3D shape and motion estimation from optical coherence tomography volumes

Nahashon O. Osinde[a], Nicolas Andreff[a,*]

[a] *Université Marie et Louis Pasteur, SUPMICROTECH, CNRS, institut FEMTO-ST, 15B Avenue des Montboucons, Besancon, 25000, France*

Abstract

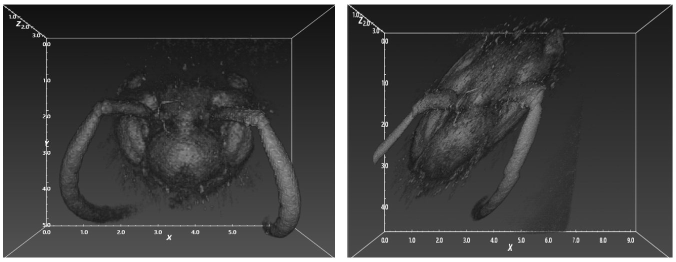
Due to the sequential, one-dimensional mechanical scanning of optical coherence tomography (OCT), a rolling shutter acquisition mode is induced that produces geometric artefacts due to the relative motion between the imaged object and the OCT probe. In this paper, we propose a technique that exploits motion artefacts in OCT volumes, to estimate both the 3D shape and motion of an arbitrary sample. Our proposed method begins by mathematically modelling motion artefacts in OCT in the 3D Cartesian space $SE(3)$. Subsequently, we develop a methodology to estimate the motion using a volume similarity method that compares a reference OCT volume with a motion-distorted OCT volume, as a whole, rather than pointwise through 3D-to-3D point matching. The estimated motion is then utilised to correct any artefacts present in the motion-distorted volume. This methodology is then enhanced to estimate both the shape and motion from only one OCT volume acquired using the rosette non-Cartesian scan trajectory (without prior knowledge of the shape). Our results reveal that accurate shape and motion estimation are achievable using both numerically simulated and experimental OCT data, demonstrating the robustness of our method.

Keywords: OCT, rolling shutter, kinematic model, motion estimation, shape estimation, volume similarity

1. Introduction

1.1. Background

For the past three decades, optical coherence tomography (OCT) has been a significant tool in carrying out minimally invasive optical biopsies in ophthalmology and lately in oncology [1, 2, 3, 4, 5, 6, 7]. OCT is an optical diagnostic imaging modality that utilises low-coherence interferometry to obtain high-resolution, one-dimensional axial scans (A-scan), cross-sectional (B-scan) images and volumetric (C-scan) datasets of internal tissue micro-structure in the micrometer resolution range by measuring echoes of backscattered light [8, 9, 10, 11, 12, 13, 14]. During OCT data acquisition, the OCT device sequentially records the intensity and delay of reflections along the path of the light beam through an object [15] to form an A-scan. Then, to form images (B-scans) or 3D-volumes (C-scans), a mechanical scanner conventionally follows a raster trajectory within a scan coordinate system covering the x and y positions of the scanning mechanism, and the axial depth along the light beam ($z - axis$) [16, 17]. Thereby, an OCT image or volume does not represent a single moment in time, but rather, it is a concatenation of one-dimensional data (A-scans) captured over a period of time. When there is relative movement between the imaged object and the OCT probe, distortions or inconsistencies referred to as motion artefacts are introduced in the volume. Motion artefacts are thus inherent to OCT imaging due to the principle of data acquisition, as can be observed in Fig 1.



a OCT data acquired when a bee-head is stationary

b OCT data acquired when a bee-head is in translatory motion In the $x - axis$

Fig. 1: Motion artefacts of a bee-head due to motion during OCT scanning

In the realm of medical imaging, motion artefacts can arise from patient movements during the image acquisition process. Involuntary movements like respiration, and voluntary actions due to patient discomfort, can manifest as blurring, streaking, or shading within the captured images [18]. Ignoring patient movement (consequently tissue motion) during image acquisition and later on processing, degrades the image quality which can result in inaccurate diagnoses by medical professionals [19].

Before proceeding further, it shall be noted that motion artefacts are not restricted to OCT, but also to standard cameras working in the so-called rolling shutter acquisition mode [18, 20, 21]. Indeed, one major drawback of rolling shutter cameras is the introduction of geometric distortions in acquired images, when either the camera or the target is in motion. These distortions occur because pixels are acquired at different points in time [22, 23, 24], and there has been a tremendous research on that topic in parallel to OCT motion artefact compensation. The latter serves as an inspiration to this work, namely for the compensation for a full motion in the Cartesian space $SE(3)$ (*i.e.,*

---

[*] Corresponding author: nicolas.andreff@femto-st.fr

[1] Emails:nahashon.osinde@edu.univ-fcomte.fr(N.O.Osinde), nicolas.andreff@femto-st.fr (N. Andreff)

[2] Orcid:0000-0002-9545-8967) (N.O.Osinde)

Preprint submitted to Machine Vision and Applications *April 28, 2026*

3 dimensions in translation and 3 dimensions in rotation). In particular, rolling shutter acquisition has been applied to high-speed robot control [25]. As a consequence, motion correction techniques not only aim to ensure that the A-scans in corrected OCT data accurately represent their expected locations, irrespective of any motion that may have occurred during data acquisition but also could be used in conjunction with microrobotics in view of intracorporeal robot-assisted optical biopsy and surgery.

There are two main ways to correct for motion artefacts in OCT; intra-processing and post-processing methods. Intra-processing methods aim to mitigate motion artefacts during data acquisition through specialised or improved hardware designs [26, 27, 28]. One approach involves incorporating advanced optics for rapid scanning, which would reduce acquisition time and the likelihood of motion-induced artefacts [29]. Another approach employs real-time tracking of OCT measurement deviations. This is achieved through an additional imaging modality such as a scanning laser ophthalmoscope [30]. Real-time tracking allows dynamic adjustments to the scanning mechanism positions during data acquisition, compensating for motion-related inconsistencies and enhancing the accuracy and reliability of OCT measurements [16, 31]. However, introducing or modifying specialised hardware can be challenging, especially in miniaturised scenarios like endoscopy, where space limitations exist. Post-processing methods aim to mitigate OCT motion artefacts during the image processing stage, primarily through software approaches, ie., after data acquisition [32, 33]. Approaches include correcting OCT data using images/data from another modality unaffected by motion artefacts [34, 35], or employing cross-correlation techniques on successively acquired OCT data to filter out motion effects [36, 37, 38]. Since cross-correlation techniques take time, it was recently proposed to rely on feature detectors and matching to develop fast compensation [39]. The last class of post-processing methods relies on deep-learning methods [40, 41, 42]. The use of another imaging modality may pose challenges in miniaturised scenarios. Moreover, existing cross-correlation approaches typically use 2D images not 3D volumes, and are primarily focused on ophthalmology [32, 36, 43, 44, 45]. If ophthalmology represents the largest application domain of OCT motion artefact compensation [46], it is nevertheless a specific case of the latter since the considered eye motion in existing approaches involves translation and tilt (in-plane rotation) without depth, i.e., 2D Cartesian motion ($SE(2)$) only. Finally, deep learning is associated to concerns about the trust one can give to the databases used for learning and the inexplicability of the results - especially in the clinical regulatory context - as well as environmental costs.

These different methods offer valuable solutions for mitigating motion artefacts in OCT data, allowing for improved quality and accurate interpretation of the underlying structures. The choice of the specific method depends on the nature of the motion artefacts and the available imaging resources.

1.2. Contributions

In this paper, our contribution is the proposal of a novel numerical method that leverages OCT motion artefacts, and a volume similarity method to estimate the motion in all six dimensions of the Cartesian space $SE(3)$ that caused the artefacts, drawing inspiration from robotics and computer vision concepts. The method employs a kinematic model to simulate motion artefacts in an OCT point cloud volume, which is subsequently used to estimate the corresponding instantaneous linear and angular velocities responsible for the motion artefacts without using any feature detector nor matching. These estimated velocities are then used in this paper to correct for the artefacts, resulting in a motion-free corrected volume. Notably, the proposed method is efficient, robust, and doesn't necessitate any additional hardware or instrumentation. The proposed method is also more explicable than deep-learning methods [40, 41] without the environmentally unaffordable need of large training datasets. The proposed motion estimation method extends its applicability beyond 2D images with discontinuities to addressing artefacts in 3D OCT volumes. Correcting for motion artefacts in 3D OCT volumes is particularly challenging due to their high dimensionality and the likelihood of distortions occurring within the volume. This represents a significant advancement over existing methods, which are typically limited to 2D OCT images and/or predominantly applied in ophthalmology applications. Moreover, this paper presents a novel approach for shape and motion estimation based on a single OCT volume obtained using the rosette non-Cartesian/non-raster trajectory, or an R-scan for short. This is an approach that marks a paradigm shift from conventional OCT volume motion artefacts compensation which typically relies on two sequential OCT volumes. By eliminating the need for two consecutive volumes, this approach reduces the amount of data required and, consequently, the computation costs needed. Actually, our method is more frugal and affordable to the environment than most of the existing techniques since it uses few thousands of A-scans (equivalent to few B-scans) instead of full C-scans, reducing drastically the amount of data to be treated.

The structure of the paper is as follows: Section 2 addresses the mathematical modelling of motion artefacts in OCT and introduces the proposed motion estimation methodology for OCT volumes affected by such artefacts. Section 3 presents the numerical simulations conducted to validate the proposed methodology. Section 4 describes the experimental setup, along with the results and discussion. Section 5 discusses the computation time and efficiency, Section 6, the limitations of the study, and Section 7 concludes the paper and outlines future perspectives.

2. 3D shape and motion estimation from OCT volumes

In this section, we begin by formulating a comprehensive mathematical model that describes the occurrence and characteristics of motion artefacts in OCT imaging. Motion in the proposed model may manifest through pure translations, pure rotations, or a combination of both. Thereafter, we detail our proposed method for 3D motion estimation from OCT volumes affected by motion artefacts having a reference shape and leveraging on volume similarity. This approach is further developed to a scenario where prior knowledge of the shape is not necessary, thus enabling the estimation of both 3D shape and motion.

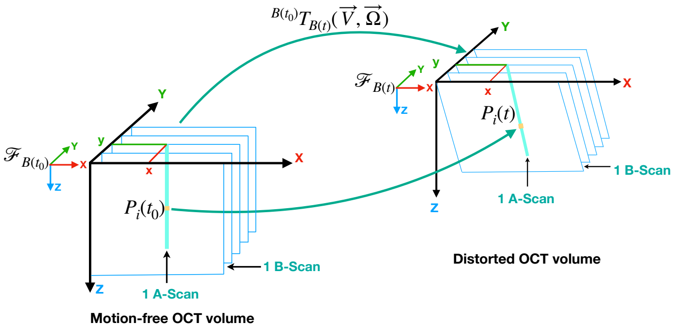


Fig. 2: Schematic of a distorted OCT volume due to motion during data acquisition

2.1. *Modelling of motion artefacts in OCT*

As a result of the sequential nature of OCT data acquisition, all the voxels of the scanned object are not acquired simultaneously, as in the case of a global shutter camera. Instead, each voxel is acquired one after the other in a sequential manner. Therefore, when there is instantaneous motion during data acquisition, distortions are introduced in the resulting OCT volume just as distortions are introduced in a rolling shutter camera image. To develop a generic model of rolling shutter in OCT, one voxel is considered in the A-scan of an OCT volume of an object given by $P_i(t_0)$ in the object frame $\mathcal{F}_{B(t_0)}$, as shown in Fig 2 (t_0 is the initial time before any motion). At time t, the object would have undergone a displacement due to the instantaneous motion, linear velocity, $\vec{V}$ and angular velocity, $\vec{\Omega}$. The motion and the time delay between the acquisition of voxels introduces a transformation (${}^{B(t_0)}T_{B(t)}$) that shifts the location of the object frame, and thus of each voxel, to a different location (taking into account motion during time t_0 to t), as illustrated in Fig 2.

The homogeneous coordinates of one voxel at time t_0 in the object frame, $\mathcal{F}_{B(t_0)}$ are:

$$ {}^{B(t_0)}\tilde{P}_i(t_0) = \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{B(t_0)}P_i(t_0) \\ 1 \end{bmatrix} \tag{1} $$

At time t, the voxel has undergone a transformation due to the instantaneous motion during time $t - t_0$:

$$ {}^{B(t)}\tilde{P}_i = {}^{B(t_0)}T_{B(t)}{}^{-1}\ {}^{B(t_0)}\tilde{P}_i \tag{2} $$

It is necessary to estimate the motion model; ${}^{B(t_0)}T_{B(t)}$. Solving for all voxels, $P_{i...n}$, individually is an underconstrained problem, we therefore make two regularizing assumptions;

- Rigid body motion (the volume is non-deforming) i.e., ${}^{B(t)}\tilde{P}_i(t) = {}^{B(t_0)}\ \tilde{P}_i(t_0)$
- Constant velocity of the object in its local frame i.e., ${}^{B(t)}\vec{V}_{B(t)} = \vec{V}$, ${}^{B(t)}\vec{\Omega} = \vec{\Omega}$

From now on P_i will be represented as P for clarity and readability. Due to the object motion, the total velocity of the single voxel is given by:

$$ {}^{B(t)}\dot{P}(t) = {}^{B(t)}\vec{V}_{B(t)} + {}^{B(t)}\ \vec{\Omega} \wedge {}^{B(t)}\ P(t) \tag{3} $$

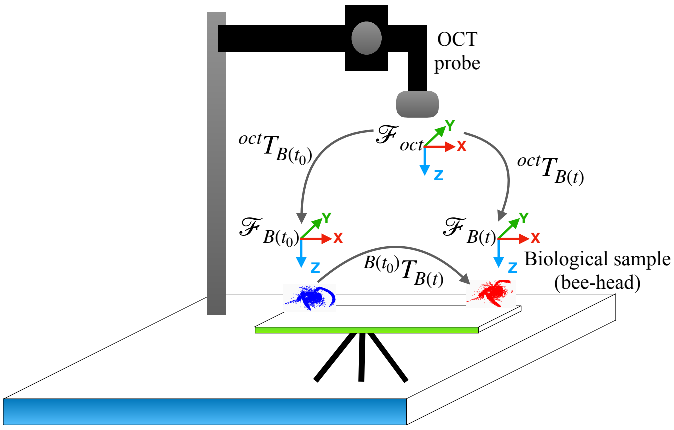


Fig. 3: Schematic of the OCT and object frames

where $\vec{V}$ is the linear velocity and $\vec{\Omega}$ is the angular velocity of the rigid object at time t, and $\wedge$ is the cross product symbol. Re-writing equation 3 in homogeneous coordinates results in:

$$ {}^{B(t)}\dot{\tilde{P}}(t) = \begin{bmatrix} {}^{B(t)}\dot{P}(t) \\ 0 \end{bmatrix} = \begin{bmatrix} {}^{B(t)}\vec{V}_{B(t)} + {}^{B(t)}\ \vec{\Omega} \wedge {}^{B(t)}\ P(t) \\ 0 \end{bmatrix} \tag{4} $$

Re-arranging equation 4 using the skew-symmetric matrix $[\vec{\Omega}]_\wedge$ yields,

$$ {}^{B(t)}\dot{\tilde{P}}(t) = \begin{bmatrix} [{}^{B(t)}\vec{\Omega}]_\wedge & {}^{B(t)}\vec{V}_{B(t)} \\ 0 & 0 \end{bmatrix} {}^{B(t)}\tilde{P}(t) \tag{5} $$

Equation 5 is a 1st order ODE in $\mathbb{R}^4$. Leveraging on the concept of the exponential map (Lie theory) as detailed in [47, 48], it solves, such that:

$$ {}^{B(t)}\tilde{P}(t) = expm\left(\begin{bmatrix} [{}^{B(t)}\vec{\Omega}]_\wedge & {}^{B(t)}\vec{V}_{B(t)} \\ 0 & 0 \end{bmatrix} (t - t_o) \right) {}^{B(t_0)}\tilde{P}(t_o) \tag{6} $$

Therefore, the motion model (transformation of the object at time t_0 to object at time t) i.e., during movement and OCT scanning is:

$$ {}^{B(t_0)}T_{B(t)} = expm\left(\begin{bmatrix} [{}^{B(t)}\vec{\Omega}]_\wedge & {}^{B(t)}\vec{V}_{B(t)} \\ 0 & 0 \end{bmatrix} (t - t_o) \right) \tag{7} $$

So far, we haven't considered the observation of the object from the OCT frame (the OCT volumes are processed in the OCT frame) as shown in Fig 3. Taking the analogy of rolling shutter in OCT, the equation of motion of the voxel at time t in the OCT frame, $\mathcal{F}_{oct}$ is thus given by:

$$ {}^{oct}\tilde{P}(t) = {}^{oct}T_{B(t)}\ {}^{B(t)}\tilde{P}(t) \tag{8} $$

The transformation of the object from the object frame to the OCT frame is as given in equation 9:

$$ {}^{oct}\tilde{P}(t) = {}^{oct}T_{B(t_0)}\ {}^{B(t_0)}T_{B(t)}\ {}^{B(t)}\tilde{P}(t) \tag{9} $$

Knowing the instantaneous linear and angular velocities, simulated motion-distorted OCT volumes can be obtained using equation 9.

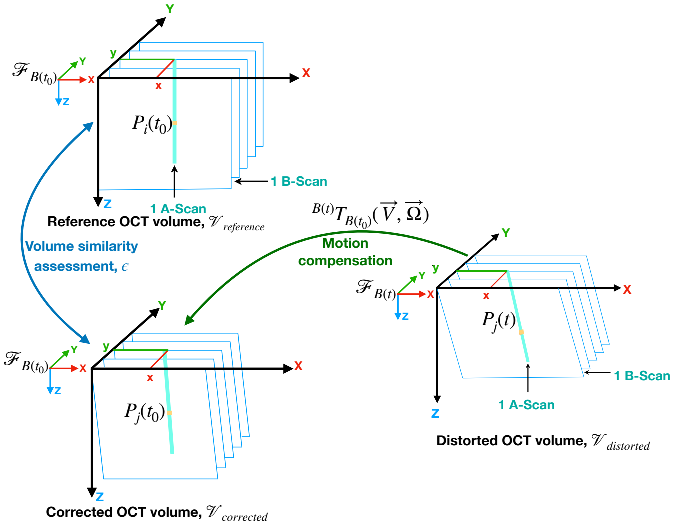


Fig. 4: Illustration of the proposed motion estimation process knowing the OCT volume reference shape

2.2. 3D motion estimation from an OCT volume with motion artefacts

Motion estimation in volumes differs from motion estimation in 2D images due to differences in dimensionality (voxels instead of pixels), and most notably, the complexity involved in performing 3D feature detection and 3D-3D matching. Matching between two volumes necessitates estimating the distance between them rather than calculating the sum of square distances between matched points. As a result, the process would involve computing the full corrected volume and subsequently comparing it to the reference volume. This comparison can be conducted B-scan by B-scan or slice (*en face* image in the $x-y$ plane) by slice but is rather complex for volume by volume. Therefore, in this paper, we propose to use a top down approach in addition to a volume-wise criterion of similarity. Implementing a volume-wise criterion of similarity, is one novelty of our proposed method. Assuming that the imaged object is inherently smooth, jumps between consecutively acquired OCT volumes are induced by motion only [49]. To perform motion correction of OCT data as illustrated in Fig 4, it is necessary to estimate the instantaneous linear ($\vec{V}$) and angular velocity ($\vec{\Omega}$) in addition to the pose ${}^{oct}T_{B(t_0)}$. This is achieved by minimising an objective function that reduces the similarity error between a corrected volume and a corresponding reference volume. To compute this similarity error ϵ, equation 10 is used:

$$\epsilon(\vec{V},\vec{\Omega},{}^{oct}T_{B(t_0)}) = similarity\ error(\mathcal{V}_{reference},\mathcal{V}_{corrected}) \tag{10}$$

where the reference volume, $\mathcal{V}_{reference}$, is defined by:

$$\mathcal{V}_{reference} = \underbrace{\left\{{}^{B(t_0)}P_i(t_0) \quad \forall i \in [1,n]\right\}}_{\text{reference volume}}, \tag{11a}$$

the corrected volume, $\mathcal{V}_{corrected}$, is calculated by:

$$\mathcal{V}_{corrected} = \underbrace{\left\{ {}^{B(t_0)}T_{B(t)}^{\ -1}\ {}^{oct}T_{B(t_0)}^{\ -1}\ \underbrace{{}^{oct}P_j(t) \quad \forall j \in [1,m]}_{\text{distorted volume due to } V,\Omega} \right\}}_{\text{corrected volume due to } V,\Omega}, \tag{11b}$$

and n and m are the number of voxels in the respective volumes. Also, ${}^{B(t_0)}T_{B(t)}$ depends on $(\vec{V},\vec{\Omega})$ by equation 7 and equation 11b (which performs motion compensation) derives by the inversion of equation 9 such that:

$${}^{B(t_0)}\tilde{P}(t_0) = {}^{B(t)}\tilde{P}(t) = {}^{B(t_0)}T_{B(t)}^{\ -1}\ {}^{oct}T_{B(t_0)}^{\ -1}\ {}^{oct}\tilde{P}(t) \tag{12}$$

To compute ϵ, we employ the objective quantitative metric, PointPCA, detailed in [50]. PointPCA is a full-reference statistical metric that assesses local shape and appearance by comparing a point cloud under evaluation with its reference. To compute volumetric similarity, PointPCA considers geometric descriptors (eigenvalues, linearity, planarity, sphericity, anisotropy, omnivariance, eigenentropy, surface variation, roughness, parallelity) based on principal component analysis (PCA) to capture local shape properties relative to topology. In addition, textural descriptors (luminance measurement) account for appearance properties. The volume similarity score is then computed using statistical functions (mean and standard deviation) applied to the geometric and textural descriptor values, capturing local relationships between the two point clouds. These statistical features effectively capture dependencies within local neighborhoods yielding measurements that are more perceptually relevant. In particular, they are well aligned with key characteristics of the human visual system: the local mean reflects low-pass filtering behavior, while the local standard deviation quantifies sensitivity to high-frequency variations. The comprehensive technical and scientific details of the PointPCA metric are provided in [50], and the corresponding software implementation is publicly accessible at `https://github.com/cwi-dis/pointpca`.

Formally, the determination of the optimal motion correction values is expressed as a nonlinear minimisation problem:

$$(\vec{V}^*,\vec{\Omega}^*,{}^{oct}T^*_{B(t_0)}) = \underset{\{\vec{V},\vec{\Omega},{}^{oct}T_{B(t_0)}\}}{argmin}\ \epsilon(\vec{V},\vec{\Omega},{}^{oct}T_{B(t_0)}) \tag{13}$$

This problem involves a total of 12 unknowns, which consist of the linear and angular velocities of the observed motion, as well as the transformation pose that encompasses translation and rotation between the OCT frame and the object frame. Therefore, starting with an initial guess for $\vec{V}$, $\vec{\Omega}$ and ${}^{oct}T_{B(t_0)}$, an iterative optimisation process will execute motion correction to address the induced distortion.

For the optimisation process, we compute the inverse transformation for each voxel in the distorted OCT volume using equation 11b to obtain a corrected OCT volume. Equation 11b takes into account the previously estimated instantaneous linear and angular velocity values as well as the transformation pose. By applying this inverse transformation to every voxel in the distorted OCT volume, we can effectively correct the motion artefacts and generate a desired artefact-free corrected volume. This process ensures that each voxel is appropriately adjusted based on the corresponding motion information, resulting in an artefact-free representation of the volume. The quality of this correction is then assessed using equation 10 to further update the estimation of $\vec{V}$, $\vec{\Omega}$ and ${}^{oct}T_{B(t_0)}$. Upon successful iterations, the process determines the optimal $\vec{V}^*$, $\vec{\Omega}^*$, ${}^{oct}T^*_{B(t_0)}$, and does the compensation using them, making the corrected volume similar to the known reference one.

2.3. 3D shape and motion estimation from OCT volumes with motion artefacts

In contrast to sub-section 2.2 where motion estimation was performed using a known reference OCT volume, in this sub-section, we propose a method to estimate the shape and motion from a single unknown motion-distorted OCT volume without prior knowledge of the shape. The distorted OCT volume is acquired during motion using the rosette non-raster trajectory and provides an R-scan. As detailed in [51], the rosette trajectory is recommended as a non-raster scanning approach for OCT volumetric acquisitions. Employing the rosette trajectory allows for significant time savings without a substantial sacrifice in OCT volume quality [51]. Moreover, while the rosette trajectory could currently have hardware constraints (since most scanners are optimised for raster scanning), the design of the mechanical characteristics of the scanner could be simplified to some degree. Rather than having two galvanometric scanners for the $x-axis$ and $y-axis$, an optimal design with only one actuator that traces the rosette path could be developed. Such a design would likely consume less energy and occupy less space since only one actuator would be needed instead of two [51]. This could potentially be further beneficial in terms of clinical applications, when OCT probes that are mounted on a robotic endoscope need to be miniaturized.

2.3.1. The rosette scan trajectory and its peculiarities

The rosette scan trajectory is an epitochroid where the distance between the centre of the rolling circle and the rotating point is equal to the sum of the radii of a fixed and a rolling circle [52]. It is a periodic pattern that repeats itself over a certain time interval intersecting itself at specific points. As shown in Fig 5a, we see the intersections of three successive petals of the rosette pattern highlighted by different colours along with their midlines and bisectors. The pattern symmetry of the rosette trajectory indicates that the petals intersect not only at the origin but also at their midlines as well as at the bisector of two consecutive midlines. In Fig 5b, the intersections are shown by × and the edge points by $*$. We see that the intersections lie on the concentric circles. The intersections that are close to the periphery are located on the first circle and the ones close to the origin remain on the innermost (fourth) concentric circle [52].

The authors in [52] elaborate that the rosette scan trajectory doesn't require back and forth motions and that the self-intersections of large temporal separation provides information that may be used to detect and remove height drift (image artefacts where sample features are partially removed or tilted due to the thermal fluctuations of instrument parts, specifically, slow expansion and contraction) in atomic force microscopy images [53]. Moreover, a study by Bush *et al.* [54], demonstrated that by acquiring magnetic resonance images using the rosette trajectory, motion artefacts were reduced and thus there was no need for motion compensation. In their image reconstruction model, motion was therefore not taken into account. However, in OCT, we observe a different phenomena. Leveraging the periodicity and the intersecting nature of the rosette pattern, we propose to estimate the 3D shape and motion from a single OCT motion-distorted volume acquired using the rosette trajectory.

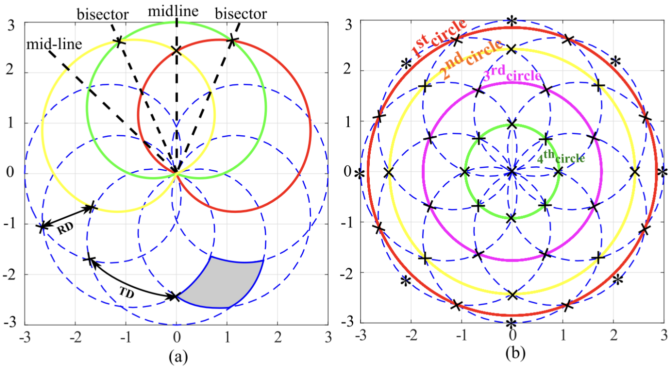


Fig. 5: Illustration of the rosette scan pattern; (a) Three successive petals of the pattern depicted by different colours, (b) Concentric circles crossing the intersections - the intersections are shown by × and the edge points by $*$ [52]

2.3.2. Proposed shape and motion estimation method from a single OCT volume

As can be seen in Fig 5, the rosette scan trajectory has several intersections within itself. Therefore, during data acquisition, data will be acquired repeatedly on some scan positions. During the point cloud image processing step, we can extract two volumes from the OCT data collected using the rosette trajectory. If the data was acquired while the object is stationary, the two extracted volumes (they will be referred to as sub-volumes from now on) should be completely identical. However, if the data was acquired while the object is in motion, the two sub-volumes will be different to some degree (depending on the magnitude of the motion). As mentioned in sub-section 2.2, assuming rigid body motion, we can correct for the motion distortion if we know the linear and angular velocities. What is remarkable is that, when we perform motion correction on the two distorted sub-volumes separately, they revert back to being identical once more. Therefore, if we can extract two sub-volumes from a single OCT volume, motion estimation can be performed. Since the rosette trajectory is an optimal scan trajectory [52], we can further exploit its nature to perform motion estimation by obtaining data only once.

Similarly to how the distortion motion was estimated using two consecutive OCT volumes in sub-section 2.2, we can employ a slightly modified version of the method. To begin, Fig 6 illustrates an OCT volume R_s (obtained using the rosette trajec-

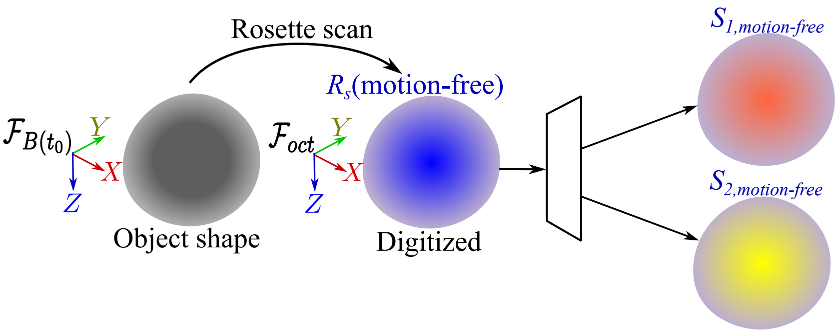


Fig. 6: Extraction of two sub-volumes from one static OCT volume

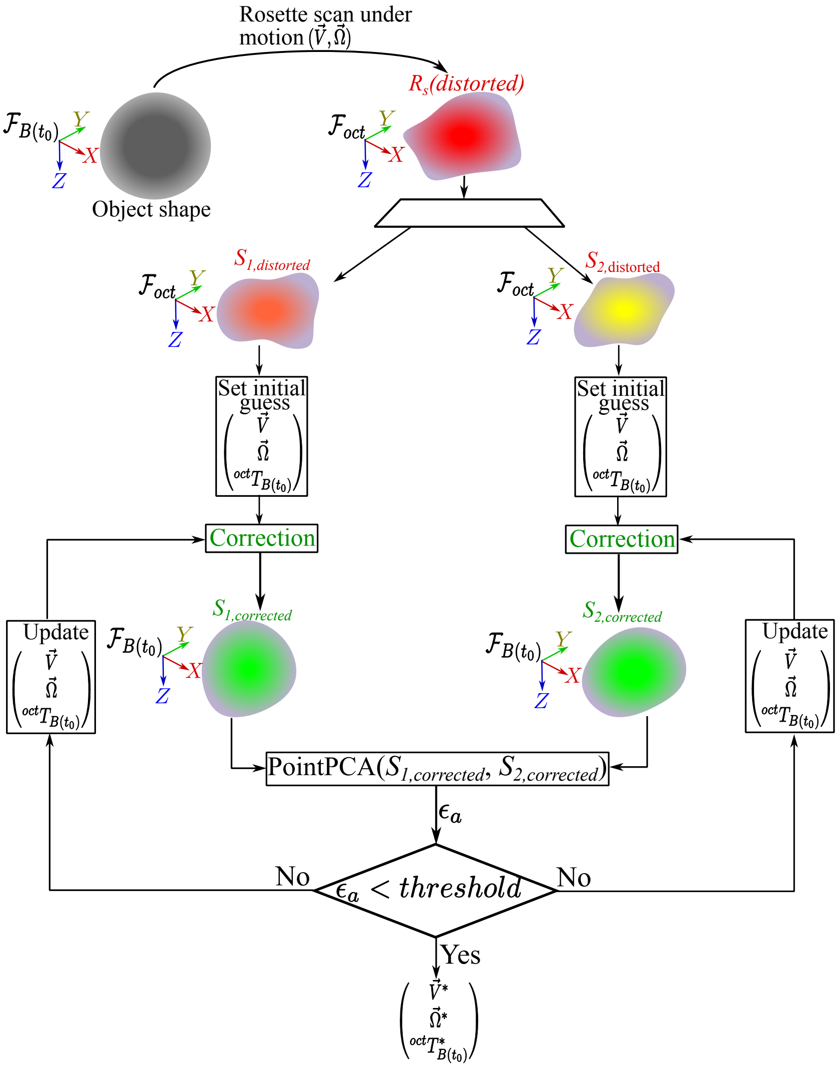


Fig. 7: Illustration of the proposed method to estimate the distortion motion from one OCT volume

tory; also referred to as R-scan), from which two sub-volumes, S_1 and S_2, have been extracted from. The sub-volumes $S_{1,motion-free}$ and $S_{2,motion-free}$ each contain a subset of points from R_s, such that; $S_{1,motion-free} = \{P_k(t_k), \quad \forall t_k \in [t_0, t']\}$, $S_{2,motion-free} = \{P_l(t_l), \quad \forall t_l \in [t_0, t'']\}$, where P_k and P_l are voxels of the sub-volumes S_1 and S_2 respectively, t_0 is the time at the beginning of OCT data acquisition, t' and t'' is the time at which the last voxel of S_1 and S_2 respectively has been acquired.

Fig 7 illustrates the proposed method to estimate the distortion motion from one OCT volume. From Fig 7, the objective is to estimate the same linear ($\vec{V}$) and angular ($\vec{\Omega}$) velocity that can be used to correct the two distorted sub-volumes and make them identical i.e., two subsets of one unique set, in the object shape, in the same frame $\mathcal{F}_{B(t_0)}$ which isn't known beforehand. This is achieved by minimising an objective function that reduces the similarity error between the corrected sub-volumes. The error ϵ_a is computed using equation 14. Even though the acquired distorted sub-volumes appear different, when accurately corrected they should be similar. We can thus form the following similarity error:

$$\epsilon_a(\vec{V}, \vec{\Omega}, {}^{oct}T_{B(t_0)}) = similarity\ error(S_{1,corrected}, S_{2,corrected}) \tag{14}$$

where the two corrected subsets are obtained as in equation 11b:

$$S_{1,corrected} = \underbrace{\left\{ {}^{B(t_0)}T_{B(t_k)}^{\ -1} \ {}^{oct}T_{B(t_0)}^{\ -1} \ {}^{oct}P_k(t_k) \quad \forall t_k \in [t_0, t'] \right\}}_{\text{corrected sub-volume 1 due to } V, \Omega} \tag{15a}$$

$$S_{2,corrected} = \underbrace{\left\{ {}^{B(t_0)}T_{B(t_l)}^{\ -1} \ {}^{oct}T_{B(t_0)}^{\ -1} \ {}^{oct}P_l(t_l) \quad \forall t_l \in [t_0, t''] \right\}}_{\text{corrected sub-volume 2 due to } V, \Omega} \tag{15b}$$

Again, to compute ϵ_a, the objective quality metric PointPCA [50] is used. The resulting optimisation problem that needs to be solved using a numerical algorithm to obtain motion correction values is:

$$(\vec{V}^*, \vec{\Omega}^*, {}^{oct}T^*_{B(t_0)}) = \underset{\{\vec{V}, \vec{\Omega}, {}^{oct}T_{B(t_0)}\}}{argmin} \ \epsilon_a(\vec{V}, \vec{\Omega}, {}^{oct}T_{B(t_0)}) \tag{16}$$

Note that, we assume here that the velocity is constant and equal for each of the two sub-volumes.

2.4. Summary of the method

The algorithm described above can be summarized by the pseudo-code in Algorithm 1. This algorithm was implemented in MATLAB with function ga.

3. Numerical simulations

To validate the proposed methods, simulation studies were performed. The frame $\mathcal{F}_{B(t_0)}$ is arbitrary and in practice it can often be chosen as $\mathcal{F}_{B(t_0)} = \mathcal{F}_{oct}$. Therefore, the pose ${}^{oct}T_{B(t_0)}$, is

Algorithm 1 Shape and motion estimation algorithm

```
Require: Trajectory    ▹ 2D Rosette scan trajectory points
Require: $N = 200$    ▹ Population size
Require: $MaxIter = 100$    ▹ Maximum generations
Require: $Threshold = 10^{-6}$    ▹ Function tolerance
Require: $MaxStall = 50$    ▹ Stall generations limit
Require: $EliteCount = 10$, $CrossoverFraction = 0.8$
Require: Selection: stochastic uniform, Crossover: scattered, Mutation: adaptive feasible
  Data acquisition:
  $R{-}scan, t_0, t'' \leftarrow$ Scan along Trajectory
  Estimation initialisation:
  $\{{}^{oct}P_j, \forall t_j \in t_0, t''\} \leftarrow$ Eliminate low intensity voxels from R-scan
  $S_{1,distorted} \leftarrow \{{}^{oct}P_k, \forall t_k \in t_0, t'\}$
  $S_{2,distorted} \leftarrow \{{}^{oct}P_l, \forall t_l \in t_0, t''\}$    ▹ Extract two sub-volumes from $R{-}scan$
  $\vec{V}, \vec{\Omega} \leftarrow \mathbf{0}, iter \leftarrow 0, stall \leftarrow 0$
  Optimization:    ▹ Genetic algorithm
  $\vec{V}_1, \vec{\Omega}_1 \leftarrow$ random initial guess    ▹ Seed one individual
  for $s \leftarrow 2..N$ do    ▹ Create remaining population randomly
      $\vec{V}_s, \vec{\Omega}_s \leftarrow$ random values
  end for
  repeat
      for $s \leftarrow 1..N$ do    ▹ Update the population
          Update $\vec{V}_s, \vec{\Omega}_s$ using genetic operations
      end for
      for $s \leftarrow 1..N$ do    ▹ Assess the new population
          ${}^{B(t_0)}T_{B(t)} \leftarrow \int \vec{V}_s, \vec{\Omega}_s \, dt$    ▹ from equation 7
          Compute $S_{1,corrected_s}$ from equation 15a
          Compute $S_{2,corrected_s}$ from equation 15b
          Compute $\epsilon_{a\,s}$ from equation 14
          Store $\vec{V}_s, \vec{\Omega}_s, S_{1,corrected_s}, S_{2,corrected_s}, \epsilon_{a\,s}$
      end for
      Preserve top $EliteCount$ individuals
      Choose $(\vec{V}, \vec{\Omega}, S_{1,corrected}, S_{2,corrected})$ with lowest $\epsilon_a$
      $iter++$
      Update $stall$    ▹ Increment if no improvement, else reset
  until $(\epsilon_a < Threshold)$ or $(iter > MaxIter)$ or $(stall > MaxStall)$
  if $(\epsilon_a < Threshold)$ then
      Best motion estimate $(\vec{V}^*, \vec{\Omega}^*) \leftarrow (\vec{V}, \vec{\Omega})$
      Corrected OCT volume $\leftarrow S_{1,corrected} \cup S_{2,corrected}$
  else
      Check validity of the solution
  end if
```

considered as the identity matrix and the objective function is simplified to:

$$(\vec{V}^{*}, \vec{\Omega}^{*}) = \underset{\{\vec{V},\vec{\Omega}\}}{argmin}\ \epsilon(\vec{V}, \vec{\Omega}) \tag{17}$$

The objective function in equation 13 and 16 were implemented in MATLAB and the genetic algorithm function *ga* selected for the optimisation process. The genetic algorithm can be applied to solve problems that are not well suited for standard optimisation algorithms, particularly problems in which the objective function is discontinuous, non-differentiable, stochastic or highly non-linear [55] as is the case in our objective function. To assess the stability as well as the consistency of the optimisation process, and consequently the motion estimation, random design of experiments were conducted [56].

3.1. 3D motion estimation from OCT volumes with motion artefacts

The following steps were taken to implement the random design of experiment:

- → The minimum and maximum boundaries for the instantaneous linear velocities in the x, y and z axes were selected. The lower bound was $-3mm/s$ and upper bound $3mm/s$. Uniform random instantaneous linear velocities to introduce distortions in an OCT volume were then generated with respect to the mentioned bounds for 50 runs.
- → The minimum and maximum boundaries for the instantaneous angular velocities in the x, y and z axes were selected. The lower bound was $-0.03rad/s$ and upper bound $0.03rad/s$. Uniform random instantaneous angular velocities to introduce distortions in an OCT volume were then generated with respect to the mentioned bounds for 50 runs.
- → Initial values for the optimal motion estimation process were also randomly generated for the linear and angular velocities with respect to the mentioned bounds in the previous two steps.
- → The motion estimation process was then run until a convergence threshold was met (if the error ϵ between successive generations fell below 10^{-6}), but bounded by a maximum of 100 iterations. Optimal values for the linear and angular velocities were then obtained.

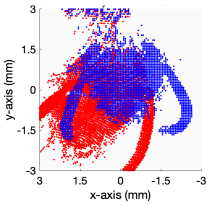
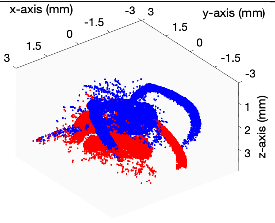
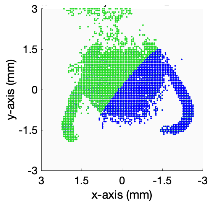
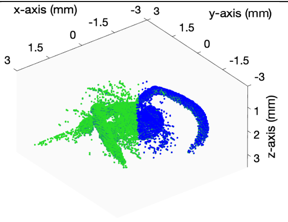
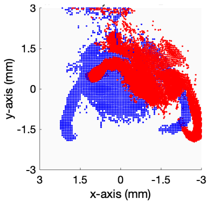
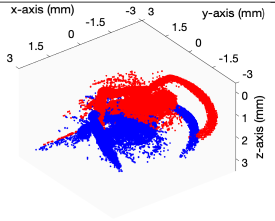
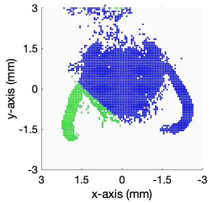
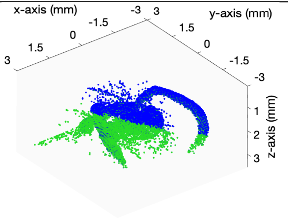
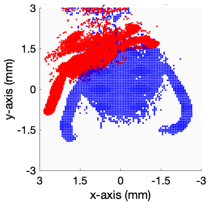
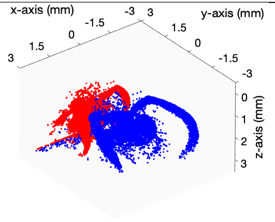
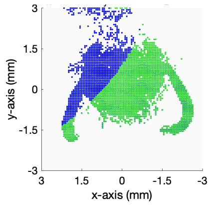
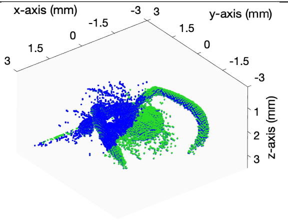
Furthermore, Gaussian noise with a mean of 0 and a standard deviation of 10 was introduced to 50% of the data. This noise was applied to the intensity value of each voxel in the distorted OCT volume.

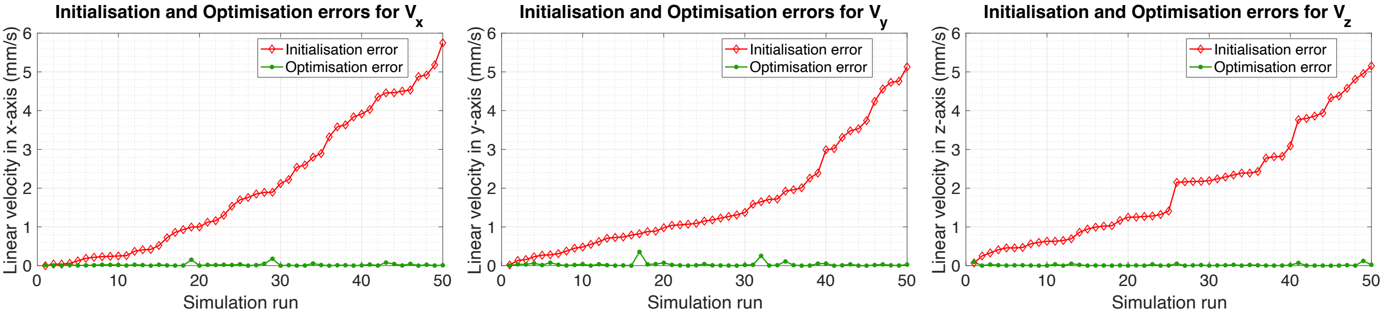
Table 1 shows the OCT volumes with motion artefacts and the corrected OCT volumes for some of the optimisation runs. The linear and angular velocities that were used to simulate the motion distorted volumes are provided as well as the estimated velocities after the motion estimation process.

3.1.1. Qualitative analysis

From the columns of distorted volumes (top view and 3D view) in Table 1, we see the significant effects of motion artefacts. In most of the images, the distorted OCT volume is barely recognisable from the motion-free OCT volume (see for instance the third and fourth row, the second and third column in Table 1). However after motion estimation and later on correction, the volumes are closely similar to the motion-free OCT volume (as illustrated in the fifth and sixth column of Table 1). These results show that the proposed motion estimation method works.

Table 1: Results from a few motion estimation runs (blue points - reference volume, red points - motion-distorted volume, green points - corrected volume). The units for the linear velocities, V, are in mm/s and angular velocities, Ω, are in $1e-2rad/s$

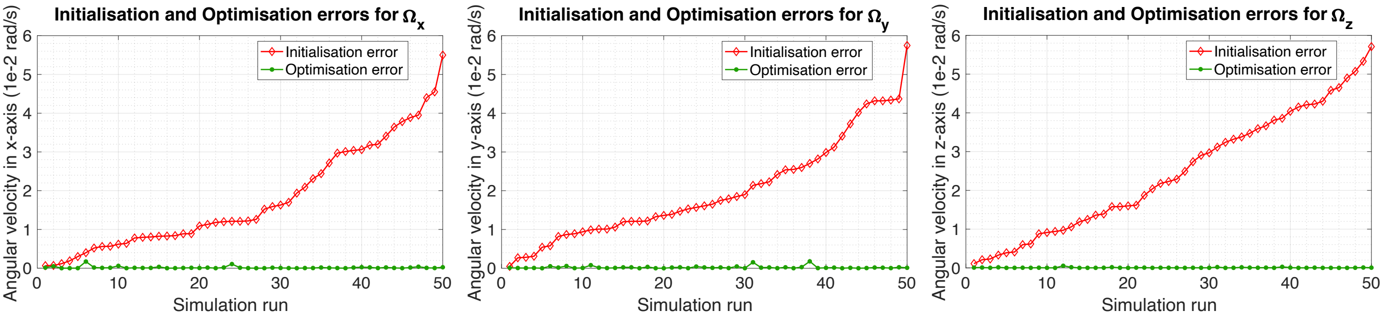
Applied motion	Reference and motion-distorted volume (top view)	Reference and motion-distorted volume (3D view)	Estimated applied motion	Reference and corrected volume (top view)	Reference and corrected volume (3D view)
$V_x = -1.77$ $V_y = 2.05$ $V_z = -2.97$ $\Omega_x = -1.39$ $\Omega_y = -1.46$ $\Omega_z = -1.26$			$V_x = -1.79$ $V_y = 2.03$ $V_z = -2.97$ $\Omega_x = -1.40$ $\Omega_y = -1.45$ $\Omega_z = -1.26$		
$V_x = 2.58$ $V_y = 1.18$ $V_z = 2.69$ $\Omega_x = 2.22$ $\Omega_y = 0.89$ $\Omega_z = -2.19$			$V_x = 2.58$ $V_y = 1.19$ $V_z = 2.69$ $\Omega_x = 2.23$ $\Omega_y = 0.89$ $\Omega_z = -2.19$		
$V_x = -1.88$ $V_y = -2.91$ $V_z = 0.95$ $\Omega_x = 1.81$ $\Omega_y = -1.13$ $\Omega_z = 1.37$			$V_x = -1.86$ $V_y = -2.88$ $V_z = 0.92$ $\Omega_x = 1.83$ $\Omega_y = -1.15$ $\Omega_z = 1.37$		



a Error for linear velocity in the $x-axis$

b Error for linear velocity in the $y-axis$

c Error for linear velocity in the $z-axis$



d Error for angular velocity in the $x-axis$

e Error for angular velocity in the $y-axis$

f Error for angular velocity in the $z-axis$

Fig. 8: Optimisation errors

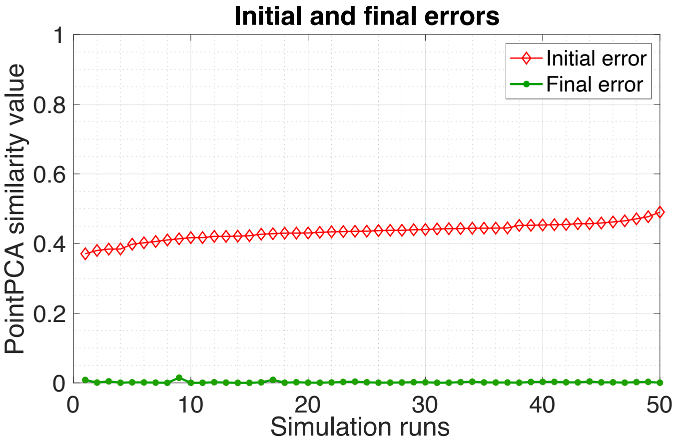


Fig. 9: Initial vs final (after optimisation) PointPCA error

3.1.2. *Quantitative analysis*

The absolute difference between the expected and the initial linear and angular velocities was computed for each axis; in addition, the difference between the expected and the final optimised values was also computed. These errors are plotted in Fig 8, in the order of increasing initialisation error. From Fig 8, we observe a steady optimisation error of almost zero in all the axes for all of the simulation runs even though the initialisation error increases. This validates that the proposed motion estimation method is robust to the initialisation guesses.

Moreover, both the initial and final PointPCA similarity error value, ϵ that is minimised during the optimisation process is plotted as illustrated in Fig 9. We observe that the similarity error after optimisation is nearly zero in all the runs. This is indicative of the accurate motion estimation from an OCT volume having geometric motion artefacts.

3.2. *3D shape and motion estimation from OCT volumes with motion artefacts*

For this method, the following steps were followed to implement the random design of experiment:

- → The minimum and maximum boundaries for the instantaneous linear velocities in the x, y and z axes were selected. The lower bound was $-2mm/s$ and upper bound $2mm/s$. Uniform random instantaneous linear velocities to introduce distortions in an OCT volume were then generated with respect to the mentioned bounds for 50 runs.

- → The minimum and maximum boundaries for the instantaneous angular velocities in the x, y and z axes were selected. The lower bound was $-0.02rad/s$ and upper bound $0.02rad/s$. Uniform random instantaneous angular velocities to introduce distortions in an OCT volume were then generated with respect to the mentioned bounds for 50 runs.

- → Initial values to the optimal motion estimation process were also randomly generated for the linear and angular velocities with respect to the mentioned bounds in the previous two steps.

- → The motion estimation process was then run until a convergence threshold was met (if the error ϵ_a between successive generations fell below 10^{-6}), but bounded by a maximum of 100 iterations. Optimal values for the linear and angular velocities were then obtained.

Table 2 highlights a comparison between the distorted OCT volumes containing motion artefacts and the corresponding corrected OCT volumes for a selection of the optimisation runs. The table includes information about the linear and angular velocities used to simulate the distorted volumes and the estimated velocities obtained through the motion estimation process.

3.2.1. *Qualitative analysis*

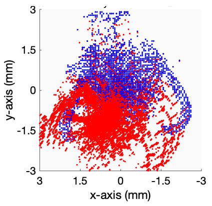
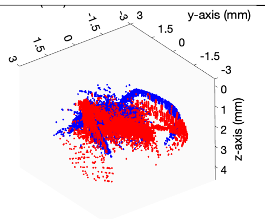
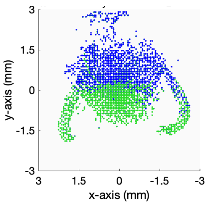
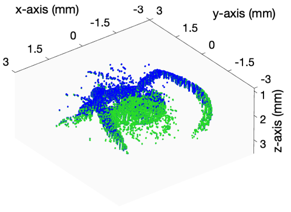
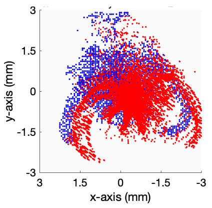
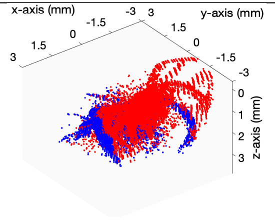
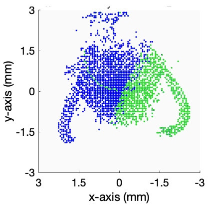
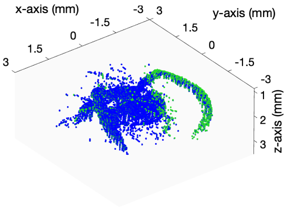
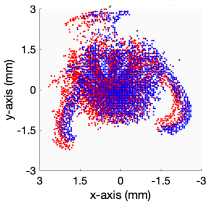
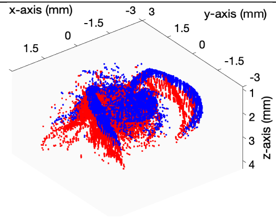
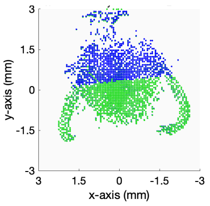
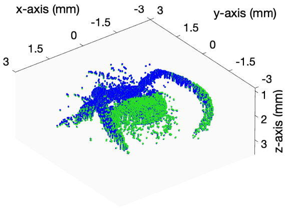
Upon examining the columns displaying the motion-distorted volumes from both top and 3D views, the pronounced impact of motion artefacts is evident. In most cases, the distorted OCT volume bears little resemblance to the motion-free OCT volume. However, following the motion estimation and subsequent correction, the volumes closely resemble the motion-free OCT volume. For instance, in the second and third row of Table 2, the motion-distorted OCT volumes exhibit a significant deterioration compared to the motion-free volume. Despite this considerable geometric distortion, the 3D motion in $SE(3)$ that caused the distortion is estimated, and a correction applied. Upon comparing the corrected volume with the motion-free volume, they closely resemble each other as seen in the fifth and sixth columns of Table 2. These results convincingly demonstrate the effectiveness of the proposed shape and motion estimation method from just one distorted OCT volume obtained using the rosette scan trajectory.

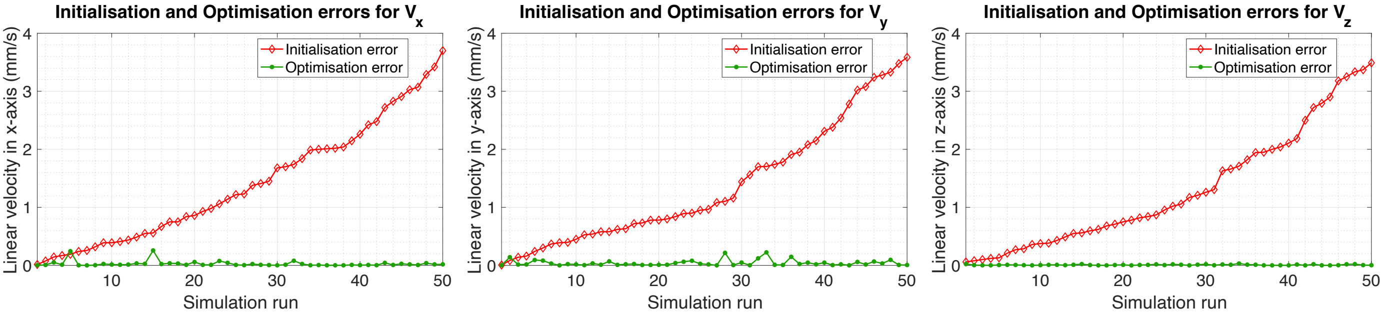
3.2.2. *Quantitative analysis*

The absolute difference between the expected and initial linear and angular velocities were calculated individually for each axis, and likewise, the discrepancies between the expected and the final optimised values were computed. These deviations were then graphically illustrated in Fig 10, sorted in ascending order based on the initial error magnitude. Fig 10 demonstrates a consistent final error that approaches nearly zero in all axes across all the simulation runs, despite the initial error increasing. This observation affirms the robustness of the proposed motion estimation method to variations in the initial guess values.

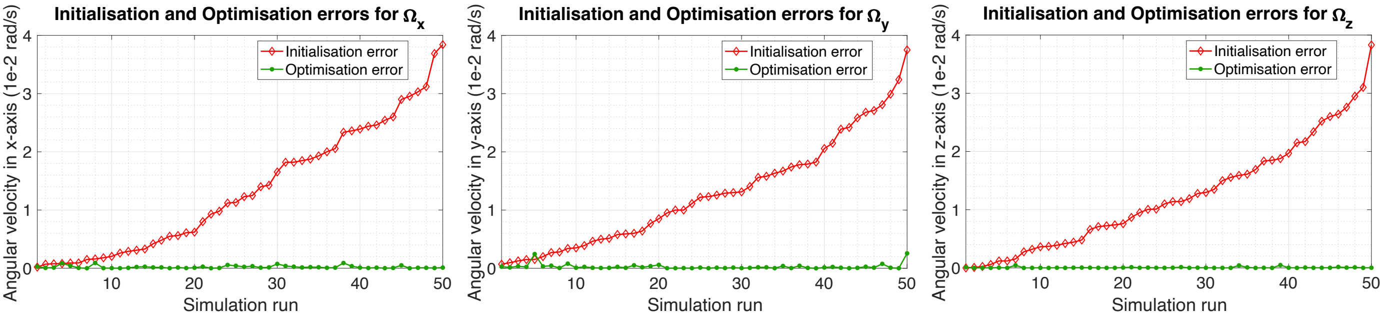
On plotting both the initial and final PointPCA similarity error value, ϵ_a that is minimised during the optimisation process as illustrated in Fig 11, we observe that the similarity error after optimisation is nearly zero in all the runs. This is indicative of the accurate shape and motion estimation from one OCT volume obtained using the rosette scan trajectory and having motion artefacts.

Table 2: Results from a few motion estimation runs from a single distorted OCT volume (blue points - motion-free volume, red points - motion-distorted volume, green points - corrected volume). The units for the linear velocities, V, are in mm/s and angular velocities, Ω, are in $1e-2rad/s$

Applied motion	Motion-free and motion-distorted volume (top view)	Motion-free and motion-distorted volume (3D view)	Estimated applied motion	Motion-free and corrected volume (top view)	Motion-free and corrected volume (3D view)
$V_x = 0.70$ $V_y = 1.36$ $V_z = -0.68$ $\Omega_x = -1.90$ $\Omega_y = -1.05$ $\Omega_z = -1.40$			$V_x = 0.69$ $V_y = 1.42$ $V_z = -0.66$ $\Omega_x = -1.87$ $\Omega_y = -1.04$ $\Omega_z = -1.40$		
$V_x = 0.07$ $V_y = 0.03$ $V_z = 1.88$ $\Omega_x = -1.02$ $\Omega_y = 1.17$ $\Omega_z = 0.22$			$V_x = 0.08$ $V_y = 0.05$ $V_z = 1.89$ $\Omega_x = -1.01$ $\Omega_y = 1.16$ $\Omega_z = 0.22$		
$V_x = -1.90$ $V_y = -0.27$ $V_z = -1.59$ $\Omega_x = -0.09$ $\Omega_y = 0.17$ $\Omega_z = 0.57$			$V_x = -1.88$ $V_y = -0.15$ $V_z = -1.61$ $\Omega_x = -0.03$ $\Omega_y = 0.16$ $\Omega_z = 0.57$		



a Error for linear velocity in the $x - axis$ b Error for linear velocity in the $y - axis$ c Error for linear velocity in the $z - axis$



d Error for angular velocity in the $x - axis$ e Error for angular velocity in the $y - axis$ f Error for angular velocity in the $z - axis$

Fig. 10: Optimisation errors

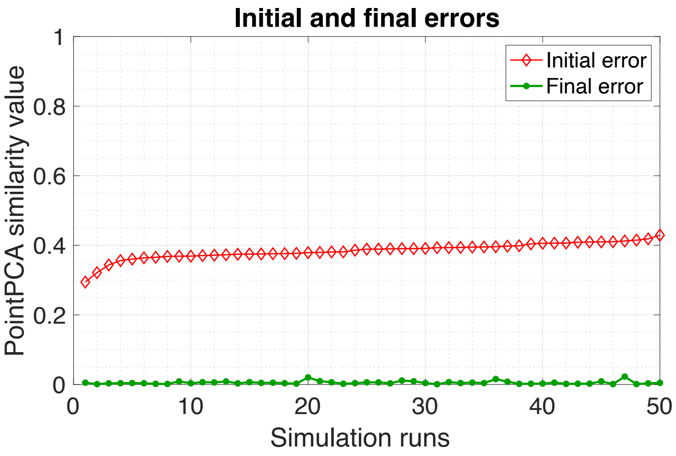


Fig. 11: Initial vs final (after optimisation) PointPCA error

4. Experimental validation

For the experimental validation, an OCT device was configured in an eye-to-hand setup on a parallel robotic platform, as depicted in Fig 12. The biological sample was positioned on the platform during the experiments. Translation and rotary motions were then introduced to the object during data acquisition using the parallel robot. This setup enabled the execution of realistic motion scenarios and facilitated an accurate assessment of the effectiveness of the motion estimation method.

4.1. Experimental Setup Description

To obtain OCT distorted volumes due to motion, the OCT device used is the Telesto II 1325nm spectral domain from Thorlabs as shown in Fig 12. It provides 1D depth (A-Scan), 2D cross-section (B-Scan) and 3D volumetric (C-Scan) images with 5.5μm axial resolution, 7μm lateral resolution and 3.54mm depth penetration. It also provides a so-called free-form B-scan mode, where A-scans are acquired along a trajectory of any shape. This mode was used to scan the volume along a rosette trajectory (R-scan). The device has a maximum A-Scan (optical core) acquisition rate of 76kHz with a maximum field-of-view of $10 \times 10 \times 3.54 mm^3$. The object to be scanned is placed on a fixed robotic platform; a 6-DoF 3PPSR parallel robot from Physical Instruments (SpaceFAB SF-3000 BS). It has the following characteristics: a maximum space reachable in translation of $(t_x, t_y, t_z) = (50mm, 100mm, 12.7mm)$ and a maximum space reachable in rotation of $(r_x, r_y, r_z) = (10°, 10°, 10°)$. Its linear resolution is 0.2μm with a repeatability of ±0.5μm and its angular resolution is 0.0005° with a repeatability of ±0.0011°. Two computers are used to carry out the experiments; the first (a 3.70 GHz Intel Xeon W-2145 CPU with 32Gb RAM and Windows 10 OS) is used for OCT data acquisition as well as data processing, and the second (a 3.20 GHz Intel Core CPU with 4Gb RAM and Windows 7 OS) is dedicated for the control of the parallel robot. These two computers communicate with each other asynchronously via the TCP/IP protocol.

4.2. 3D motion estimation from experimentally obtained OCT volumes with motion artefacts

For our experiments, the field of view was constrained in the $x - y$ plane to $6mm \times 6mm$ and a depth of 3.54mm along the z axis. The number of scan positions was 10000 and each A-scan contained 1024 elements, each encoded on 8 bits. Therefore, the C-scans had 10000×1024 voxels. The acquisition time of a C-scan was about 350ms. This time includes the initialization of the scanner and the actual scan time, upon which we have no control (the OCT system being commercial, we did not have access to the low-level control of the scanner needed to optimise the scan time). The processed point cloud was then thresholded to filter out low intensity voxels.

As in simulation, we split here the analysis into a qualitative, and then a quantitative one.

4.2.1. Qualitative Analysis

Table 3 shows the experimental OCT volumes with motion artefacts and the corrected OCT volumes. From the qualitative analysis results, it is observed that the corrected OCT volumes closely resemble the motion-free volumes. Notably, for the experimentally distorted volume (second and third column) and subsequently in the second and third rows of Table 3, some details, such as part of the antennae of the bee-head, are truncated and elongated respectively. Nonetheless, after motion estimation and correction, the corrected volume aligns and recovers the shape of the motion-free OCT volume, as observed in the sixth column of Table 3. Also, despite the experimentally distorted volume containing unwanted data (robot platform) as observed in the fourth row, second and third column of Table 3, the motion causing the distortion is able to be estimated, and the shape recovered, as seen in the fifth and sixth column.

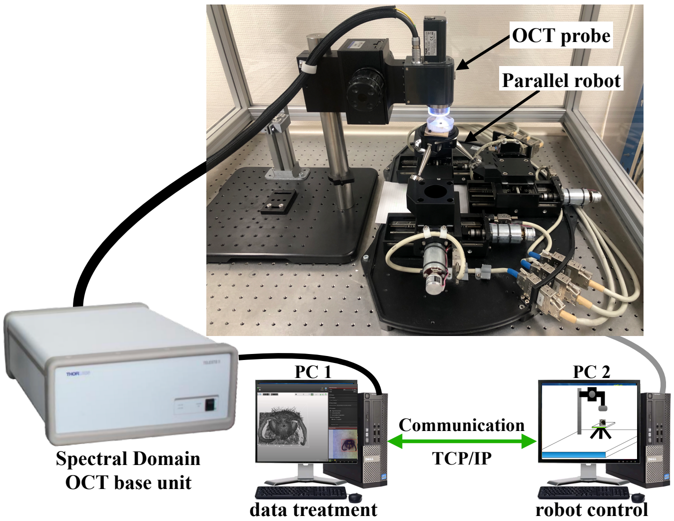
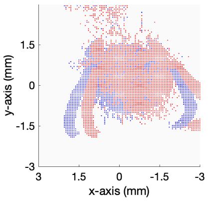
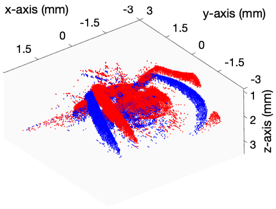
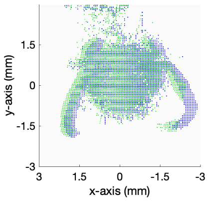
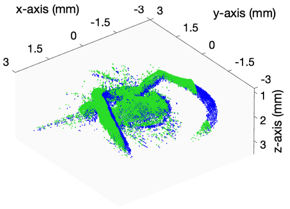
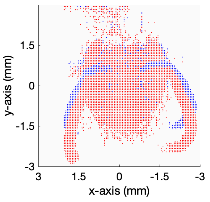
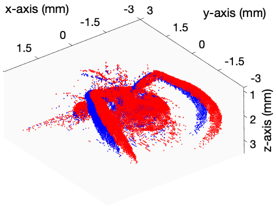
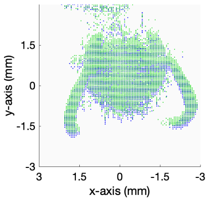
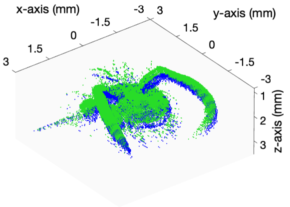
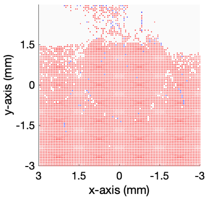
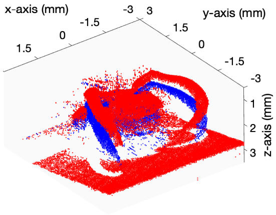
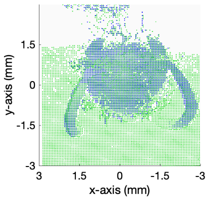
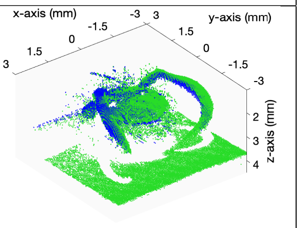


Fig. 12: Illustration of the experimental setup used to execute motions during OCT volume data acquisition

Table 3: Results from the motion estimation process using experimental data (blue points - reference volume, red points - motion-distorted volume, green points - corrected volume). The units for the linear velocities, V, are in mm/s and angular velocities, Ω, are in $1e-2rad/s$

Applied motion	Reference and motion-distorted volume (top view)	Reference and motion-distorted volume (3D view)	Estimated applied motion	Reference and corrected volume (top view)	Reference and corrected volume (3D view)
$V_x = 3.00$ $V_y = 0.00$ $V_z = 0.00$ $\Omega_x = 0.00$ $\Omega_y = 0.00$ $\Omega_z = 0.00$			$V_x = 2.79$ $V_y = 0.00$ $V_z = 0.00$ $\Omega_x = 0.00$ $\Omega_y = 0.00$ $\Omega_z = 0.00$		
$V_x = 0.00$ $V_y = 3.00$ $V_z = 0.00$ $\Omega_x = 0.00$ $\Omega_y = 0.00$ $\Omega_z = 0.00$			$V_x = 0.00$ $V_y = 2.82$ $V_z = 0.00$ $\Omega_x = 0.00$ $\Omega_y = 0.00$ $\Omega_z = 0.00$		
$V_x = 0.00$ $V_y = 0.00$ $V_z = 3.00$ $\Omega_x = 0.00$ $\Omega_y = 0.00$ $\Omega_z = 0.00$			$V_x = 0.00$ $V_y = 0.00$ $V_z = 2.52$ $\Omega_x = 0.00$ $\Omega_y = 0.00$ $\Omega_z = 0.00$		

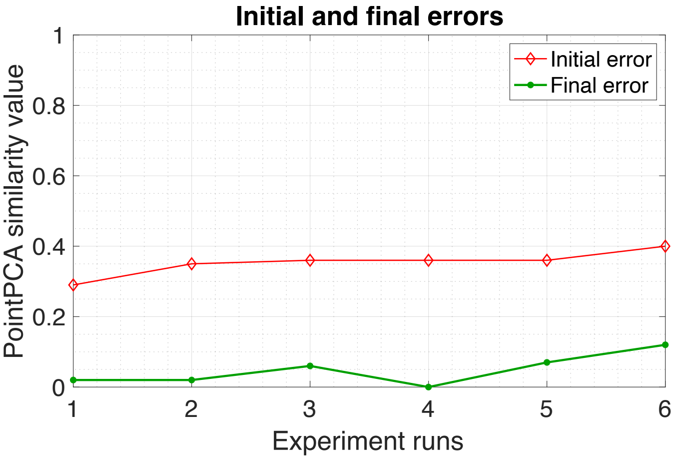


Fig. 13: Initial vs final (after optimisation) PointPCA error

4.2.2. *Quantitative Analysis*

Both the initial and final PointPCA similarity error value, ϵ (obtained using equation 10) that is minimised during the motion estimation process for the experimentally motion distorted volumes is plotted as illustrated in Fig 13. We observe that the similarity error after optimisation is minimised for all the experimental runs. However, there are instances when the final similarity error is not zero. Reasons for this discrepancy may be due to;

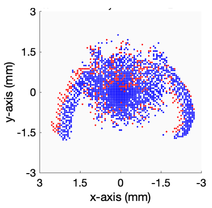
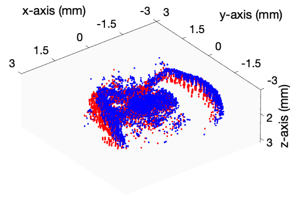
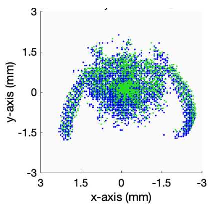
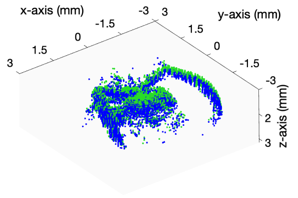
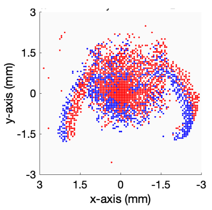
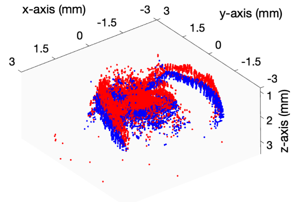
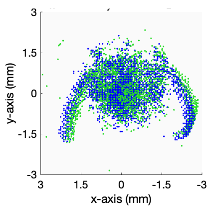
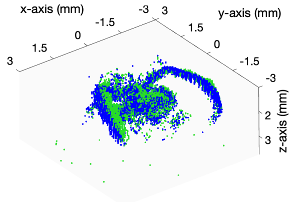
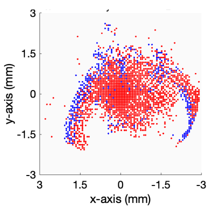
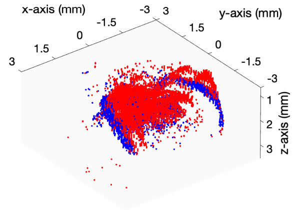
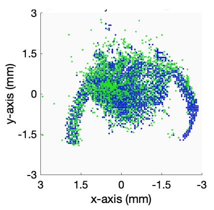
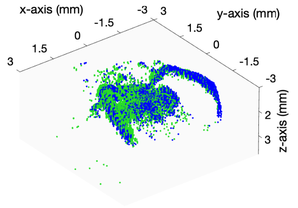
- ✗ Occurrences of missing data. This may occur when the object moves out of the field of view, resulting in missing data within the obtained volume.
- ✗ Introduction of unnecessary data. The OCT device may capture data from undesired regions, as seen in the fourth row of Table 3.
- ✗ The numerical optimisation process getting stuck at a local minimum.
- ✗ Modelling errors.

Nevertheless, despite these challenges, the analysis demonstrates that the motion estimation works fairly well.

4.3. *3D shape and motion estimation from experimentally obtained OCT volumes with motion artefacts*

To perform experimental validation, the experimental setup as detailed in sub-section 4.1 was used. Table 4 shows experimentally distorted OCT volumes obtained using the rosette scan trajectory and the corrected OCT volumes after shape and motion estimation.

Table 4: Results from a few motion estimation runs from a single distorted OCT volume using experimental data (blue points - motion-free volume, red points - motion-distorted volume, green points - corrected volume). The units for the linear velocities, V, are in mm/s and angular velocities, Ω, are in $1e-2rad/s$

Applied motion	Motion-free and motion-distorted volume (top view)	Motion-free and motion-distorted volume (3D view)	Estimated applied motion	Motion-free and corrected volume (top view)	Motion-free and corrected volume (3D view)
$V_x = -1.90$ $V_y = -0.27$ $V_z = -1.59$ $\Omega_x = -0.09$ $\Omega_y = 0.17$ $\Omega_z = 0.57$			$V_x = -1.81$ $V_y = -0.21$ $V_z = -1.81$ $\Omega_x = -0.27$ $\Omega_y = 0.19$ $\Omega_z = -0.04$		
$V_x = 1.65$ $V_y = -1.52$ $V_z = 1.88$ $\Omega_x = -0.91$ $\Omega_y = 0.94$ $\Omega_z = 1.34$			$V_x = 1.38$ $V_y = -1.50$ $V_z = 1.49$ $\Omega_x = -0.52$ $\Omega_y = 0.90$ $\Omega_z = 1.04$		
$V_x = 2.00$ $V_y = 2.00$ $V_z = 2.00$ $\Omega_x = 2.00$ $\Omega_y = 2.00$ $\Omega_z = 2.00$			$V_x = 2.17$ $V_y = 2.05$ $V_z = 2.28$ $\Omega_x = 2.37$ $\Omega_y = 2.14$ $\Omega_z = 1.51$		

4.3.1. *Qualitative Analysis*

Qualitatively analysing the results in Table 4, we observe that the corrected OCT volume closely resemble the motion-free volumes. Notably, in the experimentally motion-distorted volume (third column) and subsequently in the second and third rows of Table 4, the antennae of the bee-head is skewed from the motion-free volume. However, after motion estimation and correction, the corrected volume recovers the shape and the antennae is aligned to the motion-free OCT volume as observed in the sixth column of Table 4. Similarly, in Table 4, in the fourth row, third column, the bee-head appears to be significantly distorted. However, in the third row, in the sixth column of Table 4, the corrected volume recovers the shape.

4.3.2. *Quantitative Analysis*

The initial and final PointPCA similarity error value, ϵ_a (obtained using equation 14) that is minimised during the motion estimation process for the experimentally motion-distorted volumes is plotted as illustrated in Fig 14. It is observed that the similarity error after optimisation is minimised for all the runs, however, the final similarity error is not zero in all instances. This discrepancy in accurate motion estimation likely arises due to;

- ✗ Noisy experimental OCT data (occurrences of missing or introduction of unnecessary data).
- ✗ The numerical optimisation process getting stuck at a local minimum.
- ✗ Modelling errors.

Nonetheless, despite these challenges, the analysis demonstrates that the 3D shape and motion estimation works fairly well.

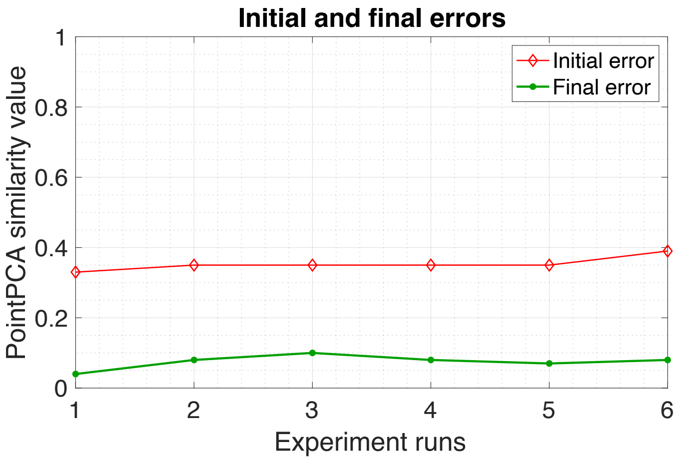


Fig. 14: Initial vs final (after optimisation) PointPCA error

5. Computation time and efficiency

Our method is implemented in MATLAB and runs primarily single-threaded. The dominant cost is the genetic-algorithm search over the six motion parameters; each evaluation computes the point-cloud features and the similarity metric before applying the correction transform. Using a maximum of 100 generations and early stopping via a convergence threshold of 10^{-6}, the end-to-end runtime was typically tens of minutes per volume on a standard workstation (Intel Xeon W-2145, 3.70 GHz; 32 GB RAM; Windows 10). This is adequate for offline processing or assisted workflows (e.g., applying the correction between scans or during a procedural pause).

The method itself is not intrinsically compute-heavy beyond repeated similarity evaluations, and there is substantial headroom for acceleration. The number of optimisation variables is six, so the problem is a low-dimensional one, even though the loss function is costly. Translating core routines from MATLAB to optimized C/C++ and/or leveraging GPU kernels for feature computation and fitness evaluation, together with parallel evaluation of genetic algorithm populations, should reduce runtimes dramatically - plausibly to seconds per volume on modern hardware. Such engineering optimisations would make the approach compatible with more time-critical scenarios (e.g., intraoperative robotic guidance), where even 30–60 s latency may be too slow. Current timings reflect a feasibility-focused approach; the pathway to near-real-time performance is primarily an implementation-efficiency problem rather than a methodological limitation.

6. Limitations

6.1. System-level

We conducted our experiments on a semi-open commercial system. In particular, the scan hardware consists of two galvanometer scanners for the $x-axis$ and $y-axis$, which we did not alter. As we pointed out earlier, rather than using two galvanometric scanners, an optimal design with one rotating axis and one radial axis could be implemented. In addition, a design using only one actuator to trace the rosette trajectory could be developed. Such a further design would likely consume less energy and occupy less space, since only one actuator would be needed instead of two. In addition to the hardware limitation, we did not have access to the low-level control of the scanner. While we were able to use a free-form B-scan mode to obtain A-scans along the rosette scan trajectory, we could not optimise the scan time.

6.2. Velocity model

A constant-velocity model with the same velocity on both sub-volumes is not always valid. A straightforward first extension is to optimise two different Cartesian velocity vectors, one for each sub-volume. This would double the number of unknowns from six to twelve while maintaining a very reasonable dimensionality for the numerical problem. Furthermore, the constant-velocity model could be replaced by a piecewise constant velocity model as in [57], or by a piecewise constant acceleration (or jerk) model to account for varying velocities during acquisition. If this is not sufficient, one could use a parametric model $V, \Omega = f(t, \vec{p})$ to handle even more complex motion. For instance, sine waves could be used for periodic motions; a function basis (such as Fourier series or wavelets) to capture arbitrary irregular motion; or any model dedicated to a specific domain. In these cases, the optimisation in equation 16 would become

$$\min_{\vec{p}, {}^{oct}T_{B(t_0)}} \quad \epsilon_a(\vec{V}(t, \vec{p}), \vec{\Omega}(t, \vec{p}), {}^{oct}T_{B(t_0)}) \tag{18}$$

where $\vec{p}$ is the set of parameters of the function defining the velocity.

6.3. Singular configurations and reconstruction ambiguities

Singular configurations are unlikely to occur as long as the number of optimisation variables remains low and the observed shape is geometrically complex enough. However, as Lao *et al.,* [24] showed in one of their papers, there may be an ambiguity between shape and motion (which can be considered a singular case) when reconstructing them in a stereoscopic rolling shutter setup. In our case, the risk is reduced because the volume is scanned twice, not simultaneously (two rolling shutter views from different viewpoints at the same time, as in [24]), but sequentially (two views from different viewpoints at different times).

Local minima are very likely to occur. In fact, because of high-frequency noise, a simple geometric difference between the two reconstructed sub-volumes would be full of them. The use of volume similarity smooths out most of the noise, but not completely. This is why genetic algorithms perform better than gradient-based methods in our case.

7. Conclusion and perspectives

This paper introduces a shape and motion estimation approach based on a single OCT volume (obtained using the non-Cartesian rosette scan trajectory) and without prior knowledge of the shape. This represents a novel strategy that departs from the conventional method of compensating for OCT volume motion artefacts, which would typically rely on two sequential OCT volumes. By eliminating the need for consecutive volumes, this approach reduces data requirements and consequently, computation costs. The methods feasibility have been numerically and experimentally demonstrated and validated. Our contributions mark a significant advancement over existing techniques, which are primarily limited to 2D OCT images and/or largely constrained by the specific context of ophthalmology. Moreover, the proposed shape and motion estimation method opens the way to robot-assisted motion tracking, robot-assisted optical biopsy, and consequently, image stabilisation during robot-assisted surgery. A number of limitations were identified, and perspectives on how to circumvent them were outlined. Among these, two main ones can be recalled here. In this research, for 3D shape and motion estimation from one OCT volume, only two sub-volumes were used. However, depending on the design of the scan trajectory, more than two sub-volumes could be extracted. Incorporating these extra sub-volumes into the existing methodology could improve the accuracy in estimating the motion, consequently enabling to free certain assumptions such as constant velocity. Other potential paths for future improvements relate to the motion model, which has been kept minimal in this study but could be made more complex to better fit physiology or to segment the sample shape and motion into consistent sub-components.

Acknowledgments

The INSERM ROBOT Project: [ITMO Cancer no 17CP068-00]; EIPHI Graduate School: [ANR-17-EURE-0002] and the Grand Prix Scientifique 2018, Fondation Charles Defforey, Institut de France.

References

[1] B. J. Vakoc, D. Fukumura, R. K. Jain, B. E. Bouma, Cancer imaging by optical coherence tomography: Preclinical progress and clinical potential, Nature Reviews Cancer 12 (5) (2012) 363–368. doi:10.1038/nrc3235.

[2] T.-H. Tsai, J. Fujimoto, H. Mashimo, Endoscopic Optical Coherence Tomography for Clinical Gastroenterology (2014). doi:10.3390/diagnostics4020057.

[3] J. Wang, Y. Xu, S. A. Boppart, Review of optical coherence tomography in oncology, Journal of Biomedical Optics 22 (12) (2017) 1. doi:10.1117/1.jbo.22.12.121711.

[4] T.-H. Tsai, C. L. Leggett, A. J. Trindade, Optical coherence tomography in gastroenterology: a review and future outlook, Journal of Biomedical Optics 22 (12) (2017). doi:10.1117/1.jbo.22.12.121716.

[5] M. J. Gora, M. J. Suter, G. J. Tearney, X. Li, Endoscopic optical coherence tomography: technologies and clinical applications [Invited], Biomedical Optics Express 8 (5) (2017) 2405. doi:10.1364/boe.8.002405.

[6] L. Van Manen, J. Dijkstra, C. Boccara, E. Benoit, A. L. Vahrmeijer, M. J. Gora, J. S. D. Mieog, The clinical usefulness of optical coherence tomography during cancer interventions, Journal of Cancer Research and Clinical Oncology 144 (10) (2018) 1967–1990. doi:10.1007/s00432-018-2690-9.

[7] Z. Yang, J. Shang, C. Liu, J. Zhang, Y. Liang, Identification of oral cancer in OCT images based on an optical attenuation model, Lasers in Medical Science 35 (9) (2020) 1999–2007. doi:10.1007/s10103-020-03025-y.

[8] M. E. Brezinski, G. J. Tearney, S. A. Boppart, E. A. Swanson, J. F. Southern, J. G. Fujimoto, Optical biopsy with optical coherence tomography: feasibility for surgical diagnostics, The Journal of surgical research 71 (1) (1997) 32–40. doi:10.1006/jsre.1996.4993.

[9] S. R. Chinn, E. A. Swanson, J. G. Fujimoto, Optical coherence tomography using a frequency-tunable optical source, Opt. Lett. 22 (5) (1997) 340–342. doi:10.1364/OL.22.000340.

[10] B. Golubovic, B. E. Bouma, G. J. Tearney, J. G. Fujimoto, Optical frequency-domain reflectometry using rapid wavelength tuning of a Cr4+:forsterite laser, Opt. Lett. 22 (22) (1997) 1704–1706. doi:10.1364/OL.22.001704.

[11] M. Szkulmowski, M. Wojtkowski, T. Bajraszewski, I. Gorczyńska, P. Targowski, W. Wasilewski, A. Kowalczyk, C. Radzewicz, Quality improvement for high resolution in vivo images by spectral domain optical coherence tomography with supercontinuum source, Optics Communications 246 (4-6) (2005) 569–578. doi:10.1016/j.optcom.2004.11.024.

[12] W. Drexler, Optical coherence tomography, Elsevier (2010).

[13] J. Fujimoto, W. Drexler, Introduction to OCT, in: W. Drexler, J. G. Fujimoto (Eds.), Optical Coherence Tomography, Technology and Applications, 2nd Edition, Springer International Publishing, Vienna, 2015, pp. 3 – 64. doi:DOI 10.1007/978-3-319-06419-2_1.

[14] J. F. Bille, S. W. Hell, R. N. Weinreb, High Resolution Imaging in Microscopy and Ophthalmology, Springer Nature Switzerland AG, 2019. doi:10.1007/978-3-030-16638-0.

[15] J. A. Izatt, M. A. Choma, Optical Coherence Tomography: Technology and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, Ch. Theory of Optical Coherence Tomography, pp. 47–72. doi:10.1007/978-3-540-77550-8_2.

[16] Y. Rao, N. P. Sarwade, R. Makkar, Modeling and simulation of Optical Coherence Tomography on Virtual OCT, in: Procedia Computer Science, Vol. 45, Elsevier B.V., 2015, pp. 644–650. doi:10.1016/j.procs.2015.03.121.

[17] L. Pan, F. Shi, D. Xiang, K. Yu, L. Duan, J. Zheng, X. Chen, Octrexpert: A feature-based 3D registration method for retinal OCT images, IEEE Transactions on Image Processing 29 (2020) 3885–3897. doi:10.1109/TIP.2020.2967589.

[18] P. E. Forssén, E. Ringaby, Rectifying rolling shutter video from handheld devices, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) (2010) 507–514doi:10.1109/CVPR.2010.5540173.

[19] J. Walther, A. Krüger, M. Cuevas, E. Koch, Effects of axial, transverse, and oblique sample motion in FD OCT in systems with global or rolling shutter line detector, Journal of the Optical Society of America 25 (11) (2008) 2791–2802. doi:10.1364/josaa.25.002791.

[20] C. K. Liang, L. W. Chang, H. H. Chen, Analysis and compensation of rolling shutter effect, IEEE Transactions on Image Processing 17 (8) (2008) 1323–1330. doi:10.1109/TIP.2008.925384.

[21] L. Magerand, A. Bartoli, A generic rolling shutter camera model and its application to dynamic pose estimation, in: 5th International Symposium 3D Data Processing, Visualization and Transmission (3DPVT), 2010.

[22] O. Ait-Aider, A. Bartoli, N. Andreff, Kinematics from lines in a single rolling shutter image, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2007. doi:10.1109/CVPR.2007.383119.

[23] W.-h. Cho, D.-W. Kim, K.-S. Hong, CMOS Digital Image Stabilization, IEEE Trans. on Consum. Electron. 53 (3) (2007) 979–986. doi:10.1109/TCE.2007.4341576.

[24] Y. Lao, O. Ait-Aider, A. Bartoli, Solving Rolling Shutter 3D Vision Problems using Analogies with Non-rigidity, International Journal of Computer Vision 129 (2020) 100–122. doi:10.1007/s11263-020-01368-1.

[25] R. Dahmouche, N. Andreff, Y. Mezouar, O. Ait-Aider, P. Martinet, Dynamic visual servoing from sequential regions of interest acquisition, International Journal of Robotics Research 31 (4) (2012) 520–537. URL `http://dx.doi.org/10.1177/0278364911436082`

[26] D. Ferguson, D. Hamer, L. Paunescu, B. Siobahn, S. Joel, Tracking optical coherence tomography, Optics letters 29 (18) (2004) 2139–2141. doi:10.1364/ol.29.002139.

[27] K. V. Vienola, B. Braaf, C. K. Sheehy, Q. Yang, P. Tiruveedhula, D. W. Arathorn, J. F. de Boer, A. Roorda, Real-time eye motion compensation for OCT imaging with tracking SLO, Biomedical Optics Express 3 (11) (2012) 2950–2963. doi:10.1364/boe.3.002950.

[28] Y. Guo, R. Morishita, I. A. El-Sadek, P. Mukherjee, Y. Zhu, Y. Yasuno, In vivo dynamic optical coherence tomography with hardware- and software-based motion correction, in: R. A. Leitgeb, Y. Yasuno (Eds.), Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XXVIII, Vol. 12830, International Society for Optics and Photonics, SPIE, 2024, p. 128300Y. doi:10.1117/12.3005623.
URL `https://doi.org/10.1117/12.3005623`

[29] T. Klein, W. Wieser, R. Andre, T. Pfeiffer, E. C. M, R. Huber, Multi-MHz FDML OCT: snapshot retinal imaging at 6.7 million axial-scans per second, in: Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XVI, Vol. Xvi, SPIE, California, 2012, p. 8213. doi:10.1117/12.908798.

[30] M. Pircher, B. Baumann, E. Gotzinger, H. Sattmann, H. C. K, Simultaneous SLO/OCT imaging of the human retina with axial eye motion correction, Optics Express 15 (2007) 16922 – 16932.

[31] Z. Yaqoob, J. Wu, C. Yang, Spectral domain optical coherence tomography: a better OCT imaging strategy., BioTechniques 39 (6) (2005). doi:10.2144/000112090.

[32] R. J. Zawadzki, A. R. Fuller, S. S. Choi, D. F. Wiley, B. Hamann, J. S. Werner, Correction of motion artifacts and scanning beam distortions in 3D ophthalmic optical coherence tomography imaging, in: B. E. Stuck, M. B. M.D., F. Manns, P. G. Söderberg, A. Ho (Eds.), Ophthalmic Technologies XVII, Vol. 6426, International Society for Optics and Photonics, SPIE, 2007, p. 642607. doi:10.1117/12.701524.

[33] R. R. Pappuru, C. Briceno, Y. Ouyang, A. C. Walsh, S. R. Sadda, Clinical significance of B-scan averaging with SD-OCT, Ophthalmic Surgery Lasers and Imaging 43 (1) (2012) 63–68. doi:10.3928/15428877-20110908-02.

[34] S. Ricco, M. Chen, H. Ishikawa, G. Wollstein, J. Schuman, Correcting motion artifacts in Retinal Spectral Domain Optical Coherence Tomography via Image Registration, National institute of health 12 (1) (2009) 100–107. arXiv:NIHMS150003, doi:10.1007/978-3-642-04268-3_13.

[35] A. G. Capps, R. J. Zawadzki, Q. Yang, D. W. Arathorn, C. R. Vogel, B. Hamann, J. S. Werner, Correction of eye-motion artifacts in AO-OCT data sets, in: Proc. SPIE 7885, Ophthalmic Technologies XXI, 2011, p. 78850. doi:10.1117/12.874376.

[36] H. J. Kraus Martin, OCT Motion Correction, in: F. J. Drexler Wolfgang (Ed.), Optical Coherence Tomography, Technology and Applications, Springer International Publishing, 2015, pp. 459–476. doi:10.1007/978-3-319-06419-2_16.

[37] G. Liao, O. Caravaca-Mora, B. Rosa, P. Zanne, D. Dall'Alba, P. Fiorini, M. de Mathelin, F. Nageotte, M. J. Gora, Distortion and instability compensation with deep learning for rotational scanning endoscopic optical coherence tomography, Medical Image Analysis 77 (2022) 102355. doi:10.1016/j.media.2022.102355.

[38] S. Y. Ksenofontov, P. A. Shilyagin, V. M. Gelikonov, G. V. Gelikonov, Motion artifact suppression method for the clinical application of otoscopic spectral-domain optical coherence tomography, Photonics 10 (7) (2023). doi:10.3390/photonics10070736.
URL `https://www.mdpi.com/2304-6732/10/7/736`

[39] X. Chen, Z. Ma, J. Zhu, C. Wang, Fast motion correction in optical coherence tomography with image feature matching, in: Q. Luo, X. Li, Y. Gu, D. Zhu (Eds.), Optics in Health Care and Biomedical Optics XIII, Vol. 12770, International Society for Optics and Photonics, SPIE, 2023, p. 127701U. doi:10.1117/12.2686448.
URL `https://doi.org/10.1117/12.2686448`

[40] J. Ren, K. Park, Y. Pan, H. Ling, Self-supervised bulk motion artifact removal in optical coherence tomography angiography, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 20617–20625.

[41] Z. Lin, Q. Zhang, G. Lan, J. Xu, J. Qin, L. An, Y. Huang, Deep learning for motion artifact-suppressed octa image generation from both repeated and adjacent oct scans, Mathematics 12 (3) (2024). doi:10.3390/math12030446.
URL `https://www.mdpi.com/2227-7390/12/3/446`

[42] Y. Wang, C. Galang, W. R. Freeman, A. Warter, A. Heinke, D.-U. G. Bartsch, T. Q. Nguyen, C. An, Retinal oct layer segmentation via joint motion correction and graph-assisted 3d neural network, IEEE Access 11 (2023) 103319–103332. doi:10.1109/ACCESS.2023.3317011.

[43] B. Antony, M. D. Abràmoff, L. Tang, W. D. Ramdas, J. R. Vingerling, N. M. Jansonius, K. Lee, Y. H. Kwon, M. Sonka, M. K. Garvin, Automated 3-D method for the correction of axial artifacts in spectral-domain optical coherence tomography images, Biomedical Optics Express 2 (8) (2011) 2403. doi:10.1364/boe.2.002403.

[44] R. Zuo, K. Irsch, J. U. Kang, Higher-order regression three-dimensional motion-compensation method for real-time optical coherence tomography volumetric imaging of the cornea, Journal of Biomedical Optics 27 (06) (2022). doi:10.1117/1.jbo.27.6.066006.

[45] X. Zhang, H. Zhong, S. Wang, B. He, L. Cao, M. Li, M. Jiang, Q. Li, Subpixel motion artifacts correction and motion estimation for 3d-oct, Journal of Biophotonics 17 (9) (2024) e202400104. doi:https://doi.org/10.1002/jbio.202400104.

[46] S. S. Gao, Y. Jia, D. Huang, Artifacts in optical coherence tomography angiography, in: Optical Coherence Tomography Angiography of the Eye, CRC Press, 2024, pp. 31–38.

[47] J. Solà, J. Deray, D. Atchuthan, A micro Lie theory for state estimation in robotics (2018). arXiv:1812.01537.
URL `http://arxiv.org/abs/1812.01537`

[48] J. Deray, J. Solà, Manif: A micro Lie theory library for state estimation in robotics applications, Journal of Open Source Software 5 (46) (2020) 1371. doi:10.21105/joss.01371.

[49] M. F. Kraus, B. Potsaid, M. A. Mayer, R. Bock, B. Baumann, J. J. Liu, J. Hornegger, J. G. Fujimoto, Motion correction in optical coherence tomography volumes on a per A-scan basis using orthogonal scan patterns, Biomedical Optics Express 3 (6) (2012) 1182–1199. doi:10.1364/boe.3.001182.

[50] E. Alexiou, X. Zhou, I. Viola, P. Cesar, Pointpca: Point cloud objective quality assessment using pca-based descriptors, EURASIP Journal on Image and Video Processing 2024 (1) (2024) 20.

[51] N. O. Osinde, N. Andreff, Multicriteria assessment of optical coherence tomography using non-raster trajectories, Journal of microscopy 298 (1) (2025) 27–43. doi:10.1111/jmi.13383.

[52] N. Nikooienejad, M. Maroufi, S. O. Moheimani, Rosette-scan video-rate atomic force microscopy: Trajectory patterning and control design, Review of Scientific Instruments 90 (7) (2019) 1–10. doi:10.1063/1.5098499.

[53] T. R. Meyer, D. Ziegler, C. Brune, A. Chen, R. Farnham, N. Huynh, J. M. Chang, A. L. Bertozzi, P. D. Ashby, Height drift correction in non-raster atomic force microscopy, Ultramicroscopy 137 (2014) 48–54. doi:10.1016/j.ultramic.2013.10.014.

[54] A. M. Bush, C. M. Sandino, S. Ramachandran, F. Ong, N. Dwork, E. J. Zucker, A. B. Syed, J. M. Pauly, M. T. Alley, S. S. Vasanawala, Rosette Trajectories Enable Ungated, Motion-Robust, Simultaneous Cardiac and Liver T2* Iron Assessment, Journal of Magnetic Resonance Imaging 52 (6) (2020) 1688–1698. doi:10.1002/jmri.27196.

[55] D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, United States, 1989.

[56] T. Caliński, S. Kageyama, Block designs: A Randomization approach, Volume I: Analysis, Springer, New York, 2000. doi:10.1007/978-1-4612-1192-1.

[57] L. Magerand, A. Bartoli, O. Ait-Aider, D. Pizarro, Global optimization of object pose and motion from a single rolling shutter image with automatic 2d-3d matching, in: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (Eds.), Computer Vision – ECCV 2012, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 456–469. doi:10.1007/978-3-642-33718-5_33.