Towards Early-Stage Malignant Melanoma Detection Using Consumer Mobile Devices

John Breneman
Stanford Center for Professional Development
jbrenem@gmail.com

Abstract—Digital dermatoscopy techniques for diagnosing malignant melanoma are adapted to consumer-grade mobile devices with macro photography lenses and a prototype application for Android devices is presented.

Keywords—dermatoscopy; mobile image processing; segmentation; biomedical image processing

I. INTRODUCTION

Melanoma is the most deadly variety of skin cancer. Although less common than other skin cancers, it is responsible for the majority of skin cancer related deaths globally [3]. Most cases are curable if detected early and several standardized screening techniques have been developed to improve the early detection rate [1], [2]. Such screening techniques have proven useful in clinical settings for screening individuals with a high risk for melanoma, but there is considerable debate on their utility among large populations due to the high workload on dermatologists [3] and the subjectivity in the interpretation of the screening [11].

In addition to deriving a set of computer vision algorithms to automate popular skin self-examination techniques, this project developed a mobile phone application that provides a pre-screening tool for individuals in the general population to help assess their risk. No computer application can provide a concrete diagnosis, but it can help inform the individual and raise the general awareness of this dangerous disease.

A. Skin-Self Evaluations using the ABCDE method

Studies have shown that self-performed skin examinations can greatly improve early detection and survivability rates of melanoma [11]. The most established method for skin self-examinations to date is the “ABCDE” promoted by the American Academy of Dermatology [7]. A detailed tutorial for conducting skin self-exams including example images for each feature is available in [7]. The “ABCDE” test provides a widely accepted, standardized set of lesion features to examine. The features are designed for members of the general public, but variability in the interpretation of the features weakens the overall utility of the test [11].

B. Image Processing for Digital Dermatoscopy and Digital Macro Photography

Epiluminescence Microscopy (ELM), also known as dermatoscopy, is a noninvasive technique for improving the early detection of skin cancer [6]. In dermatoscopy, a set of polarized light filters or oil immersion render selected epidermal layers transparent and macro lenses magnify small features not visible to the naked eye. Most dermatoscopes also include features to control lighting and focal conditions. Dermatoscopy is frequently combined with digital imaging technology and a large body of research is devoted to developing computerized processing techniques operating on the digital images produced. An adaptation of the “ABCDE” method for skin self-examinations to dermatoscopic images was first presented in 1994 [1].

The digital images processed in this project differ from conventional digital dermatoscopy in that they do not employ optical filters. Instead, they operate on the visible light images produced by standard digital cameras equipped with macro lenses for viewing objects at close range. In this sense they are analogous to the conventional skin self-exam. In general, macro photography images of skin lesions are taken under loosely controlled lighting and focal conditions. Many of the images acquired using less expensive consumer-grade macro lenses also exhibit a certain degree of barrel distortion toward the edges of their field of view.

III. RISK SCORE CALCULATION

The techniques used to compute relative risk scores follow an image-processing pattern similar to that of any feature extraction algorithm. Images are preprocessed, a region of interest (feature) is identified, and then individual feature descriptors are run on the region of interest. The results from a
MATLAB implementation of the score computation were analyzed by a group of trained dermatologists.

A. Preprocessing

Once a magnified image of a skin lesion is captured it is passed to a preprocessor. The preprocessor performs global image binarization via Otsu’s method [10]. Following binarization, a connected components analysis is performed and small region removal for both positive and negative regions removes most of the image noise. A sample output from the preprocessing stage is shown in Figure 3.

A modified version of the preprocessing algorithm was required for the live mobile application images and is described in Section IV below.

B. Image Segmentation

Following binarization and denoising in the preprocessor, the region representing the skin lesion is segmented from the rest of the image. There are many techniques in the technical literature for extracting lesion regions in digital images, including a difference-of-Gaussians (DoG) and support-vector machine (SVM) approach in [4]. However, both the static test images and the images acquired live by the mobile application were taken under controlled conditions by an expert user and did not require such complex segmentation techniques. Instead, the segmentation algorithm searches for the largest remaining connected component. Most medical imaging is designed to maximize the percentage of an image’s field of view devoted to its subject, so the largest component in the preprocessed images is nearly always the lesion. Occasionally large shadows in the images were mistaken for a lesion so a maximum allowable lesion size restriction was imposed.

C. Morphological Feature Descriptors

After preprocessing and segmentation, the images are ready for analysis by the individual feature descriptors. Three feature descriptors inspired by the “ABCDE” skin self-examination form the core of the risk assessment. Those features are asymmetry, border, and color.

1) Asymmetry

In [7], a lesion is considered potentially cancerous if “one half is unlike the other half.” This guidance is relatively vague, so techniques developed for dermatoscopy were used for inspiration. The asymmetry score calculation is based on the symmetry map technique presented in [5]. Symmetry maps encode a measure of a region’s symmetry, known as a symmetry metric, relative to a range of axes of symmetry defined by angle. Lesion color and texture comparisons were used to encode symmetry in [5]. Commonly the symmetry metric is a function of distance $R$ from a region’s center. In [5], a symmetry map is created for the range of symmetry axes passing through a region’s center with angles ranging from 0 to 180 degrees. An example symmetry map taken from [5] is shown below in Figure 4. In this map both the texture and color symmetry metrics are shown. To derive a scalar symmetry score from the symmetry map, the global maximum is used.

The symmetry map technique is attractive because it is able to achieve a degree of rotational invariance via the $max$ operator. However, calculating symmetry maps with such a high resolution in angles is computationally expensive and color and texture can vary depending on the image’s lighting and focus. Lighting and focus are not traditionally major
factors in dermatoscopy but they have a large impact in macro photography. To streamline the symmetry scoring process a symmetry metric more suitable to the loosely controlled imaging conditions of this project was selected and a “subsampled” symmetry map approach was adopted.

Unlike color or texture, borders exhibit a better invariance to imaging conditions. To exploit this property an inclusion metric was devised. For each pixel belonging to a lesion region, a counter was incremented if that pixel’s mirror about a specified symmetry axis was also included within the lesion region. After each pixel was examined the inclusion counter was normalized by the number of pixels in the lesion region. The inclusion metric is analogous to a probability score between 0 and 1 that a given pixel will have a mirror. It is also not a function of distance R from the region’s center. A diagram illustrating the inclusion calculation is shown in Figure 5.

The inclusion symmetry metric was computed for the axes passing through a region’s centroid at angles of 0, 45, and 90 degrees as shown in to form a “subsampled” symmetry map. At these angles, the inclusion metric can be computed quickly using matrix index arithmetic and a binary mask of the region of interest. As with the conventional symmetry maps, the maximum symmetry score over all three angles was taken as the lesions symmetry. For convenience, the score was converted to units of asymmetry by subtracting it from 1. In general a lesion with a high asymmetry score is at a higher risk of being cancerous.

2) Border
In [7] the shape and strength of a region’s border are considered collectively when assessing risk but the automated algorithm examines only border strength. This is because the simple segmentation techniques used were a relatively noisy measure of a lesion’s boundary and the segmentation noise quickly corrupts any border shape metric. However, border strength is relatively easy to compute. Using the gradient-of-Gaussian edge detection kernel developed by Canny in [8], a smoothed map of an image’s intensity gradient is computed via 2D convolution. The intensity gradient map can also be computed using a two-stage filter combination of Sobel and Gaussian kernels. An example image gradient map from the live feed of a mobile phone is shown in Figure 7.

Once the image gradient map is computed, the gradient magnitude values at each pixel along the lesion’s border are summed and normalized by the border’s size to calculate the average gradient magnitude along the lesion’s border. This average gradient metric forms the border strength risk value. In general lesions with poorly defined borders (ie: with low average gradient magnitude) are a higher cancer risk.

Proper choice of the Gaussian smoothing kernel is important given the relative inaccuracy of the lesion segmentation. If too small a kernel is used, the border pixels may not fall directly over pixels with a high gradient magnitude as in Figure 8. If a large kernel is used the effects of the local gradient will be diluted. For a 600x800 pixel image as in Figure 7 and Figure 8, a Gaussian kernel size of 30x30 pixels provides adequate smoothing.

3) Color
In [7], lesions are potentially cancerous if they exhibit variations in color from one area to another. They may also be cancerous if they have shades of tan, black, or brown. In macro photography, color values may be highly variable due to the many implementations of color balancing algorithms.
employed by camera manufacturers. To reduce variability, all lesion images were converted to grayscale before scoring. The standard deviation of the grayscale intensity values of all the pixels belonging to lesion regions was calculated. This standard deviation value was taken as the color variation risk. In general, lesions with a higher color standard deviation are considered to be a higher cancer risk.

D. Descriptor Validation

An initial MATLAB implementation of the feature descriptors was run on a set of test images containing both cancerous and non-cancerous skin lesions. The results were presented to a group of trained dermatologists both to determine the individual value of each metric and to help quantify the potential risk scales for each feature. A detailed study involving many hundreds of samples of cancerous lesions would be required to assign an absolute scale to each metric. Lacking such a study, the lesions were simply ranked relative to each other based on the relative risk predicted by each of the three features. The dermatologists were asked to comment on the relative ranking for each feature.

The asymmetry metric was especially useful in sorting lesions relative to risk. A simple threshold applied to the symmetry metric accurately separated cancerous from non-cancerous regions. The border strength metric appeared to rank lesions correctly by border definition, but failed to separate cancerous from non-cancerous lesions. This may be a fundamental flaw in the border feature or perhaps a reflection of the fact that the border strength metric ignores the “scalloped” and “irregular” properties of borders suggested by [7]. Color intensity proved least useful, failing to rank lesions in the test set correctly by color variation. This is perhaps due to non-uniform lighting and could be corrected by proper gamma-space image normalization and scaling. The color metric is also calculated on pixel intensity values alone, without placing higher weighting on tan or black regions as prescribed by [7]. A more effective approach may be to score color in the 2-dimensional hue/chrominance space after compensating for color balancing variations.

In future work, an overall risk score would be calculated by a weighted combination of the asymmetry, border, and color risk scores. Work from digital dermatoscopy image processing suggests a combination scheme in which the risk scores from each of the features considered are summed to compute an overall risk score [6]. If the overall risk score passes a specified threshold an alarm is triggered. In addition to summing individual scores, if any of the individual risk scores passes a certain threshold an alarm is triggered as well. Unfortunately the small sample size of melanoma images available to this project, in combination with the poor performance of the border and color metrics, prevented the derivation of risk threshold values for overall or individual feature risk scores.

IV. IMPLEMENTATION ON MOBILE DEVICES

A version of the risk score calculation algorithms was implemented for Android devices and tested on a Samsung Galaxy tablet device. To use the application, a user attaches a macro lens adapter to a mobile device as in Figure 9 and aims it at a skin lesion. If a lesion is present it is identified and risk scores are computed and presented in real-time. A screenshot from the mobile application is shown in Figure 10.

The image processing for the mobile application was performed in C++ using the open source OpenCV image processing library. The code for the application draws heavily from the “CVCam” example provided with the OpenCV distribution.

A. Image Segmentation Modifications

The largest challenge in porting the scoring algorithm to a mobile platform was segmenting the lesion from the background in the live images. Unlike the relatively well-composed images in the static test set, the live images are extremely inconsistent in both focus and lighting. In many cases the mobile device’s own shadow is confused as a large lesion when using the global histogram binning technique. Global histogram thresholding techniques failed to properly binarize an image so a local adaptive thresholding approach was adopted. Images were broken into a grid of 15x15 blocks and each block was thresholded individually based on a threshold derived from the block and its 8-connected neighbors. This approach handles the device shadow artifacts well, but can occasionally introduce blocking artifacts of its own as shown in Figure 11.
B. Mobile Application Results

It was beyond the scope of this project to test the performance of the mobile implementation of the risk score calculation with live in situ melanoma cases. Initial testing was attempted via printed versions of the static test images used to validate the feature descriptors’ performance. Unfortunately the limited resolution of color printers prevented an accurate rendering of the lesions in the test images. Despite printing at the highest DPI settings possible, pixilation artifacts from the printing process created irregular results. An example of pixilation in a printed test image is shown in Figure 12.

In place of “live” analog melanomas a series of “melanoma phantoms” were used. They consisted of a set of dark features drawn on index cards or a volunteer’s forearm. One phantom was designed to test for asymmetry. It is shown in Figure 13.

Another phantom was designed to test for color variations, shown in Figure 14.

Testing with the melanoma phantoms verifies the operation of the feature descriptors on the mobile device but additional testing against live melanomas will be required before the scoring algorithm can be considered a reasonable interpretation of the ABCDE skin self-exam.

V. Future Work

The results of this small initial study are encouraging but inconclusive. Before a mobile application is ready for release the asymmetry, border, and color metrics must be tested against a large dataset of melanoma images to derive both risk scoring thresholds and receiver operating characteristic curves. It is likely the border and color metrics must be revised to compensate for the focus and color space issues discussed in section III.D. A modified lens with a fixture to help a user position the mobile device at a precise standoff may also increase the accuracy of the results.

Perhaps the most fruitful future work involves the components of the ABCDE metric not considered in this project. In particular, the evolution of a lesion over time is an extremely powerful indicator of melanoma. A mobile application using SIFT[9]-like feature vectors to compare images of lesions with each other would be able to quickly quantify how much a potential melanoma has changed over time. Another mobile application using the mole localization scheme in [4] could help users build digital maps of mole locations on their bodies and identify if new lesions have appeared. Many dermatologists will also confirm that the ABCDE method is not necessarily a comprehensive collection of all the features they consider when diagnosing a melanoma. In this case, features derived from computer vision schemes like Bag-of-Words[12] modeling and machine learning classifiers would be able to surpass the performance the original ABCDE method.

ACKNOWLEDGMENTS

Susan Swetter, M.D., Derek Peng, M.D. – Stanford Department of Dermatology.
REFERENCES


