Near real-time classification of optical coherence tomography data using principal components fed linear discriminant analysis

Florian Bazant-Hegemark
Nicholas Stone
Cranfield University at Silsoe
Cranfield Health
Bedfordshire MK45 4DT
United Kingdom
   and
Gloucestershire Royal Hospital
Biophotonics Research Group
Great Western Road
Gloucester GL1 3NN
United Kingdom
E-mail: n.stone@medical-research-centre.com

Abstract. An optical coherence tomography (OCT) prediction algorithm is designed and tested on a data set of sample images (taken from vegetables and porcine tissues) to demonstrate proof of concept. Preprocessing and classification of data are fully automated, at a rate of 60,000 A-scans/min on a standard computer and can be considered to deliver in near real-time. A data set consisting of nine groups was classified correctly in 82% of cases after cross-validation. Sets of fewer groups reach higher rates. The algorithm is able to distinguish groups with strong visual similarity among several groups of varying resemblance. Surface recognition and normalizing to the surface are essential for this approach. The mean divided by the standard deviation is a suitable descriptor for reducing a set of surface normalized A-scans. The method enables grouping of separate A-scans and is therefore straightforward to apply on 3-D data. OCT data can reliably be classified using principal component analysis combined with linear discriminant analysis. It remains to be shown whether this algorithm fails in the clinical setting, where interpatient variation can be greater than the deviations that are investigated as a disease marker. © *2008 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.2931079]

Keywords: classification model; optical coherence tomography; automated discrimination; medical imaging; preprocessing; image processing.

Paper 07241RR received Jul. 5, 2007; revised manuscript received Jan. 8, 2008; accepted for publication Jan. 15, 2008; published online May 22, 2008.

1 Introduction

1.1 Optical Coherence Tomography

Optical coherence tomography (OCT) is a relatively novel imaging modality first described by Huang et al.,[1] which is becoming established as a technique for noninvasive, real-time subsurface imaging at resolutions of 2 to 10 μm. Nondestructive probing of this type is of interest specifically in medical imaging[2,3] and also in other fields such as art conservation,[4,5] quality assurance, or homeland security (e.g., finger prints). Compared with other emerging imaging modalities, OCT provides high resolution and imaging speed for detailed structural information—in vast data sets. We investigate automated classification, which is quicker than visually assessing images and likely to be more reproducible.

1.1.1 Principal component analysis and linear discriminant analysis

Principal component analysis (PCA) and linear discriminant analysis (LDA) are mathematical methods that, on a statistical basis, enable data reduction and clustering into groups.[6] There are papers on applying PCA for OCT related analysis. Weissman et al.[7] used PCA for comparing the agreement of their surface recognition algorithm with the line drawn by human operators. Qi et al.[8] used PCA for preprocessing. They were reducing six statistical measures for texture features to two principal components. However, using this approach they reported no benefit over using the 6-D data.

Huang and Chen[9] also applied PCA and LDA for classification, but not on the OCT image itself. Rather, they were using a set of 25 morphological parameters to characterize ophthalmologic pathologies. These parameters were thickness and size measures that could be extracted from the OCT image. In nonophthalmic OCT imaging, such defined layers are usually not observed.

Zysk and Boppart[10] reported three methods for classifying three relevant groups (adipose, cancerous, stroma) in A-scans of breast cancer biopsies. The first was taking the Fourier transform (FT) of an A-scan and using its difference from a reference (the mean of a training set) for the analysis. The second was periodicity analysis, which looks at the distance between high-intensity subsurface peaks to the surface peak. The third was a combination of both. The periodicity analysis seems to benefit from the "bubbly" structural pattern in OCT images of adipose breast tissue, which, however, is not present in every type of tissue. Part of this method, taking the FT of an A-scan, we compare with our method. (We did not implement the periodicity technique.) This enabled us to test the Fourier domain technique on sets containing more than three groups.
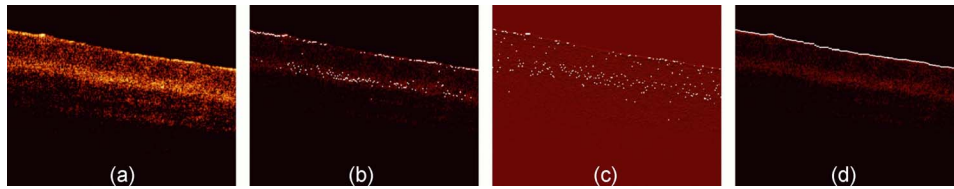
Fig. 1 (a) Denoised B-scan of a biological sample (porcine esophagus), (b) white dots mark the point of highest intensity within one A-scan (overlay over dimmed intensity image), (c) white dots mark maximum rate of change in intensity within one A-scan, and (d) surface pixels as recognised from a binary threshold mask. (b) and (c) show that the maximum intensity and derivative do not necessarily correlate with the surface.

1.1.2 Surface recognition

For OCT there is no established method for unsupervised surface recognition. This is, however, not straightforward as the point with the highest intensity within one A-scan is not necessarily correlated with the sample surface. This is also true for taking the first derivative of the intensity; the largest change in intensity is not necessarily at the surface. Were this the case, surface recognition could be done quickly and elegantly using a weighing of the magnitude and the derivative of the signal intensity over the single A-scan. Specifically for biological samples, this is prone to errors (cf. Fig. 1).

There is a small amount of literature about surface recognition in OCT and research groups seem to have methods that seem fit for the purpose of their publication, but are not always useful to implement for a different specimen type. We found a variety of reports, e.g., shapelet-based boundary recognition;[7] rotating kernel transformation, which emphasizes straight lines within an image;[11] erosion/dilation techniques on a binary threshold image;[8] or polynomial fitting through median-blurred intensity peaks.[12]

We chose binary thresholding because it is accurate and fast. A visualization of the full process is shown in Fig. 2.

1.2 Research Questions

In this paper, we investigate whether OCT data can be classified in real time on the basis of A-scans without any input from a human operator. We investigate the classification performance on a small demonstrator group, and plan to assess the impact of surface normalization. We further investigate whether representative measures can be established for larger data sets, which strain computer resources. Finally, we investigate the performance on animal tissue.

2 Methods

2.1 Measurements

2.1.1 OCT device

The OCT device used in this study was a time-domain benchtop system, which was similar in design to the system described by Yang et al.[13] The light source is a superluminescent diode (SLD, Superlum) with a central wavelength of 1310 nm that provided an axial resolution of $\sim$15 μm in air and a 10-μm lateral resolution. The rapid delay line is realized using a double-pass mirror and a reference mirror mounted on a galvanometer (Cambridge Instruments). The data acquisition was performed using LabVIEW.

2.1.2 Samples

As a demonstrator of proof of concept, fruit and vegetables served as readily available samples with varying structure (Fig. 3). Ten scans were taken of each vegetable; three vegetables of one kind form a group. In addition to this, fresh porcine specimens were obtained from a local butcher. Twenty different samples were prepared from each investigated organ, although all specimens originate from one animal. The size and shape of the samples enabled them to be placed in petri dishes for the measurements. Where necessary, samples were kept stable with needles.

2.2 Preprocessing

2.2.1 Image

All images used in this study measure 300×500 pixels, covering a 3-mm width and an approximately 5-mm depth. The log of the intensity data (time domain signal) was stored as a matrix. Values were handled in double precision format. For the automated noise reduction, a ribbon of "air," intensity in-
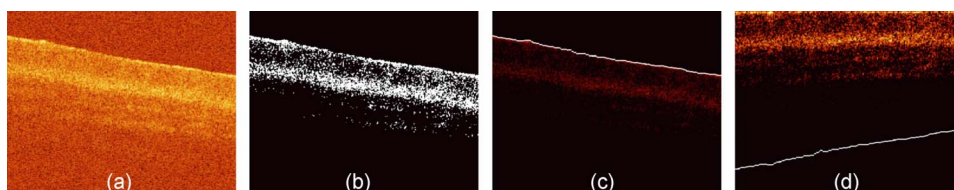

Fig. 2 (a) Original noisy OCT image (false color log of intensity envelope data), representative for many *ex vivo* scans, where the sample surface, although somewhat straight, is not exactly parallel to the scanning direction; and (b) binary image using a 0.90 quantile threshold; (c) surface recognized from (b); and (d) image after wrapping.
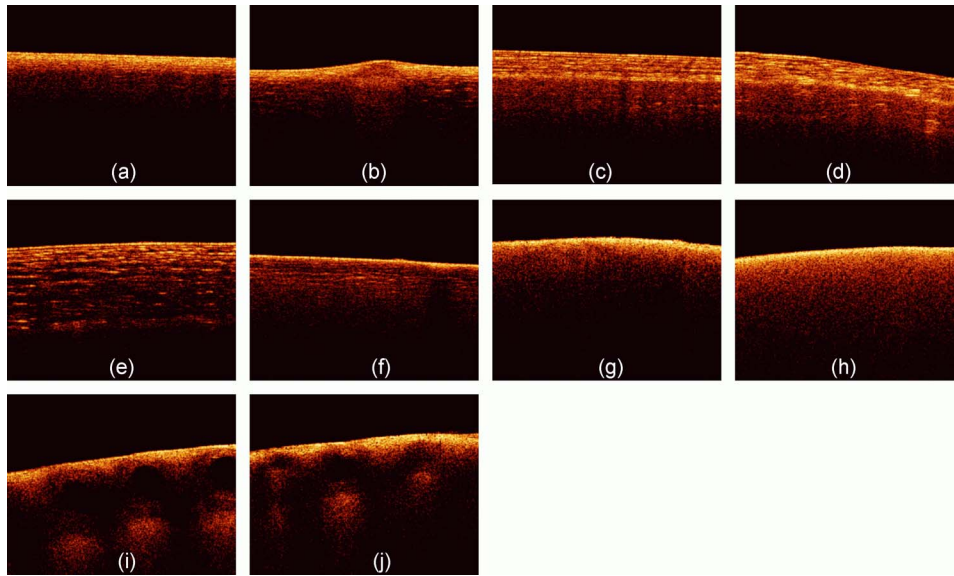
Fig. 3 Representative OCT images of vegetables (after denoising, to be fed into surface recognition): (a) celery, longitudinal; (b) Celery, axial; (c) leek, longitudinal; (d) leek, axial; (e) onion; (f) potato; (g) mushroom stem; (h) mushroom top; (i) lemon; and (j) lime. All B-scans in this paper have a width of 3 mm. Note the somewhat similar appearance of (c) and (d), (g) and (h), and (i) and (j).

formation above the surface, was regarded as background noise level.[14] The mean of this plus $7\times$ the standard deviation of this background signal was used as the intensity threshold. This relatively high threshold does crudely cut off low-level signals from deeper areas. We found it, however, to be more important for our subsequent surface recognition to have a noise-free image, with the surface not being obscured by artefacts or layers.

The scanning protocol therefore required an area of air above the surface. We aimed to cover approximately 1/5 to 1/3 of the image height, as this was convenient for manual handling. The surface of the sample was positioned to ensure that the full penetration depth could be depicted on the image. Specimens were placed so that the beam angle would not cause excessive reflections. While these limitations for the distance of the probe to the sample were chosen with respect to easy handling, no measurement was taken from the same site, as that would spoil the validity of the subsequent correlation.

After thresholding, the intensity values of one B-scan were normalized to the maximum intensity peak, by scaling to the maximum peak being unity.

2.2.2 *Surface recognition and normalizing*

Detecting the surface based on a binary threshold mask of the image, as shown in Fig. 2, is an approach we found to be robust for the purpose of our preprocessing. After finding no substantial influence on the result for threshold levels between 0.85 and 0.975, we arbitrarily defined a 0.95 quantile threshold for creating a binary image. The first pixel of this mask was taken as the surface. No smoothing or correction of eventual outliers (e.g., reflection artefacts) was performed. Normalizing to the surface was achieved by shifting each A-scan according to the surface vector.

Of these images, we calculated several representative measures of the depth profile of the intensity variation within one

image: the mean B-scan, median B-scan, standard deviation, or the mean divided by the standard deviation[15] (M/SD). Furthermore, we looked into combinations of these, e.g., M/SD and mean as two representative measures for one B-scan. We also used the FT of an A-scan to be able to roughly compare our results with the approach of Zysk and Boppart.[10]

2.3 *Data Analysis*

Data handling, preprocessing, PCA, and LDA were performed using MATLAB V6.1.0.450, Release 12.1 (The MathWorks, Inc.), extended by the "PLS_Toolbox" (Eigenvector Research, Inc.) and our in-house tools ("Classification_Toolbox"). Emphasis was put on speed, and therefore execution times were assessed using the tic-toc functions and the "profile" function in MATLAB. Bottlenecks, e.g., the routine to normalize A-scans to the surface, were vectorized.

We did not perform mean centering of the data as we found this to have no effect on the quality of results (data not shown). Using PCA, the data was reduced to the first 25 principal components (PCs), which were then sorted in order of their statistical significance, characterized by their respective F values determined by an analysis of variance[15] (ANOVA) ($p < 0.001$). [The number of PCs (25) was chosen arbitrarily and is not fundamental to the method. The number of PCs does affect the results (data not shown), but for the purpose of this paper, whereby different approaches are studied, a constant of 25 PCs is a fair representation.]

These 25 PCs were then used in their entirety for LDA, which is inherently able to suppress loadings of lower significance for the model. The training model was tested using leave-one-out cross-validation. For a small group (30 scans), one measure was left out: one A-scan when classifying separate A-scans, or one representative measure when classifying the representative measures (leave-on-scan-out, LOSO). While for the LOSO cross-validation, each prediction data set is "unknown" to the algorithm, the training model can still
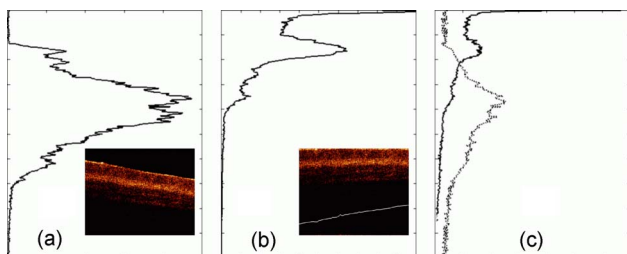
Fig. 4 (a) Mean A-scan of original data, (b) mean A-scan of surface normalized image, and (c) M/SD along row of A-scans: (—) surface normalized and (- - -) original.

include data from the very same sample. A better cross-validation approach is therefore to test a classification model with a prediction set whose samples are "unknown" in their entirety. This is achieved by splitting the data into a training group and a prediction group (leave-one-group-out, LOGO). For the large sample groups (330 scans) we created a training model from 220 randomly chosen scans, and evaluated the prediction performance for the remaining 110 scans, which are unknown to the test model. These 110 scans were from different samples than the 220 training scans. In this way, cross-validation could be performed on three different combinations. The mean of the results for the three cross-validations for one set of samples we accept as classification performance. For comparing pairs of similar vegetables and the animal samples, a prediction model was created using a randomly assigned half of the B-scans of each group and cross-validated using the other half of the B-scans. Also here, test and prediction groups were from different samples, and the mean of the two results was taken as overall performance.

3 Results

The images appear typical of common OCT image quality. Representatives of fruit and vegetable structures are shown in Fig. 3. Distinguishing such images, when they are presented next to each other, is fairly easy. However, when working

through a large quantity, some of these could start to look similar to an insufficiently trained or overworked eye.

The results section is divided into two groups. Part 1 shows the effects of preprocessing and demonstrates the rationale for choosing our parameters. Part 2 shows the actual classification results.

3.1 Part 1: Preprocessing

3.1.1 Surface recognition and normalizing

Vectorization reduced the processing time roughly by a factor of 4 to 5. The visual results of the unsupervised surface normalizing were effectively shown in Fig. 2(d). Figure 4 shows how normalizing to the surface affects the mean and the M/SD of a B-scan.

3.1.2 Vegetables—separate A-scans

In a first approach, we used 10 images of three vegetables, resulting in 30 images or 9000 A-scans. Performing PCA on the surface normalized A-scans [Fig. 5(A)] or nonsurface normalized FT curves [Fig. 5(B)] show a tendency to group, but also considerable overlap. Note that the plots in Fig. 5 show only the two most significant (ANOVA) PCs for classification against each other. The subsequent LDA separates on the basis of a pool of several PCs (25 in the examples of this paper).

Table 1 shows the performance of classification models for this case. For the amplitude of the frequency content, surface normalization does not improve the classification. For classifying time domain data (spatial information), surface normalization yields a higher rate.

3.1.3 Data reduction prior to PCA

The classification rates in Table 1 are not disappointing, but take a long time to calculate. Hence we reduced the B-scan to a representative measure, as shown in Fig. 6. From the mean of these images, the separation is clearer than in Fig. 5. For this approach, surface normalization has a stronger impact
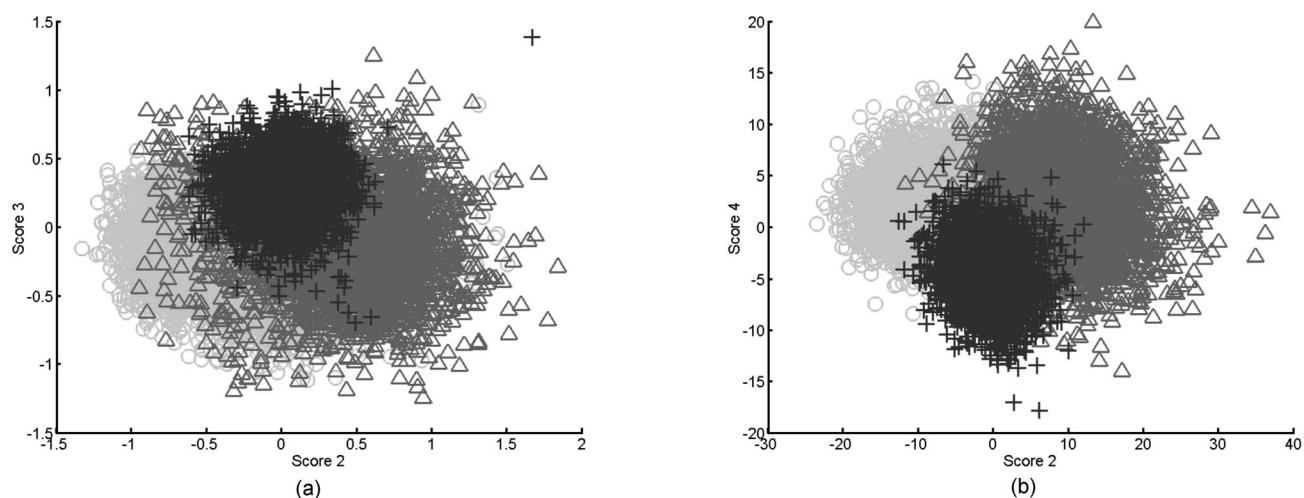

Fig. 5 PCA of three vegetables, 10 images each. The plots show the PCs of 9000 A-scans with the two highest scores (determined by ANOVA) out of 25: (a) surface normalized and (b) FT of A-scans (no surface normalization). Legend: ○=celery; △=onion; +=mushroom.

Table 1 Comparison for classification of three A-scan representations: the FT, the amplitude of the FT, and time domain data.

Preprocessing	FT	Amplitude of FT	Depth profile (A-scan)
Without surface normalization	65.4	96.1	65.4
Surface normalization	92.8	95.8	92.8

The data set consisted of three groups of 10 images each (9000 A-scans). Scores are percentages correctly classified by the algorithm, after LOSO cross-validation.

than before: classification was correct in 93.3% for the surface-normalized data [Fig. 6(A)], but only 33.3% for not-surface-normalized data [Fig. 6(B)].

3.1.4 Data reduction—large dataset

A large data set consisting of 330 images, categorized into nine groups, was used to assess various representative measures. These 99,000 A-scans could not be fed into analysis on a Pentium type processor with 1 Gbyte of RAM, so it was essential to extract a representative measure. Table 2 shows the results for data reduction. For time domain data, models created using the mean normalized by the standard deviation (M/SD) had the best results, as shown in Table 2. For models from the FT of an A-scan, the standard deviation (SD) has the highest classification score. After LOGO cross-validation, the model constructed from the mean values has a higher score.

3.2 Part 2: Classification

From the results so far, combinations with high scores were chosen for further comparison: surface normalization with M/SD, and the mean and the SD of the amplitude of the FT. We abbreviate these as SN-MSD, AFT-M, and AFT-SD, respectively.
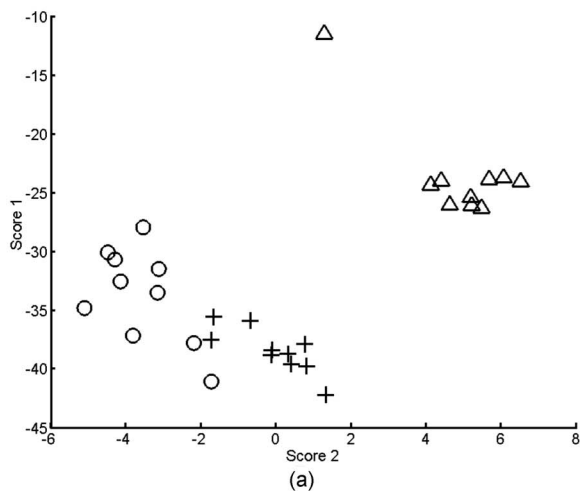
Table 2 Training performance (TP) and cross-validation (CV) of a data set consisting of nine groups; seven groups of 30 images each and two groups of 60 images each (330 representative measures).

Representative Measure	A-Scan[a]			FT of A-Scan[b]		
	TP[c]	LOSO[d]	LOGO[e]	TP[c]	LOSO[d]	LOGO[e]
Median	86.7	80.0	81.6	88.8	83.3	82.7
Mean	88.2	84.2	81.6	90.0	**86.4**	**86.7**
SD	86.4	77.6	72.9	**91.2**	85.5	85.1
M/SD	**92.1**	**86.1**	**82.0**	90.0	85.5	81.8

Scores are percentages correctly classified by the algorithm, before and after cross-validation. The highest scores are bold.
[a]Surface-normalized time domain intensity data,
[b]amplitude of FT (nonsurface-normalized) A-scan,
[c]training performance,
[d]leave-one-scan-out cross-validation,
[e]leave-one-group-out cross-validation.

3.2.1 Vegetables

A different set consisting of 330 images, categorized into seven groups, was classified correctly (after LOGO cross-validation) as follows: SN-MSD, 91.8%; AFT-M, 93.9%; and AFT-SD, 79.1%. For comparison, the non-surface-normalized approach reached 32.0% in this case.

3.2.2 Similar image types

In the next step, we assessed the classification of similar images, i.e., measurements of one type of vegetable with different orientation, or similar images (lemon versus lime). Representative B-scans are shown in Figs. 3(G)–3(J). Classification results for 30 images each are shown in Table 3.
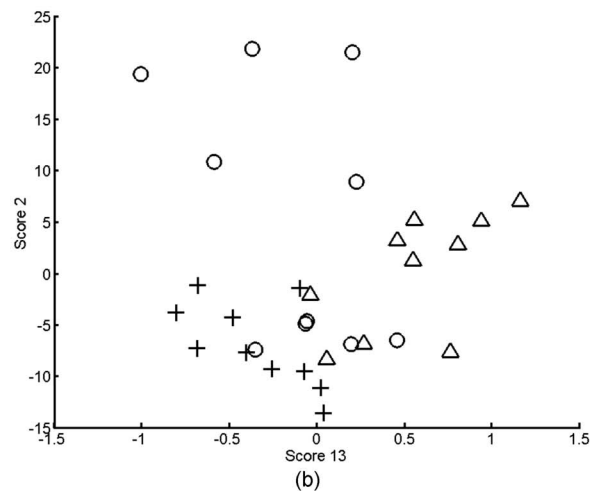

Fig. 6 PCA of three vegetables, 10 images each. The data points show the two highest scoring PCs (ANOVA) of the 30 representative measures: (a) surface normalized and (b) no surface normalisation. ○=celery; △=onion; +=mushroom.

Table 3 Classification rates after LOGO cross-validation; training models were created for pairs of similar images (2×30 images for a pair).

Type of Similar Images	Representative Measure (%)		
	SN-MSD	AFT-M	AFT-SD
Celery (two orientations)	56.7	68.3	86.7
Leek (two orientations)	61.7	63.3	80.0
Mushroom (two orientations)	96.7	90.0	90.0
Lemon versus Lime	83.3	95.0	95.0

Table 4 Classification rates after LOGO cross-validation; training models were created for sets of pairs of similar images (2×30 images for a pair).

Type of Similar Images	Representative Measure (%)		
	SN-MSD	AFT-M	AFT-SD
Two mushroom orientations, lemon versus lime	90.8	91.7	86.7
All four pairs (eight groups) from Table 3	67.1	74.2	82.1

3.2.3 Similar images in a large group

If groups are similar, further added groups do decrease the power of the separation. To show this, we grouped two vegetables to yield four groups. Separation is affected when these two groups must be distinguished not only from each other, but also among other data, as shown in Table 4.

3.2.4 Preliminary tissue pilot study

Extending on the vegetable findings, we used porcine samples for a preliminary tissue pilot study. We measured 20 sites each of four different tissue types (tongue, esophagus, tracheae, and bronchioles). Representative images are shown in Fig. 7. The results for PC-fed LDA classification are encouraging. After LOGO cross-validation, the four, albeit clearly distinct tissue types, are classified 90.0% (SN-MSD), 82.5% (AFT-M), and 70.0% (AFT-SD) correctly.

To see whether more subtle changes can be picked up, we looked at samples from one tissue type (esophageal) measuring them in a fresh state, after degradation for 5 days, and after application of 5% acetic acid, a contrast agent that is used in clinical settings. These three groups can be separated well between each other in 91.7% (SN-M/SD), 93.3% (AFT-M), and 76.7% (AFT-SD) of the cases (LOGO cross-validated).

4 Discussion

OCT provides structural images that are intuitive to look at and this might even be an important aside for this technology to become accepted by many users. Evaluation of such structural data seems straightforward; however, it remains a challenge in detailed areas. This is evident for an imaging technology that enables one to look closely at areas that previously have been impossible to observe.

Acquisition speed enables generation of high-resolution volumetric real-time data; this is difficult to handle and display meaningfully and will remain complex to interpret in the future. Searching through a vast voxel space can hardly be imagined to be an enjoyable task for a human operator.

4.1 Preprocessing

4.1.1 Time domain data and frequency content

Classification requires preprocessing. Both methods, the amplitude of the FT of an A-scan and the surface normalized A-scan, give good results. The FT we used to create a model differed in four points from that of Zysk and Boppart.[10] First, our denoising and normalizing, as already described, was different. Second, we were using the FT of the A-scan for classification, and not the difference from a reference. Third, Zysk and Boppart report that they did perform surface normalizing ("truncating") prior to FT, whereas we found that to be beneficial only when feeding the real and the imaginary part of the FT. However, we achieved better results when using the amplitude of the FT directly of the non-surface-normalized A-scan. Finally, we took a representative measure for a set of A-scans and classified the set, rather than classifying single A-scans.

It then probably depends on the sample structure, e.g., the presence of distinctive periodic patterns, whether the frequency information or time domain data enable a better classification. Time domain data must be surface normalized. This requires a robust and quick surface recognition and means an additional computational step.

We are aware that the data for this paper are suitable for binary thresholding as surface recognition, and that the converse is not necessarily always the case. Such scenarios could be imagined where the surface of interest is covered by a
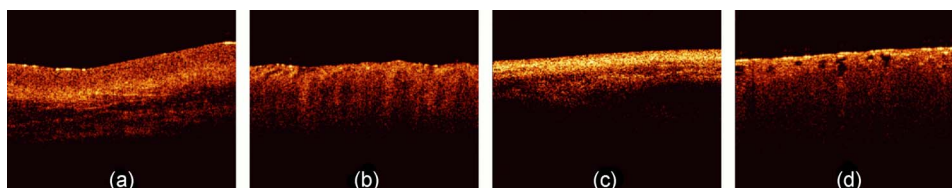

Fig. 7 Representative OCT images of porcine samples: (a) squamous lining of the esophagus, (b) tongue, (c) tracheae, and (d) bronchioles. Each image covers 3 mm in width.
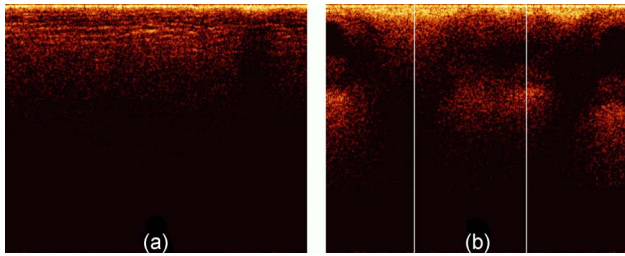
Fig. 8 (a) Surface-normalized B-scans, where the depth profile is homogeneous across the scan width and can be well described by a mean A-scan, and (b) structures that result in different A-scans, e.g., A-scans at the two white lines will be different. In this situation, a standardized mean seems to better represent the B-scan.

layer of mucus, a sheath, a balloon, or something similar. Research into more robust and reliable surface detection algorithms would therefore be beneficial.

4.1.2 Representative measure

Reducing the data to a representative A-scan is necessary; our Pentium type processor with 1 Gbyte of memory was not able to handle data sets of tens of thousands of A-scans. More powerful computers are of course available, but it is conceivable that OCT data will stretch these resources to the limit for the near future. We found that the M/SD of a surface-normalized B-scan or the mean of the frequency content yield good results. For instance, at separating two similar celery orientations, the SN-MSD approach practically fails (56.7%), whereas the AFT-SD reaches over 80%. For classifying porcine tissue, the SN-MSD approach had the best result. In the case of variations within a B-scan, such as in Fig. 8, standardizing using the SD does improve results for the subsequent classification.

In terms of preprocessing, we also investigated mean centering, and using more than one representative measure (data not shown). We did not perform any correction of the surface vector in this study, although that is beneficial (data not shown). Further, the performance of LDA can be improved by feeding an optimum number of PCs. However, this optimum depends on the number and kind of groups in the classification model. The different sample groups in this study do have different optimum numbers of PCs. For instance, for classifying four tissue types, the classification performance, using 25 PCs, was 90.0% (SN-MSD) and 82.5% (AFT-M). For these groups, the optimum for SN-MSD would be reached by feeding 11 PCs (97.5% LOGO cross-validated), and for AFT-M 7 PCs (91.3%). For the purpose of this demonstration, we decided to keep a fixed number of PCs. It might well be worth investigating a combination of two or more measures, and automatically refining extracted measures. However, this is beyond the scope of this paper and is probably only worthwhile when refining the technique for a dedicated purpose. Summarizing, it is possible to improve results by fine-tuning and combining parameters.

4.1.3 Speed

One criterion for processing is speed. OCT acquires data in real time and hence the subsequent analysis should not impact on this characteristic. Preprocessing does slightly delay re-

sults: on a 2.80-GHz Pentium 4, classification of 330 B-scans takes approximately 85 s, of which 73 s are required for surface normalization and extracting a representative measure. That corresponds to roughly 60,000 A-scans/min and makes us confident that the claim of providing a near real-time modality remains justified. It would be quite remarkable for a human operator to classify 330 images in $1\frac{1}{2}$ min. However, we were only using a run-length-encoded programming language for proof of principle; compiled programs run much faster.

4.2 Classification

Robust cross-validation methods are important for testing LDA. It is a powerful approach, and for a small enough set of samples some separation into groups can always be expected. For instance, in the example of 30 images, the model created by LDA was able to classify 100% correctly even without surface normalization. After cross-validation, the surface-normalized model was classified in 93.3 versus 33.3% for the not-surface-normalized model.

4.2.1 Relevance for medical data

The fact that the algorithm reaches over 90% accuracy in certain cases is reassuring when looking at obviously different images, but for our data set the displayed approach is also able to differentiate images reasonably well that are difficult to tell apart by the eye. Further, the algorithm is still able to classify similar images among a pool of other structural data.

This is important as we plan to apply this algorithm to medical data. There, we expect a larger variation among healthy individuals, and only eventually the small, but significant alterations for early, nonsymptomatic disease. As an example, we note epithelial cancers, carcinomas, where the epithelial thickness of the healthy population can vary considerably due to various factors. However, the subtle alterations of proliferative epithelial cells close to the basal layer are probably the feature that can be picked up by OCT scans. It still remains to be assessed whether the algorithm would fail in such scenarios.

4.2.2 Limitations and benefit

In this study the imaging protocol is strict in the sense that it requires the penetration depth to be fully exploited and to leave an area of air above the surface. However, most imaging must adhere to a protocol. As long as the operator is made aware of the requirements, we believe this limitation not to be substantial. After all, we have acquired thousands of images in this way.

The advantage we see in the presented algorithm is that the few parameters utilized here are robust enough for the analysis to be automated and in near real-time. It does not require assigning areas of interest. It does not rely on specific structures such as previously reported methods, but can still be used in combination with these algorithms. Hence, this algorithm is straightforward to implement for existing OCT data.

Note that, although we were taking representative measures of B-scans, this approach will work on volumetric data. For example, taking a representative measure of a volume

consisting of 20×20 A-scans on a system with 10-μm lateral resolution will enable classification of 25 "tiles"/mm^2, which is still superb for targeting clinical treatment.

4.2.3 *Further work*

Previous studies[16,17] have shown that expert histopathologists do not always fully agree on classification of tissue specimens, and an algorithm with lower than 100% correct classification can still outperform subjective human assessment. In clinical work, OCT data would have to be compared to histopathologic results. Even though there are significant discrepancies found between expert pathologists on particular pathology groups,[16,17] histopathology is defined as the gold standard. Therefore, we must base our analysis on the assumption that the sample pathology defined by histopathology is correct. In this study, we did not investigate the performance of the algorithm against a human operator, and hence assumed a 100% correct classification as a reference baseline. In the future, we intend to compare classification results with those achieved by histopathologists on the same OCT images from biological specimens.

The next step is testing this algorithm on medical OCT images. We expect it to distinguish structurally different pathologies correctly. In ongoing work, we are investigating how well it classifies disease groups with quite similar appearance, such as mild and moderate dysplasia in early cancer development.

References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, "Optical coherence tomography," *Science* **254**, 1178–1181 (1991).
2. P. H. Tomlins and R. K. Wang, "Theory, developments and applications of optical coherence tomography," *J. Phys. D* **38**, 2519–253 (2005).
3. F. I. Feldchtein, G. V. Gelikonov, V. M. Gelikonov, R. V. Kuranov, A. M. Sergeev, N. D. Gladkova, A. V. Shakhov, N. M. Shakhova, L. B. Snopova, A. B. Terent'eva, E. V. Zagainova, Y. P. Chumakov, and I. A. Kuznetzova, "Endoscopic applications of optical coherence tomography," *Opt. Express* **3**(6), 257–270 (1998).
4. P. Targowski, B. Rouba, M. Wojtkowski, and A. Kowalczyk, "The application of optical coherence tomography to non-destructive examination of museum objects," *Stud. Conserv.* **49**, 107–114 (2004).
5. H. Liang, M. Cid, R. Cucu, G. Dobre, A. Podoleanu, J. Pedro, and D. Saunders, "En-face optical coherence tomography—a novel application of non-invasive imaging to art conservation," *Opt. Express* **13**, 6133–6144 (2005).
6. S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987).
7. J. Weissman, T. Hancewicz, and P. Kaplan, "Optical coherence tomography of skin for measurement of epidermal thickness by shapelet-based image analysis," *Opt. Express* **12**, 5760–5769 (2004).
8. X. Qi, M. V. Sivak, Jr., G. Isenberg, J. E. Willis, and A. M. Rollins, "Computer-aided diagnosis of dysplasia in Barrett's esophagus using endoscopic optical coherence tomography," *J. Biomed. Opt.* **11**(4), 044010 (2006).
9. M. L. Huang and H. Y. Chen, "Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography," *Invest. Ophthalmol. Visual Sci.* **46**(11), 4121–4129 (2005).
10. A. M. Zysk and S. A. Boppart, "Computational methods for analysis of human breast tumor tissue in optical coherence tomography images," *J. Biomed. Opt.* **11**(5), 054015 (2006).
11. J. Rogowska, C. M. Bryant, and M. E. Brezinski, "Cartilage thickness measurements from optical coherence tomography," *J. Opt. Soc. Am. A* **20**, 357–367 (2003).
12. Y. Hori, Y. Yasuno, S. Sakai, M. Matsumoto, T. Sugawara, V. Madjarova, M. Yamanari, S. Makita, T. Yasui, T. Araki, M. Itoh, and T. Yatagai, "Automatic characterization and segmentation of human skin using three-dimensional optical coherence tomography," *Opt. Express* **14**, 1862–1877 (2006).
13. Y. Yang, A. Dubois, X.-P. Qin, J. Li, A. El Haj, and R. K. Wang, "Investigation of optical coherence tomography as an imaging modality in tissue engineering," *Phys. Med. Biol.* **51**, 1649–1659 (2006).
14. D. C. Fernández, H. M. Salinas, and C. A. Puliafito, "Automated detection of retinal layer structures on optical coherence tomography images," *Opt. Express* **13**(25), 10200–10216 (2005).
15. R. G. Brereton, "Experimental design and Pattern recognition," Chaps. 1 and 4 in *Chemometrics–Data Analysis for the Laboratory and Chemical Plant*, pp. 15–53, 191–224, John Wiley & Sons Ltd. (2003).
16. C. Kendall, N. Stone, N. Shepherd, K. Geboes, B. Warren, R. Bennett, and H. Barr, "Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in Barrett's oesophagus," *J. Laser Appl.* **200**, 602–609 (2003).
17. N. A. Shepherd, "Barrett's esophagus: its pathology and neoplastic complications," *Esophagus* **1**(1), 17–29 (2003).