Non-rigid Motion Detection for Motion Tracking of the Head

David Henry, *Student Member, IEEE*, Roger Fulton, *Senior Member, IEEE*, Julian Maclaren, Murat Aksoy, Roland Bammer, and Andre Kyme, *Member, IEEE*

***Abstract*—Optical motion tracking systems are effective tools for measuring head motion during MRI and PET scans in order to correct for motion. Most systems rely on the attachment of fiducial markers which can slip or become decoupled from the head, causing erroneous motion estimates which can introduce further image artifacts. In this work, we investigated two methods of detecting non-rigid motion which can be easily incorporated into a feature-based motion tracking system using optical stereo. The tracking system tracks detected features on small patches of the forehead, which remain completely coupled to the skin surface. By monitoring these features, surface deformations on parts of the face that deform non-rigidly with respect to the rest of the head can be detected and potentially characterised. We investigated two methods of detecting non-rigid deformations: one involved measuring distances between detected landmarks and comparing these distances to previous frames; the other used a neural network to classify a group of landmarks as either 'rigid' or 'non-rigid'. A simulation tool was also developed to aid in the characterization of non-rigid motion and its effects. Landmark distance discrepancies were found to be closely correlated to errors in the feature based motion tracking system, suggesting it is a useful metric for detecting non-rigid motion. The trained neural network was able to classify a collection of landmarks as 'rigid' with 99.8 % accuracy and classified the 'non-rigid' case with 93.3% accuracy.**

I. Introduction

OPTICAL motion tracking systems are effective tools for measuring head-motion during medical imaging procedures to correct for patient motion, a well-known source of image artifacts in MRI and PET [1], [2]. However, most of these systems rely on attaching fiducial markers to the head which can slip or become decoupled. This decoupling is harmful as it cannot be easily detected or characterized, and can introduce further image artifacts, as measured motion does not reflect the motion of the brain.

We have recently proposed a motion tracking approach which tracks detected features on a small patches on the forehead using optical stereo [3]. Detected features could include native skin features, or artificial features added to the skin with a temporary tattoo or an ink stamp. Tracked targets therefore remain completely coupled to the skin surface, and do not affect patient comfort. By monitoring targets coupled to the skin surface, surface deformations on parts of the face that are non-rigid with respect to the rest of the head can be detected and potentially characterised. This would allow the identification of corrupted rigid-motion estimates, which could then be discarded, or used to reacquire data in the case of prospective MRI motion correction.

Previously, the feature-based algorithm has been validated on human volunteers undergoing continuous, 6-dof (degrees of freedom) rigid-body head motion [4] and on volunteers with their head in discrete positions inside an MRI scanner [3]. In this work, we investigate two methods of detecting non-rigid deformations which can be easily incorporated into the feature based motion tracking algorithm: one involving monitoring the distances between detected landmarks on the forehead, and the other using a neural network to classify a collection of landmarks as either 'rigid' or 'non-rigid'. We have also developed a simulation tool which will aid in the characterization of non-rigid motion and its effects.

II. Methods

A. Simulation

The principles of the feature-based algorithm are described in detail in [5]. For each frame, distinctive image points are detected using a feature detector algorithm. Image points are matched across the views of a stereo camera system and their 3D positions triangulated. Over time, a database of these 3D landmarks is formed which forms a sparse point cloud representation of the object being tracked - in our case the forehead. For subsequent frames, 2D image keypoints are matched with 3D landmarks in the database, and a 2D-3D registration is used to calculate the 6-dof pose of the head in the current frame.

For simulation, a single, known rigid-body transformation was applied to a sparse 3D point cloud at each simulated frame. The transformed point cloud was then projected into two different 2D pixel spaces representing the two views of a virtual stereo camera system using two known projection matrices defining the stereo camera geometry, and camera parameters such as the focal length. Once projected into pixel space, Gaussian noise was added and the feature-based algorithm was used to derive an estimate of the current pose of the point cloud.

Non-rigid motion was modelled by applying small (approx. 2mm) movements to half the points in the point cloud. The simulation was run for 1000 frames, with the non-rigid deformations being applied in frames 100-150. Simulated data was compared to instances of known non-rigid motion.

B. Distances between landmarks

Motion data from a previous study was used [4]. Fourteen volunteers wore a tightly fitted fabric cap with 5 IR (infrared) reflective markers tracked by the OptiTrack motion tracking system (NaturalPoint, Inc). A temporary tattoo was applied to their foreheads, which was viewed by a stereo camera system (HobbitVew, Inc.). OptiTrack motions traces were used as the ground truth. Volunteers were instructed to perform slow, low amplitude motion for 2 minutes, with brief non-rigid forehead movements at 30s, 60s and 90s.

The feature based algorithm was applied to images from the stereo camera system, estimating volunteer motions from detected features in each of the stereo views. For each frame, as well as performing motion tracking, the distance between triangulated landmarks were calculated. In the case of rigid-body motion only, it is expected that the mutual distances between observed landmarks will remain constant, whereas under non-rigid motion, they would change. To detect non-rigid motion at each frame, mutual landmark distances in the current frame were compared with their corresponding distances in the landmark database. The difference between these corresponding distances in mm was calculated, and

the median of these differences was calculated for each frame. Motion estimates from the feature-based algorithm and motion measurements for the OptiTrack were applied separately to a test point in each frame. The Euclidean distance between these transformed points was therefore proportional to the discrepancy between our feature-based motion estimates and the ground-truth OptiTrack motions. The median landmark distance difference was then compared with the Euclidean error at each frame.

C. PointNet

A single volunteer was instructed to perform low amplitude rigid-body only head motion for 200s, followed by 200s of head motion while continuously deforming the forehead. This was captured by the stereo camera system (100 fps), with detected features being triangulated in each frame, forming 40000 point clouds (1 for each frame). These point clouds were labelled as 'non-rigid' if they had been acquired when forehead deformations were being performed, or 'rigid' if they had been acquired with rigid-body only head motion. For uniformity, each point cloud was then reduced to 50 randomly sampled points. Of these point clouds, 30000 (50-50 split between 'rigid' and 'non-rigid' clouds) were fed into PointNet [6], a neural network architecture specifically designed to classify sparse point clouds. The rest of the data was used for evaluation.

III. Results and Discussion

Results for the simulation in which non-rigid deformations were applied to the 3D point cloud at frames 100-150 are shown in Fig. 1. The z-component of the controlled, rigid-body input point cloud transformations is plotted in black. The feature-based algorithm pose estimates are plotted in green. Deviation of the simulated estimates from the ground truth can be clearly seen in frames 100-150, where non-rigid deformations were applied to the point cloud.

Fig. 2 shows the z-translational motion of one of the volunteers undergoing intermittent non-rigid motion at 30, 60 and 90s. The ground truth OptiTrack motions are in black and the feautre-based algorithm estimates are in green. The deviation from ground truth is seen at three points in Fig. 2 corresponding to when non-rigid motion occurred. Our simulated results from Fig. 1 showed similar behaviour when non-rigid motion was modelled. The large deviations between ground truth and feature-based motions are therefore very likely due to non-rigid motions.

Median landmark distance differences between un-deformed 3D landmark database and triangulated landmarks in each frame are shown in Fig. 3 in red for a single volunteer take. This is overlaid the Euclidean distance error (green). The initial rise in distance difference close to frame 3000 (when the volunteer performed non-rigid motion) corresponds almost exactly to increase in Euclidean distance error. This suggests the first derivative of median distance difference calculated at each frame could be a useful metric to detect non-rigid motion. Also of interest is that the distance difference does not return to its baseline level after the non-rigid motion has ceased, suggesting that the feature based tattoo used on the volunteers slightly deforms in an irreversible way after non-rigid motion.

Training of PointNet on our training data (30000 point clouds) took approximately 1.5 hours on a single NVIDIA V100 GPU. Classification accuracy of 'rigid' point clouds was 99.8%, and classification accuracy of 'non-rigid' point clouds was 93.3%. It is important to note that the only input to the network is raw point cloud data; no temporal information was included. In reality, we expect 'non-rigid' point clouds to be observed frame after frame, so modifying the network architecture to include some form of recurrence could see improved results.

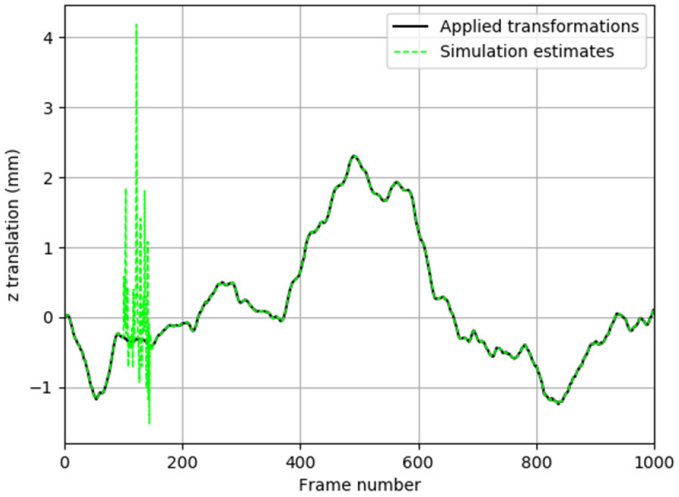

Fig. 1. z translation comparing applied transformations (black) on point cloud, with motion estimates from our algorithm simulation (green). Non-rigid deformations were applied in frames 100-150, resulting in discrepancies between motions estimates and ground truth.

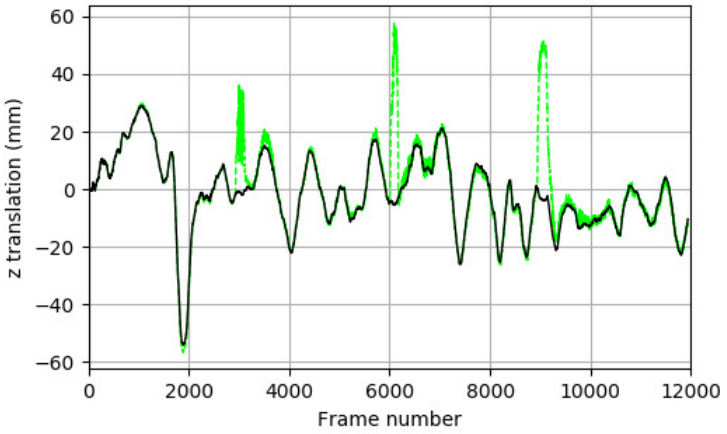

Fig. 2. z (depth) translation for a single volunteer who was instructed to perform non-rigid motions at 30s, 60s and 90s. Ground truth OptiTrack motions are plotted in black, feature based motions estimates are in green. The sharp discrepancies between the two plots are similar the discrepancies we saw in our non-rigid motion simulation.

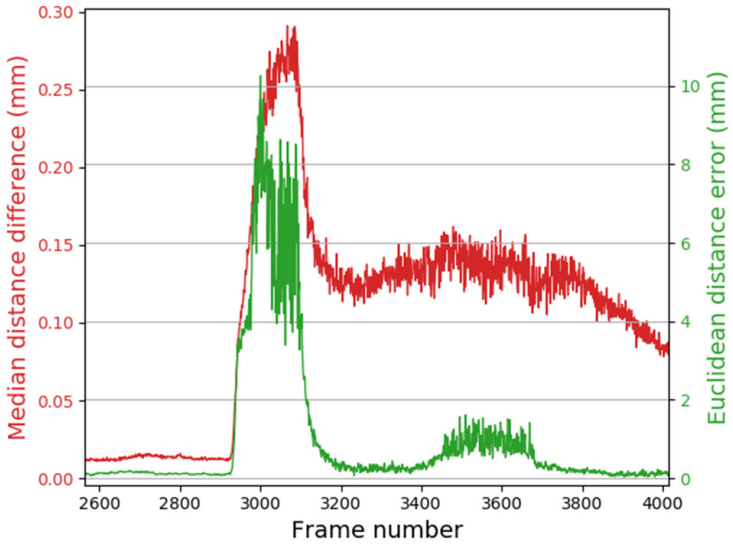

Fig. 3. Median landmark distance difference (red) plotted over Euclidean distance error (green) for a single volunteer. The first occurrence of non-rigid motion occurs close to frame 3000, where both the median distance measure and Euclidean error sharply increase.

IV. Conclusions and Future Work

We have developed a simulation tool that will help us to model and potentially detect and correct for non-rigid deformations during head motion tracking. Two methods for detecting non-rigid motion were investigated: one involved calculating mutual triangulated landmark distances each frame and comparing them to the database; the other used a neural network architecture specially designed for point cloud classification. The trained network approach appears the more promising of the two, as incorporating it into the current feature-based algorithm would be simpler as it requires less calculation than the mutual landmark distance approach. Future work will aim to incorporate temporal information into the neural network architecture, and a direct comparison between the two non-rigid motion detection methods on simulated and real data.

References

[1] J. Maclaren *et al.*, *Magnetic Resonance in Medicine*, vol. 69, no. 3, pp. 621–636, 2013.

[2] M. Zaitsev *et al.*, *NeuroImage*, vol. 31, no. 3, pp. 1038–1050, 2006.

[3] A. Kyme *et al.*, in *IEEE Nuclear Science Symposium, Medical Imaging Conference*, 2016.

[4] D. Henry *et al.*, in *IEEE Nuclear Science Symposium, Medical Imaging Conference*, 2018.

[5] A. Kyme *et al.*, *IEEE Transactions on Medical Imaging*, vol. 33, no. 11, pp. 2180–2190, 2014.

[6] C. R. Qi *et al.*, *arXiv preprint arXiv:1612.00593*, 2016.