Generalizing Common Tasks in Automated Skin Lesion Diagnosis

Paul Wighton, Tim K. Lee, Harvey Lui, David I. McLean, and M. Stella Atkins

***Abstract*—We present a general model using supervised learning and MAP estimation that is capable of performing many common tasks in automated skin lesion diagnosis. We apply our model to segment skin lesions, detect occluding hair, and identify the dermoscopic structure *pigment network*. Quantitative results are presented for segmentation and hair detection and are competitive when compared to other specialized methods. Additionally, we leverage the probabilistic nature of the model to produce receiver operating characteristic curves, show compelling visualizations of *pigment networks,* and provide confidence intervals on segmentations.**

***Index Terms*—Automated skin lesion diagnosis (ASLD), computer-aided diagnosis (CAD), dermoscopy, hair detection, melanoma, pigment network, segmentation.**

I. Introduction

FOR the past four decades, malignant melanoma has steadily increased its burden on health care in the western world. It is the seventh most common malignancy in women, the sixth most common in men and its incidence rates are increasing faster than any other cancer [1]. Since therapy for advanced melanoma is poor [2], a large emphasis is placed on regular screening and early diagnosis. In an attempt to reduce this burden, there has recently been a considerable amount of research into automated skin lesion diagnosis (ASLD) from digital dermoscopic images. Dermoscopy employs either polarized light or oil applied to the skin, which renders the stratum corneum translucent and otherwise undetectable morphological structures become visible [3]. An image is then acquired with a digital camera under consistent lighting. Proper interpretation of these dermoscopic images leads to increased clinical diagnostic accuracy [4].

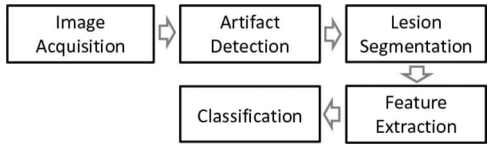


Fig. 1. Typical CAD pipeline usually adopted for ASLD. It is usually the case that supervised learning is only applied in the final stage: classification. Here, we apply supervised learning to create a generalized method capable of performing other tasks in the pipeline (artifact detection, segmentation, and feature extraction).

Most ASLD methods adopt the standard computer-aided diagnosis (CAD) pipeline illustrated in Fig. 1, which consists of five general stages. After the image is acquired, occluding artifacts (such as hair and oil bubbles) which could bias downstream processes are identified. Next, the lesion is segmented from the surrounding healthy skin. After segmentation, discriminative features are extracted from the lesion. Features are usually inspired by the ABCD rule [5] (asymmetry, border irregularity, color, and diameter). More recently, however, features have been inspired by dermoscopic diagnosis algorithms including the identification of specific patterns within the lesion known as dermoscopic structures [6]. Finally, these features are fed into a classifier and supervised learning is typically used to diagnose unseen images. Supervised learning is a general technique of estimating model parameters given a set of training examples.

It is usually the case that supervised learning is only performed in the final stage: classification; it is seldom found elsewhere in the pipeline. In the few cases where supervised learning has been applied to other stages in ASLD, it is usually done with the intent of solving one specific subproblem. For example, in [7], logistic regression is used to segment lesions; however, the model is not extended or applied to other tasks (such as hair detection or feature extraction). There is only one study in which a general supervised method is applied to various tasks in ASLD [8].

In this paper, we apply supervised learning techniques and arrive at a generalized method capable of performing many common tasks in ASLD. We apply our method to three crucial tasks: segmentation (lesion segmentation task), hair detection (artifact detection task), and the identification of the dermoscopic structure called *pigment network* (feature extraction task).

Lesion segmentation is crucial to ensure that surrounding healthy skin is not included in the computation of features, as well as to compute features off the lesion border directly [9]. We compare our method's ability to segment lesions to five other recently published methods [10]–[14] and achieve either superior or comparable results.

Manuscript received July 20, 2010; revised February 9, 2011; accepted April 21, 2011. Date of publication May 5, 2011; date of current version July 15, 2011. This work was funded by the Canadian Institutes of Health Research Skin Research Training Centre (CIHR-SRTC) and by the Canadian Health Research Program (CHRP).

P. Wighton is with the Department of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: pwighton@sfu.ca).

T. K. Lee is with the Cancer Control Research Program and Integrative Oncology Department, BC Cancer Research Centre, Vancouver, BC V5Z 4E6, Canada, with the Department of Dermatology and Skin Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada, and also with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: tlee@bccrc.ca).

H. Lui is with the Department of Dermatology and Skin Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada, with the Canadian Institutes of Health Research Skin Research Training Centre, Ottawa, ON K1A 0W9, Canada, and also with the Vancouver General Hospital Photomedicine Institute, Vancouver, V5Z 1M9, Canada (e-mail: harvey.lui@ubc.ca).

D. I. McLean is with the Dermatology and Skin Science, University of British Columbia, BC V6T 1Z4, Canada (e-mail: david.mclean@vch.ca).

M. S. Atkins is with the Department of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada and also with the Dermatology and Skin Science, University of British Columbia, BC V6T 1Z4, Canada (e-mail: stella@cs.sfu.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITB.2011.2150758

Accurately detecting occluding artifacts is also crucial to ensure biases are not induced into downstream processes. We compare our method's ability to detect hair to the popular software DullRazor [15] and achieve comparable results.

Accurately identifying the dermoscopic structure *pigment network* seems to be a promising way to enhance current diagnostic methods for two reasons. First, of all local dermoscopic structures, *atypical pigment network* is most highly correlated with malignant melanoma [16]. Second, the identification of a *pigment network* is the first stage of the two-step algorithm for differentiating between melanocytic and nonmelanocytic lesions [17]. If a lesion can be determined to be nonmelanocytic, then the diagnosis of melanoma can be ruled out entirely. This is of importance since a recent survey of automated diagnostic systems determined that all systems studied tended to "overdiagnose" most seborrheic keratoses (which are nonmelanocytic) as melanoma [18]. We report encouraging qualitative visualizations for the identification of *pigment networks*.

The remainder of this paper is organized as follows. In Section II, we review previous work. Section III describes the method. In Section IV, we present and discuss results. Finally, we conclude in Section V.

II. Previous Work

The closest related work is by Debeir *et al.* [8] in which they begin by computing 168 features for each pixel in a digital dermoscopic image (11 features computed at 5 scales across 3 channels, as well as the original RGB values). For each task, some pixels are randomly selected ($n = 2392$) from 97 images, labeled and used to train two classifiers (a decision tree and a linear discriminant model using stepwise feature selection). Tenfold cross validation is used to validate their method in each task. Their first task is lesion segmentation, where they achieved a sensitivity/specificity pair of 90%/91% and 91%/93% using decision trees and linear discriminant analysis (LDA), respectively. Although they apply their models to entire images (of size 768×512) the validation is only performed on the subset of randomly selected pixels. Next, they apply their method to a five-class problem for detecting various dermoscopic structures (diffuse pigmentation, brown globules, black dots, blue-white veil, and other). Their method remains the same and they achieve impressive results, especially after merging two difficult to discriminate classes (brown globules and black dots). While this approach proved promising, and while others have employed similar techniques, no follow-up work on a *generalized* model has been reported by either Debeir *et al.* or other research groups in over ten years. Recently, however, Celebi *et al.* [19] employed a similar supervised technique to detect the dermoscopic structure *blue–white veil* (and related structures) and achieve an overall accuracy of 82.94%.

In the rest of this section, we briefly review other previous work in hair detection, lesion segmentation, and pigment network detection paying particular attention to the use of supervised learning.

A. Lesion Segmentation

Of all the subproblems in ASLD, lesion segmentation is perhaps the most studied. Supervised methods, however, are very rarely employed. A recent survey of dermoscopic segmentation methods [20] that examines 16 articles includes only 2 which use supervised learning. Here, we review these methods [7], [21] as well as [22]. Additionally, we review the algorithms that we compare against: K-means++ (KPP) [10], J-image segmentation (JSEG) [11], dermatologist-like tumor area extraction algorithm (DTEA) [12], statistical region merging (SRM) [13], and threshold fusion (FSN) [14].

Donadey *et al.* [21] begin by using heuristics to select a point within the lesion. 1-D intensity profiles are created radially from this point on which the ground truth position of the lesion border is marked. These profiles are used to train a neural network which then predicts the lesion border on unseen profiles. They do not report any quantitative or comparative results. Tenenhaus *et al.* [7] use intensity values at multiple scales and logistic regression to segment images. They achieve an accuracy of 75%. Roberts and Claridge [22] employ many standard image processing primitives (morphological operations, logical operations, thresholding, edge filtering, etc.) and genetic programming to "evolve" segmentation algorithms. Quantitative results are reported graphically, but it appears that approximately 60% of the lesions are segmented with a sensitivity and specificity greater than 90%.

The KPP algorithm [10] begins by spatially clustering an image's pixels based on its color as well as its distance from the center of the image using the K-means++ algorithm [23]. It then finds the subset of clusters that minimizes an objective function which measures the distance between color histograms of the clusters as well as the textural gradient. The JSEG algorithm [11] first quantizes the image into 20 classes based on color. Next, scatter (ratio of interclass to intraclass variance) is computed locally at various scales. Multiscale methods are then used to merge the resulting "J-images" into a final segmentation. The DTEA segmentation algorithm [12] first obtains an initial segmentation by finding high-frequency components via filtering and the Otsu method of image thresholding [24]. This subdivides the image into many small regions which are then merged until they are of sufficient size (at least 5% of the image). A rather elaborate series of rules is then used to select the subset of regions which is considered to belong to the lesion. Finally, to mimic dermatologist's tendency to conservatively segment the lesion, the border is slightly expanded. To account for interobserver agreement in experts' segmentations, ground truth is derived via agreement from the annotations of five experts. The same research group has since proposed a metric that operates directly on multiple ground truths [25]. The SRM algorithm [13] is based on the statistical region merging algorithm [26] which iteratively tests neighboring regions to see if they should be merged based on the inter- and intraregion variance. The FSN algorithm [14] takes the results of several thresholding algorithms and fuses them using a Markov model to arrive at a final segmentation.

B. Hair Detection

The most popular hair detection method is Lee *et al.*'s freely available image processing software DullRazor [15]. DullRazor applies a grayscale morphological closing operation with a linear structuring element to detect dark hair. The operation is performed 3 times (with the structuring element oriented at

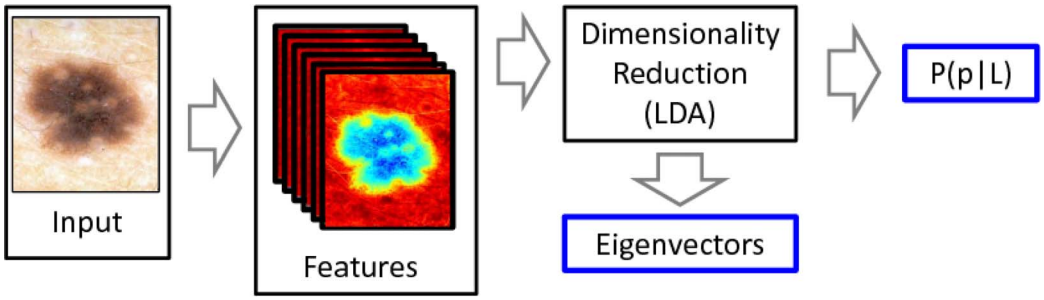


Fig. 2. Training phase of our method. Features are computed over a training set whose dimensionality is then reduced via LDA. The posterior probabilities of the labels are modeled as multivariate Gaussian distributions and saved along with the eigenvectors obtained from LDA.

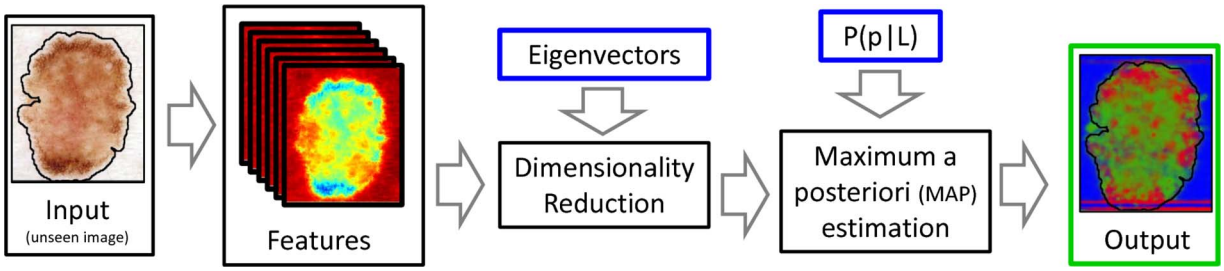


Fig. 3. Labeling phase of our method. For a previously unseen image, features are computed as before and the dimensionality of the featurespace is reduced using the eigenvectors obtained from the training phase. MAP estimation is then used along with the posterior probabilities from the training phase to label the image.

$0°, 45°$, and $90°$, respectively) and a mask is created by thresholding over the maximal difference between the 3 resulting images and the original. Such a mask is created for each channel (in RGB space) and the hair mask is obtained by a union of these three masks. Schmid-Saugeon *et al.* [27] and Xie *et al.* [28] employ similar techniques. Recently, Nguyen *et al.* [29] take the absolute value of a matched filter to detect both light and dark hairs. To date, aside from [8], there has been no other hair detection method that employs supervised learning.

C. Pigment Network Detection

There is only one known work where supervised learning has been used to detect the dermoscopic structure *pigment network*. Serrano and Acha [30] use Markov random fields in a supervised setting to classify 100 tiles (sized 40×40) that have been labeled with one of five global patterns: reticular, globular, cobblestone, homogeneous, and parallel. In the context of the study, a *reticular pattern* can be considered equivalent to *pigment network*. Using tenfold cross validation, they achieve an impressive overall accuracy of 86%, considering the difficulty of a five-class problem. It is unclear, however, how the tiles were selected. It could be the tiles were difficult, real-world examples, or that they were text-book-like definitive exemplars.

III. Method

In this section, we describe our general method designed to perform several tasks in ASLD. Our method, which is sufficiently general to identify arbitrary structures, is divided into two stages. In the training stage, parameters for multivariate Gaussian distributions of each class to be learnt are estimated. In the labeling stage, individual pixels from previously unseen images are assigned a label using MAP estimation. An overview of the training and labeling stages are illustrated in Figs. 2 and 3, respectively.

A. Training Phase

For each task, we begin with a set of N_L mutually exclusive labels $L = \{l_1, \ldots, l_n\}$ as well as a set of labeled images. Images are first converted to CIE L*a*b* and all operations are performed in this space. CIE L*a*b* was chosen because it is approximately perceptually uniform and is, therefore, a natural choice if attempting to mimic human behaviour. Additionally, CIE L*a*b* separates intensity information (in the L* channel) from color (in the a* and b* channels). Other colorspaces may yield marginal improvements; however, the emphasis of this paper is not on discriminative features, thus this was not considered. Pixel-based features are then computed over this training set. Dermoscopic images have no intrinsic orientation; therefore, all features should be rotationally invariant. Since the emphasis of this paper is on the general model rather than discriminative features, we consider two modest feature sets. The first is filtering with a series of Gaussian and Laplacian of Gaussian filters at various scales ($\sigma = \{1.25, 2.5, 5, 10, 20\}$) in each color channel (for a total of 30 features); we call this feature set G-LoG. The second is a histogram of oriented gradients [31] computed over the L* channel with nine orientations and one spatial bin over an 11×11 rectangular window. In order to make this feature set rotationally invariant, this histogram is shifted so that the dominant bin is in the first position; we call this feature set HoG. While the choice of window size was made rather arbitrarily, methods exists for determining this empirically [32]. We evaluate the effectiveness of our features in Section IV-A by comparing against the raw L*a*b* values.

After extracting features, the spatial relationship between pixels is discarded and, henceforth, will be considered to be independent; it is assumed that all relevant dependencies have been captured by the features. The dataset is now viewed as a set of labeled pixels. For larger datasets, sampling is performed on this set to make the computation tractable. Even very aggressive sampling (0.1%) was shown to have no effect on results.

LDA is then used to reduce the dimensionality of these pixels. LDA finds the subspace which maximizes the separability of the labels. This is achieved through an eigenvalue decomposition of a scatter matrix, which represents the separability of the classes (between class variance divided by within class variance) with respect to each feature. Let N_{li} represent the number of observations of class l_i, x_i^{lj} represent the ith observation (pixel) of the class l_j, μ_j the mean of class l_j, and μ the overall mean across all classes. LDA finds the eigenvectors of the scatter matrix $S_w^{-1}S_b$ where [33]

$$S_w = \sum_{j=1}^{N_L}\sum_{i=1}^{N_{lj}}(x_i^{lj}-\mu_j)(x_i^{lj}-\mu_j)^T \tag{1}$$

$$S_b = \sum_{j=1}^{N_L}(\mu_j-\mu)(\mu_j-\mu)^T. \tag{2}$$

After applying LDA, a pixel p is represented as a $1\times(N_L-1)$ vector, where N_L is the number of mutually exclusive labels. The posterior probabilities $P(p|l_i)$ in this subspace are modeled as multivariate Gaussian distributions. We evaluate the effectiveness of LDA in Section IV-A by comparing it to its unsupervised counterpart principal component analysis (PCA), we also evaluate performance when the dimensionality reduction step is skipped entirely.

B. Labeling Phase

To label an unseen image, features are computed as in the training phase and dimensionality of the featurespace is reduced using the saved eigenvectors. We then wish to estimate $P(l_i|p)$ that is the probability that a pixel is labeled l_i given the observations p. We then assign the most probable label

$$l^* = \underset{l_i\in L}{\operatorname{argmax}}\left(P(l_i|p)\right). \tag{3}$$

Applying Bayes' rule ($P(l_i|p)=P(p|l_i)P(l_i)/P(p)$), observing that $P(p)$ is constant with respect to l_i, then performing a log transformation for computational convenience, we arrive at the standard equation for maximum likelihood estimation

$$l^* = \underset{l_i\in L}{\operatorname{argmax}}\left(\log P(p|l_i)+\log P(l_i)\right). \tag{4}$$

While we could have estimated and saved the prior probabilities $P(l_i)$ in the training phase, in practice, we consider a range of priors in order to sweep out an receiver operating characteristic (ROC) curve and evaluate the method over the entire range of sensitivities. We are, therefore, performing MAP estimation. For the two-class case ($L=\{l_1,l_2\}$), an ROC curve is created by considering constraint $P(l_1)+P(l_2)=1$ and varying the values of $P(l_1)$ and $P(l_2)$ accordingly. Equivalently, one could generate an ROC curve by computing the likelihood of a class [see (5)] and considering a range of thresholds. In Section IV-A, we compare MAP estimation to the simplest of segmentation techniques: histogram thresholding.

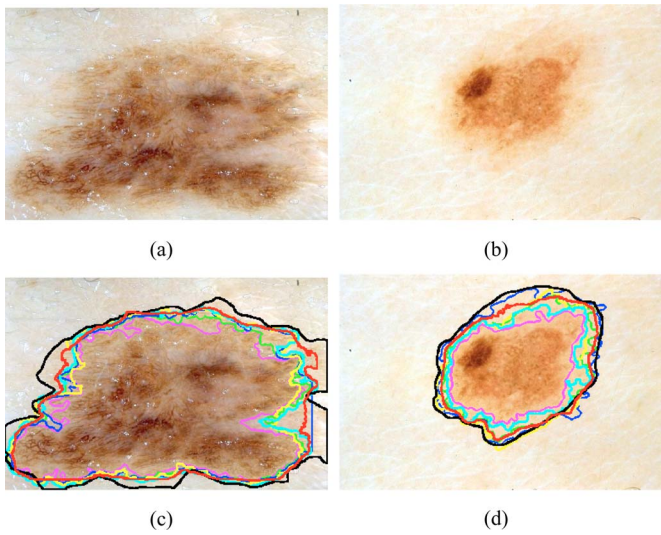


Fig. 4. Dermoscopic images (first row) and resulting segmentations (second row) comparing our method (red) to expert ground truth (black) as well as the KPP (green), JSEG (blue), DTEA (purple), SRM (yellow), and FSN (cyan) algorithms. *First column*: lesion with multiple colors and textures. *Second column*: lesion where the border is not clearly defined.

Compelling visualizations are also possible by calculating the likelihood of each label

$$\mathcal{L}_i(p) = \frac{P(p|l_i)}{\sum_{j=1}^{N_L}P(p|l_j)} \tag{5}$$

and plotting each labels' likelihood in separate color channels of the image.

IV. Results and Discussion

We now apply our method to three common tasks in ASLD: lesion segmentation, hair detection, and the detection of the dermoscopic structure *pigment network*.

A. Lesion Segmentation

To begin, we apply our model to segment lesions. A dataset is created using images from atlases [34], [35]. Eighteen typical as well as 100 challenging images are selected. Challenging images are ones that are sometimes excluded from other studies [11]. An image is considered challenging if one or more of the following conditions is met: 1) the contrast between the skin and lesion is low; 2) there is significant occlusion by either oil or hair; 3) the entire lesion is not visible; 4) the lesion contains variegated colors; or 5) the lesion border is not clearly defined. Examples of challenging images are shown in Fig. 4. Each image is segmented by a dermatologist and pixels are labeled from the set $L=\{$"lesion," "background"$\}$. The G-LoG feature set is used and tenfold cross validation is employed to segment the dataset. LDA, as well as the entire testing phase, is included within the cross validation loop.

We compare our method to five other lesion segmentation techniques: KPP [10], JSEG [11], DTEA [12], SRM [13], and

TABLE I
COMPARISON OF OUR METHOD'S ABILITY TO SEGMENT LESIONS TO FIVE PREVIOUSLY PUBLISHED METHODS

		Theirs		Ours (nearest pt.)	
Method	n	Sens	Spec	Sens	Spec
KPP [10]	118	0.765	0.770	0.868	0.843
JSEG [11]	93	0.627	0.987	0.621	0.973
DTEA [12]	118	0.597	0.986	0.596	0.977
SRM [13]	114	0.790	0.946	0.776	0.925
FSN [14]	118	0.808	0.934	0.793	0.915

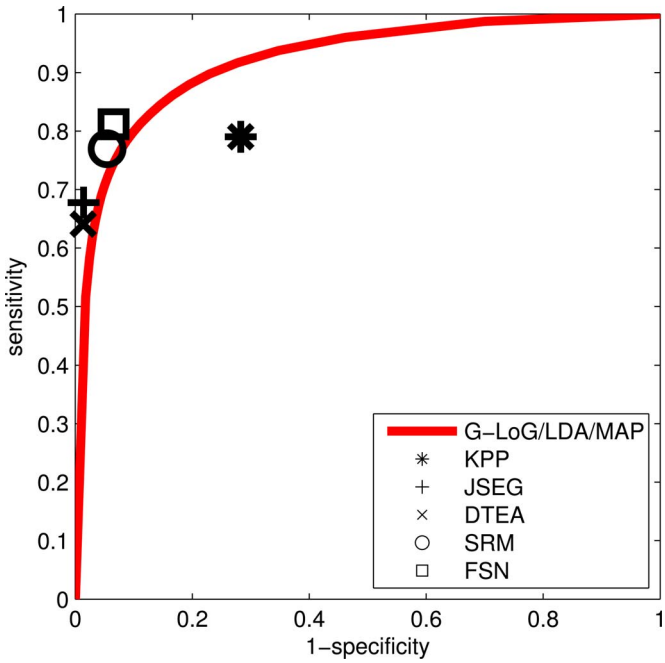


Fig. 5. ROC curve comparing our method's ability to segment lesions to five other published methods: KPP [10], JSEG [11], DTEA [12], SRM [13], and FSN [14].

FSN [14]. We present sample segmentations in Fig. 4 as well as quantitative results in Table I and Fig. 5. Since the other algorithms return binary segmentations, we compare the resulting sensitivity/specificity points, to the closest point on the ROC curve for our method. Sensitivity is defined as the fraction of correctly labeled lesion pixels. Specificity is defined as the fraction of correctly labeled background pixels.

The JSEG, DTEA, and our algorithm give comparable results with the JSEG algorithm performing marginally better; however, the JSEG algorithm was only able to segment 93 of the 118 images, presumably due to the difficulty of the dataset. The SRM algorithm slightly outperforms our method; however, it was only able to segment 114 of the 118 images. The FSN algorithm performed comparably to SRM; however, unlike SRM it managed to segment all 118 images. Our method significantly outperforms the KPP algorithm, which appears to perform poorly due to occluding hair. Results from the KPP, JSEG, DTEA, and SRM algorithms were generously provided by Professor Emre Celebi.

Additional improvements to this method are achieved by using resulting likelihood maps to initialize the random walker

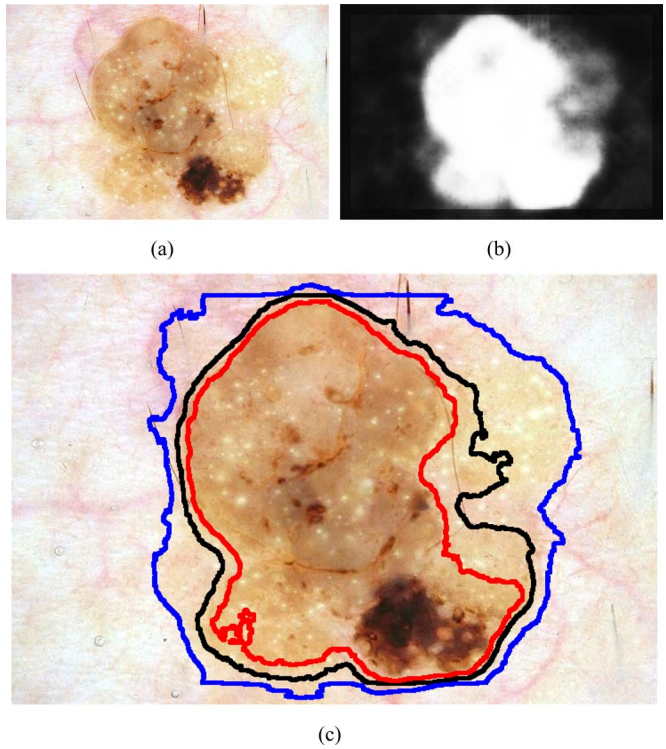


Fig. 6. Confidence Interval based segmentation. (a) Dermoscopic image of a seborrheic keratosis. (b) Resulting likelihood map ($\mathcal{L}_{\text{lesion}}$) c) MAP segmentation (black: $\mathcal{L}_{\text{lesion}} \geq 0.5$) and 80% confidence interval segmentation (blue: $\mathcal{L}_{\text{lesion}} \geq 0.1$; red: $\mathcal{L}_{\text{lesion}} \geq 0.9$).

algorithm [36]. The details and subsequent results are not reported here, but can be found in [37].

We also examine the computational time required. A standard image in this dataset is 750 $\times$ 500 pixels. For a typical PC operating on images of this size, filtering, and sampling takes 75 s per image. Performing LDA and estimating the posterior probabilities takes less than a second, and labeling a previously unseen image takes 10 s. Both the filtering/sampling stage and the labeling stage are amenable to parallelization.

Furthermore, due to the probabilistic nature of the model, confidence intervals on segmentations are possible by computing $\mathcal{L}_{\text{lesion}}$ and thresholding accordingly. For example, an 80% confidence interval segmentation can be realized by considering the segmentations $\mathcal{L}_{\text{lesion}} \geq 0.1$ and $\mathcal{L}_{\text{lesion}} \geq 0.9$. An example of this is shown in Fig. 6. Due to the diagnostic importance of the characteristics of a lesion's border, this appears to be an elegant way to explicitly define a "border region" from which subsequent border features can be extracted.

Next, we evaluate the relative contribution of different aspects of our model by systematically substituting components. The full model is named G-LoG/LDA/MAP since it employs the Gaussian and Laplacian of Gaussian filters, uses LDA for dimensionality reduction and MAP estimation to perform the labeling. We evaluate the effect of LDA by comparing it to PCA (G-LoG/PCA/MAP). For this task, LDA reduces the dimensionality of the featurespace to 1 (since $N_L = 2$). We, therefore, apply PCA and use only the single largest mode of variation. We also evaluate performance when the dimensionality reduction

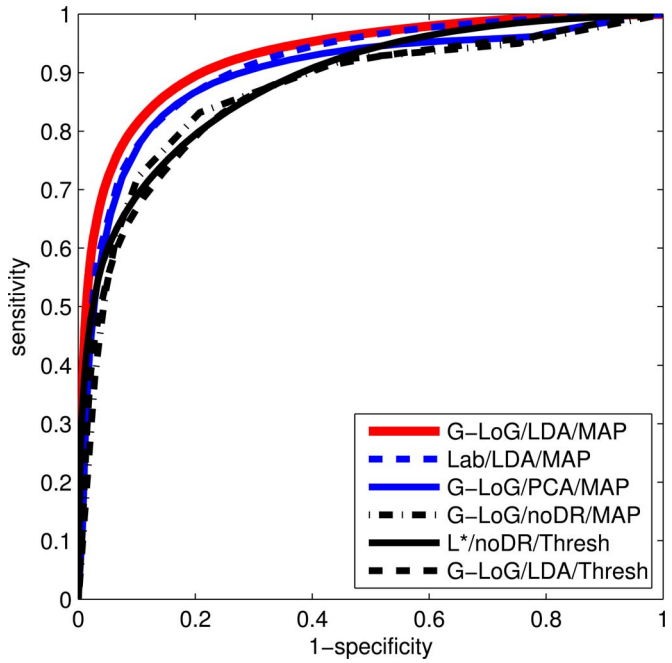


Fig. 7. ROC curves of the relative contribution of various aspects of the model for the segmentation task; see text for details.

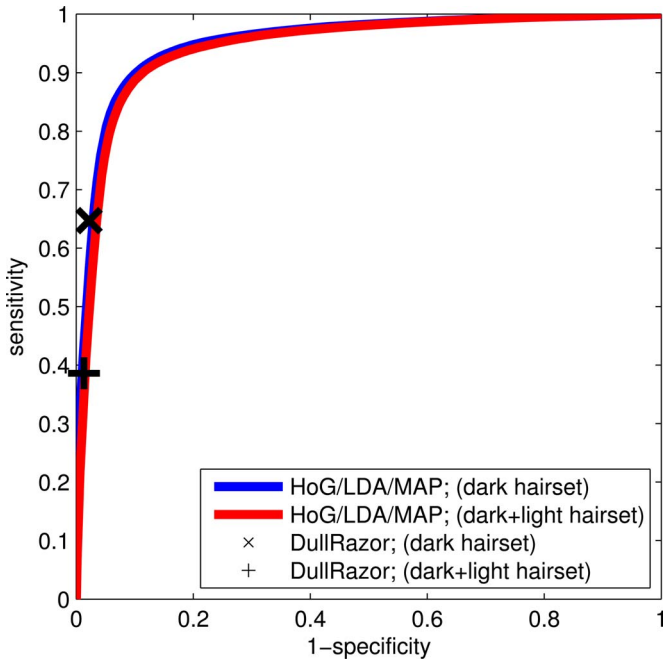


Fig. 8. Comparison of our method's (curves) ability to identify occluding hair to DullRazor [15]. DullRazor only identifies dark hair, hence the large discrepancy in sensitivity between the two datasets.

TABLE II
SYSTEMATIC COMPARISON OF VARIOUS ASPECTS OF OUR MODEL VIA SUBSTITUTION OF VARIOUS TECHNIQUES FOR THE SEGMENTATION TASK

Model	AUC	% of G-LoG/LDA/MAP
G-LoG/LDA/MAP	0.922	1.000
G-LoG/PCA/MAP	0.889	0.964
G-LoG/noDR/MAP	0.833	0.903
G-LoG/LDA/Thresh	0.858	0.931
Lab/LDA/MAP	0.906	0.983
L*/Thresh	0.879	0.953

TABLE III
COMPARING OUR METHOD'S ABILITY TO IDENTIFY OCCLUDING HAIR TO DULLRAZOR [15]

Method	Data set	n	Sens	Spec	AUC
Ours	Dark + Light	32	0.377	0.991	0.941
Ours	Dark Only	32	0.648	0.965	0.942
Drazor [15]	Dark + Light	32	0.386	0.987	-
Drazor [15]	Dark Only	32	0.647	0.978	-

(a) (b) (c)

Fig. 9. Comparison of our method to DullRazor. (a) Sample phantom hair image (L* channel). (b) Sample result from our method. (c) Sample result from DullRazor.

step is removed entirely (G-LoG/noDR/MAP). Next, we evaluate the contribution of MAP estimation by comparing the model to one that simply employs thresholding (G-LoG/LDA/Thresh). We then evaluate the contribution of the featurespace by comparing the model to one that simply operates on the L*a*b* values of each pixel (Lab/LDA/MAP). Finally, we compare to the simplest segmentation algorithm: thresholding of the intensity channel (L*/thresh). Results are presented in Fig. 7 and Table II.

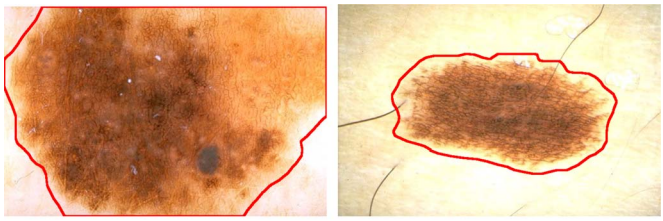
This relative evaluation admits several conclusions: 1) our final model (G-LoG/LDA/MAP) outperforms similar counterparts; 2) the dimensionality reduction step in the model is essential—removing it causes a drastic reduction in accuracy; and 3) for the task of lesion segmentation, textural, or spatial information (as captured by the G-LoG feature set) is not very useful as it barely outperforms the simpler L*a*b* feature set.

B. Hair Detection

To evaluate our method's ability to detect occluding hair, we employ dermoscopic images overlaid with randomly generated phantom hairs [38] and compare our results to DullRazor [15]. We consider two sets of 32 images. The first imageset contains both light and dark colored hair. Since DullRazor is only capable of detecting dark hairs, a second imageset is considered consisting of entirely dark colored hair. Each pixel is labeled from the set $L = \{$"hair," "background"$\}$. The HoG feature set is used and fourfold cross validation is employed to label images.

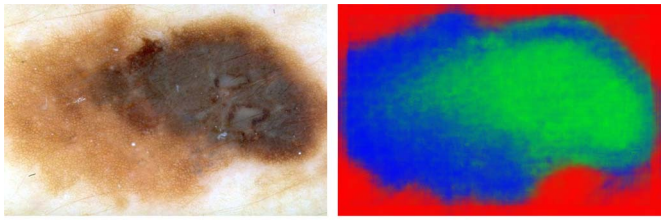
Fig. 8 compares the accuracy of our method to DullRazor and quantitative results are presented in Table III. DullRazor returns a binary mask; therefore, AUC is not reported. Reported sensitivity/specificity pairs for our method is the point on the ROC curve closest to those of DullRazor. Although DullRazor outperforms our method on both datasets, as in the segmentation task, the difference in performance is quite small.

Fig. 9 illustrates a typical result. DullRazor contains a post-processing step that eliminates any candidate regions based on



(a) (b)

Fig. 10. Sample images from the *pigment network* training set along with segmentations. Images were chosen because the structure occurs throughout the lesion. Pixels outside the lesion are assigned the label "background" while pixels inside the lesion are assigned the label "present" if the image contains a pigment network or "absent" otherwise.



(a) (b)

Fig. 11. Qualitative results of our method's ability to identify the dermoscopic structure *pigment network*. *First column*: original dermoscopic images. *Second column*: red, green, and blue channels encode the likelihood that a pixel is labeled as "background," "absent," and "present," respectively.

shape which explains the "smoother" results obtained by DullRazor. Incorporating an appropriate shape prior [39] into our model would be an elegant way to replicate this behavior within our framework.

C. Pigment Network Detection

The final task considered applies our method to detect the dermoscopic structure *pigment network*. A pigment network is defined as a "light- to dark-brown network with small, uniformly spaced network holes and thin network lines distributed more or less regularly throughout the lesion and usually thinning out at the periphery" [16]. The dataset consists of 734 images from a dermoscopy atlas [34]. Labels of either "absent" or "present" for the structure *pigment network* are derived from the atlas. Since pixel-based ground truth is not available, a custom training set is created consisting of 20 images where the pigment network is present across the entire lesion and 20 images absent of pigment network. Examples of training images containing a *pigment network* are shown in Fig. 10. Pixels are assigned a label from the set $L = \{$"background," "absent," "present"$\}$ as follows: for each image, pixels outside the segmentation are assigned the label "background," while pixels inside the segmentation are assigned either the label "absent" or "present." By considering these three labels, we are simultaneously segmenting the lesion and detecting the structure *pigment network*. The G-LoG feature set is employed. Since in this task, ground truth is not defined on a per-pixel basis, we cannot objectively evaluate performance. We, therefore, present visual results in Fig. 11 by plotting $\mathcal{L}_{\text{background}}$, $\mathcal{L}_{\text{absent}}$, and $\mathcal{L}_{\text{present}}$ in the red, green, and blue channels, respectively. We are currently in the process of creating a expertly annotated dataset to provide quantitative validation.

V. Conclusion and Future Work

We have presented a model based on supervised learning and MAP estimation that is sufficiently powerful and general to perform competitively on a variety of common tasks in ASLD. Since the model employs supervised learning, it can quickly be applied to a variety of tasks and the resulting model parameters are guaranteed to be optimal.

Although the model yields competitive results, there are many directions in which it could be extended in order to obtain even further improvements. Most hair detection and lesion segmentation methods contain a "post-processing" step which attempts to minimize any local discontinuities or noise. Extending the model to include a suitable shape prior [39] seems to be a promising way to generalize this "post-processing" step within the context of supervised learning. Alternatives to the MAP model could be explored such as the use of Markov models [40] to weaken the assumption of the independence of pixels. Nonlinearities (such as those provided by SVMs) can be incorporated by applying the so called kernel-trick to LDA [41] to expand the set of concepts the model is capable of learning.

Finally, it is hoped the model will be successfully applied to other tasks in ASLD such as the detection of black frames, oil bubbles, other dermoscopic structures, as well as to other problem domains entirely.

Acknowledgment

The authors would like to thank Prof. G. Argenziano for making his digital dermoscopy images available to us.

References

[1] C. Erickson and M. Driscoll, "Melanoma epidemic: Facts and controversies," *Clin. Dermatol.*, vol. 28, no. 3, pp. 281–286, 2010.
[2] M. Lens and M. Dawes, "Global perspectives of contemporary epidemiological trends of cutaneous malignant melanoma," *Br. J. Dermatol.*, vol. 150, no. 2, pp. 179–185, 2004.
[3] M.-L. Bafounta, A. Beauchet, P. Aegerter, and P. Saiag, "Is dermoscopy (Epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests," *Arch. Dermatol.*, vol. 137, no. 10, pp. 1343–1350, 2001.
[4] M. Vestergaard. P. Macaskill *et al.*, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting," *Br. J. Dermatol.*, vol. 159, no. 3, pp. 669–676, 2008.
[5] R. Friedman, D. Rigel, and A. Kopf, "Early detection of malignant melanoma: The role of physician examination and self-examination of the skin," *CA: A Cancer J. Clinicians*, vol. 35, no. 3, pp. 130–151, 1985.
[6] H. Pehamberger, A. Steiner, and K. Wolff, "*In vivo* epiluminescence microscopy of pigmented skin lesions. I: Pattern analysis of pigmented skin lesions," *J. Amer. Acad. Dermatol.*, vol. 17, no. 4, pp. 571–583, 1987.
[7] A. Tenenhaus, A. Nkengne, J. Horn, C. Serruys, A. Giron, and B. Fertil, "Detection of melanoma from dermoscopic images of naevi acquired under uncontrolled conditions," *Skin Res. Technol.*, vol. 16, no. 1, pp. 85–97, 2009.
[8] O. Debeir, C. Decaestecker, J. Pasteels, I. Salmon, R. Kiss, and P. Van Ham, "Computer-assisted analysis of epiluminescence microscopy images of pigmented skin lesions," *Cytometry*, vol. 37, no. 4, pp. 255–266, 1999.
[9] T. K. Lee, D. I. McLean, and M. S. Atkins, "Irregularity index: A new border irregularity measure for cutaneous melanocytic lesions," *Med. Image Anal.*, vol. 7, no. 1, pp. 47–64, 2003.

[10] H. Zhou, M. Chen, L. Zou, R. Gass, L. Ferris, L. Drogowski, and J. Rehg, "Spatially Constrained Segmentation of Dermoscopy Images," in *Proc. 5th IEEE Int. Symp. Biomed. Imag.: Nano Macro*, 2008, pp. 800–803.
[11] M. E. Celebi, Y. Aslandogan, W. Stoecker, H. Iyatomi, H. Oka, and X. Chen, "Unsupervised border detection in dermoscopy images," *Skin Res. Technol.*, vol. 13, no. 4, pp. 454–462, 2007.
[12] H. Iyatomi, H. Oka, M. E. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa, "An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Comput. Med. Imag. Graph.*, vol. 32, no. 7, pp. 566–579, 2008.
[13] M. E. Celebi, H. Kingravi, H. Iyatomi, Y. Aslandogan, W. Stoecker, R. Moss, J. Malters, J. Grichnik, A. Marghoob, and H. Rabinovitz, "Border detection in dermoscopy images using statistical region merging," *Skin Res. Technol.*, vol. 14, no. 3, pp. 347–353, 2008.
[14] M. Celebi, S. Hwang, H. Iyatomi, and G. Schaefer, "Robust border detection in dermoscopy images using threshold fusion," in *Proc. 17th IEEE Int. Conf. Image Process.*, 2010, pp. 2541–2544.
[15] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "Dullrazor®: A software approach to hair removal from images," *Comput. Biol. Med.*, vol. 27, no. 6, pp. 533–543, 1997.
[16] G. Argenziano, H. P Soyer *et al.*, "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the internet," *J. Amer. Acad. Dermatol.*, vol. 48, no. 5, pp. 679–693, 2003.
[17] A. Perrinaud, O. Gaide, L. French, J.-H. Saurat, A. Marghoob, and R. Braun, "Dermoscopy of pigmented skin lesions," *J. Amer. Acad. Dermatol.*, vol. 52, no. 1, pp. 109–121, 2005.
[18] R. P. Braun, H. S. Rabinovitz, M. Oliviero, A. W. Kopf, and J.-H. Saurat, "Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems," *Br. J. Dermatol.*, vol. 157, no. 5, pp. 926–933, 2007.
[19] M. E. Celebi, H. Iyatomi, W. Stoecker, R. Moss, H. Rabinovitz, G. Argenziano, and H. Soyer, "Automatic detection of blue-white veil and related structures in dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 32, no. 8, pp. 670–677, 2008.
[20] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. Stoecker, "Lesion border detection in dermoscopy images," *Comput. Med. Imag. Graph.*, vol. 33, no. 2, pp. 148–153, 2009.
[21] T. Donadey, C. Serruys, A. Giron, G. Aitken, J. Vignali, R. Triller, and B. Fertil, "Boundary detection of black skin tumors using an adaptive radial-based approach," in *Proc. SPIE Med. Imag. Conf.*, 2000, vol. 3979, pp. 810–816.
[22] M. Roberts and E. Claridge, "An artificially evolved vision system for segmenting skin lesion images," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.,* 2003, pp. 655–662.
[23] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.
[24] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[25] M. E. Celebi, G. Schaefer, H. Iyatomi, W. Stoecker, J. Malters, and J. Grichnik, "An improved objective evaluation measure for border detection in dermoscopy images," *Skin Res. Technol.*, vol. 15, no. 4, pp. 444–450, 2009.
[26] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1452–1458, Nov. 2004.
[27] P. Schmid-Saugeon, J. Guillod, and J. Thiran, "Towards a computer-aided diagnosis system for pigmented skin lesions," *Comput. Med. Imag. Graph.*, vol. 27, no. 1, pp. 65–78, 2003.
[28] F. Xie, S. Qin, Z. Jiang, and R. Meng, "PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma," *Comput. Med. Imag. Graph.*, vol. 33, no. 4, pp. 275–282, 2009.
[29] N. Nguyen, T. Lee, and M. Atkins, "Segmentation of light and dark hair in dermoscopic images: A hybrid approach using a universal kernel," in *Proc. SPIE Med. Imag. Conf.*, 2010, vol. 7623, pp. 76 234N-1–76 234N-8.
[30] C. Serrano and B. Acha, "Pattern analysis of dermoscopic images based on Markov random fields," *Pattern Recognit.*, vol. 42, no. 6, pp. 1052–1057, 2009.
[31] N. Dalai, B. Triggs, I. Rhone-Alps, and F. Montbonnot, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 886–893.
[32] B. Hong, S. Soatto, K. Ni, and T. Chan, "The scale of a texture and its application to segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*,, 2008, pp. 1–8.
[33] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.
[34] G. Argenziano, H. Soyer *et al.*, *Interactive Atlas of Dermoscopy* [Book and CD-ROM]. Milan, Italy: Edra medical publishing and new media, 2000.
[35] H. Soyer, G. Argenzian *et al.*, *Dermoscopy of Pigmented Skin Lesions. An Atlas based on the Consesnsus Net Meeting on Dermoscopy*. Milan, Italy: Edra medical publishing and new media, 2000.
[36] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
[37] P. Wighton, M. Sadeghi, T. Lee, and M. Atkins, "A fully automatic random walker segmentation for skin lesions in a supervised setting," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, 2009, pp. 1108–1115.
[38] N. H. Nguyen, "A hybrid approach to segmenting hair in dermoscopic images using a universal kernel," Master's thesis, Dept. Comput. Sci., Simon Fraser University, Burnaby, B.C. Canada, 2009.
[39] G. Charpiat, O. Faugeras, and R. Keriven, "Shape statistics for image segmentation with prior," in *Proc. 2007 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1–6.
[40] S. Li, "Markov random field models in computer vision," in *Proc. 3rd Eur. Conf. Comput. Vis.*, 1994, pp. 361–370.
[41] D. Cai, X. He, and J. Han, "Efficient kernel discriminant analysis via spectral regression," in *Proc. 7th IEEE Int. Conf. Data Mining*, 2007, pp. 427–432.

Paul Wighton received the B.Sc. degree in engineering from the University of Guelph, ON, Canada, in 2003. He is currently working toward the Ph.D. degree in computing science at Simon Fraser University, Burnaby, BC, Canada. His thesis focuses on methods to improve automated skin lesion diagnosis.

Tim K. Lee received the Ph.D. degree in computer science from Simon Fraser University Burnbay, Canada, in 2001.

He is currently a Senior Scientist at the Cancer Control Research Program, BC Cancer Agency, Vancouver, Canada, an Assistant Professor in the Department of Dermatology and Skin Science, University of British Columbia, Vancouver, and an Adjunct Professor in the School of Computing Science, Simon Fraser University. His research interests include computer-aided diagnosis and prevention of skin cancer.

Harvey Lui received the B. Sc., M.D. and FRCPC degrees from the University of British Columbia, Canada.

He is a Professor and the Head of the Department of Dermatology and Skin Science at University of British Columbia, Vancouver, Canada, the Director of the Canadian Institutes of Health Research Skin Research Training Centre, Ottawa, ON, Canada, and the Vancouver General Hospital Photomedicine Institute, Vancouver, Canada. He has been the principal investigator in over 35 research projects and has published over 90 journal articles and over 100 abstracts. His clinical and research interests include lasers, photomedicine, skin cancer, and pigmentary disorders.

David I. McLean studied microbiology at the University of Manitoba, and received the M.D. and FRCPC degrees at the University of British Columbia, Canada.

He is a Professor of Dermatology and Skin Science at the University of British Columbia, Vancouver, Canada. He is the former Head of Dermatologic Oncology at the BC Cancer Agency, Vancouver, where he is now the Head of Cancer Prevention. His clinical practice has been focused on pigmented lesions, including familial dysplastic nevi and melanoma.

M. Stella Atkins received the B.Sc. degree in chemistry from Nottingham University, Nottingham, U.K. in 1966 and the Ph.D. degree in computer science from the University of British Columbia, Vancouver, Canada, in 1985.

She is a Professor in the School of Computing Science and the Director of the Medical Computing Laboratory at Simon Fraser University, Vancouver. Her research interests include medical image display and analysis, particularly for skin imaging, image denoising and perception, radiology workstation design, and telehealth applications. She is also interested in the use of eyetrackers for improving surgery training.