Probabilistic Intra-Retinal Layer Segmentation in 3-D OCT Images Using Global Shape Regularization

Fabian Rathke[a,*], Stefan Schmidt[a,b], Christoph Schnörr[a]

[a]*Image & Pattern Analysis Group (IPA) and Heidelberg Collaboratory for Image Processing (HCI), University of Heidelberg, Speyerer Str. 6, 69126 Heidelberg, Germany*
[b]*Heidelberg Engineering GmbH, Tiergartenstrasse 15, 69121 Heidelberg, Germany*

Abstract

With the introduction of spectral-domain optical coherence tomography (OCT), resulting in a significant increase in acquisition speed, the fast and accurate segmentation of 3-D OCT scans has become evermore important. This paper presents a novel probabilistic approach, that models the appearance of retinal layers as well as the *global* shape variations of layer boundaries. Given an OCT scan, the *full* posterior distribution over segmentations is approximately inferred using a variational method enabling efficient probabilistic inference in terms of computationally tractable model components: Segmenting a full 3-D volume takes around a minute. Accurate segmentations demonstrate the benefit of using global shape regularization: We segmented 35 fovea-centered 3-D volumes with an average unsigned error of $2.46 \pm 0.22 \, \mu m$ as well as 80 normal and 66 glaucomatous 2-D circular scans with errors of $2.92 \pm 0.53 \, \mu m$ and $4.09 \pm 0.98 \, \mu m$ respectively. Furthermore, we utilized the inferred posterior distribution to rate the quality of the segmentation, point out potentially erroneous regions and discriminate normal from pathological scans. No pre- or postprocessing was required and we used the same set of parameters for all data sets, underlining the robustness and out-of-the-box nature of our approach.

Keywords: Statistical shape model, Retinal layer segmentation, Pathology detection, Optical coherence tomography

*Corresponding author, *Phone*: +49 6221 548787, *Address*: Image & Pattern Analysis Group (IPA), Speyerer Str. 6, 69126 Heidelberg, Germany, *Email*: fabian.rathke@iwr.uni-heidelberg.de

Email addresses: `fabian.rathke@iwr.uni-heidelberg.de` (Fabian Rathke), `sschmidt@HeidelbergEngineering.com` (Stefan Schmidt), `schnoerr@math.uni-heidelberg.de` (Christoph Schnörr)

1. Introduction

Optical coherence tomography (OCT) is an *in vivo* imaging technique, measuring the delay and magnitude of backscattered light. Providing micrometer resolution and millimeter penetration depth into retinal tissue (Drexler and Fujimoto, 2008), OCT is well suited for ophthalmic imaging. Since no other method can perform noninvasive imaging with such a resolution, OCT has become a standard in clinical ophthalmology (Schuman et al., 2004). Several studies showed the applicability for the diagnosis of pathologies such as glaucoma or age-related macular degeneration (Bowd et al., 2001; Zysk et al., 2007). The recent introduction (de Boer et al., 2003; Wojtkowski et al., 2002) of spectral-domain OCT dramatically increased the imaging speed and enabled the acquisition of 3-D volumes containing hundreds of B-scans. Since manual segmentation of retinal layers is tedious and time-consuming, automated segmentation becomes evermore important given the growing amount of gathered data. Furthermore, a probabilistic model that enables to infer uncertainties of estimates, provides essential information for practitioners, in addition to the segmentation result.

Various approaches for the task of retina segmentation in OCT images were published. All have in common that they generate appearance terms based either on intensity or gradient information. On top of that regularization is applied, which makes predictions more robust to speckle noise or shadowing caused by blood vessels. In order to provide a systematic overview over this vast field of approaches, we choose to distinguish them by the method used for regularization.

One major class is composed of rule-based heuristic techniques (Ahlers et al., 2008; Fernández et al., 2005; Ishikawa et al., 2005; Mayer et al., 2010), which for example apply outlier detection along with linear interpolation to account for erroneous segmentations. Other approaches (Baroni et al., 2007; Yang et al., 2010) use dynamic programming for single Markov chains per boundary and constrain the maximal vertical distance between neighboring boundary positions. Vermeer et al. (2011) classify pixels using support vector machines and regularize the output using level-set techniques. None of these approaches incorporates shape prior information.

Active contour approaches include gradient respectively intensity-based methods (Mishra et al., 2009; Yazdanpanah et al., 2009, 2011). Yazdanpanah et al. (2009, 2011) augment the classical active contour functional by a simple circular shape prior. All three approaches were only tested on OCT-scans

that exclude the foveal region, thus contain mainly flat boundaries with rather simple shapes.

A series of more advanced approaches (Antony et al., 2010; Dufour et al., 2013; Garvin et al., 2009; Song et al., 2013) construct a geometric graph to simultaneously segment all boundaries in a 3-D OCT volume. Unlike previously presented approaches, they take into account the interaction of neighboring boundaries to mutually restrict their relative positions. This shape prior information is encoded into the graph as hard constraints (Antony et al., 2010; Garvin et al., 2009) or, as recently introduced by Song et al. (2013) and subsequently extended by Dufour et al. (2013), as probabilistic soft constraints. However, due to computational limitations, only *local* shape information is included and boundaries are segmented in stages.

Finally, Kajić et al. (2010) apply the popular active appearance models that match statistical models for appearance and shape, to a given OCT scan. Although non-local shape modeling is in the scope of their approach, they only use landmarks, i.e. sparsely sampled boundary positions instead of the full shape model. Furthermore, only a maximum likelihood point estimate is inferred, instead of a distribution over shapes.

Contribution. We present a novel probabilistic approach for the OCT retina segmentation problem. Our probabilistic graphical model combines appearance models with a global shape prior, that comprises local as well as long-range interactions between boundaries. The discrete part of the model features a highly parallelizable column-wise discrete segmentation, that nevertheless takes into account all other image columns. In order to infer the posterior probability of this model, we utilize variational inference, a deterministic approximation framework.

To our knowledge this is the only work, where a full *global* shape prior is employed for the task of OCT retina segmentation. Moreover, we are not aware of any other segmentation approach that infers a full probability distribution. Our approach offers excellent segmentation performance, outperforming approaches relying on local or no shape regularization, as well as pathology detection and an assessment of segmentation quality. Fig. 1 illustrates the segmented boundaries, but additional boundaries like the external limiting membrane (ELM) could easily be incorporated if ground truth is available.

This work evolved out of preliminary ideas presented in a previous conference paper (Rathke et al., 2011).

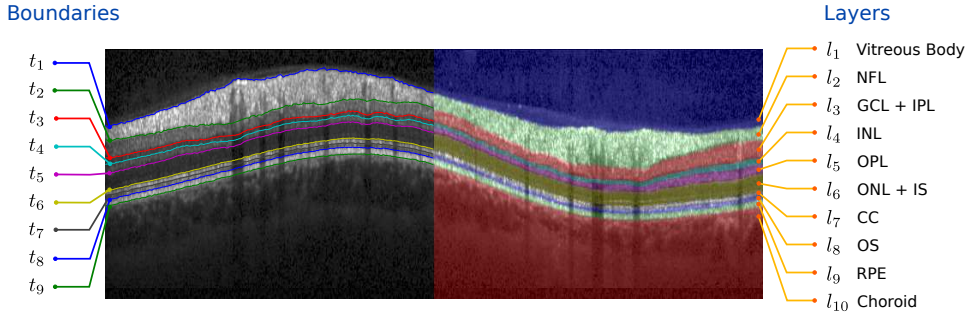**Organization.** The next section will introduce our probabilistic graphical

3

Fig. 1: Overview of retinal layers segmented by our approach and their corresponding anatomical names. The used abbreviations correspond to nerve fibre layer (NFL), ganglion cell layer and inner plexiform layer (GCL + IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer and inner segment (ONL + IS), connecting cilia (CC), outer segment (OS), retinal pigment epithelium (RPE).

model. Section 3 evaluates the posterior distribution via variational inference, and we solve the corresponding optimization problem in Section 4 in terms of efficiently solvable convex subproblems. Section 5 and 6 present the data sets we used for evaluation and the corresponding results. We conclude in Section 7 with a discussion and possible directions for future work.

2. Graphical Model

This section presents our probabilistic graphical model, statistically modeling an OCT scan y and its segmentations b and c respectively. We introduce c, the discretized version of the continuous boundary vector b, to make mathematically explicit the connection between the discrete pixel domain of y and the continuous boundary domain of b. Our ansatz is given by

$$p(y, c, b) = p(y|c)p(c|b)p(b), \tag{1}$$

where the factors are

$p(y|c)$ appearance, data likelihood term,
$p(c|b)$ Markov Random Field regularizer, determined by the shape prior and
$p(b)$ global shape prior.

In what follows we will detail each component, thereby completing the definition of our graphical model. Fig. 2 illustrates our graphical model in terms of the connectivity of the individual model layers.

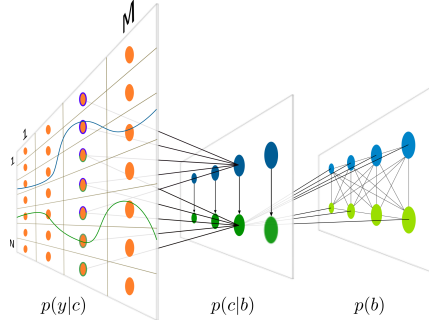**Notation.** The following notation is used throughout the paper:

Fig. 2: Illustration of our graphical model for $M = 4$, $N = 7$ and $N_b = 2$. The connectivity from b to c is only displayed for node $c_{2,3}$. Similarly, connectivity for c to y via x is only displayed for nodes in the third image column and additionally illustrated by the boundary colors of the y-nodes.
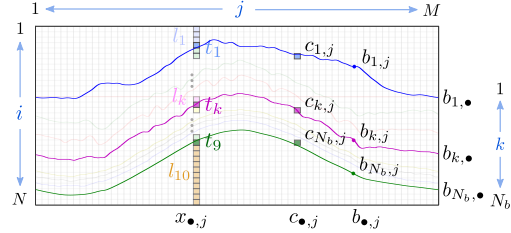
Fig. 3: Illustration of important variables used throughout the paper. Note the difference between real valued position $b_{k,j}$ of boundary k in column j and its discretized equivalent $c_{k,j}$.

N, M	OCT scan dimensions (rows, columns);
N_b	number of segmented boundaries; $N_b = 9$ in this paper;
i, j, k	corresponding indices:
	$i = 1, \ldots, N, \ j = 1, \ldots, M, \ k = 1, \ldots, N_b$;
$b_{k,j} \in \mathbb{R}$	real-valued location of boundary k in column j;
$c_{k,j} \in \{1, \ldots, N\}$	integer-valued boundary variables analogous to b,
	but specifying row-positions on the pixel grid;
$x_{i,j} \in \mathcal{X}$	class variables indicating membership to
	layer or transition classes;
$y_{i,j}$	observed data; here patches around pixel (i, j)
	projected onto a low-dimensional manifold
Δ_N	standard (N-1)-simplex: $\forall \theta \in \Delta_N : \sum_{i=1}^{N} \theta_i = 1$

The symbol $\bullet$ denotes the set of all elements of the respective index, for example $b_{k,\bullet} \in \mathbb{R}^M$ is the location vector for boundary k. By $b_{\setminus j}$ we denote the set $b \setminus b_{\bullet,j}$, with similar notations used for μ and Σ. See Fig. 3 for an illustration of most of the notation introduced here.

2.1. Appearance Models

We utilize Gaussian distributions to model the appearance of retinal layers as well as boundaries. Given a segmentation hypothesis c, we can

assign class labels $x_{i,j} \in \mathcal{X}$ to each pixel, their range being given by

$$\mathcal{X} = \{\mathcal{X}_l, \mathcal{X}_t\}, \qquad \mathcal{X}_l = \{l_1, \ldots, l_{10}\}, \quad \mathcal{X}_t = \{t_1, \ldots, t_9\},$$

which represent classes of observations corresponding to tissue layers $l_1, \ldots, l_{10}$ and transitions $t_1, \ldots, t_9$ separating them. To obtain a valid mapping $c \mapsto x$, we require c to satisfy the ordering constraint ,

$$1 \le c_{1,j} < c_{2,j} < \cdots < c_{N_b,j} \le N, \quad \forall j = 1, \ldots, M, \tag{2}$$

and point out that the real-valued counterpart b may violate this constraint.

Since OCT scans display a high variability in brightness and contrast within and between scans, each patch is first normalized by subtracting its mean. We then project each patch $y_{i,j}$ onto a low-dimensional manifold. Applying the technique of Principle Component Analysis (PCA), we draw patches randomly from the OCT scans in the training set, estimate their empirical covariance matrix and calculate its eigenvalues and eigenvectors. The projection can then be carried out using the first q_{pca} eigenvectors sorted by their eigenvalues.

We define the probability of the projected patch $y_{i,j}{}^1$ around pixel (i,j) belonging to class $x_{i,j}$ as

$$p(y_{i,j}|x_{i,j}(c)) = \mathcal{N}(y_{i,j}; \mu_{x_{i,j}}, \Sigma_{x_{i,j}}). \tag{3}$$

The class-specific moments $\mu_x, \Sigma_x \; \forall x \in \mathcal{X}$ are learned offline using patches from the respective class. Regularized estimates for Σ_x are obtained by utilizing the graphical lasso approach (Friedman et al., 2008), which augments the classical maximum likelihood estimate for Σ with an ℓ_1-norm on the precision matrix $K = \Sigma^{-1}$. This leads to sparse estimates for K, where the degree of sparsity is governed by a parameter α_{glasso}, c.f. Section 5.3.

We define the appearance of a scan y to factorize over pixels (i,j). Finally, we introduce switches $\beta^t \in \{0,1\}$ and $\beta^l \in \{0,1\}$, that turn on/off all terms belonging to any transition class t_k or layer class l_k, which yields the final appearance model

$$p(y|c) = \prod_{j=1}^{M} \prod_{i:x_{i,j}\in\mathcal{X}_l} p(y_{i,j}|x_{i,j}(c))^{\beta^l} \prod_{i:x_{i,j}\in\mathcal{X}_t} p(y_{i,j}|x_{i,j}(c))^{\beta^t}. \tag{4}$$

---

[1]For ease of notation, we will make no difference between the patch $y_{i,j}$ and its low-dimensional projection.

As we point out in the section about inference, our model can handle discriminative terms as well. We can convert generative terms (3) into discriminative ones by renormalizing

$$p(x_{i,j}(c)|y_{i,j}) = \frac{p(y_{i,j}|x_{i,j}(c))p(x_{i,j}(c))}{\sum_{x_{i,j}\in\mathcal{X}} p(y_{i,j}|x_{i,j}(c))p(x_{i,j}(c))},$$ (5)

where we use a uniform prior $p(x_{i,j}(c))$. The factorization of $p(c|y)$ is the same as in (4).

2.2. Shape Prior

As a model of the typical shape variation of layers due to both biological variability as well as to the image formation process, we adopt a joint Gaussian distribution[2]. We denote the continuous height values of all boundaries k for image columns j by the $N_b M$-dimensional vector $b = (b_{k,j})_{k=1,\ldots,N_b;\,j=1,\ldots,M}$. Hence,

$$p(b) = \mathcal{N}(b; \mu, \Sigma)$$ (6)

where parameters μ and Σ are learned offline from labeled training data. We regularize the estimation of Σ by Probabilistic Principal Component Analysis (PPCA) (Tipping and Bishop, 1999). PPCA assumes that the high-dimensional observation b was generated from a low-dimensional latent source $s \in \mathbb{R}^q$ via

$$b = Ws + \mu + \epsilon,$$

where $s \sim \mathcal{N}(0, I)$ and $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ is isotropic Gaussian noise[3].

The moments of $p(b)$ are given by $\mathbb{E}[b] = \mu$ and $\mathbb{V}[b] = WW^T + \sigma^2 I = \Sigma$. Likewise, Σ^{-1} can be decomposed into W and $\sigma^2 I$ too. Making use of these decompositions, one can reduce complexity of most operations related to Σ or Σ^{-1} as well as memory requirements, since only W and σ^2 have to be stored. The parameters for $p(b)$ can be estimated via maximum log-likelihood. W is composed of the (weighted) q_{ppca} eigenvectors with largest eigenvalues, computed from the empirical covariance matrix. For more details, we refer to Tipping and Bishop (1999).

---

[2]For circular scans, a wave-like distortion pattern is observed due to the conic scanning geometry and the spherical shape of the retina, which we capture statistically rather than modeling it explicitly.

[3]PPCA can be considered as a generalisation of classical Principle Component Analysis (PCA), which assumes the deterministic relation $b = Ws + \mu$.
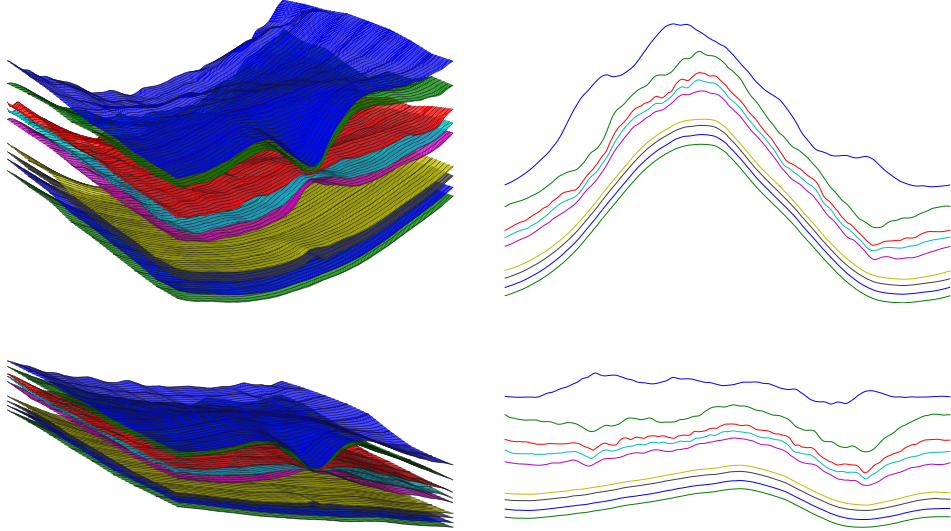
Fig. 4: Samples generated by the shape prior distribution $p(b)$ trained on volumes (left) and circular scans (right). Only one half of the volume is shown.

Fig. 4 shows samples drawn from $p(b)$, modeling fovea-centered 3-D volumes (left) and circular scans (right). Additionally, as supplementary material we provide a video that visualizes the relevance of each component of W exemplary for the 3-D setup, like translation, rotation, thickness of layers or position and form of the fovea.

2.3. Shape-Induced Regularizers

The third component of our model consists of a prior for discrete boundary assignments c, regularizing the data likelihood terms $p(y|c)$. We define $p(c|b)$ as column-wise acyclic graphs

$$p(c|b) = \prod_{j=1}^{M} p(c_{\bullet,j}|b), \qquad p(c_{\bullet,j}|b) = p(c_{1,j}|b) \prod_{k=2}^{N_b} p(c_{k,j}|c_{k-1,j}, b), \quad (7)$$

i.e. the communication of the model *between* image columns j is governed by the shape prior $p(b)$.

In order to define the conditional marginals in (7), we need a couple of prerequisites. With $b_{\bullet,\backslash j}$ denoting the subset of variables b after removing variables $b_{\bullet,j}$ of column j, and with $p(b_{\bullet,j}|b_{\bullet,\backslash j})$ denoting the corresponding conditional Gaussian distribution computed from the shape prior $p(b)$, then

8

the marginal distributions are specified in terms of b by

$$p(c_{1,j}=n|b) = \Pr\left(n-\frac{1}{2} \leq b_{1,j} \leq n+\frac{1}{2}\right),$$

$$p(c_{k,j}=n|c_{k-1,j}=m,b) = \tag{8}$$

$$\Pr\left(n-\frac{1}{2} \leq b_{k,j} \leq n+\frac{1}{2} \,\Big|\, m-\frac{1}{2} \leq b_{k-1,j} \leq m+\frac{1}{2}\right),$$

where the probabilities on the right-hand side are computed using the conditional marginal densities $p(b_{1,j}|b_{\bullet,\backslash j})$ and $p(b_{k,j}|b_{\bullet,\backslash j})p(b_{k,j}|b_{k-1,j})$ respectively, for all configurations of c conforming to $(2)^4$. The marginal $p(b_{k,j}|b_{\bullet,\backslash j})$ provides a way to introduce global shape knowledge into our column-wise graphical models $p(c|b)$.

2.4. 2-D vs. 3-D

Our description so far considered OCT scans y of dimensionality two. Nevertheless, our approach is equally applicable to 3-D volumes. We use the very same notation, since adding additional B-Scans will only increase the number of image columns M. Similarly, the connectivity of the graphical model $p(y,c,b)$ can be transferred one-to-one.

The shape prior $p(b)$ which is fully connected since Σ^{-1} is dense, can be extended to an arbitrary dimension. We exploit the fact that both, Σ and Σ^{-1}, have an explicit low-rank decomposition (see section 2.2), such that memory consumption is not an issue and complexity of operations is reduced as well. For the shape regularization term $p(c|b)$, each node $c_{k,j}$ is connected to nodes $b_{\backslash j}$ of all columns except the current one, which now additionally includes columns of all other B-scans. Finally, the data likelihood $p(y|c)$ continues to fully factorize over pixels (i,j). Each pixel (i,j) remains connected to at most two nodes $c_{k,j}$ from the same column j, determining it's label $x_{i,j}$. Furthermore, we use separate sets of appearance models for each B-scan in the volume to capture possible variations.

3. Variational Inference

Based on the model presented in the previous section and given observed data y, we wish to infer the posterior

$$p(b,c|y) = \frac{p(y|c)p(c|b)p(b)}{p(y)}. \tag{9}$$

---

[4]This computation is straightforward for Gaussian distributions (6), see Section 3.2.

One major issue is that calculating the marginal likelihood $p(y)$ would require integrating $p(y, c, b)$ over b and c. This turns out to be intractable, since we lack a closed form solution and the problem at hand is high-dimensional. We cope with this problem by applying an established variational method: approximating the posterior by a tractable distribution $q(b, c)$ by minimizing the Kullback-Leibler (KL) distance $KL(q\|p)$ with respect to q (cf., e.g., Attias (2000)). We point out that unlike in related work (e.g. McGrory et al. (2009)) where the subproblem of inferring the discrete decision variables has to be approximated as well, our model has been designed such that by choosing q properly all subproblems are tractable and can be solved efficiently.

We choose a factorized approximating distribution

$$q(b, c) = q_b(b)q_c(c). \tag{10}$$

This merely decouples the continuous shape prior and the discrete order-preserving segmentation component of the overall model, but otherwise represents both components exactly. The Kullback-Leibler distance between q and p is given by

$$\begin{aligned}
KL\left(q(b, c)\|p(b, c|y)\right) &= \int_b \sum_c q(b, c) \log \frac{q(b, c)}{p(b, c|y)} db \\
&= -\int_b \sum_c q(b, c) \Big(\log \left(p(y|c)p(c|b)p(b)\right) - \log p(y) - \log q(b, c) \Big) db.
\end{aligned} \tag{11}$$

Dropping the constant term $\log p(y)$, we may obtain our objective function.

Alternatively, we can use the marginal likelihood $\log p(y)$ to introduce *discriminative* appearance terms into the model, using $\log \frac{p(y|c)}{p(y)} = \log \frac{p(y|c)p(c)}{p(y)} - \log p(c) = \log p(c|y) - \log p(c)$. Since $p(b)$ already contains prior knowledge about the shape of boundary positions, we assume an uninformative prior for c. Dropping thus $p(c)$ and taking into account the factorization of q, we obtain the objective function

$$J(q_b, q_c) = -\int_b \sum_c q_b(b)q_c(c) \log \left(p(c|y)p(c|b)p(b)\right) db - H_{q_b}(b) - H_{q_c}(c), \tag{12}$$

where $H_p(x)$ denotes the entropy of the distribution p. It turned out that discriminative appearance terms yielded the best performance. A discussion of this issue will be given in Section 5.2.

Definitions of q_c and q_b. For $q_c(c)$, we adopt the same factorization as for $p(c|b)$, that is

$$q_c(c) = \prod_{j=1}^{M} q_{c;1,j}(c_{1,j}) \prod_{k=2}^{N_b} \frac{q_{c;k \wedge k-1,j}(c_{k,j}, c_{k-1,j})}{q_{c;k-1,j}(c_{k-1,j})} \tag{13}$$

where $q_{c;k,j} \in \Delta_N$ are discrete probability distributions. Similarly, by $q_{c;k \wedge k-1,j} \in \Delta_{N^2}$ we denote discrete probability distributions over pairs of variables $c_{k-1,j}, c_{k,j}{}^5$. For $q_c(c)$ to be a valid distribution, additional marginalization constraints have to be satisfied:

$$\sum_{c_{k-1,j}} q_c(c_{k,j}, c_{k-1,j}) = q_c(c_{k,j}), \quad \sum_{c_{k,j}} q_c(c_{k,j}, c_{k-1,j}) = q_c(c_{k-1,j}), \tag{14}$$

for all $k = 2, \ldots, N_b$ and $j = 1, \ldots, M$. Note that we ignore the set of valid configurations (2) here, because this has already been taken into account when defining $p(c|b)$. As for q_c and $p(c|b)$, we let q_b adopt the same factorization as $p(b)$, thus

$$q_b(b) = \mathcal{N}(b; \bar{\mu}, \bar{\Sigma}). \tag{15}$$

In what follows, we make the expectations with respect to q_c and q_b explicit. This will provide us below with a closed-form expression of the objective function $J(q_b, q_c)$.

3.1. First Summand $\log p(c|y)$ of $J(q_b, q_c)$

The term $p(c|y)$ does not depend on b, so q_b integrates out. Moreover, both $p(c|y)$ and q_c factorize over columns. Hence we can rewrite the first summand of (12) as

$$-\int_b \sum_c q_b(b) q_c(c) \log p(c|y) = -\sum_{j=1}^{M} \sum_{c_{\bullet,j}} \left(q_c(c_{\bullet,j}) \sum_{i=1}^{N} \log p(x_{i,j}(c_{\bullet,j})|y_{i,j}) \right),$$

where the second sum ranges over all combinations of boundary assignments for $c_{\bullet,j}$. We can further simplify this equation by noting that each label $x_{i,j}$ depends at most on two $c_{k,j}$, as illustrated in Fig. 5. This allows us to split the inner sum into $k+1$ sums, each summing over pixels with labels l_k and t_k or l_{N_b+1} respectively, and sum out all $c_{k,j}$ independent of these labels.

---

[5]To enhance readability, we will subsequently omit indices k, j of q_c, if they are determined by the input variable(s) $c_{k,j}$.
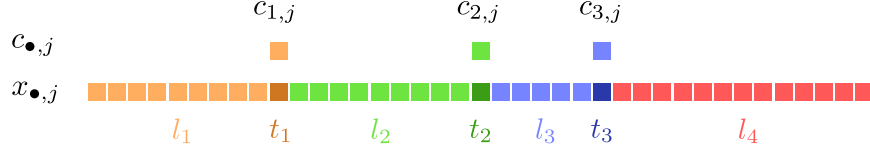
Fig. 5: Distribution of labels of $x_{i,j}$, determined by the position of boundaries $c_{k,j}$, for $N_b = 3$. Each label depends at most on two boundaries.

For each pair (l_k, t_k) of labels we define matrices $\Psi_{k,j}$, whose entries equal the sum over pixel $y_{i,j}$ with $x_{i,j} \in \{t_k, l_k\}$

$$(\Psi_{k,j})_{m,n} = \sum_{i=m+1}^{n-1} \beta^l \log p(x_{i,j} = l_k | y_{i,j}) + \beta^t \log p(x_{n,j} = t_k | y_{n,j})$$

for $k = 2, \ldots, N_b$, $j = 1, \ldots, M$ and $1 \leq m < n \leq N$. Entries for $n \leq m$ are not defined and set to negative infinity. Accordingly, we introduce vectors $(\psi_{1,j})_n$ and $(\psi_{N_b,j})_n$, representing sums over pixels with labels l_1, t_1 and l_{N_b+1} depending on $c_{1,j}$ and $c_{N_b,j}$ respectively. We now can write

$$\sum_{c_{k-1,j}} \sum_{c_{k,j}} q_c(c_{k,j}, c_{k-1,j})(\Psi_{k,j})_{c_{k-1,j}, c_{k,j}} = \langle (q_{c;k \wedge k-1,j})^T, \Psi_{k,j} \rangle$$

where $\langle A, B \rangle$ denotes the trace of AB and determines the form of $q_{c;k \wedge k-1,j}$. Finally, we write the first summand of (12) in vector form:

$$-\sum_{j=1}^{M} \left((q_{c;1,j})^T \psi_{1,j} + \sum_{k=2}^{N_b} \langle (q_{c;k \wedge k-1,j})^T, \Psi_{k,j} \rangle + q_{c;N_b,j}^T \psi_{N_b,j} \right). \tag{16}$$

3.2. Second Summand $\log p(c|b)$ of $J(q_b, q_c)$

The second term $p(c|b)$ depends on q_c *and* q_b, so we have to take care of both expected distributions. We start out this section by making explicit the expectation with respect to q_b. As a prerequisite, we define the two marginal distributions of $p(b)$ introduced in (8). Using the simplifying notation $p(b_j|b_{\setminus j}) = p(b_{\bullet,j}|b_{\bullet,\setminus j})$, by the standard rule for conditional Gaussian distributions (Rasmussen and Williams, 2006, p. 200) we obtain

$$p(b_j|b_{\setminus j}) = \mathcal{N}(b_j; \mu_{j|\setminus j}, \Sigma_{j|\setminus j})$$
$$\mu_{j|\setminus j} = \mu_j - \Sigma_{j|\setminus j} K_{j,\setminus j}(b_{\setminus j} - \mu_{\setminus j}), \qquad \Sigma_{j|\setminus j} = (K_{jj})^{-1}, \tag{17}$$

the marginal distribution of the N_b boundary positions in column j, conditioned on the $N_b(M-1)$ remaining boundary positions $b_{\setminus j}$. The 1-dimensional densities for $b_{k,j}|b_{\setminus j}$ are obtained by marginalizing over (17), with mean $(\mu_{j|\setminus j})_k$ and variance $(\Sigma_{j|\setminus j})_{k,k}$.

Similarly, we define $p(b_{k,j}|b_{k-1,j})$ the density of boundary position $b_{k,j}$ given the position of the neighboring boundary $k-1$ in column j. Its mean $\mu_{k|k-1,j}$ and variance $\sigma^2_{k|k-1,j}$ are calculated in the same fashion as in (17). We now can express the probabilities $p(c_{1,j}|b)$ and $p(c_{k,j}|c_{k-1,j}, b)$ introduced in (8) in terms of integrals

$$p(c_{k,j} = n|c_{k-1,j} = m, b) =$$

$$\int_{n-\frac{1}{2}}^{n+\frac{1}{2}} \int_{m-\frac{1}{2}}^{m+\frac{1}{2}} \mathcal{N}(\tau; (\mu_{j|\setminus j})_k, (\Sigma_{j|\setminus j})_{k,k}) \mathcal{N}(\tau; \mu_{k|k-1,j}, \sigma^2_{k|k-1,j})\, d\tau d\nu.$$

and accordingly for $p(c_{1,j} = n|b)^6$. These terms depend on $b_{\setminus j}$ through $(\mu_{j|\setminus j})_k$, hence on q_b too. It suffice to adopt the most crude numerical integration formula (integrand = step function) in order to make this dependency explicit: $\int_{a-1/2}^{a+1/2} f(x)dx \approx f(a)$.

Applying the logarithm to $p(c|b)$, we obtain a representation that is convenient for $\int_b \cdots q_b db$. We defined q_b as a Gaussian distribution (15), therefore the moments of $b_{\setminus j}$ with respect to q_b are given by

$$\mathbb{E}_{q_b}[b_{\setminus j}] = \bar{\mu}_{\setminus j} \qquad \text{and} \qquad \mathbb{V}_{q_b}[b_{\setminus j}] = \bar{\Sigma}_{\setminus j, \setminus j} + \bar{\mu}_{\setminus j} \bar{\mu}_{\setminus j}^T. \tag{18}$$

As a result, we established all necessary prerequisites to write the terms $\mathbb{E}_{q_b}[\log p(c_{1,j}|b)]$ and $\mathbb{E}_{q_b}[\log p(c_{k,j}|c_{k-1,j}, b)]$ in an explicit form, that is suitable for an optimization with respect to $\bar{\mu}$ and $\bar{\Sigma}$. Details are provided in Appendix A.

We now address the expectation with respect to q_c. Similar arguments as for $p(c|b)$ hold for $p(c|y)$ too: We can split the sum over $c_{\bullet,j}$ into parts depending (at most) on two neighboring boundaries $c_{k-1,j}$ and $c_{k,j}$. We define matrices $\Omega_{k,j}$ as

$$(\Omega_{k,j})_{m,n} = \mathbb{E}_{q_b}[\log p(c_{k,j} = n|c_{k-1,j} = m, b)],$$

for $k = 2, \ldots, N_b$, $j = 1, \ldots, M$ and $1 \leq m < n \leq N$, and vectors $(\omega_{1,j})_n$ for terms $\mathbb{E}_{q_b}[\log p(c_{1,j} = n|b)]$ accordingly. Finally, we can write the expecta-

---

[6] Note that the dependency on ν is contained in $\mu_{k|k-1,j}$ as $b_{k-1,j}$.
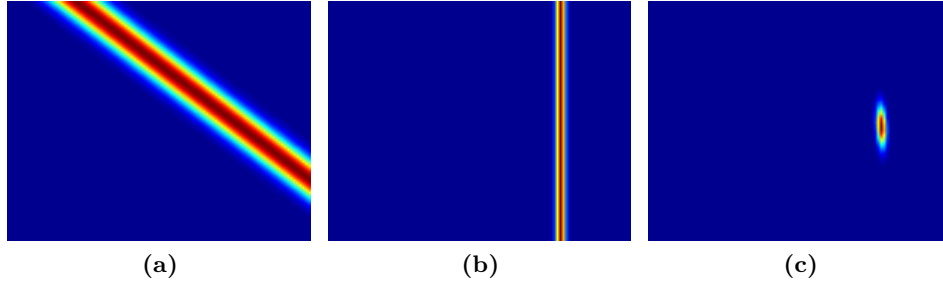
Fig. 6: Illustration of a transition matrix $\exp(\Omega_{k,j})$ (c) and the local (a) and global (b) shape information it is composed of. The plots show the exponential version that is used during the optimization of q_c (see Section 4.1), in order to illustrate the inherent sparsity that we utilize to speed up the calculation of q_c.

tion of the second term in vectorized form as

$$-\sum_{j=1}^{M}\left((q_{c;1,j})^T\omega_{1,j} + \sum_{k=2}^{N_b}\langle(q_{c;k\wedge k-1,j})^T, \Omega_{k,j}\rangle\right). \qquad (19)$$

Fig. 6 shows a transition matrix $\Omega_{k,j}$ (c) and it's two components $\mathbb{E}_{q_b}[\log p(b_{k,j}|b_{k-1,j})]$ (a) and $\mathbb{E}_{q_b}\left[\log p(b_{k,j}|b_{\backslash j})\right]$ (b) for $m, n = 101, \ldots, 200$. Since the sum-product algorithm used to find the optimal q_c (see Section 4.1) requires $\exp(\Omega_{k,j})$, our plots show the exponential version too, in order to illustrate the inherent sparsity. We see how $\Omega_{k,j}$ is build by combining prior information about the relative distance between $b_{k,j}$ and $b_{k-1,j}$ (a) with the distribution of $b_{k,j}$ conditioned on information from all other columns via $\mathbb{E}_{q_b}[b_{\backslash j}] = \bar{\mu}_{\backslash j}$ (b).

3.3. Third Summand $\log p(b)$ of $J(q_b, q_c)$

Concerning the third summand, q_c sums out. Rewriting the Gaussian using the trace and making the expectation explicit, we obtain

$$-\int_b q_b(b)\log p(b)db = C + \frac{1}{2}\langle K, \bar{\Sigma} + \bar{\mu}\bar{\mu}^T - 2\bar{\mu}\mu^T + \mu\mu^T\rangle \qquad (20)$$

i.e. a function depending on the parameters $\bar{\mu}$ and $\bar{\Sigma}$ of q_b.

3.4. Entropy Terms of $J(q_b, q_c)$

Finally, we make explicit the negative entropy of q_b and q_c.

$$-H_{q_b}(b) = \int_b q_b(b) \log q_b(b) db = C - \frac{1}{2} \log |\bar{\Sigma}| . \tag{21}$$

$$-H_{q_c}(c) = \sum_{j=1}^{M} \Bigg(\sum_{k=1}^{N_b} \sum_{c_{k,j}} q_c(c_{k,j}) \log q_c(c_{k,j}) \tag{22}$$

$$+ \sum_{k=2}^{N_b} \sum_{c_{k-1,j}, c_{k,j}} q_c(c_{k,j}, c_{k-1,j}) \log \frac{q_c(c_{k,j}, c_{k-1,j})}{q_c(c_{k-1,j}) q_c(c_{k,j})} \Bigg)$$

The first summand of H_{q_c} is comprises singleton entropies whereas the second one comprises the mutual information (Cover and Thomas, 2006, p. 19) between the random variables $c_{k,j}$ and $c_{k-1,j}$.

3.5. Explicit Formulation of the Objective Function $J(q_b, q_c)$

Combining all terms, we can reformulate (12) into a functional that can be optimized with respect to q_c and the parameters $\bar{\mu}$ and $\bar{\Sigma}$ of q_b

$$\min_{q_c, \bar{\mu}, \bar{\Sigma}} - \Bigg(\sum_{j=1}^{M} (q_{c;1,j})^T \theta_{1,j} + \sum_{k=2}^{N_b} \langle (q_{c;k \wedge k-1,j})^T \Theta_{k,j} \rangle + (q_{c;N_b,j})^T \theta_{N_b,j} \Bigg)$$

$$+ \frac{1}{2} \langle K, \bar{\Sigma} + \bar{\mu}\bar{\mu}^T - 2\bar{\mu}\mu^T \rangle - \frac{1}{2} \log |\bar{\Sigma}| - H_{q_c}(c) + C \tag{23}$$

$$\text{s. t.} \quad q_{c;k,j} \in \Delta_N \quad k = 1, \ldots, N_b, j = 1, \ldots, M \qquad \text{and} \qquad (14)$$

where we combined the terms of (16) and (19) into $\theta_{k,j}$ and $\Theta_{k,j}$. Note that the necessary constraint $\bar{\Sigma} \in \mathcal{S}_{++}$ is implicitly given by the logarithmic barrier term $\log |\bar{\Sigma}| = \sum_i \log \bar{\lambda}_i$, where $\bar{\lambda}_i$ is the i-th eigenvalue of $\bar{\Sigma}$, hence it has not to be enforced explicitly.

4. Optimization

We alternatingly optimize the objective function (23) with respect to q_c and the parameters of q_b. Optimization of q_c corresponds to inference of chain graphs and can be accomplished by the sum-product algorithm (Bishop et al., 2006, p. 402), whereas the optimization of q_b can be done in closed form. Both subproblems are convex, thus by alternatingly optimizing with respect to q_b and q_c, the functional $J(q_b, q_c)$, being bounded from below over the feasible set of variables, is guaranteed to converge to some minimum.

4.1. Optimization of q_c

Fixing all terms in (23) that depend on parameters of q_b, we obtain an optimization problem that can be split into column-wise *convex* subproblems, since each is composed solely of linear terms and the negative entropy of a chain graph, subject to simplex constraints. Making the constraints explicit using Lagrange multipliers, and derivating with respect to all $q_{c;k,j}$, we obtain a set of update equations which can be shown to correspond to sum-product updates (Wainwright and Jordan, 2008, p. 83). Thus iteratively optimizing q_c for each column j is guaranteed to converge to some fix point q_c^{opt}, which corresponds to the global optimum.

4.2. Optimization of q_b

Considering in (23) only terms depending on $\bar{\Sigma}$, we obtain the optimization problem

$$\min_{\bar{\Sigma}} -\frac{1}{2}\log|\bar{\Sigma}| + \frac{1}{2}\langle K + \tilde{P}, \bar{\Sigma}\rangle \tag{24}$$

which has the closed-form solution: $\bar{\Sigma}_{\text{opt}} = (K + \tilde{P})^{-1}$. The newly introduced matrix $\tilde{P}$ contains the dependencies on $\bar{\Sigma}$ of terms $\omega_{1,j}$ and $\Omega_{k,j}$. Being independent of q_c, we only have to calculate it once. Furthermore, since it is composed out of linear combinations of submatrices of K, it can be expressed implicitly in terms of W and $\sigma^2 I$. Details are provided in Appendix B.

For $\bar{\mu}$, we obtain

$$\min_{\bar{\mu}} \frac{1}{2}\langle K + \tilde{P}, \bar{\mu}(\bar{\mu} - 2\mu)^T\rangle + \tilde{p}^T \bar{\mu} \tag{25}$$

and $\bar{\mu}_{\text{opt}} = \mu - \tilde{p}^T \bar{\Sigma}_{\text{opt}}$. Again details for $\tilde{p}$, concerning the dependencies of $\omega_{1,j}$ and $\Omega_{k,j}$, are given in Appendix B. To minimize (25), we use conjugate gradient descent which enables us to calculate $\bar{\mu}_{\text{opt}}$ using $(K + \tilde{P})$ instead of $(K + \tilde{P})^{-1}$.

4.3. Initialization

We start the optimization of (23) by initializing the distribution q_c. This is done by setting distributions $p(b_{k,j}|b_{\setminus j})$ to a uniform distribution, since we yet lack the distribution q_b. Afterwards, we can initialize q_b via (24) and (25). Subsequently, we iterate both optimizations alternatingly until $J(q_b, q_c)$ converges.

5. Experiments

5.1. Data Acquisition

Circular B-scans measured around the optic nerve head were acquired from 80 healthy as well as from 66 glaucomatous subjects using a Spectralis HRA+OCT device (Heidelberg Engineering, Germany). Each scan had a diameter of 12°, corresponding to approximately $3.4\,mm$, and consisted of $M = 768$ A-scans of depth resolution $3.87\mu m/$pixel ($N = 496$ pixels), see Fig. 7 (a). Ground truth for the crucial boundary separating NFL and GCL as well as a grading for the pathological scans was provided by a medical expert: *pre-perimetric* glaucoma (PPG), meaning the eye is exhibiting structural symptoms of the disease but the visual field and sight are not impaired yet, as well as *early*, *moderate* and *advanced* primary open-angle glaucoma (PGE, PGM and PGA). Ground truth for the remaining eight boundaries was produced by the first author. To measure interobserver variability, a second set of labels for the healthy B-scans was obtained by the second author.

The second data set consisted of fovea-centered 3-D volumes, acquired from 35 healthy subjects using the same device as above. Each volume was composed of 61 B-Scans of dimension 500×496, covering an area of approximately $5.7 \times 7.3\,mm$. Ground truth was obtained as follows: Each volume was divided in 17 regions, and a B-scan randomly drawn from each region was labeled with the previously introduced nine boundaries. Fig. 7 (b) depicts the location of all 61 B-Scans and their partition into regions indicated by color.

5.2. Generative vs. Discriminative, Transition vs. Boundary Appearance Terms

Following (11), we described the introduction of discriminative appearance models as an alternative to generative ones. Furthermore, we introduced switches β^l and β^t in (4) to enable or disable layer and boundary appearance terms, respectively. This section explains why we settled for discriminative boundary terms.

Using the set of healthy circular scans, we tested the model with generative layer as well as boundary terms, i.e. $\beta^t = \beta^l = 1$. This configuration turned out to be sensitive to distortions of the texture caused for example by blood vessels. The result were initializations above the actual retina, since the model misinterpreted the shaded area as parts of the choroid, as shown in Fig. 8 (a). We then disabled the layer appearance terms, i.e. set $\beta^l = 0$. This solved the previous issue, but spuriously led to some columns being
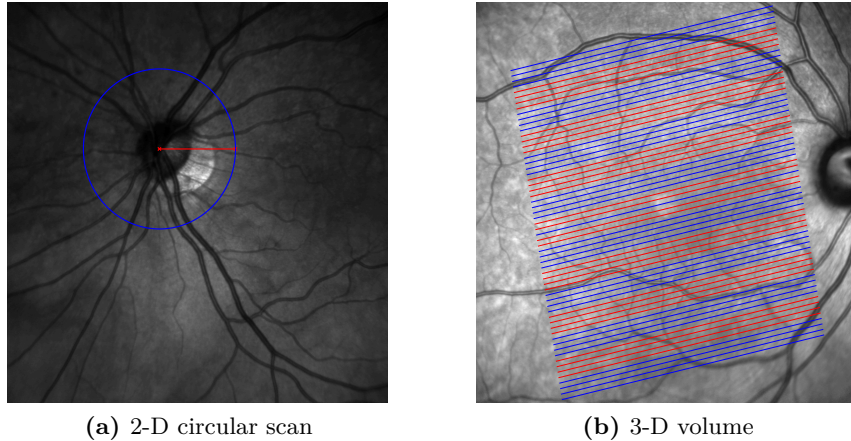
(a) 2-D circular scan   **(b)** 3-D volume

Fig. 7: SLO fundus images that exemplarily depict (a) the trajectory and radius of a 2-D circular scan centered around the optical nerve head and (b) the area covered by a 3-D volume consisting of 61 B-Scans centered at the fovea. Different colors illustrate the partitioning into 17 different regions.

Table 1: Set of model parameter values used throughout all experiments.

Parameters	Appearance			Shape	Inference
	α_{glasso}	q_{pca}	Patch-Size	q_{ppca}	Variance of $p(b_{k,j}|b_{\backslash j})$
Value	0.01	20	15×15	20	10

initialized below the retina, due to very high probabilities for some boundary classes caused by relatively small class model variances, i.e. narrow and steep normal distributions. For patches close to the mean, the probabilities for those classes happened to be up to 100 times larger than for other classes. This caused false positive class responses in the choroid to displace the whole initialization for these columns, as displayed in Fig. 8 (b).

Switching to discriminative probabilities solved this issue as well, since the local normalization limits all probabilities to 1 and gives each appearance class the same influence. Thus false-positives did not possess the probability mass any more to displace the whole column segmentation, see Fig. 8 (c). Notice that the layer terms, although switched off by setting $\beta^l = 0$, are utilized indirectly, since they contribute to the normalization of terms $p(x_{i,j}(c)|y_{i,j})$, see (5). Thus strong layer appearance terms can rule out certain parts of the OCT scan for segmentation.
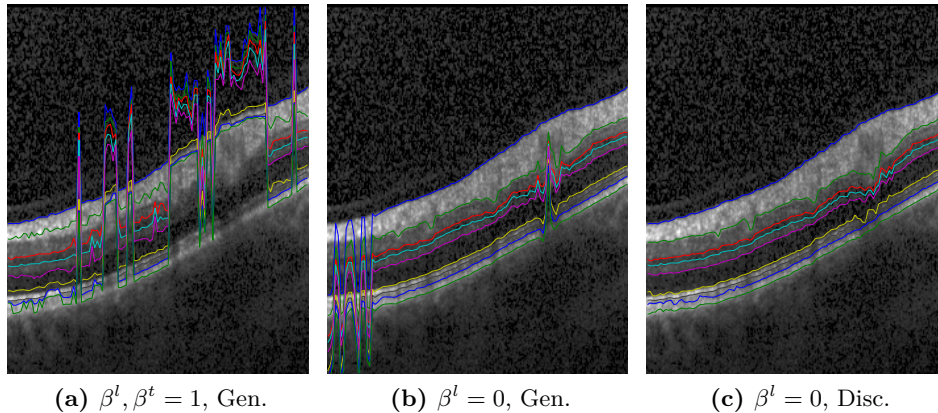
(a) $\beta^l, \beta^t = 1$, Gen. **(b)** $\beta^l = 0$, Gen. **(c)** $\beta^l = 0$, Disc.

Fig. 8: (a)-(c) Close-up view of initialization results for different configurations of appearance terms. Switches β^l and β^t include or exclude layer and transition appearance terms.

5.3. Model Parameters

Table 1 summarizes the model parameters and how they were set. For the appearance model we set α_{glasso} to 0.01, which resulted in sparse covariance matrices $\Sigma_{x_{i,j}}$ that speed up computations significantly. A patch-size of 15×15 and the projection onto the first $q_{\mathrm{pca}} = 20$ eigenvectors resulted in smooth segmentation boundaries. Similar, we used $q_{\mathrm{ppca}} = 20$ eigenvectors to build the shape prior model, after examining the eigenvalue spectrum.

An important parameter during the inference is the variance of $p(b_{k,j}|b_{\setminus j})$: it balances the influence of appearance and shape. Artificially increasing this parameter results in broader normal distributions (cf. Fig. 6 (b)), that allows q_c to take into account more observations around the mean of $p(b_{k,j}|b_{\setminus j})$. At the same time the influence of the appearance terms on q_b is reduced, which results in a smoother mean $\bar{\mu}$. Thus increasing the variance loosens the coupling between q_c and q_b and vice versa. A ten-fold increased variance turned out to provide a good balance between local appearance terms and shape regularization as well as between run-time and prediction accuracy.

We used the very same set of parameter values for all our experiments and performed no fine tuning separately for each data set. Hence it is plausible to assume that these values perform well on a broad range of data sets.

5.4. Error Measures and Test Framework

For each boundary we computed the unsigned distance E^k_{unsgn} in μm between estimates $\hat{c}_{k,j} = \mathbb{E}_{q_c}[c_{k,j}]$ and manual segmentations $\tilde{c}_{k,j}$ (ground truth) as

$$E^k_{\mathrm{unsgn}} = M^{-1} \sum_{j=1}^{M} |\hat{c}_{k,j} - \tilde{c}_{k,j}|, \qquad E_{\mathrm{unsgn}} = N_b^{-1} \sum_{k=1}^{N_b} E^k_{\mathrm{unsgn}}.$$

For volumes we additionally have to average over regions. For each data set, we provide the mean error $\bar{E}_{\mathrm{unsgn}}$ and it's standard deviation (SD) $\sigma_{E_{\mathrm{unsgn}}}$.

Results were obtained via cross-validation: After splitting each data set into a number of subsets, each subset in turn is used as a test set, while the remaining subsets are used for training. This provides an estimate of the ability to segment new (unseen) test scans. We used 10-fold cross-validation for the set of non-pathological circular scans and leave-one-out cross-validation for the volumes, to maximize the number of training examples in each split. For the set of glaucomatous scans, we used a single model trained on all healthy scans.

5.5. Implementation and Running Time

We implemented our approach in MATLAB. The main bottle-neck, the sum-product algorithm used to find an optimal solution for $q_c(c)$, was implemented in C and incorporated into MATLAB via the Mex-interface. To further decrease running time, we exploited the inherent sparsity of the transition matrices $\Omega_{k,j}$, as illustrated in Fig. 7. Also, wherever possible we transferred expensive matrix-vector multiplications to the GPU, using a wrapper for MATLAB called GPUmat (Messmer et al., 2008). Segmenting all 61 B-Scans of a 3-D volume took $60\,\mathrm{s}$, with memory requirements of about $2\,\mathrm{GB}$, measured on a Core i7-2600K 3.40GHz.

6. Results

6.1. Circular Scans

Average boundary-wise results are summarized in Table 2. In general, boundaries 1 and 6 to 9 turned out to be easier to segment than boundaries 2 to 5. For boundary 1 this stems from easily detectable textures, whereas boundaries 6-9 with their regular shape profit disproportionately from regularization by the shape prior. Boundaries 2-5 on the other hand pose a harder challenge with their high variability of texture and shape. The

Table 2: Results in $\mu m\pm$SD $(3.87\mu m \,\widehat{=}\, 1px)$ for 2-D circular scans (separately for healthy eyes as well as the different degrees of glaucoma, pre-perimetric, early, moderate and advanced) and 3-D scans of healthy subjects. Numbers within brackets denote the respective data set size.

|	2-D Healthy	2-D Glaucoma				3-D Healthy
k	*All (80)*	*PPG (22)*	*PGE (22)*	*PGM (13)*	*PGA (9)*	*All (35)*
1	2.06±0.57	2.60±0.85	3.76±1.42	4.51±1.18	6.53±2.76	1.36±0.18
2	4.68±1.13	6.66±2.41	5.65±1.66	6.74±1.64	9.95±4.74	3.32±0.37
3	3.67±0.84	4.57±1.18	5.37±1.33	5.49±1.00	8.80±3.03	3.17±0.44
4	3.31±0.78	4.43±1.09	5.78±1.48	5.44±1.19	8.30±2.21	3.23±0.56
5	3.30±0.75	4.34±1.63	4.40±1.14	4.15±0.68	5.05±0.92	3.27±0.66
6	2.10±0.76	2.67±1.37	2.76±0.97	2.88±1.62	2.99±1.92	1.61±0.23
7	2.34±1.05	2.59±1.11	2.95±1.27	2.21±0.68	2.42±0.44	1.86±0.32
8	2.81±1.42	2.82±1.00	3.40±1.22	2.94±1.40	4.19±1.97	2.27±0.40
9	2.01±1.14	2.06±0.65	1.63±0.48	1.64±0.25	2.36±1.18	2.07±0.48
∅	**2.92±0.53**	**3.64±0.68**	**3.97±0.73**	**4.00±0.53**	**5.62±1.25**	**2.46±0.22**
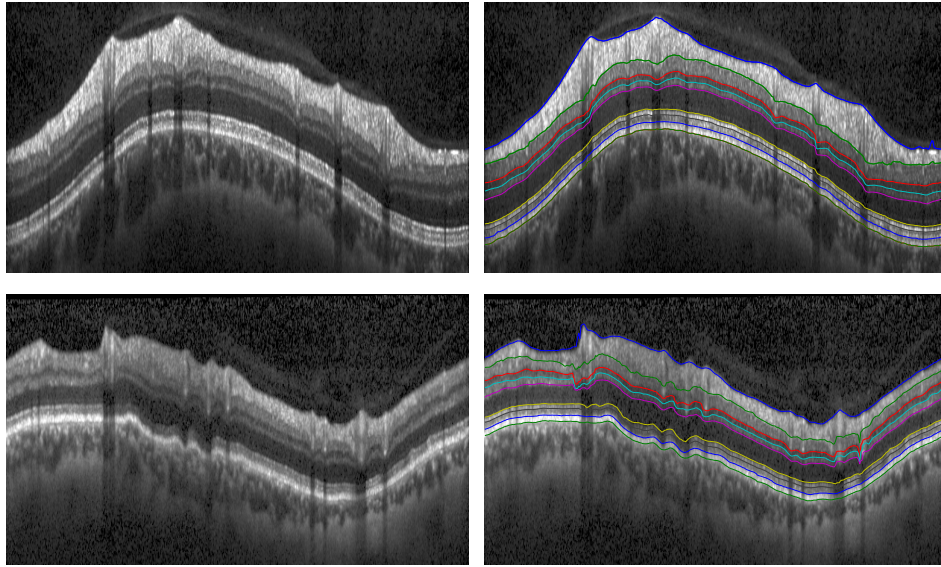
Fig. 9: Top: Segmentation ($E_{\mathrm{unsgn}} = 2.97\,\mu m$) of a non-pathological circular scan. Bottom: Segmentation ($E_{\mathrm{unsgn}} = 5.09\,\mu m$) of an advanced glaucomatous scan.

upper row in Fig. 9 shows an example close to the average segmentation performance with $E_{\mathrm{unsgn}} = 2.97\,\mu m$.

For the pathological scans segmentation performance was comparable to the healthy scans, but decreased with the progression of the disease. This happened for two reasons: Since glaucoma is known to cause a thinning of the nerve fiber layer (NFL) (Schuman et al., 1995; Bowd et al., 2001), the shape prior trained on healthy scans may encounter difficulties adapting to very abnormal glaucomatous shapes. Furthermore, we observed a reduced scan quality for glaucomatous scans, also reported by others (Ishikawa et al., 2005; Stein et al., 2006; Mayer et al., 2010), which in turn reduced the quality of the data terms. For advanced primary open-angle glaucoma, the NFL can even vanish at some locations. The appearance model for this layer, trained on healthy data, is not able to detect these extreme anomalies, which resulted in a comparatively low performance for some scans. We discuss possible modifications to overcome this problem in Section 7.

The bottom panels in Fig. 9 show an example of a PGA-type scan and its segmentation. The scan exhibits the discussed reduced scan quality. Furthermore, the segmentation proves that the shape model can generalize well to pathological shapes as well as scan artifacts.

6.1.1. Interobserver Variability

A second set of labels was created for the healthy circular B-scan data set by the second author. For training and testing we utilized the same set-up as described earlier (10-fold cross-validation, parameters as in Table 1), but used the average of both labelings for training. In Table 3 we compare the predicted segmentations with the two labelings individually and with their average. Furthermore, we report the average absolute distance between both observers, the interobserver variability.

We see, that the resulting prediction errors are well within the range of the interobserver variability. The performance using the averaged labels improves compared to the case when using only one set of labels, c.f. first column of Table 2. This suggests an increased robustness of the averaged ground truth towards scan artifacts, ambiguous image regions and labeling bias.

6.1.2. Qualitative Evaluation

A key property of our model is the *inference of full probability distributions over segmentations* q_c and q_b, instead of only modes thereof. This allows us to rate the quality of the prediction as a whole as well as indicate regions with low certainty, or classify a scan as normal or potentially

Table 3: Interobserver variability as well as prediction performance of our segmentation approach compared to ground truth of observer 1 and 2 for the set of 80 healthy circular scans ($\mu m \pm$SD ($3.87\mu m \,\widehat{=}\, 1px$)). The algorithm was trained on the averaged ground truth.

	Obs.1 vs. Obs.2	Algo. vs. Obs.1	Algo. vs. Obs.2	Algo. vs. Avg. Obs.
1	2.86±0.46	2.35±0.62	3.69±0.76	2.74±0.66
2	7.57±1.06	5.51±1.30	6.15±1.35	4.56±1.00
3	4.62±1.13	3.74±0.91	4.26±0.85	3.25±0.74
4	3.63±0.65	3.31±0.75	3.35±0.74	2.74±0.73
5	3.39±0.66	3.31±0.75	3.36±0.75	2.83±0.70
6	1.87±0.59	2.09±0.73	2.05±0.73	1.82±0.71
7	2.36±1.14	2.33±0.99	2.55±1.03	2.08±0.92
8	3.54±1.78	3.23±1.44	2.51±1.33	2.21±1.15
9	1.37±0.51	1.94±1.03	2.17±1.02	1.91±1.01
⌀	**3.47±0.37**	**3.09±0.50**	**3.34±0.52**	**2.68±0.50**

pathological. To this end, we evaluated the different terms of the objective function (12). Fig. 10 reports average function values of four terms (b-e) and compares them to the unsigned error (a). Singleton entropy (b) and mutual information (c) are the two summands of the negative entropy of q_c, given in (22). The data (d) and shape (e) terms represent the first two summands of (12), introduced in Section 3.1 and 3.2.

The shape term, which measures how much the data-driven distribution q_c differs from the shape-driven expectation $\mathbb{E}_{q_b}[\log p(c|b)]$, is highly discriminative between healthy and pathological scans. The mutual information on the other hand exhibit a good correlation with the unsigned error. It measures the dependence between variables $c_{k,j}$ and $c_{k-1,j}$. Imaging two variables $c_{k,j}$ and $c_{k-1,j}$ each having a single strong peak in $q_{c;k,j}$ and $q_{c;k-1,j}$. Their joint probability $q_{c;k\wedge k-1,j}$ will show almost no dependency. On the other hand, if we have several possibilities for each variable caused e.g. by poor data terms, then their dependency increases and thereby the mutual information. We will use these two terms in the forthcoming evaluation.

Classification. A state-of-the-art method for the clinical diagnosis of glaucoma is based upon NFL thickness, averaged for example over the whole scan or one of its four quadrants (Bowd et al., 2001; Leung et al., 2005; Chang et al., 2009; Leite et al., 2011). Estimates of the NFL thickness for all circular scans were obtained using the software of the Spectralis OCT device, version 5.6. We compared this established method against the second summand of the objective function (12), as discussed above. Using the same setup as in Bowd et al. (2001), we report sensitivities for specificities of 70%
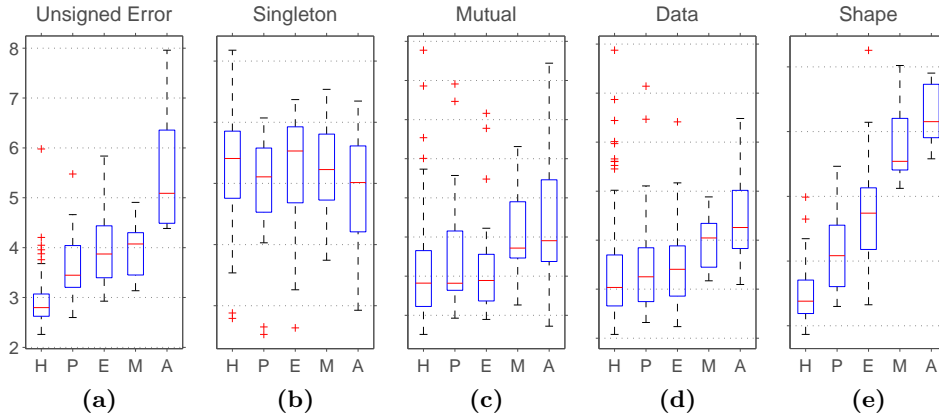
Fig. 10: Compares different terms of the objective function (b-e) with the unsigned error (a) for **h**ealthy as well as glaucomatous scans (**P**PG, PG**E**, PG**M** and PG**A**). While "Shape" is very discriminative for the glaucomatous scans, "Mutual" and "Data" correlate well with the unsigned error.

and 90%, as well as the area under the curve (AUC) of the receiver operating characteristic (ROC)[7], see Table 4. In all cases, our shape-based discriminator performs at least as good as the best thickness-based one. Especially for pre-perimetric scans, which feature only subtle structural changes, our approach improves diagnostic accuracies significantly. For this most interesting group, Fig. 11 (a) provides ROC curves of the two overall best performing NFL measures and our shape measure.

Global Quality. We obtained a global quality measure, by combining the mutual information and the shape term. Given the values for all scans, we re-weighted both terms into the ranges $[0, 1]$ and took their sum. Thereby we could establish a quality index that had a very good correlation of 0.82 with the unsigned segmentation error. See Fig. 11 (b) for a plot of all quality index/error pairs and a linear fit thereof. The estimate of this fit and the true segmentation error differs on average by only 0.51 μm. This shows that the model is able to additionally deliver the quality of its segmentation.

Local Quality. Finally, we determined a way to distinguish locally between regions of high and low model confidence. This could for example point out regions where a manual (or potentially automatic) correction is necessary.

---

[7]The AUC can be interpreted as the probability, that a random pathological scan gets assigned a higher score than a random healthy scan.

Table 4: Comparison of sensitivities for NFL-based features, measuring average thickness in different parts of the scan, and our global shape based feature. Bold numbers indicate the highest detection rate for the respective specificity and glaucoma class.

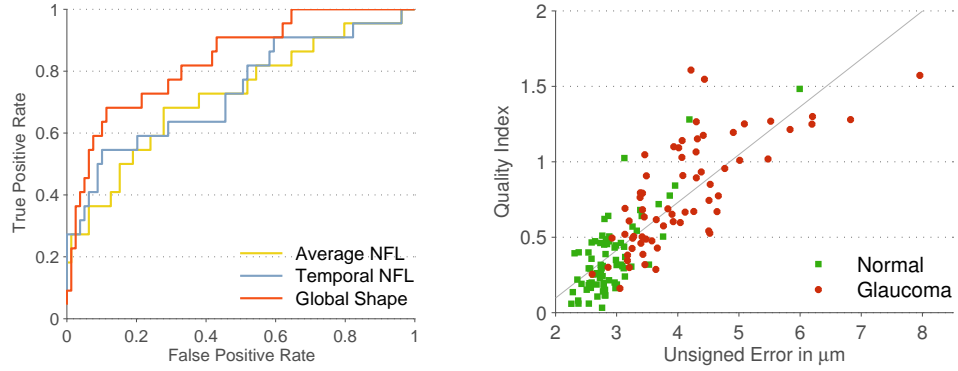| Specificity	**70%**			**90%**			**AUC**		
Type	*PPG*	*PGE*	*PGM*	*PPG*	*PGE*	*PGM*	*PPG*	*PGE*	*PGM*
Average	68.2	90.9	**100.0**	36.4	86.4	**100.0**	0.72	0.93	**1.00**
Superior	63.6	81.8	92.3	45.5	77.3	76.9	0.78	0.84	0.90
Inferior	45.5	72.7	92.3	13.6	31.8	53.8	0.69	0.77	0.89
Temporal	63.6	**95.5**	**100.0**	54.5	90.9	**100.0**	0.74	**0.95**	0.99
Nasal	36.4	63.6	92.3	18.2	45.5	61.5	0.51	0.74	0.89
Shape	**77.3**	**95.5**	**100.0**	**63.6**	**95.5**	**100.0**	**0.84**	**0.95**	**1.00**

Fig. 11: (left) ROC curves for the two overall best performing NFL-based classifiers and our shape prior based approach for pre-perimetric scans. (right) High correlation of our quality index, obtained by combining terms (c) and (e) from Fig. 10, with the actual unsigned error. On average, the estimated error (linear fit) differed by only $0.51\,\mu m$ from the true segmentation error.
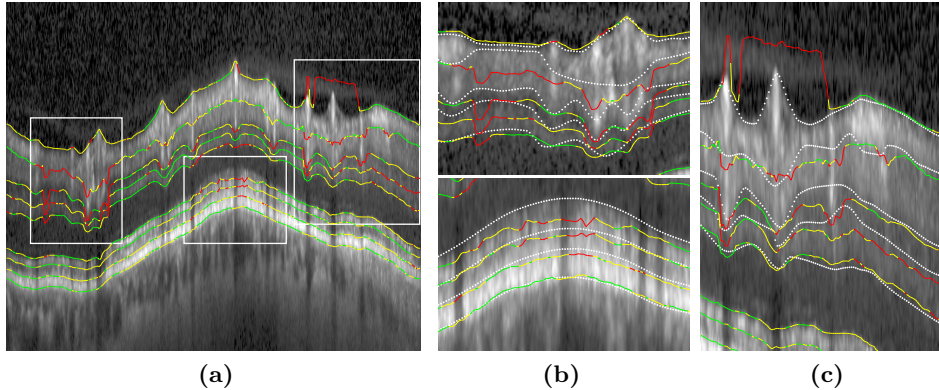
Fig. 12: (a) An advanced primary open-angle glaucoma scan and the segmentation thereof ($E_{\mathrm{unsgn}} = 6.81\,\mu m$), augmented by the local quality estimates of the model, with red representing the most uncertain class. (b) and (c) Close-ups of the three areas, the model is (correctly) most insecure about. White dotted lines represent ground truth.

To this end we examined the local correlation (i.e. on a column-wise level) of the mutual information terms with the unsigned error. We calculated its mean for instances with segmentation errors smaller than 0.5 and bigger than 2 pixels. This yielded three ranges of confidence in the quality of the segmentation. For each image we fine-tuned these ranges by dividing by max(Quality Index(CurrentImage), 1).

Fig. 12 (a) shows a PGA-type scan with annotated segmentation, whose error is $6.83\,\mu m$. The advanced thinning of the NFL and the partly blurred appearance caused the segmentation to fail in some parts of the scan. Close-ups (b) and (c) show that the model correctly identified those erroneously segmented regions. The average errors of the three categories are 4.67, 5.43 and $18.36\,\mu m$ respectively. Fig. 13 (a), on the other hand, shows a scan from a healthy eye with a segmentation error of $2.83\,\mu m$, that is accompanied by a throughout positive quality rating.

We examined the accuracy of the local quality index numerically for all scans. Fig. 13 (b) reports the average unsigned error for normal (H) and glaucomatous scans (P-A) as well as all three grades of certainty, and compares it to the average segmentation error for each data set, given as black lines. As for the global quality index, also locally the model reflects very well the distinction between correct and erroneous regions. Fig. 13 (c) shows the ratio between the three quality ratings.
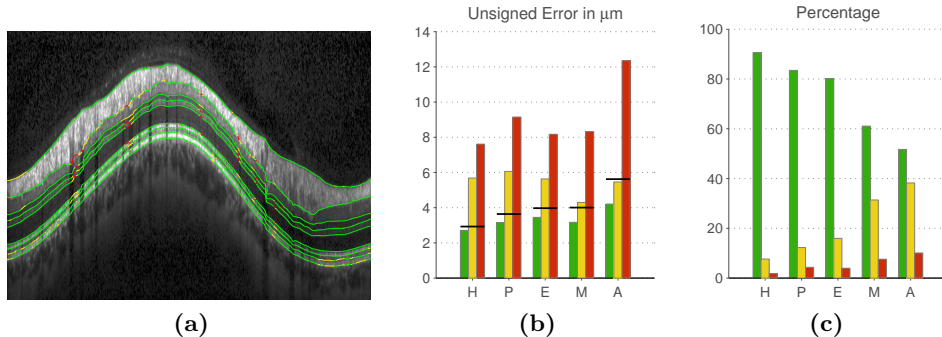
Fig. 13: (a) A healthy scan ($E_{\text{unsgn}} = 3.05\,\mu m$) accompanied by a very high model certainty. (b) Average segmentation error for healthy as well as glaucomatous scans for each quality class. Black lines donate the average segmentation error for each data set (c.f. Table 2). (c) Average partition into the three classes.

6.2. Volumetric Scans

In contrast to 2-D scans, the labeling of OCT volumes is very time consuming, hence our data set only consisted of 35 samples. Thus we were left with less data points to train a shape model of much higher dimension. Consequently, we observed a reduced ability of $p(b)$ respectively $q_b(b)$ to generalize well to unseen scans. We tackled this problem by reducing the dimensionality of $p(b)$ and by interpolating it for intermediate columns, which fixed the problem only to some extent.

We further pursued this idea and suppressed the connectivity between different B-scans, which corresponds to a block-diagonal covariance matrix Σ, where each block is obtained separately using PPCA. This significantly reduced the amount of parameters that had to be determined, and improved accuracy significantly. The last column in Table 2 reports results for all boundaries.

The average segmentation error of $2.46\,\mu m$ is significantly smaller than for circular scans, as well as the standard deviation of $0.22\,\mu m$. Reasons are smoother boundary shapes and less severe texture artifacts caused by e.g. blood vessels. Representative for the average segmentation performance, Fig. 14 shows B-scans of the same volume from four different regions, with an error of $2.53\,\mu m$ averaged over the all scans in the volume.
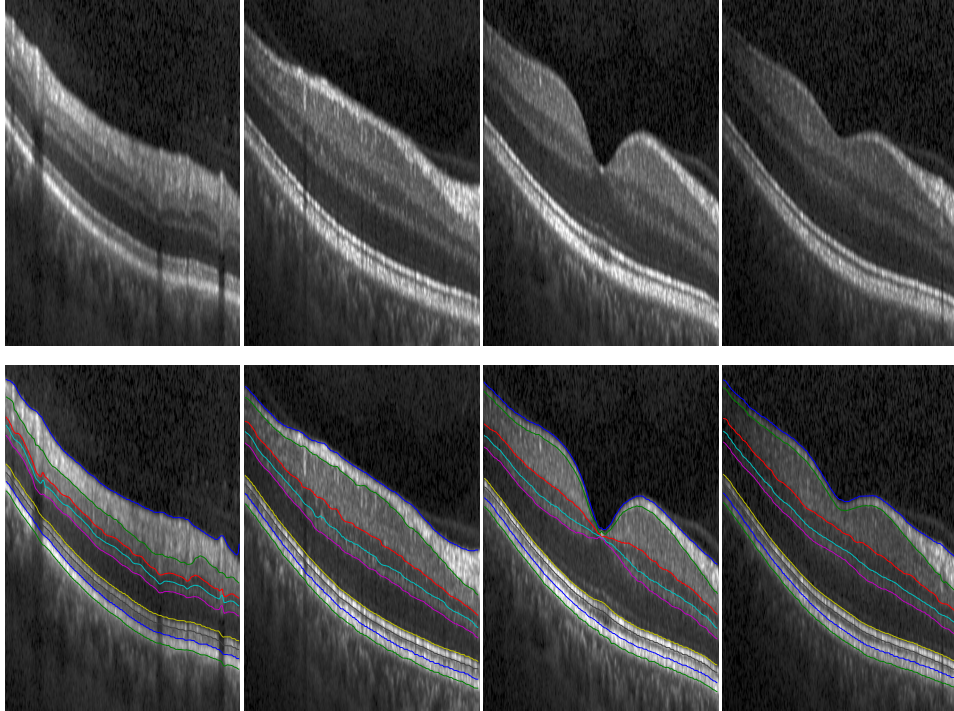
27

Fig. 14: Four segmented B-Scans from regions 2, 6, 9 and 11 of the same volume ($E_{\mathrm{unsgn}} = 2.53\,\mu m$).

7. Discussion and Conclusion

A novel probabilistic approach for the segmentation of retina layers in OCT scans was presented. It incorporates *global* shape information, which distinguishes it from most other approaches relying solely on local shape information. To obtain an approximate of the *full* posterior distribution $p(c, b|y)$, we employed variational methods, which entail efficiently solvable optimization problems. We demonstrated the applicability of our approach for a variety of different OCT scans as well as the benefit of inferring full probability distributions over segmentations rather than segmentations as point estimates.

Especially for 3-D OCT volumes, our segmentation performance was significantly better than recently reported results from approaches that use no shape information (Vermeer et al., 2011; Yang et al., 2010), local hard-constrained shape information (Garvin et al., 2009), local probabilistic shape information (Dufour et al., 2013; Song et al., 2013) or sparse global shape

28

information (Kajić et al., 2010). Taking into account, for better comparability, only publications that used data sets obtained from the same OCT device as in this publication, the following trend evolves:

- While *no shape information* lead to only mediocre results: $6.20\,\mu m$ and $5.28\,\mu m$ for healthy and moderate glaucomatous data respectively (Vermeer et al., 2011),

- adding *local shape information via hard constraints* yielded improved segmentation performance: $3.54 \pm 0.56\,\mu m$ as evaluated by Dufour et al. (2013) but comparable to the model proposed by Garvin et al. (2009).

- Additionally using *probabilistic local constraints*, Dufour et al. (2013) recently again boosted performance to $3.03 \pm 0.54\,\mu m$.

- Finally, by adding *global shape information*, we could in turn improve segmentation performance to $2.46 \pm 0.22\,\mu m$.

Although this clearly seems to support the use of global shape information for regularization, keep in mind that a concluding comparison can only be carried out using the same data set. Nevertheless, we believe that these results highlight the usefulness of global shape regularization for the segmentation of retinal layers in OCT images. Reported time requirements vary greatly, and our running time of $60\,\mathrm{s}$ is slower than the $18\,\mathrm{s}$ and $15\,\mathrm{s}$ reported by Dufour et al. (2013) and Yang et al. (2010), but faster than the remaining approaches cited above.

We also evaluated the performance for healthy and pathological 2-D circular scans and, in both cases, obtained good results. The only exception was the group of most advanced glaucomatous scans, which was caused mainly by the appearance models. Being trained on healthy data, the Gaussian distribution that models the NFL is not able to recognize instances with near-zero layer thickness. Therefore, a useful extension could be to define a mixture of Gaussians for each appearance class, adding patches centered below or above pixel (i, j), which model its surrounding but not the layer/boundary itself. Additionally, given more pathological examples especially for PGM and PGA, one could learn a pathological shape prior and let the model choose the more probable shape prior given the initialization. Future work will try to improve performance for these extreme cases and also test the approach on other pathologies like age-related macular degeneration, if training data gets available.

Finally, we investigated different ways to utilize the inferred distributions q_c and q_b. Experiments showed, that the model is quite sensitive to abnormal shapes and thus can act as a detector of glaucoma, with a higher sensitivity than established methods solely based on NFL thickness. This could relate to recent findings, that glaucoma causes a thinning of *all* inner retinal layers: NFL, GCL, IPL and (to a lesser extent) INL (Tan et al., 2008). To confirm these promising results, further studies with more patients enrolled will be needed. Another benefit of our approach is the ability to estimate the quality of the segmentation, altogether for the whole scan or for each boundary position separately. In the context of screening large patient databases, the former could be a valuable tool to minimize the effort of the physician in reassessing the results. The latter could facilitate a automatic or manual post-processing, targeted specifically at regions with a high error probability. A thorough investigation of these regions could reveal a suitable approach, and will be part of our future work.

To facilitate further research in the area of OCT segmentation and related areas, we publish our source code together with documentation on our project page: `http://graphmod.iwr.uni-heidelberg.de/Project-Details.132.0.html`.

Acknowledgments

Appendix A. Calculating $\Omega_{k,j}$ and $\omega_{1,j}$

In Sec. 3.2 we outlined the steps necessary to make explicit the expectation w.r.t q_b for the terms of $\log p(b|c)$, represented by $(\omega_{1,j})_n$ and $(\Omega_{k,j})_{m,n}$. This section will derive both terms, starting with $(\omega_{1,j})_n$:

$$
\begin{aligned}
(\omega_{1,j})_n &= \mathbb{E}_{q_b}[\log p(b_{1,j} = n | b_{\setminus j})] \\
&= \int_b q_b(b) \log \mathcal{N}\big(b_{1,j} = n; (\mu_{j|\setminus j})_1, (\Sigma_{j|\setminus j})_{1,1}\big)\, db \\
&= C - \frac{1}{2(E_{j|\setminus j})_{1,1}} \left(n^2 - 2n(\mathbb{E}_{q_b}[\mu_{j|\setminus j}]_1) + \mathbb{E}_{q_b}[((\mu_{j|\setminus j})_1)^2] \right).
\end{aligned}
$$

30

Using the definition of (17) for $(\mu_{j|\backslash j})_1$, abbreviating the k-th row of $\Sigma_{j|\backslash j} K_{j,\backslash j}$ with A_k^j and moving all terms independent of $\bar{\mu}$ to C

$$(\omega_{1,j})_n = C - \frac{1}{2(E_{j|\backslash j})_{1,1}} \left(2(n - \mu_{1,j}) A_1^j \mathbb{E}_{q_b}[b_{\backslash j}] + A_1^j (\mathbb{E}_{q_b}[b_{\backslash j} b_{\backslash j}^T] - 2\mu_{\backslash j} \mathbb{E}_{q_b}[b_{\backslash j}])(A_1^j)^T \right),$$

and finally replacing the expectations with the respective moments given in (18) yields

$$= C - \frac{1}{2(E_{j|\backslash j})_{1,1}} \left(2(n - \mu_{1,j}) A_1^j \bar{\mu}_{\backslash j} + A_1^j (\bar{\Sigma}_{\backslash j, \backslash j} + \bar{\mu}_{\backslash j} \bar{\mu}_{\backslash j}^T - 2\mu_{\backslash j} \bar{\mu}_{\backslash j}^T)(A_1^j)^T \right). \tag{A.1}$$

The terms $(\Omega_{k,j})_{m,n} = \mathbb{E}_{q_b}[\log p(c_{k,j} = n | c_{k-1,j} = m, b)]$ are the product of two Gaussians and therefore again Gaussian, modulo normalization. A lengthy derivation, using the formula for the product of two Gaussians, yields the same equation as above, with indices 1 replaced by k and a different constant C.

Appendix B. Optimization of $q_b(b)$

We showed in Sec. 4.2, that the optimization of the objective function (23) w.r.t. to the parameters of q_b can be done in closed form. This section will detail terms $\tilde{p}$ and $\tilde{P}$ introduced there, which capture the dependencies of $\Omega_{k,j}$ and $\omega_{1,j}$ on the parameters of q_b.

Appendix B.1. Derivation of $\tilde{P}$

Only considering terms in (A.1) depending on $\bar{\Sigma}$, we obtain

$$(\omega_{1,j})_n(\bar{\Sigma}) = -\frac{1}{2(E_{j|\backslash j})_{k,k}} A_k^j \bar{\Sigma}_{\backslash j, \backslash j} (A_k^j)^T$$

and similar for $(\Omega_{k,j})_{m,n}(\bar{\Sigma})$. Since A_k^j is of dimension $1 \times N_b M - N_b$ we introduce an expanded version $\tilde{A}_k^j \in \mathbb{R}^{N_b M}$ padded with zero entries, such that $\tilde{A}_k^j \Sigma (\tilde{A}_k^j)^T = A_k^j \Sigma_{\backslash j, \backslash j} (A_k^j)^T$. Note that $(\Omega_{k,j})_{m,n}(\bar{\Sigma})$ and $(\omega_{1,j})_n(\bar{\Sigma})$ are independent of m and n and thus have identical entries for all (m, n), and therefore $\mathbb{E}_p[c] = c$. Using

$b^T B b = \langle b b^T, B \rangle$, we obtain for (19)

$$
-\sum_{j=1}^{M} \Big((q_{c;1,j})^T \omega_{1,j}(\bar{\Sigma}) + \sum_{k=2}^{N_b} \big\langle (q_{c;k \wedge k-1,j})^T, \Omega_{k,j}(\bar{\Sigma}) \big\rangle \Big)
$$

$$
= \frac{1}{2} \Big(\sum_{j=1}^{M} \| q_{c;1,j} \|_1 \big\langle \frac{1}{(E_{j|\backslash j})_{1,1}} (\tilde{A}_1^j)^T \tilde{A}_1^j, \bar{\Sigma} \big\rangle
$$

$$
+ \sum_{k=2}^{N_b} \| q_{c;k \wedge k-1,j} \|_1 \big\langle \frac{1}{(E_{j|\backslash j})_{k,k}} (\tilde{A}_k^j)^T \tilde{A}_k^j, \bar{\Sigma} \big\rangle \Big)
$$

$$
= \frac{1}{2} \big\langle \sum_{j=1}^{M} \sum_{k=1}^{N_b} \tilde{P}_{k,j}, \bar{\Sigma} \big\rangle = \frac{1}{2} \langle \tilde{P}, \bar{\Sigma} \rangle
$$

Appendix B.2. Derivation of $\tilde{p}$

Again, we begin by singling out terms of (A.1), dependent on $\bar{\mu}$

$$
(\omega_{1,j})_n(\bar{\mu}) = -\frac{1}{2(E_{j|\backslash j})_{1,1}} \big(2n A_1^j \bar{\mu}_{\backslash j} - 2\mu_{1,j} A_1^j \bar{\mu}_{\backslash j} + A_1^j (\bar{\mu}_{\backslash j} \bar{\mu}_{\backslash j}^T - 2\mu_{\backslash j} \bar{\mu}_{\backslash j}^T)(A_1^j)^T \big)
$$

and accordingly for $(\Omega_{k,j})_{m,n}(\bar{\mu})$. The first term is dependent on n and thereby on q_c, whereas the remaining terms are again independent on n and q_c sums out. Using $\tilde{A}_1^j$ as introduced above, we plug $\omega_{1,j}(\bar{\mu})$ and $\Omega_{k,j}(\bar{\mu})$ into (19)

$$
-\sum_{j=1}^{M} \Big((q_{c;1,j})^T \omega_{1,j}(\bar{\mu}) + \sum_{k=2}^{N_b} \big\langle (q_{c;k \wedge k-1,j})^T, \Omega_{k,j}(\bar{\mu}) \big\rangle \Big)
$$

$$
= \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{N_b} \frac{1}{(E_{j|\backslash j})_{k,k}} \Big(2 \big(\mathbb{E}_{q_c}[c_{k,j}] - \mu_{k,j} \big) \tilde{A}_k^j \bar{\mu} + \big\langle (\tilde{A}_k^j)^T \tilde{A}_k^j, \bar{\mu}(\bar{\mu} - 2\mu)^T \big\rangle \Big)
$$

$$
= \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{N_b} 2 \tilde{p}_{k,j}^T \bar{\mu} + \big\langle \tilde{P}_{k,j}, \bar{\mu}(\bar{\mu} - 2\mu)^T \big\rangle
$$

$$
= \tilde{p}^T \bar{\mu} + \frac{1}{2} \langle \tilde{P}, \bar{\mu}(\bar{\mu} - 2\mu)^T \rangle,
$$

where $\tilde{P}$ was defined in the previous section.

References

Ahlers, C., Simader, C., Geitzenauer, W., Stock, G., Stetson, P., Dastmalchi, S., Schmidt-Erfurth, U., 2008. Automatic segmentation in three-dimensional analysis of fibrovascular pigmentepithelial detachment using high-definition optical coherence tomography. Brit. J. Ophthalmol. 92, 197–203.

Antony, B.J., Abràmoff, M.D., Lee, K., Sonkova, P., Gupta, P., Kwon, Y., Niemeijer, M., Hu, Z., Garvin, M.K., 2010. Automated 3d segmentation of intraretinal layers from optic nerve head optical coherence tomography images, in: SPIE Medical Imaging, International Society for Optics and Photonics. p. 76260U.

Attias, H., 2000. A Variational Bayesian Framework for Graphical Models, in: Adv. Neural Inf. Proc. Systems (NIPS), pp. 209–215.

Baroni, M., Fortunato, P., Torre, A.L., 2007. Towards quantitative analysis of retinal features in optical coherence tomography. Med. Eng. Phys. 29, 432–441.

Bishop, C.M., et al., 2006. Pattern recognition and machine learning. Springer New York.

de Boer, J.F., Cense, B., Park, B.H., Pierce, M.C., Tearney, G.J., Bouma, B.E., 2003. Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. Opt. Lett. 28, 2067–2069.

Bowd, C., Zangwill, L.M., Berry, C.C., Blumenthal, E.Z., Vasile, C., Sanchez-Galeana, C., Bosworth, C.F., Sample, P.A., Weinreb, R.N., 2001. Detecting early glaucoma by assessment of retinal nerve fiber layer thickness and visual function. Invest. Ophthalmol. Vis. Sci. 42, 1993–2003.

Chang, R.T., Knight, O., Feuer, W.J., Budenz, D.L., 2009. Sensitivity and specificity of time-domain versus spectral-domain optical coherence tomography in diagnosing early to moderate glaucoma. Ophthalmology 116, 2294–2299.

Cover, T.M., Thomas, J.A., 2006. Elements of information theory. Wiley-Interscience.

Drexler, W., Fujimoto, J., 2008. State-of-the-art retinal optical coherence tomography. Prog. Retin. Eye. Res. 27, 45–88.

Dufour, P., Ceklic, L., Abdillahi, H., Schroder, S., De Dzanet, S., Wolf-Schnurrbusch, U., Kowal, J., 2013. Graph-based multi-surface segmentation of OCT data using trained hard and soft constraints. IEEE Trans. Med. Imaging 32, 531–543.

Fernández, D.C., Salinas, H.M., Puliafito, C.A., 2005. Automated detection of retinal layer structures on optical coherence tomography images. Opt. Express 13, 10200–10216.

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441.

Garvin, M.K., Abràmoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M., 2009. Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. IEEE Trans. Med. Imaging 28, 1436–1447.

Ishikawa, H., Stein, D.M., Wollstein, G., Beaton, S., Fujimoto, J.G., Schuman, J.S., 2005. Macular segmentation with optical coherence tomography. Invest. Ophthalmol. Vis. Sci. 46, 2012–2017.

Kajić, V., Povazay, B., Hermann, B., Hofer, B., Marshall, D., Rosin, P.L., Drexler, W., 2010. Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis. Opt. Express 18, 14730–14744.

Leite, M.T., Rao, H.L., Zangwill, L.M., Weinreb, R.N., Medeiros, F.A., 2011. Comparison of the diagnostic accuracies of the Spectralis, Cirrus, and RTVue optical coherence tomography devices in glaucoma. Ophthalmology 118, 1334–1339.

Leung, C.K.s., Chong, K.K.L., Chan, W.m., Yiu, C.K.F., Tso, M.y., Woo, J., Tsang, M.K., Tse, K.k., Yung, W.h., 2005. Comparative study of retinal nerve fiber layer measurement by StratusOCT and GDx VCC, II: structure/function regression analysis in glaucoma. Invest. Ophthalmol. Vis. Sci. 46, 3702–3711.

Mayer, M.A., Hornegger, J., Mardin, C.Y., Tornow, R.P., 2010. Retinal nerve fiber layer segmentation on FD-OCT scans of normal subjects and glaucoma patients. Biomed. Opt. Express 1, 1358–1383.

McGrory, C., Titterington, D., Reeves, R., Pettitt, A., 2009. Variational Bayes for estimating the parameters of a hidden Potts model. Stat. Comput. 19, 329–340.

Messmer, P., Mullowney, P.J., Granger, B.E., 2008. Gpulib: Gpu computing in high-level languages. Comput. Sci. Eng. , 70–73.

Mishra, A., Wong, A., Bizheva, K., Clausi, D.A., 2009. Intra-retinal layer segmentation in optical coherence tomography images. Opt. Express 17, 23719–23728.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. MIT Press.

Rathke, F., Schmidt, S., Schnörr, C., 2011. Order preserving and shape prior constrained intra-retinal layer segmentation in optical coherence tomography., in: Fichtinger, G., Martel, A.L., Peters, T.M. (Eds.), MICCAI (3), Springer. pp. 370–377.

Schuman, J., Puliafito, C., Fujimoto, J., 2004. Optical Coherence Tomography of Ocular Diseases. Slack Incorporated.

Schuman, J.S., Hee, M.R., Puliafito, C.A., Wong, C., Pedut-Kloizman, T., Lin, C.P., Hertzmark, E., Izatt, J.A., Swanson, E.A., Fujimoto, J.G., et al., 1995. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. Arch. Ophthalmol. 113, 586.

Song, Q., Bai, J., Garvin, M., Sonka, M., Buatti, J., Wu, X., et al., 2013. Optimal multiple surface segmentation with shape and context priors. IEEE Trans. Med. Imaging 32, 376–386.

Stein, D., Ishikawa, H., Hariprasad, R., Wollstein, G., Noecker, R., Fujimoto, J., Schuman, J., 2006. A new quality assessment parameter for optical coherence tomography. Br. J. Ophthalmol. 90, 186–190.

Tan, O., Li, G., Lu, A.T.H., Varma, R., Huang, D., et al., 2008. Mapping of macular substructures with optical coherence tomography for glaucoma diagnosis. Ophthalmology 115, 949.

Tipping, M.E., Bishop, C.M., 1999. Probabilistic principal component analysis. J. R. Stat. Soc. 61, pp. 611–622.

Vermeer, K.A., van der Schoot, J., Lemij, H.G., de Boer, J.F., 2011. Automated segmentation by pixel classification of retinal layers in ophthalmic OCT images. Biomed. Opt. Express 2, 1743–1756.

Wainwright, M.J., Jordan, M.I., 2008. Graphical models, exponential families, and variational inference. Found. Trends Mach. Learn. 1, 1–305.

Wojtkowski, M., Leitgeb, R., Kowalczyk, A., Bajraszewski, T., Fercher, A.F., 2002. In vivo human retinal imaging by fourier domain optical coherence tomography. J. Biomed. Opt. 7, 457–463.

Yang, Q., Reisman, C.A., Wang, Z., Fukuma, Y., Hangai, M., Yoshimura, N., Tomidokoro, A., Araie, M., Raza, A.S., Hood, D.C., Chan, K., 2010. Automated layer segmentation of macular oct images using dual-scale gradient information. Opt. Express 18, 21293–21307.

Yazdanpanah, A., Hamarneh, G., 0002, B.S., Sarunic, M., 2009. Intra-retinal layer segmentation in optical coherence tomography using an active contour approach., in: Yang, G.Z., Hawkes, D.J., Rueckert, D., Noble, J. Alison and, C.J.T. (Eds.), MICCAI (1), Springer. pp. 649–656.

Yazdanpanah, A., Hamarneh, G., Smith, B.R., Sarunic, M.V., 2011. Segmentation of intra-retinal layers from optical coherence tomography images using an active contour approach. IEEE Trans. Med. Imaging 30, 484–496.

Zysk, A.M., Nguyen, F.T., Oldenburg, A.L., Marks, D.L., Boppart, S.A., 2007. Optical coherence tomography: a review of clinical development from bench to bedside. J. Biomed. Opt. 12, 1–21.