatically reduce the artifacts and quantitative degradation associated with voluntary and involuntary subject head motion during tomographic scans (Fulton *et al* 2002, Kyme *et al* 2003, 2012, Montgomery *et al* 2006, Yu *et al* 2007, Jin *et al* 2013, Kim *et al* 2015). Interest in these methods has grown in parallel with the steadily improving spatial resolution of modalities such as positron emission tomography (PET), single photon emission computed tomography (SPECT) and computed tomography (CT) since the improved resolution increases the sensitivity to motion artifacts. Interestingly, however, motion compensated brain imaging protocols have not found their way into routine clinical use. We postulate that a key reason for this poor translation of the science is the lack of an accurate and practical method for head motion tracking.

Motion tracking accuracy in motion-compensated imaging should be better than half the intrinsic (motion-free) reconstructed spatial resolution of the imaging modality to keep the image degradation due to motion tracking below 10%. Therefore, the motion tracking accuracy requirement for state-of-the-art clinical PET, SPECT and CT systems is approximately 2 mm, 4 mm and 0.25 mm, respectively.

Optical motion tracking is by far the most popular method for obtaining the motion information needed for motion-compensated imaging because of the potential for very high accuracy (Barnes *et al* 2008, Schmidt *et al* 2009, Forman *et al* 2011, Kyme *et al* 2012). Most optical motion tracking systems available involve the rigid attachment of markers to the patient's head, a procedure which can be time consuming and which requires training. A significant drawback of these methods is potential slippage of the attached markers due to non-rigid fixation. Although many attachment methods such as dental molds, adhesive bandages, modified neoprene caps and goggles have been used with some success (e.g. Bloomfield *et al* (2003), Ooi *et al* (2013) and Zaitsev *et al* (2006)), none have been adopted wholesale, and selecting a marker-based optical tracking method remains a basic trade-off between rigidity and convenience.

The challenges associated with marker-based optical tracking have motivated a recent push to cheaper and more convenient motion estimation methods, in particular markerless systems. In a markerless approach, features used for determining object pose usually constitute abstracted elements such as points, lines, contours and silhouettes from one or more camera images collected under either uniform or patterned visible or infra-red illumination (Szeliski 2011). Standard computer vision methods are then used for tasks such as matching features between images, reconstructing 3D points from features, registering point clouds and computing object pose.

Structured light techniques, in which patterned light is projected onto the object to assist feature detection and matching, have been used recently to track human head and neck motion in radiotherapy and PET (Santhanam *et al* 2011, Noonan *et al* 2012, Olesen *et al* 2012, 2013), and animal head motion in preclinical scans (Miranda *et al* 2017). The static accuracy of the first generation Microsoft Kinect (Santhanam *et al* 2011, Noonan *et al* 2012) appears to be on the order of a few millimeters for head translations and a few degrees for rotations. By comparison, the system in Olesen *et al* (2013) is considerably more accurate than the first generation Kinect, but suffers a heavy computational burden to register point clouds from successive frames. The second generation Kinect functions as a time-of-flight camera, collecting a complete depth map in every frame. This system was adapted for close-range tracking within a clinical PET scanner and shown to have improved accuracy and noise performance compared to the first generation Kinect (Noonan *et al* 2015).

Previously we developed and reported a markerless optical tracking system for motion-compensated PET imaging of rats (Kyme *et al* 2014). This method relies on the detection and matching of native object features using an ensemble of cameras, without the need to project patterned light. Conceptually, the approach is similar to the simultaneous localization and mapping (SLAM) problem in mobile robotics where robot-mounted sensors are used to obtain a consistent set of landmarks (a map) in the environment at the same time as the robot's motion is estimated (Se *et al* 2002). In this paper we describe the first stage in translating this method to clinical brain imaging. The two aims of our study were:

(i) To determine the feasibility of accurately estimating head pose using markerless facial tracking for a camera/subject geometry typical of diagnostic imaging in PET, SPECT and CT, and for an ethnically diverse subject population that would be encountered in practice;
(ii) To investigate strategies to optimize the accuracy of markerless tracking of humans, in particular the robust handling of non-rigidity of the face.

To address these aims we conducted a volunteer study in a mock imaging scenario consistent with typical PET, SPECT and CT geometries. This work represents a proof-of-principle that sparse native feature-based motion tracking in humans, without structured light, can be used to conveniently obtain the head motion information needed for accurate motion-compensated imaging performed using these ubiquitous imaging technologies.

2. Materials and methods

2.1. Markerless motion tracking system

The markerless motion tracking system used in this work is based on the small animal tracking method described in Kyme *et al* (2014). Here we provide a brief overview of the hardware and algorithm and its adaptation for clinical imaging.

2.1.1. System hardware

The markerless motion tracking system comprises four spatially calibrated CCD cameras (Flea2, Point Grey Research) arranged in two pairs directed at opposite sides of the face (see figure 2(ii)). The cameras (without lens attached) have dimensions 29 mm × 29 mm × 42 mm (w × h × l) and the lens adds a further 32 mm to

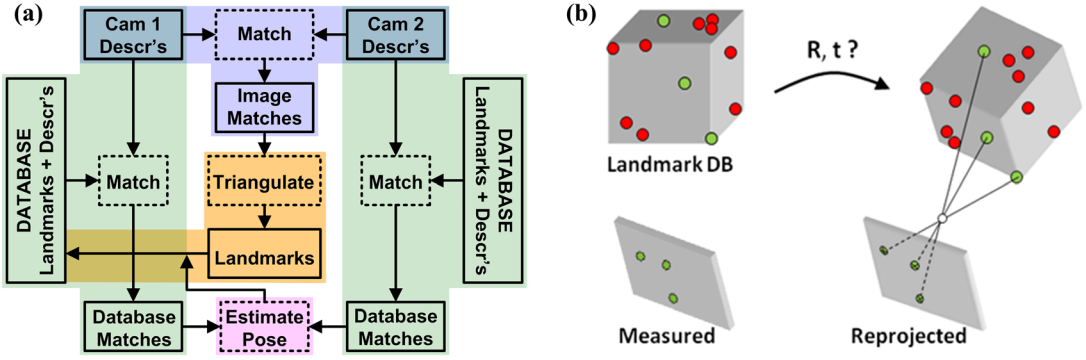

Figure 1. Principles of markerless motion estimation. (a) Overview: features (descriptors) in images from a stereo camera pair are matched (blue) to enable determination of 3D head landmarks using triangulation (orange). Image descriptors are also matched to descriptors stored in a database (green) providing the information needed for pose estimation (pink). New landmarks are then transformed to the initial pose and stored in the database with their descriptor (orange). Note that dashed boxes refer to processes and non-dashed boxes to data (inputs/outputs). Note also that we only show two cameras here for clarity, but in practice this can be scaled upwards arbitrarily. (b) Pose estimation: after identifying feature matches (green) between all possible features (red) in the landmark database and image, pose is estimated by finding the rotation, **R**, and translation, **t**, minimizing the reprojection error between measured and estimated feature locations.

the length. During data acquisition, frames consisting of four synchronized images (one from each camera) are continuously streamed to disk via a Firewire interface for offline processing. Synchronization accuracy is $<10\,\mu$s.

2.1.2. Camera calibration

The intrinsic parameters for each camera, including lens distortion coefficients, are computed using a series of images of a grid pattern (Bouguet 2018). Together with a set of synchronized images of a robotically controlled point light source, these parameters are input into a multi-camera self-calibration algorithm to compute the relative orientation (extrinsic camera calibration) of the camera network (Svoboda *et al* 2005). These calibrations only need to be performed once.

2.1.3. Pose estimation algorithm

For each acquired frame we detect highly distinctive native features in the four camera images (Lowe 2004). Features are defined by a 128-element 'descriptor' vector which enables reliable matching across images. Since the cameras are spatially calibrated, 3D landmarks can be estimated from these feature matches using triangulation. By amassing a database of landmarks with their associated descriptors, it is then possible to estimate the changing object pose. Namely, if a new camera frame shares feature descriptors with the database, we use a Gauss-Newton algorithm to solve for the rigid-body transformation (i.e. translation and orientation) that optimally aligns the database descriptors with the frame descriptors.

2.1.4. Outlier rejection

Spurious data arising in the processing pipeline, including features with poor discrimination, incorrect feature matches, and unrealistic pose changes, can limit the reliability of pose estimates. Therefore, several levels of outlier rejection are built into the algorithm. During feature matching between cameras, incorrect matches are detected and removed based on the epipolar geometrical constraint (Hartley and Zisserman 2004). During pose estimation, outliers are rejected in a two-stage process: in the early iterations of the Gauss-Newton algorithm, outliers are discriminated statistically (Iglewicz and Hoaglin 1993) based on the distribution of reprojection errors; in the late iterations, outliers are discriminated based on absolute reprojection error using a 2-pixel threshold.

The principles of the markerless tracking method are summarized in figure 1 and further details on the principles and implementation can be found in (Kyme *et al* 2014).

2.2. Experimental setup

The experimental setup for tracking the human face in a mock clinical brain imaging scenario is shown in figure 2. Each component is described below.

2.2.1. Subjects

Sixteen subjects (age range 20–48 year) were recruited for this study in accordance with an approved University of Sydney human ethics protocol. The volunteers reclined in a chair in front of an optical bench on which the 4-camera markerless tracking system was mounted (figure 2(i)). Unlike marker-based motion tracking methods,

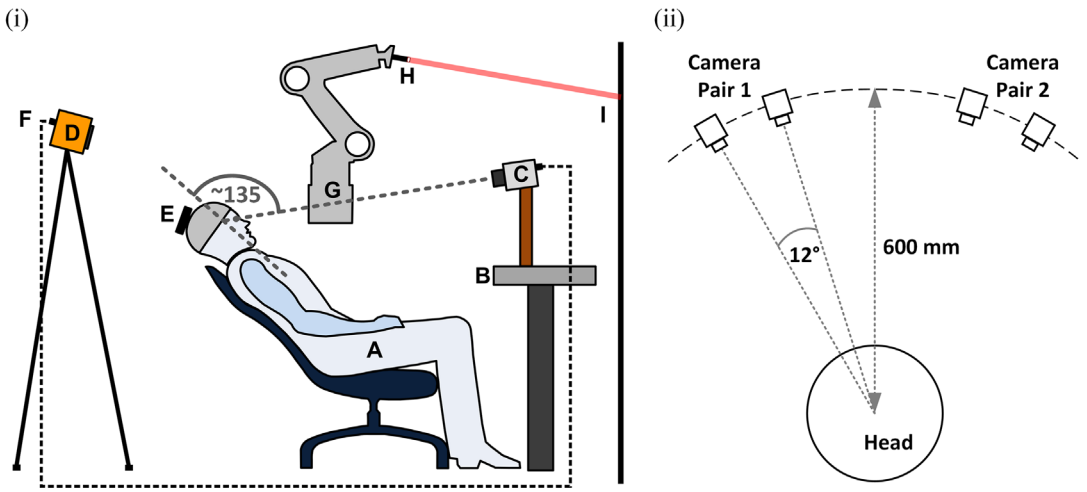

Figure 2. Experimental setup. (i) Volunteers (A) reclined in a seat in front of an optical bench (B) on which the markerless tracking system (C) was mounted. A marker-based tracking system (D) situated behind the volunteer was used to track a marker (E) affixed to the back of the head on each trigger from the markerless system (F). A robot (G) was used to project the beam of a laser pointer (H) onto the wall in front of the volunteer in a controlled motion sequence (I). (ii) Top view of the cameras showing their configuration with respect to the head.

native feature-based motion tracking is susceptible to differences in skin tone and texture. Therefore, to properly test the efficacy of feature-based tracking, we tested volunteers representing a broad range of ethnic background.

2.2.2. *Markerless motion tracking*

The four cameras were mounted to fixed brass posts on an optical bench using pan/tilt gimbals. The cameras were adjusted so that each pair subtended an angle of approximately 12 deg at a nominal distance from the face of 600 mm (figure 2(ii)). The line-of-sight of the cameras with respect to the long axis of the head was approximately 135 deg (figure 2(i)). This was to replicate the oblique viewing angle that is often necessary inside the bore of a PET or CT gantry so that cameras do not obstruct detectors in the system. During the experiment, camera frames were collected at 30 Hz with a fixed shutter time of 5–8 ms. One of the four synchronized cameras (figure 2(i)(C)) strobed a 5 V transistor-transistor logic (TTL) pulse at the start of each exposure to trigger the marker-based tracking system (see figure 2(i)(F) and section 2.2.3 below), thereby ensuring synchronization of the two systems.

2.2.3. *Marker-based motion tracking*

The MicronTracker marker-based motion tracking system (ClaronNav, Canada) (Kyme *et al* 2008) was mounted on a fixed tripod behind the volunteer and triggered via its general purpose input/output (GPIO) interface using the 30 Hz strobe signal from the markerless tracking system (figures 2(i)(D)–(F)). A large marker for high-accuracy pose tracking was printed on a rigid substrate and affixed to a neoprene swim cap (figure 3) ($n = 14$) or headband ($n = 2$). The headband was used for two female volunteers (S12, S16) who wore head scarves for religious reasons; in each case the headband was placed over the head scarf.

The two tracking systems were cross-calibrated using a paper-thin, double-sided marker sharing the same origin. Based on the location of the marker origin at 15 positions in the common field-of-view, we obtained the rigid-body transform relating the two coordinate systems using a closed-form, least-squares method (Horn 1987, Kyme *et al* 2008).

2.2.4. *Head motion sequence*

Each volunteer performed a specific sequence of head movements over a period of 2 min under steady laboratory lighting while 4000 frames were collected simultaneously from both motion tracking systems. To achieve consistent head motion across the volunteers, a 6-axis robot (Epson C3, SEIKO Corp., Japan) with a laser pointer mounted to the end-effector was used to execute pre-programmed motion during each study. The laser spot was projected onto a 3 m $\times$ 1.5 m wall in front of the volunteers, who were instructed to follow the spot using directed head movements rather than eye motion. Motion patterns of the laser spot are shown in figure 4 and included zig–zag, circular and square motion trajectories. The different motion patterns were designed to vary the speed and range of head motion and to excite all degrees-of-freedom. A video showing the setup for the laser spot motion sequence is included in the supplementary data (stacks.iop.org/PMB/00/0000/mmedia).

Figure 3. Neoprene cap with a printed 8-point MicronTracker marker attached for marker-based tracking. Each intersection of the black/white regions on this marker corresponds to a point that is tracked by the MicronTracker and used for pose estimation (Kyme *et al* 2008).

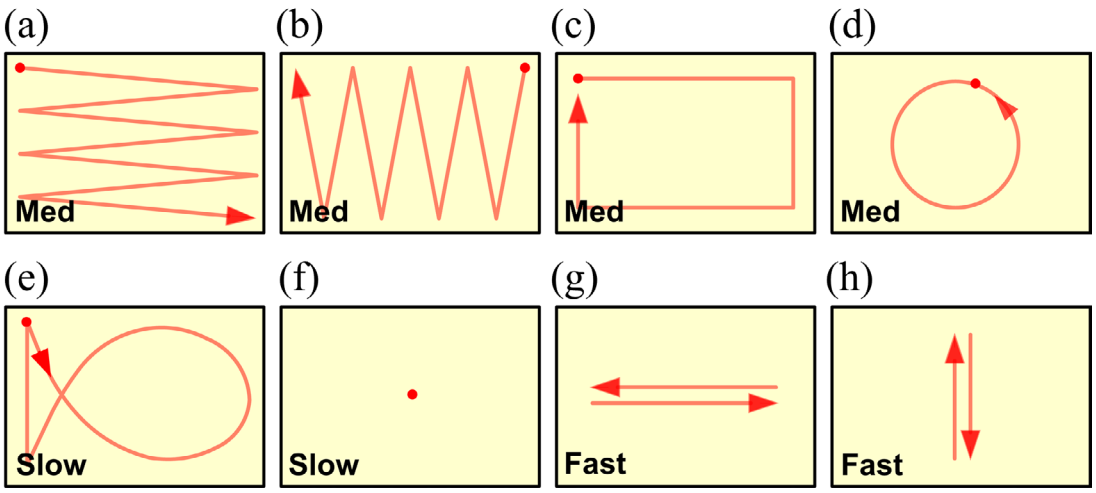

Figure 4. Sequence of motion patterns projected onto the wall in front of the subject. (a) Horizontal zig–zag, medium speed; (b) vertical zig–zag, medium speed; (c) large rectangle CW then CCW, medium speed; (d) large circle CCW then CW, medium speed; (e) large fish-shaped pattern, slow; (f) slow left and right head tilt followed by forward lean and hold with eyes focused on center dot; (g) fast left–right head-shaking; and (h) fast up–down head nodding.

2.3. Algorithm optimizations

We investigated several methods to optimize the accuracy of the markerless tracking algorithm for human subjects. These methods addressed background masking, feature pruning, handling non-rigid motion, and frame capping.

2.3.1. Background masking

The 'background' is defined as any part of the image other than the face itself. This includes parts of the neck, clothing and hair, and objects in the environment such as the chair and walls. In general, background points will be partially or completely uncorrelated with head motion, thereby contributing erroneous information to pose estimation if included. We investigated two background masking approaches to exclude such features from pose computation and studied the impact of this exclusion on motion tracking performance.

2.3.1.1.Strip masking

A rudimentary background mask was formed by rejecting fixed margins around the edge of the images. The number of image rows and columns to reject at the top, bottom, left and right of the images was determined manually for each camera, resulting in four binary masks per study. Only features inside the mask were used during pose processing. An example of the strip mask is shown in figure 6.

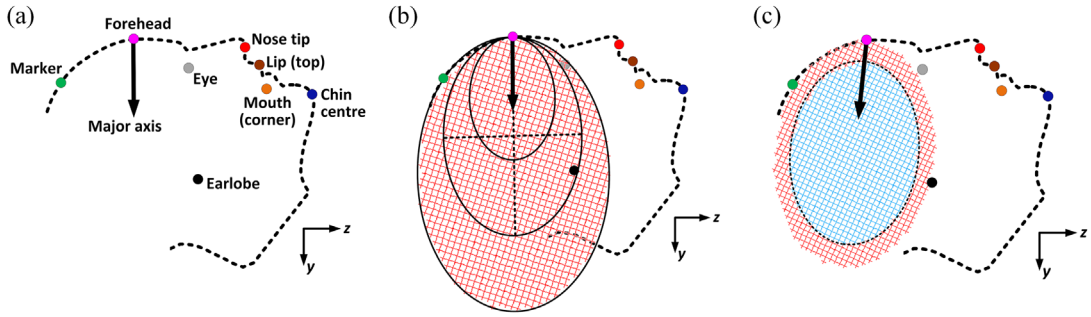

Figure 5. Generating the elliptical point cloud in the central sagittal plane of the brain. (a) Estimate major axis of the ellipse given known profile points; (b) grow the ellipse from the forehead along the major axis until it subsumes the location of the head marker (green point); (c) repeat steps (a) and (b) until a realistically oriented head ellipse is obtained, then finally fill a scaled brain ellipse with a regular grid of points (blue). The positive *x*-direction is out of the page.

2.3.1.2.Facial masking

A more restrictive mask was determined using 16 salient landmarks on the face: left/right eye, left/right mouth corner, top of the upper lip, nose tip, chin center, left/right earlobe, left/right temple area, left/right hairline, left/right eyebrow, and forehead center. The 3D location of these landmarks was reconstructed for the first frame after manually determining their image-space locations in each of the four camera images. For subsequent frames the facial mask was computed by (i) transforming the 16 landmark locations according to the latest pose estimate; (ii) reprojecting the transformed landmarks onto the four camera images; and (iii) computing the convex hull of the reprojected points for each camera. The region within the convex hull was used as the facial mask. An example of the facial mask is shown in figure 6.

2.3.2. Feature pruning

During the feature matching process, features with low specificity are more likely to find multiple matches, some of which must necessarily be false. Such features are inherently less reliable for inclusion in pose estimation. Therefore, during feature matching between the database and current frame, any database features matching >1 image feature were systematically pruned from the database. This process maintained a trimmed database containing features with higher specificity.

2.3.3. Handling non-rigidity

It is not only background features whose motion may be uncorrelated with head motion; this also applies to features on the face itself which move non-rigidly. Usually, once a landmark is added to the database, its location relative to other landmarks is fixed from that frame onwards. Therefore, if the landmark moves non-rigidly, subsequent matches to the respective database entry will introduce an error into the pose estimation. We investigated two methods to address this potential source of error.

2.3.3.1.Manual exclusion

For each volunteer study we manually identified camera frames affected by obvious non-rigid motion, including smiling, talking or facial expressions resulting in obvious facial movement. Pose estimation was performed after excluding these frames.

2.3.3.2.Updating landmarks

We hypothesized that updating the location of database landmarks each time they were re-observed could mitigate error in the pose estimation caused by non-rigid motion. This approach was tested against the manual exclusion approach.

2.3.4. Frame capping

In normal pose processing (figure 1), each new frame f contributes new landmarks to the database which can subsequently be used for pose estimation. However, it is possible that this regime results in more landmarks than are actually needed for effective pose estimation. We hypothesized that a saturation point can be reached at which nothing is gained by adding further landmarks. To test this hypothesis we implemented a frame cap, $f_c \leqslant 4000$, after which no more new landmarks were generated. Pose estimation for frames $f > f_c$ only used landmarks collected from frames $f \leqslant f_c$. All volunteer datasets were first processed without a frame cap and then with a frame cap of $f_c = 1000$ (i.e. ¼ of the total frames). Several studies (S03, S05, S07, S12) were also processed using 18 increasing frame caps: $f_c = 25, 50, 75, 100, 150, 200, 250, 300, 400, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500$ and 4000.

2.4. Validating motion tracking accuracy

The main goal of this work was to validate the accuracy of markerless motion tracking in a clinical imaging mock-up scenario by comparing the markerless motion estimates to marker-based motion estimates. Unfortunately, characterizing rigid-body motion tracking accuracy is not a straightforward task due to the non-linear nature of 3D rotations. Rotations belong to the special orthogonal group of order 3, or SO(3), which means that naïve arithmetic comparisons of Euler-angle or pitch-yaw-roll representations of 3D rotations does not represent a well-defined accuracy metric for the SO(3) topology (Stavdahl *et al* 2005, Sharf *et al* 2010). A better approach, which respects the non-linear topology of SO(3), is to consider how the estimated motion (obtained from a particular motion tracking system) transforms an arbitrary point in space. This enables us to characterise the agreement between two motion tracking systems in terms of a Euclidean distance—that is, the discrepancy in the location of a test point after transformation. A zero distance implies identical accuracy performance. Since this comparison represents a physical distance, it is far more intuitive than a comparison of the (inter-dependent) Euler-angle or pitch-yaw-roll components of a 3D rotation. The latter does not easily reveal the impact of a motion on specific locations in the region of interest.

To validate the accuracy of markerless motion tracking, we applied our distance metric to points in the brain, i.e. the primary region of interest in clinical brain imaging. For each volunteer we determined a point cloud of uniform density in the central sagittal slice of the brain and computed the motion tracking accuracy as the root-mean-square-error (RMSE) of the displacement (in mm) between the estimated (markerless) point locations and reference (marker-based) point locations, calculated over all points in the cloud:

$$\mathrm{RMSE}_i = \sqrt{\frac{\sum_j \mathbf{D}(\mathbf{P}_j, \mathbf{P}'_j)^2}{N}} \tag{1}$$

where i indexes pose number, j indexes points in the cloud, N is the total number of cloud points, $\mathbf{P}$ and $\mathbf{P}'$ are corresponding point clouds estimated using the marker-based and markerless systems, respectively, and $\mathbf{D}$ is a Euclidean distance operator providing the vector distance between two points in space. The RMSE was also computed over all poses to give a single metric of accuracy for each study.

To generate the point cloud within the central sagittal slice of the brain for each volunteer, we made use of the head landmarks estimated for the facial background masking approach (section 2.3.1). A five-step iterative process was used, illustrated in figure 5:

(i) Estimate the direction of the major axis of the brain ellipse based on the known landmark at the center of the forehead (figure 5(a)).
(ii) Grow the 2D ellipse from the forehead landmark, along the major axis direction, until it subsumes the known location of the head marker (figure 5(b)). The ratio of the minor and major axis lengths was fixed at 0.7.
(iii) Superimpose the estimated ellipse on the known head landmarks projected onto the sagittal (y-z) plane. Note that the sagittal plane was defined as the plane passing through the forehead landmark and perpendicular to the line joining the two orbital landmarks.
(iv) Iterate steps (i)–(iii) until a realistic head fit is obtained, i.e. when the ellipse is suitably bounded by the head marker, orbit and ear landmarks (figure 5(c)).

Once a realistic head ellipse was achieved for a given volunteer, a brain ellipse was obtained by scaling the head ellipse by 0.8 about its center (figure 5(c)). Finally, a point cloud was generated by filling the brain ellipse with a uniform point density of 4 cm^{-2}.

3. Results

3.1. Subjects

The 16 volunteers (9M, 7F) represented European, Asian, Middle Eastern and African American ethnicities (table 1). Factors such as facial hair or spectacles, which could potentially influence the tracking, are listed in column 5. Frames affected by non-rigid motion (e.g. smiling, talking and facial expressions resulting in significant facial movement) are listed in column 6.

3.2. Motion

Average motion characteristics for all studies are shown in table 2. The average range of rotation was $>40°$ about each axis and the average speed of rotation approximately 10 deg s^{-1}, up to a maximum of >200 deg s^{-1}. The average translational speed was 20–30 mm s^{-1}, up to a maximum of >0.5 m s^{-1}. Figure 8(a) shows examples of different head poses reached by the volunteers during the motion sequence.

Table 1. Subject-specific data.

Subject	Ethnicity	Country of origin	Sex	Comment	Excluded frames[a]
S01	European	Australia	M	Stubble	880–905, 1400–1530
S02	European	USA	M		2850–3080
S03	European	France	M		N/A
S04	European	Australia	M		N/A
S05	Asian	India	M		N/A
S06	European	Greece	M	Stubble	0–159, 3880–3950
S07	European	Germany	F		175–460, 2480–2520, 2700–2740
S08	Middle Eastern	Turkey	M		1080–1149, 2457–2660, 2990–3298
S09	European	Ireland	F	Spectacles	2695–2838, 3870–3999
S10	European	Poland	M		3610–3999
S11	European	USA	F		1800–1920, 2590–2853, 3525–3632
S12	Asian	Indonesia	F	Head scarf	230–375, 790–870, 1145–1257, 2950–3185, 3390–3448
S13	European	Australia	F		2640–3050
S14	European	Australia	M	Facial hair	2750–3176
S15	African American	USA	F		N/A
S16	Asian	Indonesia	F	Head scarf	2405–2793, 3794–3999

[a] Total number of frames was 4000.

Table 2. Summary of volunteer motion characteristics[a,b].

DoF	Range (deg, mm)	Speed (deg s^{-1}, mm s^{-1})
x-rot	44.2 (25.2–64.9)	10.2 (0–161)
y-rot	39.3 (17.1–66.6)	8.41 (0–196)
z-rot	63.1 (44.4–106)	14.6 (0–213)
x	167 (68.5–279)	23.2 (0–521)
y	175 (90.0–270)	34.3 (0–567)
z	81.0 (22.0–135)	15.3 (0–312)

[a] Values shown are mean (range).
[b] See figure 5 for coordinate system.

3.3. Feature detection, feature matching and landmark estimation

Figure 6 shows representative examples of the features detected in a single camera frame for two of the volunteers. In general, although features were detected across the face, they tended to be more concentrated in areas exhibiting greater contrast and texture (e.g. stubble). Figure 7 shows the average number of landmarks used for pose estimation (after outliers and duplicate features were removed) for each study. These data correspond to using strip background masking with no other optimizations (section 2.3) and indicate that obtaining 50–70 facial landmarks was typical for pose computation. An example of feature matching is shown at the bottom of figure 6. Feature matching was extremely reliable, with very few false matches recorded (only one false match is evident in the example in figure 6). Figure 6 also shows the effect of the facial and strip background masking methods in columns 3 and 4, respectively. Both methods were effective in removing outlier features on clothing.

3.4. Motion tracking accuracy

Figure 8 shows the motion tracking results for volunteer S04. Figure 8(a) shows several camera frames to illustrate the range of head motion for this subject, and figure 8(b) shows the markerless pose estimate for one of the rotational and translational DoF compared with the reference (marker-based) pose estimate. The pose agreement was clearly very close throughout the experiment.

Figure 9 shows an example of how the markerless tracking method maintained continuous tracking through large amplitude motion. In this study, the volunteer (S11) was not centered in the FoV leading to truncation of the head by the first camera pair (figure 9, top row). Despite this, robust tracking was maintained—evidenced by the green triangle lying nicely between the eyes and top of the mouth. This was because the second camera pair (figure 9, bottom row) provided the necessary feature data for pose estimation.

Table 3 shows the quantitative motion estimation accuracy results for all studies and optimization tests we performed. The global RMSE results are color-coded to aid interpretation: for the optimization tests within each background masking method (3A–6A for facial masking and 3B–6B for strip masking), green indicates that the

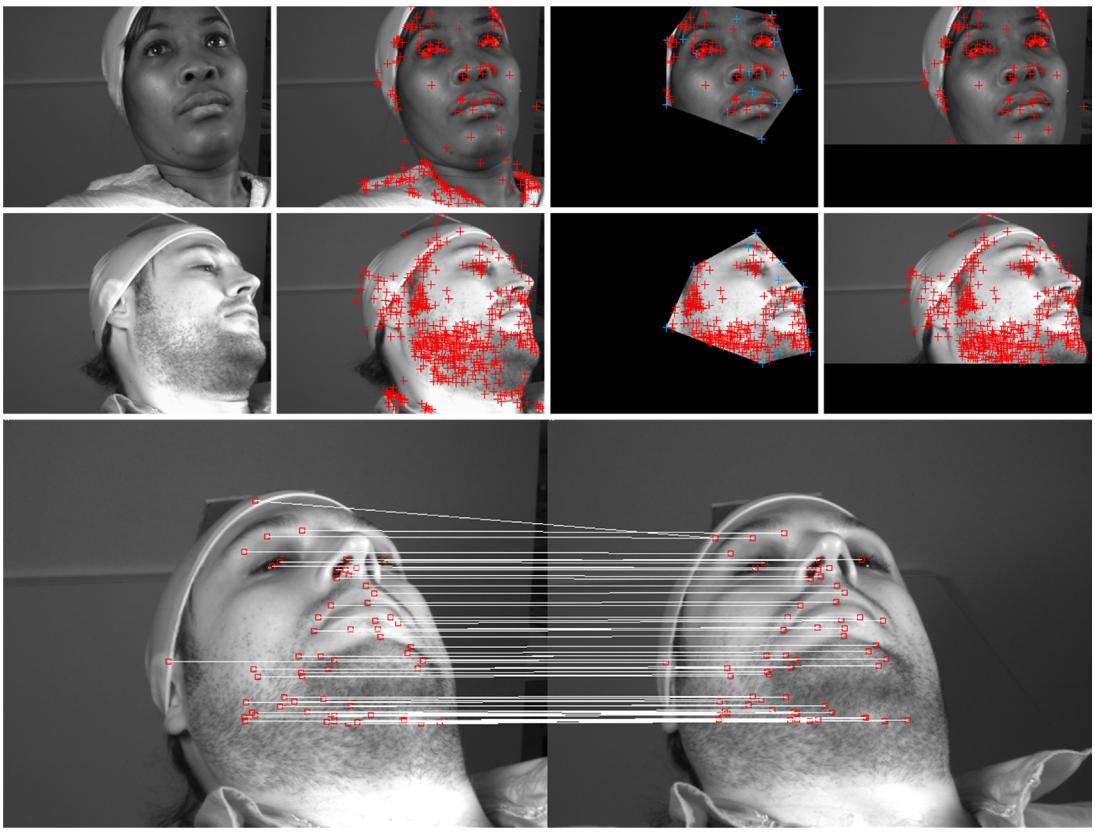

Figure 6. Feature detection, masking and matching. The top two rows show two examples (S16 and S01, respectively) of feature detection (red) with no background masking (column 2), facial masking (column 3; note that the blue crosses represent the landmarks used for computing the facial mask, section 2.3.1) and strip masking (column 4). The raw images are shown in column 1. The bottom panel shows an example of feature matching for study S01, after applying the facial mask. One false match can be seen on the forehead.

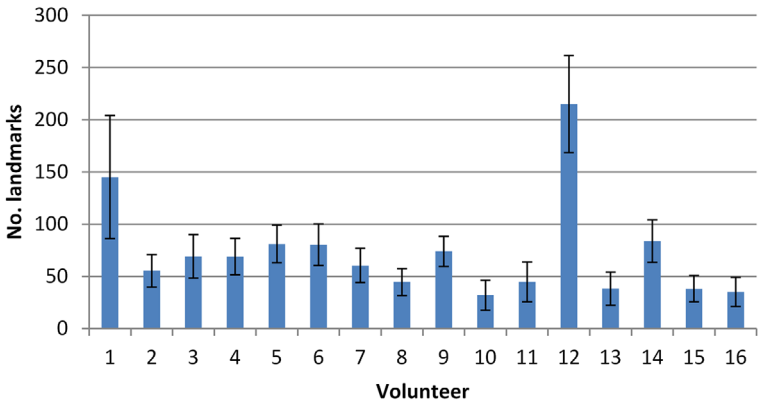

Figure 7. Number of landmarks used for markerless pose estimation. Data shown are the mean number of landmarks used to estimate the frame pose in each of the 16 volunteer studies. Error bars represent ± 1 standard deviation.

RMSE was equal to or better than the value without optimization (2A, 2B), orange indicates that the RMSE was worse by ⩽20%, and red indicates that the RMSE was worse by >20%. The RMSE data are discussed for each specific optimization in the sub-sections below.

3.4.1. Background masking

Omitting background masking (table 3, test 1) had little influence on the RMSE in 12/16 studies. For the remaining studies (S08, S13, S14, S15) the RMSE was 2–18 times worse than the best achievable RMSE. This resulted from an excess of uncorrelated background features on the neck, hair or collar (e.g. figure 6) which could not be properly excluded using the standard outlier rejection strategies (section 2.1.4). The problem was resolved

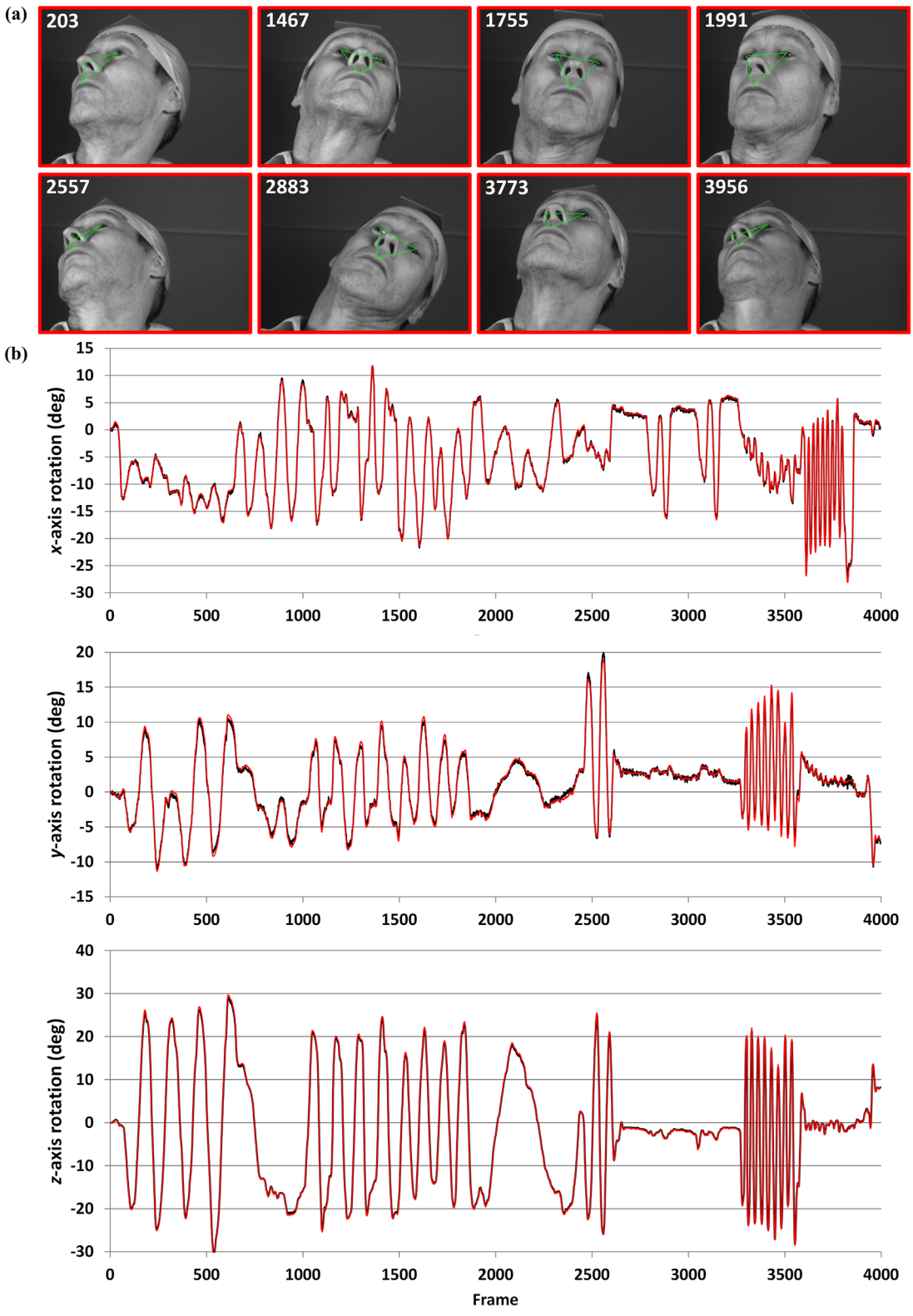

Figure 8. Motion tracking results for volunteer S04. (a) Example frames showing the range of head motion during the experiment. Accurate tracking is evidenced by alignment of the vertices of the green triangle with the eye centers and tip of the upper lip. (b) Comparison of the markerless (red) and marker-based (black) estimates of the three rotational DoFs over 4000 frames.

in all cases by including background masking. Moreover, comparing the table 3 results for tests 2A/2B, 3A/3B, 4A/4B, 5A/5B and 6A/6B indicates that simpler strip background masking consistently outperformed the more restrictive facial masking approach.

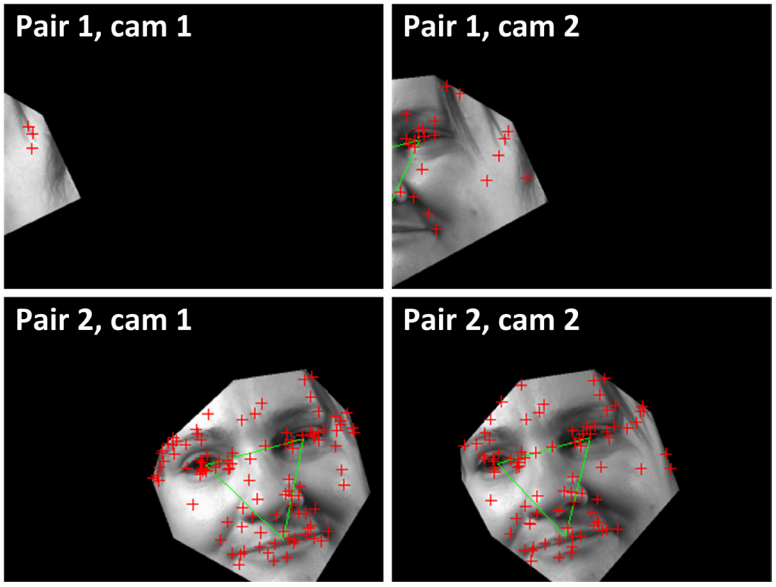

Figure 9. Continuous tracking through large amplitude motion for S11. Robust tracking is evidenced by the green triangle joining the center of the eyes and the tip of the upper lip.

3.4.2. Frame capping

Comparing table 3 tests 2A/3A and 2B/3B shows that the 1000-frame cap gave inconsistent results. For some studies (e.g. S01–S06) it had very little impact on the RMSE; for other studies (e.g. S08, S10, S11, S15) it caused a moderate (<20%) to large (>20%) worsening of the RMSE as evidenced by the prevalence of orange and red values in rows 3A and 3B.

3.4.3. Handling non-rigidity

Manually excluding frames that exhibited obvious non-rigid motion resulted in the same or better RMSE in nearly every case regardless of the background masking method (compare table 3 tests 2A/4A and 2B/4B). Using facial masking, updating the landmarks to account for non-rigidity of features (test 6A) performed better than manual exclusion of frames (test 4A) in about half the studies (S01–S06, S12, S14). For the remaining studies (notably S10), updating the landmark locations degraded RMSE. Using strip masking, updating the landmarks (test 6B) performed the same or marginally better than manual frame exclusion (test 4B) for all studies. Test 6B also resulted in the best agreement with the marker-based tracking across all studies (RMSE 1.72 $\pm$ 0.69 mm).

3.4.4. Auto pruning

With facial masking, auto pruning unreliable database features (test 5A) usually improved the RMSE, most notably for S08. With strip masking, auto pruning (test 5B) resulted in approximately the same or marginally lower RMSE for all studies compared to test 2B. Test 5B also resulted in the second best overall (RMSE 1.75 $\pm$ 0.69 mm).

3.5. Pose jitter

Pose jitter was characterized by the standard deviation of the residuals between the measured poses and the true (smoothly varying) signal (table 4). In all degrees-of-freedom except z-axis rotation (which is the least prone to jitter because it is the most robust parameter to estimate), marker-based measurements exhibited roughly twice the jitter of markerless measurements.

4. Discussion

In this study we have demonstrated the feasibility of accurately estimating rigid-body head motion of human volunteers using a multi-view optical tracking method without any attached markers and for a geometry consistent with brain imaging using the major wide-bore modalities PET, SPECT and CT. The method was tested in an ethnically diverse subject group and pose estimates were compared against a validated marker-based stereo-optical tracking system used in image-guided surgery. Image degradation due to patient motion remains a common problem in tomographic brain studies but has no solution that is in widespread use. We believe an important reason for this is the lack of an accurate and versatile marker-free motion tracking system to simplify integration into a busy clinical workflow. Unlike marker-based methods, the markerless tracking algorithm we

Table 3. RMSE (mm) for different processing regimes for all volunteers[a].

	Processing						Study														Statistics	
Test ID	Facial mask	Strip mask	1k frame cap	Manual exclude	Auto prune	Auto update	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	Mean	SD
1							1.7	1.2	1.2	1.0	3.9	2.0	1.9	4.0	1.4	2.3	1.7	1.9	37	27	12	2.0	6.38	10.55
2 A	X						**2.3**	**1.8**	**1.8**	**1.2**	**4.0**	**1.7**	**3.5**	**15**	**1.4**	**3.8**	**2.8**	**2.1**	**3.6**	**2.4**	**1.5**	**4.0**	3.31	3.27
3 A	X		X				2.6	1.8	1.9	1.2	4.3	1.7	4.8	26	1.8	17	7.0	2.4	3.9	2.6	3.6	4.4	5.44	6.61
4 A	X			X			2.3	1.8	1.8	1.2	4.0	1.5	3.3	3.3	1.3	3.5	2.6	1.9	3.2	2.4	1.5	3.9	2.48	0.95
5 A	X				X		1.9	2.0	2.1	1.2	4.0	1.7	2.7	3.5	1.4	3.4	3.3	2.9	3.2	2.1	1.8	5.6	2.67	1.13
6 A	X					X	1.9	1.5	1.3	1.2	4.0	1.6	4.3	4.1	1.8	130	3.5	1.8	4.1	2.3	1.8	6.1	2.75[b]	1.46[b]
2B		X					**1.6**	**1.3**	**1.3**	**1.1**	**3.9**	**1.3**	**1.7**	**1.7**	**1.4**	**2.6**	**1.6**	**2.0**	**2.2**	**2.3**	**1.3**	**2.0**	1.84	0.69
3B		X	X				1.8	1.3	1.3	1.1	3.9	1.3	2.5	10.4	1.7	11.2	2.5	2.1	2.3	2.3	3.8	1.8	3.21	3.08
4B		X		X			1.6	1.3	1.3	1.1	3.9	1.3	2.1	1.3	1.3	2.5	1.6	2.0	2.0	2.3	1.3	1.6	1.79	0.69
5B		X			X		1.5	1.3	1.4	1.0	3.9	1.3	1.6	1.7	1.3	2.3	1.7	1.7	2.3	2.1	1.3	1.5	1.75	0.69
6B		X				X	1.2	1.2	1.2	1.1	3.9	1.6	1.8	1.6	1.7	2.4	1.7	1.4	2.2	1.8	1.1	1.8	1.72	0.69

[a] RMSE color-coding: for each block of tests (2A–6A and 2B–6B), green denotes ⩽ reference RMSE (gray), orange denotes ⩽ 20% worse than the reference RMSE, red denotes >20% worse than the reference RMSE.

[b] S10 was treated as an outlier for test 6A and not included in the mean and standard deviation calculation.

Table 4. Pose jitter comparison[a].

DoF	Markerless	Marker-based
x-rot	0.10	0.20
y-rot	0.08	0.18
z-rot	0.10	0.07
x	0.92	1.93
y	1.28	2.18
z	0.37	0.65

[a] Values represent the standard deviation of measured poses with respect to the true signal, which is assumed to be smoothly varying like the laser spot trajectory. Units are deg for rotations and mm for translations.

present here estimates pose using only native features of the face. Improvement in accuracy and reduction in jitter are achieved by making use of many such features compared to the handful of features typical of marker-based systems.

Our work is most closely related to that of Noonan *et al* (2015) and Olesen *et al* (2012, 2013) who used a Kinect sensor and structured light system, respectively, to try to overcome the limitations of attached markers in motion-compensated PET imaging. Although the consumer-grade Kinect is cheap and accessible, it does not easily scale to multiple devices for improved accuracy due to interference of the systems. The structured light approach requires a calibrated light projector and for best accuracy seems to require areas with a strong change in topology (such as the nose bridge). The method we have presented is highly flexible in terms of the camera AQ1 type, size and resolution, lens choice and camera positioning. It is also highly scalable in terms of the number of cameras. The one-time calibration procedure remains the same irrespective of these choices. Since the accuracy of optical tracking systems scales with distance and/or sensor resolution, this flexibility is very useful for tailoring a setup for a particular scanner and study. The method allows us to exploit features all over the face for pose estimation, not just in limited areas like the bridge of the nose. Determining pose from a limited set of sparse features is inherently more efficient than registering dense point clouds of a surface, as is done in a structured light approach. This is the main reason why processing is generally slower and more intensive for the latter. Finally, our algorithm is also highly modular, allowing alternative feature detection, feature matching and outlier methods to be easily swapped in and out for comparison.

4.1. Accuracy

Our chief aim in this study was to evaluate the accuracy of our markerless tracking algorithm against a suitable reference system. The RMSE metric we used to characterize motion tracking accuracy has two important advantages over methods which compare estimated rotation and/or translation parameters directly (e.g. by wrongly assuming they are independent). Firstly, it is a well-defined and intuitive Euclidean distance metric for the non-linear SO(3) manifold. Secondly, it has intrinsic clinical relevance because it quantifies accuracy at or near the organ of interest rather than at points used to derive the motion (which are often far from the organ of interest).

Notwithstanding the soundness of the metric for quantifying motion tracking accuracy, there are several reasons why the accuracies we report here probably underestimate the performance of the markerless tracking algorithm in a clinical situation. Firstly, typical rotations and shifts of the head during tomographic brain imaging procedures are reported to be $\leqslant$5 deg and $\leqslant$7 mm (Ruttimann *et al* 1995, Beyer *et al* 2005, Dinelle *et al* 2006), respectively, although values for children, the elderly and patients with disorders impacting their control of movement could potentially be larger (Dinelle *et al* 2006). In our study, however, the average and maximum range and rate of head motion (table 2) far exceeded these normal reported values. This was a deliberate experimental design choice so we could test the performance of markerless tracking under extremely challenging conditions. In particular, it forced the algorithm to find viable features over a much larger portion of the face than would usually be required in practice. As a result, the accuracies we obtained are probably conservative, and the performance for typical clinical motion would likely improve. Secondly, as for all systems in which markers are attached to the head non-invasively, there was the potential for marker slippage in our marker-based reference motion tracking system. We expect the slight non-rigidity of the neoprene cap arrangement (and head band) to manifest as a baseline (dc) error when comparing the reference and markerless estimates. Thirdly, the cross-calibration between the reference and markerless systems similarly introduces a baseline level of error into the comparison of pose estimates. It was not possible to quantify the relative contribution of these errors from our data, however it is instructive that we have previously observed an accuracy of ~0.1 mm for the markerless algorithm in an animal phantom study (Kyme *et al* 2014). Overall, the mean accuracy of 1.7 mm we report here for markerless tracking easily satisfies the motion tracking requirements for PET and SPECT. And for individual studies, the conservative accuracy estimate was as low as 1 mm. Further work is needed to determine if the accuracy of markerless tracking is also acceptable for CT with its higher spatial resolution.

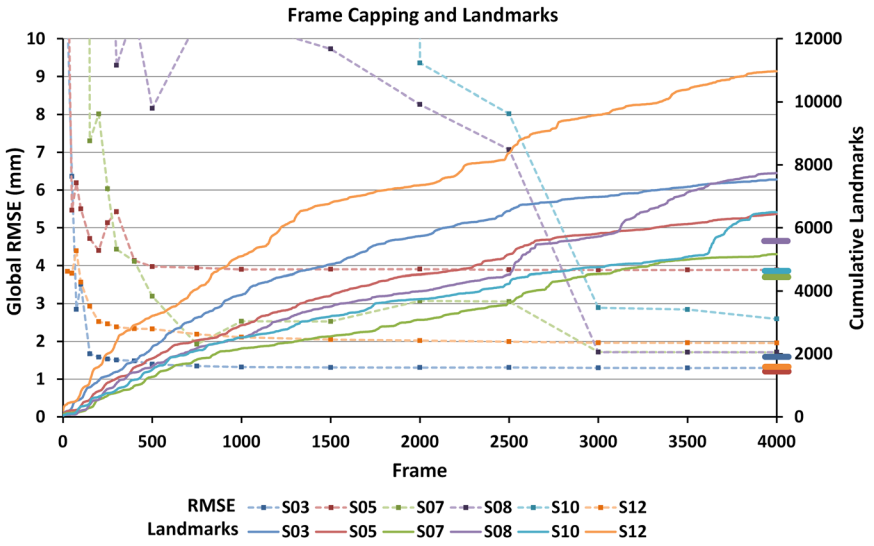

Figure 10. RMSE and number of landmarks as a function of the frame cap, shown for six studies. The left and right vertical axes show global RMSE (broken lines) and cumulative landmarks (unbroken lines), respectively. On the right vertical axis, the short colored bars indicate the number of landmarks for each study at the lowest frame cap corresponding to a stable RMSE.

4.2. Algorithm optimization

We investigated a variety of strategies to optimize the accuracy performance of markerless tracking of human faces in the brain imaging mock-up geometry. Our results suggest that background masking is an important pre-processing step for markerless tracking but that a highly subject-specific mask is unnecessary—and in some cases is counter-productive. A simple strip mask sufficed to remove some of the background features, improving the relative proportion of facial features and thereby allowing our outlier rejection methods to function successfully. In contrast, pose estimation occasionally failed if background masking was omitted or if a more restrictive background mask was used due to unreliable features becoming dominant. Our outlier rejection methods in this case were insufficient to recover a robust pose (e.g. S07, S16).

Manually excluding frames corrupted by non-rigid motion generally improved the RMSE as expected. However, it is somewhat surprising that the RMSE improved by so little in most cases (study S08 was the exception). This suggests an important role played by the normal outlier rejection methods we have built into the algorithm (section 2.1.4) to nullify the impact of non-rigid features in the pose estimation. Information contributed by non-rigid feature locations seems to be easily rejected during our routine checks of reprojection error by virtue of it being inherently inconsistent with a rigid-body pose solution. Further evidence of the effectiveness of our outlier rejection methods was the limited benefit we observed for database pruning—spurious feature matches were effectively discarded without requiring any dedicated pruning. When pruning was combined with the 1000-frame cap (data not shown), the RMSE was much more likely to degrade, suggesting that pruning the database can have the unwanted effect of removing reliable features. Therefore, a better approach may be to prune a feature only if it exhibits consistent unreliability, evidenced by multiple matches over multiple frames (Se *et al* 2002).

Imposing a 1000-frame cap for populating the database with landmarks showed inconsistent results across the 16 studies. We explored this further by studying the RMSE as a function of frame cap for six of the studies. Figure 10 shows the results: the RMSE plateaued beyond a certain frame in all cases. The plateau occurs when the landmarks in the database support estimation of the full range of head motion sampled by the cameras during the study. For some volunteers this point was reached as early as frame 150/4000 and there was little advantage in continuing to add more landmarks to the database from subsequent frames. Conversely, in about half the studies there were insufficient landmarks in the database by frame 1000 to estimate the full range of motion during the acquisition, thus leading to an increase in the RMSE when a 1000-frame cap was imposed. In some volunteers (e.g. S08 and S10) the plateau did not occur until much later (frame 3000–4000); only then had a sufficient number of landmarks accumulated to support accurate pose estimation for all frames. The frame-capping behavior has an important implication for motion tracking: the ability to cap at an earlier frame means fewer features in the database (figure 10, right axis) which in turn means faster feature matching and improved efficiency of motion estimation. In practice, one can imagine performing a brief 'priming' scan in which the patient is instructed to move their head through a range of motion which is larger than expected. During this scan a small but sufficient database of features would be amassed to allow efficient and accurate pose estimation during the ensuing imaging study. Clearly, such a priming scan would only be possible for compliant patients.

In summary, our investigation of optimization strategies suggests that consistently accurate results for markerless tracking are achieved by applying the base algorithm in conjunction with a simple background masking procedure. Other strategies such as updating landmarks each time they are re-observed can provide marginal gains in accuracy by better handling non-rigidity of the face. Capping the database is certainly advantageous to speed up pose estimation.

4.3. Other performance aspects

The reduced jitter of pose estimates derived from the markerless approach is consistent with the smooth robot-controlled motion profile of the laser spot being tracked by the volunteers. Lower variance compared to marker-based estimates likely results from the fact that there are considerably more landmarks (~50–70, figure 7) available for computing pose compared to the 8 fixed landmarks on the marker (figure 3) (Gao *et al* 2007).

We made no attempt in this work to optimize the speed of pose processing. Nevertheless, it was possible to estimate the pose at several frames per second provided the database was not overly populated. For a long clinical scan (up to 1 h) with frames acquired at 20–30 Hz, some optimization would be needed so that processing is completed in an acceptable time. Limiting the size of the database (as discussed above) is one option; GPU acceleration of the feature detection and matching processing is also feasible. We are currently also exploring the feasibility of improving the speed and accuracy of pose estimation using alternative feature detection and matching approaches.

All of our experiments were performed under stable laboratory lighting. Achieving stable lighting in a clinical setting is usually feasible, however further work is required to test the stability of our method under changing lighting conditions and the presence of shadows.

4.4. Clinical outlook

It was beyond the scope of this work to implement the markerless motion tracking method in a clinical imaging system (PET, SPECT or CT). However, the promising results we have reported suggest this is the obvious next step. Future work will involve implementing and validating the method in clinical PET, SPECT and CT systems, including a quantitative comparison of event-by-event based motion correction using markerless and marker-based motion estimates.

The relative positioning of the head and tracking system in our experimental design was chosen to reflect the geometry needed for a practical implementation in a clinical PET, SPECT or CT system. However, as mentioned above, changing the number and/or positioning of cameras to meet the specific requirements of a particular scanner is eminently feasible. The geometric setup we describe in this work does not directly address magnetic resonance imaging (MRI), which presents some unique challenges for motion tracking (Forman *et al* 2011). To achieve the highest possible tracking accuracy in MRI requires the use of MRI-compatible in-bore cameras (Maclaren *et al* 2012). This produces an extremely close range tracking situation which differs from the longer range we simulated here for PET, SPECT and CT. Secondly, the presence of a head coil in MRI to detect small signal changes in the brain dramatically limits the area of the face visible to a camera. Thus pose must be determined from features confined to small regions of the face within the line-of-sight of the cameras. We are currently investigating ways of adapting our markerless approach to these particular geometrical constraints in MRI (Kyme *et al* 2016).

It is acknowledged that considerable challenges may restrict head tracking in a clinical setting. For example, patients may have gross facial changes due to trauma, surgery, bandages or attached medical devices that limit the application of our approach. Patients may also exhibit gross facial motion (e.g. due to fear, pain, anxiety) that may be outside the range of expressions that we tested here.

5. Conclusion

In conclusion, we have adapted a motion tracking method originally developed for animals for use in motion-compensated human brain imaging. Compared to traditional marker-based motion tracking approaches, the method relies only on the detection of native facial features, with no requirement for attached markers. Modeling the potentially oblique motion tracking geometry encountered in PET, SPECT and CT imaging, we obtained pose agreement within 2 mm of a marker-based system used as a gold standard. We also demonstrated robust performance of pose estimation in the presence of non-rigid motion. Our results indicate the potential to incorporate this versatile motion tracking technique into motion-compensated imaging protocols. The versatility of performing motion tracking without attached markers could help to overcome the barriers that have prevented motion-compensated imaging from being widely adopted in research and clinical studies, despite much evidence indicating its effectiveness. Our future efforts will focus on implementing the motion tracking method on clinical systems.

Acknowledgment

The authors gratefully acknowledge Johan Nuyts for input on the pose comparison method. All experimental work was performed at the Sydney-ANSTO node of the Australian National Imaging Facility (www.anif.org.au) which is supported by the Commonwealth Government of Australia, the New South Wales Government, the University of Sydney and the Australian Nuclear Science and Technology Organization (ANSTO). During the analysis and write-up stage of this work Andre Kyme was supported by a Cassen Postdoctoral Fellowship awarded by the Education and Research Foundation of the US Society of Nuclear Medicine and Molecular Imaging.

ORCID iDs

Andre Z Kyme https://orcid.org/0000-0003-3297-5390
Steven R Meikle https://orcid.org/0000-0001-7397-0364
Roger R Fulton https://orcid.org/0000-0003-2536-2190

References

Barnes P J, Baldock C, Meikle S R and Fulton R R 2008 Benchmarking of a motion sensing system for medical imaging and radiotherapy *Phys. Med. Biol.* **53** 5845–57

Beyer T, Tellmann L, Nickel I and Pietrzyk U 2005 On the use of positioning aids to reduce misregistration in the head and neck in whole-body PET/CT studies *J. Nucl. Med.* **46** 596–602

Bloomfield P M, Spinks T J, Reed J, Schnorr L, Westrip A M, Livieratos L, Fulton R and Jones T 2003 The design and implementation of a motion correction scheme for neurological PET *Phys. Med. Biol.* **48** 959–78

Bouguet J-Y 2018 *Matlab Calibration Toolbox* (Online) (www.vision.caltech.edu/bouguetj/calib_doc/)

Dinelle K, Blinder S, Cheng J C, Lidstone S, Buckley K, Ruth T J and Sossi V 2006 Investigation of subject motion encountered during a typical positron emission tomography scan *Proc. 2006 IEEE Nuclear Science Symp. Conf. Record* vol 1–6 pp 3283–7

Forman C, Aksoy M, Hornegger J and Bammer R 2011 Self-encoded marker for optical prospective head motion correction in MRI *Med. Image Anal.* **15** 708–19

Fulton R R, Meikle S R, Eberl S, Pfeiffer J and Constable C J 2002 Correction for head movements in positron emission tomography using an optical motion-tracking system *IEEE Trans. Nucl. Sci.* **49** 116–23

Gao X W, Anishenko S, Shaposhnikov D, Podlachikova L, Batty S and Clark J 2007 High-precision detection of facial landmarks to estimate head motions based on vision models *J. Comput. Sci.* **3** 528–32

Hartley R and Zisserman A 2004 *Multiple View Geometry in Computer Vision* 2nd edn (Cambridge: Cambridge University Press)

Horn B K P 1987 Closed-form solution of absolute orientation using unit quaternions *J. Opt. Soc. Am.* A **4** 629–42

Iglewicz B and Hoaglin D 1993 *How to Detect and Handle Outliers* (Milwaukee, WI: ASQC Quality Press)

Jin X, Mulnix T, Gallezot J D and Carson R E 2013 Evaluation of motion correction methods in human brain PET imaging—a simulation study based on human motion data *Med. Phys.* **40** 102503

Kim J H, Nuyts J, Kyme A Z, Kuncic Z and Fulton R R 2015 A rigid motion correction method for helical computed tomography (CT) *Phys. Med. Biol.* **60** 2047–73

Kyme A Z, Hutton B F, Hatton R L, Skerrett D W and Barnden L R 2003 Practical aspects of a data-driven motion correction approach for brain SPECT *IEEE Trans. Med. Imaging* **22** 722–9

Kyme A Z, Maclaren J, Aksoy M and Bammer R 2016 Feasibility of marker-free motion tracking for motion-corrected MRI and PET-MRI *Proc. 2016 IEEE Nuclear Science Symp. Conf. Record* vol 1–3 (https://doi.org/10.1109/NSSMIC.2016.8069629)

Kyme A Z, Se S, Meikle S R, Angelis G, Ryder W, Popovic K, Yatigammana D and Fulton R R 2014 Markerless motion tracking of awake animals in positron emission tomography *IEEE Trans. Med. Imaging* **33** 2180–90

Kyme A Z, Zhou V, Meikle S R and Fulton R R 2008 Real-time 3D motion tracking for small animal brain PET *Phys. Med. Biol.* **53** 2651–66

Kyme A, Meikle S, Baldock C and Fulton R 2012 Refraction-compensated motion tracking of unrestrained small animals in positron emission tomography *Med. Image Anal.* **16** 1317–28

Lowe D 2004 Distinctive image features from scale-invariant keypoints *Int. J. Comput. Vis.* **60** 91–110

MacLaren J *et al* 2012 Measurement and correction of microscopic head motion during magnetic resonance imaging of the brain *PLoS One* **7** 1–9

Miranda A, Staelens S, Stroobants S and Verhaeghe J 2017 Markerless rat head motion tracking using structued light for brain PET imaging of unrestrained awake small animals *Phys. Med. Biol.* **62** 1744–58

Montgomery A J, Thielemans K, Mehta M A, Turkheimer F, Mustafovic S and Grasby P M 2006 Correction of head movement on PET studies: comparison of methods *J. Nucl. Med.* **47** 1936–44

Noonan P J, Howard J, Cootes T, Hallett W A and Hinz R 2012 Realtime markerless rigid body head motion tracking using the Microsoft Kinect *Proc. 2012 IEEE Nuclear Science Symp. Conf. Record* pp 2241–6

Noonan P J, Howard J, Hallett W A and Gunn R N 2015 Repurposing the microsoft kinect for windows v2 for external head motion tracking for brain PET *Phys. Med. Biol.* **60** 8753–66

Olesen O V, Paulsen R R, Hojgaard L, Roed B and Larsen R 2012 Motion tracking for medical imaging: a nonvisible structured light tracking approach *IEEE Trans. Med. Imaging* **31** 79–87

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

Ooi M, Aksoy M, Maclaren J, Watkins R and Bammer R 2013 Prospective motion correction using inductively coupled wireless RF coils *Mag. Res. Med.* **70** 639–47

Ruttimann U E, Andreason P J and Rio D 1995 Head motion during positron emission tomography: is it significant? *Psychiatry Res. Neuroim.* **61** 43–51

Santhanam A, Low D and Kupelian P 2011 TH-C-BRC-11: 3D tracking of interfraction and intrafraction head and neck anatomy during radiotherapy using multiple Kinect sensors *Med. Phys.* **38** 3858

Schmidt J, Berg D and Ploeg H-L 2009 Precision, repeatability and accuracy of Optotrak optical motion tracking systems *Int. J. Exp. Comput. Biomech.* **1** 114–27

Se S, Lowe D and Little J 2002 Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks *Int. J. Rob. Res.* **21** 735–58

Sharf I, Wolf A and Rubin M 2010 Arithmetic and geometric solutions for average rigid-body rotation *Mech. Mach. Theory* **45** 1239–51

Stavdahl O, Bondhus A, Pettersen K and Malvig K 2005 Optimal statistical operators for 3-dimensional rotational data: geometric interpretations and application to prosthesis kinematics *Robotica* **23** 283–92

Svoboda T, Martinec D and Pajdla T 2005 A convenient multi-camera self-calibration for virtual environments *PRESENCE: Teleoperators Virt. Environ.* **14** 407–22

Szeliski R 2011 *Computer Vision: Algorithms and Applications* 1st edn (Berlin: Springer)

Yu H Y and Wang G 2007 Data consistency based rigid motion artifact reduction in fan-beam CT *IEEE Trans. Med. Imaging* **26** 249–60

Zaitsev M, Dold C, Sakas G, Hennig J and Speck O 2006 Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system *Neuroimage* **31** 1038–50