Universidade Federal do Estado do Rio de Janeiro
Escola de Informática Aplicada
Bacharelado em Sistemas de Informação

Juliana Louback

Dermatologic Diagnosis through Computer Vision and Pattern Recognition

Rio de Janeiro, Brazil
2014

Juliana Louback

Dermatologic Diagnosis through Computer Vision and Pattern Recognition

Monograph presented at the Universidade Federal do Estado do Rio de Janeiro, developed upon completion of the Bachelor's in Information Systems.
Concentration: Computer Vision
Mentor: Sean Wolfgand Matsui Siqueira

Rio de Janeiro, Brazil
2014
Juliana Louback

Dermatologic Diagnosis through Computer Vision and Pattern Recognition

Monograph presented at the Universidade Federal do Estado do Rio de Janeiro, developed upon completion of the Bachelor's in Information Systems.

Concentration: Computer Vision

Date of defense: June 12th, 2014

Result: _____

Examining Committee

John Doe _____
Universidade Federal do Estado do Rio de Janeiro

John Doe _____
Universidade Federal do Estado do Rio de Janeiro

John Doe _____
Universidade Federal do Estado do Rio de Janeiro

I would like to dedicate this article to prof. Rob Fergus who first introduced me to the field of Computer Vision; and Dr. Jefferson Braga Louback, my father, whose devoted efforts in enabling access to medical care have long inspired my admiration.

Acknowledgements

I would like to express my gratitude to my mentor, prof. Sean Wolfgand Matsui Siqueira without whose guidance this article would never have been completed; prof. Rob Fergus, prof. Richard Staunton and prof. Li Ma who provided invaluable advice regarding the development of the algorithms used; Tom Gibara for his most effective implementation of the Canny Corner Detector and Yuri Pourre for permitting the use of his Quick Hull algorithm; Dr. Jefferson B. Louback and Dr. Carlos Martins who acted as consultants for the medical aspects of the article, which are beyond my domain of knowledge.

Summary

This project endeavors to quantify the effectivity of computer vision and pattern recognition techniques in the diagnosis of melanoma. In dermatology, the ABCD guide is a widely known method for the detection of a malignant melanoma. This guide is a set of four identifying traits: Asymmetry; Border irregularity; Color changes; Diameter greater than 6 mm. Over the past decade, various algorithms that examine digital images of skin lesions to detect the presence of malignant melanomas have been developed and tested. The algorithms use the analysis of border irregularity to provide a diagnosis. The most recent of these was developed by Dr. Richard Staunton of the University of Warwick and Li Ma of the Hangzhou Dianzi University and presents the best results with regards to accuracy and performance. The experiment proposed is to attribute a metric to all 4 categories of the ABCD melanoma diagnosis criterium so that they may be used in conjunction and increase levels of accuracy. This strategy is evaluated in comparison to Staunton and Ma's algorithm to ascertain the optimal solution. Both algorithms are run on a labeled dataset of 480 images and the specificity, the number of accurately diagnosed images of melanoma, is calculated. The algorithm's sensitivity, though of lesser importance, is also taken into consideration in the final evaluation. In conclusion, a practical application of these techniques is explored with the intent of providing solutions to the prevalent issue of limited access to medical care. The process, maturation and findings of the undertaking are detailed in this monograph.

List of Tables

List of Figures

Index

INTRODUCTION

The first section of this article provides a high-level understanding of the established objectives in addition to the tools and components necessary to reach them. First the scope of analysis is defined in order to demonstrate the importance and potential value which may be added. A brief overview of existing solutions is then presented, in addition to a description of the novel proposal this experiment is based on.

1.1 The Scope of Analysis

There is a noted undersupply of dermatologic services worldwide, common to both developing and developed countries. The effects of such a deficit in the workforce is clearly shown in the excessive mean wait times for appointment availability; it was recently discovered that a patient must schedule a dermatologic consult an average of 33 days in advance in the United States[8] and 26 days in advance in Brazil[16] .

Estimates of the current number of practicing dermatologists range from 8,000 to 8,5000 in the United States and approximately 6,000 in Brazil. This is a pitifully inadequate number of professionals to cater to the 300 million americans and 200 million brazilians. Researchers from the Case Western Reserve University and Yale University performed a study that showed the direct correlation between dermatologist density and melanoma mortality rates[1]. This finding is already significant in countries like the United States and Brazil where there is an estimated ratio of 35,000 people per dermatologist and even more severe in countries like South Africa with 3 to 4 million people per dermatologist[14].

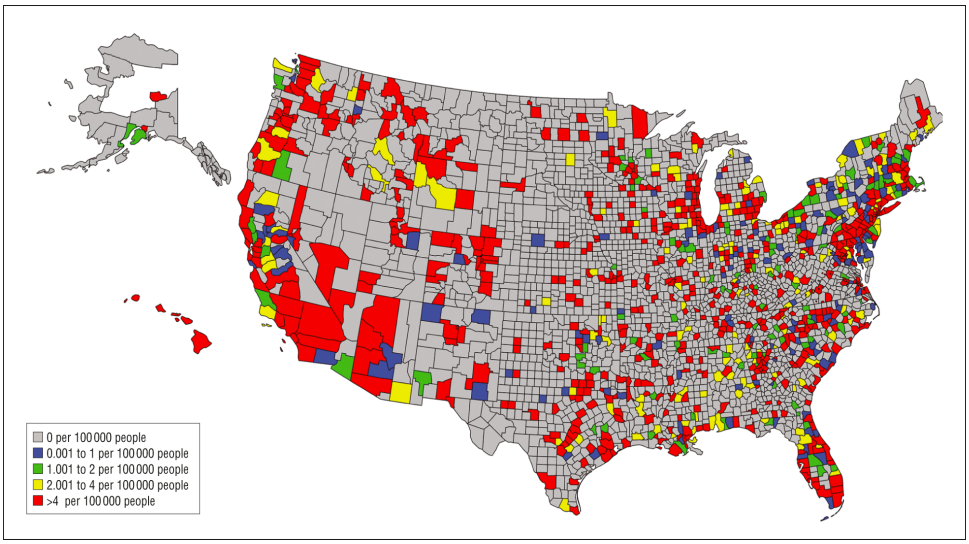
Dermatology specialists are mainly based in capitals and major cities, limiting greatly the access to a dermatologist. Technology has previously been applied to other fields to overcome geographical barriers; this project evaluates the efficiency and effectiveness of a series of computerized methods that may be applied to the dermatologic diagnosis process, in an effort to assuage this disequilibrium in supply and demand.

In the field of dermatology, the vast majority of medical cases are diagnosed

visually, as the affected regions are at least partially superficial. According to CDC's National Ambulatory Medical Care Survey in 2009, the top 5 reasons given by patients for visiting dermatologists were actinic and seborrheic keratosis, benign neoplasm, acne, malignant neoplasms and contact dermatitis and other eczema. These conditions account for near 80% of medical consults. Of the 5 listed, 4 can be diagnosed through a visual examination. Only in the case of a malignant neoplasm would it be necessary to perform a biopsy to confirm the diagnosis, but the potential malignancy is recognized during the initial (visual) examination. Due to this rather unique situation, image analysis systems may be a viable solution to perform a triage or possibly a full dermatologic diagnosis.

The scope of this project will be limited to neoplasm diagnosis, using binary classification algorithms with malignant/benign as prediction categories. This study will implement and evaluate two image analysis systems to establish the optimal solution. These systems are separated into two different categories based on the strategy employed for classification. The first strategy involves analyzing images individually based on a formula obtained from the algorithm. This approach includes the re-implementation of Staunton and Ma's algorithm detailed in the article *Analysis of the contour structural irregularity of skin lesions using wavelet decomposition.* This algorithm used border irregularity to determine malignancy. On the same approach, four systems are developed to analyse all four identifying traits used to diagnose malignant neoplasms. A metric for each of the four traits is attributed to the image of the skin lesion, which are run through a linear regression to estimate the likelihood of the presence of melanoma. The final algorithm adopts a different strategy using a supervised learning model to perform pattern recognition on a substantial image database. This algorithm is included in the study as it is capable of incorporating new data. It is proposed that with the expansion of the training database and further refining of the model's algorithm, this method may be applied to a broader spectrum of skin lesion types, enabling a specific diagnosis of diseases and conditions beyond neoplasm diagnosis.

Figure 1 - Mean dermatologist density among US counties, 2002-2006



Source: ANEJA, S; ANEJA, S; BORDEAUX J.S. [1]

1.2 Prior Art

In dermatology, the ABCD guide is a widely known method for the identification of a malignant melanoma[17]. This guide is a set of 4 traits common to malignant melanomas: Asymmetry; Border irregularity; Color changes; Diameter greater than 6 mm. Extensive research has been performed regarding the analysis of border irregularity in relation to melanoma diagnosis. Some of these studies indicate that the diagnosis of malignant melanomas can be based on the analysis of the shape of the lesion alone[10].

A study performed by K.M. Clawson of the University of Ulster uses the Harmonic Wavelet Transform to analyze lesion border irregularity, claiming maximum classification accuracy of 93.3% with 80% sensitivity[4]. However, the classification algorithms for differentiating between benign and malignant tumors were applied to only 30 cutaneous lesions. A more recent study performed in 2012 by Li Ma of the Hangzhou Dianzi University and Richard C. Staunton of the University of Warwick[12] involved 134 images of skin lesions, of these 72 were of melanomas and 62 of moles. The algorithm could be summarized as a two-step procedure: multi-scale wavelet decomposition of the extracted contour followed by the selection of significant sub-bands.

Wavelet decomposition was used to extract the structure from the contour which were then modeled as signatures with scale normalization to give position and

frequency resolution invariance. Energy distributions among different wavelet sub-bands were then analyzed to extract those with significant levels and differences to enable maximum discrimination. A set of statistical and geometric irregularity descriptors were applied at each of the significant sub-bands, followed by an effectiveness evaluation to select which descriptors contribute to an accurate diagnosis. The effectiveness of the descriptors was measured using the Hausdorff distance between sets of data from melanoma and mole contours. The best descriptor outputs were input to a back projection neural network to construct a combined classifier system.

Li and Staunton's optimum combination resulted in an area under ROC curve of 0.89 with 90% specificity and 83% sensitivity, similar to Clawson's findings. These results were obtained from a small data set consisting of 18 images, 9 of which were of melanomas and 9 of moles. With a larger training set of 67 images (31 moles and 36 melanomas), these numbers fell to 0.83 ROC with 83% sensitivity and 74% specificity, which signifies that a greater number of non-malignant moles would be classified as melanomas.

Table 1 - Comparison of classification performances

Scheme	Specialty	Sensitivity	Area of ROC
Single scale features with small sample set	0.64566	0.69444	0.69534
Single scale features with large sample set	0.87903	0.55556	0.75179
Original multi-scale features with small sample set	0.83871	0.69444	0.81541
Original multi-scale features with large sample set	0.80645	0.69444	0.81989
Selected multi-scale features with large sample set	0.74194	0.83333	0.83333
Selected multi-scale features with small sample set	0.90323	0.83333	0.89068

Source: MA, L.; STAUNTON, R.C. [12]

These findings are incredibly significant given the high levels of accuracy and reduced computational cost. It will be of interest to analyse this algorithm in comparison to others that base their classification system on the remaining three features: asymmetry, color change and diameter size.

1.3 Proposed Objectives

The end goal of this experiment is to determine which image recognition algorithm or what combination of algorithms presents the most accuracy for the diagnosis of melanoma, and how this solution may be applied to the dermatologic practice..

The three algorithms, the ABDC Linear Regression, Staunton and Ma's Border Irregularity Analysis and the Support Vector Machine model, are measured for sensitivity and specificity levels in addition to overall accuracy.

The Support Vector Machine model is used to analyze the images as a whole, requiring a substantial training data set for its classification system. The motive behind using Support Vector Machines is to explore the possibility of applying incremental learning to the diagnosis algorithm. If this were to prove effective, it is likely that the algorithm will evolve as its training data set increases.

Sensitivity will be given a higher priority than specificity, as the correct identification of a malignant or potentially malignant tumor is of greater consequence than a false positive diagnosis, within reasonable levels. However, the inconvenience of an incorrect positive diagnosis is by no means irrelevant, therefore despite specificity being a secondary priority, ideally the solution proposed will result in no less than 80% specificity.

Once the algorithms have been compared and measured for accuracy, a proposal for practical application will be formulated. As 100% accuracy is an unrealistic expectation, specialist confirmation is of importance. This project will study whether it is feasible and advantageous to incorporate computer vision algorithms to current traditional diagnosis methods.

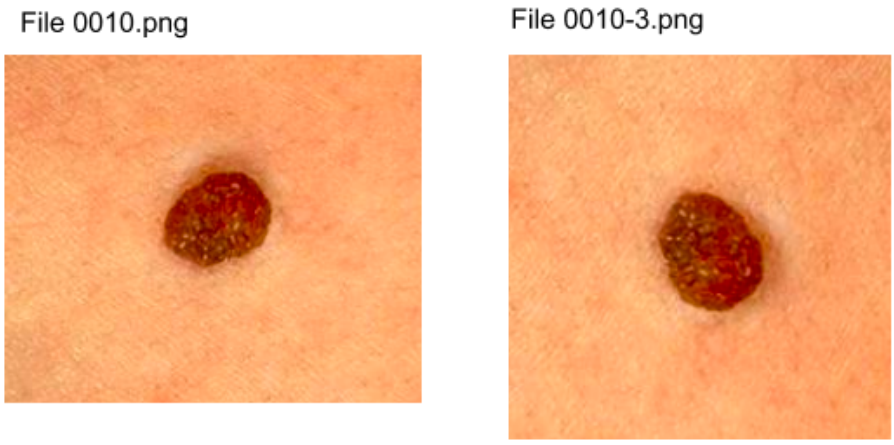
DEVELOPMENT

The following chapter details the development of the algorithms used in this experiment. Difficulties encountered and modifications to the original plan of action are included due to their contribution to the overall knowledge acquired and possibly influence over final results. First the implementation of the ABCD Linear Regression algorithm is described, followed by the re-implementation of Ma and Staunton's Border Irregularity algorithm[12].

2.1 Image pre-processing

In preparation for processing, each of the images is run through an application that rotates the image 30* counter-clockwise a total of 12 times, storing a copy of the image after each rotation. This is done by suggestion of prof. Rob Fergus as a strategy to boost the image dataset to be used in the evaluation of the algorithms. The files of the rotated images are named after the original image ID for control purposes, see Figure 2.

Figure 2 - Image rotation



Original file 0010.png and processed image 0010-3.png, representing the image after 3 consecutive rotations of 30°.
Image source: www.onecaremedicalcenter.com

Source: Processing performed on the 22nd of February 2014

2.2 ABCD Linear Regression

2.2.1 Canny Corner Detector

The first step is to single out the skin lesion from the image. In doing so, the amount of data to be processed reduced considerably. Furthermore, considering the fact that Border Irregularity is one of the four identifying traits of melanoma, extracting the border of the lesion is essential to the success of the 'B' component of the ABCD Linear Regression algorithm. For this phase of the process, Tom Gibara's Java implementation of the Canny Corner Detector[5] algorithm is used. The algorithm, detailed in the article *A Computational Approach to Edge Detection* by John Canny, is notably effective and as such it is ubiquitously used. It can be resumed in four main steps: Image filtering; gradient magnitude computing; Non-maximum suppression and hysteresis thresholding to trace the edges.

1 - Image filtering

The Canny Edge detector is highly susceptible to noise; the first step in the algorithm is filtering the image to reduce the misleading effect of noise pixels through a step edge detector. Canny calculated the optimal filter for this task, named 'Filter number 6' in his article[2]. However, he observed that although the first derivative of Gaussian operator performed approximately 20% worse than the optimal operator in the performance evaluations, this difference is hardly noticeable when visualizing their effects on real images. As the optimal operator requires much more computational effort than the first derivative of the Gaussian, the image is convolved with a Gaussian filter to obtain the desired noise reduction.

Figure 3 - Comparison of optimal and Gaussian operator.

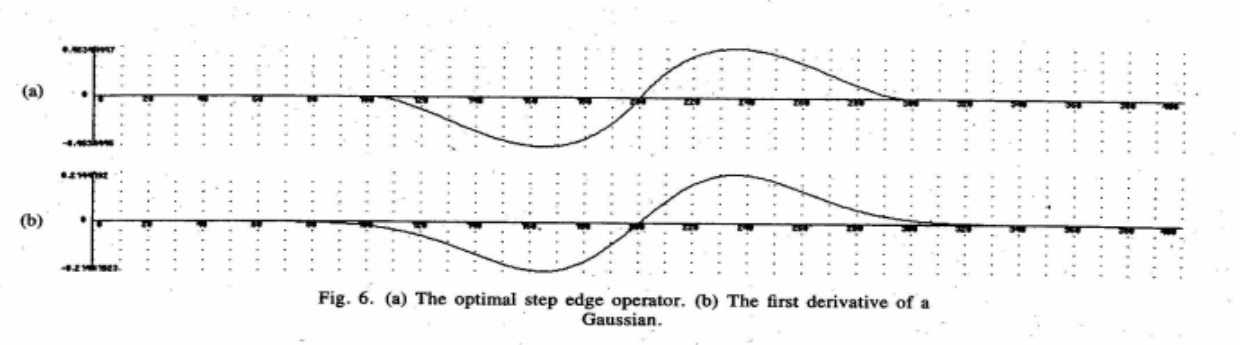


Fig. 6. (a) The optimal step edge operator. (b) The first derivative of a Gaussian.

Source: CANNY, John. [2]

2 - Gradient magnitude computing

The gradient of an image indicates the direction of a rapid change in intensity and as such provides information regarding the orientation of the edge, whether it is horizontal, vertical or diagonal. In a two dimensional coordinate system as is this case, the gradient is given by

$$\nabla f = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}$$

the vector composed of the partial derivatives of f. A horizontal gradient is given by

$$\nabla f = 0 + \frac{\partial f}{\partial y}$$

as the change is in the y direction; a vertical gradient is given by

$$\nabla f = \frac{\partial f}{\partial x} + 0$$

as the change is in the x direction. The magnitude of the gradient is given by

$$||\nabla f|| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

and is calculated for every pixel in the image.

3 - Non-maximum suppression

Once the gradient magnitude calculation is performed on each pixel, its value is then verified to determine whether it assumes a local maximum in the gradient direction. The gradient direction is given by

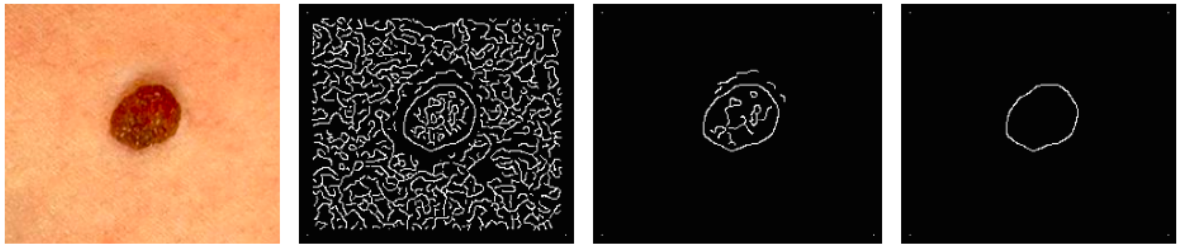
$$\theta = tan^{-1}\left(\frac{\partial f}{\partial x} / \frac{\partial f}{\partial y}\right)$$

The implementation of the algorithm does not compute the gradient direction so as to avoid performing a division calculation. Instead, the two derivatives are checked for the same sign and then the largest of the two derivatives is singled out. The pixel magnitude is then compared to its two neighbor's values in the four possible directions, these being north - south, east - west, northeast - southwest and northwest-southeast. Linear interpolation is used between the two neighbor pixels for greater accuracy. Only local maximums will be considered edge candidates, therefore if the central pixel's magnitude is not greater than that of its two neighbors it will be 'suppressed' by setting its edge strength value to zero.

4 - Hysteresis thresholding

The pixels that correspond to a local maximum and in effect possess a high edge strength value are set aside as the edges detected in the image. Yet the selected edge pixels can be further refined through specifying a threshold as the final comparison to determine an existing edge. Hysteresis thresholding is done using a low and a high threshold. The low threshold detects weak edges and the high threshold the strong edges. It was necessary to experiment to determine the optimal combination of thresholds for the situation in question, that of singling out the skin lesion in the image.

Figure 4 - Threshold performance comparison



The original image | Border detection with a low threshold of 0.5 and high threshold of 1. | Border detection with a low threshold of 2 and high threshold of 5. | Border detection with a low threshold of 2 and high threshold of 9.5.

Source: Testing performed on March 2nd, 2014.

The low threshold detected texture traces and fine edges; as the skin lesion has a considerably distinct edge a high threshold was proven to be the most adequate. With melanoma lesions the contour is not continuous; this of course is expected due to the lesion's characteristic asymmetry, border irregularity and color change. Ergo, it is presumed that this will not be misleading in the final results.

Figure 5 - Threshold performance on image of melanoma

The original image | Border detection with a low threshold of 2 and high threshold of 9.5.

Source: Testing performed on March 2nd, 2014.

During the course of testing, it became apparent that the algorithm as it is will only function properly if applied to images with a single lesion. Additionally, the images may need to be adapted to eliminate non-skin elements. In Figure 5, the original image shows what appears to be the hem of an item of clothing in the upper right corner. This is classified as a strong border as displayed in the image tracing the detected borders; if not treated, it will be inaccurately handled as part of the skin lesion in the subsequent phases of the algorithm.

The implementation of the Canny Border Detector developed by Tom Gibara was slightly modified so as to return a list of coordinates of the detected edge pixels as opposed to an image of the traced edges.

2.2.2 Asymmetry metric

Asymmetrical skin growths, in which one part is different from the other, may indicate melanoma.[22]

Symmetry is defined as invariability regardless of transformations. It is an absolute characteristic and cannot be measured in degrees. Due to this most rigid definition, rarely if ever is it possible to label a figure as symmetrical. As the images to be analyzed are of the human body, it is safe to assume that all the figures will be asymmetric. Instead of defining a figure as symmetric or asymmetric, the asymmetry will be measured and compared to an established threshold.

Symmetry in medical imaging has played an important role in contributing to diagnosis in other fields of medicine. An example of this is measuring asymmetry on mandibles from children with cleft lip and palate and children with plagiocephaly syndrome,[7] measuring the asymmetry of the hippocampi to classify schizophrenic patients[7] and using the asymmetry principle in the detection of breast tumors[9].

In examining skin blemishes, rotation and reflection symmetry is of greater relevance. Two of the more recent rotation/reflection symmetry detection algorithms are G. Loy and J. Eklundh's Detecting symmetry and symmetric constellations of features[11] and V. Prasad and L. Davis's Detection rotational symmetries[19]. G. Loy and J. Eklundh's algorithm is feature based; it uses pairwise matching and voting for symmetry foci in a Hough transform to identify asymmetry. V. Prasad and L. Davis

created an algorithm that filters in an input color image into a gradient vector flow field, extracting and matching the features into the gradient vector flow field, using a voting scheme for symmetry detection.

A comparative analysis of both algorithms[3] shows that the first algorithm has a higher sensitivity rate than the second; however it also presents very high false-positive rates which are of serious consequence in the medical field.

Tom Gibara's Symmetry Detection Algorithm[6] provides a method to measure asymmetry using a comparative threshold to identify rotational symmetries, more easily applicable to the analysis of the skin blemishes.

The algorithm proceeds as follows:

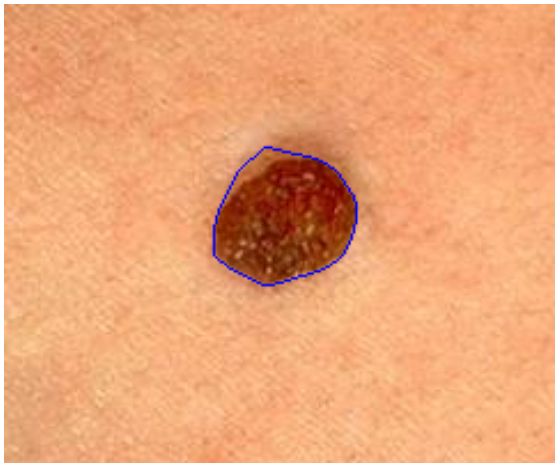
1. Identify the centroid of the object.
2. Measure the object radius (distance from centroid to most distant object pixel).
3. Choose a set of circles centered on the centroid with radius less than the object radius.
4. Sample the image at a fixed angular resolution (an even number of equiangular points) to create a vector of pixel values for each circle.
5. 'Convolve' each vector with itself to create a new set of vectors. These loosely measure the reflectional symmetry across the angle associated with each element.
6. Sum all of the resulting vectors to obtain an overall symmetry score for each angle considered.
7. Disregard any angle that does not exceed a predetermined threshold and which is not a local maximum.
8. Of the remaining angles, calculate the score-weighted average of adjacent angles (subject to a predefined threshold).
9. The resulting angles, together with the centroid, define a set of axes along which reflectional symmetry is high.

Source: GIBARA, TOM. [6]

Before running the Symmetry Detector on the images, it is necessary to delimit the region to be examined. The border traced by the Canny Edge Detector will likely not provide a continuous contour for images of melanoma lesions as they are generally multi-colored and without distinct borders (*See Figure 5)*. As it happens, there is no guarantee that even the analysis of benign lesions will result in a continuous contour.

The solution to this issue is found in Yuri Pourre's Quick Hull algorithm[18], which draws the smallest polygon possible given a set of coordinates. Once the Canny Edge Detector is run on the image, the resulting list of coordinates of the detected edge pixels is fed into the Quick Hull program.

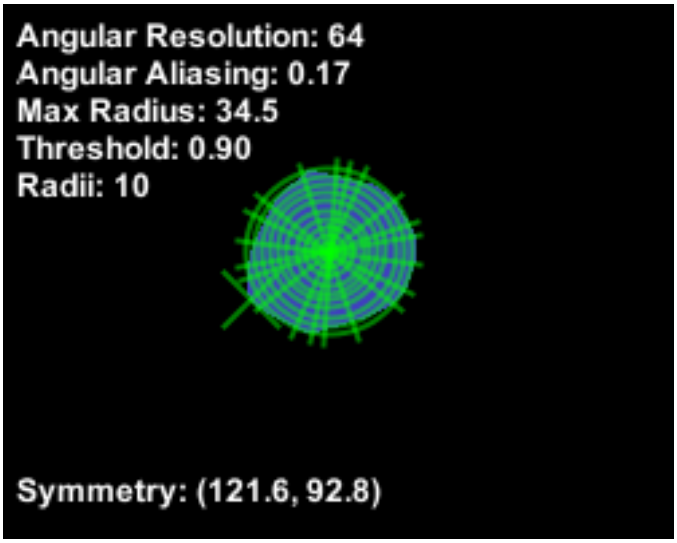
Figure 6 - Quick Hull generated polygon containing the skin lesion



Source: Testing performed on March 9th, 2014.

An image with the polygon outline is then filled in and converted to grayscale for compatibility with Gibara's Symmetry Detector algorithm. The important algorithm parameters[6] are Angular Resolution, Angular Aliasing, Radius Count ('Radii' in the sample output in Figure 8), and Threshold. The Angular Resolution is the number of sample arcs in the circle. Angular Aliasing is the smallest angle permitted between identified axes of symmetry. Angles closer together than this value are combined into a single angle. Radii indicates the number of different radii at which samples are taken. Threshold is the proportion of the maximum possible score that an angle must obtain to be considered. Once the detector is run, a summary of the execution is displayed. The above parameters and their respective values are listed. Blue pixels indicate pixels identified with the object, in this case the polygon surrounding the skin lesion. The green 'plus' indicates the position of the centroid, currently not visible due to the many traced lines of reflectional symmetry. The green 'cross' indicates a pixel at maximum distance from the centroid. Green circles indicate the circles from which samples of image data were taken.Green lines indicate the lines of reflectional symmetry identified by the algorithm.

Figure 7 - Summary of Symmetry Detector results



Source: Testing performed on March 9th, 2014.

2.2.3 B - Border Irregularity metric

Melanomas may have borders that are vaguely defined. Growths with irregular, notched or scalloped borders need to be examined by a doctor.[22]

As benign lesions are of a circular nature, it is expected that the radius of the lesion will be constant, or more specifically very nearly so, throughout the lesion. The logic behind the calculation of this border irregularity metric is very straightforward. Succinctly, it entails calculation the coefficient of variance of the lesion's radii.

1. A list of coordinates of the lesion's border is obtained from the Canny Corner Detector;
2. The centroid of the lesion would calculated by finding the mean x and y values; however as the contour of melanoma lesions is not continuous, it was necessary to correct this formula. The coordinates of the centroid are the minimum value x and y plus an offset of the difference between the maximum and minimum value of x and y divided by 2.

$$C_x = x_m + (x_M - x_m)/2 \qquad C_y = y_m + (y_M - y_m)/2$$

3. The mean radius is obtained by calculating the average of the euclidean distance between the centroid and each point in the list of border coordinates;

4. The standard deviation is given by

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \bar{r})^2}$$

with n as the number of coordinate pairs (points) in the list, r_i as the radius at point i and $\bar{r}$ as the mean radius. Bias correction is unnecessary given the sample is equal to the population.

5. Obtain the coefficient of variance given by

$$c_v = \frac{\sigma}{\mu}$$

being that μ is the mean radius calculated in step 3.

Figure 9 - Comparison of metric on malignant and nonmalignant lesions

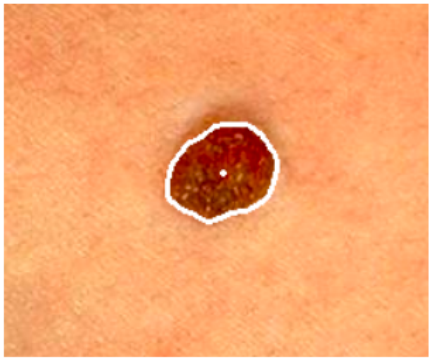


Image of a benign growth and the tracing in white of the detected edge and centroid.
Border Irregularity metric: 0.107

Image of a melanoma lesion and the tracing in white of the detected edge and centroid.
Border Irregularity metric: 0.132

Source: Testing performed on March 5th, 2014.

C - Color Changes

Multiple colors or uneven distribution of color may indicate cancer.[22]

Depending on the resolution and lighting of the image taken of the skin lesion, color variance may be noticeable even in benign lesions. This makes it necessary once again to measure the difference between colors in the lesion and compare this difference to a defined threshold. The difference between colors is represented by

Delta-E, a metric established by the International Commission on Illumination to quantify color differences. Using this metric, the algorithm will be quite straightforward:

1. Identifying the lesion within the image using the Canny edge detector;
2. Calculating the Delta-E of the selected portion of the image;
3. Establish a color change threshold using a tree model;

Functions to perform the Delta-E calculation are included in many existing image manipulation libraries. The Delta-E measure can then be compared to the threshold to assist in the final diagnosis.

D - Diameter greater than 6mm

A skin growth's large size may be an indication of cancer.[22]

This final criteria presents a challenge for the application of computerized image analysis; the images will be taken at different ranges with no comparative figure, impairing an accurate calculation of real size. A proposed solution is to identify the pores in the image and measure the average distance between them in pixels; this would provide a scale of the lesion's diameter. An adaptation of Q. Zhang and T. Whangbo's Skin Pores Detection Algorithm[21] will be implemented to find the pores in the image and note their coordinates so as to determine the distance between them. The algorithm is based on image segmentation. A preprocessing algorithm to balance the illumination of the image must be run before the pore detection algorithm. The Global Luminance Proportion algorithm is noted below.

Using the original image of MxN size:

1. a. Calculate the average luminance of the image;
   b. Split the image S into V sub-blocks, calculating the average luminance;
   c. Obtain luminance difference matrix D;
2. Interpolation algorithm for matrix D until element number in matrix equals MxN;
3. Merge matrix D and original image S into new image sized MxN.

When the image has balanced luminance, the segmentation is performed using the Fuzzy C-Means Algorithm. After this segmentation, pixels in the image are labeled to 8 connectivity and skin pores can be classified from the 8-connectivity

labeled image by calculating the quadratic moment and the ration between row and column moments.

2.3 Analysis of the contour structural irregularity of skin lesions using wavelet decomposition

Li Ma and Richard C. Staunton of the Hangzhou Dianzi University and Warwick University[12] developed an algorithm using multi-scale wavelet decomposition of the extracted lesion contour followed by the selection of significant sub-bands. Staunton and Ma have generously agreed to having their algorithm included in this study. The steps in this procedure are as follows:

1. Perform a 1D wavelet decomposition to produce a series of detailed contour signals for all the lesion contours in the data set;
2. Calculate energy measures for each of the detail, that is high frequency, sub-bands;
3. Perform a Hausdorff Distance analysis based on energy values in each sub-band to determine which detail coefficients are significant for distinguishing the two sets;
4. Select the significant sub-bands;
5. Label the sub-bands with scales below these as containing the textural component of the contour and [reject] them from further consideration;
6. Sum the approximation coefficients and the remaining detail coefficients to generate the connected structural component of the contour.

This procedure approximately isolates the structural portion of the original contour, and then the irregularity descriptors are used to refine this and partition the lesions appropriately.

2.4 Pattern recognition using Support Vector Machines

Support Vector Machines (SVMs) are non-linear statistical data modeling tools that can be used to model complex relationships between inputs and outputs[13]. It is an algorithm for binary classification developed by Cortes and Vapnik in 1995, noted among other reasons for its effectiveness even with limited training sets. David Meyer described this process as the search for “the optimal separating hyperplane between the two classes by maximizing the margin between the classes closest points - the points lying on the boundaries [of the margin] are called *support vectors*, and the

middle of the margin is the optimal separating hyperplane."[15]

A small number of training samples may degrade the performance of the classifier[20]. As new medical cases arise, data may be added to the training set to increase the number of samples and enhance the effectiveness of the classification algorithm. To incorporate this new data, the system would be adapted to integrate incremental learning. This adaption would entail significant changes and add complexity to the process, and may not be included in this experiment and instead be delegated as ongoing work.

References

[1] ANEJA, S; ANEJA, S; BORDEAUX J.S. **Association of Increased Dermatologic Density With Lower Melanoma Mortality.** Case Western Reserve University, Yale University School of Medicine, 2012.

[2] CANNY, John. **A Computational Approach to Edge Detection.** IEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, NO. 6, November 1986.

[3] CHEN, Po-chen et. al. **A Quantitative Evaluation of Symmetry Detection Algorithms.** 17 pages. Penn State University & Carnegie Mellon University, 2007.

[4] CLAWSON, K.M. et al. **Analysis of Pigmented Skin Lesion Border Irregularity Using the Harmonic Wavelet Transform.** Machine Vision and Image Processing Conference, 2009. IMVIP '09.

[5] GIBARA, Tom. **Canny Edge Detector Implementation.** Available at: <http://www.tomgibara.com/>. Accessed on January 23rd, 2014.

[6] GIBARA, Tom. **Symmetry Detection algorithm.** Available at: <http://www.tomgibara.com/>. Accessed on January 23rd, 2014.

[7] GLERUP, Nanna. **Asymmetry measures in medical image analysis.** 174 pages. Department of Innovation, IT University of Copenhagen, 2005.

[8] KIMBALL, A.B.; RESNECK, J.S. The US dermatology workforce: a specialty remains in shortage. Journal of the American Academy of Dermatology, 2008 Nov; 59(5):741-5.

[9] KURUGANTI, P.T.; HAIRONG, Qi. **Asymmetry analysis in breast cancer detection using thermal infrared images.** Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint.

[10] LEE, Tim K. et al. **Irregularity index: A new border irregularity measure for cutaneous melanocytic lesions.** Vancouver: Elsevier, 2003.

[11] LOY, G.; EKLUNDH, J. **Detecting symmetry and symmetric constellations of features.** European Conference on Computer Vision (ECCV'04), Part II, LNCS 3952, pages 508,521, 2006.

[12] MA, L.; STAUNTON, R.C. **Analysis of the contour structural irregularity of skin lesions using wavelet decomposition.** Hangzhou Dianzi University & University of Warwick, 2012.

[13] MA, Zhen et al. **A review on the current segmentation algorithms for medical images.** 6 pages. Faculty of Engineering, University of Porto.

[14] McKOY, Karen. **The Importance of Dermatology in Global Health.** Harvard Medical School Department of Dermatology.

[15] MEYER, DAVID. **Support Vector Machines.** 8 pages. Technische Universität Wien, Austria, 2012.

[16] MONTEIRO, Fábio. Saúde privada não consegue administrar o aumento da procura. **Correio Braziliense**, May 2011.

[17] NACHBAR, Franz et al. The ABCD rule of dermatoscopy. **Journal of the American Academy of Dermatology**, Volume 30, Issue 4 , Pages 551-559, April 1994.

[18] POURRE, Yuri. **Quick Hull Algorithm.** Available at: <https://github.com/yuripourre/>. Accessed on March 5th, 2014.

[19] PRASAD, V.;DAVIS, L. **Detection rotational symmetries.** IEEE International Conference on Computer Vision (ICCV), pages 346–352, 2005.

[20] RAUDYS J.S; JAIN, A.K. Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners. **IEE Transactions on Pattern Analysis and Machine Intelligence,** Vol. 13, NO. 3, March 1991.

[21] ZHANG, Qian; WHANGBO, TaegKeun. **Skin Pores Detection for Image-Based Skin Analysis.** 5 pages. Department of Computer Science, Kyungwon University, 2008.

[22] **Melanoma pictures to help identify skin cancer.** Available at: <http://www.mayoclinic.com/health/melanoma/DS00575>. Accessed on October 19th 2013.