




Efficient and high accuracy 3-D OCT angiography motion correction in pathology

STEFAN B. PLONER,^{1,2,*}  MARTIN F. KRAUS,¹ ERIC M. MOULT,²
LENNART HUSVOGT,^{1,2}  JULIA SCHOTTENHAMML,¹ A. YASIN
ALIBHAI,³ NADIA K. WAHEED,³ JAY S. DUKER,³ JAMES G.
FUJIMOTO,² AND ANDREAS K. MAIER¹

¹Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Martensstr. 3, Erlangen, 91058, Germany

²Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA

³New England Eye Center, Tufts Medical Center, 800 Washington St. Box 450, Boston, MA 02111, USA
*stefan.ploner@fau.de

Abstract: We describe a novel method for non-rigid 3-D motion correction of orthogonally raster-scanned optical coherence tomography angiography volumes. This is the first approach that aligns predominantly axial structural features such as retinal layers as well as transverse angiographic vascular features in a joint optimization. Combined with orthogonal scanning and favorization of kinematically more plausible displacements, subpixel alignment and micrometer-scale distortion correction is achieved in all 3 dimensions. As no specific structures are segmented, the method is by design robust to pathologic changes. Furthermore, the method is designed for highly parallel implementation and short runtime, allowing its integration into clinical workflow even for high density or wide-field scans. We evaluated the algorithm with metrics related to clinically relevant features in an extensive quantitative evaluation based on 204 volumetric scans of 17 subjects, including patients with diverse pathologies and healthy controls. Using this method, we achieve state-of-the-art axial motion correction and show significant advances in both transverse co-alignment and distortion correction, especially in the subgroup with pathology.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Optical coherence tomography (OCT) is a non-invasive 3-D optical imaging modality that is a standard of care in ophthalmology [1,2]. Since the introduction of Fourier-domain OCT [3], dramatic increases in imaging speed became possible, enabling 3-D volumetric data to be acquired. Typically, a region of the retina is scanned line by line, where each scanned line acquires a cross-sectional image or a B-scan. Since B-scans are acquired in milliseconds, slices extracted along a scan line, or the fast scan axis, are barely affected by motion. In contrast, slices extracted orthogonally to scan lines, i. e. in slow scan direction, are affected by various types of eye motion occurring throughout the full, multi-second volume acquisition time. The most relevant types of eye movements during acquisition are (micro-)saccades, which can introduce discontinuities or gaps between B-scans, and slow drifts, which cause small, slowly changing distortion [4]. Additional eye motion is caused by pulsatile blood flow, respiration and head motion. Despite ongoing advances in instrument scanning speed [5,6] typical volume acquisition times have not decreased. Instead, the additional scanning speed is used for dense volumetric scanning or wider fields of view [7]. OCT angiography (OCTA) [8–11] multiplies the required number of scans by at least two, and even more scans are needed to accommodate recent developments in blood flow speed estimation which are based on multiple interscan times [12,13]. As a consequence, there is an ongoing need for improvement in motion compensation especially in pathology [14–16].

One approach for correcting such motion is the use of additional hardware, e.g. a scanning laser ophthalmoscope, for eye tracking and correction of the scan position [17–19] or retrospective alignment with a reference image [20]. While only eye tracking provides the capability of a gap-free acquisition with a single scan, disadvantages are e.g. increased system complexity and cost, increased scan duration if scanning is interrupted due to saccadic eye motion, and limited accuracy, which is especially important in the context of OCT angiography. A second approach is to acquire multiple OCT scans and to perform a software-based alignment retrospectively. While this increases scan time and delays the display of the corrected result, accuracy down to below OCT resolution is possible and individual scans can be merged to create higher quality images [21,22]. Some methods additionally allow mosaicing of partially overlapping volumes [23,24], but are limited to alignment in the *en face* plane. Others allow full 3-D alignment including the axial dimension [21,25–28]. A fundamental difference between software methods is related to the underlying scan pattern. Methods that use multiple raster scans with B-scans oriented along the same direction easily allow the registration of non-square field-of-views [22,24,27]. However, since there is no motion-free reference in the direction orthogonal to the B-scans, they are limited to remove uncorrelated saccadic motion. Spatially correlated saccadic motion may occur when patients involuntarily fixate on the scanning beam as it traverses the fovea. Drift motion can be aligned with a reference but not corrected, because distortion in the reference propagates to the result. In contrast, methods based on orthogonally oriented B-scans [21,23,25,26] or lissajous scans [28,29] have nearly motion-free slices available along all dimensions and thus have the potential for the most accurate motion correction. Finally, it is common to do at least simple software-based registration before the OCT angiography computation irrespective of the use of eye tracking, and it was reported that the combination of both approaches can be superior to hardware or software motion correction alone [30].

This paper extends the orthogonal raster scan based method by Kraus et al. [25]. Distinct features of this method are the joint motion correction of all dimensions in a combined optimization and a displacement field model that is designed along the OCT scanning process: First, instead of having a single displacement field that maps a *moving* volume to a *reference*, one displacement field exists for each volume. This enables the method to correct the motion in all volume scans, potentially avoiding propagation of drift motion to the final result. Secondly, the displacement fields are enforced to be smooth not isotropically, as common in image registration, but instead along the scanning trajectory underlying each individual scan. This favors estimation of more realistic displacement fields. Combined, these design decisions enable and effect the separation of the displacement into the motion underlying each raster scan, allowing each individual scan to be undistorted in contrast to simple co-alignment that may reproduce distortion from the input.

This technique is used in commercial clinical instruments, however, since its initial publication in 2014, the requirements for motion correction have increased with the demand for repeated scanning in OCTA and widefield imaging. In contrast to other methods, OCT angiography data was not yet incorporated into the motion correction algorithm, and details, such as masking areas with saturated OCTA signal due to saccadic motion, were not covered. The fine structure of vessel networks, as visualized by OCTA, at the same time requires but also provides potential for further improved registration accuracy. Conversely, the longer scan times for OCTA and higher densities increase distortion from drift and saccadic motion in the uncorrected scans. It is thus the goal of this paper to extend the previous motion correction method to utilize the potential of OCTA with minimal impact on runtime, and to demonstrate improvements in distortion correction capability on dense OCTA scans in a comprehensive and clinically relevant study cohort.

2. Methods

This section proposes an extension to the motion correction approach of Kraus et al. [21,25] to additionally employ OCTA data during registration. In contrast to the OCT volume, the OCTA

data is projected to generate an *en face* image before it is used in the registration. This is because the layered structure of the eye already provides ample features for subpixel axial registration with standard OCT alone. Due to the retinal curvature especially around the foveal avascular zone and optic nerve, these features are also useful for a global transverse registration. However, these features are continuously changing in the *en face* plane, limiting the achievable accuracy. In contrast, OCTA has high contrast visualization of fine vessel structures and can provide local features. Thus, while the *en face* approach maintains the relevant benefits of the OCTA signal, the increased runtime for handling the 2-D OCTA data during optimization is negligible compared to the 3-D OCT data, provided that the *en face* image can be computed quickly.

Figure 1 shows an outline of the algorithm as a whole with terminology used throughout this paper. Section 2.1 briefly introduces the *OCT only* algorithm of Kraus et al. [21,25] and introduces basic notations. It is followed by detailed descriptions of the *OCT + OCTA* extension, starting with the detection of motion artifacts in the OCTA data in Section 2.2, continuing with the efficient generation of *en face* OCTA images in Section 2.3 and their incorporation into the alignment optimization in Section 2.4. Finally, merging of the motion corrected OCTA volumes into a single result is described in Section 2.5.

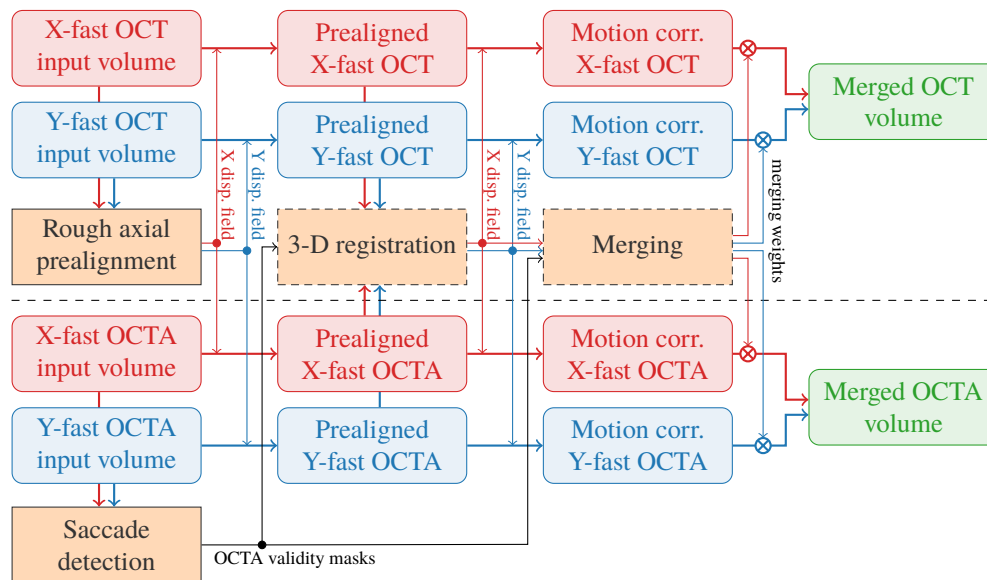


Fig. 1. Algorithm data flow. The part above the dashed line describes the existing *OCT only* method: First, a rough prealignment and tilt correction is performed in the axial dimension. Then, the volumes are registered jointly in all dimensions to obtain a motion corrected image (*motion corr.*) and subsequently merged by a weighted average. The new *OCT + OCTA* approach adds saccade detection, the 3-D registration is extended to project and use OCTA data, and merging excludes white line artifacts in OCTA data.

2.1. Basic OCT only algorithm

The basic method uses at least two raster scans from an arbitrary OCT instrument, of which at least one volume must have B-scans oriented orthogonally to other volumes, such that scans with minimal motion are available in both lateral dimensions. For simplicity, this paper is limited to the case of exactly two orthogonal scans unless otherwise stated. For a straightforward generalization to more than two volumes see Section 2.4 in [21].

The motion corrected volume should resemble a hypothetical scan that was performed without motion. It is scanned over an equiangular grid with width w and height h with A-scans of depth d such that the voxel positions are located at $\mathbf{p}_{ijk} = (x_i, y_j, z_k)$ with $i \in \{1, \dots, w\}$, $j \in \{1, \dots, h\}$, $k \in \{1, \dots, d\}$. According to typical viewing of OCT data, voxel spacing along each axis is assumed to be uniform. Since the actual acquisitions are distorted by motion, they need to be resampled at these locations to become aligned with the motion-free scan. Thus, the goal of the registration is to estimate the displacements \mathbf{D} composed of vectors $\mathbf{d}_{ij}^V = (\Delta x_{ij}^V, \Delta y_{ij}^V, \Delta z_{ij}^V)$, where (i, j) are the horizontal and vertical A-scan indices and $V \in \{X, Y\}$ identifies the X- or Y-fast scan. It is assumed that there is no distortion within A-scans.

First, the registration steps *rough axial prealignment* and *3-D registration* (see Fig. 1) are described, which follow the same concept. The estimated displacements are iteratively refined by minimizing an objective function

$$\arg \min_{\mathbf{D}} S_0^{\text{str}}(\mathbf{S}^X, \mathbf{S}^Y, \mathbf{D}) + \alpha R(\mathbf{D}) + M(\mathbf{D}) + T(\mathbf{S}^X, \mathbf{S}^Y, \mathbf{D}), \quad (1)$$

composed of a data term S that describes the dissimilarity of the motion corrected scans, a regularizer R that penalizes fast displacement changes, the mean shift regularizer M and the tilt normalization term T . The constituents are further described in the following.

Alignment is quantified based on the residual difference

$$R_{ijk}(\mathbf{S}^X, \mathbf{S}^Y, \mathbf{D}) = I\left(\mathbf{S}^X, x_i + \Delta x_{ij}^X, y_j + \Delta y_{ij}^X, z_k + \Delta z_{ij}^X\right) - I\left(\mathbf{S}^Y, x_i + \Delta x_{ij}^Y, y_j + \Delta y_{ij}^Y, z_k + \Delta z_{ij}^Y\right) \quad (2)$$

of each resampled voxel i, j, k , where $I(\mathbf{S}^V, x, y, z)$ is an interpolator that samples \mathbf{S}^V at position (x, y, z) via Hermitian interpolation and $\mathbf{S}^X, \mathbf{S}^Y$ are the preprocessed structural OCT data of the X- and Y-fast input volumes. The purpose of the preprocessing is to normalize the data to become invariant to noise and illumination changes from e.g. vitreous opacities. After applying a loss function L , the sum over all voxels is computed to form the structural similarity

$$S_0^{\text{str}}(\mathbf{S}^X, \mathbf{S}^Y, \mathbf{D}) = \sum_i \sum_j \sum_k L\left(R_{ijk}(\mathbf{S}^X, \mathbf{S}^Y, \mathbf{D})\right). \quad (3)$$

L was chosen to be the slightly modified pseudo-huber-loss

$$L_{H, \epsilon_H}(x) = \epsilon_H \cdot \left(\sqrt{1 + \left(\frac{x}{\epsilon_H}\right)^2} - 1 \right) \quad (4)$$

with $\epsilon_H = 0.001$, which is smooth and stable towards outliers like speckle noise.

The primary regularization term R favors more plausible displacement fields by penalizing change in displacement along the scan trajectory, leading to a more uniform displacement along the fast scan axis and limiting the change between B-scans. It is implemented as an additional cost term based on absolute displacement differences, where differences between B-scans are weighted lower to take the flyback time into account. Its influence is controlled by the weighting factor α . Since alignment is translation-invariant, 3 degrees of freedom are yet undetermined. They are fixed by the mean displacement term M , which penalizes the sum of displacements and thereby centers the alignment to the origin of the coordinate system. T is only used in stage 1 of the optimization and normalizes tilt within B-scans. Since regularization is the same as in the previous method, we refer the reader to the publications of Kraus et al. for details [21,25].

The objective function is minimized with a gradient-based L-BFGS solver. To avoid local minima, both optimizations are performed in a factor 2 multi resolution pyramid, beginning with

the most downsampled volumes and zero-initialized displacements. The *rough axial prealignment* only estimates axial displacements and tilt at B-scan granularity, whereas the *3-D registration* estimates 3-D displacements for, in its last multi resolution level, all A-scans. The resulting displacements are then used for Hermitian resampling to obtain the, in terms of Fig. 1, *prealigned* or *motion corrected* volumes, respectively.

In the final *merging* step, the motion corrected volumes are fused into a single *merged volume* by a weighted average. Voxels with a sampling density higher than normal are assigned exponentially decaying weights. Weights below a small threshold are clipped to zero, while the remaining weights are normalized to achieve uniform brightness. The sampling density-based weighting is done in order to achieve negligible weights for locations in the motion corrected volumes that correspond to areas not covered by the corresponding input volume due to transverse motion. All motion corrected A-scans of these areas end up being sampled from one of the acquired input B-scans next to the gap, which would be sampled more frequently and replicated along the gap. At positions where gaps of both scans overlap, the replicated data of the B-scan with less oversampling remains, which typically originates from the scan with the smaller gap.

2.2. Handling of saccadic motion artifacts in OCT angiography data

OCTA data cannot be used for registration at locations where either of the two co-registered scans are affected by severe saccadic motion. Such motion causes the OCT angiography signal to saturate, resulting in uniformly high intensities across the whole B-scan, as can be seen in Fig. 4(a, b). This means that anatomical features vanish and registration could be misguided. Instead, these B-scans are detected and excluded by a weighting factor of zero at these locations.

For this purpose, 1-D angiography validity masks $\mathbf{v}^X, \mathbf{v}^Y$ were defined for each input volume with values corresponding to B-scans, which are 1 (valid) if neither the corresponding, nor the previous and succeeding 2 B-scans had a mean decorrelation greater than 3 standard deviations above the mean, and 0 (invalid) otherwise. To match the image data during the multi-scale registration, this signal is downsampled in the same way as the intensity signals.

Validity \tilde{v}_{ij}^V of the *motion corrected* OCTA A-scan (i, j) of scan $V \in \{X, Y\}$ is based on the validity of the sampling location in the *input volume*, which is smoothly interpolated according to

$$\tilde{v}_{ij}^X(\mathbf{D}) = I\left(\mathbf{v}^X, y_j + \Delta y_{ij}^X\right) \quad \tilde{v}_{ij}^Y(\mathbf{D}) = I\left(\mathbf{v}^Y, x_i + \Delta x_{ij}^Y\right) \quad (5)$$

respectively, where the interpolator I again uses hermite interpolation and thus provides continuous first order derivatives when \tilde{v}_{ij}^V is used during optimization. Since the input validities only vary between B-scans, interpolation only depends on a single coordinate, the slow scan axis.

2.3. En face OCTA image generation

Because *en face* image generation must be fully automated, yet robust with respect to pathology, we opted against the typical approach of projecting between segmented retinal layer-boundaries. Instead, we developed a new approach which determines the voxels used for projection based on the contrast among neighboring OCTA intensities. A binary mask includes all voxels with relevant OCTA features in their neighborhood. The dependencies are only local, making the approach highly parallelizable and thus suitable for GPU processing. The projection algorithm is designed for amplitude decorrelation angiography data. Such data has high OCTA values in areas of low OCT signal like the vitreous or deeper choroid, which are typically removed via OCT signal-based thresholding (see Section 4.1 in [11]). We defer this thresholding, thus low OCTA signals only occur in the retina, between vessels (see Fig. 2(a)). The projection is performed on the axially prealigned volumes (see Fig. 1).

After an initial 3-D median filter with radius 1 for noise reduction, 1-D rank-3 filters with width 15 are applied in the fast and slow scan directions. By choosing a low rank, these filters act similar

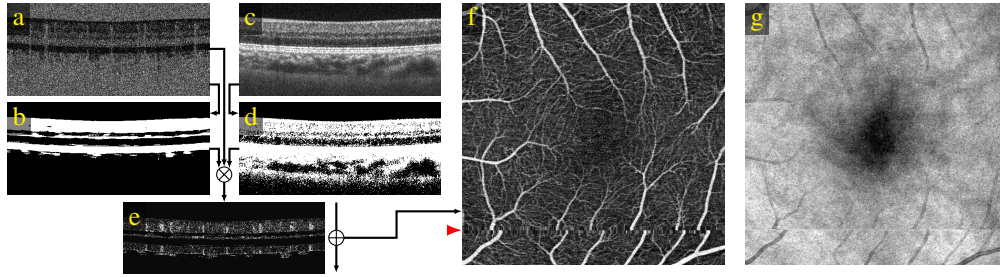


Fig. 2. *En face* OCTA image generation example from a 28 y/o healthy subject. (a) OCTA B-scan and derived mask (b). (c) OCT B-scan and derived amplitude decorrelation mask (d). (e) OCTA B-scan masked with both masks. These B-scans are axially averaged to generate an *en face* image (f). (g) *En face* OCT image.

to an erosion / minimum filter with added robustness toward outliers. The filter size is chosen to match the radius of large vessels, such that most of these vessels are completely removed. The resulting volume typically has low values in the retina and high values elsewhere. In valid B-scans according to \mathbf{v}^X and \mathbf{v}^Y , the 5th and 95th percentiles are computed. The intensities of the full image are then normalized by linearly mapping these values to 0 and 1. Thresholding is performed at 0.1 to form a binary voxel mask, see Fig. 2(b). This mask essentially divides the volume in areas with and without relevant lateral contrast. Both this mask and the typical amplitude decorrelation mask derived via thresholding the OCT volume (d) are applied to the original OCTA B-scan (a), resulting in (e). Note that the OCT-based thresholding alone is insufficient, because it does not consistently remove the amplitude decorrelation angiography signal from the choroid, which would cause varying brightness in the projection. An *en face* OCTA image is formed by averaging the non-excluded voxels along depth, and B-scan median subtraction is applied to compensate for increased OCTA signal from decorrelated B-scans as caused by bulk motion. Finally, to prevent saturated OCTA values from being used for downsampling and interpolation, invalid B-scans according to \mathbf{v}^X or \mathbf{v}^Y were replaced with their nearest valid neighbor (red triangle), forming (f). Note that the contribution of these B-scans to the similarity term is mostly suppressed, as described in the following sections. Figure 3 shows representative data from an eye affected by age-related macular degeneration with geographic atrophy. In this case, due to increased light penetration in the area of atrophy, choroidal vasculature is included in the projection, providing additional features for registration.

In contrast to typical layer segmentation-based approaches, this method is highly parallelizable and thus suitable for GPU processing. Rank filters were implemented similarly to the median filter implementation of Perrot et al. [31], where the low rank filters disproportionately benefit from the forgetful selection algorithm described in their study. When using data from a specific system, the normalization mapping can be fixed to precomputed percentiles to reduce runtime.

2.4. Incorporation of an angiography-based similarity measure

We introduce a new similarity measure which is a combination of a structural $S_{\delta_0}^{\text{str}}$ and an angiographic $S_{\delta_0}^{\text{ang}}$ similarity term, which are defined according to

$$S_{\delta_0}^{\text{str}}(\mathbf{S}^X, \mathbf{S}^Y, \mathbf{D}) = \sum_i \sum_j (1 - \delta_{ij\delta_0}^{XY}(\mathbf{D})) \cdot \sum_k L(R_{ijk}(\mathbf{S}^X, \mathbf{S}^Y, \mathbf{D})) \quad (6)$$

$$S_{\delta_0}^{\text{ang}}(\mathbf{A}^X, \mathbf{A}^Y, \mathbf{D}) = \sum_i \sum_j \delta_{ij\delta_0}^{XY}(\mathbf{D}) \cdot d \cdot L(R_{ijk}(\mathbf{A}^X, \mathbf{A}^Y, \mathbf{D})), \quad (7)$$

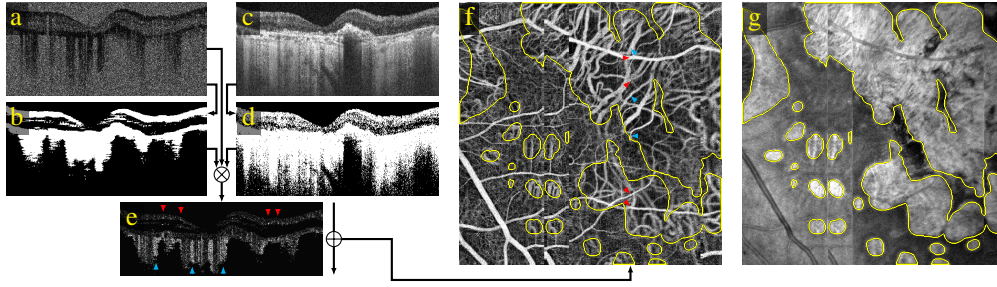


Fig. 3. *En face* image example from an 83 y/o patient with geographic atrophy (yellow). (a) OCTA B-scan and derived mask (b). (c) OCT B-scan and typical derived amplitude decorrelation mask (d). (e) OCTA B-scan masked with both masks. B-scans are axially averaged to generate an *en face* image (f), which shows choroidal vasculature in areas of atrophy. Corresponding superficial and choroidal vessels are marked with red and cyan triangles, respectively. (g) Corresponding *en face* OCT image.

where A^X, A^Y are the projected X- and Y-fast OCTA *en face* images (according to 2.3). The loss L is computed consistently with structural data by the pseudo huber norm as defined in Eq. (4).

$$\delta_{ij\delta_0}^{XY}(\mathbf{D}) = \delta_0 \cdot \tilde{v}_{ij}^X(\mathbf{D}) \cdot \tilde{v}_{ij}^Y(\mathbf{D}). \quad (8)$$

is a weighting factor based on the OCTA validity of scans X and Y at the *motion corrected* A-scan (i, j) . Thus, the similarity measures are identical except for the contrary weighting factors and that the summation along depth is replaced by a multiplication for the *en face* OCTA image. $\delta_{ij\delta_0}^{XY}(\mathbf{D})$ becomes the default value δ_0 , if both angiography validities are valid; 0, if at least one is invalid and thus the difference based on saturated data is not meaningful for registration; and fades smoothly for in between sampling. Note that when $\delta_{ij\delta_0}^{XY}(\mathbf{D})$ is zero, the angiography term becomes zero and the structural term becomes identical to the *OCT only* method. Due to regularization, these B-scans still benefit from the use of OCTA data in their neighborhood.

In its current form, the sum of both terms is still dependent on δ , because the high contrast in the *en face* OCTA image leads to higher residuals. Thus, when simply summing both terms, lower δ values, as they occur in saturated B-scans, would entail a lower loss, biasing the displacement towards sampling from these B-scans. To avoid this bias, a scaling factor η is introduced to equalize both terms. The combined similarity term is defined as

$$S_{\delta_0}(S^X, S^Y, A^X, A^Y, \mathbf{D}, \eta) = \eta \cdot S_{\delta_0}^{\text{str}}(S^X, S^Y, \mathbf{D}) + \frac{1}{\eta} \cdot S_{\delta_0}^{\text{ang}}(A^X, A^Y, \mathbf{D}). \quad (9)$$

The scaling is based on the initial displacement field \mathbf{D}_0 of the current multi resolution level.

$$\eta = \sqrt{S_{0.5}^{\text{ang}}(A^X, A^Y, \mathbf{D}_0) / S_{0.5}^{\text{str}}(S^X, S^Y, \mathbf{D}_0)} \quad (10)$$

Basing the normalization on the initial displacement field avoids recomputing the normalization in every iteration. This normalizes the structural and angiographic data terms to their geometric mean. With this choice, both data terms with their distinct sensitivity for misalignments equally influence the ratio between the overall similarity and the regularizers. Since both similarity terms are normalized to the same value, the contrary weighting factors provide a bias-free exclusion of saturated OCTA data. $S_{\delta_0}(S^X, S^Y, A^X, A^Y, \mathbf{D}, \eta)$ replaces the OCT only similarity term in Eq. (1) in the *3-D registration* (Fig. 1). OCTA data is not used for the *rough axial prealignment*.

2.5. OCTA volume merging

The merging approach for angiography data is based on the merging of OCT volumes described in Section 2.1. However, OCTA B-scans with saturated signal due to saccadic motion need to be excluded. Following the method for *en face* OCTA image generation, the B-scans with invalid data according to v^X, v^Y are first substituted by their nearest valid B-scan. This removes the influence of saturated B-scans during interpolation. The merging itself is extended by multiplying the averaging weights with $\tilde{v}_{ij}^V(D)$ as previously described in Section 2.2 using the final displacements, to exclude the affected OCTA B-scans from the merged result.

Figure 4 shows the effect of volume merging on *en face* projected OCTA volumes. Exclusion of white line B-scans removes the artifact. However, at these locations, only data from the other scan remains. At the crossings of white lines from two orthogonal scans, valid data is not available. While it would be possible to display saturated OCTA data, we wanted to avoid confusion with vasculature, and displayed the data gap in black. The gaps in data can be avoided with high probability by registering additional scans in either or both scanning directions. The degree of noise suppression depends on the actual number of merged volumes at a given region.

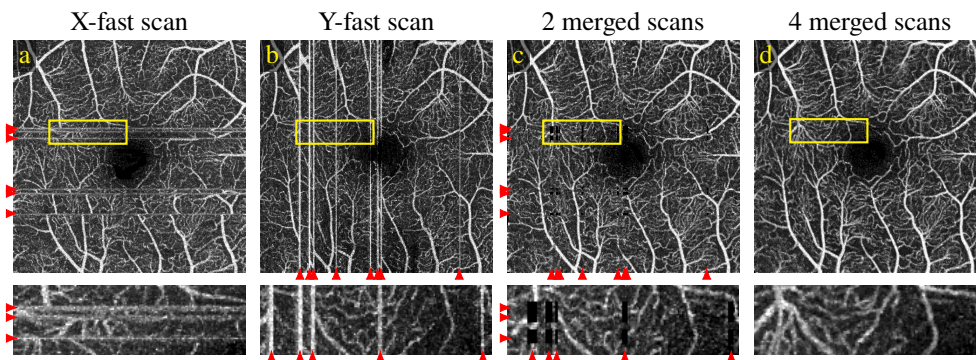


Fig. 4. Removal of white line artifacts in OCTA data from saccadic eye motion. The *en face* images show the superficial vascular plexus. Arrows point to white lines in the original scans, zoomed images show enlargements of the yellow rectangles. (a) Original X-fast and (b) Y-fast scan with white line artifacts. Saccades are in-plane in the X-fast volume and thus partially corrected before the angiography computation. (c) Merged volume from one pair of orthogonal scans with B-scans corresponding to white lines excluded, data gaps appear as black boxes at the intersection of removed B-scans. (d) Merged volume from two pairs of orthogonal scans (4 total) which fills in data gaps.

3. Evaluation framework

Since ground truth motion is not available, the motion correction error cannot be quantified directly. Instead, an evaluation framework is used, of which orthogonal evaluation aspects are described in each following subsection. Section 3.1 introduces two objectives of the registration and their evaluation: *alignment* between the co-registered scans, and *reproducibility*, which evaluates if residual uncorrected distortion remains. Section 3.2 describes a method to quantitatively evaluate the error separately in axial and transverse directions, by computing the *ILM position disparity* and *vessel disparity* metrics. The OCT system, the imaging protocol and the study cohort, which contains both *elderly pathologic subjects* and *young healthy controls*, is described in Section 3.3. Section 3.4 concludes with a comparison of the *OCT only* and *OCT + OCTA* algorithms.

3.1. Objectives of the registration

This section describes two objectives of the motion correction: alignment between the co-registered scans and low residual distortion. A schema is described for each objective that allows its evaluation by mapping the data of interest or derived features to a co-registered space, which allows a pointwise comparison across the volume. If the displacement field maps to a location outside the compared scans, the respective A-scans were left black in the qualitative evaluation and were extrapolated as constant in the quantitative evaluation to minimize their influence.

3.1.1. Alignment

Alignment refers to the agreement of the two input scans after motion correction. This is a relevant objective because it is necessary to achieve a sharp merged image. Improper alignment will result in blurred or even doubled features, such as thin retinal layers or vessels. For evaluation, in terminology of Fig. 1, *input volumes* or derived features are mapped to *motion corrected* space by applying the estimated displacement fields, and then compared. Alignment was evaluated for multiple acquired scan pairs for each eye and field size. This evaluation is important because it is closely related to the perceived image quality of the merged result, but it does not quantify whether distortion was corrected and also has the disadvantage that it is prone to favor overfitted registrations. Overfitting can occur when features that are only visible in one of the scans, such as small capillaries or noise, are aligned with non-corresponding features in the other image. This can lead to a reduction of disparity metrics despite increased distortion.

3.1.2. Reproducibility

The fundamental objective of motion correction is to minimize distortion. Repeated distortion-free OCT scans of the same eye should (*re*)produce structurally consistent results, when differences in perspective caused by variations in beam alignment through the pupil are accounted for. Based on these observations, the following evaluation schema was derived: Scan pairs of the same eye were acquired independently and independently motion corrected and merged. Scan alignment differences are estimated by registering each merged volume to corresponding other volumes using a constrained affine registration. In addition to the translational and rotational components of rigid transforms, tilt-induced axial sheering along the transverse dimensions is compensated. Tilt occurs when the laser beam is not exactly aligned with the pupil [25]. This registration was performed consistently with the previous paper [25] on illumination corrected OCT data using a parameterization that resembles the aforementioned relevant degrees of freedom. Figure 5 outlines the quantitative reproducibility analysis.

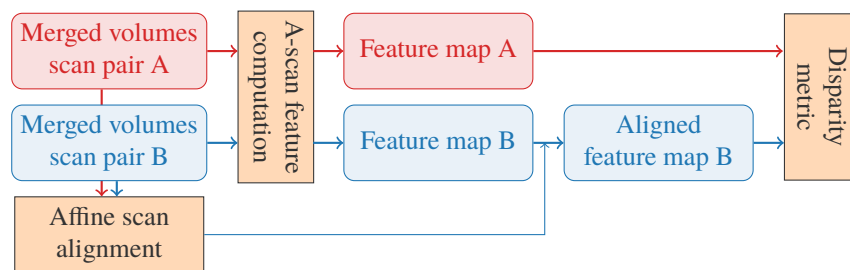


Fig. 5. Data flow of the quantitative reproducibility evaluation of independently acquired and independently motion-corrected volumes. After their computation, the A-scan features (see Section 3.2) are aligned affinely before their disparity can be measured.

Due to the rigid nature of the affine registration, there is no possibility for an overfitted co-registration and independent eye motion during acquisition of each scan pair ensures that the

residual distortion is different. The noise reduction and artifact removal during merging does not bias the algorithm comparison because it was performed consistently, as described in Section 3.4.

3.2. Feature-based, direction dependent disparity metrics

Several considerations were taken into account when choosing the metrics used for quantitative evaluation. To reduce the confounding influence of noise, due to speckle as well as variations in focus and illumination, which is more pronounced in elderly and pathologic eyes, we chose a comparison based on retinal structure features rather than an intensity-based comparison. In addition, the selected features are largely dependent on displacement along specific axes, so the evaluation can separately differentiate these components of registration error. We selected two features, the algorithmically determined likeliness of an A-scan to show a vessel, and the position of the inner limiting membrane (ILM) along an A-scan, in order to separate transverse and axial components. Evaluating these features only once per A-scan is sufficient to quantify the correction of motion-induced distortion, because this distortion only manifests between A-scans, not within them. The selected features are also advantageous because they relate to clinically relevant features and can be computed with high reliability. The final disparity score \mathcal{M} of scans X, Y is computed by averaging the differences of features m_{ij}^X, m_{ij}^Y co-registered according to 3.1.1 or 3.1.2 to the A-scan location i, j . A border of 5% of the transverse image area $w \times h$ is omitted on each side to exclude most data gaps at the boundary (see, e. g., Fig. 9).

$$\mathcal{M}(X, Y) = \frac{1}{0.9w} \cdot \frac{1}{0.9h} \cdot \sum_{i=0.05w}^{0.95w} \sum_{j=0.05h}^{0.95h} |m_{ij}^X - m_{ij}^Y| \quad (11)$$

Both chosen features depend on layer boundaries. Our segmentation approach is similar to the method of Chiu et al. [32]. Our modified approach finds multiple boundaries at once to improve reliability by allowing the incorporation of minimal and maximal layer depth constraints directly within the graph search. The graph construction is modified such that each node represents the pixels $z_1, \dots, z_{n_{\text{boundaries}}}$ with $z_1 < \dots < z_{n_{\text{boundaries}}}$ of the A-scan x , where pixel (x, z_i) corresponds to the boundary position of the i -th segmented layer. Edges are added between nodes if the corresponding pixels of all boundaries are neighbors. The edge weight is then determined by the sum of the edge weights of the individual layers.

Based on this framework, layer segmentation is performed as follows: Initial ILM and posterior retinal pigment epithelium (RPE) boundary estimates are found by a joint segmentation of the blurred and by factor 5 downsampled A-scans. Both layer estimates are subsequently smoothed with a Gaussian filter. Next, the volume is flattened with respect to the smooth RPE estimate and a joint segmentation of an estimate of the line immediately anterior to the ellipsoid zone (EZ) / inner segment-outer segment (IS-OS) junction and the final posterior RPE boundary is performed at full resolution. On the same flattened volume, the final EZ / IS-OS segmentation is performed in combination with the OS-RPE boundary. The ILM is segmented individually on a volume flattened according to the smooth ILM estimate. Finally, the layer positions are inversely shifted such that they again correspond to the true shape of the non-flattened volumes. Figure 6 shows a segmentation of the ILM and EZ / IS-OS on a representative B-scan.

3.2.1. ILM position disparity

An ILM position map is obtained by arranging the segmented ILM positions in all A-scans in an *en face* image as shown in Fig. 6. This map is used to evaluate the axial registration accuracy. For this purpose, the ILM position maps of both compared scans are mapped to registered space before computing their A-scan by A-scan difference and scaling it with the axial pixel spacing. The error metric describes the distance between the ILM segmentations in μm .

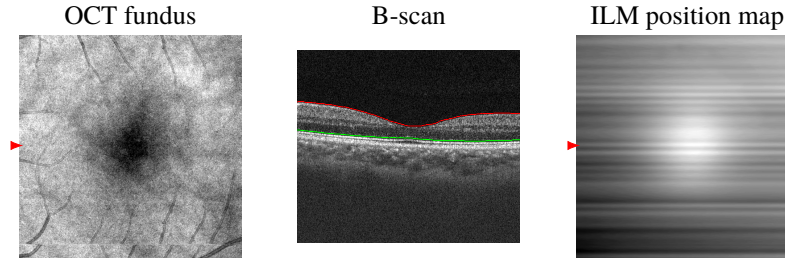


Fig. 6. Layer segmentation in a representative X-fast scan from a 28 y/o healthy subject. From left to right: OCT fundus image, B-scan at the position of the arrow in the fundus image, with segmented ILM and EZ / IS-OS, ILM position map showing deeper ILM positions in white. The image is consistent with the expected foveal contour.

3.2.2. Vessel disparity

A vessel map was used to quantify transverse registration performance, which was derived from a layer segmentation-based *en face* OCTA image. The *en face* image was computed by filtering the B-scans with a 3×3 median filter, followed by a mean projection between the ILM and the EZ / IS-OS (see Fig. 6). The EZ / IS-OS was selected because it can be reliably segmented in pathologies and because small errors in this avascular area have minimal effect on the projection. Similar to the OCTA data similarity term, white lines caused by saccadic B-scans would be wrongly detected as vasculature. These B-scans were detected based on the standard deviation of the original OCTA B-scans and replaced with their nearest valid neighbor in the *en face* image, following the method described in Section 2.2. This step is not performed when evaluating reproducibility, because in the compared merged volumes, the B-scans were already removed during volume merging (Section 2.5). Finally, in both evaluations, the background noise is normalized by mapping the 10 and 95 percentiles to black and white, before computing the vesselness according to Frangi et al. [33]. This filter is designed to emphasize bright tubular structures in multiple scales, corresponding to vessels with varying diameters. The filter result describes the likeliness of pixels (here: A-scans) showing a vessel, in a range from 0 to 1.

Figure 7 shows the original mean projected *en face* OCTA image, the processed *en face* OCTA image after replacement of saccadic B-scans and normalization, and the resulting vessel map of a representative X-fast scan, as it is used in the alignment evaluation. The error metric then describes the average difference in the co-registered vessel maps and is thus sensitive to displacement-induced blurring and vessel doubling artifacts. The range of actual scores is smaller than 1, because even perfect vessel overlaps differ by OCTA signal variations. In addition, since individual vessels are not distinguished, misaligned vessels are not detected if they happen to overlap with other vasculature.

3.3. Imaging protocol & study cohort

This study protocol was approved by the Institutional Review Boards at the Massachusetts Institute of Technology (MIT) and Tufts Medical Center. All participants were imaged in the ophthalmology clinic at the New England Eye Center (NEEC) and written informed consent was obtained prior to imaging. The research adhered to the Declaration of Helsinki and the Health Insurance Portability and Accountability Act. Data acquisition was performed with a research prototype SS-OCT instrument developed at MIT, which was described in [34] and is briefly summarized here. The system uses a swept-source vertical cavity surface-emitting laser (VCSEL) light source with a 400 kHz A-scan rate at a central wavelength of 1050 nm. Raster scanning was performed in a grid of 500×500 A-scans with 5 repeats per B-scan, for a total

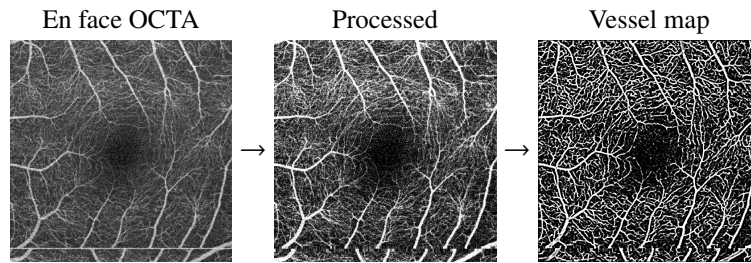


Fig. 7. Vessel map computation from the same scan as in Fig. 6. From left to right: *En face* OCTA, the image after processing and white line removal, and the vessel map.

scan time of ~ 3.9 seconds including scanner flyback. Field sizes of both 3×3 mm and 6×6 mm were acquired for each eye, thus the A-scan spacing is $6 \mu\text{m}$ and $12 \mu\text{m}$ respectively and the axial pixel spacing is $4.5 \mu\text{m}$ in tissue. Repeated scans were registered rigidly with a cross-correlation based approach according to [35] with an accuracy of $1/32$ pixels before the OCTA signal was computed using amplitude decorrelation [11].

Three pairs of orthogonal scans with 3×3 and 6×6 mm field sizes centered at the fovea were acquired in each eye. After each scan pair, the subject sat back and the instrument was realigned to ensure that reproducibility is evaluated on independent acquisitions. If a subject blinked during an acquisition, that specific scan was repeated up to two times. In total, 18 eyes from 18 subjects were scanned. The study cohort, described in detail in Table 1, included subjects with non-arteritic anterior ischemic optic neuropathy (NAION), dry age-related macular degeneration (AMD), AMD with geographic atrophy (GA), diabetes mellitus without diabetic retinopathy (DM no DR) and proliferative diabetic retinopathy (PDR) with diabetic macular edema (DME). An additional case with non-proliferative diabetic retinopathy (NPDR) was excluded from evaluation because large axial motion caused the retina to move out of the axial imaging range for wide areas in most scans. It is thus not included in the table.

Table 1. Pathology and age statistics of the study cohort. Mean age and standard deviation are given, with minimum and maximum in brackets. Exclusion of a 60 y/o NPDR patient and of specific scans in quantitative alignment and reproducibility evaluation are not related to motion correction and justified in the text.

Category	#cases	Age	All #pairs	Alignment #pairs	Reproducibility #merged pairs
NAION	1	61	6	6	6
Dry AMD	1	61	6	5	5
AMD with GA	1	83	6	5	6
DM no DR	3	62.0 ± 10.2 (52 – 76)	18	17	18
PDR with DME	1	67	6	6	6
Pathologic	7	65.5 ± 10.0 (52 – 83)	42	39	41
Healthy	10	27.1 ± 2.7 (23 – 32)	60	58	59
All	17	43.8 ± 19.9 (23 – 83)	102	97	100

The pathologic cases are from significantly older patients. Age and the presence of pathology make motion correction in this group harder due to higher prevalence of opacities, blinks and saccades. Whereas the healthy subgroup had below 3 saccades per scan on average, the subgroup with pathology had more than 4, with several scans containing 10 saccades or more.

A small fraction of scans needed to be excluded from the quantitative analysis due to failures of the layer segmentation. Severe opacities caused the ILM segmentation to erroneously segment the RPE in one dry AMD and two healthy scans. However, in both subjects, the RPE was still visible, which seemed to suffice for registration purposes: visually there were no severe issues in axial registration in the excluded cases. For one scan each of the DM no DR and AMD with GA cases, no scan without a blink could be acquired. Consequently, the ILM could not be segmented properly in the areas of the blink. One healthy scan and all scans with blinks could be segmented after volume merging due to reduced severity of opacities and filled gaps. In total, 97 of 102 scan pairs remained for alignment evaluation, and 100 of 102 merged scan pairs remained for reproducibility evaluation. This indicates that reliability and thus reproducibility of post processing steps such as layer segmentation can be improved via volume merging.

3.4. Comparison of algorithms

The *OCT only* method corresponds to the basic algorithm described in Section 2.1, whereas the *OCT + OCTA* method includes the novel extensions described in this paper. Two adaptations were made to assure a fair and meaningful comparison. First, in order to separate the evaluation of registration accuracy from the effects of white line removal as described in Section 2.5, both the *OCT only* and the *OCT + OCTA* method use white line removal when merging OCTA data. Secondly, since the ratio between data term and regularization has a major impact on registration quality (see Fig. 11), the same scaling factor η is also multiplied with the structural similarity term in the basic method analogously to Eq. (9). This ensures that identical regularization weights α correspond to equal ratios between data and regularization terms, and thus that the influence of this ratio is removed as a confounding variable when comparing the registration quality of both methods at identical α values. Note that these adjustments make the basic algorithm equivalent to the proposed method with an angiography weight $\delta_0 = 0$. Statistical testing was performed with a two-sided Wilcoxon signed rank test at the 0.01 (**) and 0.001 (***) significance levels.

4. Results

We begin with a qualitative visualization of alignment in a representative case from the elderly and pathologic subgroup, where we demonstrate better co-registration, fewer artifacts and increased sharpness resulting from the approach with OCTA data. Furthermore, we show a visualization of residual uncorrected distortion and volume renderings. We then quantitatively verify generality of the observations using registration aspects highlighted in the previous chapter. This includes clinically relevant metrics from the literature and an investigation of differences between the healthy and elderly pathologic subgroups. We conclude the results with a description of runtime.

4.1. Qualitative results

Figure 8 demonstrates the potential of OCTA data for improving registration quality by reducing or preventing artifacts that would otherwise compromise volume merging. Red triangles mark 5 saccades that occurred in rapid succession during the X-fast scan.

The yellow box outlines an area having a small misregistration in the *OCT only* approach ($\alpha = 1$, justified below). Due to the misalignment, the larger horizontal vessel appears blurred, because at the boundary the vessel of one scan is merged with background in the other scan. Tiny structures like the vertical capillaries become hard to identify. A more severe misregistration in the *OCT only* merged volume is outlined by the orange box. This originates from an area with relatively low contrast in the *OCT* input volumes. Since the misalignment exceeds the vessel diameter, the same vessel appears twice (vessel doubling), once from each merged scan. In the composite image, the coloring reveals the origin of both appearances, and that the red X-fast data is misaligned, which contains a saccade in close proximity. The corresponding boxes in the *OCT*

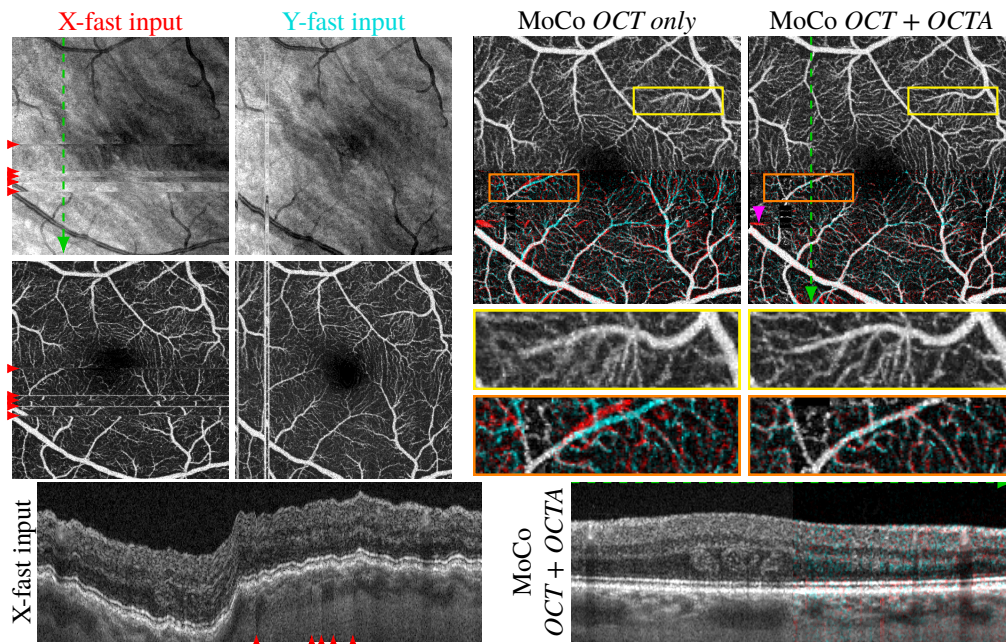


Fig. 8. Increased sharpness in merged volume data when using OCTA data during motion correction (MoCo) in the 61 y/o NAION subject. The merged images are split vertically and are displayed in regular grayscale in the upper half and in composite color in the lower half. Here, the X-fast scan is displayed in red and the Y-fast scan in cyan. If the contributions from both scans are equal, the colors sum to a gray value. Areas with contribution from only one volume are also colored gray in order to emphasize misalignments. The composite images are median filtered to reduce color from speckle noise, and contrast enhanced to emphasize the structural differences. The OCT B-scans are extracted along the dashed green arrows.

+ OCTA merged volume ($\alpha = 0.1$) do not show these artifacts. However, as pointed out by the pink triangle, a small misregistration of the X-fast data between the two lowest saccades remains.

The B-scan slices are extracted along the vertical axis (dashed arrows in *en face* images). Since the X-fast slice is oriented in slow scanning direction, it is severely distorted. In the motion corrected slice, the red X-fast data does not show distortion and agrees with the Y-fast B-scan.

In the 3×3 mm field size scans, results from 1 healthy and 7 pathologic eyes had a double vessel artifact resulting from a single, small size misregistered area between saccadic motions. The distance between the double vessels depends mostly on saccade amplitude. Furthermore, 3 results from eyes with pathology were heavily misregistered and not usable because of large amplitude saccades or a large number of saccades (≥ 12 in both scans) combined with opacities. Occasional blurring occurred in the 6×6 mm results as well, however double vessels did not occur.

Figure 9 shows the remaining distortion between the OCT + OCTA merged volume in Fig. 8 and an additional merged volume from an independently scanned and motion corrected scan pair (left), which was, as a whole, affinely registered to the first as described in Section 3.1.2. The *en face* images of the merged volumes differ at the boundaries of certain vessels (orange markers). These differences reveal a 2 – 3 pixel amplitude, low frequency residual transverse distortion in the peripheral areas. The pink triangle points to a location of misalignment in the first scan pair (see Fig. 8), which also contributes to the difference between the merged volumes. The B-scan extracted along the dashed green arrow shows the highest degree of axial distortion,

which changes direction roughly at the center of the scan and might be related to incompletely corrected tilt differences caused by different pupil alignment. Note that the affine registration achieves its best alignment in the center, where the extracted B-scan (dashed yellow arrow) shows excellent agreement. However, like affine registration of a single B-scan only, this is insufficient to capture the full distortion in the volume, as apparent from the distortion in the other B-scan.

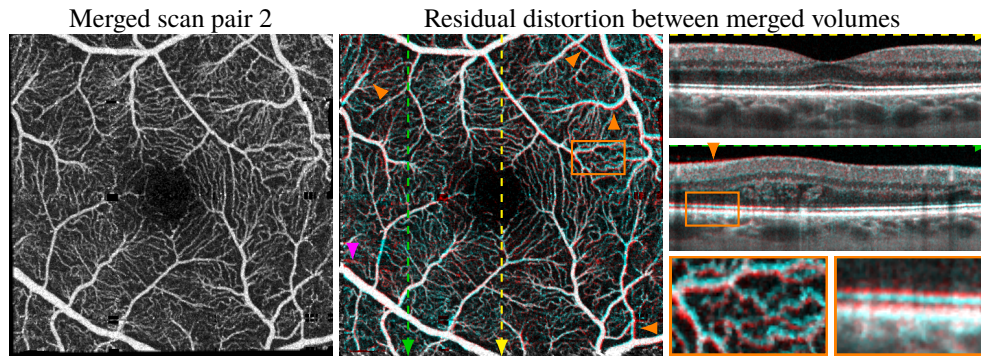


Fig. 9. Residual distortion between independently acquired and independently *OCT + OCTA* motion-corrected scan pairs of the same eye. The compared volumes are shown in red (merged volume in Fig. 8) and cyan (left image). Zooms show distorted areas.

We conclude the qualitative evaluation with volumetric renderings of a 12×12 mm, 800×800 A-scans scan pair and the merged results (Fig. 10). The *OCTA* volume is limited along depth to only show retinal vasculature and uses shadow artifact removal as described in [36].

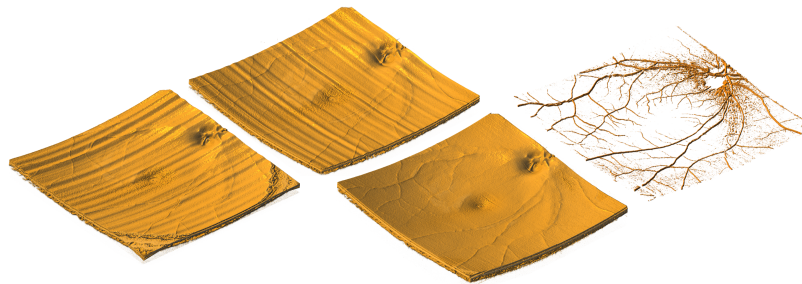


Fig. 10. From left to right: X-fast and Y-fast inputs, and merged *OCT* and *OCTA* volumetric renderings of a healthy eye at equal axial and transverse scale.

4.2. Quantitative results

We begin by presenting algorithm performance with respect to varying regularization weight α to justify our chosen algorithm configurations. We then continue with a boxplot subgroup analysis between young healthy and elderly pathologic cases. In both figures, we show the evaluation objectives side by side, the metrics one below the other and plot both algorithms to give insights on how co-alignment vs. distortion correction (horizontally), axial vs. transverse error components (vertically) and the algorithms are influenced by control parameters.

Figure 11 shows the mean and standard deviation (over all scan pairs) of mean feature disparity (between the co-registered volumes) for motion correction with varying regularization weight α . Both approaches clearly outperform the uncorrected data in terms of ILM position disparity. For reasonable parameter choices this result also holds for vessel disparity. Only for the smallest α

value 0.001 does the *OCT only* approach show slightly worse reproducibility. Furthermore, for equal regularization weights, the *OCT + OCTA* motion correction consistently outperforms the *OCT only* approach in terms of vessel disparity, and shows equivalent or lower ILM position disparity. In terms of alignment, both disparity metrics show improvement with decreasing regularization weight. However, reproducibility results reveal that residual distortion increases again below a certain regularization weight, suggesting that improvements in alignment beyond this point stem from overfitting with implausible motion fields or wrong convergence. In addition to the general reduction of vessel disparity, the *OCT + OCTA* approach also seems to have improved stability. The mean vessel disparities varied in a smaller range compared to *OCT only* throughout the evaluated regularization weights, with lower standard deviations. The slightly lower best mean vessel disparities in reproducibility compared to alignment may result from the fact that the compared merged volumes only have small residual gaps, whereas the original scans used for alignment evaluation use interpolated data for whole white line B-scans (Fig. 7).

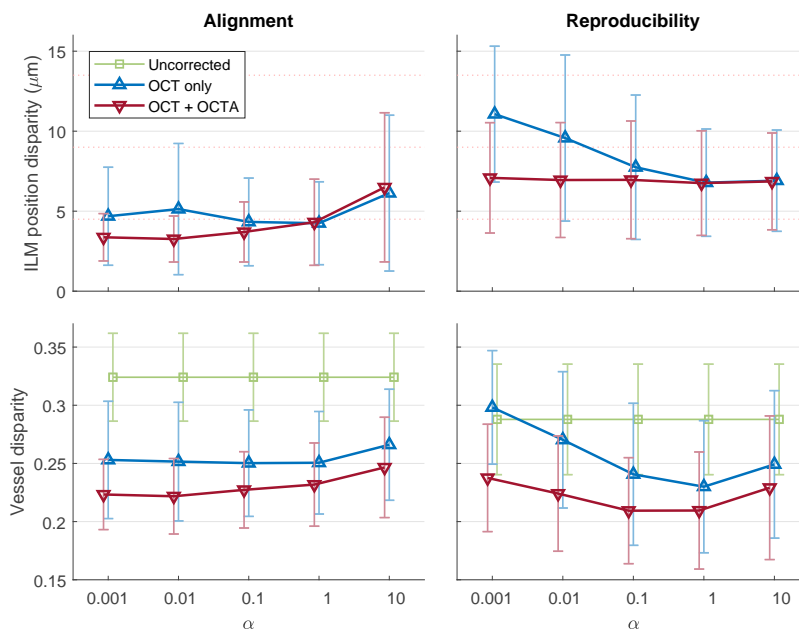


Fig. 11. Mean of mean feature disparity under 5 degrees of regularization, specified by the weight α (slightly offset along the x-axis for visibility). Error bars indicate one standard deviation. For ILM position disparity, dotted grid lines show the 4.5 μm pixel spacing. The uncorrected ILM position disparities, 140 and 45 μm , are omitted for reasonable scaling. Uncorrected reproducibility was evaluated on the X-fast scans.

In the *OCT + OCTA* evaluations of Fig. 11, δ was fixed to 0.5. We also evaluated corresponding parameter choices with $\delta = 0.25$ and 0.75, but the difference was marginal (within 7% standard deviation of $\delta = 0.5$). This is no surprise, because δ primarily affects the value of the objective function, but the location of the minimum, which corresponds to the optimal displacement field, remains where both OCT and OCTA data are aligned. Only for δ values close to 0 does the contribution of the OCTA component become negligible compared to the noise of the OCT term and registration reduces to OCT only registration. The opposite happens for δ close to 1, with the additional drawback that, due to the lack of an axial dimension in the projected OCTA data, axial motion is no longer corrected. Thus, $\delta = 0.5$ is the natural choice.

Figure 12 compares the healthy and the pathologic subgroups after motion correction with the optimal parameter configurations of the two algorithms. Since the α values 0.1 and 1 performed similarly in terms of reproducibility, we chose $\alpha = 0.1$ for the *OCT + OCTA* approach for its better alignment scores. For the *OCT only* approach, we chose $\alpha = 1$ because the marginal improvement in vessel alignment using $\alpha = 0.1$ does not justify the loss of reproducibility. Again, for ILM disparity, the results are very similar. In the alignment evaluation, only the pathological subgroup showed a notable improvement for the *OCT + OCTA* approach. However, the improvement does not transfer to reproducibility and should thus be considered carefully. Vessel disparity shows a consistent improvement, which was found significant in both subgroups for both alignment and reproducibility. The boxplot also reveals that larger improvements in mean disparity are predominantly achieved in cases with comparatively worse results in the *OCT only* approach: While the lower quartiles in general, and, for the healthy subgroup, the median, show a small improvement when using OCTA data, the improvement is much more pronounced in the upper quartiles and the median of the pathological subgroup. Although using OCTA data reduced the performance gap between healthy and pathological cases, a significant difference remains. In order to relate the vessel disparity metric with a visual example, we marked the scores of the scan pair shown in Figs. 8 and 9 with white crosses.

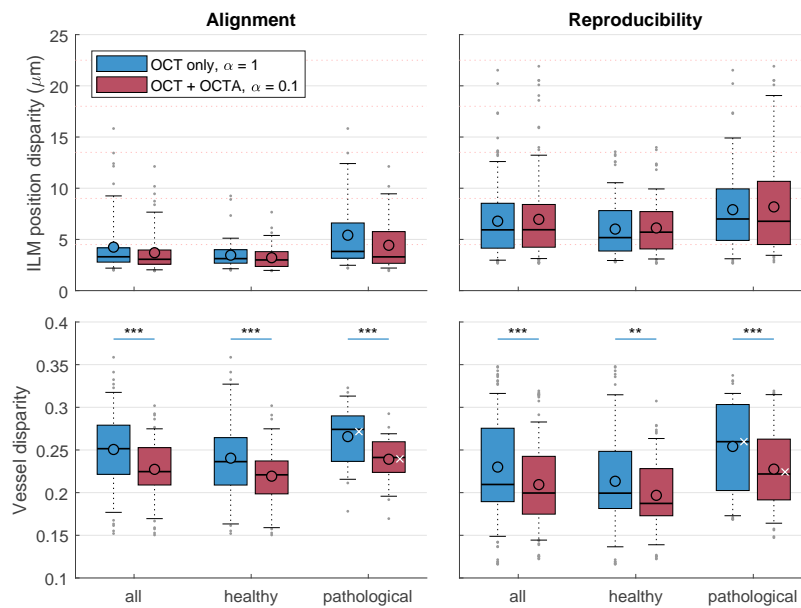


Fig. 12. Box plot subgroup analysis of mean feature disparity. Whiskers represent the 5th and 95th percentiles. Circles show the mean of mean feature disparity analogously to Fig. 11. For ILM disparity, dotted grid lines show the $4.5 \mu\text{m}$ pixel spacing. White crosses mark the scores of the scans shown in Figs. 8 and 9.

Tables 2 and 3 report metrics on alignment and reproducibility for the *OCT + OCTA* algorithm, again using $\alpha = 0.1$ and $\delta = 0.5$. Naturally, median errors are lower due to the low influence of outliers, which is why we use the more comprehensive mean errors in our plots. Mean ILM position disparity in terms of alignment using the *OCT + OCTA* approach is below the pixel spacing ($3.7 \mu\text{m}$ vs. $4.5 \mu\text{m}$). Metrics are higher for reproducibility ($7.0 \mu\text{m}$), and reveal that there is residual distortion in addition to the misalignment.

Table 2. Disparity metrics based on the X-fast and Y-fast volume alignment without (uncorrected) and with motion correction (new) specified as mean \pm standard deviation over the volume-aggregated metrics. The metrics are median and mean ILM position disparity, both in μm , and mean vessel disparity, which is scaled by 10.

Alignment	all		healthy		pathologic	
	uncorrected	new	uncorrected	new	uncorrected	new
median ILM	135 \pm 106	2.9 \pm 1.3	126 \pm 101	2.7 \pm 1.1	148 \pm 111	3.1 \pm 1.5
mean ILM	140 \pm 104	3.7 \pm 1.9	128 \pm 101	3.2 \pm 1.1	157 \pm 109	4.4 \pm 2.5
mean vessel	3.2 \pm 0.4	2.3 \pm 0.3	3.2 \pm 0.4	2.2 \pm 0.3	3.4 \pm 0.3	2.4 \pm 0.3

Table 3. Disparity metrics based on the reproducibility of merged volumes (new) specified as mean \pm standard deviation over the volume-aggregated metrics. Uncorrected results are based on the X-fast scans. The metrics are median and mean ILM position disparity, both in μm , and mean vessel disparity, which is scaled by 10.

Reproducibility	all		healthy		pathologic	
	uncorrected	new	uncorrected	new	uncorrected	new
median ILM	35 \pm 25	5.8 \pm 3.3	23 \pm 15	5.2 \pm 2.3	52 \pm 27	6.7 \pm 4.1
mean ILM	45 \pm 33	7.0 \pm 3.7	28 \pm 16	6.1 \pm 2.5	69 \pm 36	8.2 \pm 4.7
mean vessel	2.9 \pm 0.5	2.1 \pm 0.5	2.7 \pm 0.5	2.0 \pm 0.4	3.1 \pm 0.3	2.3 \pm 0.5

4.3. Runtime

Table 4 lists the runtime of the optimal configuration for different volume sizes on a test system with an Intel Core i7-6700K CPU, 32 GB RAM and an Nvidia GeForce GTX 960 GPU (GM206, 4 GB), which corresponds to current entry-level GPUs. The runtime statistic of the $500^2 \times 465$ voxel volumes is based on the full evaluation set when using the optimal configuration. Except for the GPU-implemented data similarity term evaluation and few steps during preprocessing, calculations are performed on the CPU and were not optimized. This includes the tilt estimation, which is only performed during prealignment. Thus, the relative scaling between the two registration steps is likely misleading. Furthermore, despite the fact that some memory copy bottlenecks were removed, the ongoing need for memory copies between CPU and GPU especially during optimization causes unnecessary overhead. In total, the GPU compute time comprises only $\sim 19\%$ of the runtime. There is thus potential for further speedup.

Table 4. Runtime of selected processing steps and volume sizes. Read / write includes the time necessary to read the input from disk and to write the merged volumes back. In a clinical setting, input data would be available in memory and display would be more important than writing to disk. Registration timings include their respective preprocessing. Total includes additional steps such as saccade detection and merging.

#voxels	read / write	prealignment	3-D registration	total
$500^2 \times 465$	11.5 \pm 0.9 s	26.7 \pm 0.2 s	31.4 \pm 8.0 s	86.1 \pm 8.0 s
$800^2 \times 433$	16.5 \pm 0.2 s	49.6 \pm 0.3 s	83.7 \pm 18.8 s	181.8 \pm 19.6 s

While the 2-D *en face* OCTA data term evaluation is negligible compared to the volumetric OCT term, the runtime of the *en face* projection itself must still be shown to be small. We stopped optimizing its GPU implementation at a runtime of ~ 209 ms per volume, of which a large fraction of time is used to sort voxel intensities for the quantile-based normalization. Precomputing the normalization mapping for the specific system reduced the runtime to ~ 66 ms. We also ran the algorithm on 5 scan pairs with $800^2 \times 433$ voxels (2 B-scan repeats) over a 12×12 mm field-of-view to prove feasibility for even larger wide-field datasets with the same

system. We used the same algorithm configuration, up to a correspondingly larger displacement field (800^2 per volume) and a fixed normalization mapping for the OCTA projection.

5. Discussion

The results presented in this paper show that a sophisticated displacement model combined with the fine vascular features of OCTA can improve post processing-based motion correction with negligible additional computational cost. Overall scan distortion is reduced to micrometer scale and subpixel alignment is achieved improving visibility of capillaries and thin retinal layers. While both compared methods performed similarly well in the subgroup of younger healthy subjects, the *OCT only* method had occasional blurred or misaligned vessels appear predominantly in the elderly and pathologic group. Both these misalignment artifacts were reduced in severity and frequency in the new *OCT + OCTA* approach, which produces consistent, sharp images with higher reliability, especially in the clinically relevant subgroup with pathology.

Interpretation of the quantitative results must consider the whole evaluation pipeline, comprised of the acquisition with its underlying motion, the OCT scanner with its resolution and signal-to-noise, layer segmentation and vessel map computation, the motion correction itself, the mapping of the feature maps, and the affine registration used for reproducibility evaluation. Thus, even perfect motion correction with a zero registration error cannot achieve zero disparity metrics.

As the purpose of regularization is to enforce a more plausible displacement field, the selection of the optimal regularization weight α should primarily be based on reproducibility, not alignment. Since ILM disparity did not vary significantly for the *OCT + OCTA* approach (see Fig. 11), we chose $\alpha = 0.1$ based on vessel disparity as the best configuration. However, since $\alpha = 1$ shows a similar low disparity, the optimal choice is likely in between. In contrast, the *OCT only* approach has a clear minimum at $\alpha = 1$. This suggests that when using OCTA data, the influence of regularization can be slightly lowered. Potential explanations for this phenomenon are the use of brightness normalized amplitude decorrelation OCTA and that the finer transverse features in OCTA data strengthen the importance of the data term in general. However, it is evident that both approaches drastically lose reproducibility when reducing α below 0.1, emphasizing the importance of using appropriate regularization in OCT motion correction.

Although the performance difference between the young healthy versus elderly pathologic groups could be reduced using OCTA data, as shown in Fig. 12, a significant difference between these groups remains. Because the new method uses a layer segmentation-free, intensity-based similarity metric, structural changes due to pathologies should not impede registration accuracy. While there are specific diseases that reduce blood flow in larger regions and thereby cause a reduction of OCTA features, many pathologies introduce new structures, e. g. neovascularizations, or increase the visibility of choroidal vasculature, like geographic atrophy, which can be used for registration. Instead, we believe that the performance difference is caused by a combination of the following reasons: First, the accuracy of feature computations underlying the two disparity metrics could be negatively influenced by pathologies and the higher incidence of vitreous opacities in elderly subjects. Secondly, since vitreous opacities can cause OCT / OCTA signals to change between repeated scans, matching is complicated. And thirdly, elderly patients often have more saccades which not only increases the chance for misregistrations, but also makes the registration of the thinner saccade-free regions harder. This is why some approaches reject small regions based on size, disagreement with co-registered data or combinations thereof [23,28,37]. These factors emphasize the importance of evaluating registration methods on elderly patients and in pathology. This is also the majority of data encountered in actual clinical practice.

Kraus et al. [25] used a similar evaluation pipeline. However, in addition to different patient demographics, their OCT acquisition times were shorter (approximately one half the scan time in our study) and transverse sampling was sparser but axial pixel spacing was finer ($3.1 \mu\text{m}$ vs $4.5 \mu\text{m}$). The finer axial pixel spacing is likely the main reason why they achieved a mean of mean

ILM reproducibility error of slightly below $7\ \mu\text{m}$, compared to $7.0\ \mu\text{m}$ in this paper. A further difference is that their blood vessel maps were extracted from the shadow artifacts in the inner segment-RPE region, which do not show capillary scale vasculature. *En face* OCTA images show these fine structures and thus necessitate higher transverse registration accuracy (yellow box in Fig. 8). Combined with denser transverse sampling, our vessel disparity metric should be more sensitive, but it is not directly comparable to the metric in Kraus et al. [25].

In contrast to the reproducibility-based evaluations, Lezama et al. base their residual distortion evaluation on simulated distortion [26]. Differences in evaluation, like comparison of individual scans to ground truth, a differently computed vessel metric, motion simulation that is closely in line with the correction method and restricted to at most 4 saccades do not allow a direct comparison. In terms of alignment, a mean of median ILM disparity of $5.0\ \mu\text{m}$ was reported in wide-field scans from healthy subjects, which is slightly above their pixel spacing. Their approach is based on an ILM segmentation, where the alignment depends on a few pixels that define a single layer transition. Our substantially lower disparity ($2.7\ \mu\text{m}$ in healthy subjects) might be a consequence of using intensity-based alignment as introduced in [21], which depends on all pixels and thus all layer transitions along the A-scan. This improves robustness and avoids loss of spatial resolution by transverse smoothing. Although Lezama et al. did not use OCTA data for registration, qualitative results show good transverse agreement with occasional discontinuities and double vessel artifacts in their composite OCT images. In their B-scan composite images of co-aligned scans, occasional bands of green and pink pixels above and below layer transitions such as around the RPE reveal axial misalignment not present in our results (Fig. 8).

Chen et al. qualitatively evaluated a lissajous scan pattern based, OCT only, 3-D software motion correction approach on healthy subjects [28]. In addition to good transverse alignment without double vessel misregistration, they achieved high reproducibility in *en face* OCT images. Although the distortion correction could be transferred to *en face* OCTA images in their follow-up paper on 2-D *en face* motion correction [29], smaller capillary-scale vasculature was blurred.

Kim et al. [38] used a scanner that acquires both orthogonal volumes at the same time. At the cost of increased system complexity, this simplifies the motion correction, because only half the number of displacements need to be estimated. Still, compared with their results, the reproducibility of the representative pathological case shown in Fig. 9 using our method is higher.

In addition to motion correction accuracy, clinical use poses additional requirements on acquisition and computation time to reduce patient fatigue and maintain patient throughput. Both orthogonal and lissajous scans have a similar acquisition time due to oversampling by a factor of ~ 2 to enable registration. However, the reported average computation times for similar volume size 3-D motion corrections, 70 min (Chen et al.) and 86 s (ours) for dense, standard field-of-view scans and 14 min (Lezama et al.) and 3.0 min (ours) for wide-field scans vary substantially.

Despite the overall high accuracy and clinical applicability, our method has limitations. Compared to the methods of Kraus et al. or Lezama et al., our algorithm can only be used on OCTA datasets. Furthermore, the current implementation is restricted to square volumes. However, since rectangular raster scans typically extend further in the fast scan direction, the (square) overlapping part of two orthogonal scans could be registered normally and the displacement field could be extended along the B-scans with minimal error. Finally, our method can have a low amount of distortion and, as mentioned above, pathologic cases tend to have worse motion correction results, which show up as occasional blurring or double vessels.

6. Conclusion

We extended the orthogonal scan-based 3-D OCT motion correction method by Kraus et al. to handle and utilize OCTA data. We quantitatively compared the new method with the previous approach using a dataset with 102 scan pairs containing both healthy and a wide range of pathologic cases. Metrics were related to visual evaluation by showing composite

images of the most representative pathologic case. We achieved a significant improvement in transverse alignment, which also enabled significantly reduced residual distortion of the motion corrected results, especially in the clinically relevant elderly pathologic subgroup. In this group, larger improvements were achieved, which demonstrate reliability of the new method in pathology. Performance approaches the control group despite more frequent and larger eye motion. Furthermore, we showed that the additional steps have marginal impact on runtime, preserving the clinical utility. This new method has potential to further increase reproducibility of OCT(A)-derived disease metrics and to minimize inconsistencies between follow-up scans.

Funding

Deutsche Forschungsgemeinschaft (MA 4898/12-1); National Institutes of Health (5-R01-EY011289-31); Retina Research Foundation; Beckman-Argyros Award in Vision Research; Champalimaud Vision Award; Massachusetts Lions Eye Research Fund; Macula Vision Research Foundation; Research to Prevent Blindness.

Acknowledgements

We thank Tobias Geimer, Ben Potsaid, ByungKun Lee, and Chen Lu for valuable discussions.

Disclosures

SBP: IP related to VISTA-OCTA (P). MFK: Optovue (C, P), currently with Siemens Healthineers. EMM: IP related to VISTA-OCTA (P). NKW: Optovue (C), Carl Zeiss Meditec (F), Heidelberg Engineering (F), Nidek (F). JSD: Optovue (C, F), Topcon (C, F), Carl Zeiss Meditec (C, F). JGF: Optovue (I, P), Topcon (F), IP related to VISTA-OCTA (P).

References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," *Science* **254**(5035), 1178–1181 (1991).
2. J. Fujimoto and E. Swanson, "The development, commercialization, and impact of optical coherence tomography," *Invest. Ophthalmol. Visual Sci.* **57**(9), OCT1 (2016).
3. M. Wojtkowski, R. Leitgeb, A. Kowalczyk, T. Bajraszewski, and A. F. Fercher, "In vivo human retinal imaging by fourier domain optical coherence tomography," *J. Biomed. Opt.* **7**(3), 457 (2002).
4. S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," *Nat. Rev. Neurosci.* **5**(3), 229–240 (2004).
5. B. Potsaid, B. Baumann, D. Huang, S. Barry, A. E. Cable, J. S. Schuman, J. S. Duker, and J. G. Fujimoto, "Ultrahigh speed 1050nm swept source / fourier domain oct retinal and anterior segment imaging at 100,000 to 400,000 axial scans per second," *Opt. Express* **18**(19), 20029–20048 (2010).
6. T. Klein and R. Huber, "High-speed OCT light sources and systems," *Biomed. Opt. Express* **8**(2), 828–859 (2017).
7. J. P. Kolb, T. Klein, C. L. Kufner, W. Wieser, A. S. Neubauer, and R. Huber, "Ultra-widefield retinal MHz-OCT imaging with up to 100 degrees viewing angle," *Biomed. Opt. Express* **6**(5), 1534–1552 (2015).
8. S. Makita, Y. Hong, M. Yamanari, T. Yatagai, and Y. Yasuno, "Optical coherence angiography," *Opt. Express* **14**(17), 7821–7840 (2006).
9. R. K. Wang, S. L. Jacques, Z. Ma, S. Hurst, S. R. Hanson, and A. Gruber, "Three dimensional optical angiography," *Opt. Express* **15**(7), 4083–4097 (2007).
10. A. Mariampillai, B. A. Standish, E. H. Moriyama, M. Khurana, N. R. Munce, M. K. K. Leung, J. Jiang, A. Cable, B. C. Wilson, I. A. Vitkin, and V. X. D. Yang, "Speckle variance detection of microvasculature using swept-source optical coherence tomography," *Opt. Lett.* **33**(13), 1530–1532 (2008).
11. Y. Jia, O. Tan, J. Tokayer, B. Potsaid, Y. Wang, J. J. Liu, M. F. Kraus, H. Subhash, J. G. Fujimoto, J. Hornegger, and D. Huang, "Split-spectrum amplitude-decorrelation angiography with optical coherence tomography," *Opt. Express* **20**(4), 4710–4725 (2012).
12. J. Tokayer, Y. Jia, A.-H. Dhalla, and D. Huang, "Blood flow velocity quantification using split-spectrum amplitude-decorrelation angiography with optical coherence tomography," *Biomed. Opt. Express* **4**(10), 1909–1924 (2013).
13. S. B. Ploner, E. M. Moul, W. Choi, N. K. Waheed, B. Lee, E. A. Novais, E. D. Cole, B. Potsaid, L. Husvogt, J. Schottenhamml, A. Maier, P. J. Rosenfeld, J. S. Duker, J. Hornegger, and J. G. Fujimoto, "Toward quantitative optical coherence tomography angiography," *Retina* **36**, S118–S126 (2016).
14. L. Sánchez Brea, D. Andrade De Jesus, M. F. Shirazi, M. Pircher, T. van Walsum, and S. Klein, "Review on retrospective procedures to correct retinal motion artefacts in OCT imaging," *Appl. Sci.* **9**(13), 2700 (2019).

15. C. Enders, G. E. Lang, J. Dreyhaupt, M. Loidl, G. K. Lang, J. U. Werner, and D. G. Vavvas, "Quantity and quality of image artifacts in optical coherence tomography angiography," *PLoS One* **14**(1), e0210505 (2019).
16. J. L. Laueremann, A. K. Woetzel, M. Treder, M. Alnawaiseh, C. R. Clemens, N. Eter, and F. Alten, "Prevalences of segmentation errors and motion artifacts in OCT-angiography differ among retinal diseases," *Graefe's Arch. Clin. Exp. Ophthalmol.* **256**(10), 1807–1816 (2018).
17. R. D. Ferguson, D. X. Hammer, L. A. Paunescu, S. Beaton, and J. S. Schuman, "Tracking optical coherence tomography," *Opt. Lett.* **29**(18), 2139–2141 (2004).
18. K. V. Vienola, B. Braaf, C. K. Sheehy, Q. Yang, P. Tiruveedhula, D. W. Arathorn, J. F. de Boer, and A. Roorda, "Real-time eye motion compensation for OCT imaging with tracking SLO," *Biomed. Opt. Express* **3**(11), 2950–2963 (2012).
19. F. LaRocca, D. Nankivil, S. Farsiu, and J. A. Izatt, "Handheld simultaneous scanning laser ophthalmoscopy and optical coherence tomography system," *Biomed. Opt. Express* **4**(11), 2307–2321 (2013).
20. S. Ricco, M. Chen, H. Ishikawa, G. Wollstein, and J. Schuman, "Correcting motion artifacts in retinal spectral domain optical coherence tomography via image registration," in *Medical image computing and computer-assisted intervention - MICCAI 2009*, vol. 5761 of *Lecture Notes in Computer Science* G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, eds. (Springer, Berlin, 2009), pp. 100–107.
21. M. F. Kraus, B. Potsaid, M. A. Mayer, R. Bock, B. Baumann, J. J. Liu, J. Hornegger, and J. G. Fujimoto, "Motion correction in optical coherence tomography volumes on a per a-scan basis using orthogonal scan patterns," *Biomed. Opt. Express* **3**(6), 1182–1199 (2012).
22. M. Heisler, S. Lee, Z. Mammo, Y. Jian, M. Ju, A. Merkur, E. Navajas, C. Balaratnasingam, M. F. Beg, and M. V. Sarunic, "Strip-based registration of serially acquired optical coherence tomography angiography," *J. Biomed. Opt.* **22**(3), 036007 (2017).
23. H. C. Hendargo, R. Estrada, S. J. Chiu, C. Tomasi, S. Farsiu, and J. A. Izatt, "Automated non-rigid registration and mosaicing for robust imaging of distinct retinal capillary beds using speckle variance optical coherence tomography," *Biomed. Opt. Express* **4**(6), 803–821 (2013).
24. P. Zang, G. Liu, M. Zhang, C. Dongye, J. Wang, A. D. Pechauer, T. S. Hwang, D. J. Wilson, D. Huang, D. Li, and Y. Jia, "Automated motion correction using parallel-strip registration for wide-field en face OCT angiogram," *Biomed. Opt. Express* **7**(7), 2823–2836 (2016).
25. M. F. Kraus, J. J. Liu, J. Schottenhamml, C.-L. Chen, A. Budai, L. Branchini, T. Ko, H. Ishikawa, G. Wollstein, J. Schuman, J. S. Duker, J. G. Fujimoto, and J. Hornegger, "Quantitative 3D-OCT motion correction with tilt and illumination correction, robust similarity measure and regularization," *Biomed. Opt. Express* **5**(8), 2591–2613 (2014).
26. J. Lezama, D. Mukherjee, R. P. McNabb, G. Sapiro, A. N. Kuo, and S. Farsiu, "Segmentation guided registration of wide field-of-view retinal optical coherence tomography volumes," *Biomed. Opt. Express* **7**(12), 4827–4846 (2016).
27. P. Zang, G. Liu, M. Zhang, J. Wang, T. S. Hwang, D. J. Wilson, D. Huang, D. Li, and Y. Jia, "Automated three-dimensional registration and volume rebuilding for wide-field angiographic and structural optical coherence tomography," *J. Biomed. Opt.* **22**(2), 026001 (2017).
28. Y. Chen, Y.-J. Hong, S. Makita, and Y. Yasuno, "Three-dimensional eye motion correction by Lissajous scan optical coherence tomography," *Biomed. Opt. Express* **8**(3), 1783–1802 (2017).
29. Y. Chen, Y.-J. Hong, S. Makita, and Y. Yasuno, "Eye-motion-corrected optical coherence tomography angiography using Lissajous scanning," *Biomed. Opt. Express* **9**(3), 1111–1129 (2018).
30. A. Camino, M. Zhang, S. S. Gao, T. S. Hwang, U. Sharma, D. J. Wilson, D. Huang, and Y. Jia, "Evaluation of artifact reduction in optical coherence tomography angiography with real-time tracking and motion correction technology," *Biomed. Opt. Express* **7**(10), 3905–3915 (2016).
31. G. Perrot, S. Domas, and R. Couturier, "Fine-tuned high-speed implementation of a GPU-based median filter," *J. Sign. Process Syst.* **75**(3), 185–190 (2014).
32. S. J. Chiu, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in sdopt images congruent with expert manual segmentation," *Opt. Express* **18**(18), 19413–19414 (2010).
33. A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," *Lecture Notes in Comput. Sci.* **1496**, 130–137 (1998).
34. W. Choi, K. J. Mohler, B. Potsaid, C. D. Lu, J. J. Liu, V. Jayaraman, A. E. Cable, J. S. Duker, R. Huber, and J. G. Fujimoto, "Choriocapillaris and choroidal microvasculature imaging with ultrahigh speed OCT angiography," *PLoS One* **8**(12), e81499 (2013).
35. J. R. Fienup and A. M. Kowalczyk, "Phase retrieval for a complex-valued object by using a low-resolution image," *J. Opt. Soc. Am. A* **7**(3), 450–458 (1990).
36. S. B. Ploner, E. M. Moul, J. Schottenhamml, L. Husvogt, C. D. Lu, C. B. Rebhun, A. Y. Alibhai, J. S. Duker, N. K. Waheed, A. K. Maier, and J. G. Fujimoto, "Hybrid OCT-OCTA vessel visualization for projection-free display of the intermediate and deep retinal plexuses," *Invest. Ophthalmol. Vis. Sci.* **58**, 638 (2017).
37. S. B. Ploner, J. Schottenhamml, E. Moul, L. Husvogt, A. Y. Alibhai, N. K. Waheed, J. S. Duker, J. G. Fujimoto, and A. K. Maier, "Correction of artifacts from misregistered B-scans in orthogonally scanned and registered OCT angiography," *Invest. Ophthalmol. Vis. Sci.* **60**, 3097 (2019).
38. H.-J. Kim, B. J. Song, Y. Choi, and B.-M. Kim, "Cross-scanning optical coherence tomography angiography for eye motion correction," *J. Biophotonics* **13**(9), e202000170 (2020).