

Resilient Subclass Discriminant Analysis with Application to Prelens Tear Film Interferometry*

Kim L. Boyer¹ and Dijia Wu^{1,2}

¹ Signal Analysis and Machine Perception Laboratory
Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY, USA

² Siemens Corporate Research
Princeton, NJ, USA

kim@ecse.rpi.edu, dijia.wu@siemens.com

Abstract. The study of tear film thickness and breakup has important implications for understanding tear physiology and dynamics. We have developed a complete end-to-end automated system for robust and accurate measurements of the tear film thickness from interferometric video as a function of position and time (following a blink). This paper will primarily address the problem of identifying dry regions on the surface of the contact lens, which is one of the four major components of the system. (The others are motion stabilization, image normalization, and phase demodulation to infer absolute thickness and map the surface. To address the challenging wet/dry segmentation problem, we propose a new Gaussian clustering method for feature extraction in high dimensional spaces. Each class is modeled as a mixture of Gaussians, clustered using Expectation-Maximization in the lower-dimensional Fisher's discriminant space. We show that this approach adapts to a wide range of distributions and is insensitive to training sample size. We present experimental results on the real-world problem of identifying regions of breakup (drying) of the prelens tear film from narrowband interferometry for contact lens wearers *in vivo*.)

Keywords: Mixture of Gaussians, Expectation-Maximization, Feature Extraction, Clustering, Prelens Tear Film, Interferometry, Dry Eye Syndrome.

1 Introduction

The thickness of the tear film on the surface of the human eye has important implications in the study of tear film physiology and its fluid dynamics. Knowledge of the tear film thickness as a function of time and position over the eye following a

* This paper corresponds to an invited keynote address and contains some material that previously appeared in the proceedings of the 2009 *IEEE International Conference on Computer Vision*.

blink is necessary to develop and/or verify models of tear film deposition [1] and to analyze the flow in the tear film. For example, the surface tension gradients of the tears will pull the tears upward toward regions of greater surface tension, which is thought to account for the upward drift of the film surface after a blink. For given surface tension gradients and tear viscosity, the tear velocity will be proportional to tear film thickness and the total flow will be proportional to the square of the thickness [2]. In addition, the curvature of the outer surface of the tear film generates pressure in the film and variations in the surface curvature will therefore cause tangential tear flow. For given pressure gradients and viscosity, the tear velocity will be proportional to the square of tear film thickness, and the total rate of flow will be proportional to the cube of thickness [1, 3]. The same dependence on tear film thickness also applies to the downward flow arising from gravity, but this is a small effect [4] for a normal tear film thickness of about 3 to 4 μm [5-7].

In this study we focus on the thickness measurements of one particular type of tear film, the prelens tear film (PLTF) on the surface of a contact lens. That is, the thickness we will measure is the distance between the air surface of the tears and the anterior contact lens surface (more on this below). The PLTF is particularly important for several specific reasons. First, the outer layer of the tears provides a uniform coating over the contact lens, making it a smooth (*i.e.* low distortion) optical surface. If the outer layer becomes rough or irregular as a result of tear drying and breakup, light will be scattered and image (vision) quality impaired. Another function of the PLTF is to provide comfort and lubrication to the palpebral conjunctiva, especially during the blink. In addition, the superficial lipid layer of the tear film reduces evaporation of the film, maintaining contact lens hydration. Dryness and discomfort with contact lens wearers have been reported by as many as 50% of contact lens wearers and are two major reasons associated with premature contact lens discontinuation [8]. The increased evaporation of the PLTF, followed by contact lens dehydration and depletion of the post-lens tear film by absorption into the contact lens, may be the mechanism of contact lens related dry eye in these subjects.

Despite the importance of tear film thickness to understanding its behavior with an eye to developing therapies for dry eye problems, the true thickness of the tear film under various conditions (prelens, post lens, precorneal) remains controversial; widely different results have appeared in the literature [9]. The underlying reason for these highly variable results arises from the significant challenges presented in making such a measurement. Current methods of measuring human tear film thickness can be categorized as invasive or non-invasive. Invasive methods require the insertion of substances or objects into the tear film, such as absorbent paper discs [4] and fluorescein [10, 11] and are generally inconsistent. Non-invasive approaches are limited to various forms of interferometry. Among these, angle-dependent fringes [12, 13] are well-suited only for films thicker than the lipid, tear film layers; wavelength-dependent fringes [14, 15] can measure only at a single location each time; but thickness-dependent fringes [16] can provide a two-dimensional distribution of thickness over the full surface and for a properly chosen wavelength can properly handle films as thin as the tear layer. However, thickness-dependent fringes produce only *relative* depth information, and even the depth gradient orientation is ambiguous, unless a reference level can be provided. This approach is unsuitable for studying a flat, unchanging surface – but that is not a concern in this domain.

This paper shows how to provide a reference level (zero, corresponding to a dry lens surface) by segmenting a video frame into wet and dry regions. The larger body of work of which this paper describes only a part solves the phase ambiguity problems and the complete system produces a time-varying map of tear film depth over the eye's surface.

We present a novel contribution to pattern recognition and computer vision in this paper, motivated by the challenging problem of interpreting interferometric video of the surface of the human eye and the associated prelens tear film. The contribution addresses a difficult wet/dry texture segmentation problem – but represents a more fundamental contribution to Gaussian mixture models generally.

The popularity of Gaussian mixture models (GMM) for density estimation derives from their flexibility in representing the wide variety of distributions encountered in real applications [17-19]. Approximating the distribution of each class as a weighted sum of Gaussians, as in Subclass Discriminant Analysis (SDA) [20] or Mixture Discriminant Analysis (MDA) [21, 22], renders the resulting classifier more adaptable. When each class comprises multiple disjoint clusters (subclasses), the gain over traditional discriminant analysis techniques (*e.g.* Fishers' DA or Linear DA) can be especially significant.

2 Resilient Subclass Discriminant Analysis: Clustering

A key problem that arises in applying GMM is that of clustering the data to identify the individual component Gaussians that represent the subclasses. SDA [20] uses a fast nearest neighbor (NN) clustering method that works well even for small sample sizes – but it assumes equal subclass priors, which is unrealistic in most applications. Moreover, in using the Euclidean metric, NN is sensitive to the number of irrelevant (uninformative) features. Expectation-Maximization (EM) is a popular and powerful algorithm with good convergence properties that has found success with MDA [22] in a number of applications [19]. However, EM requires a larger training set to produce accurate estimates of the mixture parameters, and becomes unstable when the within-subclass scatter matrices approach singularity, which is typical in problems having high dimensionality and small training sets.

This paper presents a simple, reliable clustering method we call Resilient Subclass Discriminant Analysis (RSDA) that can estimate the Gaussian mixture parameters irrespective of training set size. The underlying concept is that, in each EM iteration LDA is first used to project the data onto a much lower dimensional space with maximum class separability, and the data are then clustered in this new space. When compared to the conventional EM approach, the proposed technique offers improved numerical stability because the new subclass covariance matrices are much smaller and therefore more likely to be invertible for a given sample size. Moreover, this approach often reduces the computational complexity, despite the added projection step at each iteration, because the assignment of samples to subclasses, the most computationally demanding step in conventional EM, is now accomplished on a space of far lower dimensionality.

We point out that our algorithm differs from the Gaussian parsimonious clustering models [25, 26], which are also supposed to improve robustness with respect to small

training sample size. As opposed to the adaptability of RSDA, these GMM parameterization models impose specific restrictions on the class covariance matrices and, therefore, can fit only certain classes of distributions.

2.1 RSDA Computational Procedure

We describe the RSDA clustering procedure below. In the following, N_c and N_t are the sample sizes for classes c and t , respectively, and N is the total sample size. The total number of classes is C , each with M_c subclasses.

- 1 Given an assumption for the number of subclasses (clusters) M_c in class c , use the K-means algorithm and accept the clustering results $\{d_{ji}: d_{ji} = 1 \text{ if } x_j \in \text{cluster } i; 0 \text{ else}\}$ to initialize the subclass priors, mean vectors, and covariance matrices:

$$\rho_{ci}^{[0]} = \frac{\sum_{j=1}^{N_c} d_{ji}}{N_c} \quad (1a) \quad \mu_{ci}^{[0]} = \frac{\sum_{j=1}^{N_c} d_{ji} x_j}{\sum_{j=1}^{N_c} d_{ji}} \quad (1b)$$

$$\Sigma_{ci}^{[0]} = \frac{\sum_{j=1}^{N_c} d_{ji} (x_j - \mu_{ci}^{[0]})(x_j - \mu_{ci}^{[0]})^T}{\sum_{j=1}^{N_c} d_{ji}} \quad (1c)$$

- 2 At iteration k , calculate the averaged within subclass and between subclass scatter matrices:

$$\hat{S}_W = \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^{M_c} N_c \rho_{ci}^{[k]} \Sigma_{ci}^{[k]} \quad (2)$$

$$\hat{S}_B = \frac{1}{N^2} \sum_{c=1}^{C-1} \sum_{i=1}^{M_c} \sum_{t=c+1}^C \sum_{l=1}^{M_t} N_c N_t \rho_{ci}^{[k]} \rho_{tl}^{[k]} (\mu_{ci}^{[k]} - \mu_{tl}^{[k]})(\mu_{ci}^{[k]} - \mu_{tl}^{[k]})^T \quad (3)$$

Eq. (3) differs from the traditional definition of between class scatter in that it emphasizes class separability over intra-subclass scatter. For comparison, the more traditional definition is:

$$\tilde{S}_B = \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^{M_c} N_c \rho_{ci}^{[k]} (\mu_{ci}^{[k]} - \mu^{[k]})(\mu_{ci}^{[k]} - \mu^{[k]})^T \quad (4)$$

where $\mu^{[k]}$ is the total mean vector. Note that, because we use Fisher's discriminant variates for dimensionality reduction rather than direct LDA classification, the subclass covariance matrices are not constrained to be the same so long as the dimension reduction model (DRM) and linear design condition (LDC) assumptions are valid [23].

- 3 Compute $p = (\sum_c M_c) - 1$ discriminant vectors V by solving for the generalized eigenvectors of the between subclass scatter matrix with respect to the within subclass scatter matrix:

$$\hat{S}_B V = \hat{S}_W V \Lambda \quad (5)$$

Λ is a $p \times p$ diagonal matrix of generalized eigenvalues sorted by size. We next use V to project the original (high-dimensional) samples $\{x_j\}$ onto the p -dimensional Fisher's space: $x_j' = V^T x_j$. The subclass means and sample covariance matrices are also projected onto the lower dimensional space:

$$\mu_{ci}^{[k]'} = V^T \mu_{ci}^{[k]} \quad \Sigma_{ci}^{[k]'} = V^T \Sigma_{ci}^{[k]} V$$

- 4 With the estimated parameters (means, covariances) and the projected discriminant coordinates in hand, we can calculate the probability of each sample x_j' being drawn from each subclass i in class c using Bayes' rule and the law of total probability (the E-step):

$$p(z_j = i | x_j', \theta_c^{[k]}) = \frac{\rho_{ci}^{[k]} g(x_j'; \mu_{ci}^{[k]'}, \Sigma_{ci}^{[k]'})}{\sum_{i=1}^{M_c} \rho_{ci}^{[k]} g(x_j'; \mu_{ci}^{[k]'}, \Sigma_{ci}^{[k]'})} \quad (6)$$

- 5 Now we can update the new subclass prior probabilities, mean vectors, and covariance matrices by maximizing the conditional expectation of the log-likelihood function (the M-step):

$$\begin{aligned} \rho_{ci}^{[k+1]} &= \frac{1}{N_c} \sum_{j=1}^{N_c} p(z_j = i | x_j', \theta_c^{[k]}) \\ \mu_{ci}^{[k+1]} &= \frac{1}{N_c} \sum_{j=1}^{N_c} \frac{x_j}{\rho_{ci}^{[k+1]}} p(z_j = i | x_j', \theta_c^{[k]}) \\ \Sigma_{ci}^{[k+1]} &= \frac{1}{N_c} \sum_{j=1}^{N_c} \frac{1}{\rho_{ci}^{[k+1]}} p(z_j = i | x_j', \theta_c^{[k]}) (x_j - \mu_{ci}^{[k+1]})(x_j - \mu_{ci}^{[k+1]})^T \end{aligned} \quad (7)$$

- 6 Return to Step 2 (recalculate the scatter matrices) and repeat to convergence.

As just shown, RSDA uses FDA to reduce the data dimension before each E-step based on the between and within subclass scatter matrices from the previous M-step. It is easy to show that the discriminant vectors obtained from Eq. (5) are the same as those obtained using the total covariance matrix, which remains constant throughout, in place of the average within subclass covariance matrix:

$$\hat{S}_B V = \hat{S}_T V \Lambda, \text{ where } \hat{S}_T = \hat{S}_B + \hat{S}_W$$

The null space of the total covariance matrix provides no (sub)class discrimination information [27]. Therefore, we can safely discard it prior to RSDA. As long as the subclass conditional covariance matrices in the reduced dimension space have full rank, numerical problems can be avoided entirely. It is usually far easier to meet the full rank condition in the lower dimensional space because $p \ll d$ in general. But because the estimates of the covariance matrices are biased prior to convergence, the resulting discriminant vectors are also biased. Accordingly, the class means and the covariance matrices must be updated back on the original high (d) dimensional feature space in each M step.

2.2 Computational Complexity

Relative to the conventional EM algorithm, the additional computational cost presented by RSDA arises from two stages of the procedure. First is the generalized eigenvector decomposition of the between subclass scatter matrix with respect to the within subclass scatter matrix in Step 3. This is usually solved by Cholesky factorization with a computational complexity of $O(d^3)$. The second source of additional cost is the projection of the original d -dimensional feature vectors onto the p -dimensional Fisher's discriminant space, which calls for $O(Npd)$ multiplications and additions. However, the calculation of the posterior probability in Step 4 (the E-step) is now much less costly because the Mahalanobis distance in Eq. (6) is computed in the lower dimensional space, saving $O(Nd^2 - Np^2)$ computations. Therefore, when the total sample size N is sufficiently large and the original feature space dimension d is significantly larger than the reduced one p – typical in many problems of interest, including our application (below) – substantial computational savings can be realized by RSDA vs. conventional EM.

2.3 Comparison with Other GMM Parameterization Models

RSDA improves the robustness of EM in Gaussian mixture estimation when the sample size is low relative to the dimensionality of the feature space. Other methods have been proposed to address this situation, including the so-called Parsimonious Gaussian Mixture Models [25, 26]. These models place restrictions on the class conditional covariance matrices, reducing the number of free parameters to estimate and, thereby, stabilizing the numerical process.

Our results [28, 29] in comparing these models show that RSDA – by not enforcing such constraints – is capable of effectively modeling a far greater range of data sets, with corresponding improvements in classification. Moreover, RSDA can accomplish this while also being more computationally efficient in most cases, especially with respect to those models requiring iterative calculation of adjustable diagonal and adjustable positive semidefinite symmetric matrices, respectively.

3 Segmentation of the PLTF from Narrowband Interferometry

We demonstrate the capabilities of RSDA with a set of experiments on our own problem in computer vision: texture-based segmentation of narrowband interferometry of the prelens tear film *in vivo*. We established the optimal number of

clusters per class using 10-fold cross validation. The quadratic classifier based on the Mahalanobis distance from each (sub)class centroid is then used.

Image segmentation based on visual texture is a classic problem in computer vision. In this experiment (which motivated our developing RSDA in the first place), we begin with a set of prelens tear film (PLTF) images acquired using a modified Doane's interferometer [30, 31]. This custom instrument superposes two (or more) coherent light waves, one reflected from the surface of the tear film and the other from the anterior surface of the contact lens, as shown in Fig. 1. The resulting oscillations in net reflected intensity as the waves constructively and destructively interfere, analogous to Newton's rings, indicate the thickness of the tear film as a function of position and time (we collect video), modulo half the optical wavelength.

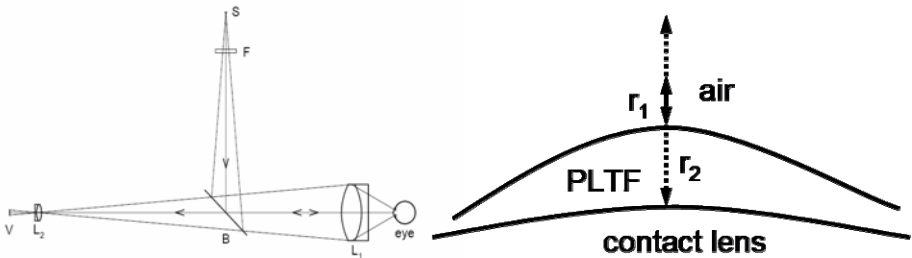


Fig. 1. [Left] Simple schematic of a modified Doane's interferometer, used to image the prelens tear film (PLTF). [Right] Principle of operation. An incident wave is reflected by two surfaces; r_1 is the reflection from the tear film surface and r_2 is the reflection from the contact lens.

Fig. 2 presents a pair of typical PLTF frames, each including dry (zero thickness) areas in which the tear film has broken up and areas that remain wet. The detection of dry areas and the assessment of the tear film thickness as a function of position and time (post-blink) is important to the study of tear physiology and fluid dynamics associated with dry eye syndromes in contact lens wearers. As we see in the images, the wet and dry regions present distinct visual textures. It follows that detecting and measuring the location and extent of tear film breakup can be formulated as a texture segmentation problem. We begin by computing a standard feature set of Gabor filter bank responses on a pixel-by-pixel basis, followed by RSDA to classify the extracted feature vectors; more detail is available in [32].

An inspection of the images reveals that wet regions are relatively uniform in texture, displaying sinusoidal fringes whose contours and spatial frequency correspond to the local gradient of the tear film surface. Our experiments show that wet regions can be classified as arising from a single Gaussian distribution. However, two distinct texture patterns exist in dry regions, arising primarily from differences in the contact lens material from one manufacturer to another, or from one lens type to another. The "dry class" is therefore multimodal, requiring multiple Gaussian subclasses to capture the complete distribution. Some form of GMM is therefore desirable. In particular, owing to the ratio of dimensionality to training data, RSDA is an excellent tool to use against this classification problem.

We determined the number of subclasses by running RSDA with different trial values for the numbers of subclasses: $M_{DRY} \times M_{WET} = \{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\}$. Increasing the number of dry subclasses from 1 to 2 improved classification performance by about 6% (e.g., from 84% to 90%); this improvement was relatively insensitive to the number of wet subclasses assumed. However, so long as $M_{DRY} \geq 2$, performance was relatively insensitive to the value of M_{WET} , ranging from a low of 89.6% to a high of 91.0% for $M_{DRY} = 4$. Although we believe this range is so small as to be statistically insignificant (and thus confirming our hypothesis of two dry subclasses), we assumed four dry subclasses for the results to follow – thus forcing RSDA to build a richer model of the dry class conditional distribution.

The classification rate for $M_{DRY} = M_{WET} = 1$, in which case RSDA reduces to LDA, was relatively poor at just 84.0%. The poorest performance among all 36 combinations was achieved with $M_{DRY} = 6$ and $M_{WET} = 1$, corresponding to over modeling the wet class conditional distribution (more subclasses than needed) while under modeling the dry (insufficient number of subclasses). But, given the observations in the preceding paragraph, the real culprit is clearly the under modeling of the dry class conditional distribution. Tests with the other parsimonious GMM parameterizations failed to achieve the same level of performance; those assuming diagonal subclass conditional covariance matrices achieved the best rate among these models at just over 88%.

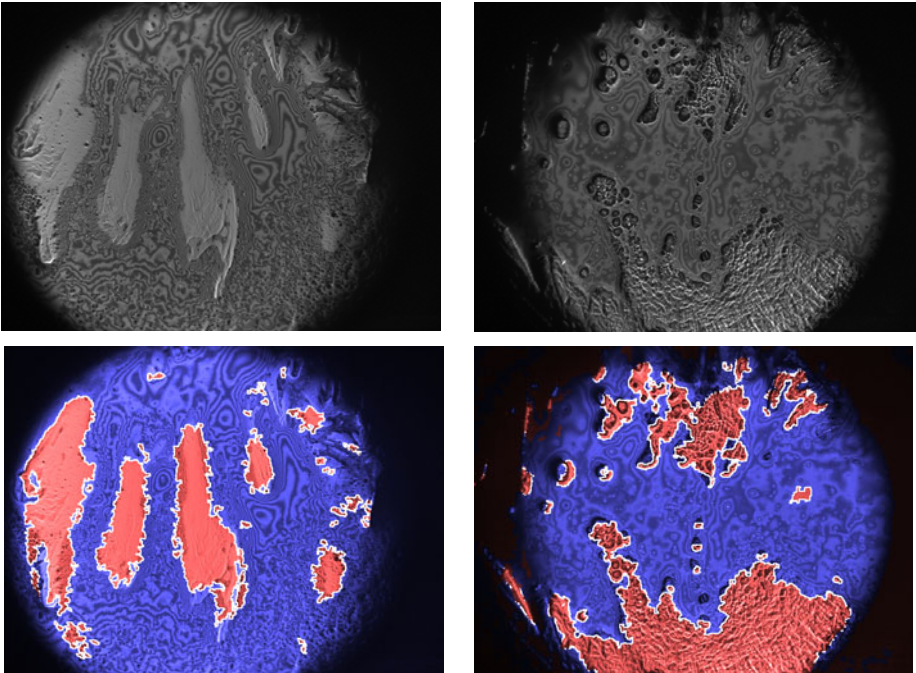


Fig. 2. PLTF interferometry. Originals (top) and segmented (bottom) using RSDA feature selection from Gabor filter outputs. Dry regions are indicated in red; wet regions in blue. Ground truth for training and classification rate calculation verified by faculty in the Ohio State University College of Optometry. Note the differences in the texture of the dry regions: very smooth for the left example and very rough for the example on the right.

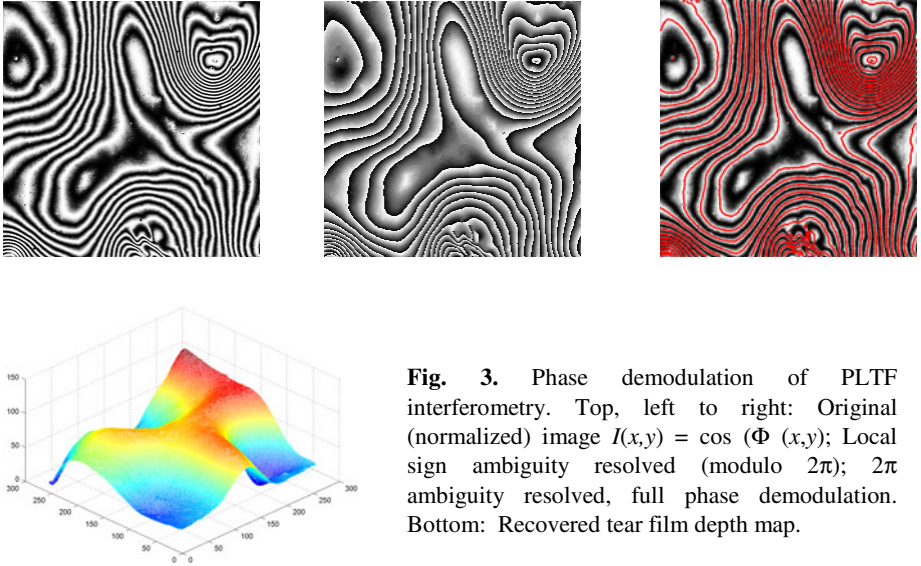


Fig. 3. Phase demodulation of PLTF interferometry. Top, left to right: Original (normalized) image $I(x,y) = \cos(\Phi(x,y))$; Local sign ambiguity resolved (modulo 2π); 2π ambiguity resolved, full phase demodulation. Bottom: Recovered tear film depth map.

4 Final Comments

Our new Gaussian clustering method is based on the simple idea of conducting the E-step in the lower dimensional Fisher space. It offers improved numerical stability compared to conventional EM for small training sets, while effectively accommodating a wider variety of distributions without restrictions on the subclass conditional covariance matrices. Using this approach, we successfully tackled the challenging problem of detecting dry spots on the surfaces of contact lens wearers *in vivo* from interferometric video. With pixel-based texture segmentation, we detect the tear film breakup from feature vectors extracted using a bank of 30 Gabor filters tuned to five different spatial frequencies and six orientations. The E-step, however, is accomplished in a four dimensional space, based on the five total subclasses (one wet, four dry), assumed for these tests. The results are quite accurate: 91.0% of pixels are correctly classified (wet or dry) using RSDA, compared to 84.0% using the well-known LDA, or 84.9% using traditional EM followed by a Bayesian classifier. A three subclass (two dry) solution, using a two dimensional Fisher space would likely work just as well.

Owing to space limitations, the discussion here has been confined to outlining RSDA and demonstrating its effectiveness in wet/dry segmentation of the prelens tear film. However, to show something about “the rest of the story” we present one reconstructed frame of a prelens tear film depth map in Fig. 3. The original normalized image is first decoded with the inverse cosine (not shown) and a sign image is computed (not shown) in an optimization step based on the reasonable assumption that the *phase gradient* (as opposed to the *image gradient*) is smooth everywhere, except at critical points. A wrap count image (not shown) is then produced in an optimization step based on the Itoh condition [33] to resolve the 2π

ambiguity and complete the phase demodulation. Scaling according to the wavelength of the light source then produces the depth map. For details and more results see [34, 35].

References

1. Wong, H., Fatt, I., Radke, C.J.: Deposition and thinning of the human tear film. *J. Colloid and Interface Science* 184(1), 44–51 (1996)
2. Berger, R.E., Corrsin, S.: A surface tension gradient mechanism for driving the pre-corneal tear film after a blink. *J. Biomechanics* 7, 225–238 (1974)
3. Sharma, A., Tiwari, S., Khanna, R., Tiffany, J.M.: Hydrodynamics of meniscus-induced thinning of the tear film. *Adv. Exp. Med. Biol.* 438, 425–431 (1998)
4. Ehlers, N.: The thickness of the precorneal tear film. *Acta Ophthalmol (Copenh)* 81, 92–100 (1965)
5. King-Smith, P.E., Fink, B.A., Fogt, N., Nichols, K.K., Hill, R.M., Wilson, G.S.: The thickness of the human precorneal tear film: Evidence from reflection spectra. *Invest. Ophthalmol. Vis. Sci.* 41(11), 3348–3359 (2000)
6. Fogt, N., King-Smith, P.E., Tuell, G.: Interferometric measurement of tear film thickness by use of spectral oscillations. *J. Opt. Soc. Am. A* 15(1), 268–275 (1998)
7. Wang, J., Fonn, D., Simpson, T.L., Jones, L.: Precorneal and pre- and postlens tear film thickness measured directly with optical coherence tomography. *Invest. Ophthalmol. Vis. Sci.* 44, 2524–2528 (2003)
8. Doughty, M., Fonn, D., Richter, D., Simpson, T., Coffrey, B., Gordon, K.: A patient questionnaire approach to estimating the prevalence of dry eye symptoms in patients presenting to optometric practices across Canada. *Optom. Vis. Sci.* 74, 624–631 (1997)
9. King-Smith, P.E., Fink, B.A., Hill, R.M., Koelling, K.W., Tiffany, J.M.: The thickness of the tear film. *Current Eye Research* 29(4-5), 357–368 (2004)
10. Benedetto, D.A., Clinch, T.E., Laibson, P.R.: In vivo observation of tear dynamics using fluorophotometry. *Arch. Ophthalmol.* 102(3), 410–412 (1984)
11. Benedetto, D.A., Shah, D.O., Kaufman, H.E.: The instilled fluid dynamics and surface chemistry of polymers in the precocular tear film. *Invest. Ophthalmol. Vis. Sci.* 14, 887–902 (1975)
12. Green, D.G., Frueh, B.R., Shapiro, J.M.: Corneal thickness measured by interferometry. *J. Opt. Soc. Am.* 65, 119–123 (1975)
13. Prydal, J.I., Artal, P., Woon, H., Campbell, F.W.: Study of human precorneal tear film thickness and structure using laser interferometry. *Invest. Ophthalmol. Vis. Sci.* 33, 2006–2011 (1992)
14. King-Smith, P.E., Fink, B.A., Fogt, N.: Three interferometric methods for measuring the thickness of the layers of the tear film. *Optom. Vis. Sci.* 76, 19–32 (1999)
15. Nichols, J.J., King-Smith, P.E.: Thickness of the pre- and post-contact lens tear film measured in vivo by interferometry. *Invest. Ophthalmol. Vis. Sci.* 44, 68–77 (2003)
16. Doane, M.G., Gelason, W.J.: Tear layer mechanics. *Clinical Contact Lens Practice*, 1–17 (1994)
17. McLachlan, G.J., Basford, K.E.: *Mixture Models: Inference and Applications to Clustering*. M. Dekker, New York (1988)
18. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical Pattern Recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(1), 4–37 (2000)

19. Ju, J., Kolaczyk, E.D., Gopal, S.: Gaussian mixture discriminant analysis and sub-pixel land cover classification in remote sensing. *Remote Sensing of Environment* 84(4), 550–560 (2003)
20. Zhu, M., Martinez, A.M.: Subclass discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(8), 1274–1286 (2006)
21. Hastie, T., Tibshirani, R.: Discriminant analysis by Gaussian mixtures. *Journal of the Royal Statistical Society, Series B (Methodological)* 58(1), 155–176 (1996)
22. Hastie, T., Tibshirani, R., Buja, A.: Flexible discriminant and mixture models. In: Kay, J., Titterton, D. (eds.) *Statistics and Neural Networks: Advances at the Interface*. Oxford University Press, Oxford (1999)
23. Fisher, R.A.: The statistical utilization of multiple measurements. *Annals of Eugenics* 8, 376–386 (1938)
24. Rao, C.R.: *Linear Statistical Inference and its Applications*, 2nd edn. Wiley Interscience, Hoboken (2002)
25. Celeux, G., Govaert, G.: Gaussian parsimonious clustering models. *Pattern Recognition* 28(5), 781–793 (1995)
26. Halbe, Z., Aladjem, M.: Model-based mixture discriminant analysis – an experimental study. *Pattern Recognition* 38(3), 437–440 (2005)
27. Zhang, S., Sim, T.: Discriminant subspace analysis: A Fukunaga-Koontz approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 29(10), 1732–1746 (2007)
28. Wu, D., Boyer, K.L.: Resilient subclass discriminant analysis. In: *12th International Conference on Computer Vision*, Kyoto, Japan, October 2009, pp. 389–396 (2009)
29. Wu, D., Boyer, K.L.: A new Gaussian clustering method for high dimensional classification problems. In: *International Conference on Pattern Recognition and Information Processing*, Minsk, Belarus (May 2009) (Invited keynote)
30. Doane, M.G.: An instrument for in vivo tear film interferometry. *Optom. Vis. Sci.* 66(6), 383–388 (1989)
31. King-Smith, P.E., Fink, B.A., Nichols, J.J., Nichols, K.K., Hill, R.M.: Interferometric Imaging of the full thickness of the precorneal tear film. *J. Opt. Soc. Am. A, Opt. Image Sci. Vis.* 23(9), 2097–2104 (2006)
32. Wu, D., Boyer, K.L.: Sign ambiguity resolution for phase demodulation in interferometry with application to prelens tear film analysis. In: *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA (June 2010)
33. Itoh, K.: Analysis of the phase unwrapping problem. *Appl. Opt.* 21(14), 2470 (1982)
34. Wu, D., Boyer, K.L.: Markov random field based phase demodulation of interferometric images. *Computer Vision and Image Understanding* 115(6), 759–770 (2011)
35. Wu, D., Boyer, K.L., Nichols, J.J., King-Smith, P.E.: Texture based prelens tear film segmentation in interferometry images. *Machine Vision and Applications* 21, 253–259 (2010)