MELANOMA CLASSIFICATION FROM HIDDEN MARKOV TREE FEATURES

Marco F. Duarte, m Thomas E. Matthews, d Warren S. Warren, d Robert Calderbank d

m University of Massachusetts, Amherst, MA 01003
d Duke University, Durham, NC 27707

ABSTRACT

Melanoma detection relies on visual inspection of skin samples under the microscope via a qualitative set of indicators, causing large discordance among pathologists. New developments in pump-probe imaging enable the extraction of melanin intensity levels from skin samples and provide baseline qualitative figures for melanoma detection and classification. However, such basic figures do not capture the diverse types of cellular structure that distinguish different stages of melanoma. In this paper, we propose an initial approach for feature extraction for classification purposes via Hidden Markov Tree models trained on skin sample melanin intensity images. Our experimental results show that the proposed features provide a mathematical microscope that is able to better discriminate cellular structure, enabling successful classification of skin samples that are mislabeled when the baseline melanin intensity qualitative figures are used.

Index Terms— Image processing, wavelet transform, hidden Markov tree, melanoma detection and classification.

1. INTRODUCTION

Melanomas are among the most commonly occurring cancers, but they are clinically challenging to diagnose. For example, from 1990 to 2006, while overall cancer death rates decreased by 21% in men and 12% in women, the death rate for melanoma in the United States increased more than 7% [1]. Early detection remains difficult but is critical for successful treatment; five year survival rates fall from 98% for local cancers to 16% for metastatic melanoma [2].

The current standard for diagnosis remains biopsy and histopathology, but this too results in discordant conclusions because there is no one histological criterion for melanoma — diagnosis must instead be made by subjectively weighing a series of separate indicators that may be present in atypical lesions as well. A recent study found a discordance rate of 14% among pathologists for melanoma diagnosis [3]. Doctors must therefore err on the side of caution, which leads to an excess of false positive diagnoses and increased medical costs and emotional trauma from unnecessary surgeries, lymph node biopsies, and other treatments.

Melanoma presents a promising target for optical diagnosis both because suspicious lesions are accessible and disease occurs within a few hundred micrometers of the skin surface. Since melanin carries information about the metabolism and location of melanocytes (melanin producing cells) in tissue, the distribution of eumelanin and pheomelanin, the two dominant types of melanin, could act as a marker for disease.

It was recently demonstrated that two-color pump-probe imaging of skin samples can separate eumelanin and pheomelanin, two chromophores that are key to separate melanomas from benign nevi in a highly sensitive manner [4]. Initial approaches to melanoma detection and classification based on pump-probe imaging rely on pixel-averaged eumelanin-to- pheomelanin concentration ratios, achieving remarkable performance levels [5]. However, such pixel-wise features cannot capture the morphological features that distinguish different stages of melanoma, which are evident by inspection of the obtained images (see Fig. 1).

In this paper, we leverage prior work on modeling of natural images to perform feature extraction from pump-probe skin sample images for the purpose of melanoma classification. We focus on statistical models of image wavelet coefficients that capture the scale, predominance, and intensity of the spatial features present in the image. We show empirically that the use of these statistical morphological features for classification leverages the distinct cellular features of the different stages of melanoma to improve the performance of melanoma classification over pixel-averaged methods. In a sense, wavelet analysis provides a mathematical microscope that tabulates features relevant to melanoma detection and classification.

2. BACKGROUND

Melanoma classification: We briefly describe several classes of melanoma according to the histological and chemical features evident in the skin samples, as illustrated in Figure 1.

Benign nevi exhibit a proliferation of small melanocytes arranged as single cells and/or nests usually distributed along the basal layer, the deepest layer in the epidermis. These samples usually contain no melanocytes in the superficial layers of the epidermis, indicating no pagetoid (outward) spread.
Compound nevi exhibit small matured melanocytes that involve the epidermis as well as the dermis. The intra-epidermal component of these lesions is organized as single cells and nests without confluent growth pattern or presence of upper migration of melanocytes. Dysplastic nevi feature junctional nests of melanocytes that appear as a clustering of pigmented cells at the basal layer. Pigmented keratinocytes are often seen in the upper epidermis, but there is no pagetoid spread. Melanoma in-situ samples correspond to very early forms of melanoma, where the melanocytes have proliferated only radially within the base of the epidermis. In contrast, invasive melanoma features both radial and vertical proliferation of melanocytes. Such samples also tend to have large structure across the image, including pigmentation in the dermis.

In terms of chemical features, a large quantity of pheomelanin is often found in melanocytic benign and compound nevi, with a shift to eumelanin dominance in melanoma. Because eumelanin is photoprotective and has antioxidant properties, whereas pheomelanin can act as a photosensitizer, it has been postulated that elevated amounts of pheomelanin would lead to increased damage from ultraviolet radiation and an increased risk of malignant transformation. Dysplastic nevi display atypical growth and seem to have increased pheomelanin content compared to normal skin and other melanocytic nevi [6]. However, the chemical identity of melanin in melanomas is less clear, and some evidence exists to show that eumelanin may in fact occur in increased concentration [7]. This indicates a more heterogeneous chemical signature being characteristic of melanomas in contrast to other types of lesions. The generalizations we make are drawn from only one new study (our STM paper) and in limited sample set, and that the literature does not paint a clear or definitive picture either.

Bulk analysis of the eumelanin content alone allows for the rejection of many false positive diagnoses. We calculate a weighted average of eumelanin content across the entire image by normalizing by total melanin content in a pixel-wise fashion. Regions containing surgical ink were not considered. As shown in Fig. 2, if only raw melanin content is considered, a threshold of 38% eumelanin captured all invasive melanomas and most of the melanomas in-situ while excluding > 75% of the dysplastic nevi. Although the eumelanin to pheomelanin ratio is not sufficient to diagnose melanoma, it may greatly improve diagnostic accuracy in conjunction with complementary diagnostic techniques.

Hidden Markov trees: A widely used sparse representation in signal and image processing is the wavelet transform. The wavelet transform of an image provides a multiscale time-frequency analysis of the image content, effectively encoding the locations and scales of the image features in a compact fashion. This energy compaction property is the main reason behind the popularity of wavelet transforms for image processing and compression, including the state-of-the-art JPEG2000 standard.

In a typical 2D real-valued wavelet transform of an \( \sqrt{N} \times \sqrt{N} \)-pixel image \( x \in \mathbb{R}^N \), each wavelet coefficient \( w_{o,s,i,j} \) is labeled by a scale \( s \in \{1, \ldots, S := \log_2(N)/2 \} \), orientation \( o \in \{H,V,D\} \) for horizontal, vertical, and diagonal, respectively, and offset \( (i,j) \), \( 1 \leq i,j \leq 2^{s-1} \). Additionally, a scaling coefficient \( w_0 \) captures the remaining energy of the signal. The image \( x \) can then be written as

\[
x = w_0 \varphi + \sum_{o \in \{H,V,D\}} \sum_{s=1}^{S} \sum_{i,j=1}^{2^{s-1}} w_{o,s,i,j} \psi_{o,s,i,j},
\]

where \( \varphi \) denotes the scaling function and \( \psi_{o,s,i,j} \) denotes the mother wavelet function \( \psi_o \) for orientation \( o \) dilated to scale \( s \) and translated to offset \( (i,j) \). For convenience, we also index the wavelet coefficients and wavelet functions as \( \{w_0, w_1, \ldots, w_{N-1}\} \) and \( \{\varphi, \psi_1, \ldots, \psi_{N-1}\} \) using an arbitrary ordering, e.g., lexicographic.
A coefficient \( w_{o,s,i,j} \) at scale \( s \) describes a portion of the signal of size \( O(4^{-s}) \). With \( 4^{s-1} \) such coefficients at each scale and orientation, a quad-tree provides a natural organization for the coefficients. Each coefficient at scale \( s < \log_2(N)/2 \) has 4 children at scale \( s + 1 \), and each coefficient at scale \( s > 1 \) has one parent at scale \( s - 1 \).

A large wavelet coefficient (in magnitude) generally indicates the presence of a singularity inside its support; a small wavelet coefficient generally indicates a smooth region. This energy compaction property causes wavelet coefficients to have a peaky non-Gaussian distribution. Thanks to the nesting of child wavelets inside their parents, edges in general manifest themselves in the wavelet domain as chains of large coefficients propagating across scales in the wavelet quad tree; we call this phenomenon the persistence property.

Hidden Markov Trees (HMTs) [8] offer one modeling framework that succinctly and accurately captures this joint structure in natural images. In an HMT, wavelet coefficients are modeled probabilistically using a mixture of Gaussians: one component features a large variance that models large nonzero coefficients and receives a small weight (to encourage few such coefficients), while a second component features a small variance that models small and zero-valued coefficients and receives a large weight. We distinguish these two components by associating to each wavelet coefficient \( w_n \) an unobserved hidden state \( S_n \in \{ S, L \} \); the value of \( S_n \) determines which of the two components of the mixture model is used to generate \( w_n \):

\[
f(w_n|S_n = S) = \mathcal{N}(0, \sigma^2_{S,n}), f(w_n|S_n = L) = \mathcal{N}(0, \sigma^2_{L,n}),
\]

with \( \sigma^2_{L,n} > \sigma^2_{S,n} \). To generate the mixture, we apply a probability distribution to the available states: \( p(S_n = S) = p_S \) and \( p(S_n = L) = p_L \), with \( p_S + p_L = 1 \).

To simplify the model, the coefficient-dependent parameters are made equal for all coefficients within a scale; that is, the new model has parameters \( \Theta = \{ p_S, p_L, \sigma_{S}, \sigma_{L} \} \). We can obtain estimates of all these parameters for a set of coefficients \( w \) using maximum likelihood estimation via an expectation-maximization (EM) algorithm [8].

3. FEATURE EXTRACTION

We design a feature set for a skin sample based on the parameters of an HMT trained over its pump-probe image. The goal of feature selection is to capture the scale and orientation of the dominant features in an image, given that different types of melanoma exhibit different scales and orientations for their cellular structure. This is in contrast to the features in [5], which do not consider orientation and scale, but rather global average melanin concentration. Therefore, we select as features the likelihoods of the small state for each orientation \( \{ H, V, D \} \) and scale \( s = 1, \ldots, S \). If strong features are present in an image at a given scale and orientation, the corresponding wavelet coefficients will be large, yielding a small probability of the small state for that scale and orienta-

![Fig. 3. Pictorial representation of feature vectors extracted from skin samples. Each column corresponds to a different sample from the labeled classes and each row corresponds to the likelihood of a small state for each scale. Red (dark) color denotes largest likelihood, while blue (light) colors denote smallest likelihood.](image)

4. MELANOMA CLASSIFICATION PERFORMANCE

Once the images of lesions have been reduced to an appropriate mathematical representation, they need to be classified. Support vector machines (SVMs) have been used to tackle diverse classification and regression problems and are one of the most effective tools for these tasks. To develop a clas-
Table 1. Skin sample classification results for F2 features, scales 4–9, $\sigma = 0.2$. The table shows the probability of correct classification $P_C$, sensitivity (probability of detection) $P_D$, and specificity (complement of false alarm probability) $1 - P_{FA}$ for each of the classification problems listed.

<table>
<thead>
<tr>
<th>Test</th>
<th>$P_C$</th>
<th>$P_D$</th>
<th>$1 - P_{FA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma vs. nevi</td>
<td>75%</td>
<td>72%</td>
<td>74%</td>
</tr>
<tr>
<td>Melanoma vs. nevi + SKs</td>
<td>61%</td>
<td>62%</td>
<td>60%</td>
</tr>
<tr>
<td>Invasive melanoma vs. nevi</td>
<td>57%</td>
<td>54%</td>
<td>57%</td>
</tr>
<tr>
<td><em>In situ</em> melanoma vs. nevi</td>
<td>72%</td>
<td>73%</td>
<td>72%</td>
</tr>
<tr>
<td>Melanoma vs. benign</td>
<td>59%</td>
<td>60%</td>
<td>58%</td>
</tr>
<tr>
<td>Melanoma vs. dysplastic</td>
<td>56%</td>
<td>52%</td>
<td>60%</td>
</tr>
</tbody>
</table>

![Image of Table 1](image-url)