Pathology Hinting as the Combination of Automatic Segmentation with a Statistical Shape Model

Pascal A. Dufour[1,2], Hannan Abdillahi[3], Lala Ceklic[3], Ute Wolf-Schnurrbusch[2,3], Jens Kowal[1,2]

[1] ARTORG Center for Biomedical Engineering Research, Ophthalmic Technologies, University of Bern, 3010 Bern, Switzerland
pascal.dufour@artorg.unibe.ch
[2] University Hospital Bern, Ophthalmic Department, 3010 Bern, Switzerland
[3] Bern Photographic Reading Center, 3010 Bern, Switzerland

Abstract. With improvements in acquisition speed and quality, the amount of medical image data to be screened by clinicians is starting to become challenging in the daily clinical practice. To quickly visualize and find abnormalities in medical images, we propose a new method combining segmentation algorithms with statistical shape models. A statistical shape model built from a healthy population will have a close fit in healthy regions. The model will however not fit to morphological abnormalities often present in the areas of pathologies. Using the residual fitting error of the statistical shape model, pathologies can be visualized very quickly. This idea is applied to finding drusen in the retinal pigment epithelium (RPE) of optical coherence tomography (OCT) volumes. A segmentation technique able to accurately segment drusen in patients with age-related macular degeneration (AMD) is applied. The segmentation is then analyzed with a statistical shape model to visualize potentially pathological areas. An extensive evaluation is performed to validate the segmentation algorithm, as well as the quality and sensitivity of the hinting system. Most of the drusen with a height of 85.5μm were detected, and all drusen at least 93.6μm high were detected.

Keywords: pathology hinting, statistical shape model, multi-surface segmentation, optical coherence tomography.

1 Introduction

With the recent advances in OCT, the quality and acquisition speed has increased dramatically. This has revolutionized ophthalmology, as it allows the fast and non-invasive imaging of various structures of the human eye. Today, the imaging of the retinal layers with OCT is standard clinical practice. As with other image modalities in medicine, the faster acquisition speed and increased image resolution also increases the amount of work to be done by the clinician to screen the datasets and state a diagnosis. The current clinical practice of analyzing an OCT volume is visual inspection of each individual B-scan. As such an

Presented at the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012, pp. 599-606, Nice, France, October 2012

DOI 10.1007/978-3-642-33454-2_74.
Link: http://link.springer.com/chapter/10.1007%2F978-3-642-33454-2_74

inspection takes minutes instead of seconds, in practice clinicians only quickly scroll through the B-scans and abnormalities are often missed. This calls for new analysis and visualization tools.

Segmentation of OCT datasets has become an important tool to quantitatively analyze retinal layers. Recent advances in segmentation techniques [1], [2] make it possible to segment the drusen of AMD patients with high accuracy.

The main contribution of this work is the use of a statistical model as a tool to quickly visualize possibly pathological regions in a dataset. For this we built a statistical shape model [3] from segmentations of fovea-centered OCT datasets of healthy patients. To analyze a new patient's OCT volume, the dataset is first segmented and the statistical shape model is then fitted to that segmentation. The error of the fitting is then visualized as a top down rendering (en-face map).

Key contributions: We propose the combination of a graph-based segmentation algorithm with a statistical shape model to build a quick pathology hinting system. Our key contributions are twofold:

- A graph-based method for highly accurate drusen segmentation
- A hinting system able to quickly visualize morphological abnormalities.

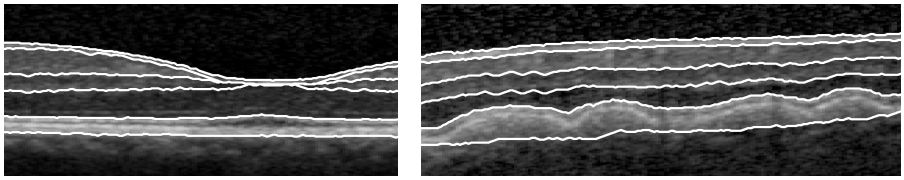
2 Methods

2.1 Segmentation

The segmentation is based on the optimal net surface problems introduced by Wu et al. [4] and extended to multiple surfaces by Li et al. [5]. Garvin et al. [1] applied this algorithm to the segmentation of retinal layers in OCT volumes. Additional soft constraints, proposed by Song et al. [2], were applied. These constraints allow us to add costs for the rigidity of a surface and costs for the distance between two surfaces. Prior information can therefore be better incorporated into the graph, resulting in an improved segmentation [2].

Furthermore, the soft constraints also enable more possibilities when segmenting pathologies such as drusen. Drusen are an accumulation of extracellular material in the Bruch's membrane (BM) of the retina [6]. The result is a displacement of the cell layers above it. In the OCT, this is most visible in the displacement of the inner and outer photoreceptor segments (IS/OS). However, the boundary between the choroid and the Bruch's membrane is left mostly intact. This makes it easy to segment the lower BM boundary with this method. See Fig.1(b) for an example of large drusen in an OCT. By adding a strong soft constraint on the rigidity of the surface of the lower BM boundary, it is possible to get a smooth and accurate segmentation.

The next step is the segmentation of the upper IS/OS boundary. Because the displacement from the drusen can be quite large, only a weak soft constraint on the rigidity is added. Additionally, a medium soft constraint on the expected distance to the lower BM boundary is added. This favors a segmentation that is close to the expected IS/OS position of a healthy RPE, but still allows the segmentation of the displaced IS/OS when a druse is present. Fig.1 shows two examples of the applied segmentation.



(a) (b)

Fig. 1. Segmented OCT B-scans: (a) Healthy retina around the fovea, (b) AMD patient with drusen.

2.2 Statistical Shape Model

Statistical shape models can capture the natural shape variation from training shapes [3]. The idea is to allow the statistical shape model to generate the whole range of healthy shapes, without being able to generate pathological shapes. By fitting the statistical shape model to a new segmentation, we will then be able to detect if and where the new segmentation cannot be accurately represented by this model.

Model Building. To build a statistical shape model, the training shapes first have to be brought into a common coordinate frame. Because we do not use 3D positions to build the model, but only the differences between the segmented layers, the rotation of the shapes can be omitted and a translation is enough to align all training shapes. The fovea was used as an anchor, which was detected by finding the lowest point on the segmentation of the inner limiting membrane (ILM). All shapes were translated so the fovea is at the coordinate origin $(0, 0)$. Each position $(i \cdot d_x, j \cdot d_y)$ relative to the fovea becomes a landmark position, where i and j are integers with predefined boundaries and d_x and d_y are the sampling spacing in x- and y-direction. This simplifies the landmarking process, as the landmark positions form a simple grid around the fovea and no anatomical landmarks are required. The statistical shape model was then built by computing the mean shape and covariance matrix, and principal component analysis (PCA) was applied for dimensionality reduction [3]. Automatic segmentations of 28 OCT volumes were used as training shapes. Fig.2 illustrates all steps required to build the statistical shape model.

Model Fitting. Given a new segmentation of a new OCT volume, the statistical shape model can be deformed so that it minimizes the distance to the shape vector of that new segmentation. See [3] for an iterative approach to deform the statistical shape model to a new shape vector.

During the fitting, the deformation of the model is limited so it is able to only represent about 99% of the variation encountered in the training datasets. This ensures that the generated shape is similar to the shapes seen in the training

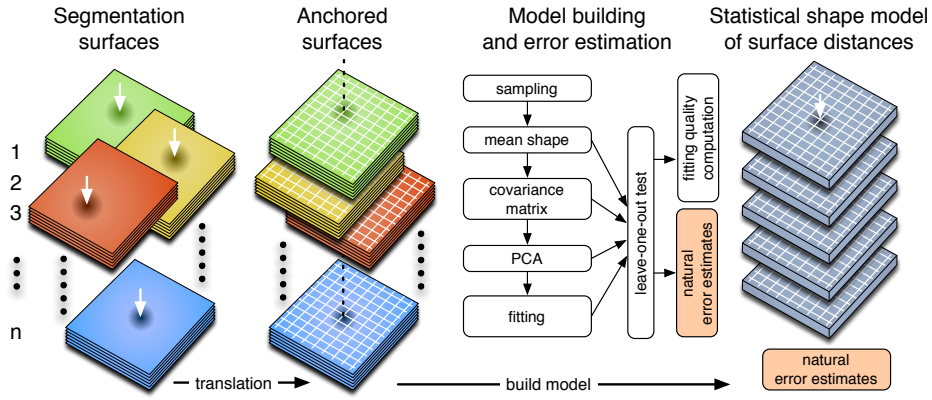


Fig. 2. Complete statistical shape model building procedure including leave-one-out test for the natural error estimate computation.

datasets and cannot deform to an extreme shape. See Fig.3 for the necessary steps to fit the model to a new segmentation.

2.3 Pathology Hinting

For every landmark, the residual absolute fitting error between the deformed model and the new segmentation serves as a measure of pathology at that position. The hinting can be further improved by normalizing that error.

When building the statistical shape model, we perform a leave-one-out test to estimate the natural residual errors. For every landmark position, the distribution of the mean unsigned error ε_μ and its variance ε_σ^2 is computed. Assuming a normal distribution of that error, we now know that 68% of the errors in the training set are within the interval $\varepsilon_\mu \pm \varepsilon_\sigma$, 95% are within $\varepsilon_\mu \pm 2\varepsilon_\sigma$, and so on.

Let's say we now fit the model to a new segmentation and encounter a residual fitting error ε at a specific landmark, for example $\varepsilon = \varepsilon_\mu + 3\varepsilon_\sigma$. We know that at that landmark, only 0.27% of the measured errors in a healthy dataset are at least this large and that therefore the landmark is highly abnormal. We formulate the measure of abnormality ψ at a specific landmark position as the actual residual fitting error normalized to the natural residual fitting error:

$$\psi^2 = \frac{(\varepsilon - \varepsilon_\mu)^2}{\varepsilon_\sigma^2} \tag{1}$$

Intuitively, ψ measures in what interval $\varepsilon_\mu \pm \psi \cdot \varepsilon_\sigma$ the error is. The role of this error normalization is also illustrated in Fig.3.

To present the computed measurements to the diagnostician, an en-face map is built for each layer thickness. Every landmark becomes a pixel in this en-face image. A color transfer function is used to map the values from (1) to a color value. Fig.4 shows en-face maps for two dataset with drusen and a healthy one.

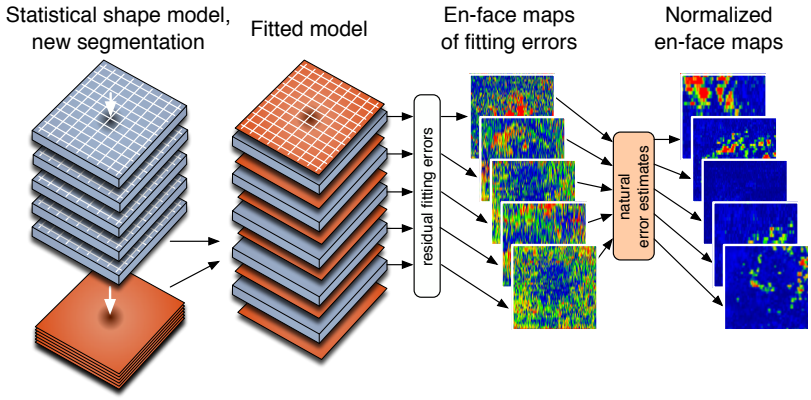
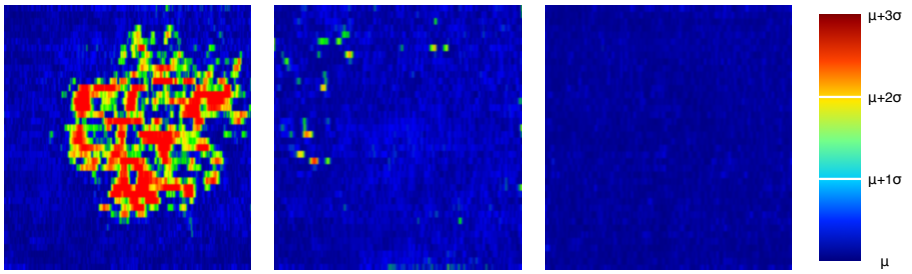


Fig. 3. Statistical shape model fitting: The statistical shape model (grey) is deformed to fit to the surfaces of a new segmentation (red) and the residual error is normalized to build the hinting en-face map for each surface.



(a) (b) (c) (d)

Fig. 4. Thickness hinting en-face maps of the RPE thickness: (a) AMD patient with large drusen, (b) small drusen, (c) healthy RPE, (d) used color transfer function.

The error can of course also be projected back into the OCT volume. Fig.5 shows an example B-scans containing drusen where the cell layers between the segmentation surfaces are colored by the hinting system. The drusen are clearly made visible by the hinting. Note also the abnormally thin cell layer above the drusen to the right of the fovea.

We also implemented a statistical shape model using the full 3D positions of the segmented surfaces. This model is less accurate in detecting small thickness abnormalities, but can reliably visualize abnormalities in the actual shape of the segmented surfaces. Fig.6 shows an example of a patient with an abnormally shaped retina.

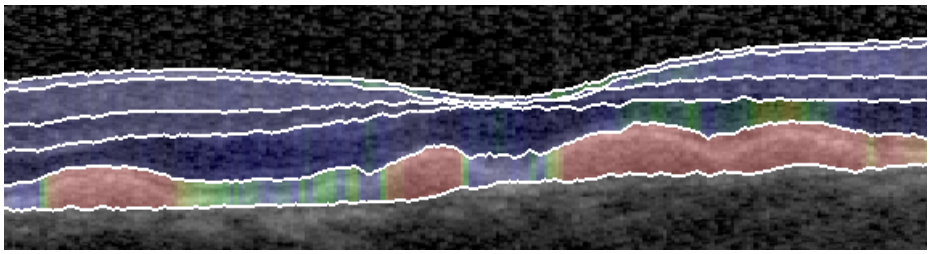
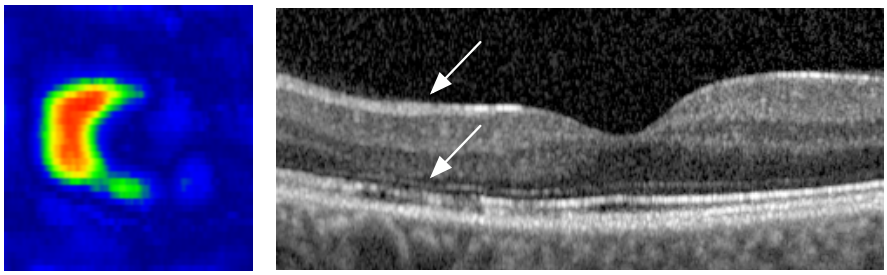


Fig. 5. B-scan with the residual fitting error projected back into the OCT volume.



(a) (b)

Fig. 6. (a) Shape deformation analysis of the ILM of a patient with a degenerated outer nuclear layer (lower arrow), leading to an abnormally shaped ILM (upper arrow), visible in the corresponding hinting en-face map in (b).

3 Experiments and Results

3.1 Segmentation Evaluation

The accuracy of the hinting system depends greatly on the accuracy of the segmentation and a thorough evaluation was therefore performed. The segmentation algorithm needs to be able to segment healthy datasets as well as datasets with drusen. We therefore evaluated the algorithm on both healthy and AMD datasets. 20 OCT volumes, each having five randomly chosen B-scans manually segmented by two experts, were used in the evaluation. The average manual segmentation from both observers was compared to the result of the automatic segmentation algorithm. As we evaluated the hinting system on drusen, the evaluation of the segmentation focused on the two layers relevant for drusen segmentation: the lower BM and upper IS/OS boundaries.

The upper IS/OS boundary was segmented with a mean unsigned error of $1.69 \pm 1.61\mu$m. The lower BM boundary was segmented with a mean unsigned error of $2.75 \pm 2.49\mu$m.

To evaluate the accuracy of the segmentation on actual drusen, 20 datasets containing drusen were used. In every dataset, the lower BM and upper IS/OS boundary of every druse up to a height of 141μm was segmented manually. The

healthy parts of the RPE were not segmented and not used in the evaluation. This allowed us to evaluate the accuracy of the algorithm only on the positions where drusen were present. For the upper IS/OS boundary, the mean unsigned error over all drusen was $5.66 \pm 10.00\mu$m. The mean unsigned error of the lower BM boundary was $4.67 \pm 7.42\mu$m over all drusen.

While the error for drusen is larger than the error for healthy tissue, it is still relatively small considering the axial pixel resolution of 3.9μm. The result is also comparable to the results by Garvin et al. [1], who reported an unsigned mean error of $3.30 \pm 1.60\mu$m for the upper IS/OS boundary on healthy datasets. Furthermore, the larger segmentation errors were observed for large drusen that did not have a clear gradient. However, in this case, a larger error is not a problem as the segmentation is still very much outside the range of segmentations of healthy eyes and the drusen are clearly visible in the hinting system.

3.2 Hinting Evaluation

To evaluate the statistical shape model, a leave-one-out test was performed. The mean unsigned error over all landmark points over all datasets was $5.71 \pm 7.86\mu$m.

To evaluate the actual hinting, we implemented a screening system where experts had to decide purely from the en-face map of the RPE thickness if a dataset contained drusen. 20 datasets with drusen were mixed with 20 datasets of healthy eyes. The datasets were presented in random order to each experts and they had to decide for each one whether it contained drusen. One dataset with only a few small drusen was missed by one of the two experts, all other datasets containing drusen were correctly identified by both experts. The combined false positive ratio of the two expert was 10% (4 out of 40). As most datasets contained very large drusen, this test was very easy for the experts. Nevertheless, it shows that the hinting system works well on real data.

As we wanted to know how sensitive the hinting system is, we implemented an evaluation procedure with artificially created small drusen. 50 datasets of 50 healthy subjects were used. When a dataset was presented to the expert, a single druse was created with a probability of 50%. The position of the druse was randomly chosen, as was the height displacement, ranging from just one pixel to 15 pixels in the OCT volume. With an axial resolution of 3.9μm, the resulting height displacement was in the range of 3.9 to 58.5μm. Table1 shows the results of the evaluation. Most drusen with a height of 85.5μm were already detected, and all drusen with a height of at least 93.6μm were detected. The experts spent an average duration of only 7.3 ± 4.8 seconds looking at the en-face map before making a decision. This included loading the segmentation, fitting of the statistical shape model and visualization. As expected, experts had difficulties identifying the smaller drusen, as they often were not sure if a slightly abnormal looking area was healthy and had to guess. In a real clinical setting, the OCT would of course be made visible as well. A click at an abnormal looking area in the hinting map could for example open the OCT volume at that position for the diagnostician to check.

Table 1. Combined results of the two experts identifying drusen from only the hinting en-face map. The column 'AMD' is the result of the test with patients with drusen.

Druse height (μm)	no druse	70.2	74.1	78.0	81.9	85.5	89.70	93.6-136.5	AMD
Identified as drusen	23	0	3	2	3	8	7	49	39
Identified as healthy	129	5	1	4	3	2	1	0	1
Correctly identified (%)	84.9	0	75	33.3	50	80	87.5	100	97.5

4 Conclusion

In this work a new method to quickly detect morphological abnormalities in medical images is proposed. By combining automatic segmentation with a statistical shape model, we were able to visualize and detect most drusen in OCT volumes. With an average time of only 7.3 ± 4.8 seconds spent on each dataset, experts were able to very quickly screen the datasets. The proposed method is completely complementary to the current clinical practice of visual inspection of each individual B-scan of the OCT volume. The accuracy and speed make this method a valuable tool for both large-scale screening systems and the daily clinical practice. Furthermore, it can be applied in virtually any field where automatic segmentation and statistical modeling of anatomy is possible. We are currently working on extending the method to include the analysis of texture information by using a statistical model of appearance.

References

1. Garvin, M.K., Abràmoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M.: Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. IEEE Transactions on Medical Imaging **28**(9) (2009) 1436–1447
2. Song, Q., Wu, X., Liu, Y., Garvin, M., Sonka, M.: Simultaneous searching of globally optimal interacting surfaces with shape priors. IEEE International Conference on Computer Vision and Pattern Recognition (2010)
3. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding **61**(1) (1995) 38–59
4. Wu, X., Chen, D.Z.: Optimal net surface problems with applications. In Widmayer, P., Eidenbenz, S., Triguero, F., Morales, R., Conejo, R., Hennessy, M., eds.: ICALP. Volume 2380 of LNCS. Springer, Heidelberg (2002) 1029–1042
5. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images - a graph-theoretic approach. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(1) (2006) 119–134
6. Ryan, S.J., Wilkinson, C.P.: Retina. Mosby (2005)