Quantitative 3D-OCT motion correction with tilt and illumination correction, robust similarity measure and regularization

Martin F. Kraus,1,2,3* Jonathan J. Liu,3 Julia Schottenhamml,1 Chieh-Li Chen,6,7 Attila Budai,1,2 Lauren Branchini,4 Tony Ko,5 Hiroshi Ishikawa,6,7 Gadi Wollstein,6 Joel Schuman,6,7 Jay S. Duker,4 James G. Fujimoto2 and Joachim Hornegger1,2

1Pattern Recognition Lab, University Erlangen-Nürnberg, D-91058 Erlangen, Germany
2School of Advanced Optical Technologies (SAOT), University Erlangen-Nürnberg, D-91058 Erlangen, Germany
3Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
4New England Eye Center and Tufts Medical Center, Tufts University, Boston, MA 02116, USA
5Optovue Inc., Fremont, CA 94538, USA
6Department of Ophthalmology, UPMC Eye Center, Pittsburgh, PA 15213, USA
7Department of Bioengineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA, USA

*martin.kraus@fau.de

Abstract: Variability in illumination, signal quality, tilt and the amount of motion pose challenges for post-processing based 3D-OCT motion correction algorithms. We present an advanced 3D-OCT motion correction algorithm using image registration and orthogonal raster scan patterns aimed at addressing these challenges. An intensity similarity measure using the pseudo Huber norm and a regularization scheme based on a pseudo L0.5 norm are introduced. A two-stage registration approach was developed. In the first stage, only axial motion and axial tilt are coarsely corrected. This result is then used as the starting point for a second stage full optimization. In preprocessing, a bias field estimation based approach to correct illumination differences in the input volumes is employed. Quantitative evaluation was performed using a large set of data acquired from 73 healthy and glaucomatous eyes using SD-OCT systems. OCT volumes of both the optic nerve head and the macula region acquired with three independent orthogonal volume pairs for each location were used to assess reproducibility. The advanced motion correction algorithm using the techniques presented in this paper was compared to a basic algorithm corresponding to an earlier version and to performing no motion correction. Errors in segmentation-based measures such as layer positions, retinal and nerve fiber thickness, as well as the blood vessel pattern were evaluated. The quantitative results consistently show that reproducibility is improved considerably by using the advanced algorithm, which also significantly outperforms the basic algorithm. The mean of the mean absolute retinal thickness difference over all data was 9.9 um without motion correction, 7.1 um using the basic algorithm and 5.0 um using the advanced algorithm. Similarly, the blood vessel likelihood map error is reduced to 69% of the uncorrected error for the basic and to 47% of the uncorrected error for the advanced algorithm. These results demonstrate that our advanced motion correction algorithm has the potential to improve the reliability of quantitative measurements derived from 3D-OCT data substantially.

©2014 Optical Society of America

OCIS codes: (170.4500) Optical coherence tomography; (170.4470) Ophthalmology; (100.2980) Image enhancement; (100.5010) Pattern recognition.
References and links

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(5045), 1178–1181 (1991).
2. J. S. Schuman, Optical Coherence Tomography of Ocular Diseases, 3rd ed. (SLACK Inc., 2013).
3. T. Klein, W. Wieser, R. Andre, T. Pfeiffer, C. M. Eigenwillig, and R. Huber, “Multi-MHz FDML OCT: snapshot retinal imaging at 6.7 million axial-scans per second,” Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XVI 8213(2012).
4. B. Potsaid, V. Jayaraman, J. G. Fujimoto, J. Jiang, P. J. S. Heim, and A. E. Cable, “MEMS tunable VCSEL light source for ultrahigh speed 60kHz-1MHz axial scan rate and long range centimeter class OCT imaging,” Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XVI 8213(2012).
5. B. Potsaid, I. Gorczynska, V. J. Srinivasan, Y. Chen, J. Jiang, A. Cable, and J. G. Fujimoto, “Ultrahigh speed spectral / Fourier domain OCT ophthalmic imaging at 70,000 to 312,500 axial scans per second,” Opt. Express 16(19), 15149–15169 (2008).
6. 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).
7. B. Braaf, K. V. Viinola, C. K. Sheehy, Q. Yang, K. A. Vermeer, P. Tiruveedhula, D. W. Arathorn, A. Roorda, and J. F. de Boer, “Real-time eye motion correction in phase-resolved OCT angiography with tracking SLO,” Biomed. Opt. Express 4(1), 51–65 (2013).
8. K. V. Viinola, 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).
9. S. Ricco, M. Chen, H. Ishikawa, G. Wollstein, and J. Schuman, “Correcting motion artifacts in retinal spectral domain optical coherence tomography via image registration,” Medical Image Computing and Computer-Assisted Intervention - Miccai 2009, Pt I, Proceedings 5761, 100–107 (2009).
10. A. G. Capps, R. J. Zawadzki, Q. Yang, D. W. Arathorn, C. R. Vogel, B. Hamann, and J. S. Werner, “Correction of eye-motion artifacts in AO-OCT data sets,” Proc. SPIE 7885, 78850D (2011).
11. 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).
12. H. C. Hendargo, R. Estrada, S. J. Choi, 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).
13. R. J. Zawadzki, A. R. Fuller, S. S. Choi, D. F. Wiley, B. Hamann, and J. S. Werner, “Correction of motion artifacts and scanning beam distortions in 3D ophthalmic optical coherence tomography imaging,” Proc. SPIE 6426, 642607 (2007).
14. N. Sebe, M. S. Lew, and D. P. Huijmsman, “Toward improved ranking metrics,” IEEE Trans. Pattern Anal. Mach. Intell. 22(10), 1132–1143 (2000).
15. M. Bashkansky and J. Reinitjes, “Statistics and reduction of speckle in optical coherence tomography,” Opt. Lett. 25(8), 545–547 (2000).
16. P. J. Huber, “Robust estimation of a location parameter,” Ann. Math. Stat. 35(1), 73–101 (1964).
17. K. Fountoulakis and J. Gondzio, “A second-order method for strongly convex L1-regularization problems,” arXiv preprint arXiv:1306.5386 (2013).
18. Z. Hou, “A review on MR image intensity inhomogeneity correction,” Int. J. Biomed. Imaging 2006, 49515S (2006).
19. S. J. Choi, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, “Automatic segmentation of seven retinal layers in SD-OCT images congruent with expert manual segmentation,” Opt. Express 18(18), 19413–19428 (2010).
20. H. Ishikawa, M. L. Gabriele, G. Wollstein, R. D. Ferguson, D. X. Hammer, L. A. Paunescu, S. A. Beaton, and J. S. Schuman, “Retinal nerve fiber layer assessment using optical coherence tomography with active optics nerve head tracking,” Invest. Ophthalmol. Vis. Sci. 47(3), 964–967 (2006).
21. M. L. Gabriele, H. Ishikawa, G. Wollstein, R. A. Bilonick, L. Kagemann, M. Wojtkowski, V. J. Srinivasan, J. G. Fujimoto, J. S. Duker, and J. S. Schuman, “Peripapillary nerve fiber layer thickness profile determined with high speed, ultrahigh resolution optical coherence tomography high-density scanning,” Invest. Ophthalmol. Vis. Sci. 48(7), 3154–3160 (2007).
22. H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, and J. S. Schuman, “Macular segmentation with optical coherence tomography,” Invest. Ophthalmol. Vis. Sci. 46(6), 2012–2017 (2005).
23. A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” Lect. Notes Comput. Sci. 1496, 130–137 (1998).

1. Introduction
Optical Coherence Tomography (OCT) [1] is a standard imaging modality in ophthalmic clinical practice and enables the detection and monitoring of retinal disease [2]. 2D and 3D OCT images are generally not acquired instantaneously, but over the course of multiple
seconds by scanning the OCT beam over the tissue. Motion artifacts occur if there is relative motion between the imaged object and the OCT system during this time. These artifacts distort the acquired OCT volume and limit the reliability of quantitative measurements that are extracted from OCT data.

There are multiple ways to deal with motion artifacts in 3D-OCT data. These approaches range from increasing the system acquisition speed [3–5], active tracking [6–8], to post processing based approaches [9–13]. We previously presented a method to correct motion in retinal 3D-OCT volumes using multiple volumes with orthogonal fast scan axis as input [11]. The different temporal sampling patterns cause motion artifacts that are complementary in the two (or more) input volumes. In each input volume, the spatial dimension which is in the fast scanning direction remains relatively intact due to the rapid time in which it is acquired. Motion is corrected on a per A-scan basis using dense displacement fields that have 3 degrees of freedom for each A-scan in the input volumes. A global objective function is optimized with respect to the parameters of the displacement fields. This function models two goals. First, after motion correction the two volumes should be similar to each other. Second, it is assumed that within very short amount of time the change in displacement that is equivalent to motion is expected to be low. The objective function is optimized using non-linear, multi-resolution optimization techniques. After the optimization, the input volumes are transformed to the motion corrected common space and can be combined to create a single motion corrected merged volume with higher image quality.

As was shown in our previous publication [11], this approach works very well for data sets of young and healthy subjects that were acquired by experienced OCT operators. In clinical practice, however, the conditions that affect image quality of the input volumes can be much more challenging. Subjects can be geriatric and have various ocular pathologies. One major factor is aberrations in the eye, which can lead to degraded transverse resolution in the OCT volume and lower signal-to-noise ratio (SNR). In addition, optical opacities such as cataracts and floaters can cause loss of retinal signal altogether. These effects can be dependent on alignment of the subject relative to the OCT device, so different alignment might cause opacities to shadow different parts of the retina in subsequently acquired volumes. An additional alignment related effect is tilting of the retina in axial direction. This effect is pronounced because OCT images are displayed with an expanded aspect ratio in the axial direction. The tilting is dependent on alignment of the OCT beam with respect to the pupil and can therefore change between subsequently acquired volumes and to a lesser extent within a single volume. Finally, yet importantly, volumes that cover a larger field of view can be especially affected by vignetting that is again dependent on alignment of the subject’s iris with respect to the OCT beam. These effects can cause systematic inconsistencies between multiple volumes of the same area, although they were acquired consecutively in time. This means that corresponding anatomical locations in two input volumes might have very different signal levels and A-scan signatures. Figure 1 shows views of an OCT volume of a young healthy subject and a volume of an elderly patient with ocular pathology. Due to lower signal, the fundus and the cross sectional view of the data set from the diseased eye is noisier. Dark spots are visible that likely originate from lens opacities or floaters in the eye. In addition, the data set of the elderly subject shows saccadic eye motion artifacts, while the higher quality volume from the younger subject does not. This shows the considerable variation in volume quality that is encountered in clinical practice.
In this paper, we demonstrate advances in the motion correction algorithm to address these issues to increase overall robustness and make the correction of the most challenging data sets possible. Starting with an overview of the basic correction algorithm, we then introduce an advanced algorithm with augmentations and modifications aimed at this goal. In addition, we perform a multifaceted quantitative evaluation to assess the impact of the advanced motion correction algorithm on the reproducibility of 3D-OCT data. We also evaluate how the advanced correction techniques described in this paper compare relative to no correction and to using the basic correction algorithm.

2. Methods

2.1 Basic algorithm

In order to put the proposed modifications and extensions into context, we first summarize the original motion correction algorithm [11] which we call the basic algorithm. It is assumed that two input volumes $I_x(x,y,z)$ and $I_y(x,y,z)$ are to be motion corrected and registered where $I_x$ denotes the volume with XFAST scan pattern and $I_y$ the YFAST type volume. Both volumes cover the same area and have the same $x$, $y$ and $z$ dimensions, denoted $w$, $h$ and $d$, respectively. It is also assumed that A-Scans are acquired instantaneously and that therefore there is no deformation within each A-Scan. In addition, the intensities of the volumes can be interpolated and are defined between voxel grid positions. To correct motion, two 2D displacement fields \( \delta_{x,i,j} \) and \( \delta_{y,i,j} \) are estimated which associate a 3D displacement vector with each A-scan grid position \( (x_i,y_j) \) of each input volume, where \( i \in \{1,2,\ldots,w\} \) and \( j \in \{1,2,\ldots,h\} \) denote the grid indices in each dimension. $\delta_{x,i,j}$ and $\delta_{y,i,j}$ can depend on different underlying parameters. The displacement fields are estimated by minimizing a non-linear objective function with respect to the parameters of the displacement fields. The objective function consists of two parts, a similarity term and a regularizer. The similarity term is based on computing a per voxel real number called residual as...
\[ R_{i,j,k}(\mathbf{d}_x, \mathbf{d}_y) = I_x(x_i + \delta x_x(x_i, y_j), y_j + \delta y_x(x_i, y_j), z_k + \delta z_x(x_i, y_j)) - \\
I_y(x_i + \delta x_y(x_i, y_j), y_j + \delta y_y(x_i, y_j), z_k + \delta z_y(x_i, y_j)) \]  

(1)

Where \( i \) and \( j \) are defined as before and \( k \in \{1, 2, \ldots, d\} \) marks an index into the volume grid in the axial dimension. Note that if the displacement fields are zero everywhere, the original volumes are interpolated and compared within the residual computation. The displacements serve to deform the input volumes to compensate the distortion caused by motion. The similarity measure is then computed by evaluating the residual at every grid point and applying a square loss function \( L^2_2(\mathbf{x}) = |\mathbf{x}|^2 \) where \( \mathbf{x} \in \mathbb{R}^N \) and summing such that

\[ S_{\text{base}}(\mathbf{d}_x, \mathbf{d}_y) = \sum_{i=1}^{w} \sum_{j=1}^{h} \sum_{k=1}^{d} L^2_2(R_{i,j,k}). \]  

(2)

The following regularization term is used to penalize the change in displacement between consecutive A-scans in time and can be expressed as

\[ \text{Reg}_{\text{base}}(\mathbf{d}_x, \mathbf{d}_y) = \sum_{i=1}^{w} \sum_{j=1}^{h} \sum_{k=1}^{d} \left( \frac{d}{d t} \frac{d}{d t} \right) \]  

(3)

where the displacements are treated as functions in time according to when the associated A-scan was acquired in the respective input volume. The two terms are weighted and combined using a factor \( \alpha \) such that

\[ F(\mathbf{d}_x, \mathbf{d}_y) = S_{\text{base}}(\mathbf{d}_x, \mathbf{d}_y) + \alpha \text{Reg}_{\text{base}}(\mathbf{d}_x, \mathbf{d}_y) \]  

(4)

which is minimized with respect to the parameters of the displacement fields \( \mathbf{d}_x \) and \( \mathbf{d}_y \) using a multi resolution approach [11].

2.2 Robust similarity measure

The use of the square loss function on the residual corresponds to the sum of squared differences (SSD) similarity measure. SSD is an optimal similarity measure if the signal has additive zero-mean Gaussian noise [14]. Using SSD implies that intensities of the OCT signal from a certain anatomical location imaged at multiple time points in multiple volumes will satisfy this assumption. In OCT imaging this assumption can be violated. For example, speckle noise is not Gaussian [15]. Furthermore, because of vignetting, floaters and lens opacities, an A-scan of a single location can have signal from the retina in one volume, or it can be completely obscured and only have background in another volume. Unfortunately, SSD is sensitive to such violations. Because of the quadratic loss, the error contribution from outliers dominates the objective function value. This can cause the optimizer to find parameters that alleviate primarily outlier error contributions, at the expense of overall performance. Areas that are not anatomically corresponding can therefore mistakenly be registered to each other.

In order to make the similarity term more robust to outliers, the quadratic loss function \( L^2_2(\mathbf{x}) \) in Eq. (2) is replaced with a slightly modified Pseudo-Huber loss function [16,17] defined as

\[ \text{Huber}(\mathbf{x}, \varepsilon_n) = \sum_{n=1}^{N} \varepsilon_n (\sqrt{1 + (x_n / \varepsilon_n)^2} - 1) \]  

(5)

where \( x_n \) is the n-th component of the vector \( \mathbf{x} \) and \( \varepsilon_n \) is set to an empirically chosen small number 0.001. For absolute values of \( x_n \) significantly larger than \( \varepsilon_n \) this function approximates
the $L_1$ norm. Therefore, outliers do not cause a disproportionately large error contribution. This enables the optimizer to more easily “ignore” outliers, improving the handling of inconsistent intensities. As opposed to the $L_1$ norm however, the chosen function is at least twice continuously differentiable, which is a prerequisite for the numerical optimization methods that are employed.

2.3 Regularization using Pseudo $L_{0.5}$ norm

As opposed to penalizing the gradient of the displacement fields with respect to time with the square loss function as in the basic motion correction algorithm, we use an approximation of the $L_{0.5}$ norm. The $L_{0.5}$ norm is defined as

$$L_{0.5}(x) = \sum_{n=1}^{N} \sqrt{|x_n|}.$$  \hspace{1cm} (6)

Application of the squared loss disproportionally penalizes individual high gradients. This favors the modeling of smoothly changing displacements with respect to time. However, there are also fast movements, especially in the transverse direction (saccades) where the motion profile is actually a sharp transition. Using the $L_{0.5}$ norm within the regularizer would not cause the disproportionate penalization of individual high gradients. In contrast, high gradient values would actually be associated with a disproportionally weaker penalty. This would favor sparse displacement fields, i.e. where there are a few high gradients with respect to time, while most of the gradients are zero. More specifically, it enables the easier modelling of saccades within the optimization, as they cause individual high gradients of the displacement fields.

In practice however, the $L_{0.5}$ norm does not have a continuous gradient at zero, nor is it convex. This would cause major problems when optimizing the objective function using standard gradient based iterative methods. Therefore, a $C^*$ continuous approximation is used, which is also convex within the value range for displacement gradient vector components that are typically occurring ($< 0.05$). The function is constructed as

$$L_{0.5,\epsilon}(x) = \sum_{n=1}^{N} \left( \sqrt{x_n^2 + \epsilon^2} - \sqrt{\epsilon^2} \right)$$  \hspace{1cm} (7)

where $\epsilon_{0.5}$ is a small value empirically chosen to be 0.1.

2.4 Two stage optimization

When using nonlinear optimization methods, in general one is only guaranteed to find a local minimum of the objective function. The basic motion correction algorithm uses multi-resolution optimization to alleviate this problem and reach a more global minimum [11].

![Fig. 2. Two stage optimization flowchart.](image)

We propose to extend this optimization approach to a two stage multi-resolution optimization. Figure 2 shows a schematic of the approach. In the first stage, a rough alignment of the input volumes limited to only the axial direction is performed using multi-resolution optimization. It is assumed that B-scans can only move in axial direction and that each B-scan is rigid. This effectively reduces the number of parameters that the residual depends on such that
\[ R^\text{axial}_{i,j,k}(d_x, d_y) = I_x(x_i, y_j, z_k + \delta z_x(y_j)) - I_y(x_i, y_j, z_k + \delta z_y(x_i)) \] (8)

where \( \delta z_x(y_j) \) and \( \delta z_y(x_i) \) are the per B-scan axial displacements to be determined in the optimization. Correspondingly, the displacement fields \( d_x, d_y \) are parameterized by these parameters. In this case, the regularization reduces to penalizing the derivative in time of only these variables.

Once the first stage optimization is finished, roughly registered volumes are constructed and used as the input for the second stage, which performs a full optimization with per A-scan 3D displacement parameters, again using multi-resolution.

2.5 Tilt correction

The basic algorithm does not account for different axial tilts between the input volumes. If there is a significant difference in tilt arising from differences in the OCT beam alignment, the tilt must be modeled as axial motion in order to register the volumes. This conflicts with the regularization that is employed because the axial displacement caused by tilt depends on the transverse coordinates and not on time. Figure 3 shows an example of how changes in alignment can produce an inconsistent axial tilt between corresponding locations in two consecutively acquired volumes.

\[ R^\text{axial+tilt}_{i,j,k}(d_x, d_y) = I_x(x_i, y_j, z_k + \delta z_x(y_j) + x_i m_x(y_j)) - I_y(x_i, y_j, z_k + \delta z_y(x_i) + y_j m_y(x_i)) \] (9)

where \( m_x(y_j) \) and \( m_y(x_i) \) are parameters that model per B-scan axial slopes that represent tilting. Based on the idea that tilt is unlikely to change significantly from one B-scan to the next, the change of the respective slopes over time is penalized using an additional regularization term. This enables the algorithm to compensate for a difference in tilt between the input volumes in stage one. However, this formulation does not consider the specific tilt that the output from stage one should have. Ideally, tilt should be removed altogether such that the retina appears aligned horizontally in the B-scans. An additional advantage of reduced tilt is that it produces less coupling of the axial and transverse degrees of freedom in the objective function, which simplifies the optimization problem. In order to reach this goal, an additional term is introduced which is based on the distribution of intensities in the axial direction after application of the displacement fields.

\[ M(k,d_x,d_y) = \sum_{i=1}^k \sum_{j=1}^l \left[ I_x(x_i, y_j, z_k + \delta z_x(y_j) + x_i m_x(y_j)) + I_y(x_i, y_j, z_k + \delta z_y(x_i) + y_j m_y(x_i)) \right] \] (10)
If the variance of $M(k, d_x, d_y)$ treated as a histogram in $k$ is low, this indicates two conditions. First, variance decreases when the volumes are in good alignment with each other. Second, variance decreases if the high intensities are concentrated in a small axial band of the volume, indicating no tilt of the retina. Therefore, the variance of the (normalized) distribution of $M(k, d_x, d_y)$ over $k$ is penalized during stage one with a scaling factor of 10 analogous to $\alpha$. Figure 4 illustrates the concept.

![Composite images before and after stage one and their corresponding axial intensity distributions. Plotted red, green, and blue are the intensity distributions of the first and second volume and the combined distribution, respectively. Using $\text{Var}(M(k, d_x, d_y))$ as a penalty term in the objective function the variance is reduced from $\sim 47400$ to $\sim 46600$ pixels$^2$, thereby aligning the volumes and removing tilt.](image)

**2.6 Mean displacement regularization**

The motion correction approach does not use a fixed reference volume, both volumes are transformed. Therefore, the objective function has degrees of freedom in the mean global displacement. Because the regularization penalizes the derivative and not the actual value of the displacement fields, it does not penalize an arbitrary 3D displacement vector that is added to all displacements in each displacement field. Therefore, the optimization can generate displacement fields that have an arbitrarily large common offset. Especially for low SNR input volumes, this unwanted effect can cause the residual function to sample the input volumes in regions where there is no signal from the sample.

The free displacement is called mean displacement and defined as

$$\text{MD}(d_x, d_y) = \sum_i \sum_j (d_x(i,j) + d_y(i,j)).$$ (11)

An additional term is added to the objective function that penalizes the square magnitude of $\text{MD}(d_x, d_y)$ scaled with a high factor of 100. This causes the mean displacement after optimization to be close to the zero vector and prevents drifting outside of the defined signal volume.
2.8 Illumination correction

Alignment dependent differences in signal intensities are caused by floaters, vignetting or lens opacities. In order to alleviate the effect of these sources of artifacts, illumination correction is performed as a preprocessing step. For each input volume $I$, illumination effects are estimated from the fundus projection

$$F_{i,j} = \sum_{k=1}^{d} I(x_i, y_j, z_k).$$

Based on the assumption that illumination effects have low spatial frequency, each fundus projection is low pass filtered using a 2D Gaussian filter with a 10 pixel standard deviation [18]. This image is called $F_{i,j}$. Next, the 75th percentile intensity value of the filtered input fundus image $\tau$ is calculated and used as a reference target for the intensity correction. A correction map $C_{i,j} = \tau - F_{i,j}$ is calculated. Assuming a multiplicative illumination model and due to the fact that images are in log scale, a correction value must be added/subtracted to the pixels of each A-scan such that the fundus projection pixel value is increased/decreased about $C_{i,j}$. However, one cannot simply correct all the pixels of an A-scan, as most of the axial pixels will contain background, which is not influenced by illumination effects. Therefore, only pixels which have an initial intensity value that is significantly higher than the background level are modified. This level was empirically chosen. Figure 5 shows an example of the illumination correction method. It can be seen that the correction compensates illumination effects on the fundus while maintaining visibility of features such as the vessels. The illumination correction method does not necessarily produce physically accurate results and creates some artifacts itself, but helps to make the appearance of multiple volumes that are affected by different illumination conditions more consistent. Note that as with all pre-processing steps, the resulting data is only used for the optimization itself. Registered and merged volumes are generated from the original input volumes.

![Fig. 5. Illumination correction example.](image)

2.9 Extrapolation changes

If the retinas in two input volumes are located at significantly different axial positions, the optimization must model large axial displacements in order to register the volumes. This
introduces the need to assign intensity values to axial coordinates outside of the defined range during the computation of the residuals. The previous method of extrapolation in this case was to repeat the last axial pixel [11]. However, due to noise, this means that for large displacements a random value will be replicated often in the interpolated A-scan that is used as input for the residual computation. The statistics of the intensity values which are extrapolated from outside of the axial range will therefore not correspond to the intensity statistics of actual background pixels. This can cause problems because a large residual might be generated when the residual between corresponding background regions would be much smaller, which potentially biases the optimization in the wrong direction.

To solve this problem, the extrapolation behavior in axial direction was changed to wrap around once, instead of repeating the last pixel. This means that when interpolating data at and beyond the top axial end of the volume grid, missing data is replaced with data from the low axial end of the same A-scan. And vice versa for the bottom axial end. In the transverse direction, the last pixel is still repeated. The algorithm modification for the axial dimension causes more realistic (noise) pixels to be introduced at the top and bottom ends of the axial range, instead of a single repeated pixel. Residuals in these regions will more closely correspond to normal background-to-background residuals. However, the OCT volume must not have a significant sensitivity rolloff for this method to work. In this context, too much rolloff means that placing background from the bottom of the range at the top would lead to a visible edge. The imaging system used in these studies did not have significant sensitivity rolloff within the axial range of interest according to this.

3. Evaluation and discussion

3.1 Introduction

OCT as an imaging modality and its clinical utility depends on its ability to represent the structure of the imaged object accurately and with good signal quality. Given this, measurements can be performed with high reliability and high spatial precision. This enables the detection of small and focal changes in anatomy, which is important for example for tracking of disease progression. Motion artifacts and low signal quality on the other hand limit this ability. The performance of motion correction algorithms can therefore be evaluated by how close the structure as depicted in the volume is to the true structure.

Unfortunately, ground truth structural information is hard to get. Other modalities such as Fundus imaging, MRI and Ultrasound are of limited use due to several factors. These factors include no 3D information, low resolution, different distortion and other kinds of motion artifacts. When designing the study to evaluate motion correction it was therefore decided against using a secondary modality for evaluation.

Instead, the similarity between multiple independent OCT volumes of the same object is used to evaluate how accurately structure is depicted in the volumes. We compare advanced motion correction using the methods presented in this paper with uncorrected data and with a basic correction algorithm. As a complementary measure, when using motion correction it is also evaluated how well the XFAST and YFAST volume are registered with each other.

The concept of the evaluation is that if multiple volumes of an area of an eye are acquired in short succession, the actual imaged object should remain the same. The volume data itself however will differ between the volumes because of noise, illumination, motion artifacts and alignment related distortions. Alignment related effects are modeled as a 3D translation of the entire volume, a possible rotation around the optical axis and tilts in the other two directions. If a set of volumes from the same location has no motion artifacts, or motion artifacts have been removed by motion correction, it should be possible to perfectly align the set of volumes with each other using quasi-rigid registrations that model the above alignment related effects. Therefore, how similar the volumes are after said quasi-rigid registration is a measure for how accurate and consistently structure is depicted in the volumes.

1 August 2014 | Vol. 5, No. 8 | DOI:10.1364/BOE.5.002591 | BIOMEDICAL OPTICS EXPRESS  2600
In order to evaluate how similar a pair of volumes is after quasi-rigid registration, the abstract similarity measure of mutual information is used. In addition, anatomical structures such as retinal layer boundaries and thicknesses and blood vessels of each volume are segmented, mapped to the common space and then compared for similarity. Smaller differences between the measurements produced by the segmentations indicate that anatomical locations are better mapped to each other, again indicating better similarity between the underlying volumes. Given a large data set with multiple, successively acquired volumes from many eyes, this approach is used to assess the performance of the different methods.

After explaining the evaluation approach in detail, key steps will be explored visually using an example after which the quantitative results will be presented, followed by a discussion.

3.2 Data acquisition

For the quantitative evaluation, \(6 \times 6\) mm, \(200 \times 200\) A-scan, 3D-OCT volumes of both the optic nerve head (ONH) and the macula regions were acquired using software modified RTVue (Optovue, Fremont, CA) devices at New England Eye Center and the University of Pittsburgh Medical Center. The axial resolution was \(~5\) um, with a pixel spacing of \(3.1\) um/pixel in tissue in axial direction and \(30\) um/pixel in the transverse direction. The study protocol was approved by the Investigational Review Boards of the New England Medical Center, University of Pittsburgh Medical Center and the Massachusetts Institute of Technology. Written informed consent was obtained from all subjects before imaging. For each subject, one eye was chosen at random for imaging. Each subject was imaged three times at two scan regions over the macula and ONH. Each time a set of two orthogonally scanned volumes was acquired. Between repeated orthogonal volume acquisitions, the device was reset and the subject re-aligned to the device. This yields three independent pairs of orthogonal raster scans per eye and location as input for correction. With 73 subjects being imaged, 876 input volumes were acquired in total.

A subject qualified as a normal subject if they had a normal Humphrey (Zeiss, Dublin, CA) 24-2 visual field test, intraocular pressure (IOP) at or below 21 mmHg, no history of diabetes, no family history of glaucoma and a normal ocular examination. Glaucoma suspect eyes were defined as those with IOP at between 22 and 34 mmHg, asymmetrical ONH cupping or an abnormal appearing OHN, all in the presence of normal visual field test results. The contralateral “healthy” eye of unilateral glaucomatous eye was defined as glaucoma suspect. This subgroup includes eyes that may manifest ocular hypertension, increased cupping or asymmetrical cupping. The third group of eyes, namely glaucomatous eyes was defined as those with at least one of the following features: Glaucomatous visual field defect, ONH cupping or a nerve fiber layer (NFL) defect on biomicroscopy. Table 1 shows statistics about the study population. Three subjects were excluded from the evaluation due to poor image quality, resulting in 70 subjects taking part in the evaluation.

| Table 1. Reproducibility data set statistics |
|--------------------------------------------|
| Count | Age                     |
|-------|-------------------------|
| Total | 73                      |
| Normal subjects | 38 | 52.0±/-19.0 (21-89) |
| Glaucoma subjects | 26 | 39.9±/-16.4 (21-74) |
| Glaucoma suspects | 6  | 63.0±/-10.7 (33-83) |
| Excluded | 3 | 63.0±/-10.6 (33-83) |
| excluded | 3 | 61.7±/-9.0 (51-73) |

3.3 Evaluation approach

For examining performance in the case without any motion correction and merging, the XFASST type volume of the orthogonal pairs. For evaluating motion correction, the basic
algorithm and the advanced algorithm were applied to each pair of orthogonal volumes. The advanced algorithm uses all the modification and enhancements that are described in this paper. The subsequent extraction of measurements and evaluation of similarity was performed on the merged, motion corrected output volumes. The regularization weighting factor $\alpha$ was stepped from 0.001 to 10 in multiples of 10. For 70 subjects, 6 volume pairs per subject, 5 $\alpha$ settings and two motion correction methods were evaluated. This corresponds to $4200 = 70 \times 6 \times 5 \times 2$ merged, motion corrected volumes, which were generated as input for the subsequent analysis. Experiments were performed on a Core i7-2600k CPU with an NVIDIA GeForce GTX 580 GPU running C++ and CUDA code, respectively. The typical run-time per pair was about 20 seconds for the basic and about 50 for the advanced algorithm. However, the advanced algorithm was relatively unoptimized, therefore the relative scaling might be misleading. In order to assess the ability of the algorithms to produce motion free, undistorted volumes, the following method was employed. Regardless of the method (uncorrected, basic algorithm, advanced algorithm), three independent volumes of the same location are available for comparison. All possible pairings of the three volumes were quasi-rigidly registered with each other. This registration models translation in 3D, rotation around the axial direction and tilting as an axial shift linearly depending on the transverse coordinates. Illumination correction was performed prior to the quasi-rigid registration. In addition, the modified pseudo Huber loss function (Eq. (5) was used as an intensity similarity measure. Similar to our original paper [11], the average mutual information between the volumes in this registered state over all possible pairs was calculated. The higher this value, the more similar the volumes are, indicating good motion correction performance, i.e. undistorted and consistent result volumes.

As a complementary measure, for each motion correction algorithm run, the difference in mutual information (MI) between the two registered, motion corrected volumes and the original input volumes was calculated. This measure assesses the ability of the correction method to register the XFAST and YFAST volume onto each other. Note that because there is no reference volume, this assessment is distinct from assessing the ability of the correction method to produce motion free, undistorted volumes.

3.3.1 Layer segmentation

Layer segmentation and its reproducibility is also employed in the quantitative evaluation in order to assess the similarity of independent volumes of the same location. To evaluate performance over the entire transverse area it is helpful to use two dimensional segmentation maps, which associate a measurement with every A-scan of the respective volume. The segmentation process itself should be as reliable as possible. Reliability in this context means...
that segmentation errors must be minimized. At the same time, the segmentation should be spatially accurate on a per A-scan basis.

Based on these considerations, layer segmentation was performed using an algorithm based on Chiu et al. [19] which segmented the positions of inner limiting membrane (ILM) inner segments (IS) and retinal pigment epithelium (RPE) and based on these positions, retinal thickness (defined to be RPE − ILM). In order to use additional knowledge and improve robustness, Chiu et al.’s algorithm was extended so that multiple layers could be found in a single Dijkstra shortest path search. Whereas in Chiu’s algorithm a node in the graph corresponds to a single pixel in the B-scan \((x, z)\) we extend this concept such that a node corresponds to a combination of two or more axial positions within an A-scan of the B-scan \((x, z_1, ..., z_n)\) with \(z_1 < ... < z_n\). This enables a node to model multiple layer boundary positions at the same time. A path through the graph then denotes the segmentation of multiple layer boundaries within the B-scan. Connections between the nodes of this “multi-graph” are based on the combination of the possible transitions between individual boundary positions from one A-scan to the next. Additionally, transitions that lead to boundary positions that are too close, or too far apart axially are discarded. The corresponding connection costs are the sum of the axial gradient based costs of the individual boundaries. Compared to the original approach, the graphs are much larger with this method because of the combinatorial explosion of the possible number of nodes. Therefore, special care had to be taken in order to optimize graph construction and shortest path search. The number of nodes per A-Scan is \(O(n_{axial}^{n_{layer}})\), where \(n_{axial}\) is the possible distinct axial positions on which a layer can be located and \(n_{layer}\) is the number of layer boundaries that are searched simultaneously. In order to keep the graph size within reasonable limits \(n_{axial}\) had to be restricted by limiting the axial search space or only considering every \(n^{th}\) axial position, or both.

Despite the added computation complexity, this approach has a principal advantage compared to segmenting multiple layers individually and in sequence. Because multiple boundaries are considered at the same time it is possible to incorporate information about minimal and maximal thickness and the fact that one boundary needs to be always below the other into the search. The algorithm still returns the globally shortest path but for more than one boundary and also with the possibility to model interactions between the boundaries directly.

For this study, segmentation was performed individually for each B-scan of each volume in the following sequence: First, candidate locations for the ILM and RPE were found simultaneously using the multi-graph method, considering every fifth axial pixel. Then, the RPE estimate was low pass filtered using a Gaussian filter and a reference layer was defined to be slightly above (17 pixels) the RPE. The volume was then shifted in the axial direction such that the reference layer would become a straight line. Subsequently, the inner segments (IS) and RPE layer positions were simultaneously searched for below the reference line. The ILM was searched for as an individual layer boundary above said reference line. By applying this segmentation algorithm to every B-scan in a volume, 2D en face layer boundary position and thickness maps for ILM, IS, RPE and retinal thickness were obtained. In order to enable maximum spatial resolution of the resulting maps, no further smoothing was performed. Figure 6 shows an example application of the segmentation step on a merged ONH volume.

To evaluate the algorithm a manual segmentation study was performed. Three human graders segmented 158 randomly chosen B-Scans from the available data. Comparison with the automatic algorithm showed the mean absolute difference between human observers to be only slightly lower than between human and automatic observers (ILM: 2.32 pixels vs. 2.96 pixels, RPE: 2.96 pixels vs. 3.48 pixels).

In addition, NFL thickness was automatically segmented per A-scan on all volumes with another method based on adaptive thresholding [20–22].
When comparing the resulting segmentation maps, the difference is sensitive to transverse distortion between the volumes for all maps. Furthermore, the individual layer maps (ILM, IS, RPE) are also sensitive to axial distortion, while retinal and NFL thickness are not.

3.3.2 Blood vessel mapping

Blood vessel (BV) likelihood maps were generated from the volume data in order to assess transverse distortion. These maps associate every A-scan with a value from zero to one that represents the continuous likelihood of there being a blood vessel at the corresponding location. In order to generate the maps, the mean intensity between the IS and RPE layer was plotted into a 2D map. This map was subsequently illumination corrected by applying a bias field that was obtained using a large standard deviation Gaussian filter (20 pixels) [18]. In addition, the map was scaled such that the median intensity was 0.5. Subsequently, a Hessian based multi-scale “vesselness” measure was applied [23], yielding a 2D vessel likelihood map. The maps were thresholded and scaled such that the maximum likelihood of 1.0 was reached when the vesselness measure was at or above 0.0001. Figure 7 shows the maps at different stages of this algorithm for three volumes of the same subject and scan region.
3.3.3 Difference map generation and error calculation

Fig. 8. Segmentation based quantitative evaluation workflow.

Given quasi-rigid registration and layer/blood vessel segmentation, the evaluation proceeds as follows. Figure 8 shows a flowchart of the process. For each pair of output volumes, the transform obtained by the quasi-rigid registration step was used to map the segmentation maps into a common coordinate system. Subsequently, for every pixel within an area of interest, excluding a border of 10 pixels, i.e. 5 percent of the transverse area on each side, several absolute difference maps were calculated. The border was introduced to account for a slight lack of overlap of acquired volumes due to changes in fixation and/or motion. Other than that, no areas were excluded, especially not the ONH. The absolute difference maps simulate a varying tolerance to an uncertainty in lateral position in the maps. This is achieved by taking the minimum absolute difference of a reference pixel in the first map \(a(x, y)\) and the values in a neighborhood \(n(x, y, tol)\) of \(\pm tol\) pixels in each direction around the corresponding position in the second map \(b(x, y)\) such that

\[
\text{absdiff}_{tol}(x, y) = \min_{(x_2, y_2) \in n(x, y, tol)} |a(x, y) - b(x_2, y_2)|. \tag{13}
\]

Values of the positional tolerance \(tol\) of 0, 1 and 2 pixels (corresponding to a tolerance of up to \(\pm 90\) microns) were evaluated, where \(tol = 0\) is the standard absolute difference operation. Figure 9 shows a schematic of the difference computation in relation to the spatial tolerance.

Fig. 9. Spatial tolerance schematic. The minimum absolute difference between a value in map 1 and a neighborhood in map 2 is calculated. The size of the neighborhood depends on the spatial tolerance.

Finally, the mean value of each difference map was calculated. Values closer to zero corresponded to more reproducible segmentation maps obtained from the volumes. Therefore,
the mean of the absolute difference map can be seen as a segmentation-based reproducibility error. For comparing errors where there was a pairing between data sets, i.e. when comparing the different methods for the same set of input volumes, a Wilcoxon signed rank test (significance level 0.01) was used to check whether the error distributions were significantly different. For independent sets of input volumes, the Mann-Whitney-U test was employed.

3.4. Visual Inspection

In order to visually examine the evaluation pipeline and the effect of advanced motion correction on volume data reliability, a random pair of XFAST volumes and their corresponding YFAST volumes from one subject was selected. Figure 10 shows fundus images from two pairs of orthogonally scanned volumes and the corresponding fundus images from the merged volumes after performing basic and advanced correction. All four input volumes show obvious motion artifacts and illumination inconsistencies. After correction using either technique using $\alpha = 0.1$ these artifacts are mostly removed. However, in the top periphery of the merged fundus views, slight artifacts due to a mis-registration are visible. Other than that, it is not immediately obvious whether advanced correction leads to more accurate results, pointing out the need for quantitative evaluation.

![Fig. 10. Comparison of fundus views from two pairs of volumes before and after basic and advanced motion correction.](image)

Figure 11 shows a comparison of central cross-sectional image data of the same set of volumes before and after quasi-rigid registration for the three methods. As shown in the composite images, compared to the uncorrected result, the rigidly registered motion corrected volumes correspond much better. This indicates that the motion correction leads to consistent volume data for independent sets of input volumes from the same scan region. Also advanced correction slightly outperforms basic correction. This can be observed when looking at the correspondence of the blood vessel shadows, for example.

After segmentation of the volumes and quasi-rigid registration, the transform obtained is used to map the segmentation maps into a common coordinate system. Subsequently, difference maps are calculated (see Eq. (13)). Figure 12 and Fig. 13 show the individual transformed segmentation maps for blood vessels and retinal thickness and the corresponding difference maps for the three methods. In both cases, the resulting difference maps have lower values for basic and advanced correction, again indicating a higher consistency in the volume data and the derived segmentation maps. Advanced correction leads to the lowest errors for both maps.
Fig. 11. Example comparison of quasi-rigid registration performance between uncorrected and motion corrected data.

Fig. 12. Example comparison between mapped blood vessel likelihood maps and the corresponding difference maps. The mean absolute difference is 0.19 for the uncorrected pair, 0.11 for basic correction and 0.06 for the pair resulting from using the advanced correction algorithm.
3.5. Quantitative results

Figure 14 shows a comparison of the two information theoretic measures between the three methods and for all the data. In subfigure A, it can be seen that the mean mutual information over all data steadily decreases as $\alpha$ increases. This is consistent with previous results (see Fig. 5A in [11]) and can be explained by the balance of similarity measure and regularization. As regularization strength is increased, volume similarity becomes relatively less important leading to lower similarity of the volumes after motion correction. In addition, the advanced algorithm leads to higher similarity than the basic algorithm, regardless of $\alpha$. This difference was found to be significant for all $\alpha$. Subfigure B shows the mean mutual
information after quasi-rigid registration of all possible corresponding pairs of output volumes for the three methods. Both motion correction algorithms lead to higher similarity after quasi-rigid registration compared to not applying correction. In addition, the advanced algorithm shows consistently higher similarity compared to the basic algorithm. Both of these differences were found to be significant regardless of $\alpha$. Looking at both of these complementary graphs shows that the advanced algorithm yields a better similarity of the volumes after motion correction and better similarity of independent volumes compared to both the basic algorithm and to no correction. In addition, the basic motion correction algorithm outperforms no correction.

Figure 15 compares the mean absolute error over all pairs of data sets between the three methods for different segmentation maps. Lower numbers indicate better reproducibility of the values contained in the respective maps. In addition, retinal thickness and blood vessel likelihood map errors are not sensitive to axial motion artifacts, while ILM and RPE position are. Applying no correction consistently leads to the largest error for all four types of maps. The advanced algorithm tends to produce the lowest errors, followed by the basic algorithm.

Looking at Fig. 14 and 15 in conjunction, especially the blood vessel likelihood errors and their dependence on $\alpha$, $\alpha = 0.1$ can be considered an optimal parameter setting because it produces the lowest overall errors for the advanced algorithm without a disadvantage in the basic algorithm. Therefore, the remaining evaluation will be performed with $\alpha$ at 0.1. For this regularization strength and most others (except 0.001), errors are significantly lower for retinal thickness, ILM, NFL thickness and blood vessel maps. The blood vessel map error is reduced to 69% of the uncorrected error for the basic and to 47% of the uncorrected error for the advanced algorithm, which was found to be statistically significant.

![Comparison of mean absolute segmentation map errors between all possible pairs of output volumes from one scan region for the three methods and different $\alpha$.](image)

Fig. 15. Comparison of the mean of mean absolute segmentation map errors between all possible pairs of output volumes from one scan region for the three methods and different $\alpha$. Error bars indicate ± one standard deviation around the mean. A: Retinal thickness error in pixels. B: ILM position error. C: NFL thickness error. D: Blood vessel map error.
The previous results considered the mean error over all available data; however, it is also interesting to look at potential differences in performance for subgroups of the data. Figure 16 shows a box plot comparison between the errors over all data sets versus all normal and versus glaucomatous subjects and glaucoma suspects combined. For all groups and the four map types, the advanced algorithm always has the lowest errors, followed by the basic algorithm and no correction. In addition, the normal subject group shows slightly lower errors than the combined glaucoma suspect and glaucoma groups.

Figure 17 shows the same type of box plots, but this time grouped according to anatomical location into all data sets, ONH only and macula only. Again, the advanced correction shows best performance. In addition, the macula subgroup shows lower errors. This can be explained by the fact that the area around the ONH has both more variability in retinal thickness and contains more blood vessels. Also, NFL thickness varies more around the ONH and the NFL
thickness segmentation algorithm is not very reliable within the optic disc. For the same amount of transverse distortion, this leads to higher errors for mismatched maps. Using advanced correction, the mean error consistently drops below two axial pixels (6.2 μm) for retinal thickness for all subgroups in the two figures.

Finally, Fig. 18 compares the errors over all data sets for different spatial uncertainty tolerance values $tol$ of 0, 1 and 2. As expected, a higher spatial tolerance consistently leads to lower errors, when other parameters are kept fixed. In addition, regardless of the tolerance value, the advanced algorithm leads to the lowest error, followed by the basic algorithm, with no correction being the worst.

3.6 Discussion

These results show that the advanced algorithm described in this paper yields significant improvements in obtaining reliable quantitative measurements from 3D-OCT volume data. Whereas the basic algorithm already shows significantly lower errors than performing no correction, the advanced algorithm yields even further improvement. It is important to keep in mind that the errors measured in this evaluation are associated with the combined reproducibility of the entire processing pipeline. This pipeline includes the OCT device and its axial and transverse resolution and sampling, its signal to noise ratio (SNR), the presence of motion artifacts in the data, the performance of the motion correction and merging algorithm, layer and blood vessel segmentation and quasi-rigid registration performance. These components also interact with each other. For example, good motion correction counteracts motion artifacts in the input data. In addition, good motion correction and registration followed by merging will improve SNR. High SNR is important for segmentation algorithm performance. Conversely, poor motion correction can introduce additional non-reproducible distortions in the volume. In addition, if the motion correction algorithm fails to register the volumes, the merged volume would have lower SNR and resolution since image information from different anatomical locations would be combined. Based on these interactions, it can be hypothesized that advanced motion correction plays an especially important role. It effectively corrects for motion artifacts and improves SNR compared to the input volume, leading to better performance of the subsequent steps and subsequently the highest reproducibility.
In addition, there is an inherent tradeoff between the precision of the segmentation and its reproducibility. By applying low pass filtering to the segmentation maps prior to difference map calculation, the reproducibility error would be lowered. However, the segmentation would not be as precise, i.e. losing its ability to capture focal changes. In this evaluation, the aim was to have the segmentation itself be as spatially precise as possible, in part also to be more sensitive to distortions caused by motion artifacts. The spatial uncertainty tolerance was introduced to be able to assess how reproducibility error would decrease when the requirements for a precise per A-scan segmentation are decreased. In this context, the results show that even for the largest spatial tolerances, when the minimum of a $5 \times 5$ grid of pixel differences is taken when computing difference maps, advanced motion correction still leads to the best results. One factor that could explain this result is the improved SNR of the merged data. Also, the motion induced distortions in uncorrected data which lead to reproducibility errors might in part be larger than what the spatial tolerance allows for.

4. Summary and conclusion

This paper presents the latest advances in 3D-OCT motion correction based on orthogonal scans. The variable data quality that occurs in clinical imaging settings motivates these improvements. More specifically, variations in tilt, illumination and signal quality pose a significant challenge.

In order to deal with the non-normally distributed noise in OCT and inconsistencies such as those caused by floaters, a pseudo Huber loss function based similarity measure for comparing the volumes within the objective function was introduced. Similarly, regularization based on a function approximating the $L_{0.5}$ norm was introduced in order to promote a more sparse solution for the displacement fields.

In order to improve robustness and address inconsistent tilt that might be present in the volumes, a two-stage optimization scheme was introduced. The first stage performs a coarse correction of axial motion and tilt. By incorporating the variance of the intensity distribution in the axial dimension into the objective function, the tilt can also be removed during this step. The result of stage one is then used as input for a full optimization with per A-scan 3D displacement fields in stage two. We also corrected for inconsistencies in illumination that can be caused by differences in alignment of the OCT device by performing an illumination correction step during preprocessing. Furthermore, an additional mean displacement term and changes in the extrapolation behavior in volume interpolation were introduced.

A multivariable evaluation with data sets from 38 normal, 26 glaucomatous and 6 glaucoma suspect subjects was performed. Three independent pairs of orthogonally scanned volume pairs were available for the macula and ONH region for every subject. The three different cases that were evaluated were no correction, use of a basic correction algorithm and use of the advanced algorithm as described in this paper. Registration performance itself was evaluated by examining the increase in mutual information from motion correction and registration. As a second component of the evaluation, reproducibility of quantitative measurements derived from the data was evaluated. ILM, RPE, retinal and NFL thickness and blood vessels of the output volumes of each method were segmented as quantitative measurements and acted as surrogates for actual clinical measurements. For each method, all possible pairs of result volumes from the same subject and scan region were quasi-rigidly registered. The resulting quasi-rigid transform was then used to map the segmentation maps into a common space and subsequently compute the absolute difference map between them. Finally, the mean of the difference map was taken as a measure of reproducibility error. In addition, the mutual information between volume pairs in the quasi-registered state was used as an abstract measure for similarity of the output volumes.

Regardless of which measures were used and which subgroups of the data was examined, the advanced algorithm consistently yielded the best results, followed by the basic algorithm and performing no correction, with the differences being very clear. For example, the mean of
the mean retinal thickness absolute difference over all data was 9.9 um without correction, 7.1 um using the basic algorithm and 5.0 um using the advanced algorithm, the mutual differences being statistically significant. The mean blood vessel map error over all data was reduced to 69% of the uncorrected error for the basic algorithm and to 47% of the uncorrected error for the advanced algorithm, with the mutual differences also being significant.

In conclusion, the algorithm advances introduced in this paper are shown to yield improvements in the reliability of quantitative measurements as shown by the evaluation of clinical OCT data. This improvement results from addressing issues of data inconsistency such as those caused by tilt, illumination etc.. Because clinical data often suffers from these effects, a significant improvement results from an algorithm that is adapted to address these. The evaluation was primarily focused on comparing 2D segmentation maps in order to assess reproducibility in the entire volume. In addition, the choice of using maps to evaluate performance was primarily motivated by relative ease of segmentation. However, these results should also pertain to standard clinical measurements such as circumpapillary NFL thickness profiles, which should also become more reproducible and therefore more sensitive for diagnosis. The use of 2D thickness maps is promising for future clinical diagnostics. Because of the importance of volumetric data, we feel that clinical practice as well as research can benefit significantly from advanced motion correction algorithms.

Acknowledgments and disclosures

Authors M. Kraus, J. Hornegger and J. Fujimoto receive royalties for intellectual property on motion correction that is owned by MIT and licensed to Optovue Inc. J. Fujimoto is a scientific advisor for Optovue and has stock options. Optovue provided a software modification which enabled orthogonal scan patterns to be used for this study, however Optovue personnel did not participate in the data analysis, interpretation or manuscript preparation. This research is sponsored in part by the National Institutes of Health R01-EY011289-28, R01-EY013178-14, R01-EY018184-05, R01-CA075289-16, R01-CA178636-01, R44-CA101067-05, Air Force Office of Scientific Research FA9550-12-1-0499 and FA9550-10-1-0551, the German Research Foundation DFG-HO-1791/11-1 and DFG Training Group 1773 “Heterogeneous Image Systems”. The authors also gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative. We thank Thomas Koehler, Andre Aichert, Chen D. Lu, Michael Giacomelli, Lennart Husvogt and Maria Polyanskaya for valuable discussions and assistance.