


Article

Concept for Markerless 6D Tracking Employing Volumetric Optical Coherence Tomography

Matthias Schlüter [†] , Lukas Glandorf [†], Martin Gromniak and Thore Saathoff and Alexander Schlaefer ^{*}

Institute of Medical Technology, Hamburg University of Technology, 21073 Hamburg, Germany; matthias.schlueter@tuhh.de (M.S.); lukas.glandorf@tuhh.de (L.G.); martin.gromniak@tuhh.de (M.G.); thore.saathoff@tuhh.de (T.S.)

^{*} Correspondence: schlaefer@tuhh.de

[†] These authors contributed equally to this work.

Received: 18 March 2020; Accepted: 5 May 2020; Published: 8 May 2020



Abstract: Optical tracking systems are widely used, for example, to navigate medical interventions. Typically, they require the presence of known geometrical structures, the placement of artificial markers, or a prominent texture on the target's surface. In this work, we propose a 6D tracking approach employing volumetric optical coherence tomography (OCT) images. OCT has a micrometer-scale resolution and employs near-infrared light to penetrate few millimeters into, for example, tissue. Thereby, it provides sub-surface information which we use to track arbitrary targets, even with poorly structured surfaces, without requiring markers. Our proposed system can shift the OCT's field-of-view in space and uses an adaptive correlation filter to estimate the motion at multiple locations on the target. This allows one to estimate the target's position and orientation. We show that our approach is able to track translational motion with root-mean-squared errors below 0.25 mm and in-plane rotations with errors below 0.3°. For out-of-plane rotations, our prototypical system can achieve errors around 0.6°.

Keywords: markerless tracking; tracking systems; optical coherence tomography; motion estimation

1. Introduction

Optical tracking is an established approach in a wide range of applications to estimate or compensate the motion of a target. One particularly important field is the navigation of medical procedures, for example, in radiation therapy, neurosurgery, laparoscopy, or orthopedic surgery. In some fields, we mainly consider the tracking of rigid structures, which can be described by their pose; i.e., position and orientation in space. A classic example is the human skull during surgeries [1]. Optical markers can be used to determine the pose with respect to some reference frame. However, marker placement is critical for the accuracy of the approach [2]. Accurate head tracking is also fundamental in cranial radiosurgery. If the patient moves his head during the treatment, we irradiate healthy areas instead of the tumor. Therefore, we have to detect any larger motion to stop treatment and avoid damage. To reduce the likelihood of such situations, different approaches to immobilize the head can be employed; for example, stereotactic frames or thermoplastic masks. These approaches, however, reduce patient comfort and can make patient positioning more difficult. Additionally, there is always some residual motion in the case of masks. For these reasons, head tracking is an important approach, especially because systems like the CyberKnife (Accuray) allow one to dynamically adjust the direction of the treatment beams. However, this requires continuous tracking of the head's pose with high accuracy and low latency. Markerless approaches to this task employ, for example, structured light [3] or time-of-flight cameras [4]. Extending point cloud models with additional information, such as tissue thickness estimation, has been proposed [5]. Another application for head tracking in

medicine is robotic transcranial magnetic stimulation (rTMS). In rTMS, we use a robot to place a coil closely to the patient's head. This coil non-invasively stimulates a brain region with magnetic pulses. Tracking the head's pose is fundamental to continuously adapting the coil pose and ensuring steady stimulation of the correct brain region.

Another class of targets constitutes moving and deforming structures within the patient. In abdominal radiosurgery, liver or prostate motion has to be tracked for correct irradiation. If respiration is the cause of the motion, external optical markers can serve as a surrogate to estimate the internal motion [6]. Direct optical tracking of organs frequently occurs in minimally invasive surgeries, such as laparoscopy. While markerless instrument tracking [7] can typically exploit clear and known geometries, this is not the case for organs. Their borders might not be within the imaging range, their shapes and appearances can substantially vary between patients, and textural features on their surfaces can be rare and sensitive to moisture, illumination, and other external conditions. Recently, there has been an increasing interest in integrating augmented reality (AR) and mixed reality (MR) in the operating rooms, especially due to the release of devices such as HoloLens (Microsoft) [8]. While virtual reality (VR) is mainly increasing the efficiency of training or planning applications, AR and MR promise to provide information more concise and intuitive to the surgeon during actual interventions [9–11]. Instead of observing visualizations and data on several monitors, information is directly fused and visualized by superimposing it to the surgeon's current view onto the patient. Therefore, they require one to solve several processing tasks, such as object detection, segmentation, 3D scene reconstructions, registration with previous data, localization, and pose estimation. Furthermore, all these tasks have to be solved online with small delays in order to provide an acceptable perception. This emphasizes that efficient solutions for high-accuracy tracking in difficult environments are still an important topic today.

In summary, highly accurate tracking is an important issue for various applications involving navigation, robotics, and VR/AR. If we want to overcome the limitations of markers, most approaches are limited to targets with clear or known textures or borders. This is problematic in cases of unknown smooth surfaces and under difficult conditions, like limited space. For these reasons, we propose to consider the imaging modality optical coherence tomography (OCT) for markerless tracking. The basis of OCT are one-dimensional interferometric measurements employing near-infrared light, typically at a center wavelength which experiences rather little absorption in water. Thereby, it is possible to acquire depth profiles (A-scans) containing information about reflecting structures below the surface, which are not accessible with superficially scanning modalities. In general, the imaging range of an A-scan covers only few millimeters, but it provides micrometer-scale resolution and acquisition rates above 1 MHz [12]. Scan heads, typically utilizing galvanometer mirrors (galvos), allow one to laterally translate the OCT beam to obtain sequences of A-scans representing cross-sectional (B-scans) or volumetric images (C-scans). With the latest systems, even high-resolution C-scans can be acquired at video rate [13] and with low motion artifacts [14].

OCT imaging is an established technique in ophthalmology, but it is also increasingly being used in other fields, for instance, in intravascular diagnosis in cardiology [15] or dermatology [16]. OCT has also been studied to guide interventions like cochleostomy [17] or to monitor medical lasers [18,19]. For intra-operative applications, integration of OCT scanning into surgical tools has been proposed [20,21]. OCT has been studied for image registration in the contexts of tracking of artificial markers [22] and video stabilization [23]. Additionally, different deep-learning approaches have been evaluated recently for learning known marker geometries [24] or evaluating optical flow [25]. However, the constantly limiting aspect when using OCT imaging is the very small field-of-view (FOV). Although experimental systems allowing for larger-scale OCT imaging by different techniques have been described [26–28], the FOV of commercial systems is still limited to some millimeters in every direction. In clinical applications, this often leads to using a second modality to guide the actual OCT imaging; for example, angiography in cardiology for catheter placement. To overcome the limited size of the FOV, different approaches have been studied. They usually involve moving the OCT scan head—for example, by integration into a motorized microscope [29] or by mounting it to a robot [30],

and some kind of volume stitching [31]. As an alternative which does not require moving the scanning system, we recently proposed a setup which uses a second pair of galvos to laterally move the FOV to different locations [32]. By additionally using a motorized reference arm, the FOV can also be shifted axially. Besides volume stitching, this setup also allows one to compensate for translational motion of a target by dynamically adjusting the FOV position and to recover its trajectory [33]. However, this approach is so far limited to 3D tracking.

In this work, we describe a concept for employing OCT to setup a generic tracking system considering the following key specifications. First, the system should be able to estimate trajectories of 6D motion, i.e., translations and rotations of a rigid target, in a range that substantially exceeds the very limited FOV size. Second, it realizes fully markerless tracking; i.e., no target-specific information is required. Third, targets can be tracked without having their borders in the FOV or even without having a clearly textured surface at all by exploiting OCT's sub-surface information.

2. Material and Methods

In this section, we first describe the hardware setup of our system, which consists of an OCT device and a setup to dynamically reposition the OCT's FOV. Repositioning the FOV allows one to follow a moving target. We first describe a single-template concept and then propose a multi-template concept which allows one to estimate 6D motion by tracking multiple points on a target. For tracking of the different templates, we employ a MOSSE filter. Finally, we describe in this section a calibration procedure to obtain a Cartesian coordinate frame and the experimental setup for our experiments and evaluations. The overall system is sketched in Figure 1.

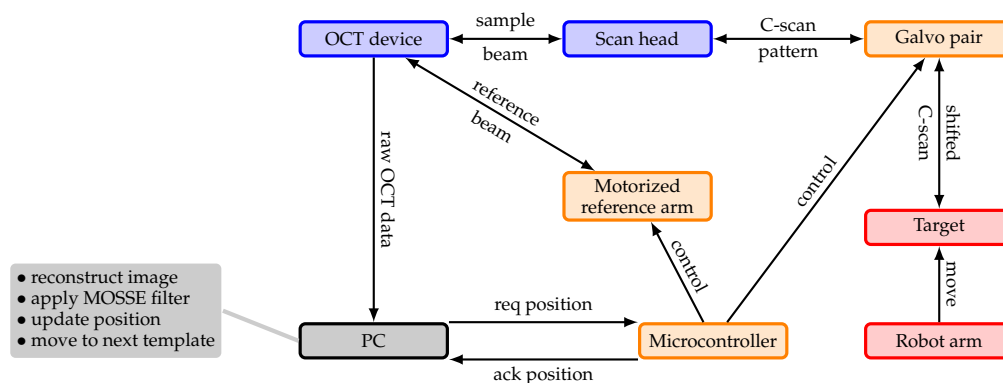


Figure 1. Components of the proposed tracking system. It includes an OCT imaging system (blue), additional hardware to reposition the field-of-view (FOV; orange), a computing device running the tracking loop (gray), and a target moved by a robot arm for evaluation (red).

2.1. Hardware Setup

For imaging, we employ a commercially available swept-source OCT device (OMES, Optores), which consists of an FDML laser, an imaging module, and a scan head. The device has a center wavelength of 1315 nm and provides an A-scan rate of 1.59 MHz, where each A-scan is reconstructed from 2432 raw data samples. By closing a shutter in the sample arm, we estimate the background signal and subtract it from the raw data. The non-linearity of the laser's sweep is estimated by branching off a recalibration arm from the reference arm and processing the recalibration signal with a correction algorithm provided by the manufacturer's SDK. Image reconstruction is fully implemented in CUDA to allow for efficient GPU processing. For acquiring C-scans consisting of 32×32 axial A-scans, we achieve a C-scan rate of 831 Hz by using a scan head with one resonant galvanometer mirror. We use a lateral scan area of approximately $2.5 \text{ mm} \times 2.5 \text{ mm}$ and crop the A-scans to 480 pixels which cover approximately 3.5 mm axially.

We placed a pair of galvos in front of the scan head, which allows one to shift the whole FOV laterally (Figure 2a). In front of these galvos, we have a lens with a focal distance of 30 mm defining the working distance between our system and the target to be tracked. Further, its diameter of 50.8 mm limits the lateral positioning range to a disk. To shift the OCT's FOV along its axial direction, we employ the motorized reference arm shown in Figure 2b. It consists of a fixed mirror and a retroreflector on a linear guide, which is driven by a stepper motor via a belt. When we move the retroreflector, the FOV is shifted by twice the distance. While this reference arm could be used to realize rather large axial shifts, we only consider shifts in this work with similar lateral and axial ranges.

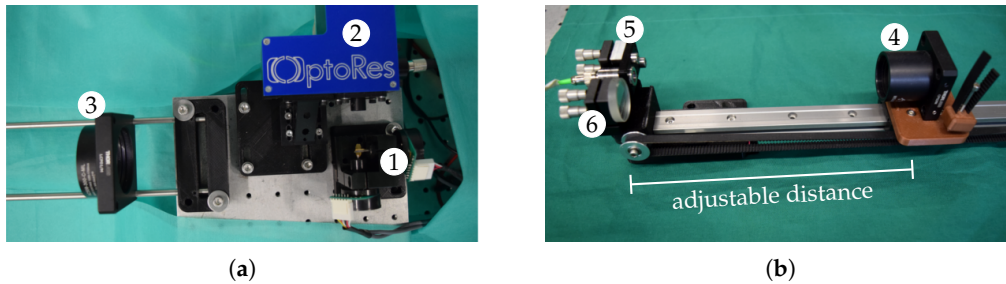


Figure 2. The OCT's FOV can be shifted laterally by a second pair of galvos (a, 1) in front of the actual scan head (a, 2) and the range is limited by the achromatic lens with diameter of 50.8 mm (a, 3). For axial shifting, a retroreflector (b, 4) is mounted to a motor. Due to the setup with the collimator (b, 5) and the mirror (b, 6), the path length of the reference arm varies as twice the traveling distance of the motor.

In summary, our setup allows one to shift the FOV in 3D space by selecting a three-dimensional vector $\mathbf{m} = (m_x, m_y, m_z)^\top$ of motor steps, where m_x and m_y refer to lateral shifting by the two galvos and m_z to axial shifting by the motorized reference arm. The motor and galvos are controlled by a microcontroller which communicates with our measurement PC. The PC has a digitizer board (ATS9373, Alazartech) for OCT data acquisition and two GPUs (GTX Titan X and GTX 980 Ti, Nvidia). We use one of the GPUs for image reconstruction, and the second one is available for further data processing; i.e., implementation of a tracking algorithm.

2.2. Single-Template Tracking

In order to track a moving target and to adjust the FOV position accordingly, we generally employ a template-based approach. Given a template C-scan $V^0(x, y, z)$ and a later C-scan $V^i(x, y, z)$, an image processing algorithm determines the translation $\mathbf{t}(V^0, V^i) = (\Delta x, \Delta y, \Delta z)^\top$ measured in voxels. This translation is converted into a new vector of motor steps \mathbf{m}^i as

$$\begin{aligned} \mathbf{m}^i &= \mathbf{m}^{i-1} + \Delta \mathbf{m}^i \\ &= \mathbf{m}^{i-1} + \text{diag}(p_{xy}, p_{xy}, p_z) \cdot \mathbf{C} \cdot \mathbf{t}(V^0, V^i) \end{aligned} \quad (1)$$

where \mathbf{C} is a matrix converting voxels into motor steps and p_{xy} and p_z are proportional gains for the movement of the galvos and the stepper motor, respectively. We assumed that \mathbf{C} is a diagonal matrix—i.e., each motor axis is exactly identified with one OCT axis—which turned out to be sufficient for the feedback loop in our setup. We estimated the lateral conversion factors to be 2.5 steps/voxel and the axial conversion to be 0.7 steps/voxel. Note that the axial voxel spacing is substantially smaller than the lateral.

2.3. Multi-Template Tracking

The pose of a rigid body can be determined from the locations of three non-collinear points. Therefore, to estimate the orientation in addition to the positions of our moving targets, we simultaneously track

multiple templates which are spatially separated. We consider N templates and their corresponding initial FOV positions $m_n^0, n = 1, \dots, N$, and arrange them on a circle with radius r around a suitable center point $c = (x^0, y^0, z^0)^T$, as shown in Figure 3. Therefore, we have

$$m_n^0 = c + \begin{pmatrix} r \cdot \cos\left(2\pi \frac{n-1}{N}\right) \\ r \cdot \sin\left(2\pi \frac{n-1}{N}\right) \\ 0 \end{pmatrix} \tag{2}$$

as the initial motor positions. For targets which are not sufficiently flat within radius r , we could also define the templates at different depths. However, we assume for now that this is not necessary.

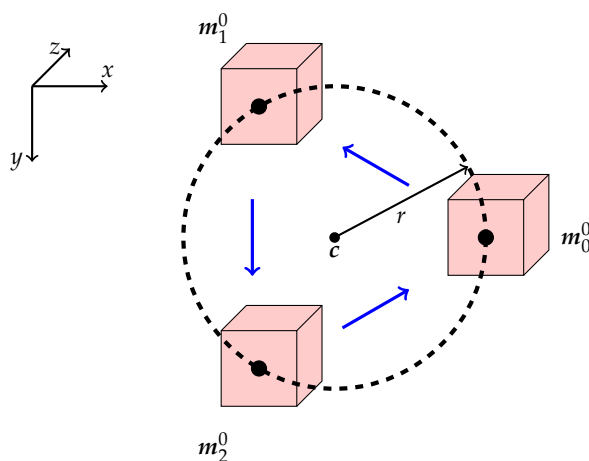


Figure 3. We define, in this example, $N = 3$ templates by moving the OCT’s FOV (red) to three positions m_i^0 on a circle with radius r . The initial circle lies within a plane parallel to the xy -plane, but during tracking this might change due to rotational motion.

We track these N templates sequentially. In iteration i for template n , we acquire a C-scan V_n^i at position m_n^{i-1} ; compute its translation with respect to V_n^0 and convert it to motor steps Δm_n^i ; and determine the next position m_n^i according to Equation (1). Afterwards, we do the same steps for the next template. However, we can pipeline these operations as sketched in Figure 4 to increase the tracking rate. Note that the most time-consuming part in our setup is moving the FOV axially by changing the length of the reference arm, especially if the employed stepper motor has to change its direction of travel.

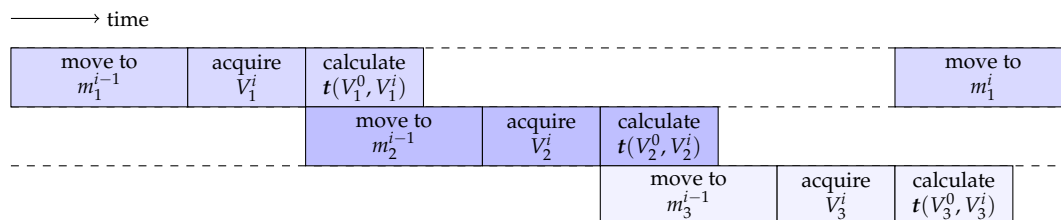


Figure 4. Sketch of pipelined multi-template tracking with three templates. Iteratively, a new volume is acquired and evaluated to determine the motor position for this template in the next iteration. Note that calculating the translations typically is faster than moving to the next position for acquisition.

We do not consider m_n^i for tracking directly, but only use it to adjust the FOV position. Considering Equation (1), we see that the appropriate FOV position for V^i is rather

$$\tilde{m}_n^i = m_n^{i-1} + C \cdot t(V_n^0, V_n^i); \tag{3}$$

i.e., without the proportional gains. As a timestamp for this position, we use the acquisition timestamp of V^i .

2.4. Tracking Algorithm

In previous work [33], we showed that the phase correlation method is a suitable and efficient tracking algorithm for the single-template setup with purely translational motion. It employs cross-correlation to estimate the translation between a template C-scan and a new template C-scan. By using a Fourier-domain approach, phase correlation has very low computational cost.

However, if we also observe substantial rotations, which we cannot compensate by shifting the FOV in our system, then we cannot expect this simple template-based approach to succeed. The rotation has a severe impact on the appearance of the target in the C-scan, and matching with the template C-scan will quickly fail due to the decaying similarity. More recently developed adaptive correlation filters extend the Fourier domain approach of the phase correlation method to gradually alter the filter's target representation and allow target tracking through appearance changes [34]. We chose to use the MOSSE (minimum output sum of squared error) filter [35] that offers high computational speed, online filter adaptation, and a straightforward extension from two-dimensional to volumetric image data.

Following Bolme et al. [35], we summarize the key concepts of the MOSSE filter for three-dimensional images. It typically does not employ a single template image $f(x)$, $x \in \mathbb{R}^3$, but constructs an optimal filter $H(\mathbf{k})$ in the Fourier domain from T training images $f_j(x)$. This reduces the noise and makes the filter more robust and stable. Let $F_j(\mathbf{k})$ be the three-dimensional Fourier transform of $f_j(x)$. The complex conjugate of the filter $H(\mathbf{k})$ is then constructed from the training images as

$$H^*(\mathbf{k}) = \frac{\sum_{j=1}^T G_j(\mathbf{k}) \cdot F_j^*(\mathbf{k})}{\sum_{j=1}^T F_j(\mathbf{k}) \cdot F_j^*(\mathbf{k}) + \lambda'} \quad (4)$$

where λ is a small regularization constant. $G_j(\mathbf{k})$ is the Fourier transform of a Gaussian function

$$g_j(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^3} \exp\left(-\frac{(\mathbf{s}_j - \mathbf{x})^\top(\mathbf{s}_j - \mathbf{x})}{2\sigma^2}\right) \quad (5)$$

which is shifted by \mathbf{s}_j and represents a smoothed version of the desired filter output. The vector \mathbf{s}_j represents the translation which the filter should determine for input $f_j(x)$; i.e., the training images do not have to be centered in this algorithm. During tracking, we then obtain the translation \mathbf{t} of a new image by transforming the image to the Fourier domain, multiplying it with $H^*(\mathbf{k})$, and applying the inverse Fourier transform to the product. The location of the highest intensity in this result corresponds to the translation of the new image with respect to the filter. The advantage of allowing for arbitrarily shifted images for training becomes even more apparent when considering updating the filter during tracking to adapt to appearance changes. The filter can be updated after each evaluation of a new image. When we evaluate the i -th image $f^i(x)$ during tracking, we obtain its translation \mathbf{t}^i from the current filter $H^{i-1}(\mathbf{k})$ as described above. Note that we use superscript indices for the image sequence during actual tracking and subscript indices for the training images. The determined translation \mathbf{t}^i defines the target output $g^i(x)$ of the i -th image, and the updated filter is recursively updated as

$$H^{i*}(\mathbf{k}) = \frac{A^i(\mathbf{k})}{B^i(\mathbf{k})} \quad (6)$$

with nominator

$$A^i(\mathbf{k}) = \eta \cdot \left(G^i(\mathbf{k}) \cdot F^{i*}(\mathbf{k})\right) + (1 - \eta) \cdot A^{i-1}(\mathbf{k}), \quad (7)$$

denominator

$$B^i(\mathbf{k}) = \eta \cdot \left(F^i(\mathbf{k}) \cdot F^{i*}(\mathbf{k}) + \lambda \right) + (1 - \eta) \cdot B^{i-1}(\mathbf{k}), \quad (8)$$

and learning rate $\eta \in [0, 1]$. The higher η , the more influence the new image has on the filter. Updating templates always leads to a critical trade-off [36]. On the one hand, it allows for compensating stronger changes in appearance, but on the other hand, it also increases the likelihood of a drift; i.e., the filter no longer tracks the original target after some time but a structure at another location.

In our implementation, we use ten training images initially acquired at the same position. We use $\sigma = 2$ for the Gaussians and further set $\lambda = 0.001$. We update the filter each time we evaluate a new image and compare different learning rates η . The algorithm is implemented in CUDA for processing on a GPU.

2.5. Cartesian Calibration

Neither the C-scans nor the motor positions form a Cartesian coordinate frame with physical units. Therefore, we need a calibration to obtain proper coordinates, especially to calculate meaningful orientations from the motor positions. For this purpose, we can use a 3D motion stage and single-template tracking with the phase correlation algorithm to acquire calibration data [33]. We move a marker slowly to different positions within a volume-of-interest with active tracking. Thereby, we obtain M pairs of positions \mathbf{p}^i of the motion stage and motor positions \mathbf{m}^i of the tracking system. We can firstly determine an affine transformation matrix $A \in \mathbb{R}^{3 \times 4}$ by solving

$$\min_A \sum_{i=0}^{M-1} \left\| \mathbf{p}^i - A \cdot \begin{bmatrix} \mathbf{m}^i \\ 1 \end{bmatrix} \right\|^2. \quad (9)$$

This matrix compensates for the overall offset, aligns the axes, and scales the motor steps to millimeters between the coordinate frame of the motor positions and the Cartesian frame of the motion stage. However, the rotating galvos introduce some spherical distortions which cannot be adequately represented by an affine transformation. Therefore, we secondly fit a three-dimensional quadratic function $\mathbf{q} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ to finally obtain Cartesian motor positions $\hat{\mathbf{m}}^i$ as

$$\hat{\mathbf{m}}^i = \mathbf{q} \left(A \cdot \mathbf{m}^i \right). \quad (10)$$

2.6. Pose Estimation

We estimate the position of the target in iteration i as the centroid of the N Cartesian motor positions $\hat{\mathbf{m}}_n^i$, $n = 1, \dots, N$. For this purpose, we resample the positions of the N templates to common timestamps. To estimate the orientation in iteration i , we apply the Kabsch algorithm [37] to the positions $\hat{\mathbf{m}}_n^i$ and the corresponding initial positions $\hat{\mathbf{m}}_n^0$ of the templates.

2.7. Phantoms and Simulated Motion

As targets for tracking, we consider the phantoms shown in Figure 5. The 40 mm \times 40 mm sized plates have a thickness of about 1 mm. They are 3D-printed using photopolymer resin and we receive OCT signals from their backsides as well. The back sides of the plates are rough and randomly structured to simulate sub-surface information. The front side of the first phantom shows the same structure, while the second phantom has a flat surface, instead.

We use a 6-axis robot arm (IRB 120, ABB) to simulate and evaluate different motion patterns. As shown in Figure 6, the end-effector has a sample holder attached to position a phantom about 300 mm in front of the lens of the tracking system. It is positioned such that the axes of its base coordinate frame are approximately aligned to the OCT's and the tracking system's axes. This allows one to systematically evaluate two types of rotations. We define in-plane rotations to be in a plane parallel to the OCT's xy -plane. Therefore, the axis of rotation corresponds to the axial scan direction

of the OCT, and hence motion can mostly be compensated by only moving the second pair of galvos while the reference arm length does not have to be adjusted. The other type is out-of-plane rotations, which we define to be in the OCT's yz -plane.

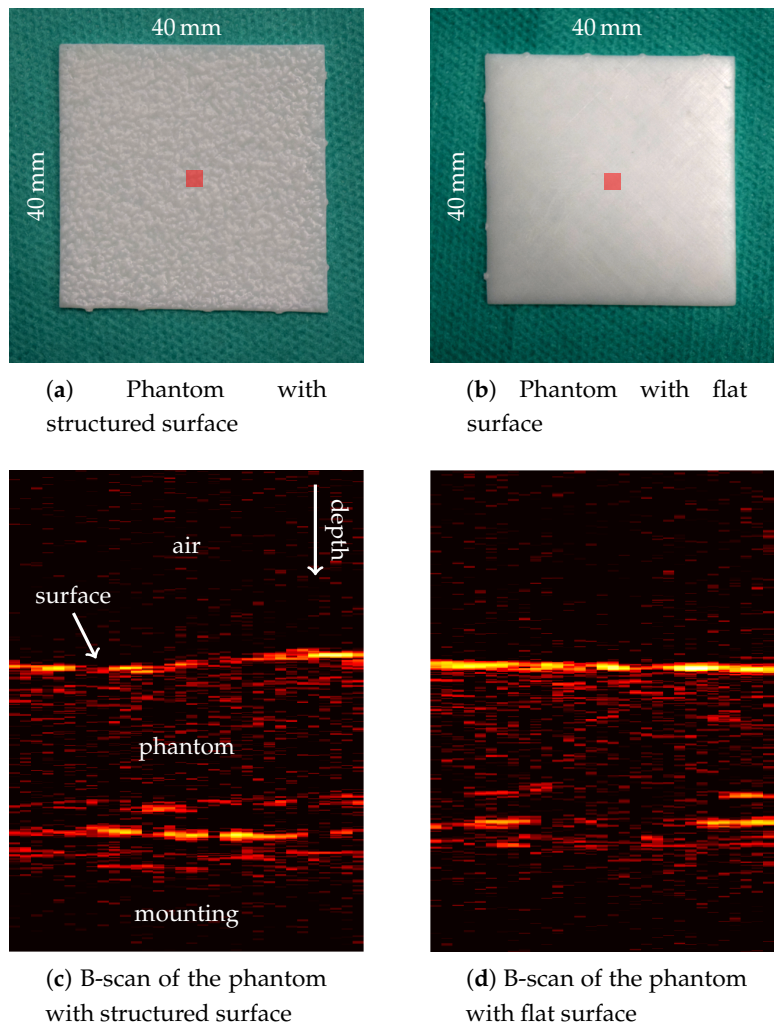


Figure 5. Top row: Pictures of the phantom with a structured surface (a) and the phantom with a flat surface (b). The backsides of both phantoms are structured too. For comparison, the size of the OCT's FOV is illustrated as red squares. Bottom row: B-scans of the structured-surface phantom (c) and the flat-surface phantom (d). Each column corresponds to one A-scan with increasing measurement depth from top to bottom. The B-scans' pixels are stretched to match their physical dimensions, which are about 2.5 mm laterally and 3.5 mm in depth direction in air.

To evaluate pure translations, we move the robot arm's end-effector back and forth along a line. For rotations, we rotate in-plane or out-of-plane starting with the original end-effector orientation. Furthermore, we consider random 6D motion. For the sequence of positions, we uniformly draw random points within a cube. For the sequence of rotations, we define a random axis of rotation and uniformly draw a random angle. The positions and the orientations are then realized simultaneously by the robot arm. Each experiment consists of a motion sequence which we record over 45 s. For each repetition, we acquire the templates at a different randomly selected spot on the phantoms before motion starts.

We also use the robot arm to perform the Cartesian calibration. In detail, we move the structured-surface phantom to 80 different positions within a cylinder with a radius of 20 mm and a

depth of 40 mm. We only move at 2 mm/s to ensure reliable tracking. At each position, we log the tracking system's position ten times.

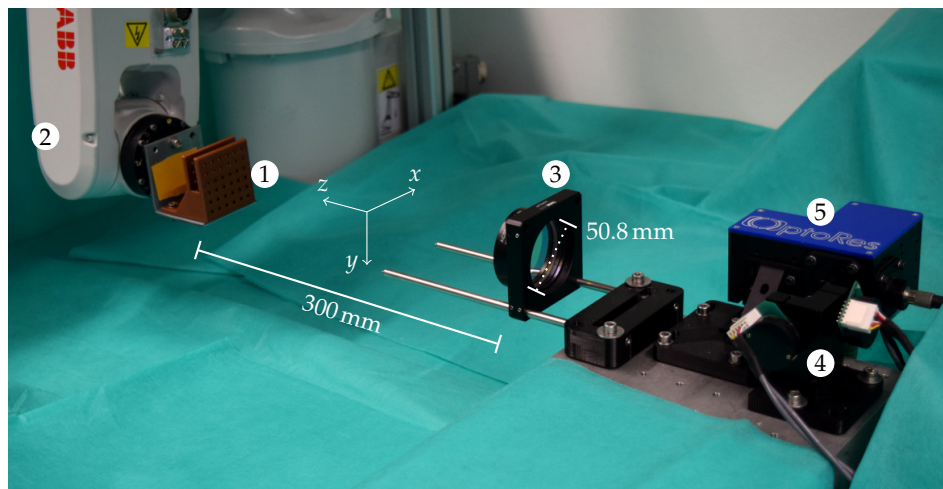


Figure 6. A sample can be mounted to the holder (1) which is attached to the robot arm (2) in front of the scan lens (3), the second pair of galvos (4), and the scan head (5).

2.8. Evaluation Metrics

For pure translations, we evaluate the Euclidean norm of the differences between the relative positions of the robot and those of our tracking system. To evaluate orientations, we use the axis-angle representation, which provides an angle θ which is independent of coordinate systems. Evaluating the six degrees of freedom individually would require one to express translations and rotations in a common frame. While hand-eye calibration could be used to relate the robot and OCT-based tracking system [30,38], it would also be limited by the same system components we use in our setup for tracking, and the residual calibration error would affect the evaluation of the tracking approach. Therefore, we consider the relative rotations in the robot frame and the tracking-system frame independently and compare their magnitudes, which include all three rotational degrees of freedom. Similarly, we established the magnitude of the translation as the Euclidian distance. We mainly consider the root-mean-squared error (RMSE) to describe the results, either with respect to the translations or to the angles.

3. Results

3.1. Cartesian Calibration

The calibration positions are located at three discrete radii within the cylindric volume, and Figure 7 shows the residual error of the calibration procedure depending on the radius. After fitting an affine transformation to the data points, the residual errors are especially high for the outer positions. After additionally fitting the quadratic function (10), the overall residual RMSE and the maximum residual error are 0.061 mm and 0.136 mm, respectively.

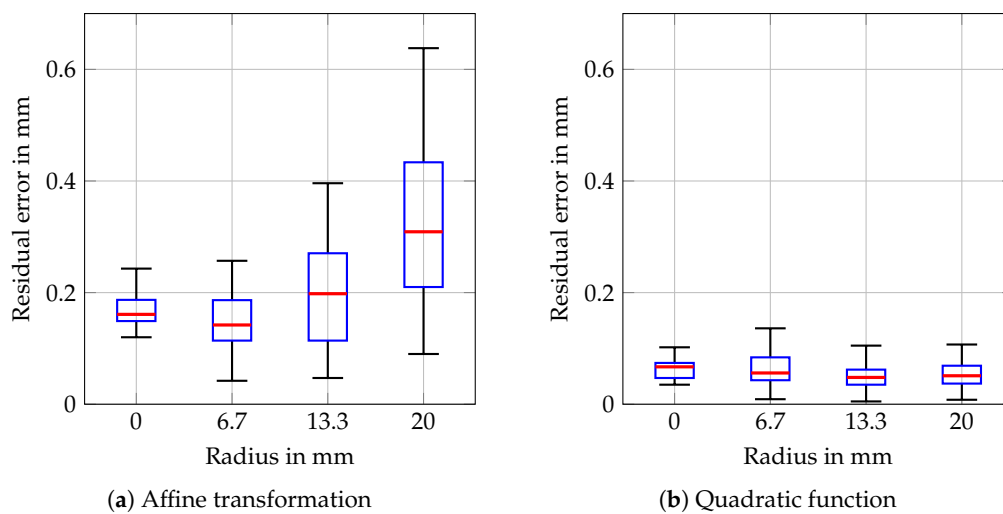


Figure 7. Boxplots of the residual calibration errors after fitting only an affine transformation (a) and after additionally fitting a quadratic function (b) depending on the radius within the cylindrical volume of the positions. The red bars indicate the medians, the blue boxes range from the 25%-quantiles to the 75%-quantiles, and the whiskers cover all data points.

3.2. In-Plane Rotations

We consider tracking of rotations and estimation of the current orientation by tracking $N = 3$ spatially separated localities with the MOSSE filter at the same time. Their spatial arrangement follows Equation (2) with a radius of $r = 200$ motor steps, to begin with. This corresponds to a radius of approximately 5.4 mm. We first evaluate in-plane rotations using the phantom with a structured surface (Figure 5a). Remember that in-plane rotations mainly require one to move the two galvos but not the stepper motor.

Figure 8 shows that learning rates between 0.01 and 0.03 for the MOSSE filter do not affect tracking accuracy for purely translational motion. For the actual in-plane rotations, we see from Figure 9b that rotations at $4^\circ/\text{s}$ and with maximum angle of $13^\circ/\text{s}$ can be tracked with RMSEs around 0.2° for the lower learning rates. When employing higher learning rates, the RMSEs increase above 0.3° . If we consider rather slow motion at $0.2^\circ/\text{s}$ in Figure 9a, however, the influence of the higher learning rates becomes much clearer. Especially for tracking motion with a maximum angle of 8° , the errors severely increase with increasing learning rate, and Figure 10 shows the deviations between estimated and actual orientations for different learning rates. Using a higher learning rate results in increased underestimation of a slow rotation.

The RMSEs shown in Figure 11 for in-plane rotations of the phantom with a flat surface (Figure 5b)—i.e., without visible superficial information—are similar to the errors for in-plane rotations of the structured-surface phantom. We again observe increased RMSEs for slow rotations when choosing 0.02 for η compared to choosing 0.01. For fast rotation and a maximum angle of 8° , a higher spread of the results compared to the structured-surface phantom can be observed.

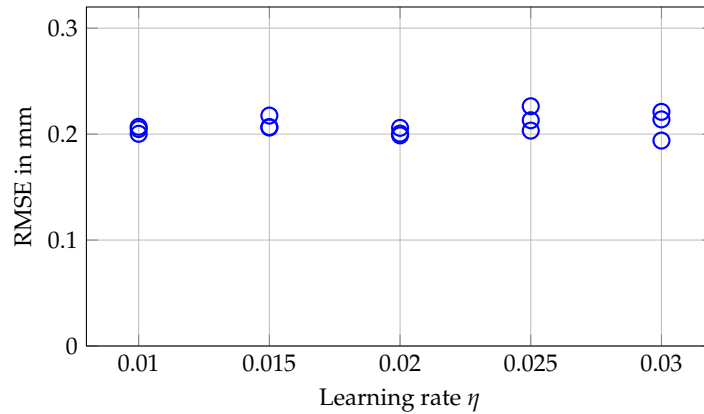


Figure 8. RMSEs for translational motion along all three axes. Three repetitions are shown for each learning rate η . Each time, the robot arm drove back and forth along a diagonal line within a cube with side length 8 mm at a velocity of 12 mm/s.

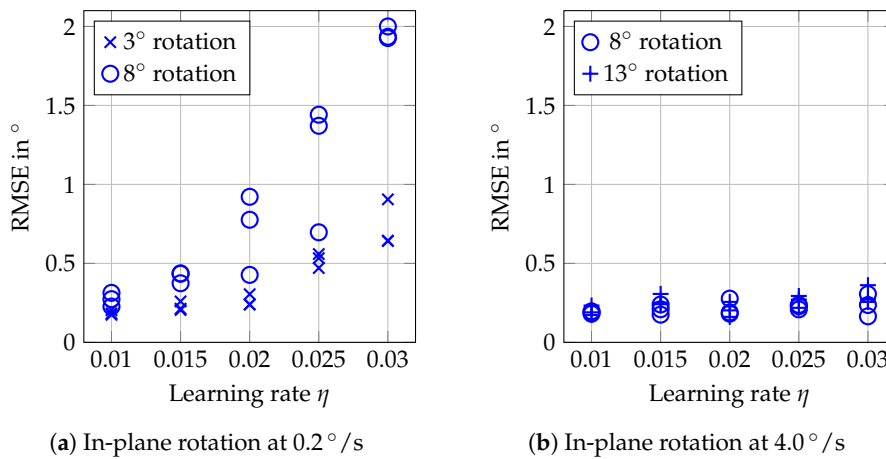


Figure 9. RMSEs for in-plane rotations with the different learning rates η from 0.01 to 0.03 of the MOSSE filter. Different maximum angles were used and the velocities were 0.2°/s (a) and 4°/s (b).

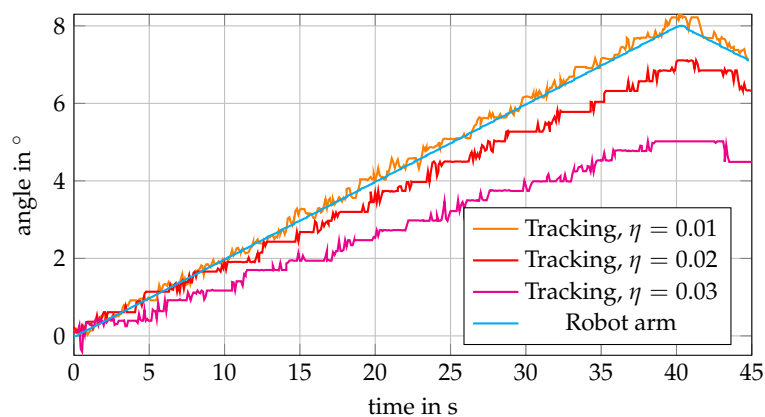


Figure 10. Exemplary curves of estimated angles and robot arm angles for in-plane rotations at 0.2°/s and a maximum angle of 8° (compare Figure 9a). When increasing the learning rate η , the angle is increasingly underestimated.

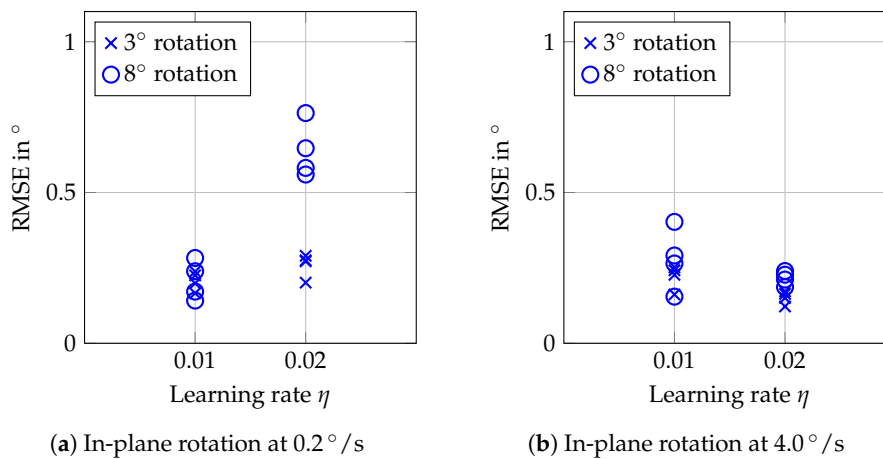


Figure 11. RMSEs for in-plane rotation of the flat-surface phantom.

3.3. Out-of-Plane Rotations

When considering out-of-plane rotations, motion of the target along the OCT’s axial direction will occur. That is, the motorized reference arm has to be used to follow the target by adjusting the reference-arm length accordingly. In previous experiments [33], this mechanical movement appeared to be much more limiting to the dynamic behavior of our system compared to movement of the galvos for lateral compensation.

The results in Figure 12 show larger RMSEs, in the order of 1° , compared to the in-plane rotations. Furthermore, the trackable velocity was more limited. With the lower learning rates, already tracking rotations at $1.5^\circ/s$ was not successful. In contrast, all reported measurements were successful for $\eta = 0.03$, for rotation with maximum angles of 3° and of 8° . The reason for the increased RMSE can be seen in the trajectories shown in Figure 13. There is an overestimation of small angles which turns into an almost constant offset for angles above 2° . The general trend, however, is reflected correctly by our tracking approach.

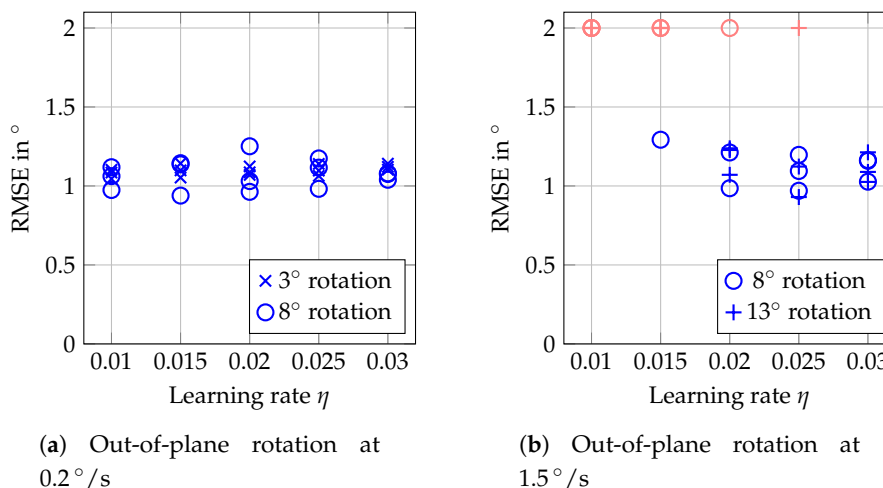


Figure 12. RMSEs of the errors for out-of-plane rotations and different learning rates η from 0.01 to 0.03 of the MOSSE filter. Three repetitions were done for each combination, but RMSEs exceeding 2° are considered as failed tracking and indicated by red marks at 2° .

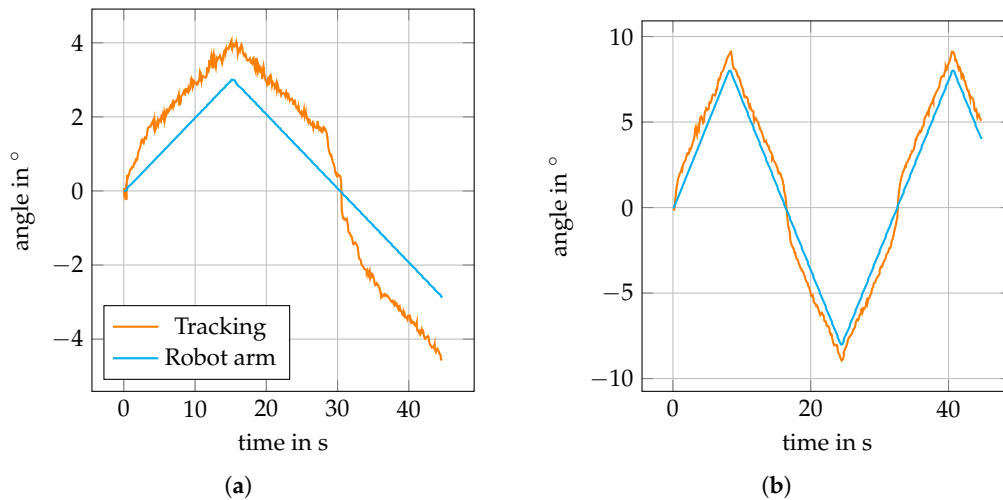


Figure 13. The tracking system (orange) systematically overestimates the robot angles (cyan) for out-of-plane rotations with different maximum angles at $0.2^\circ/s$ (a) and at $1^\circ/s$ (b). Learning rate $\eta = 0.02$ is used.

As Figure 14 shows, the errors depend on the radius of the circle on which the templates are positioned. A higher radius provides better estimations. The drawback of larger radii is a smaller range in which motions can be tracked, because we sooner reach the border of the scanning lens. If we combine the larger radius with more than three templates, we can further improve the quality of the estimates. Table 1 shows that the combination of 300 motor steps as the radius r and tracking of $N = 7$ templates results in RMSEs in the order of only 0.6° . As a drawback, tracking an 8° rotation already failed at $1^\circ/s$ in two of three cases. Moving the FOV to the next template’s position requires some time, and therefore, by setting $N = 7$, we increase the time until we evaluate the same template again; i.e., we decrease the effective sampling rate.

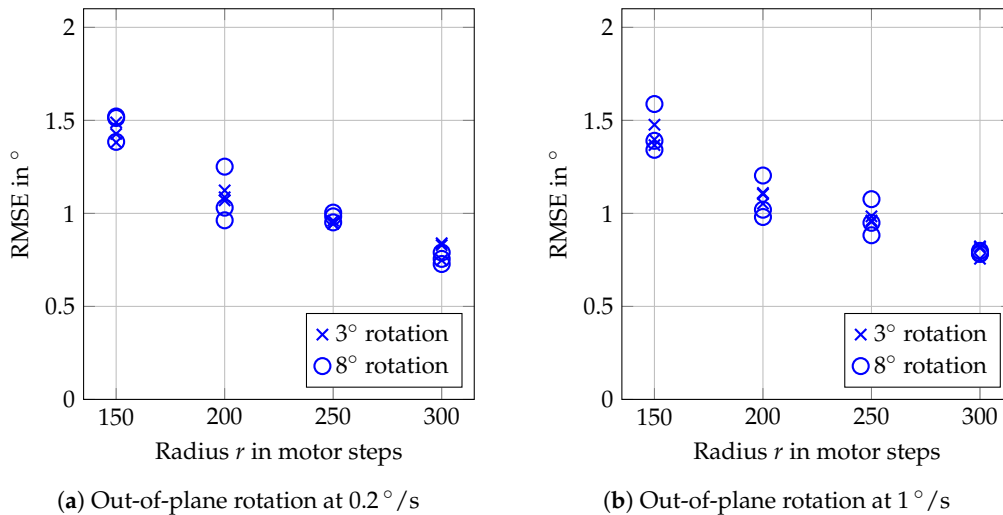


Figure 14. RMSEs for out-of-plane rotations and tracking with different radii r in motor steps. The learning rate is set to $\eta = 0.02$.

Table 1. RMSEs for out-of-plane rotations and tracking with parameters $N = 7, r = 300$, and $\eta = 0.02$.

	3° Max Angle				8° Max Angle				
	0.2°/s		1.5°/s		0.2°/s		1.5°/s		
RMSE in °	0.6224	0.6341	0.6423	0.5993	0.6202	0.6347	0.5533	0.6271	0.6281

As a final proof of concept, we consider estimating the orientation during random 6D motion; i.e., simultaneous translations along all axes and rotations about random axes. We again set $r = 200$ and $N = 3$ as in the first experiments. An exemplary sequence of translations and orientations shows Figure 15. While having an offset as observed before and exhibiting some noise, the estimated angle follows the progress of the actual angle realized by the robot arm. In more detail, Figure 16 shows the RMSEs of the orientations estimated during simultaneous translation at 2, 5, and 8 mm/s within a cube of side length 10 mm. Again, the lower learning rate fails more often for higher velocities at larger maximum angles. However a rotation with maximum angle of 3° at $1.5^\circ/\text{s}$ can be tracked with RMSEs in the order 0.75° while at the same time random translations at 8 mm/s are present.

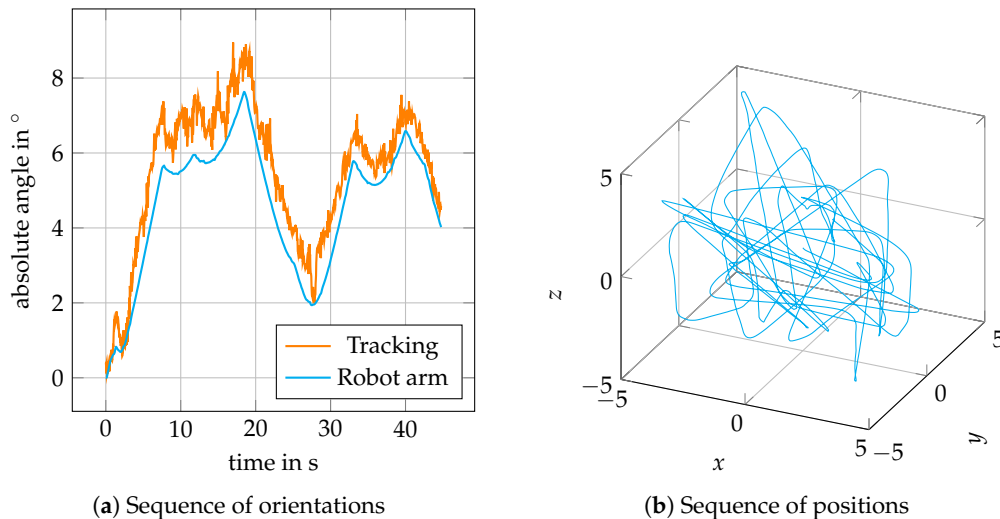


Figure 15. Exemplary trajectory for random 6D motion. The robot arm changes its end-effector orientation randomly at $1^\circ/\text{s}$ up to 8° from its initial orientation (a). At the same time, the robot arm translates the phantom randomly within a $10\text{ mm} \times 10\text{ mm} \times 10\text{ mm}$ cube at 8 mm/s (b).

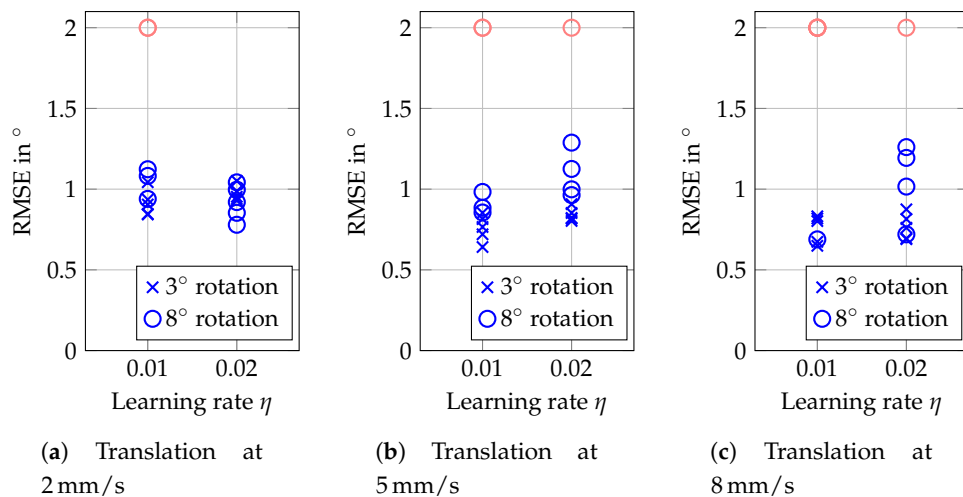


Figure 16. RMSEs for the estimated orientation during random 6D motion. Results are reported for a rotational velocity of $1^\circ/\text{s}$ and translational motion within a cube of side length 10 mm. Results of five repetitions for each parameter combination are shown, but RMSEs exceeding 2° are considered as failed tracking and indicated by red marks at 2° .

4. Discussion

Conventional optical tracking systems or cameras usually rely on either textured surfaces, or surfaces with varying height, visibility of borders, or artificial markers. In contrast, our proposed tracking approach allows one to track translations and rotations markerlessly, even for the phantom with a smooth surface and no visible borders. We are able to track translations with RMSEs below 0.25 mm (Figure 8). Tracking of in-plane rotations is feasible with errors below 0.3° (Figure 9), even for a large target with a flat surface (Figure 11). By employing OCT, we can exploit micrometer-scale volumetric image data including information about the structure below the surface. Furthermore, imaging is non-invasive and feasible at high temporal resolution.

For our tracking approach using an adaptive MOSSE filter, the choice of the learning rate η can be critical. While increasing η leads to severe underestimation of angles for in-plane rotations (Figure 10), a higher η is necessary for reliable tracking of out-of-plane rotations (Figure 12). This is a general issue, and for an application one should consider which type of motion can be expected. Depending on the velocities expected, either a smaller or a higher learning rate should be chosen. Furthermore, one could consider an approach where the learning rate is adapted online based on the last observed velocities [39].

Tracking of out-of-plane rotations generally performs worse than tracking of in-plane rotations, which only involves the pair of galvos for adjusting the FOV position. The target, however, is still tracked correctly during out-of-plane rotations from a qualitative point of view (Figure 13), even for more complex motion (Figure 15). However, the quantitative evaluation (Figure 12) is distorted by an overestimation of the angles, which occurs throughout all measurements. By arranging the template on a larger circle and by tracking more templates, this effect can be reduced (Table 1). For specific applications, this can, however, lead to deciding for a trade-off between the accuracy and the size of the area in which motion can be tracked.

5. Conclusions

In this work, we showed that 6D motion tracking with volumetric OCT is feasible. Our hardware setup allows one to follow motion at multiple locations on a target in order to estimate its position and orientation. Furthermore, we showed that an adaptive MOSSE filter is appropriate to handle the changing appearance of targets in the images caused by rotational motion. Our results indicate that adapting the learning rate with respect to the current speed of the target could be advantageous. Tracking of in-plane rotations was feasible with very promising small errors, even for targets with smooth surfaces. This highlights the value of OCT's sub-surface information for markerless tracking compared to superficially scanning modalities. Following out-of-plane rotations was feasible as well. However, the task is more tricky, and the motorized reference arm seems to not be an optimal choice. Nevertheless, the trajectories could be qualitatively tracked, and we showed that the quantitative results improved by tracking more locations on a larger radius. Therefore, our proposed concept shows that OCT imaging is promising to realize optical tracking in applications which are difficult for conventional optical tracking systems, especially scenarios not allowing for proper marker placement and providing only access to small and poorly-structured segments of the target.

Author Contributions: Conceptualization, M.S., M.G., and A.S.; methodology, M.S., L.G., M.G., and T.S.; software, M.S., L.G., and M.G.; validation, M.S. and L.G.; formal analysis, M.S.; investigation, M.S. and L.G.; writing—original draft preparation, M.S.; writing—review and editing, L.G. and A.S.; visualization, M.S.; supervision, A.S.; project administration, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by Forschungszentrum Medizintechnik Hamburg (grant 04fmthh16). We acknowledge support for the open access fees by Hamburg University of Technology (TUHH) in the funding program Open Access Publishing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kral, F.; Puschban, E.J.; Riechelmann, H.; Freysinger, W. Comparison of optical and electromagnetic tracking for navigated lateral skull base surgery. *Int. J. Med. Robot. Comput. Assist. Surg.* **2013**, *9*, 247–252. [[CrossRef](#)] [[PubMed](#)]
2. West, J.B.; Fitzpatrick, J.M.; Toms, S.A.; Maurer, C.R.; Maciunas, R.J. Fiducial Point Placement and the Accuracy of Point-based, Rigid Body Registration. *Neurosurgery* **2001**, *48*, 810–817. [[CrossRef](#)] [[PubMed](#)]
3. Frost, R.; Wighton, P.; Karahanoğlu, F.I.; Robertson, R.L.; Grant, P.E.; Fischl, B.; Tisdall, M.D.; van der Kouwe, A. Markerless high-frequency prospective motion correction for neuroanatomical MRI. *Magn. Reson. Med.* **2019**, *82*, 126–144. [[CrossRef](#)] [[PubMed](#)]
4. Noonan, P.J.; Howard, J.; Hallett, W.A.; Gunn, R.N. Repurposing the Microsoft Kinect for Windows v2 for external head motion tracking for brain PET. *Phys. Med. Biol.* **2015**, *60*, 8753–8766. [[CrossRef](#)] [[PubMed](#)]
5. Wissel, T.; Stüber, P.; Wagner, B.; Bruder, R.; Erdmann, C.; Deutz, C.S.; Sack, B.; Mani, J.; Schweikard, A.; Ernst, F. Enhanced Optical Head Tracking for Cranial Radiation Therapy: Supporting Surface Registration by Cutaneous Structures. *Int. J. Radiat. Oncol. Biol. Phys.* **2016**, *95*, 810–817. [[CrossRef](#)]
6. Ernst, F.; Bruder, R.; Schlaefer, A.; Schweikard, A. Correlation between external and internal respiratory motion: a validation study. *Int. J. Comput. Assist. Radiol. Surg.* **2012**, *7*, 483–492. [[CrossRef](#)]
7. Bouget, D.; Allan, M.; Stoyanov, D.; Jannin, P. Vision-based and marker-less surgical tool detection and tracking: A review of the literature. *Med. Image Anal.* **2017**, *35*, 633–654. [[CrossRef](#)]
8. Verhey, J.T.; Haglin, J.M.; Verhey, E.M.; Hartigan, D.E. Virtual, augmented, and mixed reality applications in orthopedic surgery. *Int. J. Med. Robot. Comput. Assist. Surg.* **2020**, e2067. [[CrossRef](#)]
9. Chauvet, P.; Bourdel, N.; Calvet, L.; Magnin, B.; Teluob, G.; Canis, M.; Bartoli, A. Augmented Reality with Diffusion Tensor Imaging and Tractography during Laparoscopic Myomectomies. *J. Minim. Invasive Gynecol.* **2019**. [[CrossRef](#)]
10. Prevost, G.A.; Eigel, B.; Paolucci, I.; Rudolph, T.; Peterhans, M.; Weber, S.; Beldi, G.; Candinas, D.; Lachenmayer, A. Efficiency, Accuracy and Clinical Applicability of a New Image-Guided Surgery System in 3D Laparoscopic Liver Surgery. *J. Gastrointest. Surg.* **2019**. [[CrossRef](#)]
11. Akladios, C.; Gabriele, V.; Agnus, V.; Martel-Billard, C.; Saadeh, R.; Garbin, O.; Lecointre, L.; Marescaux, J. Augmented reality in gynecologic laparoscopic surgery: Development, evaluation of accuracy and clinical relevance of a device useful to identify ureters during surgery. *Surg. Endosc.* **2019**, *34*, 1077–1087. [[CrossRef](#)]
12. Klein, T.; Huber, R. High-speed OCT light sources and systems [Invited]. *Biomed. Opt. Express* **2017**, *8*, 828. [[CrossRef](#)]
13. Kolb, J.P.; Draxinger, W.; Klee, J.; Pfeiffer, T.; Eibl, M.; Klein, T.; Wieser, W.; Huber, R. Live video rate volumetric OCT imaging of the retina with multi-MHz A-scan rates. *PLOS ONE* **2019**, *14*, e0213144. [[CrossRef](#)]
14. Wang, T.; Pfeiffer, T.; Regar, E.; Wieser, W.; van Beusekom, H.; Lancee, C.T.; Springeling, G.; Krabbendam-Peters, I.; van der Steen, A.F.; Huber, R.; et al. Heartbeat OCT and Motion-Free 3D In Vivo Coronary Artery Microscopy. *JACC Cardiovasc. Imaging* **2016**, *9*, 622–623. [[CrossRef](#)]
15. Bouma, B.E.; Villiger, M.; Otsuka, K.; Oh, W.Y. Intravascular optical coherence tomography [Invited]. *Biomed. Opt. Express* **2017**, *8*, 2660. [[CrossRef](#)] [[PubMed](#)]
16. Olsen, J.; Holmes, J.; Jemec, G.B.E. Advances in optical coherence tomography in dermatology—A review. *J. Biomed. Opt.* **2018**, *23*, 1. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, Y.; Pfeiffer, T.; Weller, M.; Wieser, W.; Huber, R.; Raczkowski, J.; Schipper, J.; Wörn, H.; Klenzner, T. Optical Coherence Tomography Guided Laser Cochleostomy: Towards the Accuracy on Tens of Micrometer Scale. *BioMed. Res. Int.* **2014**, *2014*, 1–10. [[CrossRef](#)] [[PubMed](#)]
18. Kim, B.; Kim, D.Y. Enhanced Tissue Ablation Efficiency with a Mid-Infrared Nonlinear Frequency Conversion Laser System and Tissue Interaction Monitoring Using Optical Coherence Tomography. *Sensors* **2016**, *16*, 598. [[CrossRef](#)] [[PubMed](#)]
19. Schlüter, M.; Fuh, M.M.; Maier, S.; Otte, C.; Kiani, P.; Hansen, N.O.; Miller, R.J.D.; Schlüter, H.; Schlaefer, A. Towards OCT-Navigated Tissue Ablation with a Picosecond Infrared Laser (PIRL) and Mass-Spectrometric Analysis. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Berlin, Germany, 23–27 July 2019; pp. 158–161. [[CrossRef](#)]

20. Park, B.; Lee, S.; Bang, H.; Kim, B.; Park, J.; Kim, D.; Park, S.; Won, Y. Image-Guided Laparoscopic Surgical Tool (IGLaST) Based on the Optical Frequency Domain Imaging (OFDI) to Prevent Bleeding. *Sensors* **2017**, *17*, 919. [[CrossRef](#)]
21. Schulz-Hildebrandt, H.; Pfeiffer, T.; Eixmann, T.; Lohmann, S.; Ahrens, M.; Rehra, J.; Draxinger, W.; König, P.; Huber, R.; Hüttmann, G. High-speed fiber scanning endoscope for volumetric multi-megahertz optical coherence tomography. *Opt. Lett.* **2018**, *43*, 4386. [[CrossRef](#)]
22. Zhang, Y.; Wörn, H. Optical coherence tomography as highly accurate optical tracking system. In Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Besacon, France, 8–11 July 2014; pp. 1145–1150. [[CrossRef](#)]
23. Laves, M.H.; Schoob, A.; Kahrs, L.A.; Pfeiffer, T.; Huber, R.; Ortmaier, T. Feature tracking for automated volume of interest stabilization on 4D-OCT images. *SPIE Med. Imaging* **2017**, 101350W. [[CrossRef](#)]
24. Gessert, N.; Schlüter, M.; Schlaefer, A. A deep learning approach for pose estimation from volumetric OCT data. *Med. Image Anal.* **2018**, *46*, 162–179. [[CrossRef](#)] [[PubMed](#)]
25. Laves, M.H.; Kahrs, L.A.; Ortmaier, T.; Ihler, S. Deep-learning-based 2.5D flow field estimation for maximum intensity projections of 4D optical coherence tomography. *SPIE Med. Imaging* **2019**, 109510R. [[CrossRef](#)]
26. Song, S.; Xu, J.; Wang, R.K. Long-range and wide field of view optical coherence tomography for in vivo 3D imaging of large volume object based on akinetic programmable swept source. *Biomed. Opt. Express* **2016**, *7*, 4734. [[CrossRef](#)] [[PubMed](#)]
27. Pfeiffer, T.; Draxinger, W.; Wieser, W.; Klein, T.; Petermann, M.; Huber, R. Analysis of FDML lasers with meter range coherence. *SPIE BiOS* **2017**, 10053, 100531T. [[CrossRef](#)]
28. Huang, Y.; Badar, M.; Nitkowski, A.; Weinroth, A.; Tansu, N.; Zhou, C. Wide-field high-speed space-division multiplexing optical coherence tomography using an integrated photonic device. *Biomed. Opt. Express* **2017**, *8*, 3856–3867. [[CrossRef](#)]
29. Finke, M.; Kantelhardt, S.; Schlaefer, A.; Bruder, R.; Lanke, E.; Giese, A.; Schweikard, A. Automatic scanning of large tissue areas in neurosurgery using optical coherence tomography. *Int. J. Med. Robot. Comput. Assist. Surg.* **2012**, *8*, 327–336. [[CrossRef](#)]
30. Rajput, O.; Antoni, S.T.; Otte, C.; Saathoff, T.; Matthäus, L.; Schlaefer, A. High accuracy 3D data acquisition using co-registered OCT and Kinect. In Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Baden-Baden, Germany, 19–21 September 2016; pp. 32–37. [[CrossRef](#)]
31. Gan, Y.; Yao, W.; Myers, K.M.; Hendon, C.P. An automated 3D registration method for optical coherence tomography volumes. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 3873–3876. [[CrossRef](#)]
32. Schlüter, M.; Otte, C.; Saathoff, T.; Gessert, N.; Schlaefer, A. Feasibility of a markerless tracking system based on optical coherence tomography. *SPIE Med. Imaging* **2019**, 1095107. [[CrossRef](#)]
33. Schlüter, M.; Glandorf, L.; Sprenger, J.; Gromniak, M.; Saathoff, T.; Schlaefer, A. High-Speed Markerless Tissue Motion Tracking Using Volumetric Optical Coherence Tomography Images. In Proceedings of the IEEE International Symposium on Biomedical Imaging, Iowa City, IA, USA, 3–7 April 2020.
34. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 583–596. [[CrossRef](#)]
35. Bolme, D.S.; Beveridge, J.R.; Draper, B.A.; Lui, Y.M. Visual object tracking using adaptive correlation filters. In Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2544–2550. [[CrossRef](#)]
36. Matthews, L.; Ishikawa, T.; Baker, S. The template update problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 810–815. [[CrossRef](#)]
37. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A* **1976**, *32*, 922–923. [[CrossRef](#)]
38. Zhou, M.; Hamad, M.; Weiss, J.; Eslami, A.; Huang, K.; Maier, M.; Lohmann, C.P.; Navab, N.; Knoll, A.; Nasser, M.A. Towards Robotic Eye Surgery: Marker-Free, Online Hand-Eye Calibration Using Optical Coherence Tomography Images. *IEEE Robot. Autom. Lett.* **2018**, *3*, 3944–3951. [[CrossRef](#)]

39. Asha, C.; Narasimhadhan, A. Adaptive learning rate for visual tracking using correlation filters. *Procedia Comput. Sci.* **2016**, *89*, 614–622. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).