Intrinsic Feature Pose Measurement for Awake Animal SPECT Imaging

J. S. Goddard, *Member, IEEE*, J. S. Baba, *Member, IEEE*, S. J. Lee, A. G. Weisenberger, A. Stolin, J. McKisson, and M. F. Smith, *Member, IEEE*

***Abstract*– New developments have been made in optical motion tracking for awake animal imaging that measures 3D position and orientation (pose) for a single photon emission computed tomography (SPECT) imaging system. Ongoing SPECT imaging research has been directed towards head motion measurement for brain studies in awake, unrestrained mice. In contrast to previous results using external markers, this work extracts and tracks intrinsic features from multiple camera images and computes relative pose from the tracked features over time. Motion tracking thus far has been limited to measuring extrinsic features such as retro-reflective markers applied to the mouse's head. While this approach has been proven to be accurate, the additional animal handling required to attach the markers is undesirable. A significant improvement in the procedure is achieved by measuring the pose of the head without extrinsic markers using only the external surface appearance. This approach is currently being developed with initial results presented here. The intrinsic features measurement extracts discrete, sparse natural features from 2D images such as eyes, nose, mouth and other visible structures. Stereo correspondence between features for a camera pair is determined for calculation of 3D positions. These features are also tracked over time to provide continuity for surface model fitting. Experimental results from live images are presented.**

I. Introduction

COLLABORATION between Oak Ridge National Laboratory (ORNL), Thomas Jefferson National Accelerator Facility (Jefferson Lab), and the University of Maryland has continued in the development of an advanced high-resolution SPECT instrument to image awake, un-anesthetized small animals. Previous work has been described here [1]-[3]. Animal brain studies are sensitive in that anesthetic agents or physical restraints can affect the results. Functional imaging studies can now be performed on animals without anesthetics, significant restraints, or extensive training. The technology can be extended directly to clinical applications on human patients unable to remain still (e.g. Parkinson's patients, Alzheimer's patients, small children) during a scan.

Manuscript received November 13, 2009. This paper was prepared by the Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37831-6285, operated by UT-Battelle, LLC for the U.S. Department of Energy under contract DE-AC05-00OR22725.

J. S. Goddard, J. S. Baba, S. J. Lee are with Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (telephone: 865-574-9034, e-mail: goddardjsjr @ornl.gov).

A.G. Weisenberger, A. Stolin, J. McKisson are with Jefferson Lab, Newport News, VA, USA, (telephone: 757-269-7090, e-mail: drew @jlab.org).

M. F. Smith. is with University of Maryland, Baltimore, MD 21201 USA (e-mail: msmith7@umm.edu).

The optical measurement and motion tracking system provides 3D position and orientation (pose) during the single photon emission computed tomography (SPECT) imaging scan. The tracking apparatus is an integral part of the scanner and is designed to measure the pose at real-time rates with sub-millimeter accuracy. The pose data is time synchronized for post processing of SPECT list mode data to enable correction of motion blur during 3D image reconstruction.

New developments are continuing in the optical tracking system. In contrast to previous results using external markers [2], intrinsic features are now being extracted and tracked from multiple camera images to compute relative pose over time. Motion tracking thus far has been limited to measuring extrinsic features such as retro-reflective markers applied to a mouse's head. While this approach has been proven to be accurate, the additional animal handling required to attach the markers is undesirable. A significant improvement is achieved by measuring the pose of the head without extrinsic markers using only the normal external surface appearance. This approach is currently being developed with initial results described here. The intrinsic features measurement first extracts discrete, sparse natural features from 2D images such as eyes, nose, mouth and other visible structures. Stereo correspondence between features for a camera pair is then determined for calculation of 3D positions. These features are also tracked over time to provide continuity for surface model fitting.

II. Tracking System

The current tracking system uses optical imaging to view an animal enclosed in a transparent burrow located between the gamma detector heads developed by Jefferson Lab. Up to three cameras along with concentric infrared (IR) ring lights provide multiple views of the animal's head. The lights are synchronized and pulsed to freeze motion while minimizing reflections. The multiple camera image acquisition system acquires images of the awake animal through a range of motions accommodated by the burrow. The camera redundancy significantly reduces loss of tracking where a 3D measurement can be calculated when a marker is visible from only two of three cameras.

Fig. 1 shows the existing SPECT system along with a close-up of the optical tracking cameras and mouse burrow. As shown, three cameras are mounted in the rear of the scanner facing the front of the transparent tubular burrow containing the animal. The cameras are raised above the centerline of the burrow and angled to look down on top of the head of the animal. The LED ring light is mounted around each camera's

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 23,2010 at 15:20:19 EDT from IEEE Xplore. Restrictions apply.

lens. Previously, retro-reflective markers were attached to the animal's head and illuminated by the lights. Measurement software on the tracking computer performs image acquisition, feature segmentation, reflection rejection, stereo correspondence, 3D measurement, and pose calculation. The pose data is also time stamped as is the detector system event data so that list mode SPECT data can be motion compensated as part of reconstruction after the scan is complete.

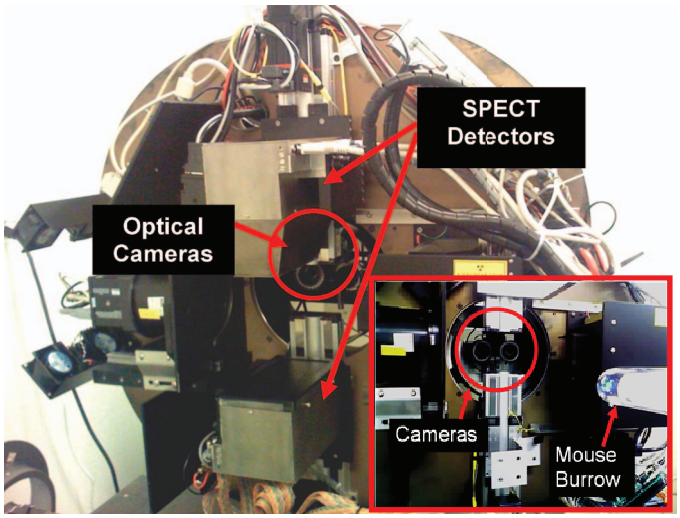

Fig. 1 Front view of scanner is illustrated showing tracking system hardware and configuration with respect to SPECT gamma detectors. Close up shows the cameras, illumination hardware, and mouse burrow.

III. Intrinsic Feature Method

The intrinsic feature extraction approach for this work is given in Fig. 2. Images from two calibrated cameras are first rectified in pairs to reduce the correspondence search computation. Next the salient features on the head of a mouse without external markers are extracted from each image using well-known feature extraction methods such as Harris [5] and Shi-Tomasi [9]. These features are then tracked from frame to frame for each camera. Correspondence matching of features between stereo images is performed with filtering to remove reflections and extraneous points. This correspondence is determined by normalized correlation or other matching methods for efficient computation. From these stereo matches, an optimal triangulation method is used for accurate 3D position calculation [4]. Tracking the 3D features over time using robust statistical methods enables the determination and fit to a local model of the animal's head between consecutive images. Pose is calculated as the 3D translation and rotation needed to align the current measurement relative to the model.

Each camera is initially calibrated offline to determine both intrinsic calibration parameters (focal length, camera center, and lens distortion) and extrinsic calibration parameters (relative coordinate transformations between the cameras).

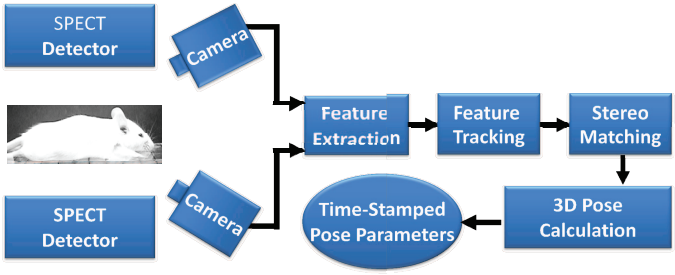

Fig. 2 Tracking system functional block diagram is shown for intrinsic feature motion tracking.

A. Calibration

Intrinsic and extrinsic offline calibration for camera parameters and SPECT detector calibration have been previously described [2]. The calibration procedure combines both calibration steps in calculating the intrinsic and extrinsic parameters of a camera and lens model. The extrinsic parameters determine the 3D position and orientation of the cameras with respect to a common reference frame. A sequence of images of a checkerboard calibration pattern is taken simultaneously from all cameras. The pattern is moved to different positions and orientations between images. Corners are automatically extracted followed by a nonlinear optimization to obtain the minimum least square error fit of the parameters to the model. Typical calibration errors in terms of the difference between actual and projected 2D image point corner locations are less than 0.25 pixel RMS.

B. Feature Detection and Tracking

Point features are extracted and tracked over time within the image sequence to identify features for each image pair that show the projected 2D image movement between consecutive images. The live mouse images shown in Fig. 3 give the tracks for two consecutive stereo pairs. Tracks are determined independently for each camera image without regard to the stereo pair and features are not always tracked through multiple frames. A variety of methods for feature extraction and tracking are being evaluated including Harris corner point, Shi-Tomasi "Good features to track", Lucas-Kanade optical flow, SIFT, and SURF [5]-[9]. The features given in Fig. 3 are extracted using the Shi-Tomasi method and tracked with LK optical flow. Tracked feature points are not required to be tracked across more than two sequential images.

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 23,2010 at 15:20:19 EDT from IEEE Xplore. Restrictions apply.

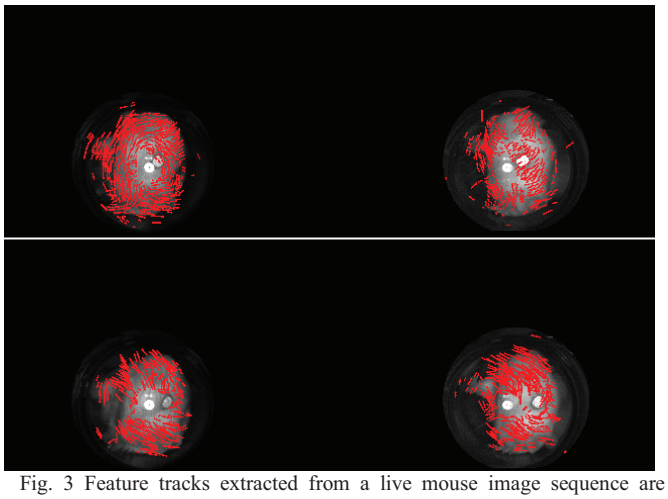

Fig. 3 Feature tracks extracted from a live mouse image sequence are shown for left and right stereo pairs. The tracked features are shown as lines connecting matching points from consecutive images. Features are extracted using the Shi-Tomasi method and tracked using Lucas-Kanade optical flow.

C. Stereo Correspondence

Correspondence of features between stereo image pairs is performed next along with filtering to remove reflections and extraneous points not located on the head. This correspondence includes normalized correlation or other matching methods for efficient computation. Corresponding point locations are determined to sub-pixel accuracy. Prior rectification of the images ensures that epipolar lines are horizontal so that matching features have the same y-coordinate. The differences in the x-coordinate determine the disparity which corresponds to range or depth from the reference image. Fig. 4 shows a rectified stereo image pair with matched points determined using normalized correlation. Example matching points are shown connected by horizontal lines. The result of this step is a set of matched (x,y) 2D image points from the left and right images.

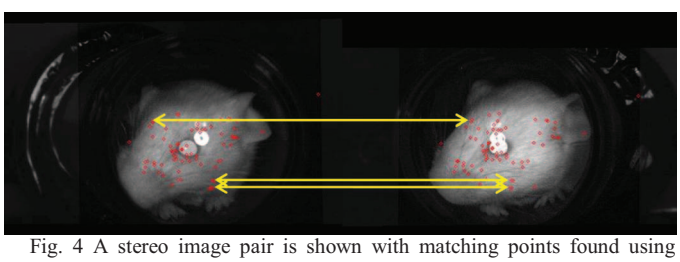

Fig. 4 A stereo image pair is shown with matching points found using normalized correlation. The images have been rectified and horizontal lines are shown connecting several points between the two images. The bright spot near the center of each image is a reflection of the illumination off the glass tube.

D. 3D Measurement

From each feature match set, an optimal triangulation method is used for accurate 3D position calculation [4]. In the case of rectified images, the optimal triangulation coordinates are the reference x-coordinate and the mean of the left and right y-coordinates. Each matched point set is described by a vector as given by the following (equation 1):

$$\begin{bmatrix} x \\ y \\ d \end{bmatrix} = \begin{bmatrix} x_R \\ \frac{(y_R - y_L)}{2} \\ x_R - x_L \end{bmatrix} \tag{1}$$

where x_R, y_R, x_L, y_L are the image coordinates of the right and left camera respectively. The variable d is the image disparity. Each disparity vector can then be used to calculate the homogeneous 3D coordinates vector given camera calibration parameters (equation 2).

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & -\frac{1}{T_x} & \frac{(c_x - c'_x)}{T_x} \end{bmatrix} \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix}, \tag{2}$$

where c_x, c_y, are the optical center coordinates of the right camera and c'_x is the optical center x-coordinate of the left camera. $\mathbf{T_x}$ is the x-axis separation distance between cameras and f is the focal length which is the same for both cameras. Dividing through by **W** gives the absolute 3D coordinates.

The calculated 3D points for a set of live mouse stereo images are shown in Fig. 5. These points correspond to the tracked matched points for one image pair.

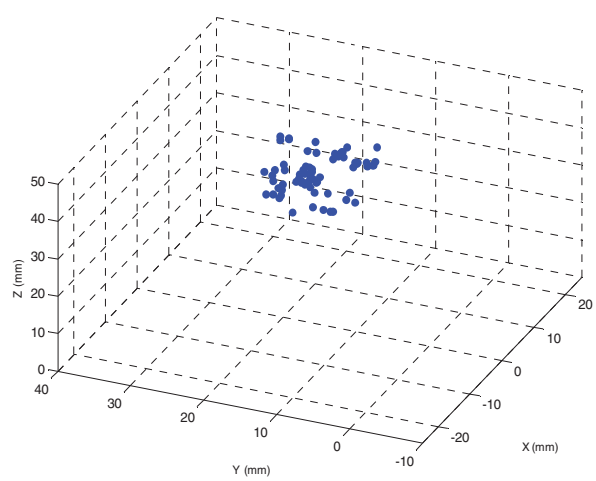

Fig. 5 3D measurements are plotted for corresponding features of mouse head from live mouse images shown in Fig. 4. The plot of extracted 3D feature locations is relative to right camera reference frame.

E. Pose Calculation

Tracking the 3D features over time using robust statistical methods enables the determination and fit to the partial head model from the previous point set. Pose is calculated as the 3D translation and rotation needed to align the current measurement relative to this rigid body model. Tracking of points through sequential images provides continuity in the pose measurement through calculation of the 3D transformation between sequential image pairs. At least 3

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 23,2010 at 15:20:19 EDT from IEEE Xplore. Restrictions apply.

point tracks are needed in order to compute the transformation. Fig. 6 illustrates the calculation of the transformation. From two sets of 3D points the rotation and translation is calculated using Horn's method [10]. This method is applicable to any number of points greater than 2.

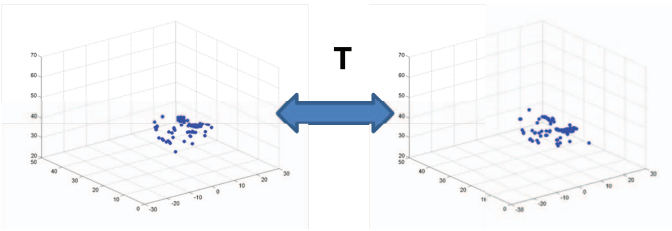

Fig. 6 The 3D points from each of two consecutive images are shown. These two sets of points are used to calculate the 3D transformation.

IV. Live Mouse Results

Live mouse image sequences without markers have been acquired on the existing SPECT system for development and testing. Software has been developed to rectify the image pairs and automatically extract point features, track features between image frames, perform feature correspondence, stereo matching, and calculate 3D locations relative to a fixed reference frame. The initial results to date are shown in Fig. 7. The pictures with coordinate axes superimposed show pose changes between consecutive images using detected and matched features from the right and left cameras. Matching features are tracked from frame to frame to automatically fit a local head model. Pose is estimated as the optimum fit to this model at each image frame.

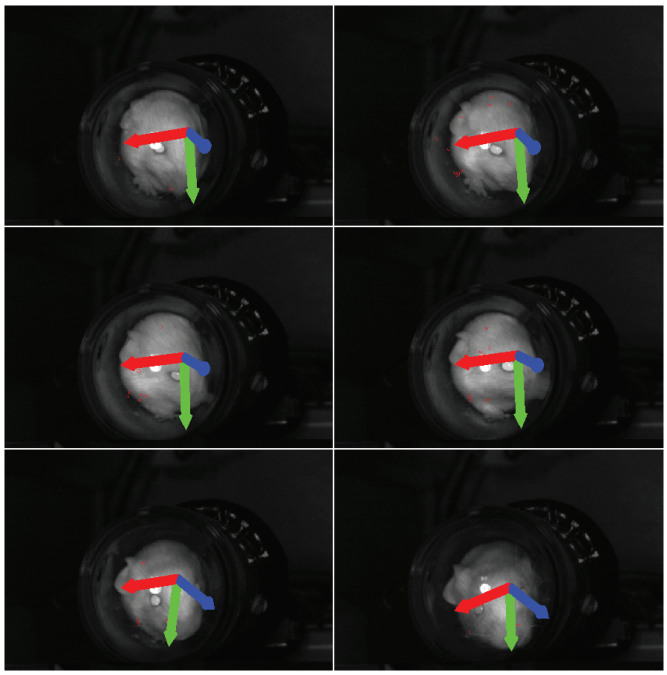

Fig. 7 Calculated pose coordinate axes are shown superimposed on an image sequence. The image sequence is shown left to right and top to bottom. The axes follow the head motion through the sequence. As previously noted, the bright spot seen in some of the images is a tube reflection that remains stationary.

V. Summary

Initial results for real-time motion tracking of intrinsic features for an awake animal SPECT scanner have been described. Thus far the method has been developed to calculate relative pose from image to image with robustness to mismatches and noise. Test results have been provided to demonstrate the functionality of pose measurement. This system is fully integrated within a commercial scanner and has been demonstrated to reliably measure head position and orientation of a mouse through multiple image sequences. Future plans are to continue studies with live mice concentrating on further improving reliability and accuracy for motion tracking. Validation of this method against markers is planned to assess performance in terms of accuracy, reliability, and processing requirements. Further evaluation is needed to optimize feature extraction, tracking, and robust matching methods.

References

[1] A. G. Weisenberger, B. Kross, S. Majewski, V. Popov, M. F. Smith, V. H. Tran, B. Welch J. Baba, J. Goddard, M. Pomper, and B. Tsui, "Instrumentation Development of a SPECT-CT System to Image Awake Mice," Proc. IEEE Medical Imaging Conference, San Diego, November 2006.

[2] S. S. Gleason, J. S. Goddard, M. J. Paulus, J. S. Baba, S. Majewski, M. Smith, T. Tran, A. Weisenberger, B. Welch, R. Wojcik "Real-time, high-accuracy 3D tracking of small animals for motion-corrected SPECT imaging," Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference, Rome, Italy, October, 2004.

[3] J.S. Goddard, J,S. Baba, A. G. Weisenberger, and M. F. Smith, "Improved Pose Measurement and Tracking System for Motion Correction of Awake, Unrestrained Small Animal SPECT Imaging" *Conference Record of the 2007 IEEE Nuclear Science Symposium and Medical Imaging Conference*, Honolulu, 2007.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge: Cambridge University Press, 2000.

[5] H. H. Baker and T. O. Binford. "Depth from edge and intensity based stereo," Proceedings 7th Int. Joint Conf. on Artificial Intelligence, pp 631--636, Vancouver, BC, Canada, August 1981.

[6] H. Bay, T. Tuytelaars and L. Van Gool "SURF: Speeded Up Robust Features", Proceedings of the 9th European Conference on Computer Vision, Springer LNCS volume 3951, part 1, pp 404--417, 2006.

[7] Lucas, B., and Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision, Proc. of 7th International Joint Conference on Artificial Intelligence (IJCAI), pp. 674-679.

[8] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* 60 (2004): 91-110.

[9] Jianbo Shi and Carlo Tomasi, "Good Features to Track," *1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, 1994, pp. 593 - 600.

[10] B. Horn, "Closed-form solution of absolute orientation using unit quaternions," Journal of the Optical Society of America, 4, pp. 629-642, April 1987.

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 23,2010 at 15:20:19 EDT from IEEE Xplore. Restrictions apply.