

# Compensating Motion Artifacts of 3D *in vivo* SD-OCT Scans

O. Müller<sup>1</sup>, S. Donner<sup>2,3</sup>, T. Klinder<sup>4</sup>, I. Bartsch<sup>3</sup>, A. Krüger<sup>2,3</sup>,  
A. Heisterkamp<sup>2,3</sup>, and B. Rosenhahn<sup>1,\*</sup>

<sup>1</sup> Institut für Informationsverarbeitung, Leibniz Universität Hannover, Germany

{omueLLer,rosenhahn}@tnt.uni-hannover.de

<sup>2</sup> Laser Zentrum Hannover e.V., Germany

<sup>3</sup> CrossBIT, Hannover Medical School, Germany

<sup>4</sup> Philips Research Hamburg, Germany

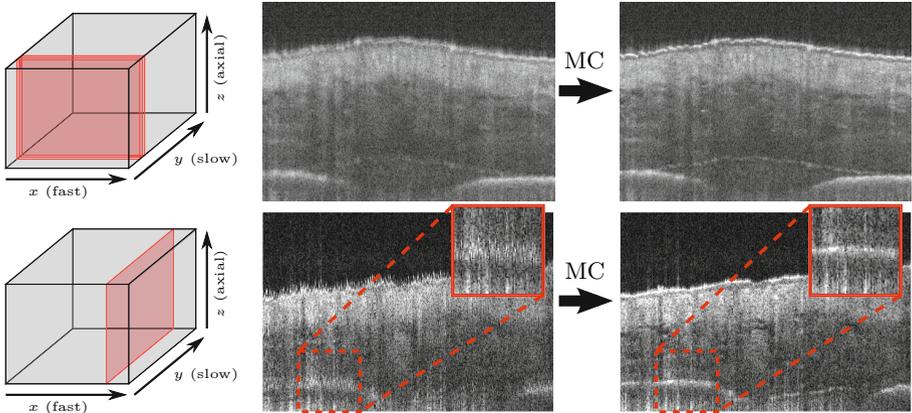
**Abstract.** We propose a probabilistic approach for compensating motion artifacts in 3D *in vivo* SD-OCT (spectral-domain optical coherence tomography) tomographs. Subject movement causing axial image shifting is a major problem for *in vivo* imaging. Our technique is applied to analyze the tissue at percutaneous implants recorded with SD-OCT in 3D. The key challenge is to distinguish between motion and the natural 3D spatial structure of the scanned subject. To achieve this, the motion estimation problem is formulated as a conditional random field (CRF). For efficient inference, the CRF is approximated by a Gaussian Markov random field. The method is verified on synthetic datasets and applied on noisy *in vivo* recordings showing significant reduction of motion artifacts while preserving the tissue geometry.

## 1 Introduction

Optical coherence tomography (OCT) is a rapidly evolving non-invasive imaging modality used for high-resolution imaging of biological tissue structures. OCT measures the backscattering profile (over time) of a light beam penetrating the sample in axial direction. Spectral-domain OCT (SD-OCT) acquires the backscattering profile in spectral domain rather than time domain, enabling shorter acquisition time. For 3D volume acquisition, single axial scan (A-scan) acquisition is combined with a lateral scanning mechanism. 2D scans (B-scans) are composed by a series of A-scans along the  $x$ -axis (fast scanning axis). 3D volume scans in turn consist of a series of B-scans along the  $y$ -axis (slow scanning axis). In our setting we are confronted with severe axial motion shift due to heart beat or breathing during *in vivo* SD-OCT (spectral-domain OCT) volume acquisition of mouse skin tissue around a percutaneous implant (see Fig. 1). Along the fast scanning axis, motion artifacts are illustrated by averaging three consecutive B-scans, resulting in noticeable image blur. In slow scanning direction, motion artifacts manifest itself in dithering in axial direction.

---

\* This work is supported by funding from the DFG (German Research Foundation) for the Cluster of Excellence REBIRTH.



**Fig. 1.** In vivo SD-OCT scan (cropped) and its motion compensated (MC) result. Top: Slice along the fast scanning axis. Three slices are averaged for visualization of the motion distortion. Bottom: One reslice along the slow scanning axis.

Yun et al. [1] have investigated motion artifacts of SD-OCT occurring during a single A-scan capturing such as signal fading, spatial distortion and blurring. These artifacts can be reduced by increasing the A-scan acquisition rate. However, image shifts in axial direction of several pixels occurring during acquisition of several thousands of A-scans (e.g. for volume acquisition) are still an issue. Later works [2,3] focus on compensating such image shifts in full volume scans using reference measures. While Ricco et al. [2] compensate transverse motion in retinal volume scans using scanning laser ophthalmoscopy (SLO) images as a reference measure, Lee et al. [3] correct motion shift in dynamic SD-OCT imaging, periodically capturing the same region over several seconds, using one of such captures as reference. Recent work in [4] correct motion artifacts by estimating a displacement vector for each A-scan using orthogonal OCT scan patterns. The method works without having a reference measure. Inference of the displacement field is done by minimizing an objective function using a gradient-descent method combined with a multi-resolution approach. Our work extends this approach by transferring the objective function to CRF notation and adding additional priors, allowing better tissue structure preservation and fast global optimization.

**Contributions:** We propose a probabilistic method for estimation and compensation of axial motion shift in *in vivo* SD-OCT without requiring a reference measure. The key challenge is to distinguish between motion shift and the natural spatial structure of the subject tissue. We tackle this problem by combining two different lateral scanning schemes for volume acquisition: The motion shift of multiple taken A-scans at the same lateral position (but at various time points) differ whereas the tissue structure remains unchanged. The motion compensation problem is formulated as an energy minimization problem using a conditional random field (CRF) notation, allowing both estimation of the motion field and the tissue structure. For inference, the CRF is simplified to a Gaussian Markov

random field (GMRF) by approximating crosscorrelation terms with a Gaussian pdf. Finally, our method is applied on in vivo SD-OCT scans of skin tissue with a percutaneous implant (see Fig. 1, dashed red rectangles indicate the subcutaneous implant base).

## 2 Motion Field Model

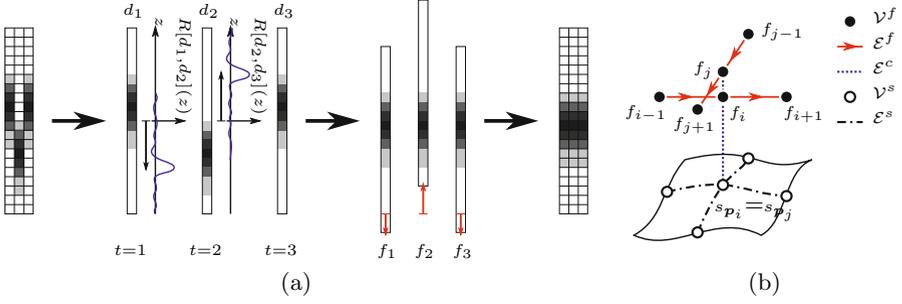
For estimation and compensation of in vivo subject movement, the following assumptions are made: A sequence of A-scans  $\{d_t\}$  is captured at discrete time points  $t$  and lateral position  $\mathbf{p}_t = (x_t, y_t)$ . For image acquisition, it is assumed that the scanned subject is somehow fixed (e.g. no freehand capturing involving transverse motion drift). Nevertheless, subject movements can not be suppressed completely, e.g., slight up-down movements caused by breathing or heart beating can still occur, but are limited in amplitude. Thus, for each A-scan  $d_t$ , we have a corresponding axial *motion shift*  $f_t$ . Since axial shift is not solely determined by  $f_t$  due to spatial tissue structure changes, the true axial shift is defined by  $f_t + s_{\mathbf{p}_t}$ , where  $s_{\mathbf{p}_t}$  is the tissue surface height at lateral position  $\mathbf{p}_t$ . In the following, we derive a CRF model  $E(f, s | d) = E(d | f, s) + E(f, s)$  given an observation model  $E(d | f, s)$  and regularizer  $E(f, s)$  with  $f = \{f_t\}$  as described below.  $E(f, s | d)$  is defined over a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  given the vertex set  $\mathcal{V}$  containing model instances and the edge set  $\mathcal{E}$  representing dependencies between instances.

**Observation Model:** The observation model is based on the assumption of structural similarity of (i) spatial neighbored A-scans  $d_i$  and  $d_j$  with  $(i, j) \in \mathcal{E}^R$ , where the structure of  $\mathcal{E}^R$  depends on the scanning schemes used (see Fig. 3(c) in Sect. 3) and (ii) A-scans taken at the same spatial position but at different time points  $d_i$  and  $d_j$  with  $(i, j) \in \mathcal{E}^c = \{(i, j) | \mathbf{p}_i = \mathbf{p}_j\}$ . As a similarity measure of adjacent A-scans, we use the crosscorrelation  $R[\tilde{d}_i, \tilde{d}_j](z) = \int \tilde{d}_i(\tau) \cdot \tilde{d}_j(z + \tau) d\tau$  of two adjacent volume gradient columns  $\tilde{d}_i, \tilde{d}_j$ , with  $z = f_i - f_j$  denoting the relative motion shift (see Fig. 2(a)). Actually, the axial shift does not only depend on the motion shift itself, but also on the spatial change of the tissue surface structure. Therefore, we introduce new model variables  $\{s_{\mathbf{p}_i}\}_{\mathbf{p}_i \in \mathcal{V}^s}$  denoting the subject surface change. Then, the relative axial shift is now determined by  $f_i + s_{\mathbf{p}_i}$ , rather than only by  $f_i$ , i.e.  $z$  becomes  $(f_i + s_{\mathbf{p}_i}) - (f_j + s_{\mathbf{p}_j})$ . If  $s$  is not known a priori, the problem is ill-posed because of ambiguities in the sum of relative motion and surface change. We solve this ambiguity by correlating A-scans taken at the same sample position at different time points. For such A-scans, it is  $z = f_i + s_{\mathbf{p}_i} - f_j - s_{\mathbf{p}_j} = f_i - f_j$ , because  $s_{\mathbf{p}_i} = s_{\mathbf{p}_j} \forall (i, j) \in \mathcal{E}^c$ . Thus, we have

$$E(d | f, s) = \gamma \sum_{(i,j) \in \mathcal{E}^R} R_{ij}(f_i + s_{\mathbf{p}_i} - f_j - s_{\mathbf{p}_j}) + \sum_{(i,j) \in \mathcal{E}^c} R_{ij}(f_i - f_j) \quad (1)$$

where  $R_{ij}(\cdot) := -\log R[\tilde{d}_i, \tilde{d}_j](\cdot)$  and  $\gamma$  is a weighting factor.

**Motion Field Prior:** For regularizing the motion estimation problem, additional assumptions are encoded in the prior energy term. Due to mass inertia of the subject, the motion field has to be smooth in time direction. In our model,



**Fig. 2.** Proposed model: (a) Motion correction workflow: Motion field  $\{f_i\}_i$  (red arrows) is estimated by maximizing the crosscorrelation  $R[d_i, d_j](z = f_i - f_j)$  (blue curves) of adjacent image rows  $d_i, d_j$ , (b) Graphical model structure (red arrows indicate temporal scanning direction).

we use first order smoothness. Additionally, we assume a Gaussian motion shift prior with zero mean, i.e.  $f_t \sim \mathcal{N}(0, \sigma_f^2)$ . The tissue surface  $s$  is regularized analogously. Thus, the prior is:

$$E(f, s) = \theta_1 \sum_{i \in \mathcal{V}^f} \frac{f_i^2}{2} + \theta_2 \sum_{(i,j) \in \mathcal{E}^f} (f_i - f_j)^2 + \theta_3 \sum_{i \in \mathcal{V}^s} \frac{s_{\mathbf{p}_i}^2}{2} + \theta_4 \sum_{(i,j) \in \mathcal{E}^s} (s_{\mathbf{p}_i} - s_{\mathbf{p}_j})^2 \quad (2)$$

with  $\mathcal{E}^f = \{(t, t+1) \mid t \in [0, 1, \dots, T]\}$  and  $\theta_1 = 1/\sigma_f^2$ ,  $\theta_2 = \lambda_f$ ,  $\theta_3 = 1/\sigma_s^2$ ,  $\theta_4 = \lambda_s$ ,  $\theta_5 = \gamma$  are the model parameters. The composed graph structure  $\mathcal{G} = (\mathcal{V}^f \cup \mathcal{V}^s, \mathcal{E}^f \cup \mathcal{E}^R \cup \mathcal{E}^s \cup \mathcal{E}^c)$  is depicted in Fig. 2(b).

## 2.1 Inference

To efficiently find a configuration  $\{f^*, s^*\}$  minimizing  $E(f, s \mid \theta)$ , we have decided to simplify  $E(f, s \mid \theta)$ . The only terms in  $E(f, s \mid \theta)$  which makes efficient inference difficult are the crosscorrelations  $R[d_i, d_j](\cdot)$ . Assuming that the A-scan intensities follow an edge-step model and is augmented with additive white Gaussian noise, the crosscorrelation of the first derivative of the A-scan intensities has a Gaussian shape with additive white Gaussian noise. For model parameter estimation, nonlinear least-squares Gaussian fitting is applied. Thus we obtain  $R[d_i, d_j](z) \approx \hat{R}[d_i, d_j](z) = \mathcal{N}(\mu_{ij}, \sigma_{ij}^2, z)$ , where  $\mu_{ij}$  and  $\sigma_{ij}^2$  are the mean and variance of the estimated Gaussian distribution  $\mathcal{N}(\mu, \sigma^2, \cdot)$ .

Using this approximation, the CRF energy function  $E(f, s \mid d)$  simplifies to a Gaussian Markov random field (GMRF), i.e.  $E(f, s \mid d)$  is a quadratic function in  $\{f, s\}$  and can be rewritten as  $E(x \mid \theta) = \frac{1}{2} x^T A_\theta x + x^T b_\theta + c_\theta$ , where  $x = \{f, s\}$  and  $A_\theta$  is sparse due to the Markov property. Its minimizer is  $x^* = -A_\theta^{-1} b_\theta$ . This can be efficiently solved by (sparse) Cholesky decomposition of  $A_\theta$  [5].

Estimation of the optimal parameter vector  $\theta^*$  is done by minimizing the mean-square-error (MSE) of  $f$  with  $\theta^* = \arg \min_\theta \|f_\theta^* - f_{\text{correct}}\|_2^2$ , where  $x_\theta^* =$

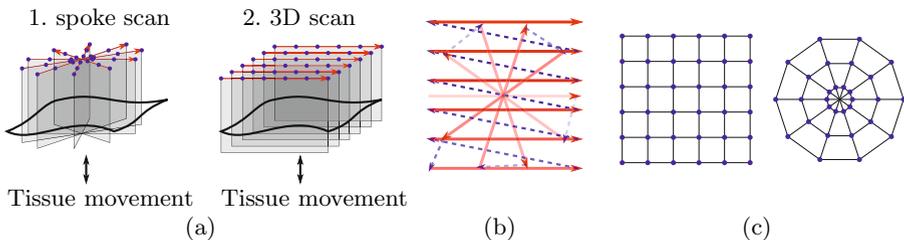
$\arg \min_x E(x | d, \theta)$  with  $x_\theta^* = \{f_\theta^*, s_\theta^*\}$  and  $f_{\text{correct}}$  is the ground truth motion field. In practice, it is sufficient to set  $\sigma_s$ ,  $\lambda_s$  and  $\gamma$  fixed (e.g.  $\sigma_s = 500$ ,  $\lambda_s = 0.01$  and  $\gamma = 1$ ) and only optimize over  $\sigma_f$  and  $\lambda_f$ , since the former parameters don't affect the estimation results much. Finally, grid search is performed for estimation of  $\sigma_f \in \{10, 100, 1000, 10000\}$  and  $\lambda_f \in \{0.0001, 0.001, 0.01, 0.1\}$ .

### 3 Experiments and Discussion

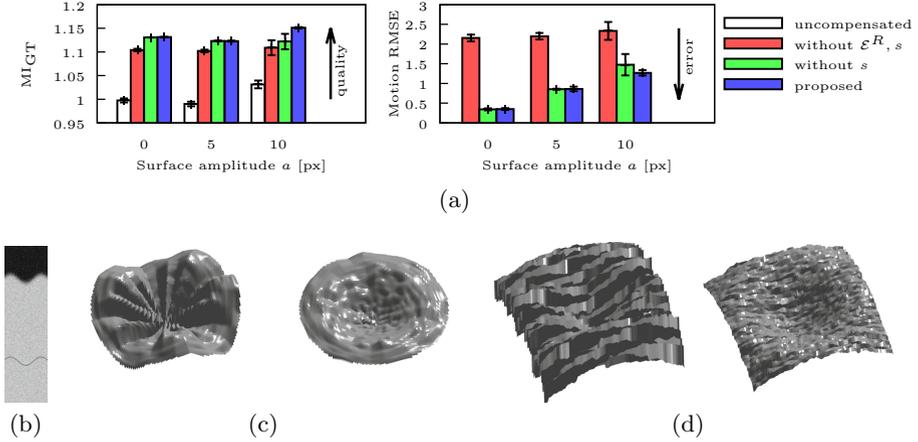
In this section, we compare three different settings of our proposed method. The first setting uses  $\gamma = 1$ ,  $\sigma_s = 500$  and  $\lambda_s = 0.01$ . In the second setting, the tissue surface is omitted, enforcing  $s \equiv 0$ , i.e.  $\sigma_s \rightarrow 0$  and  $\lambda_s \rightarrow \infty$ . The third setting additionally omits the spatial crosscorrelation ( $\mathcal{E}^R$ ) term, i.e.  $\gamma = 0$ , leading to a configuration most similar to the approach of Kraus et al. [4].

We present two different experiments involving synthetic data and real OCT acquisitions. The first experiments are done on synthetic datasets, where ground truth surface and motion fields are available. In a second part, real OCT scans of both post mortem (with artificial motion field) and in vivo (without prior known motion field) are evaluated. For synthetic data, as well as real OCT measures, the subjects are scanned consecutively with two scanning schemes for ensuring that enough surface points are scanned twice. The first scheme is a spoke pattern scanning scheme with  $N_{\text{spoke}}$  B-scans, each B-scan consists of  $N_A$  A-scans as shown in Fig. 3(a). The second scanning scheme is a dense 3D (cuboidal) scanning scheme with  $N_{3D}$  B-scans. Figure 3(b) shows a schematic of the lateral scanning positions over time of a complete subject scan. Figure 3(c) shows the spatial structure of neighboring, crosscorrelated A-scans (encoded in  $\mathcal{E}^R$ ).

**Synthetic data** is generated with  $N_{\text{spoke}} = 16$ ,  $N_{3D} = 100$ ,  $N_A = 100$  and axial resolution of  $Z = 600$  px. The tissue is modeled as a uniformly scattered medium with the tissue-air interface modeled by a step edge function convolved with a Gaussian kernel with  $\sigma_{\text{step}} = 5$ . The image intensities (with range  $[0, 1]$ )



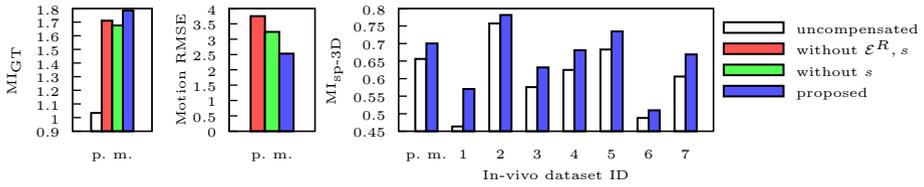
**Fig. 3.** Scanning schemes: (a) spoke and dense 3D pattern, (b) lateral positions of B-scans (red lines) over time. Dashed blue lines: connection of consecutive B-scans. Color shading and arrows depict temporal direction, (c) spatial structure of  $\mathcal{E}^R$  for spoke and 3D pattern. Blue points: A-scans  $d_i$ , black lines: neighborhood relations  $(i, j) \in \mathcal{E}^R$ .



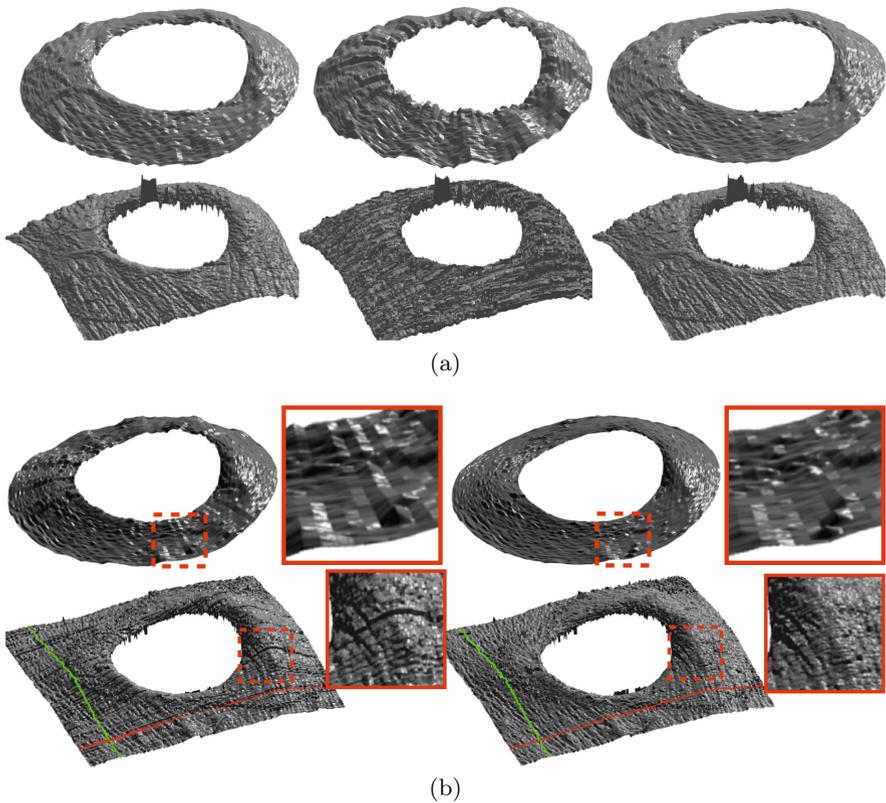
**Fig. 4.** Synthetic data results: (a) Evaluation of the mutual information towards the ground truth ( $MI_{GT}$ , left) and the RMSE of the motion fields towards ground truth for different surface amplitudes (right) and (b)–(d) example surface segmentation of a synthetic dataset with non-planar surface. (b) Example B-scan slice, (c) Spoke scans and (d) dense 3D volume scans. Left: with motion artifacts, right: motion compensated.

are corrupted with additive Gaussian noise with  $\sigma_{\text{noise}}^2 = 0.07$ . Artificial motion artifacts were generated by adding two sine waves of random amplitude and phase to simulate periodic movement. Low-frequency random shift of up to  $\pm 20$  px is added for simulation of non-periodic movement. We evaluated our motion compensation algorithm on data with sinusoidal tissue surface of amplitude  $a$  as shown in Fig. 4(b). Performance evaluation is done using mutual information (MI) inspired by [4], i.e. measuring the similarity between spoke scan volume and 3D scan volume (resliced to capture the same regions as the spoke scan), denoted with  $MI_{\text{sp-3D}}$ . Since ground truth volumes for synthetic data is available, we can also compute the MI of the ground truth volume scans to its motion compensated volume, denoted with  $MI_{GT}$ . Figure 4(a) shows  $MI_{GT}$  and motion RMSE results of 30 randomly generated datasets with varying tissue surface amplitudes (10 datasets for  $a = 0, 5,$  and  $10$  respectively) with errorbars indicating the standard deviation. The results show best performance on the first configuration, showing most increase of MI and least motion RMSE. The configuration ommitting  $\mathcal{E}^R$  and  $s$  performs worst on every dataset.  $MI_{\text{sp-3D}}$  gives nearly similar results for every configuration, since this measure only captures intra-volume similarity enforced by the  $\mathcal{E}^c$  term and cannot capture the tissue structure preservation, as noticed in [4]. Figure 4(c)–(d) shows a comparison of extracted tissue surface renderings of uncompensated to compensated volumes.

**Real OCT Scans:** Our real world application uses *in vivo* and post mortem SD-OCT scans of the percutaneous implant of an anesthetized (and fixed) mouse from [6]. The setting has following parameters:  $N_{\text{spoke}} = 72$ ,  $N_{3D} = 800$ ,  $N_A =$



**Fig. 5.** Evaluation on post mortem data (with known ground truth) using  $MI_{GT}$  and motion RMSE and in vivo data (without known ground truth) using  $MI_{sp-3D}$



**Fig. 6.** Motion compensation results: (a) post mortem dataset with ground truth (left), artificial motion (middle), and motion compensated (right) and (b) in vivo data with motion artifacts (left) and motion compensated (right). Red and green lines indicating the position of slices and reslices respectively shown in Fig. 1. Top row: spoke pattern scan, bottom row: dense 3D scan respectively.

800 and an axial resolution of  $Z = 600$  px. Acquisition time was approx. 0.1 s per B-scan. For enhancement of computation time and memory usage, a downsampling along the fast scanning axis by a factor of 8 is applied and the motion field is upsampled afterwards for providing motion compensation in full resolution. In Fig. 5, the evaluation results of one post mortem dataset (p. m.) corrupted with artificial motion (thus known ground truth) and several *in vivo* datasets (without known ground truth) are shown. For both post mortem and *in vivo* scans a increase of  $MI_{GT}$  and  $MI_{sp-3D}$ , respectively, is observed, showing a significant reduction of motion artifacts. This finding can also be observed in the surface segmentation visualization of the post mortem dataset (see Fig. 6(a)) and a typical *in vivo* data example as shown in Fig. 6(b) and Fig. 1.

## 4 Conclusion

In this work, we propose a novel probabilistic approach for motion compensation of *in vivo* SD-OCT volume scans. The motion estimation problem is reformulated as a CRF energy function and approximated by a GMRF for efficient inference. Our method reliably separates axial motion from tissue structure change by combining two scanning schemes. We use multiple A-scans taken at the same lateral position but different time points as anchor points to estimate the tissue morphology. The method is verified on synthetic data as well as *in vivo* SD-OCT volume scans. Motion artifacts are significantly reduced while the geometry of the tissue is preserved.

## References

1. Yun, S.H., Tearney, G.J., de Boer, J.F., Bouma, B.E.: Motion artifacts in optical coherence tomography with frequency-domain ranging. *Optics Express* 12, 2977 (2004)
2. Ricco, S., Chen, M., Ishikawa, H., Wollstein, G., Schuman, J.: Correcting Motion Artifacts in Retinal Spectral Domain Optical Coherence Tomography via Image Registration. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 100–107. Springer, Heidelberg (2009)
3. Lee, J., Srinivasan, V., Radhakrishnan, H., Boas, D.A.: Motion correction for phase-resolved dynamic optical coherence tomography imaging of rodent cerebral cortex. *Optics Express* 19(22), 21258–21270 (2011)
4. Kraus, M.F., Potsaid, B., Mayer, M.A., Bock, R., Baumann, B., Liu, J.J., Hornegger, J., Fujimoto, J.G.: Motion correction in optical coherence tomography volumes on a per A-scan basis using orthogonal scan patterns. *Biomedical Optics Express* 3(6), 1182–1199 (2012)
5. Rue, H., Held, L.: *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC, Boca Raton (2005)
6. Müller, O., Donner, S., Klinder, T., Dragon, R., Bartsch, I., Witte, F., Krüger, A., Heisterkamp, A., Rosenhahn, B.: Model Based 3D Segmentation and OCT Image Undistortion of Percutaneous Implants. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 454–462. Springer, Heidelberg (2011)