Characterization of the Intel RealSense D415 Stereo Depth Camera for Motion-Corrected CT Imaging

M. Dashtbani Moghari, *Student Member, IEEE*, P. Noonan, D. Lewis Henry, *Student Member, IEEE*, R. Fulton, *Senior Member, IEEE*, N. Young, K. Moore, and A. Kyme, *Member, IEEE*

***Abstract*– A combination of non-contrast CT (NCCT) and CT Perfusion (CTP) imaging is the most common regimen for evaluation of stroke patients. CTP-based image analysis is known to be compromised by patient head motion. However, there is currently no technique to compensate for intra-frame head motion during CTP acquisition. In this work, we investigated the feasibility of using the small form factor Intel RealSense D415 stereo depth camera to obtain accurate head pose estimates for intra-frame motion correction in CTP. First, we quantitatively evaluated head movement in a cohort of 72 acute stroke cases. Then we characterized the performance of the Intel D415 against ground-truth robotic motion and the clinically validated OptiTrack marker-based motion tracking system. The results showed that head motion during CTP imaging of acute stroke of patients is extremely common, with around 50% of patients moving > 5 mm and 1 deg and around 20% moving 10-100 mm and rotating 3-20 deg. The pose accuracy of the Intel for controlled robotic motion was approximately 5 mm and 2 deg. For translations and rotations, respectively. For human head motion using the OptiTrack as ground truth, the accuracy was approximately 4 mm (except for lateral motion) and 1.25 deg, respectively. Although poorer than what is needed clinically, there is a lot of potential to optimize performance and potentially achieve an accuracy consistently around 1 mm and 1 deg.**

I. Introduction

A combination of non-contrast CT (NCCT) and CT Perfusion (CTP) imaging is the most common regimen for diagnosis, treatment decision making and treatment monitoring in stroke patients [1]. CTP-based image analysis and stroke modelling is known to be compromised by patient head motion, however perfusion software packages typically only perform very basic frame-to-frame motion correction. Moreover, although intra-frame head motion is common, there are currently no techniques to perform intra-frame motion correction in CTP. Thus, the aim of this work was to investigate the feasibility of using the small form factor Intel RealSense D415 stereo depth camera to obtain accurate and rapidly sampled head pose estimates for intra-frame motion correction in CTP.

In this paper, we first quantitatively evaluate head movement in a cohort of 72 acute stroke cases. We then characterize and validate the performance of the Intel D415 against ground-truth robotic motion and the clinically validated OptiTrack marker-based motion tracking system.

II. Method

A. Extent of head motion during CTP imaging

We performed frame-to-frame motion analysis of CTP datasets from a random cohort of 72 acute stroke cases admitted to Westmead Hospital, Sydney. Head movement was classified into four categories – extreme, high, moderate and low – according to Table 1. For each patient, we rigidly registered all frames (33 time frames in total) to the first frame (reference) using SPM12 [2]. The extent of the head movement relative to the first time frame was described by 6 motion parameters (3 rotations and 3 translations).

B. Rationale for the Intel D415 as a motion tracking device

We hypothesize that the prevalence of inter-frame head motion in CTP (Section III) implies underlying continuous (intra-frame) motion which should be corrected. Therefore, to fully correct for motion in CTP data, rapidly sampled motion estimates synchronized with the CTP acquisition are needed. CTP frames are collected within a short time (1.5-3 seconds per frame). The Intel D415 supports a framerate of up to 90 Hz, requires no attached markers, and can produce a depth map with high pixel density when working at close range (<1 m). Furthermore, it is cheap, compact and can be easily integrated into the clinical scanner. The D415 has a relatively narrow field of view (FoV), 69° horizontal and 42° vertical, however with the use of multiple devices, more coverage can be obtained without interference.

C. Robotic motion testing

A realistic rubber-mask human face phantom was rigidly mounted to a UR3 robot (Universal Robots, Denmark) and

Table I. Head motion classification. "T" and "R" refer to translational and rotational motion, respectively.

Category	Motion Component
Extreme	T >= 100 mm or R >= 20°
High	100 mm > T > 10 mm or 20° > R >= 3°
Moderate	10 mm> T >= 5 mm or 3° > R >= 1°
Low	5 mm > T >= 0 mm or 1° > R >= 0°

Manuscript received May 5, 2019.

M. Dashtbani Moghari and A. Kyme are with School of Biomedical Engineering, University of Sydney, Sydney, Australia (e-mail: mdas7628@uni.sydney.edu.au; andre.kyme@sydney.edu.au).

P. Noonan is with School of Biomedical Engineering & Imaging Sciences, King's College London, UK (e-mail: philip.noonan@kcl.ac.uk).

D. Lewis Henry, R. Fulton and A. Kyme are with Faculty of Health Sciences and the Brain & Mind Centre, University of Sydney, Sydney, Australia (e-mail: dhen2714@uni.sydney.edu.au; roger.fulton@sydney.edu.au).

N. Young and K. Moore are with Department of Radiology, Westmead Hospital, Sydney, Australia (e-mail: Noel.Young@health.nsw.gov.au; Krystal.Moore@ health.nsw.gov.au).

978-1-7281-4164-0/19/$31.00 ©2019 IEEE

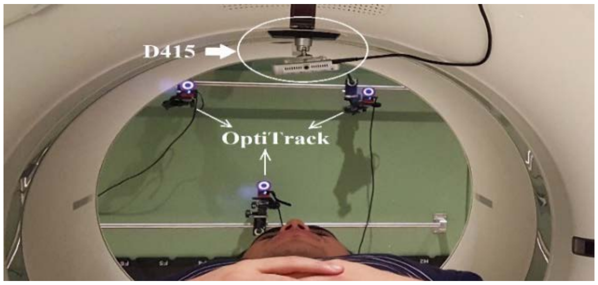

Fig. 1. Experimental setup to validate the Intel D415 using the OptiTrack system.

viewed by a single Intel D415 at a distance of 0.35 m. The Intel camera was connected to a powered USB hub to avoid frame drop-out.

The robot tool pose was extracted at 125 Hz by interrogating the robot controller MODBUS server while the Intel D415 simultaneously recorded phantom poses at 30 Hz in the Intel coordinate system. To record poses with the Intel, we used software developed using an open-source multi-platform API [3,4]. The resolution of the Intel depth frames was set to 848×480. RGB images collected by the Intel were not used.

To cross-calibrate the UR3 robot and Intel camera we collected 35 distinct static poses of the phantom and performed a hand-eye calibration based on dual quaternions [5]. Pose averaging was used to reduce noise and improve the robustness of the cross-calibration.

To quantify the pose accuracy of the Intel, we programmed the robot to execute a continuous motion trajectory over 90 seconds with motions in the “moderate” to “high” range reported in Table 1. Intel poses were directly compared to the robotic motion after applying the cross-calibration.

D. Clinical testing

We also performed a proof-of-principle motion tracking validation of the Intel inside the CT component of a clinical PET/CT scanner (Siemens Biograph mCT) by comparing with a well validated marker-based motion tracking system. The Intel D415 was mounted inside the CT bore using a 3D printed custom mounting device facilitating pan/tilt positioning. Three OptiTracks (NaturalPoint, Inc.) were clamped to the wall behind the CT scanner at a distance of ~2 m. Four retroreflective OptiTrack markers were attached to the head of the target (phantom or human volunteer) and tracked as a rigid body at a frequency of 100 Hz. The experimental setup is shown in Fig. 1.

We computed the cross-calibration between the Intel and OptiTrack systems using a face phantom moved to 35 discrete poses, similarly to Section II.C. A human volunteer wearing an OptiTrack head cap with the reflective markers affixed then lay on the scanner bed and moved his head continuously for 90 seconds while both the Intel D415 and OptiTrack recorded poses. The Intel D415 frame-rate was 30 Hz.

III. Results and discussion

The results of motion analysis of CTP data from acute stroke patients showed that head movement is very common: 53% (38/72) low motion, 26% (19/72) moderate motion, 18% (13/72) high motion and 3% (2/72) extreme motion. These data corroborate what other researchers and vendors have reported regarding the importance of correcting for motion in CTP imaging [6].

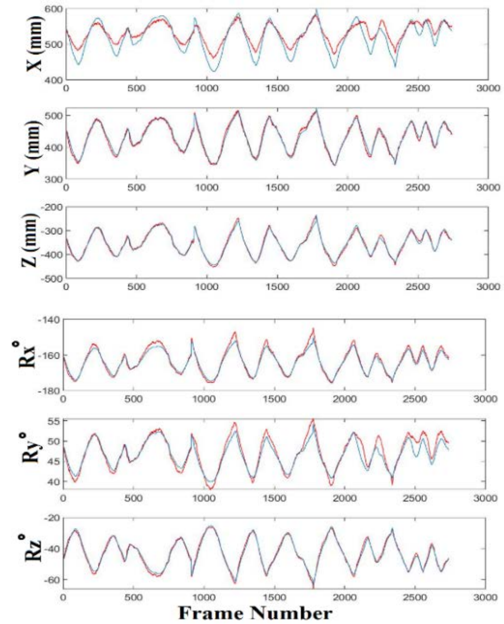

Fig. 2. Intel D415 and OptiTrack comparison in Intel coordinate system for human head motion inside the CT scanner. Red and blue graphs show motion tracks recorded by Intel and OptiTrack respectively.

Table II. RMSE (mm) between Intel D415 and robot/OptiTrack

RMSE	X (mm)	Y (mm)	Z (mm)	Rx°	Ry°	Rz°
Robotic Motion	5.39	4.39	6.86	2.18	1.93	2.40
Human Motion	17.79	3.05	4.46	1.47	1.34	1.13

The root-mean-square error (RMSE) between the Intel D415 and the ground truth motion (UR3 and OptiTrack) pose measurements are shown in Table 2. Figure 2 also shows a comparison of motion tracking by the OptiTrack and Intel D415 over 90 seconds for human head motion. Although discrepancies are larger than we would like for CT, the presence of values close to 1 mm for some degrees-of-freedom (DoF) is promising. There are several likely reasons for the reduced pose accuracy:

(i) The greatest discrepancy occurred about the x-axis (head roll) and is caused by obstruction of the face (due to the nose) in the stereo depth data. It is clear from our testing that a smaller tracking ROI introduces greater uncertainty since there are less points to perform iterative closest point (ICP) registration for pose estimation. This suggests having a larger ROI covering more of the surface is critical – however, it comes with increased likelihood of obstruction. One solution is to use multiple cameras to track motion from multiple views. We have built a trigger synchronization device to accommodate up to 4 Intel D415 cameras and will investigate the potential for multiple cameras to improve accuracy in future work.

(ii) As shown in Table 2, agreement between the Intel D415 and the robot was worse than for the OptiTrack (except for x-axis). This may be partly due to the additional DoF of the robotic motion. This extra range of motion increased the chance of partially losing the tracked ROI in these tests.

(iii) According to Table 2, the discrepancy corresponds to ~1%-2% of the distance between the object and Intel D415, which is consistent with the Intel D415 specifications. Using optimized presets for the Intel D415, such as optimized exposure, gain, depth unit and resolution settings, are likely to improve the accuracy.

Investigation on the impact of presets and the use of multiple cameras on the pose accuracy is in progress and will be addressed in our future work. We will also investigate how theIntel D415 may be used to support a recent data-driven method [7].

IV. Conclusion

Head motion is very common in CTP imaging and can compromise diagnosis and treatment management of stroke patients. The Intel D415 depth camera has the potential to provide rapidly sampled pose measurements during CTP for motion corrected imaging. Preliminary characterization of the Intel D415 showed potential for its use in CT, however further optimization is required to achieve ~1 mm accuracy consistently. Several improvements are currently under development to achieve this.

References

[1] JJ. Heit, and M. Wintermark, "Perfusion computed tomography for the evaluation of acute ischemic stroke: strengths and pitfalls," Stroke, vol. 47, no. 4, pp. 1153-8, Apr. 2016

[2] Available: https://www.fil.ion.ucl.ac.uk/spm/software/spm12/

[3] Available: https://software.intel.com/sites/products/realsense/camera/

[4] Available: https://github.com/philipNoonan/oglFusionRealsense

[5] K. Daniilidis, "Hand-eye calibration using dual quaternions," The International Journal of Robotics Research, vol. 18, no. 3, pp. 286-98, Mar. 1999.

[6] F. Fahmi, LFM. Beenen, GJ. Streekstra, NY. Janssen, HW. de Jong, A. Riordan, YB. Roos, CB. Majoie, and HA. Marquering. "Head movement during CT brain perfusion acquisition of patients with suspected acute ischemic stroke," European journal of radiology vol. 82, no. 12, pp. 2334-2341, Dec. 2013.

[7] S. Tao, JH. Kim, R. Fulton, and J. Nuyts. "An iterative projection-based motion estimation and compensation scheme for head x-ray CT," Medical physics, vol 43, no. 10, pp. 5705-5716, Oct. 2016.