# Markerless Motion Tracking of Awake Animals in Positron Emission Tomography

André Kyme*, *Member, IEEE*, Stephen Se, *Senior Member, IEEE*, Steven Meikle, *Senior Member, IEEE*, Georgios Angelis, *Member, IEEE*, Will Ryder, *Member, IEEE*, Kata Popovic, Dylan Yatigammana, and Roger Fulton, *Senior Member, IEEE*

***Abstract*—Noninvasive functional imaging of awake, unrestrained small animals using motion-compensation removes the need for anesthetics and enables an animal's behavioral response to stimuli or administered drugs to be studied concurrently with imaging. While the feasibility of motion-compensated radiotracer imaging of awake rodents using marker-based optical motion tracking has been shown, markerless motion tracking would avoid the risk of marker detachment, streamline the experimental workflow, and potentially provide more accurate pose estimates over a greater range of motion. We have developed a stereoscopic tracking system which relies on native features on the head to estimate motion. Features are detected and matched across multiple camera views to accumulate a database of head landmarks and pose is estimated based on 3D-2D registration of the landmarks to features in each image. Pose estimates of a taxidermal rat head phantom undergoing realistic rat head motion via robot control had a root mean square error of 0.15 and 1.8 mm using markerless and marker-based motion tracking, respectively. Markerless motion tracking also led to an appreciable reduction in motion artifacts in motion-compensated positron emission tomography imaging of a live, unanesthetized rat. The results suggest that further improvements in live subjects are likely if nonrigid features are discriminated robustly and excluded from the pose estimation process.**

***Index Terms*—Markerless optical motion tracking, motion compensation, positron emission tomography (PET).**

## I. Introduction

POSITRON emission tomography (PET) imaging of small animals is a vital component of basic neuroscientific research, enabling longitudinal studies of the molecular changes in the brain resulting from controlled interventions [1], [2]. The vast majority of such studies are performed on anesthetized animals to avoid any motion that might cause artifacts in the reconstructed images. However, not only do different anesthetics have varying impact on animal physiology and the resulting PET measurements [3], [4], this approach also precludes imaging an animal and concurrently studying its behavioral responses to external stimuli or administered drugs.

The ability to image animals while awake and unrestrained would greatly enhance the potential of PET in neurological investigations. For example, it would allow: 1) study of the functional response to sensory stimulation that requires the animal's conscious attention, especially when using radiotracers that do not get trapped like FDG, and therefore require that imaging commences at the time of injection to enable the application of kinetic models; 2) study of the concurrent behavioral and functional response following a cognitive, pharmacologic or environmental challenge; and 3) study of dynamic functional processes, especially those related to receptor-ligand binding. A hybrid approach providing some but not all of this capability involves delivering a challenge or stimulus to an awake animal during the tracer uptake phase, then anesthetizing and imaging the animal [5]. In this approach the impact of anesthesia on the PET measurements is avoided and net changes caused by an intervention can be detected. However, the approach is not applicable to all tracers and does not allow the study of dynamic (transient) changes or correlations. The full range of possibilities requires that animals can be imaged while awake and unrestrained.

Two recent technologies enable an animal to be imaged while awake and unrestrained. The first, applicable only in rats, is a miniature head-mounted PET tomograph, termed "RatCAP," which eliminates motion by moving rigidly with the head [6]. Using this device, Schulz *et al.* quantified levels of the neurotransmitter dopamine in the brains of active and anesthetized rats [7]. Dopamine plays an important role in regulating reward-driven learning and behavior, movement and memory, and has also been implicated in Parkinson's disease. The study showed increased dopamine levels in anesthetized rats compared with awake rats and an inverse correlation between dopamine levels and physical activity. More fundamentally, it demonstrated the feasibility of using PET to study correlations between brain function and behavior in awake animals. Notwithstanding the major technological advance RatCAP represents, the counterbalance mechanism used to support the scanner weight could potentially restrict natural movement and behavior of the rat. It

Manuscript received May 08, 2014; accepted June 14, 2014. Date of publication June 26, 2014; date of current version October 28, 2014. This work was supported by the Australian Research Council under Discovery Project Grant DP0663519 and Grant DP0988166. *Asterisk indicates corresponding author.*

*A. Kyme is with the Brain and Mind Research Institute and Faculty of Health Sciences, University of Sydney, Sydney, NSW, 2050 Australia, and also with the Department of Biomedical Engineering, University of California, Davis, CA 95616 USA (e-mail: andre.kyme@sydney.edu.au).

S. Se is with MDA Systems Ltd., Richmond, BC, V6V 2J3 Canada (e-mail: sse@mdacorporation.com).

S. Meikle, G. Angelis, W. Ryder, and K. Popovic are with the Brain and Mind Research Institute and Faculty of Health Sciences, University of Sydney, Sydney, NSW, 2050 Australia.

D. Yatigammana is with Portsmouth Hospital's National Health Services Trust, PO6 3LY Portsmouth, U.K.

R. Fulton is with the Brain and Mind Research Institute, Faculty of Health Sciences, and School of Physics, University of Sydney, Sydney, NSW, 2050 Australia, and also with the Department of Medical Physics, Westmead Hospital, Sydney, NSW 2145, Australia (e-mail: roger.fulton@sydney.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2014.2332821

also seems unlikely the technology could be scaled down further for imaging the mouse brain while maintaining sufficiently high photon detection sensitivity [8].

Another means of imaging awake rodents (rats or mice), which can utilize conventional scanner technology, involves motion compensation. Here, rigid motion of the head is recorded during the study and subsequently accounted for before or during image reconstruction [9], [10]. The feasibility of performing radiotracer imaging on fully conscious rodents in conjunction with stereo-optical head motion tracking and motion compensation has been demonstrated both for animals confined to a tube [11], [12] and for animals within an open chamber [13]. In each case, motion tracking was marker-based—that is, it involved specially designed markers attached to the head. In our experience, use of attached markers increases the risk of failed experiments since the marker and head motion can become decoupled, lengthens experimental protocols since animals must be trained to tolerate attached markers, and restricts the detectable range of motion due to limitations on the marker size and placement. By contrast, markerless motion tracking, which does not require anything to be attached to the head, avoids the risk of marker detachment and streamlines the experimental workflow since animals do not need to be acclimatized. Moreover, it can potentially provide more accurate pose estimates over a greater range of motion by exploiting many features spread across the head.

Markerless tracking of object pose has been studied extensively in a wide variety of fields, including human motion capture, robot navigation, object recognition, virtual and augmented reality, and the movement of insects in flight [14]–[17]. In a markerless approach, features usually constitute abstracted elements such as points, lines, contours and silhouettes and are determined directly from images collected under either uniform or patterned visible or infra-red (IR) illumination [18]. Sophisticated point feature detectors exhibiting invariance to a range of factors such as translation, rotation and scale have found extensive use in stereo-vision setups where it is necessary to match features robustly between disparate views [19]. The method we describe uses such features.

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 [20]–[22]. The static accuracy of the Microsoft Kinect [20], [21] over the range of useful working distances appears to be on the order of a few millimeters for head translations and a few degrees for rotations. The system in [22] is considerably more accurate than the Kinect, however the light pattern and decoding scheme suit only relatively small head movements, not the large movements often performed by rodents [23]. More fundamentally, however, structured light patterns can easily be distorted by the high local contrast of fur, making robust surface reconstructions for animals difficult [24].

In this paper, we describe a multi-camera markerless tracking system for awake rats which operates under uniform visible lighting. Conceptually, our 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 (i.e., a map) in the environment at the same time as the robot's motion is estimated [17]. Rat tracking is an analogous problem if one considers the sensor (tracking) frame fixed and the rat as the moving environment. In this case we seek to identify a consistent set of landmarks on the surface of the head that can be used for pose estimation. The SLAM framework also provides a means of incorporating information from the re-observation of landmarks in order to reduce errors and improve pose estimates over time—though we have not considered this aspect here.

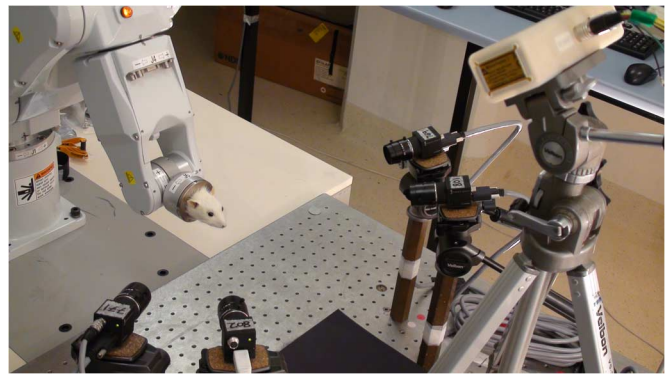


Fig. 1. Photograph of the laboratory setup showing the four cameras arranged in two pairs and mounted on brass posts attached to an optical table. On the left is the taxidermal rat phantom attached to the robot end-effector. In the top right is the MicronTracker system used for marker-based tracking. Matt black cardboard (not shown) was mounted behind the rat's head during the experiment to minimize SIFT features detected in the background.

The paper is structured as follows. Section II is the system overview where we outline the various stages of the markerless tracking method, from camera calibration and feature detection to pose estimation. In Section III, we describe the validation experiments using a realistic rat phantom and a live animal. In Sections IV and V, we report and discuss the results.

## II. System Overview

### A. Motion Tracking System

The tracking system uses $640 \times 480$ resolution monochrome Flea2 CCD cameras (Point Grey Research, Richmond, BC, Canada) fitted with 12 mm lenses (GMN21214, Goyo Optical, Japan). The number of cameras, $N$, should be $\geq 2$; for this work we used $N = 4$. Without lenses the Flea2 has dimensions 29 mm $\times$ 29 mm $\times$ 30 mm. The cameras link to a PC via the IEEE 1394b (800 Mb/s) interface. Maximum frame rate of the Flea2 is specified as 80 frames per second (fps) but in practice this is limited by the IEEE 1394b bandwidth. Using a pixel depth of 8 bits, four cameras on the same IEEE 1394 bus can stream at slightly under 60 fps, which is more than sufficient for tracking rat head motion [11]. Therefore, a single bus was used for all our experiments. Cameras on the same bus automatically synchronized frame capture to within 10 $\mu$s.

The cameras were arranged in front of the animal in two pairs, with each pair viewing a different side of the animal's head (Fig. 1). The pair-wise arrangement served to increase the spatial distribution of detected features. In the phantom validation, the cameras were rigidly mounted to an optical bench using

brass posts (Fig. 1). For the animal study they were attached to an aluminum frame mounted to the PET scanner (Fig. 6).

### B. Camera Calibration

The internal optical parameters of each camera and the exterior orientation (i.e., the rigid-body transformation relating the camera frame to the world frame) were obtained using the multi-camera self-calibration toolbox [16]. A sequence of several thousand frames of a small light source (the calibration object) moved throughout the working volume served as input. The source was rastered slowly through the working volume using a 6-axis robot (Epson C3-A601S, SEIKO Corp., Japan). Uniform sampling of the working volume in this way gave more accurate calibration results than waving the light source by hand (the approach suggested in [16]).

Assuming a pinhole camera model in which world points $\mathbf{X} = (x, y, z)^{\mathrm{T}}$ project to image points $\mathbf{x} = (u, v)^{\mathrm{T}}$ according to

$$\mathbf{x} = \mathbf{P}\mathbf{X} \tag{1}$$

where $\mathbf{P}$ is the camera calibration matrix, the set of calibration matrices $\{\mathbf{P}_k, k = 1, \ldots, N\}$ is obtained in [16] by iteratively minimizing the reprojection error between measured and estimated object locations. An advantage of this method over other multi-camera calibration methods is that the calibration object need not be visible to all cameras in any given frame. However, we enforced this requirement to improve the calibration accuracy. We also improved the accuracy by supplying independent fixed estimates of the internal camera parameters (principal point, focal length) and lens distortion coefficients (two radial, two tangential) for each camera rather than iteratively estimating these within the calibration. The independent parameter estimates were obtained using the method in [25].

Calibrations obtained using the method in [16] are determined in an arbitrary world frame, the origin of which is the centroid of the cloud of points representing the frame-by-frame location of the calibration object. Moreover, the calibration is determined only up to an unknown scale (i.e., a factor relating distances in the arbitrary world frame to real-world distances). Therefore, transformation of the calibration to a known, real-world reference frame requires that we 1) resolve the scale factor, and 2) determine the (rigid-body) alignment between the arbitrary world frame and the desired reference frame. Scale was easily resolved by comparing known real-world distances (measured using the robot) with distances measured in the markerless frame. Frame alignment is explained in Section III.

### C. Feature Detection

Features were detected using the scale-invariant feature transform (SIFT) algorithm [19]. SIFT features are highly distinctive and able to be detected and matched reliably across disparate views of the same object. They are also reported to provide reasonable discrimination in animal textures [26]. For a full description of the SIFT algorithm and assembly of the SIFT feature descriptor, see [19].

Briefly, SIFT features typically correspond to high contrast of variation across multiple image scales. Examples of SIFT features are shown in Fig. 2 (see also Fig. 11). Mathematically, SIFT features correspond to local maxima or minima in scale space, where scale space represents the image signal as a function of viewing depth. In practice, the discretized scale space can be generated efficiently for a given image using repeated application of a difference-of-Gaussian operator. Each local maximum or minimum in scale space—a so-called "keypoint"—is represented by a 1-D 128-element descriptor, the elements of which are determined from radial sampling of the local image gradients. Since sampling is performed with respect to a local coordinate system defined by the scale and the orientation of the dominant gradient, the keypoint descriptor exhibits invariance to scale, rotation and translation.

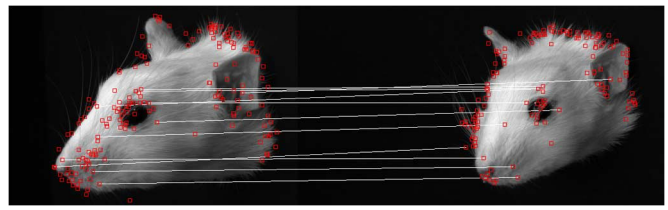


Fig. 2. Example of SIFT features (red squares) and matches (white lines) for a pair of images of a taxidermal rat head.

Using a publicly available implementation of the SIFT algorithm [19], SIFT features were detected in each of the four images (one per camera) comprising a "frame." An example of SIFT applied to images of a taxidermal rat head is shown in Fig. 2. After detecting features, the 2-D feature locations were corrected for lens distortion before further processing.

### D. Feature Matching

SIFT features were matched using a nearest neighbor approach [19]. In this approach we compute the Euclidean distance between a SIFT descriptor, $\mathbf{s}$, and each of its candidate matches. If the distance to the nearest neighbor (denoted "NN") is significantly closer than the distance to the second nearest neighbor (denoted "2NN"), there is a high probability of a true match. Mathematically, this criterion is given by

$$\frac{D(\mathbf{s}, \mathbf{s}_{\mathrm{NN}})}{D(\mathbf{s}, \mathbf{s}_{\mathrm{2NN}})} < \beta \tag{2}$$

where $D$ denotes the Euclidean distance operator and the parameter $\beta$ can take any value in the interval [0,1]. Increasing $\beta$ results in less strict matching, with a concomitant increase in outliers (false matches). A value of 0.6 for $\beta$ typically provides a good trade-off between inliers (true matches) and outliers [19].

In the frame processing pipeline shown in Fig. 3, features were matched between the two images constituting a camera pair (Fig. 3, blue shaded region). Matching *across* camera pairs was found to be unnecessary since the angular separation of the camera pairs and the shape of a rat's face meant few, if any, matches were obtained.

### E. Generating Landmarks

In a true match, the two features represent the projection of a unique 3-D location—what we term a landmark—onto the respective image planes. Neglecting noise, a landmark corresponds to the point of intersection of the back-projected rays joining each of the two feature points to its respective camera center. Determining this intersection, or an estimate of it in the case of noisy data, is called triangulation. We used the direct

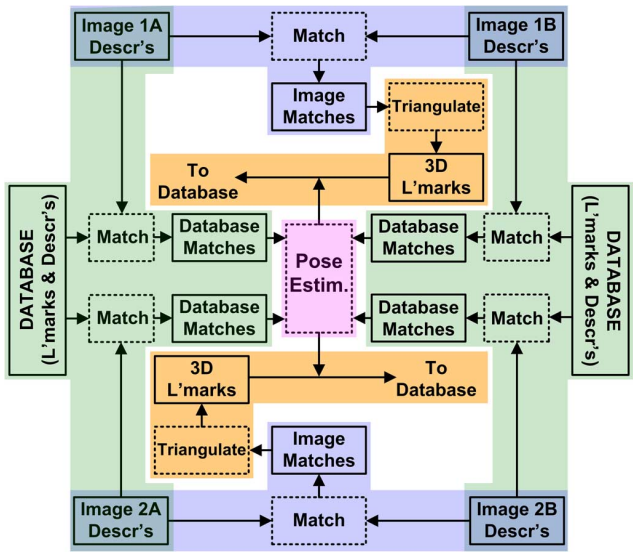



Fig. 3. Flowchart of data processing. Features (SIFT descriptors) in images from the two pairs of cameras (1A/1B and 2A/2B) were matched separately (blue shaded regions) to enable determination of 3-D head landmarks using triangulation (orange shaded regions). Image descriptors were also matched to descriptors stored in the database (green shaded regions) providing the information needed for pose estimation (pink shaded region). Once pose was estimated, new landmarks were transformed to the initial pose and stored in the database with their descriptor (orange shaded regions). Note that dashed boxes refer to processes and nondashed boxes to data (inputs/outputs).

linear transform approach [27], [28] to obtain a triangulation estimate for each feature match. The resulting landmarks were then stored in a database together with a SIFT descriptor (obtained by averaging the descriptors involved in the match). The procedure is shown in Fig. 3 (orange shaded region). Note that once the pose for the frame was computed (see below on pose estimation), any landmarks generated from that frame were spatially transformed according to the inverse pose before being stored in the database. Thus, landmarks accumulating in the database represented a sparse 3-D model of the head surface, in the initial pose, that evolved as each new frame was processed.

### *F. Pose Estimation*

The procedure for pose estimation (Fig. 3, pink region) was similar to that used in [17] to solve the SLAM problem for an indoor mobile robot. Conceptually, it amounts to a 3D-2D registration: finding the optimal alignment between the 3-D head model stored in the database, and features found in the 2-D images comprising each new frame.

For each frame we matched features found in each of the four camera images to features in the database (Fig. 3, green regions). Given at least three such matches, pose can be estimated as the rigid-body transformation minimizing the reprojection error between measured and estimated feature locations (Fig. 4). Reprojection error, $\mathbf{e}$, is the distance (in pixels) between measured and estimated feature locations and is given by

$$\mathbf{e} = \mathbf{x}_k - \mathbf{x}'_k \tag{3a}$$

where

$$\mathbf{x}'_k = \mathbf{P}_k \mathbf{H} \mathbf{X}. \tag{3b}$$

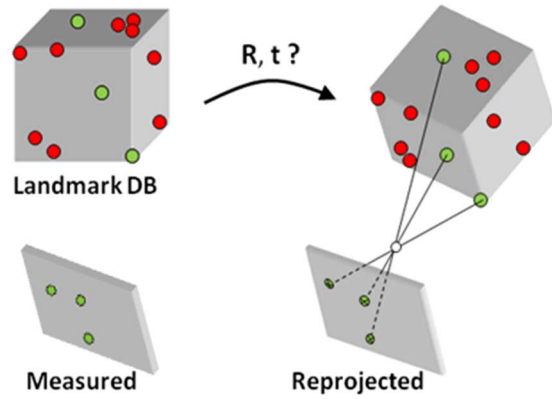



Fig. 4. Pose estimation. After identifying feature matches (shown here in green) between the landmark database and image, pose was estimated by finding the rotation, $\mathbf{R}$, and translation, $\mathbf{t}$, that minimized the reprojection error between measured and estimated feature locations.

Here, $\mathbf{x}_k$ is a measured feature location for camera $k$ and $\mathbf{x}'_k$ is the estimated feature location. The estimated location is obtained in (3b) by transforming the matched database landmark $\mathbf{X}$ according to the current pose estimate $\mathbf{H}$, then projecting it onto the image plane of camera $k$ using the camera calibration matrix operator $\mathbf{P}_k$. In (3b), $\mathbf{H}$ is represented by a $4 \times 4$ rigid-body transformation matrix and, for convenience, $\mathbf{X}$ is expressed using homogeneous coordinates, i.e., $\mathbf{X} = (x, y, z, 1)^{\mathrm{T}}$. Given $n_k$ matches between the database and an image acquired by camera $k$, if the measured landmark locations on the image are $\{\mathbf{x}_{ki} = (u_{ki}, v_{ki}), i = 1, \ldots, n_k\}$ and the corresponding estimated locations are $\{\mathbf{x}'_{ki} = (u'_{ki}, v'_{ki}), i = 1, \ldots, n_k\}$, the reprojection error vector for camera $k$ is given by

$$\mathbf{e}_k = (e_{u1}, e_{v1}, e_{u2}, e_{v2}, \ldots, e_{un_k}, e_{vn_k}) \tag{4}$$

where

$$\begin{aligned} e_{ui} &= u'_{ki} - u_{ki} \\ e_{vi} &= v'_{ki} - v_{ki}. \end{aligned} \tag{5}$$

Minimizing the reprojection errors across all matches to obtain a pose solution was performed using a Gauss-Newton (G-N) iterative approach. Rather than determining the optimal transformation directly, we followed [17], [29] and solved instead for a small correction to the previous pose, viz

$$\hat{\mathbf{h}} = \mathbf{h} + \mathbf{h}'. \tag{6}$$

Here, $\hat{\mathbf{h}}$ is the new pose, $\mathbf{h}$ is the previous pose, and $\mathbf{h}'$ is the correction. In contrast to (3b) where pose is represented using a $4 \times 4$ transformation matrix, here pose is represented as a six-element vector containing the yaw (angle $\psi$ about the $x$-axis), pitch (angle $\theta$ about the $y$-axis), roll (angle $\phi$ about the $z$-axis) and $x$, $y$, and $z$ position coordinates, respectively. This six-parameter representation of pose is equivalent to the $4 \times 4$ matrix representation used in (3b), i.e.,

$$\mathbf{h} \Leftrightarrow \mathbf{H}. \tag{7}$$

Given a vector of reprojection errors, $\mathbf{e}$, calculated according to (4) and (5), we want $\mathbf{h}'$ such that

$$\mathbf{J}\mathbf{h}' = \mathbf{e} \tag{8}$$

where $\mathbf{J}$ is the Jacobian matrix with elements

$$J_{mn} = \frac{\partial e_m}{\partial h'_n}. \tag{9}$$

The underlying assumption here is that small changes in the pose parameters result in proportional changes in the projected location. For the over-constrained case, where there are more error measurements than parameters, we solve for $\mathbf{h}'$ using least-squares minimization

$$\mathbf{J}^{\mathrm{T}}\mathbf{J}\mathbf{h}' = \mathbf{J}^{\mathrm{T}}\mathbf{e}. \tag{10}$$

The Jacobian for camera $k$ is the $6 \times 2n_k$ matrix in which the $(2i-1)$th and $2i$th rows are, respectively

$$\left[\frac{\partial u'_{ki}}{\partial \psi} \quad \frac{\partial u'_{ki}}{\partial \theta} \quad \frac{\partial u'_{ki}}{\partial \phi} \quad \frac{\partial u'_{ki}}{\partial x} \quad \frac{\partial u'_{ki}}{\partial y} \quad \frac{\partial u'_{ki}}{\partial z}\right] \tag{11a}$$

and

$$\left[\frac{\partial v'_{ki}}{\partial \psi} \quad \frac{\partial v'_{ki}}{\partial \theta} \quad \frac{\partial v'_{ki}}{\partial \phi} \quad \frac{\partial v'_{ki}}{\partial x} \quad \frac{\partial v'_{ki}}{\partial y} \quad \frac{\partial v'_{ki}}{\partial z}\right]. \tag{11b}$$

Jacobian elements were computed numerically by determining the small change in image location ($\delta u$ and $\delta v$) for small changes in each of $\psi$, $\theta$, $\phi$, $x$, $y$, and $z$. Note that the error vectors from each camera [i.e., (4)] were concatenated into a single error vector, $\mathbf{e}$, and similarly for the Jacobians [i.e., (11)] to form $\mathbf{J}$.

Equation (10) was solved using singular value decomposition [27]. A total of 10 G-N iterations were performed, with the pose solution being updated according to (6) after each iteration.

### G. Outlier Rejection Strategies

Spurious data arising in the processing pipeline, including SIFT features with poor discrimination and false matches, can limit the reliability of pose estimates. Several outlier rejection strategies were used to exclude spurious data during processing:

*Background Subtraction:* Background subtraction was applied to acquired camera images to restrict detected SIFT features to the region of the rat's head/neck (what we term the "foreground" region). For each camera, intensity thresholding was used to identify the major regions of the images collected during the first few seconds. The region showing the largest change in center-of-mass (CoM) during this time was set to the initial foreground mask. For all subsequent frames, the new foreground mask was taken to be the region with CoM nearest to that of the previous foreground mask.

*Intra-Frame Matching:* During intra-frame feature matching, incorrect matches were detected and removed based on the epipolar geometrical constraint [27]. This constraint reduces the feature matching problem from a 2-D (image) search to a line search. Furthermore, by applying an appropriate transform to the 2-D images (a process termed rectification), the line search may be confined to image rows. We computed the 2-D rectifying transform for each camera using the method in [30] and applied it to the feature points in candidate matches. If the two features in a candidate match were not within five rows of each other the match was rejected.

*Pose Estimation:* During pose estimation, outliers were rejected in a stratified process. The first stage involved discrimination based on the modified Z-score [31]. For a vector $\mathbf{e} = (e_1, e_2, e_3, \ldots, e_n)$, the modified Z-score, $M_i$, for the $i$th element of $\mathbf{e}$ is given by

$$M_i = \frac{0.6745(e_i - \tilde{\mathbf{e}})}{\mathrm{MAD}} \tag{12}$$

where

$$\mathrm{MAD} = \mathrm{median}\left\{|e_i - \tilde{\mathbf{e}}|,\ i = 1, \ldots, n\right\} \tag{13}$$

and $\tilde{\mathbf{e}}$ is the sample median. According to [31], an entry may be considered an outlier if $M_i > 3.5$. Therefore, in each G-N iteration, data points resulting in reprojection errors satisfying this inequality were rejected. The second stage of outlier rejection involved discrimination based on an absolute reprojection error threshold of two pixels. This was applied for the final four (/10) iterations.

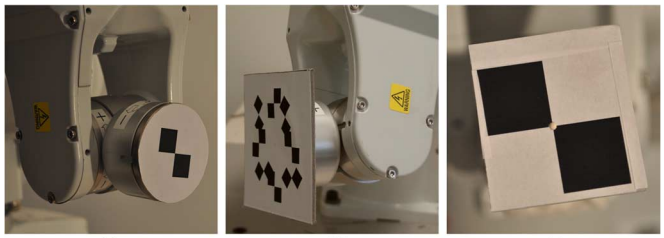
*Pose Checking:* To ensure the final pose was sensible we tested for convergence of the correction term, $\mathbf{h}'$, after 10 iterations. If this check failed, the estimated pose for the frame was set to the previous estimate and no new landmarks generated from the frame were added to the database.

## III. Experimental Validation

### A. Rat Phantom

We performed an initial validation of the markerless tracking system using a taxidermal rat head phantom derived from an adult Sprague Dawley specimen. Fig. 1 shows a photograph of the setup. The phantom was attached rigidly to the end-effector of the six-axis robot and moved according to the motion of an awake rat [11]. The original motion sequence comprised 5000 poses of a rat's head collected over approximately 3 min at a sampling rate of 30 Hz. To eliminate motion blur as a potential source of error in the validation, the sequence was executed as a series of static poses, with the robot pausing between poses to allow for a frame to be collected by the markerless tracking system. All frames were streamed to disk for offline processing. The motion sequence was then repeated with marker-based motion tracking using the MicronTracker (Claron Technology Inc., Toronto, ON, Canada) [32] in conjunction with a marker we have used previously for rat studies [11], [23]. The marker comprised three coplanar "checkerboard" points printed on an area approximately $20 \times 20$ mm$^2$. Repeatability of the robot movements was $\pm 20$ $\mu$m.

To compare pose estimates from the two tracking systems with the known applied robot movements, a spatial cross-calibration was performed between each tracking system and



(a) (b) (c)

Fig. 5. Cross-calibration rigs. Rig used to cross-calibrate the robot to the markerless tracking system (a) and the MicronTracker (b). To cross-calibrate the markerless tracking system to the microPET we used the rig in (c). This had a molecular sieve (seen at the intersection of the black/white regions) with $^{18}$F adsorbed. Cross-calibration methodology is described in [32].

Fig. 6. Integration of the markerless tracking system with the Focus 220 scanner.

the robot. This was performed similarly to the method in [32], using an appropriate calibration target for each system [Fig. 5(a) and (b)].

### B. Live Rat Validation

The markerless tracking system was fitted to the microPET Focus 220 scanner (Preclinical Solutions, Siemens Healthcare Molecular Imaging, USA) using an aluminum frame that locked into the mounting unit in place of the pallet (Fig. 6). Two flexible strips of white LEDs glued to the inside of the microPET bore provided sufficient lighting to use an image exposure time of 5 ms for each camera. The tracker and scanner data streams were synchronized by sending a TTL pulse from the tracker to the gating input of the scanner at the start of each frame exposure (see [11] for further details). Spatial cross-calibration between the tracker and scanner was performed similarly to the method in [32], using the calibration target shown in Fig. 5(c).

Under an approved institutional animal ethics protocol, an adolescent male Sprague–Dawley rat was acclimatized over several days (30 min/day) to an open-ended tube inside the scanner. The rat was injected via the tail vein with approximately 50 MBq $^{18}\mathrm{F}-\mathrm{FDG}$ and scanned 20 min later unanesthetized in the tube for 10 min in conjunction with markerless motion tracking. A second 10-min emission scan (started 45 min postinjection) was performed after adding several markings on the cheeks and around the eyes of the rat using a permanent marker pen. This was to test the benefit of amplifying the number of detectable features on relatively rigid parts of the face. For both scans, tracking was performed at 30 fps, leading to a total of 17000 frames/camera. Finally, a 20 min emission scan (started 90 min postinjection) was acquired with the rat under 1.5% isoflurane/$O_2$ gas anesthesia. Data from this scan were used as a motion-free reference.

We note that although a tracer like $^{18}\mathrm{F}-\mathrm{FDG}$ would not, in general, necessitate awake animal imaging, it was chosen here for its broad uptake in the brain and other head structures, enabling easier visualization of the effect of motion correction on the resulting images.

Tracker data collected during the awake studies were processed offline to obtain a pose estimate for each of the 17000 frames, as described in Section IIC-G. Pose estimates were used within an event-driven list mode motion compensation reconstruction algorithm [33] to obtain images of the $^{18}\mathrm{F}-\mathrm{FDG}$ distribution in the rat brain. Calculated attenuation correction was applied as a pre-correction. In the reconstruction we used an isotropic, spatially-invariant point spread function, modeled as the weighted sum of two Gaussians (FWHM 1.3 and 4.5 mm, respectively, and ratio 0.05). The reconstructed pixel size was $0.949\ \mathrm{mm} \times 0.949\ \mathrm{mm} \times 0.796\ \mathrm{mm}$.

To further assess the reliability of pose estimates, the 3-D coordinates corresponding to the center of each eye and the nose tip were obtained manually from the initial frame, then for all subsequent frames these test points were transformed according to the estimated pose and reprojected according to (3b). Robustness of the pose estimates using this method is indicated by close correspondence of the reprojected features with the eyes and nose tip.

## IV. Results

### A. Camera Calibration

The mean calibration accuracy for the multi-camera setup, measured in terms of the reprojection error, was 0.12 pixels $\pm 0.12$ pixels. At our working distance of 350 mm this represented a positional accuracy on the surface of the rat's head of approximately 0.025 mm $\pm$ 0.025 mm.

### B. Phantom Study

Fig. 7 shows the number of landmarks accumulated in the database as a function of frame number for the phantom study. Approximately 1500 landmarks were collected over 5000 frames. The rate of detection of new landmarks increased noticeably whenever the head moved to a previously unseen orientation (indicated by arrows in Fig. 7).

Fig. 8 shows the estimated rotation about the $x$-axis over the last 500 poses. There was excellent agreement between the applied robot motion (black) and the markerless estimate (green). By comparison, the marker-based estimate (red) exhibited bias of 3°–4° and also the characteristic jitter associated with small markers [11], [34]. The comparison was similar for the other degrees-of-freedom (dof).

Fig. 9 shows the distribution of database landmarks based on the marker-based [red dots, Fig. 9(a)] and markerless [green dots, Fig. 9(b)] motion estimates, in each case superimposed on the ground truth distribution (black circles). A magnified view

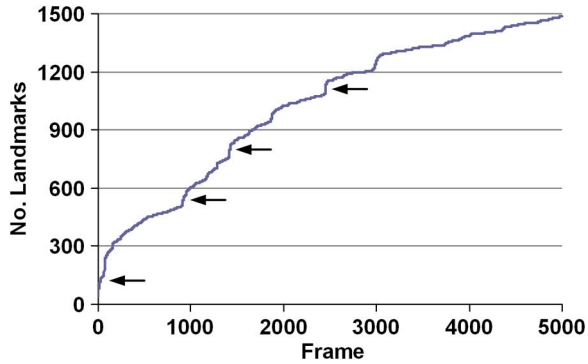



Fig. 7. Cumulative number of landmarks as a function of pose number for the taxidermal rat phantom. Arrows indicate occasions where the rate of new landmarks increased due to the head moving to a previously unseen orientation.

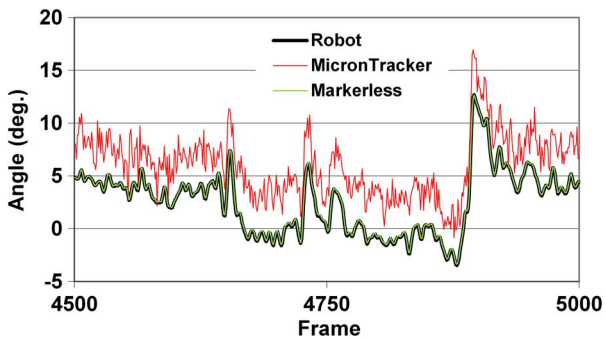



Fig. 8. Last 500 poses of the $x$-axis rotation component of motion of the phantom: applied robot motion (black), marker-based motion estimation (red), and markerless motion estimation (green). Note that the (green) markerless estimate almost exactly overlays the known motion.

of one area of the head is shown in Fig. 10 to indicate the relative mismatch between the estimated and true landmark distributions. Overall mismatch was quantified as the root mean square (rms) of the vector displacement between estimated and ground truth locations, computed over all landmarks. RMS error was 1.8 mm and 0.15 mm for the marker-based and markerless motion tracking systems, respectively.

### C. Live Rat Study

Fig. 11 shows an example of the feature detection with and without added face markings. When face markings were used, SIFT features tended to concentrate at these locations leading to an increased number of features for matching, and 7000 additional landmarks in the database (Fig. 12).

Estimated head rotation about the $x$-axis, with and without the use of face markings, is shown in Fig. 13. For each scenario, six images exemplifying the range of head motion during the scan are shown. Without face markings [Fig. 13(a)], the algorithm failed to converge to a head pose solution in 44% of frames. By comparison, with face markings we obtained a head pose estimate in 99.1% of frames [Fig. 13(b)].

Comparing the alignment of measured and estimated (reprojected) facial structures, there was frequent (59%) misalignment $>4$ pixels when face markings were not used. By contrast, with face markings the alignment was within 4 pixels for all test features in approximately 80% of frames.

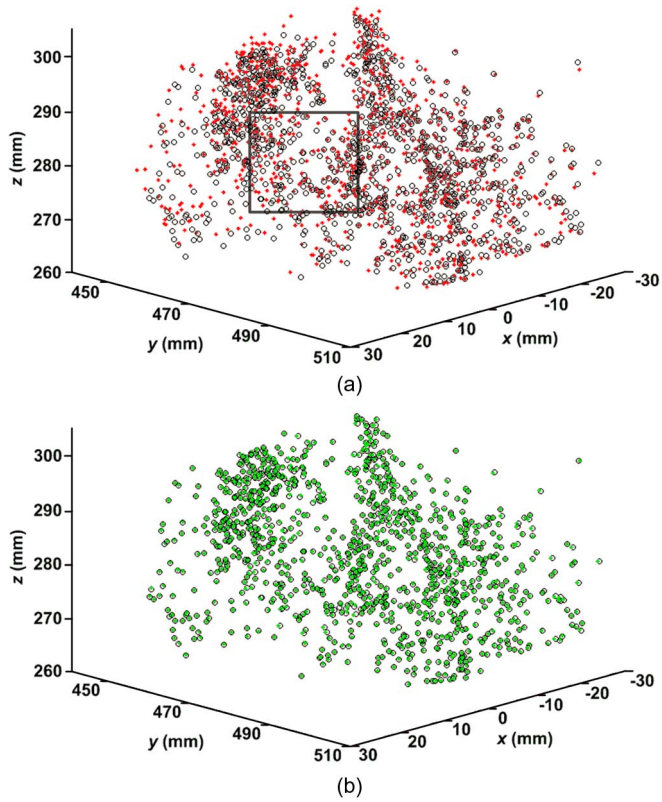



Fig. 9. Distribution of landmarks compared to ground truth for the taxidermal rat. Plots show the complete distribution of landmarks for the marker-based (a) and markerless (b) motion tracking approaches. Black circles represent ground truth.

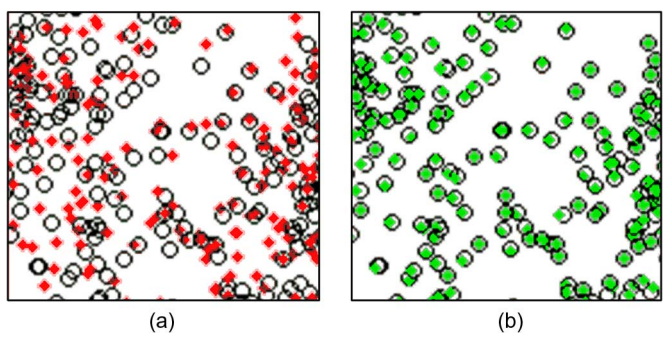



Fig. 10. Zoomed distribution of landmarks compared to ground truth. The zoomed region corresponds to the black box in Fig. 9(a). Degree of mismatch between landmarks obtained using marker-based (red dots) (a) and markerless (green dots) (b) motion tracking relative to the ground truth landmark locations (black circles).

Fig. 14 shows the reconstructed $^{18}\mathrm{F}-\mathrm{FDG}$ distribution in the rat brain with and without motion compensation, and with and without face markings. Regardless of whether face marking was used, motion compensation provided an obvious improvement compared to no motion correction. The Harderian and salivary glands, lying outside the brain, were the brightest structures in all images. However, because of their small size, these structures were more susceptible to partial volume-like effects caused by inaccuracies in the motion data. Thus, the global peak voxel values after motion correction were lower than for the motion-free case. Nevertheless, in the brain, activity and contrast recovery was promising: for the unmarked rat, peak activity and

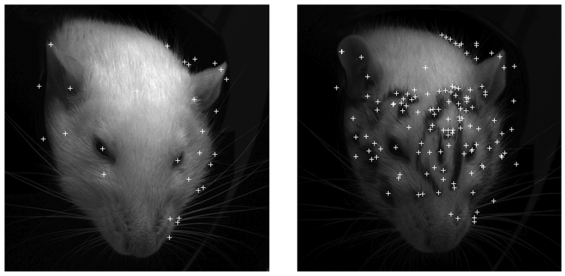


Fig. 11. SIFT features detected for the live rat without (left) and with (right) added face markings.

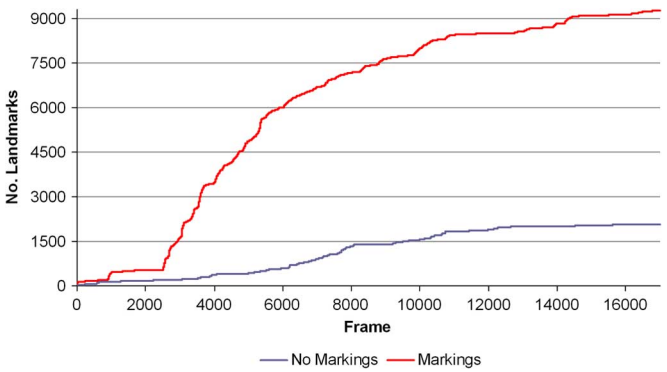



Fig. 12. Cumulative number of landmarks as a function of frame number for the live rat without (blue) and with (red) added face markings.

local contrast (measured as the local peak height) in the striatal profile shown in Fig. 15 were 88% and 41%, respectively. Corresponding values for the marked rat were 93% and 50%.

This confirms what is apparent qualitatively in Fig. 14: motion correction—and therefore the underlying motion estimates—were superior in the marked rat (see arrows in Fig. 14). This is in spite of more frequent and larger amplitude motion during this particular study. In Fig. 14, examples of the reduced motion artifact and increased contrast throughout the head obtained using facial marking are indicated by the red and blue arrows, respectively.

## V. Discussion

We have developed a stereoscopic tracking system and associated algorithms to measure continuous rigid-body head motion of an animal without the need for attached markers. A four-camera implementation of the system was validated in both a phantom study and in live animal PET imaging experiments. The method comprises mostly standard algorithms and hardware; our effort has been to integrate these disparate components into a coherent system for scalable multi-view markerless tracking of rapid and continuous, fully general rigid-body head motion of rats, and to demonstrate its feasibility for motion-compensated preclinical imaging. We believe preclinical imaging is a key area of application for motion tracking since it would allow studies on awake animals and thereby enable a range of new experiments that depend on the subject being conscious (see Section I for examples).

In the phantom study, the positional accuracy of head landmarks using markerless pose estimates was approximately ten times better than using pose estimates from a high performance commercially available marker-based tracking system. This result is less surprising when one considers that the accuracy and precision of pose estimates obtained from sparse feature-based models generally improves as the number of features and/or spread of features increases [34], [35]. The marker used in the phantom study had three features (the minimum needed for pose estimation) distributed over an area of approximately 0.6 $\mathrm{cm}^2$. By contrast, the markerless system utilized 10–100 features from each camera over a surface area of approximately 20 $\mathrm{cm}^2$ (i.e., the entire face).

In the live animal experiments, ground truth motion data were not available as in the phantom study. Instead, we relied on the quality of the motion-corrected PET images, relative to the motion-free image, as a surrogate measure of tracking accuracy. Residual errors and blurring in the motion-corrected reconstructions (Fig. 14) suggest that the tracking accuracy was somewhat poorer than the PET spatial resolution ($\sim$1.2 mm), and certainly not as accurate as for the phantom study. Overall, the motion-corrected images were comparable to images obtained previously using marker-based tracking [11].

In order for pose estimates from sparse feature-based models to improve with more features, or a greater spread of features (as described above), the extra information must outweigh the contribution of noise [36]. This helps to explain the poorer tracking obtained in the live animal studies: there was a key source of noise in these studies that was not present in the phantom study, namely nonrigid features. Nonrigid features are features whose motion is not tightly coupled to brain motion. Examples include background features, and features on the neck, ears, eyes, nose, and whiskers. Indeed, of the features remaining on the neck/face after background subtraction, we found typically 50% or more were concentrated on nonrigid structures.

Two strategies were used to reduce the impact of nonrigid features on the head. The first was to reduce the relative proportion of nonrigid features (outliers) by amplifying the number of inliers—that is, features on stable parts of the head. We did this by making arbitrary, indelible marks on the fur with a marker pen, in areas away from the eyes, nose, whiskers, and ears. The marks took only a matter of seconds to apply and had no observable impact on the animal's behavior. The approach is akin to the projected light patterns used in structured light techniques to assist feature detection and matching. Importantly, applying marks did not share the disadvantages of using attached markers (see Section I). The second strategy used to reduce the impact of nonrigid features was to incorporate stratified statistical and absolute threshold-based outlier rejection into the G-N pose estimation procedure. This method enforced a reprojection error consistency condition that eliminated features moving nonrigidly by $>2$ pixels and tended to constrain the data used for pose estimation to be highly conformant to a rigid-body model.

The effectiveness of these strategies is borne out in the results. When markings were used, reprojected test features consistently aligned to the anatomical test landmarks to within four pixels, even when only a small fraction of the head was visible (Fig. 13(b), third from left); a sensible pose solution was found for nearly every frame; and despite the rat exhibiting larger and more frequent movement during the study in which markings

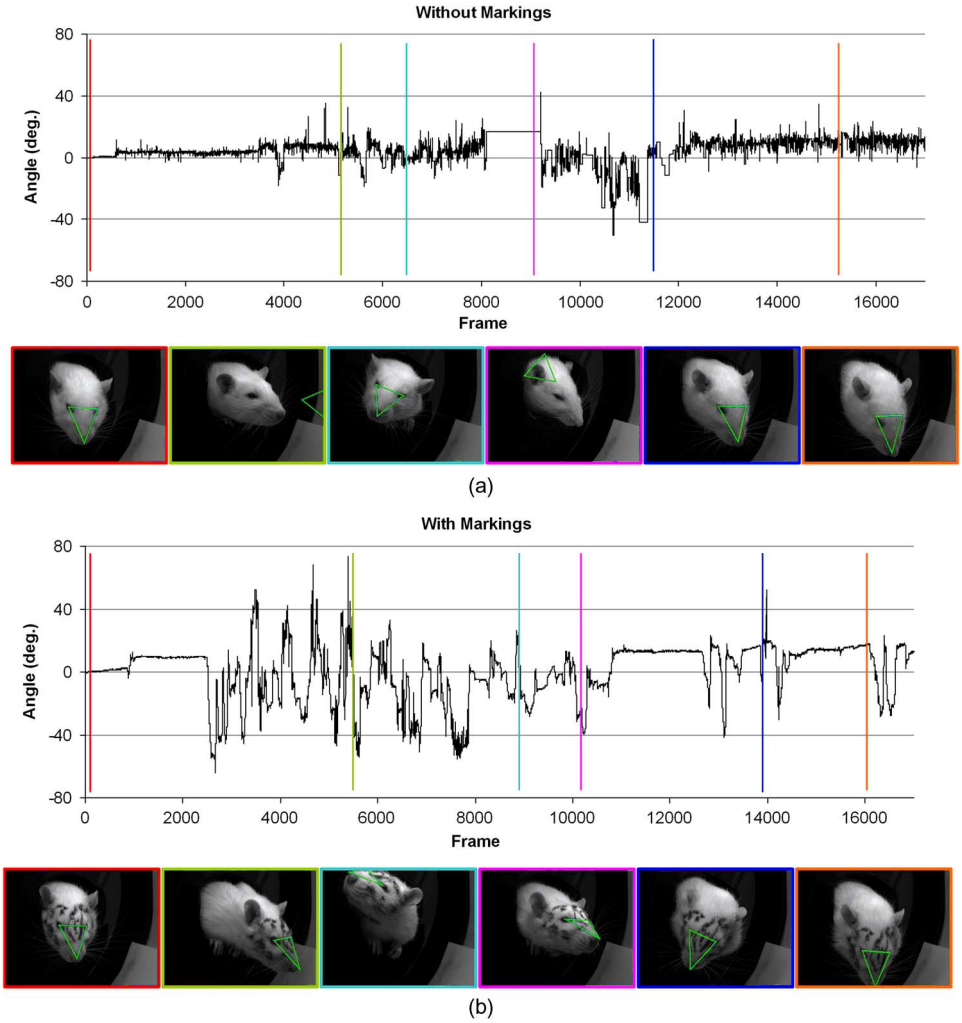



Fig. 13. *X*-axis rotation component of the estimated head motion of the awake rat without (a) and with (b) face markings. In each case six images (acquired using camera 1) at various stages in the study are shown to indicate the range of rat head movement. Image borders are color-coded with the vertical bars drawn on the corresponding graph to indicate where in the sequence the image occurred. Overlaid on these images is a green triangle which, for reliable pose estimates, should connect the two eye centers and nose tip. See main text for details.

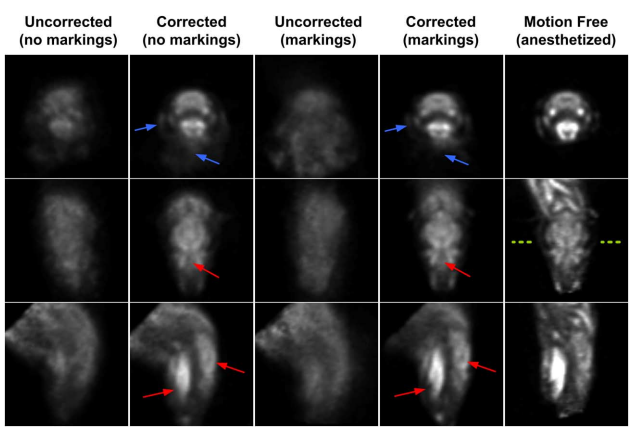



Fig. 14. Transverse (top row), coronal (middle row), and sagittal (bottom row) reconstructed slices of the $_{18}$F-FDG distribution in the rat brain. Blue arrows indicate examples of reduced motion artifact in the "markings" case compared to the "no markings" case. Similarly, red arrows indicate examples of increased contrast throughout the head for the "markings" case compared to the "no markings" case. Green bars (coronal slice of the motion-free image) indicate the level at which the coronal profiles in Fig. 15 were obtained.

were applied [compare Fig. 13(a) and (b)], the motion-compensated images were visually and quantitatively closer to the motion-free case than for the unmarked case. By contrast, when the rat was unmarked, the likelihood that very few (or even insufficient) features remained for pose estimation after applying outlier rejection increased dramatically such that a sensible pose solution was obtained in $<50\%$ of frames; and the likelihood that outliers (nonrigid features) outnumbered inliers (rigid features) also increased, leading to more incorrect poses and jitter.

Other strategies (not implemented here) could be used to reduce the impact of nonrigid features. For example, the background masking could be extended to segment specific regions of the face and restrict processing to regions containing only rigid features. Systematic pruning of unreliable features from the database would also reduce the likelihood of false matches, and reduce the size of the search when matching features to the database. Another possibility is to estimate pose using a robust estimator such as random sampling with consensus (RANSAC) [37]. Finally, a more general approach is to incorporate error modeling in the processing—keeping up-to-date information on the uncertainty of each landmark (in the form of a covariance matrix), updating this uncertainty each time the landmark is reobserved, and weighting each landmark in the pose estimation according to its uncertainty. Such an approach is typical in

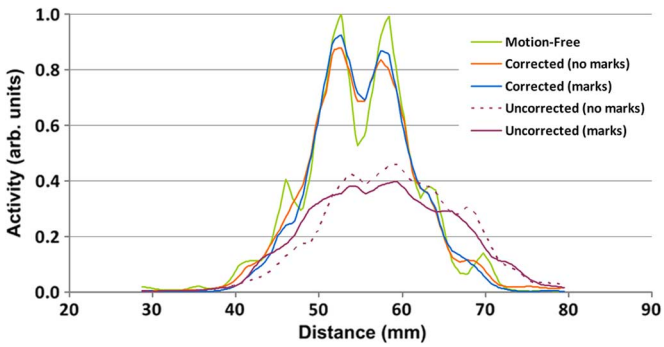



Fig. 15. Striatal profiles of $_{18}$F-FDG uptake (arbitrary units) measured on the respective coronal slices of Fig. 14, at the level indicated by the green bars in Fig. 14.

SLAM for mobile robotics (e.g., [17]) and is easily transferable to this application. In future work, we plan to explore each of these strategies as an alternative to marking the fur.

There are several aspects of the algorithms underlying our markerless tracking system that make it particularly flexible and which help to generate robust pose estimates. Firstly, the camera calibration procedure allows it to be multi-view scalable: that is, it allows for any number of cameras, positioned arbitrarily. Secondly, an important benefit of our 3D-2D registration approach for pose estimation is that even features observed by a single camera can be used to constrain the estimate. And thirdly, the collection of a landmark database means that poses are computed using information up to and including the initial frame. This makes it less prone to drift than purely frame-to-frame methods such as [43] (see below for a discussion of drift).

Drift is an important issue to consider when comparing marker-based and markerless tracking. Drift refers to steadily worsening pose estimation over time. Marker-based tracking is not, in general, susceptible to drift since each pose is computed independently. By contrast, the markerless tracking approach relies on previous information (landmarks collected since the initial frame), potentially making it susceptible to drift caused by the accumulation of error in landmark locations. However, in the phantom study there was no evidence of drift over the 5000 frames. The fact that all features were rigid would have helped here. In the live animal studies, the robust alignment of reprojected test features suggest that drift was minimal over the 20 min period. However, further work is needed to properly assess the likelihood and impact of drift in real subjects. The error modelling described above would function to counteract drift and should also be investigated in future work.

In this work, we did not make any attempt to optimize the processing speed of markerless tracking. With the current implementation, processing 5000 frames in the phantom study took approximately 3.5 h: 0.5 h for SIFT feature detection on a Dell M610 Blade with two 2.8-GHz Xeon Processors which allowed 12 cores to run 24 threads simultaneously with hyper-threading, and 3 h for landmark generation and pose estimation using a 4-core Intel Xeon 1.86-GHz processor. Processing 17 000 frames in the live animal studies took approximately 10 h.

There are several ways to improve processing efficiency. Feature detection could be sped up by using faster scale-invariant feature detectors (e.g., [38], [39]), GPU implementations [40], or non-scale-invariant feature detectors (e.g., [41]). However, to be useful for this application, the feature detector must maintain high specificity on furry textures. The pose processing component could be sped up by improving the efficiency of the database search—either by using a more directed search of the most likely match candidates [17], [19] or, alternatively, by reducing the overall size of the database. Examples of the latter include regularly pruning the database of descriptors with low specificity, or initializing a new database whenever a certain quota of landmarks is reached and globally aligning the databases post hoc [42].

The only similar markerless tracking system and application we are aware of is that described by Goddard for head tracking of mice in single photon emission computed tomography (SPECT) [43], [44], though this system does not appear to have been applied in animal studies yet. In that approach, corner features detected in one frame were tracked into the next frame using cross-correlation. This frame-frame tracking requires the concatenation of successive pose estimates to compute the change at any given time with respect to the initial pose. Therefore, it may be more susceptible to drift compared to our method which makes use of all features detected since the start of the scan.

In summary, markerless tracking appears very promising for enabling accurate motion-compensated imaging of awake animals. These initial results demonstrate encouraging accuracy and precision, and a large range of detectable motion. The results also suggest that further improvements in performance could be achieved through robust handling of nonrigid features. In addition to this optimization, and improved processing efficiency, we also plan to explore the potential for this method to be translated to the clinical setting and to other imaging modalities.

## VI. Conclusion

We have developed a markerless tracking system and demonstrated the feasibility of using this system to obtain highly accurate ($<$0.2 mm rms error) six dof pose estimates for continuous motion. The system is based on a SLAM framework and involves identifying a consistent set of landmarks on the head for pose estimation. Pose estimates obtained for realistic head motion of a taxidermal rat were more accurate by an order-of-magnitude compared to those obtained from a state-of-the-art marker-based system. The feasibility of using the approach for accurate motion compensated PET imaging of an awake rat was also demonstrated. There is considerable potential for further development to improve the accuracy, precision and computational efficiency of the system, and to adapt it to the clinical setting with PET and other modalities.

## Acknowledgment

The authors would like to thank S. Ingleby and the Australian Museum for the taxidermal rat specimen, and to J. Eisenhuth

for electronics support. The authors also thank Dr. T. Svoboda, Dr. A. Straw, Prof. D. Lowe, and Dr. S. Williams for valuable discussions during the course of this work.

## REFERENCES


[1] S. Lancelot and L. Zimmer L, "Small-animal positron emission tomography as a tool for neuropharmacology," *Trends Pharmacol. Sci.*, vol. 31, pp. 411–417, 2010.

[2] T. Massoud and S. Gambhir, "Molecular imaging in living subjects: Seeing fundamental biological processes in a new light," *Genes Develop.*, vol. 17, pp. 545–580, 2003.

[3] C. Martin, J. Martindale, J. Berwick, and J. Mayhew, "Investigating neural-hemodynamic coupling and the hemodynamic response function in the awake rat," *Neuroimage*, vol. 32, pp. 33–48, 2006.

[4] R. Hosoi, A. Matsumura, S. Mizokawa, M. Tanaka, and F. Nakamura F *et al.*, "MicroPET detection of enhanced 18F-FDG utilisation by PKA inhibitor in awake rat brain," *Brain Res.*, vol. 1039, pp. 199–202, 2005.

[5] W. Schiffer and C. Liebling *et al.*, "Cue-induced dopamine release predicts cocaine preference: Positron emission tomography studies in freely moving rodents," *J. Neurosci.*, vol. 29, no. 19, pp. 6176–6185, 2009.

[6] P. Vaska, C. Woody, D. Schlyer, S. Shokouhi, S. Stoll, J.-F. Pratte, P. O'Connor, S. Junnarkar, S. Rescia, B. Yu, M. Purschke, A. Kandasamy, A. Villanueva, A. Kriplani, V. Radeka, N. Volkow, R. Lecomte, and R. Fontaine, "RatCAP: Miniaturized head-mounted PET for conscious rodent brain imaging," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2718–2722, Oct. 2004.

[7] D. Schulz, S. Southekal, S. Junnarkar, J.-F. Pratte, and S. Junnarkar *et al.*, "Simultaneous assessment of rodent behavior and neurochemistry using a miniature positron emission tomograph," *Nature Methods*, vol. 8, pp. 347–352, 2011.

[8] S. Cherry, "Functional whole-brain imaging in behaving rodents," *Nature Methods*, vol. 8, pp. 301–303, 2011.

[9] P. Bloomfield, T. Spinks, J. Reed, L. Schnorr, A. Westrip, L. Livieratos, R. Fulton, and T. Jones, "The design and implementation of a motion correction scheme for neurological PET," *Phys. Med. Biol.*, vol. 48, pp. 959–978, 2003.

[10] V. Zhou, A. Kyme, S. Meikle, and R. Fulton, "An event driven motion correction method for neurological PET studies of awake laboratory animals," *Mol. Imag. Biol.*, vol. 10, pp. 315–324, 2008.

[11] A. Kyme, V. Zhou, S. Meikle, C. Baldock, and R. Fulton, "Optimised motion tracking for positron emission tomography studies of brain function in awake rats," *PLoS ONE* vol. 6, no. 7, p. e21727, Aug. 2011 [Online]. Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0021727

[12] J. Baba, C. Endres, C. Foss, S. Nimmagadda, H. Jung, J. Goddard, S. Lee, J. McKisson, M. Smith, A. Stolin, A. Weisenberger, and M. Pomper, "Molecular imaging of conscious, unrestrained mice with AwakeSPECT," *J. Nucl. Med.*, vol. 54, pp. 969–976, 2013.

[13] A. Kyme, J. Eisenhuth, W. Ryder, G. Angelis, M. Akhtar, K. Popovic, R. Bashir, V. Zhou, R. Fulton, and S. Meikle, "Developing a system for the molecular imaging of freely moving rats," presented at the World Mol. Imag. Conf., Dublin, Ireland, 2012.

[14] J. Gall, C. Stoll, E. de Aguiar, C. Theobalt, B. Rosenhahn, and H. Seidel, "Motion capture using joint skeleton tracking and surface estimation," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, 2009, pp. 1746–1753.

[15] D. Grover, J. Tower, and S. Tavare, "O fly, where art thou?," *J. R. Soc. Interface*, vol. 5, pp. 1181–1191, 2008.

[16] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multi-camera self-calibration for virtual environments," *Presence, Teleoperators Virt. Environ.*, vol. 14, pp. 407–422, 2005.

[17] S. Se, D. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *Int. J. Rob. Res.*, vol. 21, pp. 735–758, 2002.

[18] R. Szeliski, *Computer Vision: Algorithms and Applications*, 1st ed. Berlin, Germany: Springer, 2011.

[19] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[20] P. Noonan, J. Howard, T. Cootes, W. Hallett, and R. Hinz, "Real-time markerless rigid body head motion tracking using the Microsoft Kinect," in *Proc. 2012 Nucl. Sci. Symp. Med. Imag. Conf.*, Los Angeles, CA, 2012, pp. 2241–2246.

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

[22] O. Olesen, J. Sullivan, T. Mulnix, R. Paulsen, L. Hojgaard, B. Roed, R. Carson, E. Morris, and R. Larsen, "List-mode PET motion correction using markerless head tracking: Proof-of-concept with scans of human subject," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 200–209, Feb. 2013.

[23] A. Kyme, S. Meikle, C. Baldock, and R. Fulton, "Tracking and characterising the head motion of unanaesthetised rats in positron emission tomography," *J. R. Soc. Interface* vol. 9, no. 76, pp. 3094–3107, Jun. 2012 [Online]. Available: http://rsif.royalsocietypublishing.org/content/9/76/3094.full.pdf+html?sid=b0e35d65-4738-4af1-9856-bd78213967d0

[24] A. Kyme, "Optimised motion tracking in small animal positron emission tomography," Ph.D. dissertation, Dept. Phys., Univ. Sydney, Sydney, Australia, 2012.

[25] J.-Y. Bouguet, Matlab Calibration Toolbox [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/

[26] D. Ramanan, D. Forsyth, and K. Barnard, "Building models of animals from video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1319–1334, Aug. 2006.

[27] R. Hartley and A. Zisserman, *Multiple View Geometry*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[28] T. Werner, *Computer Vision Software Library*. , U.K.: Virtual Geometry Group, Dept. Eng. Sci., Univ. Oxford, 2001 [Online]. Available: http://www.robots.ox.ac.uk

[29] D. Lowe, "Fitting parameterized three-dimensional models to images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 5, pp. 441–450, May 1991.

[30] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Mach. Vis. Appl.*, vol. 12, pp. 16–22, 2000.

[31] B. Iglewicz and D. Hoaglin, *How to Detect and Handle Outliers*. Milwaukee, WI, USA: ASQC Quality Press, 1993.

[32] A. Kyme, V. Zhou, S. Meikle, and R. Fulton, "Real-time 3-D motion tracking for small animal brain PET," *Phys. Med. Biol.*, vol. 53, pp. 2651–2666, 2008.

[33] A. Rahmim, K. Dinelle, J. Cheng, M. Shilov, W. Segars, S. Lidstone, S. Blinder, O. Rousset, H. Vajihollai, B. Tsui, D. Wong, and V. Sossi, "Accurate event-driven motion compensation in high-resolution PET incorporating scattered and random events," *IEEE Trans. Med. Imag.*, vol. 27, no. 8, pp. 1018–1033, Aug. 2008.

[34] A. Kyme, J. Maclaren, S. Meikle, C. Baldock, and R. Fulton, "The effect of time domain pose filtering on accuracy of small marker based motion correction in awake animal PET," in *Proc. 2011 IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, Valencia, Spain, 2011, pp. 2557–2560.

[35] J. Fitzpatrick, J. West, and C. Maurer, "Predicting error in rigid-body point-based registration," *IEEE Trans. Med. Imag.*, vol. 17, pp. 694–702, Oct. 1998.

[36] T. Broida, S. Chandrashekhar, and R. Chellappa, "Recursive 3-D motion estimation from a monocular image sequence," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 4, pp. 639–656, Jul. 1990.

[37] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Graph. Image Proc.*, vol. 24, no. 6, pp. 381–395, 1981.

[38] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[39] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. 2011 IEEE Int. Conf. Comput. Vis.*, Barcelona, Spain, 2011, pp. 2548–2555.

[40] S. Sinha, J.-M. Frahm, M. Pollefeys, and Y. Genc, "Feature tracking and matching in video using programmable graphics hardware," *Mach. Vis. Appl.*, vol. 22, pp. 207–217, 2011.

[41] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. 2011 IEEE Conf. Comput. Vis.*, Barcelona, Spain, 2011, pp. 2564–2571.

[42] S. Se, D. Lowe, and J. Little, "Vision-based global localization and mapping for mobile robots," *IEEE Trans. Robot.*, vol. 21, no. 3, pp. 364–375, Jun. 2005.

[43] J. Goddard, J. Baba, S. Lee, A. Weisenberger, A. Stolin, J. McKisson, and M. Smith, "Improvements in intrinsic feature pose measurement for awake animal imaging," in *Proc. 2010 IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, Knoxville, TN, 2010, pp. 2814–2818.

[44] J. Goddard, J. Baba, S. Lee, A. Weisenberger, A. Stolin, J. McKisson, and M. Smith M, "Intrinsic feature pose measurement for awake animal SPECT imaging," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, Orlando, FL, 2009, pp. 2557–2560.