Extracting Morphological High-Level Intuitive Features (HLIF) for Enhancing Skin Lesion Classification

Robert Amelard, *Student Member, IEEE*, Alexander Wong, *Member, IEEE*,
David A. Clausi, *Senior Member, IEEE*

***Abstract*— Feature extraction of skin lesions is necessary to provide automated tools for the detection of skin cancer. High-level intuitive features (HLIF) that measure border irregularity of skin lesion images obtained with standard cameras are presented. Existing feature sets have defined many low-level unintuitive features. Incorporating HLIFs into a set of low-level features gives more semantic meaning to the feature set, and allows the system to provide intuitive rationale for the classification decision. Promising experimental results show that adding a small set of HLIFs to the large state-of-the-art low-level skin lesion feature set increases sensitivity, specificity, and accuracy, while decreasing the cross-validation error.**

I. Introduction

Melanoma causes the most deaths worldwide of all skin diseases [1]. The World Health Organization (WHO) estimates that the current rate of new melanoma cases annually is 132,000 worldwide, and one of every five American citizens are expected to develop malignant melanoma in their lifetime [2]. The mortality rate of malignant melanoma is highly correlated with the stage in which it is identified. The five-year survival rate for patients whose melanoma is diagnosed in its infancy is 98%, but decreases rapidly to 62% if the melanoma has spread to surrounding tissues, and a bleak 16% if the melanoma has spread to remote parts of the body [3].

The current clinical standard for diagnosing skin lesions is visual inspection. One of the most widely used rubrics for visually analysing a lesion is the "ABCD" rubric [4], [5]. The dermatologist assigns a score for asymmetry (A), border irregularity (B), colour patterns (C), and diameter (D). The resulting summed score is indicative of the patient's chance of having malignant melanoma. It has been reported that expert dermatologists using the ABCD rubric exhibit a sensitivity of 76.0%-87.7% and a specificity of 61.0%-77.8% with the aid of a dermatoscope [6].

Most image processing techniques that address this issue rely on an image obtained with a digital dermatoscope [8]-[10]. However, it has been reported that only 48% of US fellows of the American Academy of Dermatology use dermatoscopes [7], signifying that these methods would not be incorporated into the majority of clinical settings. Some methods have recently been proposed that use standard camera images [13], [14], but these papers have focused on preprocessing the images rather than feature extraction and classification.

The main contribution of this paper is the proposition and analytical reasoning of high-level intuitive features (HLIF) that quantitatively describe the degree of irregularity about a lesion's border. We define "high-level intuitive" features as features that have been carefully designed such that their formulation models a human-observable phenomenon, and whose score can be intuited in a natural way. In contrast, low-level features are calculations that were not designed with the intent of describing a particular human-observable phenomenon. A feature set that contains HLIFs can provide understandable justification of a classification result. This is an important characteristic of the system if doctors are to trust the system's diagnostic decisions.

We measure the success of the feature set according to the sensitivity, specificity, accuracy, and error associated with leave-one-out cross-validation (LOO CV) trials using a standard soft-margin linear support vector machine (SVM) classifier [15]. The three proposed border irregularity HLIFs are concatenated to a slightly modified version of the feature set proposed by Cavalcanti *et al.* [14] which produces better performance than the original set. The success of this new set is compared with their original set, as well as the modified set. Automated lesion detection is a problem not addressed in this paper.

II. Feature Extraction

A. Clinical "Border Irregularity" Description

Clinically, dermatologists visually divide the border delineating a skin lesion into eight segments such that the lesion is separated vertically, horizontally, and by an axis at $\pm 45^\circ$ [5]. One point is added to the score for each region that has abrupt pigment cut-off or irregular notches, thus $B \in \{0, 1, \ldots, 8\}$. This rubric is not only very subjective, but the scores themselves are very restricted. The proposed HLIFs give a more precise measurement for the irregularity of the entire border which can still be interpreted by the dermatologist.

B. HLIF for Fine Irregularities

Morphological operations are a set of techniques for processing shapes in images [11]. Given a structuring element such as a disk, one can apply a morphological operation using this structuring element to alter a shape in some deterministic way.

*This work was supported by Agfa Healthcare Inc., Ontario Centres of Excellence (OCE), the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Ontario Ministry of Economic Development and Innovation.

R. Amelard, A. Wong, and D. A. Clausi are with Department of Systems Design Engineering, University of Waterloo, Ontario, Canada, N2L 3G1 {ramelard,a28wong,dclausi}@uwaterloo.ca

Two of the most fundamental techniques are morphological opening and morphological closing. Let A be a binary image and B be a structuring element. Then, morphological opening and closing are defined as:

$$A \circ B = (A \ominus B) \oplus B \tag{1}$$

$$A \bullet B = (A \oplus B) \ominus B \tag{2}$$

where $A \ominus B$ and $A \oplus B$ are morphological erosion and dilation, defined as:

$$A \ominus B = \bigcap_{b \in B} A_{-b} \tag{3}$$

$$A \oplus B = \bigcup_{b \in B} A_b \tag{4}$$

where A_b is the translation of A by vector b.

Intuitively, morphological closing using a disk structuring element tends to smooth over sharp *exterior* valleys. That is, performing morphological closing on a shape whose border contains abrupt valleys results in an image with filled valleys and thus a larger area. Similarly, morphological opening using a disk structuring element tends to smooth over sharp *interior* peaks. That is, performing morphological opening on a shape whose border contains abrupt peaks results in an image with truncated peaks and thus a smaller area. In contrast, morphologically closing or opening a shape with a smooth border results in little structural change.

The resulting feature is the normalized change in area from morphological closing and opening:

$$f_1^B = \frac{A_{closed} - A_{lesion}}{A_{lesion}} + \frac{A_{lesion} - A_{opened}}{A_{lesion}} \tag{5}$$

where A_{closed} and A_{opened} are the areas of the resulting mask from performing morphological closing and opening on the initial lesion mask, and A_{lesion} is the area of the original lesion. The two measures are summed to account for both irregular valleys and irregular peaks (both indicative of an irregular border).

Fig. 1 depicts a lesion mask with the corresponding result of performing a morphological closing and opening with a disk of size 20. Note how the resulting mask from morphological closing tends to smooth over valleys of irregularities whilst keeping the general structure of the border. Similarly, the resulting mask from morphological opening tends to smooth the peaks of the irregularities. The difference in area of the morphed mask and the original mask thus captures the amount of change that results from smoothing over irregular structural features. Combining the two measures results in a score that represents the overall structural irregularity of the border.

C. HLIF for Coarse Irregularities

Intuitively, highly irregular borders with many notches can be viewed as a structure containing high-frequency information. Fourier descriptors are used to describe a shape in terms of frequency information. Determining a shape's Fourier descriptors is accomplished by representing the x-axis as the real axis, and the y-axis as the imaginary axis.

Fig. 1. Depiction of morphological closing/opening of the mask of a lesion diagnosed as malignant (lentigo maligna) melanoma using a disk structuring element of size 20. Note how the closing smooths the valleys of border irregularities, and the opening smooths the peaks of those irregularities. In this example, $f_1^B = 0.0910$

Thus the points of the shape border are transformed into a 1D complex number. In particular, if $x(i)$ and $y(i)$ are the coordinates of the i^{th} pixel, then let $s(i) = x(i) + jy(i)$. Applying the discrete Fourier transform results in the Fourier descriptors:

$$C(u) = \sum_{n=0}^{N-1} s(n) e^{-j2\pi kn/N} \tag{6}$$

for $u = 0, 1, \ldots, N-1$.

Fourier descriptor normalization is performed in the following three steps [12]:

1) *Translation Invariance*: set the first Fourier coefficient, the DC component, to 0.
2) *Scale Invariance*: divide each Fourier coefficient by the magnitude of the second coefficient.
3) *Rotation/Point-Order Invariance*: consider only the magnitude of the Fourier coefficients.

If each set of Fourier descriptors undergoes these three normalization steps, they can be compared in an invariant and robust manner.

Upon computing the Fourier descriptors of the lesion's border using Fast Fourier Transform (FFT), a coarse representation was reconstructed using only low-frequency components. We found that using the lowest four frequency components of a 1000-sampled border yielded good overall structural representations of the lesion. A lesion with a highly irregular border will have a larger perimeter than the low-frequency representation of the same border. Thus the perimeters of the low-frequency border and the original border were compared to yield the feature:

$$f_2^B = \frac{|P_{lesion} - P_{low}|}{P_{lesion}} \tag{7}$$

where P_{lesion} and P_{low} are the lengths of the perimeter of the original and low-frequency border. Fig. 2 depicts the lesion's border (in red) and the reconstruction of that border using the lowest four frequency components (in blue). Note how the reconstructed border produces a smooth version of the original border, with less high-frequency information. Thus, by comparing the perimeters we are measuring how "bumpy" the border is relative to its low-frequency, smooth representation.

D. HLIF for Comparing Against Average Malignant Lesion

As indicated in the clinical description, malignant lesions tend to have very distinct borders when compared with

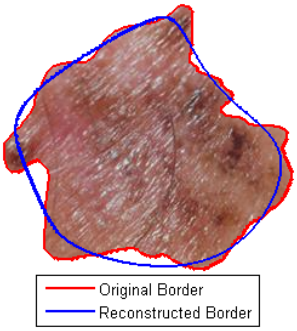

Fig. 2. Depiction of reconstructing the border of a lesion diagnosed as malignant (lentigo maligna) melanoma using the lowest four Fourier coefficients. Note how the low-frequency reconstruction captures the general shape of the lesion, but not the high-frequency information that delineates the border irregularities. In this example, $f_2^B = 0.2623$

benign cases. This can be viewed in the following way: the average malignant lesion is structurally different than the average benign lesion.

The Fourier descriptors are computed for each image in the data set, and the sum of all coefficients is stored. Let $\bar{C} = \{\bar{C}(0), \bar{C}(1), \ldots, \bar{C}(N-1)\}$ denote this set of coefficients. We normalize $\bar{C}$ with the same steps as described before:

1) *Translation Invariance*: $\bar{C}(0) = 0$,
2) *Scale Invariance*: $\bar{C}(u) = \frac{\bar{C}(u)}{|\bar{C}(1)|}$.
3) *Rotation/Point-Order Invariance*: Consider $|\bar{C}(i)|$.

The feature is then computed as the sum of squared differences between the magnitude of the complex-valued descriptors and the average descriptors:

$$f_3^B = \sum_{u=0}^{N-1} (|C(u)| - |\bar{C}(u)|)^2 \tag{8}$$

where C is the set of normalized Fourier coefficients for the image under analysis. The magnitude is used so that the comparison is invariant to any phase-related phenomena (i.e., rotation and order of sequential points).

III. Experimental Results

Our data set contains 206 images obtained with consumer-level cameras: 69 images from the Dermatology Information System [16] (43 malignant melanomas, 26 nevi), and 137 images from DermQuest [17] (76 malignant melanomas and 61 nevi). Each image contains only a single lesion of interest, which was manually segmented to create a binary mask for differentiating pixels describing the lesion from the surrounding skin. Each image underwent the illumination invariance algorithm described by Cavalcanti *et al.* [14]. For the proposed HLIFs, each image was rotated so that its major axis (i.e., the axis that passes through the lesion's centroid and is in the direction of maximal variance) aligned with the horizontal axis. This ensures rotation-invariant features. The image was then uniformly scaled so that the bounding box of the lesion fit inside a 200×200 rectangle. Due to the small number of samples, leave-one-out cross-validation (LOO CV) was used to compute the error of the classifier, as well as sensitivity, specificity, and accuracy.

For clarity, we use the following terminology to denote the different feature sets:

- $\boldsymbol{F_{HLIF}^B}$: feature set containing the three proposed HLIFs describing border irregularity (see Section II).
- $\boldsymbol{F_C}$: feature set proposed by Cavalcanti *et al.* containing 52 features describing asymmetry, border irregularity, colour variation, and differential structures [14].
- $\boldsymbol{F_{CM}}$: a modified version of F_C without the four inconsistent asymmetry features ($F_{CM} = F_C \backslash \{f_8, \ldots, f_{11}\} \subset F_C$).
- $\boldsymbol{F_T}$: the "total" superset containing $F_{CM} \cup F_{HLIF}^B$.

The rationale behind creating the set F_{CM} is as follows. In Cavalcanti *et al.*'s paper, the asymmetry features $\{f_8, \ldots, f_{11}\}$ measure the relative difference of the borders separated by two orthogonal axes. However, the mathematical formulation does not constrain the relative sizes of the borders (i.e., B_1 may refer to either the larger or smaller border). These features are inconsistent, and thus are omitted during our analysis.

The extraction of each of these feature sets was implemented in MATLAB. Upon extraction, we classified the images using SVM without a kernel. The linear SVM model was chosen to emphasize the linear separability of the data in the feature space, as opposed to the efficiency of the classifier. SVM is known as a classifier that is robust to noise and does not tend to overfit training data. This is an important trait due to the small size of our data set.

We evaluated the feature sets in the following way. The LOO CV technique was used to generate the success metrics (i.e., LOO CV error, sensitivity, specificity, and accuracy). That is, for each data point the classifier was trained on all $N_i - 1$ other data points, where N_i is the number of images. Then, the omitted data point was used as the test case. This was performed for each image, and identically for each feature set. The classification results are summarized in Table I.

A. Analysis of Results

As expected, F_{CM} performs better than F_C in all metrics. We added our three proposed HLIFs to this feature set to obtain the 51-dimensional superset F_T.

Classification using F_T attains the highest sensitivity (90.76%), specificity (82.76%), and accuracy (87.38%) of all the feature sets, and exhibits the lowest LOO CV error (12.62%). In medical imaging, sensitivity is a crucial metric because low sensitivity indicates that malignant cases have been left undetected. This can lead to patient deaths. It is therefore important to note that adding the border features F_{HLIF}^B to the set F_{CM} increases the sensitivity from 84.87% to 90.76%, and is higher than the benchmark feature set F_C which has a sensitivity of 83.19%. The importance of including HLIFs in a feature set is demonstrated by this case. Although we are only adding three features to a 48-dimensional feature set (i.e., we are only increasing the feature space by 6%), all success metrics show non-trivial

TABLE I

COMPARING CLASSIFICATION RESULTS OF DIFFERENT FEATURE SETS. LOO CV IS "LEAVE-ONE-OUT CROSS-VALIDATION".

Feature set	Description (see Section III)	# features	Sensitivity	Specificity	Accuracy	LOO CV Error
F_C	Cavalcanti *et al.* feature set [14]	52	83.19%	74.71%	79.61%	20.39%
F_{CM}	Modified F_C (see Section III)	48	84.87%	75.86%	81.07%	18.93%
F_T	Combined F_{CM} and F^B_{HLIF}	51	**90.76%**	**82.76%**	**87.38%**	**12.62%**

(a) False Negatives (b) False Positives

Fig. 3. Examples of images that were misclassified. The manually segmented lesion border has been superimposed on the image for visualization purposes.

improvements. This is due to the intuitive nature of HLIFs, which describe human-observable phenomena.

Classification using F_C attains lower results than reported in [14]. This is because the authors grew their data set using the Smoothed Bootstrap Resampling method, which does not introduce as much variability as is seen in independent clinical trials.

B. Sources of Error

Examples of misclassified images are given in Fig. 3. Notice how most of the false negative cases have a fairly regular border. The last image in the set has a very irregular border. However, it is very saturated by white pixels due to the flash, resulting in skewed colour features.

Most of the false positive cases have very irregular borders. The HLIFs for describing border irregularity are performing as intended, but other characteristics of the lesion are not being accounted for by the feature set. This emphasizes the need for HLIFs to describe the other characteristics of melanoma, namely asymmetry and colour patterns.

IV. CONCLUSION

In this paper we have proposed a set of HLIFs that describe the amount of border irregularity about a lesion image obtained using standard consumer-level cameras. HLIFs capture deterministic information about some human-observable phenomenon. The experimental findings indicate that incorporating the small set of HLIFs to a set of low-level features and classifying the data with a standard linear SVM model yields very promising results. Future work includes mapping the HLIF scores to intuitive labels for further user comprehension. The data set will also be expanded and a statistical analysis of the feature space will be conducted. HLIFs to describe asymmetry and colour patterns will also be designed, and this set will be evaluated as a feature space for a diagnostic aid system.

REFERENCES

[1] *Melanoma – PubMed Health*, U.S. National Library of Medicine [Online]. Available: `http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001853`

[2] *WHO | Skin cancers*, World Health Organization [Online]. Available: `http://www.who.int/uv/faq/skincancer/en/index1.html`

[3] "Cancer Facts & Figures 2011," American Cancer Society, Atlanta, GA, Tech. Rep. ACSPC-029771, 2011.

[4] W. Stolz *et al.*, "ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma," *European J. of Dermatology*, vol. 4, pp.521-527, 1994.

[5] F. Nachbar *et al.*, "The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions," *J. of the Amer. Academy of Dermatology*, vol. 30, no. 4, pp. 551-559, Apr. 1994.

[6] G. Argenziano *et al.*, "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet," *J. of Amer. Academy of Dermatology*, vol. 48, no. 5, pp. 679-693, May 2003.

[7] H. C. Engasser, E. M Warshaw, "Dermatoscopy use by US dermatologists: A cross-sectional survey," *J. of Amer. Academy of Dermatology*, vol. 63, no. 3, pp. 412-419, Sep. 2010.

[8] H. Iyatomi *et al.*, "Classification of melanocytic skin lesions from non-melanocytic lesions," in *IEEE Int. Conf. on Eng. in Medicine and Biology Soc.*, Buenos Aires, Argentina, 2010, pp. 5407-5410.

[9] G. Capdehourat *et al.*, "Pigmented skin lesions classification using dermatoscopic images," *Progress in Pattern Recognition, Image Anal., Comput. Vision, and Applicat.*, vol. 5856, pp. 537-544, 2009.

[10] M. E. Celebi *et al.*, "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graph.*, vol. 31, pp. 362-373, Sep. 2007.

[11] J. Serra, *Image analysis and mathematical morphology*. London: Academic Press, 1982.

[12] T. P. Wallace, P. A. Wintz, "An efficient three-dimensional aircraft recognition algorithm using normalized fourier descriptors," *Comput. Graph. and Image Process.*, vol. 13, no. 2, Jun. 1980.

[13] J. F. Alcon *et al.*, "Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis," *IEEE J. of Sel. Topics Signal Process.*, vol. 3, no. 1, Feb. 2009.

[14] P. G. Cavalcanti, J. Scharcanski, "Automated prescreening of pigmented skin lesions using standard cameras," *Computerized Medical Imaging and Graph.*, vol. 35, no. 6, pp. 481-491, Sep. 2011.

[15] C. Cortes, V. Vapnik, "Support-vector networks," *Mach. Learning*, vol. 20, no. 3, pp. 273-297, Sep. 1995.

[16] *DermIS* [Online]. Available: `http://www.dermis.net`

[17] *DermQuest* [Online]. Available: `http://www.dermquest.com`