Published in final edited form as:

Skin Res Technol. 2009 August ; 15(3): 283–287. doi:10.1111/j.1600-0846.2009.00354.x.

Detection of Basal Cell Carcinoma Using Color and Histogram Measures of Semitranslucent Areas

William V. Stoecker^{a,e,h}, Kapil Gupta^a, Bijaya Shrestha^b, Mark Wronkiewicz^{a,g}, Raed Chowdhury^{a,g}, R. Joe Stanley^b, Jin Xu^a, Randy H. Moss^b, M. Emre Celebi^c, Harold S. Rabinovitz^d, Margaret Oliviero^d, Joseph M. Malters^e, and Isabel Kolm^f

^aStoecker and Associates 1702 E. 10th St. Rolla, MO 65401-4600 ^bDepartment of Electrical and Computer Engineering Missouri University of Science and Technology Rolla, MO 65409-0040

^cDepartment of Computer Science Louisiana State University in Shreveport Technology Center 206, One University Place Shreveport, LA 71115 ^dSkin and Cancer Associates 201 N.W. 82nd Avenue Bennett Medical Plaza, Suite 501 Plantation, FL 33324 ^eThe Dermatology Center, 1702 E. 10th St. Rolla, MO 65401-4600 ^fDepartment of Dermatology University Hospital Zürich Gloriastrasse 31 CH-8091 Zurich Switzerland ^gRolla High School 900 Bulldog Run Rolla MO, 65401 ^hDivision of Dermatology Department of Internal Medicine University of Missouri Health Sciences Center Columbia, MO 65212

Abstract

Background—Semitranslucency, defined as a smooth, jelly-like area with varied, near-skin-tone color, can indicate a diagnosis of basal cell carcinoma (BCC) with high specificity. This study sought to analyze potential areas of semitranslucency with histogram-derived texture and color measures to discriminate BCC from non-semitranslucent areas in non-BCC skin lesions.

Methods—For 210 dermoscopy images, the areas of semitranslucency in 42 BCCs and comparable areas of smoothness and color in 168 non-BCCs were selected manually. Six color measures and six texture measures were applied to the semitranslucent areas of the BCC and the comparable areas in the non-BCC images.

Results—Receiver operating characteristic (ROC) curve analysis showed that the texture measures alone provided greater separation of BCC from non-BCC than the color measures alone. Statistical analysis showed that the four most important measures of semitranslucency are three histogram measures: contrast, smoothness, and entropy, and one color measure: blue chromaticity. Smoothness is the single most important measure. The combined 12 measures achieved a diagnostic accuracy of 95.05% based on area under the ROC curve.

Conclusion—Texture and color analysis measures, especially smoothness, may afford automatic detection of basal cell carcinoma images with semitranslucency.

Keywords

Basal cell carcinoma; Image analysis; Texture; Semitranslucency; Dermoscopy

I. INTRODUCTION

Dermoscopy is a noninvasive imaging technique that uses optical magnification and fluid immersion or cross-polarized lighting to allow better clinical assessment of skin lesions [1]. Dermoscopy has been shown to improve diagnostic accuracy of pigmented lesions in those with formal training [2].

The clinical feature semitranslucency, a semi-liquid or jelly-like appearance that is commonly visible in basal cell carcinomas (BCCs) with the unaided eye, is useful for the identification of these common skin cancers [3]. Semitranslucent areas in dermoscopic images of BCC also display the smooth, jelly-like appearance of the clinical counterpart [4]. The color ranges from light pink, usually seen in the thickest areas, to dull orange, often in a peripheral location, to gray [4]. Semitranslucency may be observed dermoscopically in tiny BCCs with clinical measurements of only 3 or 4 mm. Such tiny BCCs may not yet show other dermoscopy signs of BCC such as arborizing vessels. Accordingly, this sign is important for early detection of BCC, the most common skin cancer.

The smooth, jelly-like appearance of semitranslucent areas is a visible form of texture. This visible texture may be analyzed using characteristics of the normalized intensity histogram. Six histogram characteristics for analysis were chosen as shown in Table 1 [5]. The visible characteristics of brightness, contrast, and smoothness are respectively calculated as the histogram mean **H1**, variance **H2**, and a function **H3** of the reciprocal of the variance as shown in Table 1. Three other histogram characteristics – skewness **H4**, uniformity **H5**, and entropy **H6** are also defined in Table 1. For a smooth semitranslucent area, contrast, smoothness and skewness features have a low value and uniformity and entropy have a high value. For a semitranslucent area, which is bright but lacks flash and pigment, brightness lies in the middle range.

Semitranslucent areas also possess a characteristic range of colors as noted above, rather than a single distinguishing color. These colors may be analyzed by characterizing the colors in various color systems. One study noted that relative red, green and blue (RGB) and chromaticity averages over a dermoscopy feature allow good detection of a color feature (blue veil) in dermoscopy images [6]. Chromaticity in red is defined as $R/(R+G+B)$. Definitions for green and blue chromaticity are similar. The color variables chosen were **C1–C3**: R, G and B averages respectively over the semitranslucent spots and **C4–C6**: R, G and B chromaticity over the respective spots. The other six variables analyzed were **H1–H6** as defined in Table 1.

Numerous methods for extracting features from clinical skin lesion images have been proposed in the literature [7–9]. However, feature extraction in dermoscopy images is relatively unexplored. The dermoscopic feature extraction studies to date include two pilot studies on pigment networks [10][11] and globules [10], and four systematic studies on dots [12], blue-white areas [6], and blotches [13][14]. To the best of our knowledge, there is no published systematic study on the detection of semitranslucency.

It is the primary purpose of this study to determine if BCC can be differentiated from other skin lesions by application of measures to the semitranslucency feature of BCC. The measures chosen are the 12 color and histogram measures defined above, determined from the semitranslucent areas belonging to BCC and those areas in non-BCCs which most closely resemble the semitranslucent areas in BCC. A second purpose of this study is to determine which color and histogram features best separate BCC from other skin lesions.

II. METHODS

The image set used for this study includes non-contact, polarized dermoscopy images of 42 BCCs and 168 non-BCC lesions from two clinics: Skin and Cancer Associates, Plantation, FL, and the Dermatology Center, Rolla, MO, selected for dermoscopy study during January–November, 2007. All lesions with any dermoscopy features of malignancy and any benign lesions for which there was any uncertainty were biopsied. All images were non-contact polarized dermoscopy images taken with a Sony DSC-W70 7.2 megapixel digital camera with a 3Gen DermLite Pro HR dermoscopy attachment (3Gen LLC, San Juan Capistrano, CA). Non-contact, polarized dermoscopy images were used in this study because the contrast in both texture and color with surrounding areas is often better preserved and visualized with this type of dermoscopy [15]. Accordingly, semitranslucency is more easily seen with this imaging modality [4].

The 168-image competitive set to which the BCC set was compared consists of a variety of lesions encountered in the clinic during the same time the BCCs were acquired. These lesions were both malignant and benign: 2 actinic keratoses, 2 acral nevi, 1 angiokeratoma, 2 blue nevi, 4 benign collision tumors, 1 collision between malignant and benign lesions, 5 congenital nevi, 19 dysplastic nevi without significant atypia, 5 dysplastic nevi with severe atypia, 2 foreign body reactions, 4 hemangiomas, 7 lentigos, 6 lichenoid keratoses, 11 melanomas in situ, 10 invasive melanomas, 41 nevocellular nevi, 1 prurigo nodularis, 1 porokeratosis, 5 squamous cell carcinomas, 1 sebaceous carcinoma, 2 sebaceous hyperplasia, 8 squamous cell carcinomas in situ, 26 seborrheic keratoses, and 2 warts.

For the 42 BCCs, a student (JX) manually marked the semitranslucent portions of the lesions with an approximate oval shape smoothed by a second-order spline. A dermatologist (WVS) adjusted all marked areas as needed. A BCC with a typical area of semitranslucency is shown in Figure 1. The 168 non-BCC lesions were manually marked using the same procedure. The areas marked on the non-BCC lesions were similar in size and as close as possible to the same color and texture as the semitranslucent spots in the BCCs. The marked areas in both BCC and non-BCC lesions were then analyzed for RGB and chromaticity color features as well as the texture measured by the **H1–H6** intensity histogram features.

For the semitranslucent area data gathered, statistical analysis was performed using SAS software (SAS Corp., Cary, NC) [16]. The twelve measures **C1–C6** and **H1–H6** were used as inputs to a standard back propagation neural network within Matlab (MathWorks Inc, Natick MA). The network architectures used were $6 \times 6 \times 4 \times 1$, $6 \times 6 \times 4 \times 1$ or $12 \times 6 \times 4 \times 1$ (4 layers, with 6 or 12 inputs, two hidden layers with 6 and 4 nodes in each layer and one output node. Because of the relatively small data set, a leave-one-out methodology was used for training/test set generation. The neural network was trained up to 15 epochs or until root-mean-square error was less than 0.001.

III. RESULTS

When all twelve measured variables are taken together in a binary logistic regression model, **H3** has the lowest p-value, although none of the variables are found to be statistically significant at a 0.05 significance level. The lack of significance of any variable indicates that the full model can be improved by variable selection.

Using stepwise selection, five features, **H2**, **H3**, **H6**, **C1**, and **C6** are found to be statistically significant at a 5% significance level, with **H3** and **C6** having the lowest odds ratio estimates of < 0.001 . Of these two features, **H3** has the lowest p-value which reflects that the smoothness **H3** is the most significant variable in the reduced model obtained by stepwise selection as well as the full model. Analyzing the multicollinearity effects, **H2** is found to be

highly correlated with **H3** (correlation coefficient > 0.9). Therefore, keeping both **H2** and **H3** in any classification model should be avoided. Since **H3** has a lower odds ratio than **H2**, it should be retained. In the model with **H3**, **H6**, **C1**, and **C6** only **H3** and **C6** are found to be statistically significant. Therefore, the model using smoothness **H3** and blue chromaticity **C6** appears to be the best model using two variables.

Analyzing two-way and three-way interaction effects between the variables obtained from stepwise selection, the model with the three way interaction effect **H2**, **H3**, and **H6** and the variable **C6** appears to be the best. Here, both the interaction term and the single variable are statistically significant and have odds ratios of < 0.001. However, is not desirable to include interaction terms in the analysis as the physical interpretation of the interaction variables is difficult.

The area under the receiver operating characteristic (ROC) curve for detection of BCC using all twelve measures in a backpropagation neural network for the semitranslucent spots, combining both color and histogram features, is 0.9505. For the histogram features only, the area under the curve is 0.9351. For the color features only, the area under the curve is 0.9087. Figure 2 presents the ROC curves for the combined, color only, and texture only feature cases. The vertical and horizontal axes refer to the sensitivity and 1-specificity, respectively, for the ROC curve plots. For comparison purposes, the neural network (architecture 4×4×2×1) was run using the same leave-one-out approach on the four features (**H2**, **H3**, **H6**, **C6**), and the area under ROC obtained was 0.9204.

IV. DISCUSSION

These results are consistent with the clinical impression that the jelly-like semitranslucency phenomenon is distinguished more by texture, characterized here by histogram variables, than by the color, which varies considerably. Using variable selection, three of the four most important variables are histogram variables: contrast, smoothness, and entropy. Smoothness appears to be the most important semitranslucency measure in both the full model and reduced model. The high separation of BCC from non-BCC areas appears promising for early detection of these common skin cancers. The detection of BCC with a high degree of accuracy and differentiation from a wide variety of lesions, both malignant and benign, is also encouraging, as most studies focus on the more limited problem of discrimination of melanoma from benign pigmented lesions, rather than differentiation of a single non-melanoma cancer type from all other lesions.

One key limitation of this study is the possibility that other non-semitranslucent areas may have abnormal smoothness, and are not as highly correlated with BCC as semitranslucency is. The dermoscopy finding most similar in texture to semitranslucency is ulceration, which can display similar smoothness on dermoscopy. Ulceration is often bright red but sometimes displays similar colors to semitranslucency, and cannot always be distinguished from semitranslucency in dermoscopy images. Atraumatic ulceration, however, also signifies the presence of skin cancer, most commonly basal cell carcinoma, so the detection of this abnormally smooth texture in a lesion signals the need for biopsy just as semitranslucency does.

A significant limitation of the study is the use of manual selection of the semitranslucent areas. Another limitation is the relatively small number of lesions in the study. Another limitation is the use of absolute (non-normalized) color rather than relative color (normalizing the color in selected areas by the background color), which serves to minimize color variations due to lighting and due to normal variations among individuals [6,14,17]. For future work, automatic detection of semitranslucent areas will be undertaken using

texture segmentation to find the smooth areas along with relative color determination, and the techniques reported here will be applied to a larger set of lesions.

Acknowledgments

This publication was made possible by Grant Number SBIR R44 CA-101639-02A2 of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH. The authors wish to thank Jason Hagerty, who served as system administrator.

References

- [1]. Argenziano G, Soyer HP, Chimenti S, Talamini R, Corona R, Sera F, et al. Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet. *J Am Acad Dermatol*. 2003; 48:679–93. [PubMed: 12734496]
- [2]. Menzies SW, Zalaudek I. Why perform Dermoscopy? The evidence for its role in the routine management of pigmented skin lesions. *Arch Dermatol*. 2006; 142:1211–1222. [PubMed: 16983009]
- [3]. Vanker AD, Stoecker WV. An expert diagnostic program for dermatology. *Comput Biomed Res*. 1984; 17:241–7. [PubMed: 6375947]
- [4]. Stoecker WV, Kolm I, Rabinovitz HS, Oliviero MC, Xu J, Malter JM. Semitranslucency in dermoscopy images of basal cell carcinoma. *Arch Dermatol*. 2008 submitted.
- [5]. Gonzalez, RC.; Woods, RE. *Digital Image Processing*. 3rd Ed.. Pearson Prentice Hall; Upper Saddle River, NJ: 2008.
- [6]. Celebi MC, Iyatomi H, Moss RH, Stoecker WV, Rabinovitz HS, Argenziano G, Soyer HP. Automatic detection of blue-white veil in dermoscopy images. *Comput Med Imaging Graph*. 2008 In press.
- [7]. Stoecker WV, Li WW, Moss RH. Automatic detection of asymmetry in skin tumors. *Comput Med Imaging and Graph*. 1992; 16(3):191–197. [PubMed: 1623494]
- [8]. Lee TK, McLean DI, Atkins MS. Irregularity index: A new border irregularity measure for cutaneous melanocytic lesions. *Medical Image Analysis*. 2003; 7(1):47–64. [PubMed: 12467721]
- [9]. She Z, Fish PJ. Analysis of skin line pattern for lesion classification. *Skin Res and Technol*. 2003; 9(1):73–80. [PubMed: 12535288]
- [10]. Fleming MG, Steger C, Zhang J, et al. Techniques for a structural analysis of dermatoscopic imagery. *Comput Med Imaging and Graphics*. 1998; 22(5):375–389.
- [11]. Caputo B, Panichelli V, Gigante GE. Toward a quantitative analysis of skin lesion images. *Studies in Health Technology and Informatics*. 2002; 90:509–513. [PubMed: 15460746]
- [12]. Yoshino, S.; Tanaka, T.; Tanaka, M.; Oka, H. Application of morphology for detection of dots in tumor. *Proc. of the SICE Annual Conf.*; 2004. p. 591-594.
- [13]. Pellacani G, Grana C, Cucchiara R, Seidenari S. Automated extraction and description of dark areas in surface microscopy melanocytic lesion images. *Dermatology*. 2003; 208(1):21–26. [PubMed: 14730232]
- [14]. Stoecker WV, Gupta K, Stanley RJ, et al. Detection of asymmetric blotches in dermoscopy images of malignant melanoma using relative color. *Skin Res and Technol*. 2005; 11:179–184. [PubMed: 15998328]
- [15]. Benvenuto-Andrade C, Dusza SW, Agero AL, Scope A, Rajadhyaksha M, Halpern AC, Marghoob AA. Differences between polarized light dermoscopy and immersion contact dermoscopy for the evaluation of skin lesions. *Arch Dermatol*. 2007; 143:329–38. [PubMed: 17372097]
- [16]. UCLA: Academic Technology Services, Statistical Consulting Group. Introduction to SAS. From <http://www.ats.ucla.edu/stat/sas/notes2/> (accessed April 24, 2008)
- [17]. Stanley RH, Moss RH, Stoecker WV, Aggarwal C. A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images. *Comput Med Imaging and Graphics*. 2003; 27(5):387–396.

Figure 1.
BCC with semitranslucent spot marked. The jelly-like appearance can be seen in a large area, around and above the selected spot.

Figure 2. ROC curves for the combined, color-only, and texture-only feature cases using a standard back-propagation neural network with leave-one-out methodology.

Table 1**Histogram Texture Measures**

No.	Measure	Formula for measure	Description
H1	Histogram Mean (brightness)	$\mu = \sum_{i=0}^{L-1} r_i p(r_i)$	The mean is the average brightness or gray level, where r_i is the i th gray level, L is the number of gray levels, $p(r)$ is the histogram for a region, and $p(r_i)$ is the probability of occurrence of gray level r_i .
H2	Histogram Variance (Contrast)	$\sigma = \sum_{i=0}^{L-1} (r_i - \mu)^2 p(r_i)$	The variance σ measures the average contrast.
H3	Histogram Smoothness	$S = 1 - \frac{1}{(1 + \sigma^2)}$	The smoothness index S measures the relative smoothness of the gray level in a region. S approaches 0 for a very smooth region and 1 for a coarse region.
H4	Histogram Skewness	$\kappa = \sum_{i=0}^{L-1} (r_i - \mu)^3 p(r_i)$	The skewness index, κ , measures the skewness of a histogram. $\kappa = 0$ for a symmetric histogram and is positive or negative for histograms skewed right or left.
H5	Histogram Uniformity	$U = \sum_{i=0}^{L-1} p^2(r_i)$	The uniformity measure U measures uniformity and is maximum when all gray levels are equal.
H6	Histogram Entropy (Information)	$e = - \sum_{i=0}^{L-1} p(r_i) \log_2 p(r_i)$	Entropy e , as defined by Claude Shannon, measures the uncertainty or information content of a message. The higher the entropy, the greater the information.