Novel SLAM-Based Markerless Motion Tracking of Conscious Unrestrained Rodents in PET

Andre Kyme, *Student Member*, *IEEE*, Stephen Se, *Senior Member*, *IEEE*, Steven Meikle, *Senior Member*, *IEEE*, Clive Baldock, Will Ryder, *IEEE*, *Member*, Roger Fulton, *Senior Member*, *IEEE*

Abstract*– Motion-compensated positron emission tomography (PET) has the potential to improve translational investigations by allowing animals to behave normally and respond to external stimuli. Several groups have demonstrated the feasibility of performing motion-compensated brain PET on rodents, obtaining the necessary head motion data using marker-based techniques. However, markerless motion tracking would simplify animal experiments, be less invasive, and potentially provide more accurate pose estimates over a greater range of motion. We describe a markerless stereo motion tracking system and demonstrate the feasibility of using this system to obtain highly accurate (< 0.2 mm) pose estimates for realistic motion of a taxidermied rat head. The system is based on the simultaneous localization and mapping (SLAM) framework used in mobile robotics and involves building a consistent set of landmarks on the head for pose estimation. Pose measurements using the markerless system were approximately 10 times more accurate than a state-of-the-art marker-based system. Planning of experiments to validate markerless tracking in microPET imaging of awake animals is currently underway.

I. Introduction

Previously we developed a motion tracking and correction system for small animal positron emission tomography (PET) to enable brain imaging in awake rats and thereby to expand the potential applications of PET in drug and disease related research [1, 2]. Successful head motion tracking was reported using a commercially available binocular system and a synthetic head marker. In this work we describe an entirely new, in-house vision-based system to obtain head pose without markers. Compared to a marker-based approach, markerless tracking is less invasive, requires less acclimatization of the animals, and potentially enables tracking of an unlimited range of head motion. It is, therefore, an attractive alternative if sufficient accuracy can be achieved.

Our feature detection and tracking were inspired by the simultaneous localization and mapping (SLAM) problem in robotics. SLAM uses robot-mounted sensors to simultaneously determine a robot's motion together with a map of the environment (e.g. [3]). Rat tracking is an analogous problem if one considers the sensor (tracking) frame fixed and the rat as the moving environment. We demonstrate how the SLAM framework enables a sparse map of the head to be built up over time and how this can be used for pose estimation.

The new motion tracking system was tested on realistic rat head motion of a taxidermied rat head phantom. We also compared the performance to a marker-based system. The comparison was possible using a 6-axis robot to perform known, precisely controlled motion.

II. Methods

A. Motion tracking system

The tracking system consisted of four Flea2 Firewire cameras (Point Grey Research, Canada) fitted with 12 mm lenses (GMN21214, Goyo Optical, Japan). Synchronized capture rates of 60 Hz were possible. The cameras were arranged in front of the animal looking slightly down on the head. They were organized into two pairs viewing opposite sides of the head in order to ensure that detected features were well dispersed (see section IIC).

B. Calibration

Calibration of the camera system was performed using [4]. In this method images of a tiny laser light source moved throughout the working volume serve as input and the over-determined system of two-dimensional (2D) to three-dimensional (3D) correspondence equations is solved to obtain an estimate of intrinsic and extrinsic camera parameters. Although lens distortion coefficients can optionally be included, we obtained greater calibration accuracy (< 0.15 pixels) by solving for the distortion separately using [5]. The resulting calibration was in an arbitrary world frame with unknown scale. Therefore, a further calibration was performed to resolve the scale ambiguity and transform the camera calibration to the robot frame (see section IIF).

C. Feature detection and matching

Features were detected in each image using the scale-invariant feature transform (SIFT) algorithm [6]. SIFT

This work was supported in part by Australian Research Council Grant No. DP 0663519.

André Kyme is with the School of Physics and Brain & Mind Research Institute, University of Sydney, NSW 2006, Sydney, Australia (e-mail: a.kyme@physics.usyd.edu.au).

Stephen Se is with MDA Systems Ltd., 13800 Commerce Parkway, Richmond, B.C. V6Y 1B4 Canada (e-mail: sse@mdacorporation.com).

Steven Meikle is with the Faculty of Health Sciences and Brain & Mind Research Institute, University of Sydney, NSW 2006, Sydney, Australia (e-mail: steven.meikle@sydney.edu.au).

Clive Baldock is with the School of Physics, University of Sydney, NSW 2006, Sydney, Australia (e-mail: clive.baldock@sydney.edu.au).

Will Ryder is with the Brain & Mind Research Institute and Faculty of Health Sciences, University of Sydney, NSW 2006, Sydney, Australia (e-mail: will.ryder@sydney.edu.au).

Roger Fulton is with the School of Physics, Faculty of Health Sciences, and Brain & Mind Research Institute, University of Sydney, NSW 2006, Sydney, Australia, and with the Department of Medical Physics, Westmead Hospital, NSW 2145, Sydney, Australia (e-mail: r.fulton@physics.usyd.edu.au).

features are scale, rotation, translation invariant and are therefore well suited to feature detection and matching in stereo applications. SIFT features are characterised by a normalized descriptor vector which can be compared with the descriptors obtained from other images to determine matches. Features were matched between camera images for a given frame using the nearest neighbour approach [6] (see Fig. 1). Since the camera pairs were each viewing a different side of the animal's head, this intra-frame feature matching was only performed between images of a given pair and not between the pairs themselves. Identified feature matches were assigned a new SIFT descriptor equal to the normalized average of the matched descriptors. Matched features were reconstructed into 3D landmarks (LMs) and added to a LM database (see below).

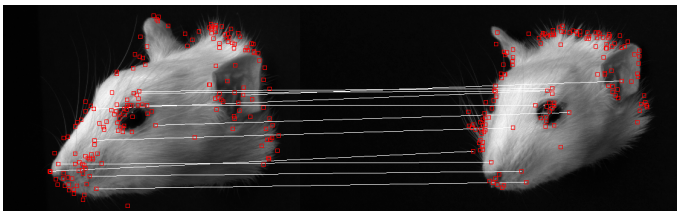

Fig. 1. Example of SIFT features (red) and matches (white lines) for a pair of images of a taxidermied rat head.

D. *Frame processing*

Frame processing was performed similarly to [3] where a database of 3D LMs (initially empty) was continuously updated as new frames were processed. For each frame, SIFT matches between the images comprising that frame were determined. Matches were verified using the epipolar constraint with a 5 pixel threshold. Based on these matches, LMs representing points on the surface of the head were reconstructed. The arrangement of our cameras in two pairs ensured that LMs obtained in each new frame were well distributed across the whole head. Each time the pose of a frame was estimated (see section IIE), any new LMs arising from that frame were added to the database – referenced to the initial pose – and could be used for subsequent pose determination. Therefore, the LM database represented a (growing) sparse model of the head in the initial pose.

E. *Pose estimation*

For any given frame, pose estimation (i.e. the motion with respect to the initial frame) was performed by finding the rotation and translation which minimized the reprojection error between image features and any matched, reprojected database LMs [3]. Minimizing reprojection error has the advantage that even features observed by a single camera can be used, provided they can be matched to a database LM. A pose solution was found using the Gauss-Newton iterative approach described in [7]. Here, instead of finding the pose directly, a small correction to the previous pose was found. We performed ten Gauss-Newton iterations with outlier rejection (based on the magnitude of the reprojection error) to exclude spurious LMs and/or incorrect matches that could bias pose estimation.

F. *Experimental validation*

Validation of the markerless tracking system was performed using a taxidermied rat head phantom, derived from an adult Sprague Dawley rat. The phantom was attached to the end-effector of a 6-axis robot (Epson C3-A601S 6-axis, SEIKO Corp., Japan) and moved according to measured rat head motion [8]. Repeatability of the robot was ±20 μm. The robot executed 5000 movements representing approximately 3 min of real-time head motion. Images from the markerless system were processed to estimate the head motion relative to the initial frame. The motion was repeated with a miniature marker attached to the phantom so we could compare marker-based estimates using the MicronTracker (Claron Technology Inc., Canada). Figure 2 shows a photo of the setup.

Fig. 2. Photo of the setup for validation showing the four cameras arranged in two pairs, and to the right the MicronTracker which was used for marker-based tracking.

III. Results and Discussion

The database grew to approximately 1500 LMs during processing of the 5000 frames. Figure 3 shows the cumulative number of LMs as a function of pose number. Arrows indicate instances of increased rate of new LMs caused by the head moving to a previously unseen orientation. In Fig. 4, a 500-pose segment of the *x*-rotation component of motion is shown. Here the black curve represents the applied robot motion, the red curve the estimated motion using the marker-based system, and the green curve the estimated motion using the markerless system. There was excellent agreement between the applied motion and the markerless system estimates, with far less jitter than in the mark-based estimates. Given that this segment of motion represented the last 500 poses, there was clearly no evidence of increasing drift error associated with inaccurate estimation of LMs over time.

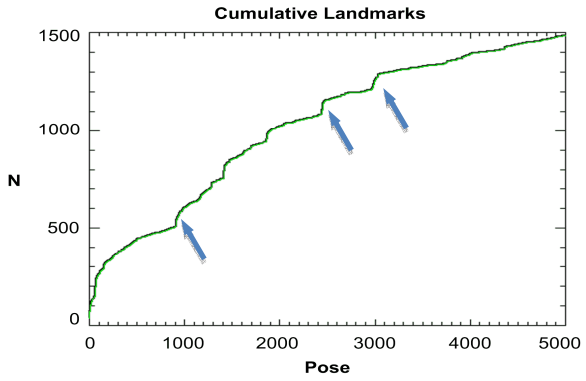

Fig. 3. Cumulative number of landmarks as a function of pose number. The arrows indicate occasions where the rate of new landmarks increased due to the head moving to a previously unseen orientation.

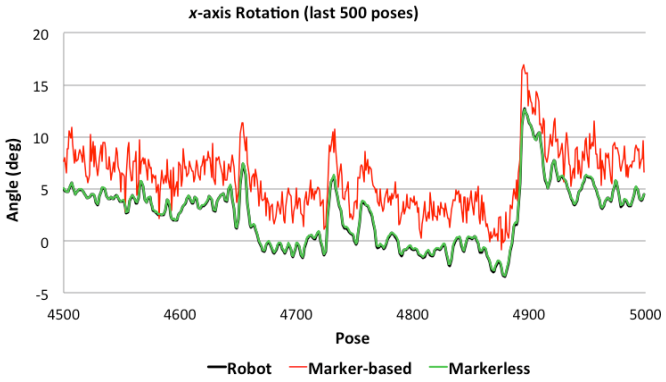

Fig. 4. Last 500 poses of the *x*-axis rotation component of motion, showing the applied robot motion (black), marker-based estimation (red) and markerless estimation (green). Note that at this scale the (green) markerless estimate overlays almost exactly the (black) applied robot motion.

Figure 5a shows the distribution of database landmarks for the marker-based (red dots) and markerless (green dots) estimates relative to the ground truth (black circles). Fig. 5b shows a magnified view of one particular region of the distribution to indicate the mismatch more clearly. There was clear discrepancy between the marker-based estimates and the known locations, but very close agreement for the markerless estimates. The discrepancy across all LMs, computed as the root mean square error, was 2.0 mm for the marker-based system and 0.16 mm for the markerless system. Therefore, the markerless system out-performed the marker-based system, providing considerably more accurate and less noisy motion estimates. We attribute this to the greater spatial distribution of points involved in the pose estimation, and the greater redundancy, as compared to the marker-based method (3 points, closely spaced).

The framework we have proposed potentially makes use of all features detected since the initial frame. In principle, this is much less susceptible to drift errors in pose estimation than frame-to-frame approaches such as [9].

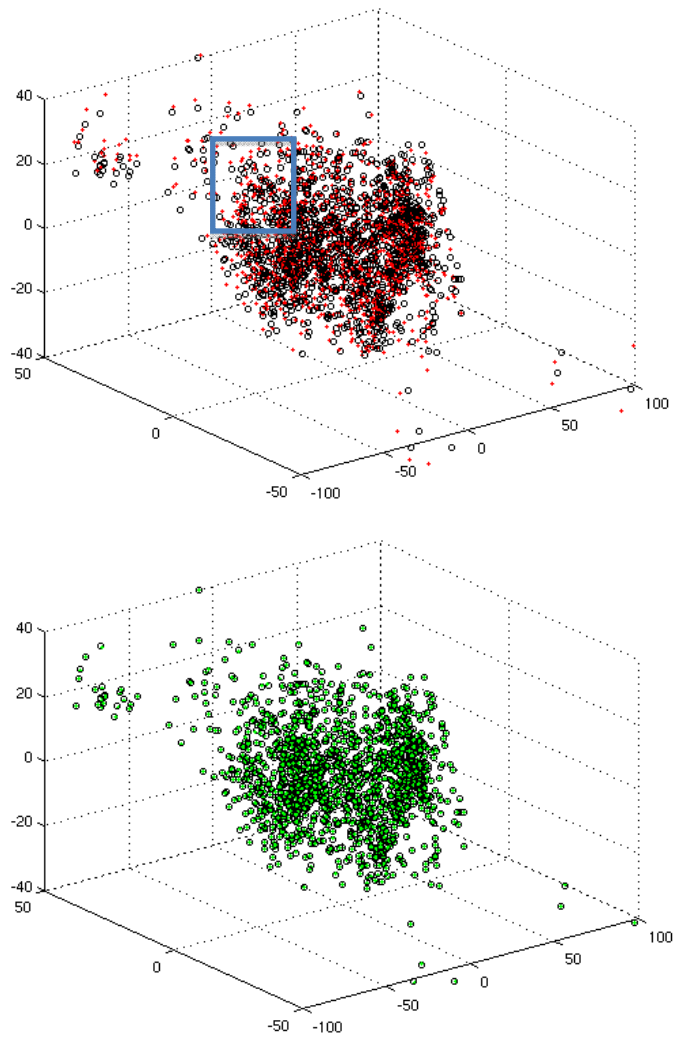

Fig. 5a. Distribution of landmarks compared to ground truth. The top two plots show the complete distribution of landmarks for the marker-based (red) and markerless (green) approaches.

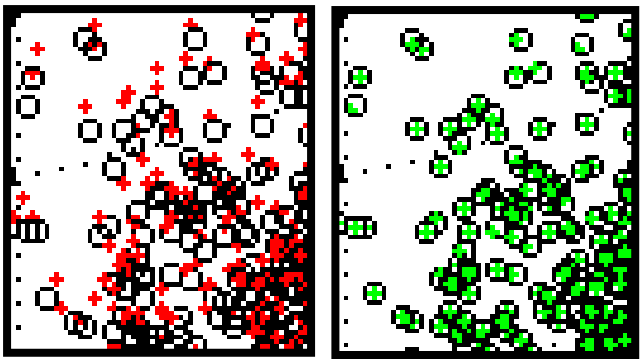

Fig. 5b. Zoomed distribution of landmarks compared to ground truth. The zoomed region corresponds to the blue box in Fig. 5a. Left: degree of mismatch between the marker-based (red) and ground truth landmark locations (black); right: degree of mismatch between the markerless (green) and ground truth landmark locations (black).

Currently processing times are quite slow, with the 5000 pose sequence taking 3.5 hours, 0.5 hours of which was SIFT feature detection. Apart from the parallelization of SIFT feature detection, no attempt was made to optimize the speed of processing. In future work we plan to investigate faster alternatives to SIFT.

The next stage of this work is to perform PET imaging on a live rat in conjunction with markerless tracking and to use these data to generate a motion-compensated reconstruction. Additional challenges are expected for a live rat due to feature detection on whiskers, ears and other non-rigid parts of the head. Outlier rejection methods we have implemented are likely to be particularly important here to avoid matches that would bias the pose estimates. The SLAM framework enables robust statistical filtering to be incorporated so that the location of re-observed LMs can be updated intelligently. This could also play an important role in addressing drift error that might occur over longer sequences than was performed here.

IV. Conclusions

We have described a novel markerless tracking system for tracking the head of a conscious rat, capable of sub-millimeter accuracy. The system outperformed a state-of-the-art marker-based system in preliminary testing involving a taxidermied rat head moved according to measured rat head motion. By using a SLAM framework, our system is potentially more robust to drift errors than other methods. A markerless approach to tracking considerably simplifies conscious animal experiments involving motion tracking as well as enabling tracking over an unlimited range of head motion. Planning of experiments to validate markerless tracking in microPET imaging of awake animals is currently underway.

Acknowledgment

We are grateful to Dr Tomas Svoboda, Dr Andrew Straw and Professor David Lowe for valuable discussions during the course of this work. The work was supported by Australian Research Council Discovery Grants DP0663519 and DP0988166.

References

[1] R. Fulton, S. Meikle, A. Kyme, V. Zhou, K. Popovic, M. Kassiou and M. Akhtar, "Motion-corrected microPET brain imaging of conscious rats," *Proc. 2009 World Molecular Imaging Conference*, Sept 23-26, Montreal, Canada, 2009.

[2] A. Kyme, V. Zhou, S. Meikle, K. Popovic, J.-P. Man, M. Akhtar, I. Karllsson, R. Fulton, "Motion tracking of fully conscious small animals in PET," *Proc. 2009 IEEE Nucl. Sci. Symp. Med. Imaging Conf.*, Florida, USA, pp. 2561-66, 2009.

[3] 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.

[4] T. Svoboda, D. Martinec and T. Pajdla, "A convenient multi-camera self-calibration for virtual environments," *PRESENCE: Teleoperators & Virtual Environments,* vol. 14, pp. 407-422, 2005.

[5] J-Y. Bouguet, *Matlab Calibration Toolbox*. URL: http://www.vision.caltech.edu/bouguetj/calib_doc/

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

[7] D. Lowe, "Robust model-based motion tracking through the integration of search and estimation," *Int. J. Comp. Vis.,* vol. 8, pp. 113-122, 1992.

[8] 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, 2011: e21727. doi:10.1371/journal.pone.0021727.

[9] 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," *Proc. 2009 Nucl. Sci. Symp. & Med. Imaging Conf.*, Orlando, Florida, pp. 2557-2560, 2009.