An Optimized Feature Detector for Markerless Motion Tracking in Motion-Compensated Neuroimaging

David Henry, *Student Member, IEEE*, Yidi Yao, Roger Fulton, *Senior Member, IEEE*, and

Andre Kyme, *Member, IEEE*

***Abstract*–Head movements during PET and MRI scans can have a detrimental effect on image quality and quantitative measurements. For both modalities, motion correction methods exist that rely on accurate characterization of head motion. In the case of prospective correction in MRI, the motion estimates also need to be delivered in real-time. Motion tracking methods that rely on attached markers are susceptible to decoupling of the head and marker, hinder clinical workflow, and have line of sight issues due to the geometry of the bore and headcoil. In this study, we aim to optimize a methodology that measures head motion by detecting and tracking SIFT features native to the forehead. These features can be extracted and described in many ways, with different algorithms offering varying levels of computational efficiency and robustness to scene changes. A phantom study was performed to assess the accuracy and speed performance of five different feature detectors: SIFT, SURF, ORB, BRISK and AKAZE. Except for ORB, position estimates obtained using the different feature detectors showed similar agreement (error <0.4 mm) with the ground-truth robot measurements. Processing time varied, with SURF, BRISK and AKAZE offering a substantial speed increase over SIFT while maintaining similar accuracy. We conclude that SURF, BRISK and AKAZE appear to be suitable alternative feature detectors to SIFT for prospective motion correction in MRI and MRI-PET.**

I. Introduction

NEUROIMAGING with PET and MRI are particularly susceptible to patient head motion due to long scan times. Motion can lead to severe artifacts in the reconstructed image, thereby reducing its diagnostic value and the quantitative accuracy of MR and PET derived measurements. To correct for these motion effects, accurate characterization of the motion is required.

Optical cameras in conjunction with computer vision algorithms are an effective way of quantifying in-bore head motion for both MRI and PET brain studies [1],[2]. This approach can provide high frequency head pose updates without affecting the image data acquisition process. In particular, it seems to be a useful choice for prospective motion correction (MC) in MRI [3]. Here, magnetic field gradients and radiofrequency (RF) waveforms are adjusted in real-time to move the imaging volume in response to motion. Prospective MC has the potential to mitigate motion artifacts that are difficult to deal with retrospectively, such as spin history effects and large rotational motions [4]. To do this without introducing further artifacts, motion estimates must be provided rapidly and accurately.

Challenges exist in adapting camera-guided prospective MC to a clinical setting. Current methodologies involve attaching markers to the head, which can become detached or exhibit motion relative to the head, leading to erroneous motion estimates. Marker attachment in general hinders clinical workflow, and is quite problematic in the case of uncooperative patients. Fitting cameras into the tight bore geometry and getting good line of sight to attached markers through the headcoil is also a major issue.

Recently, we proposed a markerless motion tracking approach for hybrid MR imaging which involves tracking small feature patches on the forehead [5]. Tracking features from images is common in computer vision applications such as object recognition, with SIFT features [6] being a notable example. Other feature detection and descriptions algorithms exist which can offer drastic improvements in processing speed, of interest for real-time applications such as prospective MC. The aim of this study was to compare the performance of different feature detectors (SIFT, SURF [7], ORB [8], BRISK [9] and AKAZE [10]) in a motion tracking context and to find the feature detector that is most suited to this application. We also investigated optimizing stereo camera geometry to maximize the field of view overlap, thus maximizing the number of feature correspondences between the two stereo views.

II. Methods

A. Simulating Head Motion Tracking in MRI

Two 640x480 resolution CCD cameras (Flea2, Point Grey Research) were mounted on an optical table with a baseline distance of 6.5 cm. High resolution photographs of human foreheads printed life-size on photographic paper were attached to the end-effector of a 6-axis robot arm (C3-A601ST, Epson America Inc., positional reproducibility 20 μm), positioned at approximately 120 mm from the two

Manuscript received November 7, 2017.

David Henry and Roger Fulton are with the Brain and Mind Centre, University of Sydney, Australia (telephone: +61 410 093 382, e-mail: dhen2714@uni.sydney.edu.au, roger.fulton@sydney.edu.au).

Yidi Yao is with Tsinghua University, Beijing, China

Andre Kyme is with Biomedical Engineering, School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, Australia (e-mail andre.kyme@sydney.edu.au).

978-1-5386-2282-7/17/$31.00 ©2017 IEEE

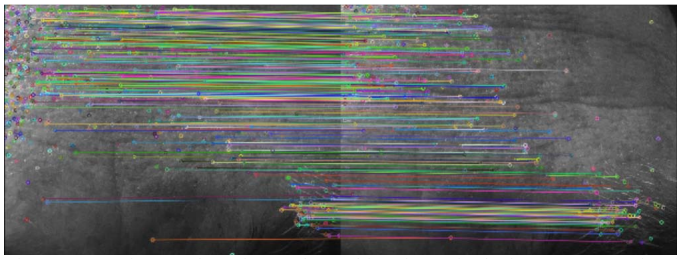

Fig. 1. SIFT feature detection and matching for a single frame stereo pair of the forehead. For both the left and right image, features (colored circles) were detected and then cross matched (colored lines). From a series of these cross matches, the pose of the forehead phantom for the current frame could be determined.

cameras. Photographs were obtained with and without an ink based feature patch stamped above the eyebrow. Adding a feature-rich stamp was found previously to amplify the number of detected features without the problem of decoupling from the forehead like traditional attached markers [5]. Intrinsic and extrinsic calibration of the stereo camera setup was performed using a checkerboard pattern of 4 mm squares, moved to 30 different positions by the robot arm. The calibration data were processed using the Matlab Calibration Toolbox [11].

B. Feature Detectors

We examined accuracy and speed of five different feature detectors (Table I), both aspects of performance are critical in this application. Each feature detector works across multiple scales or resolutions via image processing, for example, by forming an image pyramid, which involves successive Gaussian smoothing and downsampling. Features are identified in scale space as extrema of the response to some function such as the Laplacian, determinant of Hessian or the FAST algorithm [9]. Once detected and located, features are labelled with a descriptor. The feature detectors were implemented with Python OpenCV algorithms. An example of detection and matching of SIFT features for a pair of stereo images is shown in Fig. 1.

TABLE I. FEATURE DETECTOR ALGORITHMS

	SIFT	SURF	ORB	BRISK	AKAZE
Scale Space	Image Pyramid	Box filter convolution	Image Pyramid	Image Pyramid	Diffusion filtering
Detection	Laplacian	Determinant of Hessian	Intensity threshold	Intensity threshold	Determinant of Hessian
Descriptor	Integer vector	Real vector	Binary string	Binary string	Binary string
Descriptor Extraction	Orientation histogram	Wavelet response	Intensity comparison	Intensity comparison	Intensity, derivative comparison

C. Data Acquisition and pre-processing

For each set of photos, the forehead phantom was moved to 30 different, discrete poses, and images were acquired at each step. Pose estimates (section 2D) derived using the feature detectors were reported in the camera coordinate frame, whereas the ground-truth (robot-arm) poses were in the robot's coordinate frame. In order to compare these, a cross-calibration was performed by moving a marker on the end-effector to 20 positions in the camera field of view. The center of the marker was located in the camera and robot frames and the transformation relating the two coordinate systems was obtained using absolute orientation [12].

D. Pose Estimation

Pose estimates for each movement were obtained using a similar feature-based algorithm previously described in [13]. For each of the five feature detectors in Table I, keypoints detected by each camera within each frame were matched and then triangulated. These triangulated landmarks were compared to landmarks triangulated in previous frames, such that the pose change between frames could be acquired using absolute orientation [12].

Motion estimates were compared to the ground-truth robot motions. Several test points located 10 cm behind the forehead were spatially transformed using both the known robot motions and the markerless motion estimates. The average Euclidean distance between the two sets of points for each pose was used to assess accuracy of the motion estimates.

E. Geometric Optimization

A single camera was attached to the end effector of the same robot arm used previously. In the camera's field of view was a grid, with each grid point being labelled with a number. The distance between the camera and grid was 6 cm, and a picture of the pattern was taken at two points horizontally separated by baselines of 2, 4, 6, and 8 cm, thus simulating a stereo camera setup. For each of these baseline distances, geometries were also simulated where both cameras were tilted towards each other around the vertical axis by up to 30˚, with 0˚ being where the central projections of the cameras were parallel. This process repeated for camera-grid distances of 8, 10, 12 and 14 cm.

The field of view for each picture taken could be quantified by analyzing the portion of the grid visible. This could then be used to determine the amount of relative overlap for a specific stereo geometry by pairing it with another picture. Relative overlap was also calculated geometrically using the formula for angle of view (α):

$$\alpha = 2arctan\frac{d}{2f} \quad (1)$$

where d is the known sensor size and f is effective focal length. Eq. (1) is valid when lens-to-object distance is much larger than focal length.

III. RESULTS AND DISCUSSION

The average spatial discrepancy (in mm) due to differences in the ground-truth and estimated motions for the test points is shown for a forehead image (without stamp) moved to different poses in Fig. 2. Positional discrepancies for pose estimates derived using SIFT, SURF, BRISK and AKAZE were similar, <0.4 mm. For the ORB derived measurements, discrepancies were higher, but still sub-millimeter. Fig. 3 compares the distribution of test point discrepancies across poses for pose estimates derived using the different feature detectors for a forehead image without a stamp, with a small stamp, and with a large stamp. A similar spread in discrepancies can again be seen here for SIFT, SURF, BRISK and AKAZE. Given the number of features we could detect and triangulated in each frame (a few hundred), the addition of

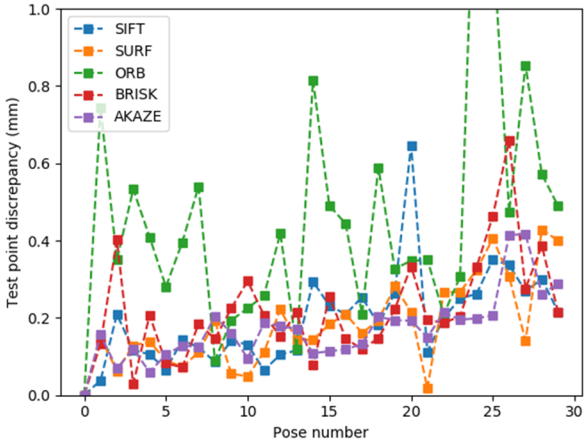

Fig. 2. Test point discrepancies for a single dataset. For each robot pose, the mean Euclidean distance between points transformed by robot motions (ground truth) and points transformed by the feature-based motion estimates was calculated.

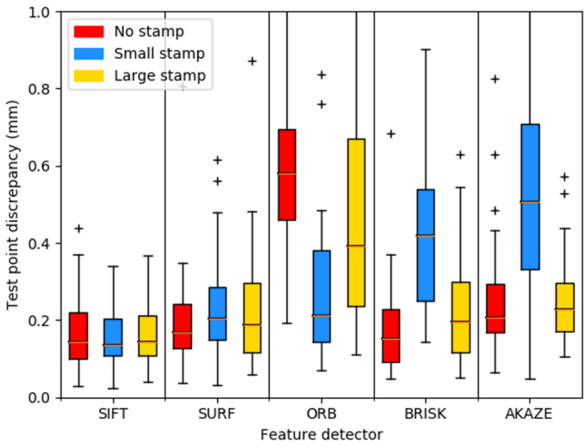

Fig. 3. Boxplot comparisons of test point discrepancy distributions across poses for a forehead image without stamp, with a small stamp, and with a large stamp and for different feature detectors. Crosses are outliers.

a feature-rich patch appears does not appear to have a measurable positive effect in our results in Fig. 3.

The average time per frame for feature detection and extraction for each feature detector is shown in orange in Fig. 4. The average total time (blue bars in Fig. 4) includes the time taken to detect, extract, match and triangulate features, and to perform pose estimation. SIFT, although accurate in detecting forehead features, is slower than the other feature detectors by a factor of ~3 in total frame processing time.

Given the speed advantage offered by the other feature detectors, they appear better suited to feature-based tracking for prospective MC in MRI and hybrid PET-MR imaging. This is especially true for SURF, BRISK and AKAZE, as the accuracy of pose measurements obtained using these detectors was similar to that of SIFT. Indeed, from Fig. 2, the positional discrepancies for SIFT, SURF, BRISK and AKAZE follow a very similar trajectory. This could suggest a limiting baseline error due to inaccuracies triangulation or cross-calibration. If this is correct, the potential accuracy of the feature detectors is better than what we report here.

For MRI, the tight geometry of the headcoil means that head movement is quite limited. In this case, cameras fixed on the forehead will not be tracking large changes in perspective.

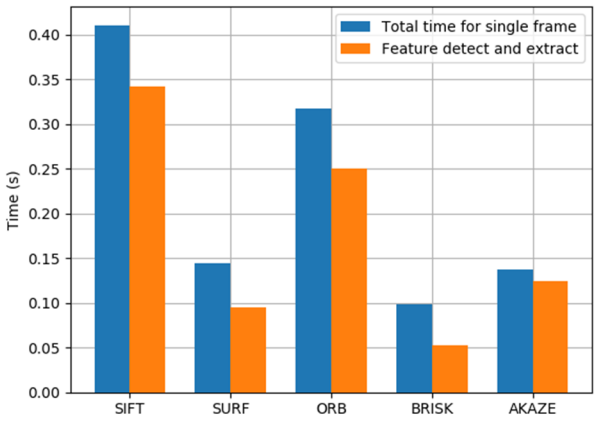

Fig. 4. Average time taken to process a single frame and the time taken to detect and extract features.

Thus, for markerless tracking in this context, the computational overhead required for scale invariance may not be necessary and ignoring this aspect should lead to further improvements in efficiency.

In our geometric optimization, we found that our experimental values for stereo overlap matched closely with simulations using Eq. (1). These data will thus aid in positioning of cameras given baselines and working distances that are imposed by the constraints of the headcoil.

IV. CONCLUSIONS AND FUTURE WORK

We have demonstrated the feasibility of accurate, markerless, motion tracking of a forehead phantom mimicking the real scenario using a variety of different feature detectors. Markerless motion tracking estimates using SIFT, SURF, BRISK and AKAZE were in close agreement with ground-truth motion. Choosing the right feature detector can offer large improvements in processing time, with processing times for SIFT being about three times longer on average than the others. This is of great importance for prospective MC in MRI, where motion estimates are required in real-time. In the future, we aim to further compare these feature detectors in tracking of human subjects, and perform in-bore experiments with optimized stereo camera geometry to minimize triangulation errors.

REFERENCES

[1] J. Maclaren et al., "Measurement and correction of Microscopic head motion during magnetic resonance imaging of the brain," PLoS One, vol. 7, no. 11, pp. 3-11, 2012.

[2] O. Olesen et al., "List-Mode PET motion correction using markerless head tracking: proof-of-concept with scans of human subject," IEEE Transactions on Medical Imaging, vol. 32, no. 2, pp. 200-209, 2013.

[3] M. Zaitsev, C. Dold, G. Sakas, J. Hennig, and O. Speck, "Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system," NeuroImage, vol. 31, no. 3, pp. 1038-1050, 2006.

[4] J. Maclaren, M. Herbst, O. Speck, and M. Zaitsev, "Prospective motion correction in brain imaging: a review," Magnetic Resonance in Medicine, vol. 69, no. 3, pp. 621-636, 2013.

[5] A. Kyme, J. Maclaren, M. Aksoy, and R. Bammer, "Feasibility of marker-free motion tracking for motion-corrected MRI and PET-MRI," IEEE Nuclear Science Symposium, Medical Imaging Conference, 2016.

[6] D. Lowe, “Distinctive image features from scale invariant keypoints,” International Journal of Computer Vision, vol.60, pp. 91-110, 2004.
[7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-Up robust features (SURF),” Computer Vision and Image Understanding, vol.110, no. 3, pp. 346-359, 2008.
[8] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: an efficient alternative to SIFT or SURF,” Proceedings of the IEEE International Conference on Computer Vision, pp. 2564-2571, 2011.
[9] S. Leutenegger, M. Chli, and R. Siegwart, “BRISK: binary robust invariant scalable keypoints,” Proceedings of the IEEE International Conference on Computer Vision, pp. 2548-2555, 2011.
[10] P. Alcantarilla, J. Nuevo, and A. Bartoli, “Fast explicit diffusion for accelerated features in nonlinear scale spaces,” Proceedings of the British Machine Vision Conference 2013, 2013.
[11] J. Bouguet, “Camera calibration toolbox for Matlab,” 2008. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/
[12] B. Horn, “Closed-form solution of absolute orientation using unit quaternions,” Journal of the Optical Society of America: Optics and Image Science, and Vision, vol. 4, no. 4, pp. 629-642, 1987.
[13] A. Kyme et al., “Markerless motion tracking of awake animals in positron emission tomography,” IEEE Transactions on Medical Imaging, vol. 33, no. 11, pp. 2180-2190, 2014.