The Median Split Algorithm for Detection of Critical Melanoma Color Features

Kaushik V. S. N. Ghantasala[1], Raeed H. Chowdhury[2], Uday Guntupalli[1], Jason Hagerty[1,2], Randy H. Moss[1], Ryan K. Rader[2] and WilliamV.Stoecker[2]
[1]*Missouri University of Science And Technology, G20 Emerson Electrical Co. Hall, Rolla, MO 65409, U.S.A.*
[2]*Stoecker & Associates, 10101 Stoltz Drive, Rolla, MO 65401, U.S.A.*

Keywords: Median Split, Melanoma, Image Analysis, Color Processing, Dermoscopy.

Abstract: Detection of melanoma remains an empirical clinical science. New tools for automatic discrimination of melanoma from benign lesions in digitized dermoscopy images may allow an improvement in early detection of melanoma. This research implements a fast version of the median split algorithm in an open source format and applied to four-color splitting of the lesion area to capture the architectural disorder apparent in melanoma colors. Our version of the median split algorithm splits colors along the color axis with maximum Range. For a set of 888 dermoscopy images, the best model for discrimination produces an area under the receiver operating characteristic curve of 0.821. Logistic regression analysis of 242 parameter variables obtained from 888 images shows that the most important features in the final model, measured by Wald Chi-square significance, are the lengths of two peripheral inter-color boundaries and one measure of boundary overlay by different colors. The median split algorithm is fast, requiring less than one second per image and only a four-color splitting, but it captures sufficient critical information regarding color disorder, with peripheral inter-color boundaries showing the highest significance for melanoma discrimination.

1 INTRODUCTION

Early detection of melanoma may be aided by analytic methods applied to dermoscopy images of melanoma, which offer the possibility of detecting potential melanomas before they are sufficiently advanced to affect life expectancy. Color methods splitting the entire lesion were investigated by Andreassi et al. Eccentricity of color components and presence of color islands were important in discriminating melanomas from benign lesions (Andreassi et al., 1999). Colors were also used to discriminate melanomas from benign lesions based on three-dimensional color probability histograms that were measured by both crisp methods (Faziloglu et al., 2003) and fuzzy logic methods (Khan et al., 2009). In this paper, we describe a technique termed the median split technique (Umbaugh, 2011), which we use to capture the architectural disorder of early *in situ* melanoma. This technique has the advantages of speed, simplification of lesion architecture yet retention of critical features, and high discriminatory power for melanoma.

2 MEDIAN SPLIT ALGORITHM

The median split algorithm has been previously applied to entire images using CVIPtools (http://cviptools.ece.siue.edu). This algorithm is based on the Heckbert color compression algorithm (Heckbert, 1982). In this research, we apply the median split algorithm to a specific region of interest (ROI)—the lesion only. The motivation is to quantize the ROI so that fewer colors are used in order to describe the ROI. The simplified image allows exact quantization of color values, color areas, and inter-color boundaries. The color space segmentation is performed by splitting the pixel histogram of a color segment. Each iteration splits this color segment into two segments with equal pixel populations. The segment with the highest range in any color axis is chosen for the subsequent split. Within the chosen segment, the split is performed along the color axis with the highest range. The division occurs at the median pixel *m* on the chosen axis. Formally, the chosen color axis and chosen color bin satisfies:

Ghantasala K., Chowdhury R., Guntupalli U., Hagerty J., Moss R., Rader R. and Stoecker W..
The Median Split Algorithm for Detection of Critical Melanoma Color Features.
DOI: 10.5220/0004304904920495
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 492-495
ISBN: 978-989-8565-47-1

$$chosen\ bin = \operatorname{argmax}_{bins}\ \max_{color\ axis\ in\ bin}(range\ of\ color\ axis) \quad (1)$$

$$chosen\ color\ axis = \operatorname{argmax}_{color\ axis\ in\ chosen\ bin}\ range\ of\ color\ axis \quad (2)$$

$$\{hist(bin) < m\} = \{hist(bin) > m\}\ along\ chosen\ axis \quad (3)$$

where {} denotes the number of pixels satisfying the condition within {}; and hist(bin) denotes the histogram of pixels within the chosen bin. The two new bins replace the old bin in the list of color ins for the ROI. After a given split, the segment (bin) list is updated, and again the segment with highest range along any color axis is chosen for the next split. The algorithm continues until the specified number of colors is reached. Umbaugh found four colors sufficed to represent the colors within a lesion (Umbaugh, 2011). Figures 1-3 show the splitting process for a melanoma image using range as the decision variable in red, green and blue (RGB) color space as performed on the ROI—the melanoma lesion only. Pixels in a bin are mapped to the largest region remaining in the final image.

2.1 Median Split Example

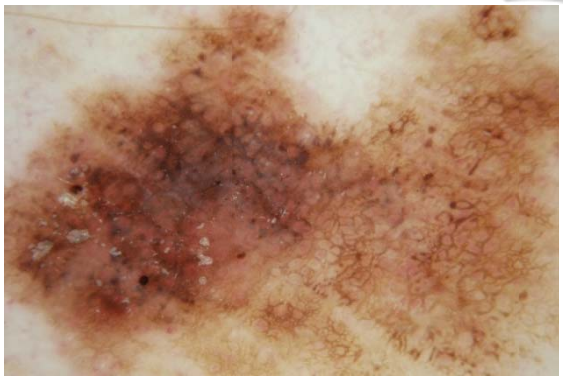

Figure 1: Original melanoma: dermoscopy image.

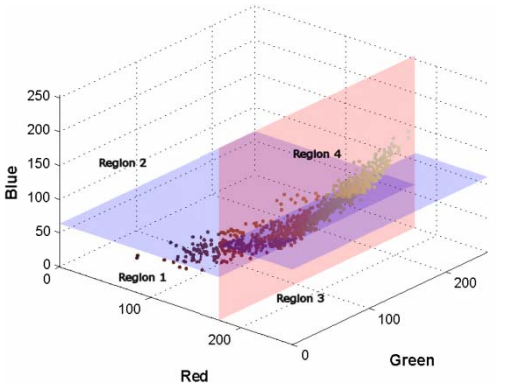

Figure 2: Median split, 3 splits (RGB) create 4 color regions.

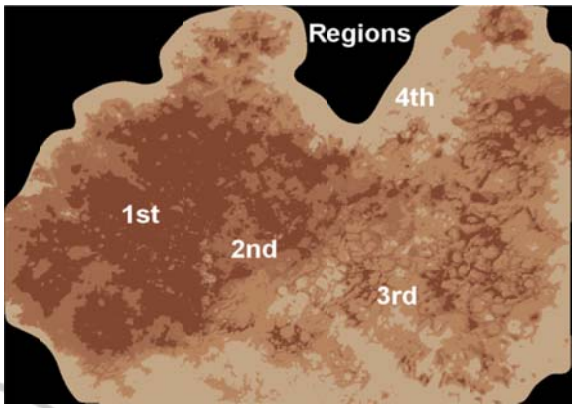

Figure 3: Median split, 3 splits create 4 colors in the lesion.

2.2 Median Split Pseudocode

for no_of_splits (outer loop begins)

for no_of_current_bins (inner loop)

-> find channel with largest range end of inner loop

-> find a bin with the largest range

-> arrange pixels in bin with largest range in order of channel with largest range

-> split bin into nearly equal bins (as equal as possible after splitting)

-> remap colors in bins to image (the pixels in bin are set to mean of bin)

end of outer loop

We set no_of_splits = 3. Three channels are used for each bin: red, green and blue, but in practice, any three channels may be used, for example, principal component axes.

2.3 Median Split Decision Variable

Range is chosen as the median split decision variable. Alternate decision variables may include variance and inter-quartile distances. For this domain, contact dermoscopy images with evenly distributed LED lighting, there is little random noise. Pre-processing allows elimination of nearly all significant artifacts, e.g. hairs protruding from the skin and bubbles from the gel interface. Accordingly, we retained the CVIPtools default decision variable: range within a given color plane. This has potential to capture small variations in the image, such as small networks that would not be captured well by variance or inter-quartile distance.

3 EXPERIMENTS

3.1 Instrumentation and Images

The contact dermatoscope used in this study is the 3Gen DermLite Fluid attachment (3Gen LLC, San Juan Capistrano, CA). This device uses bright white LED lights, 10X magnification and a gel interface.

3.2 Images

A set of 195 melanoma and melanoma *in situ* images were obtained in the study SBIR R44 CA-101639-02A2 of the National Institutes of Health (NIH). A similar benign set of 693 images were obtained for the same study. Lesion borders (ROIs) were manually drawn using second-order b-splines.

3.3 Median Split Features Obtained for Each Lesion

1. Ring Value: Fraction of pixels in the border of the lesion that overlap with the pixels in the color.

2. Total Lesion Area

3. Ratio of Each Color to Peripheral Ring

4. Average R, G, B values in Each Color Region

5. Number of Blobs before filling for each Color region: This gives the number of blobs in each region without filling holes or filtering the small blobs. These blobs are the connected contours present in that segment.

6. Number of Blobs after filling for each Color: Each color blob with area less than 26 pixels is eliminated and interior holes in the remaining blob are filled.

7. Area of Each Color

8. Area of largest Blob: Area occupied by largest blob in each color before and after filling the holes.

9. Internal/External Perimeter of each Color

10. Perimeter of largest Blob: Total internal and external perimeter of largest blob for each color

11. Normalized Perimeter: Found by dividing perimeter by square root of lesion area.

12. Perimeter Intensity Drop: Average RGB color difference between the pixels that are on the border of each color and those one pixel outside the border.

13. Normalized Intensity Drop: Perimeter intensity drop divided by square root of total lesion area.

14. Centroid of each Color: The centroid of the lesion and of each segment by Matlab region props.

15. Euclidean Distance: Distance between the centroid of each color and the centroid of the lesion.

16. Normalized Distance of each Color: The ratio of Euclidean distance and square root of color area.

17. Absolute, Background and Relative luminance for each Color: Absolute luminance is the measure of brightness of the desired segment: **Luminance = 0.30R + 0.59G + 0.11B.**

18. Average background skin R, G, B Values: Average Red Green Blue values of the skin color that surrounds the lesion part in the original image.

19. Relative RGB values for each Color

20. Average Euclidean distance of each Color

21. Red Chromaticity

4 RESULTS

4.1 Performance

The median split algorithm in C running on a Core™ 2 duo processor requires < 1 second/image.

4.2 Most Significant SAS Features

Logistic regression analysis using Statistical Analysis Software (SAS) was used to analyze the results. The significant features described in §3.3 are shown by Chi-sq and p-value statistics (Table 1).

Table 1: Significant parameters (Section 3.3).

Feature	Chi-Sq	P value
Perimeter of largest blob before filling (§ 3.3.10)	84.31	<.0001
Ring value (§ 3.3.1)	28.30	<.0001
Normalized perimeter of largest blob before filling (§ 3.3.11)	19.72	<.0001

4.3 Receiver Operating Characteristic (Roc) Curves

Accuracy of the algorithm was tested using the Receiver Operating Characteristic (ROC) curve. The ROC curve plots sensitivity versus one minus specificity. We report here the area under the curve (AUC) for five experiments. The five ROC curves (Figure 4) depict results for five cases, adding different parameters from perimeters measured on all four color segments.

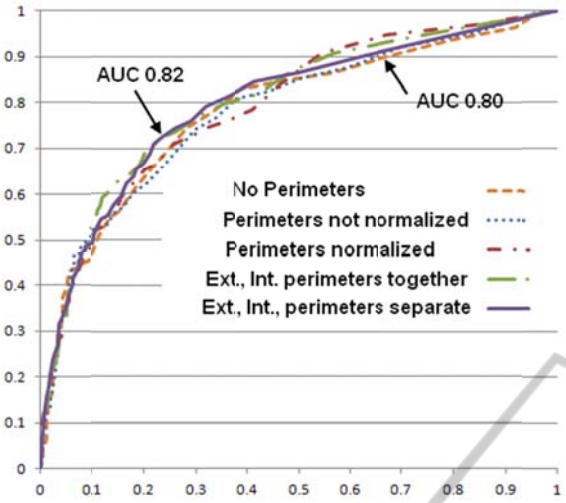

Figure 4: ROC curves, sensitivity vs. 1-specificity, for 5 perimeter cases. The best case for normalized perimeters with separate external and internal color perimeters has AUC = 0.82.

5 DISCUSSION

5.1 The Median Split Algorithm and the Goal of Capturing Architectural Disorder

A new dermoscopy classification algorithm used by clinicians emphasizes architectural disorder—the "A" in the CASH algorithm (Henning et al, 2007). CASH is an acronym for Color, Architecture, Symmetry and Homogeneity. Melanomas have different colors, haphazard architecture, color asymmetry, and are inhomogeneous. The median split images shown capture all four of these characteristics. The image is quantified into a small number of colors (four colors suffice) enabling us to make the measurement of the length of the inter-color boundaries. These inter-color boundary lengths are high in importance as measured by logistic regression analysis. The median split color algorithm appears to capture the haphazard color distribution in melanoma.

5.2 Median Split Channels and Channel Pre-processing

The choice of the RGB color space rather than alternate color spaces is arbitrary. The median split algorithm could be extended to other color spaces, including L*a*b*, HSV, and CIE-LUV color spaces. The method could also be extended to the principal component transform applied to these color spaces.

6 CONCLUSIONS

The median split algorithm was applied to 888 melanoma and benign dermoscopy images. Separation of melanomas from the benign lesions as measured by area under the ROC curve was as high as 0.82. This method was applied to a difficult set of melanomas, the majority at the *in situ* stage, and therefore is not directly comparable to earlier studies (Andreassi et al., 1999), (Khan et al., 2009). The arbitrary selection of four colors captures color disorder. The most critical features were derived from inter-color perimeters.

REFERENCES

Andreassi, L., Perotti, R., Rubegni, P., Burroni, M., Cevenini, G., Biagioli, M., Taddeucci, P., … Barbini, P. (1999). Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma: a new quantitative semiology. *Archives of Dermatology*, 135(12), 1459-65.

Faziloglu Y, Stanley RJ, Moss RH, Van Stoecker W, McLean RP. (2003). Colour histogram analysis for melanoma discrimination in clinical images. *Skin Research and Technology*. 9(2), 147-156.

Heckbert, P. (1982). Color image quantization for frame buffer display. In Proceedings of *SIGGRAPH '82*, 297.

Henning, J.S., Dusza, S.W., Wang, S.Q., Marghoob, A.A., Rabinovitz, H.S., Polsky, D., & Kopf, A.W. (2007). The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy. *Journal of the American Academy of Dermatology*, 56(1), 45-52.

Khan, A., Gupta, K., Stanley, R.J., Stoecker, W.V., Moss, R.H., Argenziano, G., … Cognetta, A.B. (2009). Fuzzy logic techniques for blotch features evaluation in dermoscopy images. *Computerized Medical Imaging and Graphics*, 33(1), 50-57.

Umbaugh, S.E. (2011). *Digital Image Processing and Analysis: Human and Computer Vision Applications with CVIPtools* (2nd ed.). Boca Raton: CRC Press.