

# Method for optical coherence tomography image classification using local features and earth mover's distance

## Yankui Sun

Tsinghua University  
Department of Computer Science and Technology  
Beijing 100084  
China  
E-mail: syk@mail.tsinghua.edu.cn

## Ming Lei

Tsinghua University  
Department of Computer Science and Technology  
Beijing 100084  
China

**Abstract.** Optical coherence tomography (OCT) is a recent imaging method that allows high-resolution, cross-sectional imaging through tissues and materials. Over the past 18 years, OCT has been successfully used in disease diagnosis, biomedical research, material evaluation, and many other domains. As OCT is a recent imaging method, until now surgeons have limited experience using it. In addition, the number of images obtained from the imaging device is too large, so we need an automated method to analyze them. We propose a novel method for automated classification of OCT images based on local features and earth mover's distance (EMD). We evaluated our algorithm using an OCT image set which contains two kinds of skin images, normal skin and *nevus flammeus*. Experimental results demonstrate the effectiveness of our method, which achieved classification accuracy of 0.97 for an EMD+KNN scheme and 0.99 for an EMD+SVM (support vector machine) scheme, much higher than the previous method. Our approach is especially suitable for nonhomogeneous images and could be applied to a wide range of OCT images. © 2009 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3251059]

Keywords: optical coherence tomography; image classification; support vector machine; local features; earth mover's distance.

Paper 09137R received Apr. 9, 2009; revised manuscript received Aug. 14, 2009; accepted for publication Aug. 20, 2009; published online Oct. 30, 2009.

## 1 Introduction

### 1.1 Optical Coherence Tomography

Optical coherence tomography (OCT)<sup>1</sup> is a recent imaging method that allows high-resolution, cross-sectional images through tissues and materials, similar to ultrasound. OCT acquires images by measuring backscattered light. Compared to traditional imaging approaches such as computed tomography (CT) and magnetic resonance imaging (MRI), OCT has higher resolution, so it can reveal the microstructure of tissues and materials. Images from OCT systems typically have a resolution of 1 to 15  $\mu\text{m}$ . Over the past 18 years, OCT has been successfully used in disease diagnosis, biomedical research, material evaluation, and many other domains.

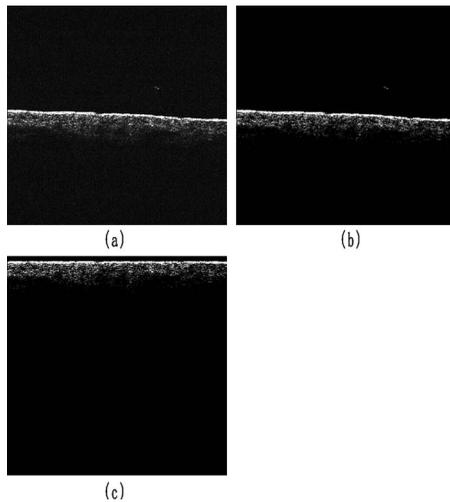
### 1.2 Automated Analysis of OCT Images

In practical applications, the number of OCT images obtained from the imaging device is usually very large. In addition, until now surgeons have limited experience in analyzing these images, so we need methods to automatically process them. Several efforts have been made in this area. Fernandez et al.<sup>2</sup> have employed nonlinear complex diffusion and coherence-enhancing diffusion to automatically detect retinal layer struc-

tures. Baroni et al.<sup>3</sup> have tested texture classification of retinal layers in optical coherence tomography images. In Ref. 4, Koozekanani et al. have devised an approach for retinal thickness measurements from OCT images using a Markov boundary model. Zysk and Boppart<sup>5</sup> have tested the automated diagnosis of breast tumor tissue in OCT images. In Ref. 6, Bazant-Hegemark and Stone proposed a near-real-time method for OCT image classification using principal component analysis (PCA) and linear discriminant analysis (LDA). They calculate row statistical values, such as mean, standard deviation, etc., for preprocessed images, these values are used as feature vectors. For classification, PCA and LDA are used.

For most cases, image classification is the most important task. In all known existing methods, this is done by texture analysis, such as cooccurrence matrix, Fourier transform, etc. Different kinds of machine learning methods, such as linear discriminant analysis and artificial neural network, are used to classify these images. As OCT images are usually structure-poor, texture analysis is the best choice. However, all these methods extract features from images only globally. For nonhomogeneous images, global features cannot describe them effectively. In these images, local regions possess different kinds of patterns. So, we can no longer process them together—we need to analyze these local regions separately.

Address all correspondence to: Yankui Sun, Tsinghua University, Department of Computer Science and Technology, Beijing 100084, China. Tel: 86-10-6278-2609; Fax: 86-10-6277-1138; E-mail: syk@mail.tsinghua.edu.cn



**Fig. 1** Preprocessing results: (a) original image, (b) background noise removed image, and (c) surface normalized image.

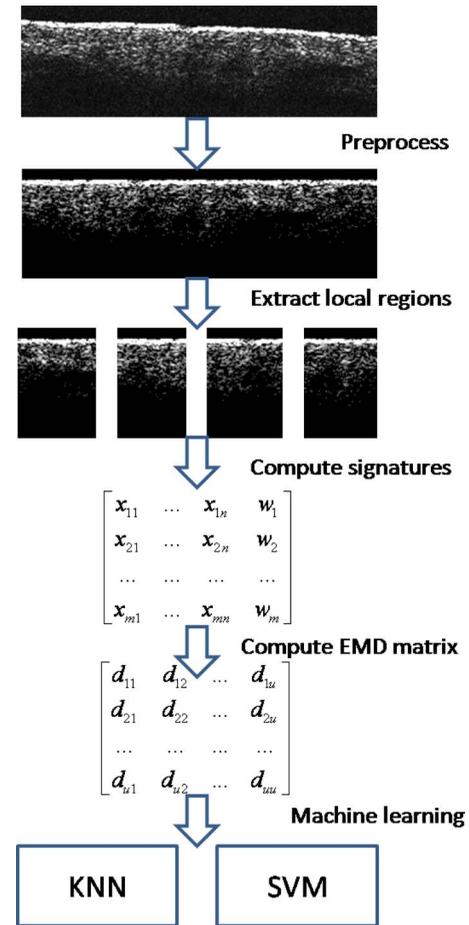
Over the past 10 years, many methods based on analysis of local regions in images have been developed<sup>7–10</sup> in the computer vision domain. In these methods, images are represented by local regions (subimages). Compared to traditional approaches, these methods are robust to background clutter and partial occurrence. When OCT images are nonhomogeneous—that is, they have different kinds of local regions, the local region–based approach is a better choice.

In this paper, we propose an effective and accurate method for OCT image classification based on local region features and earth mover’s distance (EMD). Our approach is robust, and it can handle partial abnormal problem. We divide each image into several subimages (local regions) and then extract features from these subimages and combine them to form signatures.<sup>11–13</sup> Thus, our method is local region–based, as compared to the traditional global approach, which computes features from the entire image, it is robust and can effectively handle the partial abnormal problem. For classification, we implemented both the *k* nearest neighbor (KNN) classifier and the support vector machine (SVM) classifier with the EMD kernel. We used an OCT image set to evaluate our method, which contains normal skin OCT images and *nevus flammeus* images. In experiments, we compared our methods with a baseline approach (PCA+SVM), and achieved much higher performance. We believe that our approach can be applied to other types of OCT images, and it is especially suitable for nonhomogeneous images.

The paper is organized as follows. In Sec. 2, we will describe our approach in detail, including preprocessing, the PCA+SVM approach (baseline method), and our EMD+KNN and EMD+SVM schemes. Section 3 will give experimental results and discussion. Conclusions are given in Sec. 4.

## 2 Methods

In order to demonstrate the effectiveness of our method, we compared it with the approach proposed in Ref. 6. Instead of using LDA, we chose a support vector machine (SVM), as it



**Fig. 2** Framework of our approach.

is the best classifier in almost all cases. For both methods, the preprocessing procedure is identical, and is similar to methods proposed in Ref. 6

In Sec. 2.1, we will describe the preprocessing method. Section 2.2 will describe the baseline method—that is, the PCA+SVM approach. Our new method will be introduced in Sec. 2.3.

### 2.1 Preprocessing

For preprocessing, we employed a method similar to Ref. 6, which contains the following three main steps:

1. Background noise removal.
2. Surface recognition and smoothing.
3. Surface normalization.

The first step is to remove background noise in images. We calculate the mean  $\mu$  and standard deviation  $\sigma$  of the top 20 rows for each image, and then  $\mu+7\times\sigma$  is used as intensity threshold. In experiments, we tested different threshold values, from  $\mu+2\times\sigma$  to  $\mu+9\times\sigma$ , and found that  $\mu+7\times\sigma$  is the best choice. All pixels whose intensity is less than this threshold are considered as background.

After background removal, we conducted surface recognition and normalization identical to Ref. 6. We also performed surface smoothing in order to remove noisy points in the surfaces. Figure 1 shows the preprocessing results. Figure 1(a) is the original image, and Fig. 1(b) is the preprocessed image.

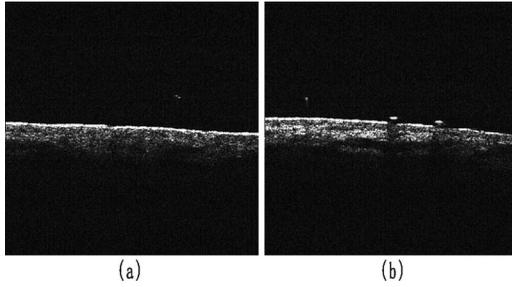


Fig. 3 OCT images of (a) normal skin and (b) *nevus flammeus*.

We can see that background noise was removed and all A-scans were aligned according to surface. In order to demonstrate the results of surface normalization better, we shift the surface point for each column to line 10, and thus the “bright bar” is not located at the top of the images.

## 2.2 Baseline Method: PCA and Support Vector Machine

In the baseline method, we calculate the mean value and standard deviation of each row for preprocessed images. These values are combined to form a feature vector. PCA was performed on these feature vectors. Among all classifiers, SVM often produces state-of-the-art results in high-dimensional problems,<sup>14–16</sup> so we chose it for our classification task instead of LDA. SVM finds a hyperplane that separates two-class data with a maximal margin. When the data set is not linearly separable, SVM uses two strategies, soft margin classification and kernel mapping. There are four kinds of kernel functions usually used: linear kernel, polynomial kernel, radius basis function (RBF), and sigmoid kernel. In experiments, we chose the RBF kernel:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0. \quad (1)$$

In our implementation, we used the source code from Ref. 15 which is a perfect implementation of the various SVM algorithms.

## 2.3 Our Methods: EMD+KNN and EMD+SVM Schemes

The earth mover’s distance<sup>11–13</sup> is a distance between two distributions (signatures) that reflects the minimal amount of work that must be performed to transform one distribution into the other by moving “distribution mass” around. It is a special case of the transportation problem from linear optimization, for which efficient algorithms are available. Signatures are defined as follows:

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_m \end{bmatrix}, \quad (2)$$

where  $x_i$  are  $n$ -dimensional vectors, and scalars  $w_i$  are weights for each vector. When used to compare distributions that have the same overall mass, the EMD is a true metric and has easy-to-compute lower bounds. If two signatures  $P$  and  $Q$  are

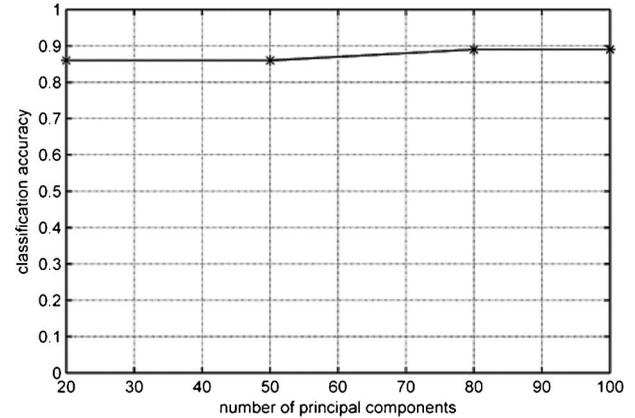


Fig. 4 Classification results of baseline method.

$$P = \{(x_1, p_1), (x_2, p_2), \dots, (x_m, p_m)\},$$

$$Q = \{(y_1, q_1), (y_2, q_2), \dots, (y_n, q_n)\}, \quad (3)$$

where  $x_i$  and  $y_j$  are  $n$ -dimensional vectors, and  $p_i$  and  $q_j$  are weights, then the EMD between  $P$  and  $Q$  is defined as:

$$\text{EMD}(P, Q) = \frac{\sum_{i=1}^m \sum_{j=1}^n f_{ij} d(x_i, y_j)}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}}, \quad (4)$$

where scalars  $f_{ij}$  are flow values, and  $d(x_i, y_j)$  is the ground distance—that is, the distance between vectors  $x_i$  and  $y_j$ . Usually, this is a Euclidean distance; in our approach, we also chose Euclidean distance as ground distance. Scalars  $f_{ij}$  are found to minimize the following function:

$$C = \sum_{i=1}^m \sum_{j=1}^n f_{ij} d(x_i, y_j). \quad (5)$$

In Ref. 13 Rubner et al. used EMD for image and texture retrieval; they represent the image as signatures. Signatures are calculated from the color distribution of images or the output of a Gabor filter applied to texture images. Then, EMD is computed for each image pair. This distance can be used for image and texture retrieval.

In our approach, we do not use color distribution or a Gabor filter; instead, we choose local region features to form signatures. We divide each OCT image into some local regions (subimages), and features are extracted from these subimages and combined to form signatures. We calculate the mean of each row in a subimage, and these values are combined to form the feature vector of the subimage. Then, all of the subimage’s feature vectors are combined to form a matrix, which is the matrix of the signature. Additionally, each feature vector has a weight, and all of these weights form the weight vector of the signature. For classification, we tested both KNN and SVM algorithms. The main framework for our methods contains the following five steps:

1. Preprocess, which is identical to baseline method.

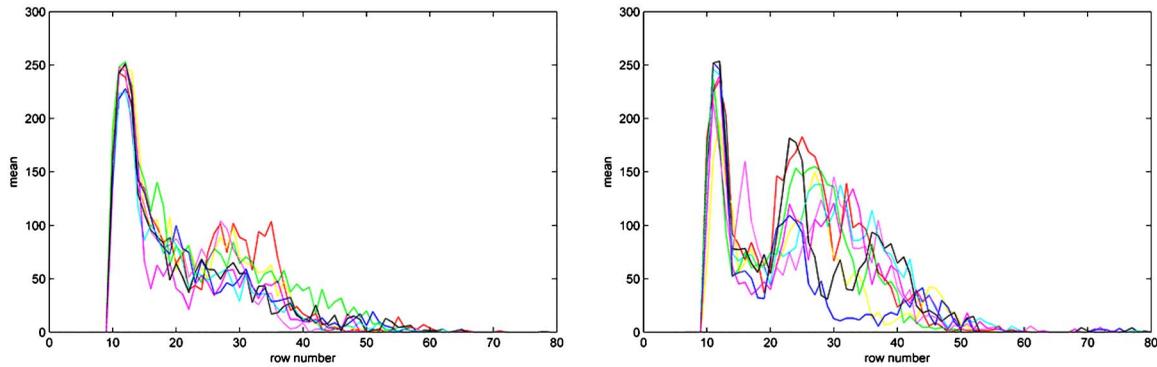


Fig. 5 Signatures of two images: (a) normal image, and (b) abnormal image. Each curve represents a row vector in signatures.

2. Extract the effective region of image, and divide it into subimages.

3. Compute the features of the subimages and combine them to form signatures.

4. Compute the EMD between image pairs.

5. Classify using KNN or SVM classifiers.

For the SVM classifier, we used the EMD kernel,<sup>17,18</sup> defined as:

$$K(S_i, S_j) = \exp\left[-\frac{1}{A}D(S_i, S_j)\right], \quad (6)$$

where  $S_i$  and  $S_j$  are two signatures,  $D(S_i, S_j)$  is the earth mover's distance between  $S_i$  and  $S_j$ , and  $A$  is a scaling parameter that could be determined by cross-validation. When using the EMD kernel, the decision function for SVM becomes:

$$f(S) = \text{sgn}\left[\sum_{i=1}^n y_i \alpha_i K(S, S_i) + b\right]. \quad (7)$$

The framework of our method is shown in Fig. 2. In our approach, we represent images with signatures. The effective region of OCT images contains only the top 80 rows in surface-normalized images. For each preprocessed image, we divide its top 80 rows into 10 subimages ( $80 \times 40$ ) and extract the features of these subimages (the mean of each row); all these feature vectors are combined to form the signature. The weights are set to be 1 for each vector, as all local regions are equally important. Once we have calculated EMD between

each training image and testing sample, we use KNN and SVM algorithms to classify the testing images. For the EMD+SVM scheme, we must calculate the EMD kernel before learning and testing.

### 3 Experimental Results and Discussion

#### 3.1 Experimental Data

We used an image set to evaluate our approach. This data set contains normal skin OCT images and *nevus flammeus* images; the total number of images is 100, of which 50 are normal and the remaining 50 are abnormal. These images are taken from 14 patients; the size of all images is  $400 \times 400$ , and the gray level is 256. The practical axial resolution of our OCT imaging system is about  $10 \mu\text{m}$ , and the transverse resolution is about  $8 \mu\text{m}$ . The actual SNR is about 80 dB, and the penetration depth into skin is more than 2 mm.

Figure 3 shows two images in our image set, Fig. 3(a) is normal skin, and Fig. 3(b) is *nevus flammeus*. In normal images, the layer structure is obvious, the epidermal layer is smooth and continuous, and the dermis is uniform. But in *nevus flammeus* images, the layer structure is destroyed and the epidermal layer is discontinuous. In addition, the dermis is not uniform. So, OCT can effectively discriminate normal skin and *nevus flammeus*. As we can see from Fig. 3, *nevus flammeus* images usually have local abnormal regions—they do not demonstrate global uniform pattern.

Table 1 Classification accuracy of our EMD+KNN method.

Round \ K	1	3	5	7	9	11	13	15
1	0.98	0.96	0.90	0.84	0.84	0.84	0.84	0.84
2	0.96	0.90	0.88	0.84	0.84	0.84	0.82	0.74
3	0.96	0.90	0.86	0.72	0.68	0.68	0.78	0.78
4	0.96	0.88	0.82	0.78	0.78	0.76	0.74	0.74
5	0.96	0.86	0.80	0.82	0.78	0.78	0.76	0.74
Average	0.97	0.90	0.85	0.80	0.78	0.78	0.79	0.77

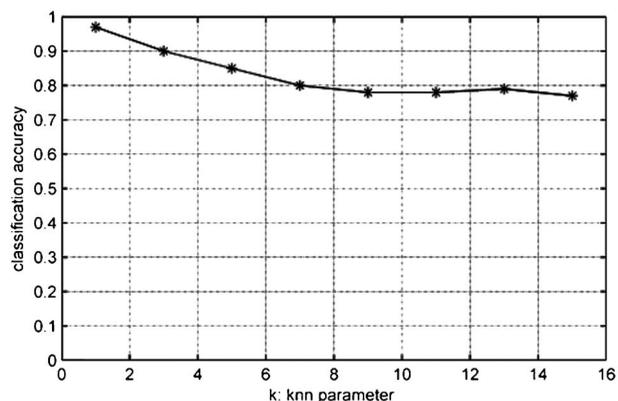


Fig. 6 Average classification accuracy of EMD+KNN method.

### 3.2 Experimental Results of Baseline Method

Our baseline method is quite similar to the approach in Ref. 6—the main difference is that we choose SVM instead of LDA. In experiments, we tested four different numbers of principal components (pc): 20, 50, 80, and 100. The classification accuracy is demonstrated in Fig. 4. We can see that the number of principal components is not crucial for classification performance; this phenomenon was also reported in Ref. 6. When the pc number is 20, the accuracy is 0.86; as the pc number increased to 100, the accuracy increased only to 0.89. The value of penalty parameter  $C$  for SVM is 8 in all experiments. All results are got from 5-fold cross-validation.

### 3.3 Experimental Results of EMD+KNN Method

In our EMD+KNN method, we represent images as signatures, which are formed by feature vectors of local regions in each image. Two signatures are shown in Fig. 5: Fig. 5(a) is normal skin, and Fig. 5(b) is *nevus flammeus*. We can see that signatures can effectively discriminate these two kinds of im-

ages, as they are formed by local region features.

In experiments, we tested a lot of values for parameter  $K$  of the KNN algorithm. The classification results are listed in Table 1. In Table 1, the values of parameter  $K$  include 1, 3, 5, 7, 9, 11, 13, and 15. We conducted 2-fold cross-validation five times; in each round, we randomly select 50 images as training samples and the rest as testing samples. The bottom row lists the average classification accuracy for each  $K$  value. We can see that as  $K$  increases, the performance decreases. When  $K$  is 1, the accuracy is 0.97, which is much higher than the baseline method.

The average accuracy is illustrated in Fig. 6, from which we can see the relation between parameter  $K$  and classification accuracy.  $K=1$  is the best choice, and as  $K$  increases, the overall accuracy will decrease. When  $K=1$ , the KNN algorithm is in fact the nearest neighbor (NN) classifier.

### 3.4 Experimental Results of EMD+SVM Method

Table 2 demonstrates the classification accuracy of the EMD+SVM scheme. In experiments, we tested different values for scaling parameter  $A$  of the EMD kernel and different values for penalty parameter  $C$  for SVM. Each row in Table 2 represents different  $A$  values, and each column represents different  $C$  values. We can see that EMD+SVM outperforms both the baseline and EMD+KNN methods; when the scaling is properly chosen, the accuracy could reach 0.99, which is a promising result. Classification accuracy with different  $A$  values is listed in Fig. 7. We can see that as  $A$  increases from 15 to 60, the accuracy increases simultaneously. But when  $A$  has reached 55, the improvement is negligible.

Compared to the baseline method, our approach achieved much higher performance. Essentially, the former is a global method, while the latter is a local approach. For most *nevus flammeus* OCT images, the pattern is not uniform; they usually contain local abnormal regions, and thus global features cannot describe this property. Unlike the global method, the

**Table 2** Classification accuracy of SVM with EMD kernel.  $A$ : scaling parameter for EMD kernel;  $C$ : penalty parameter for SVM.

$A \setminus C$	1	2	4	8	16	32
15	0.85	0.85	0.85	0.85	0.85	0.85
20	0.86	0.88	0.88	0.88	0.88	0.88
25	0.90	0.91	0.91	0.91	0.91	0.91
30	0.93	0.94	0.94	0.94	0.94	0.94
35	0.96	0.96	0.96	0.96	0.96	0.96
40	0.97	0.97	0.97	0.97	0.97	0.97
45	0.97	0.97	0.97	0.97	0.97	0.97
50	0.97	0.97	0.97	0.97	0.97	0.97
55	0.99	0.99	0.99	0.99	0.99	0.99
60	0.99	0.99	0.99	0.99	0.99	0.99

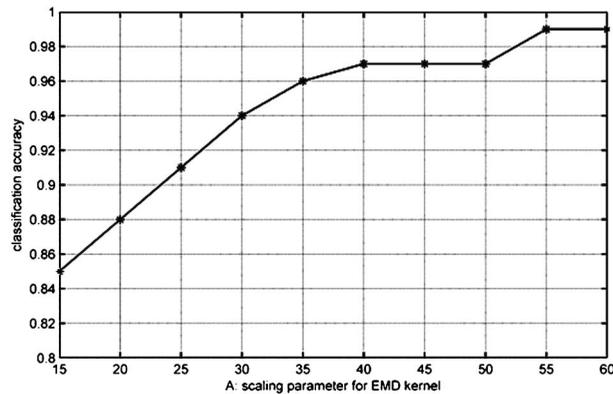


Fig. 7 Average classification accuracy of EMD+SVM method.

EMD approach represents images as signatures, which are formed by combinations of the features of several local regions, and it can completely and accurately describe different regions of images. Furthermore, EMD allows for partial matching and is robust to clutters, so our approach could achieve higher classification accuracy than the baseline method. We believe our method can also be used to classify other types of OCT images that do not demonstrate uniform, homogeneous patterns.

#### 4 Conclusion

We have proposed a new method for OCT image classification. Our approach is based on calculating the signatures of images and EMD between image pairs, which can effectively handle nonhomogeneous images. Experimental results demonstrated the effectiveness of our method, which achieved classification accuracy of 0.97 and 0.99 for EMD+KNN and EMD+SVM schemes, respectively, for appropriate parameter values. Compared to the baseline method, which achieved accuracy of 0.89, our method possesses obvious advantages, and it is especially suitable for nonhomogeneous images. We believe that our method can also be applied to classification tasks of other types of OCT images.

#### Acknowledgments

This research was supported by 863 Project of China under Grant No. 2006AA02Z472, the National Natural Science Foundation of China under Grant No. 60971006. The authors are thankful for the anonymous reviewer's helpful comments.

#### References

1. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schumann, 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. D. C. Fernandez, H. M. Salinas, and C. A. Puliafito, "Automated detection of retinal layer structures on optical coherence tomography images," *Opt. Express* **13**, 10200–10216 (2005).
3. M. Baroni, S. Diciotti, A. Evangelisti, P. Fortunato, and A. La Torre, "Texture classification of retinal-layers in optical coherence tomography," in *11th Mediterranean Conf. Medical and Biological Engineering and Computing (MEDICON 2007)*, T. Jarm, P. Kramar, and A. Županič, Eds., pp. 847–850, Springer, Berlin Heidelberg (2007).
4. D. Koozekanani, K. Boyer, and C. Roberts, "Retinal thickness measurements from optical coherence tomography using a Markov boundary model," *IEEE Trans. Med. Imaging* **20**, 900–916 (2001).
5. 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).
6. F. Bazant-Hegemark and N. Stone, "Near real-time classification of optical coherence tomography data using principal components fed linear discriminant analysis," *J. Biomed. Opt.* **13**(3), 034002 (2008).
7. L. Zhu, A. Rao, and A. Zhang, "Theory of keyblock-based image retrieval," *ACM Trans. Inf. Syst.* **20**(2), 224–257 (2002).
8. S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8), 1265–1278 (2005).
9. S. Agarwal and D. Roth, "Learning a sparse representation for object detection," in *Proc., 7th European Conf. on Computer Vision*, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., Vol. 2353, pp. 113–130, Springer, London, UK (2002).
10. G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," in *Proc., ECCV04 Workshop on Statistical Learning in Computer Vision*, A. Leonardis, and H. Bischof, pp. 1–22 (2004).
11. E. Levina and P. Bickel, "The earth mover's distance is the mallows distance: some insights from statistics," *Proc. Intl. Conf. Computer Vision*, **2**, 251–256 (2001).
12. Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.* **40**(2), 99–121 (2000).
13. Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Proc., 6th Intl. Conf. on Computer Vision*, pp. 59–66 (1998).
14. V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York (1995).
15. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines" (2001). Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
16. C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," (2009). Available at: [www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf](http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf).
17. O. Chapelle, P. Haffner, and V. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Netw.* **10**(5), 1055–1064 (1999).
18. F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "Support vector machines for region-based image retrieval," in *Proc., Intl. Conf. on Multimedia and Expo (ICME'03)*, Vol. 2, pp. II-21-4, IEEE Computer Society, Washington, DC (2003).