02 Apr 2007

TELANGIECTASIA DETECTION IN DERMOSCOPY IMAGES

David Erdos

TELANGIECTASIA DETECTION IN DERMOSCOPY IMAGES

David Erdos
Electrical Engineering Department
University of Missouri – Rolla

Dr. Joe Stanley
Research Advisor

April 2nd, 2007

Abstract

Melanoma lesions that exhibit irregular linear vessels can be misdiagnosed due to the fact that their clinical appearance can often closely resemble other skin conditions. Detection these irregular linear blood vessels, or telangiectasia could improve the accuracy of melanoma diagnosis. An algorithm that detects telangiectasia has been developed; it initially searches for color drops in an image in various directions and then uses various post processing techniques to improve the accuracy. The feasibility of using classifiers such as neural networks to further improve the accuracy has also been investigated.

Introduction

The skin is the largest organ in the body, and therefore it's not surprising that cancer of the skin is the most common type of cancer. The vast majority of skin cancer cases are composed of basal cell carcinoma and squamous cell carcinoma. Melanoma is the least common form of skin cancer, but is also the deadliest form of skin cancer, and results in 74% of skin cancer related deaths in the United States[1]. An estimated 10,850 people will die of skin cancer this year, 8,110 from melanoma, and 2,740 from other skin cancers[1].

Researchers have been able to identify various risk factors that are associated with melanoma. These risk factors are: genetic, environmental, previous melanoma, and immunosuppressive. The most common environmental factor in developing skin cancer is ultra-violet (UV) light emitted by the sun or artificial tanning devices[2]. Everybody should practice extensive sun protection, and avoid sun exposure as much as possible, it has also been noted that five or more sunburns can double your risk of developing skin cancer[3].

It has been found that Amelanotic and hypomelanotic melanomas tend to exhibit a common denominator of atypical linear vessels varying in size and shape, usually associated with a central pink to white veil. Amelanotic and hypomelanotic melanomas often lead to delayed or inaccurate diagnosis due to the fact that its clinical appearance can closely resemble other skin conditions. These atypical linear vessels are also called telangiectasia. If telangiectasia can be accurately identified, the accuracy of the diagnosis of these types of melanoma could be greatly improved and action could be taken at an early stage to reduce the chance of the cancer spreading.

The goal of this project was to develop an algorithm to identify and mask these irregular and liner blood vessels, or telangiectasia, in dermoscopy images. Given the ability to automatically detect telangiectasia, the amount of time it takes to diagnose a lesion could be greatly reduced.

Body

This problem of automatically detecting telangiectasia in a dermoscopy image was approached by first studying what telangiectasia are and how they are different from their surroundings. Telangiectasia tend to be dark red, narrow, and long. It is important to realize that color images on a computer in the RGB color model are represent by pixels (short for picture elements), these pixels represent a single point in a graphic image. Each pixel in the RGB color model is a three dimensional point composed of a red, green, and blue value, hence the name RGB (Red Green Blue). When RGB values are written in 24 bpp (bits per pixel), also called Truecolor, color values are represented by integers from 0 to 255, each representing red, green, and blue intensities respectively. For example white would be represented by (255,255,255), black by (0,0,0), red by (255,0,0), green by (0,255,0), and blue by (0,0,255). With this method or representing color images digitally by representing red, green and blue values as integers between 0 and 255, 16,777,216 different colors are possible.

When studying the color values of telangiectasia images in an image editing application it was observed that as you travel outwards from the telangiectasia the red values tended not to change significantly, while the green value increased significantly, the blue value also increased, although not quite as noticeably as the green value. Figure 1 clearly illustrates this drop in color in one specific image (Fgl044). Although it should be noted that not all images exhibit such a distinct change in color, nor do all images have the same surrounding skin color as this specific image.

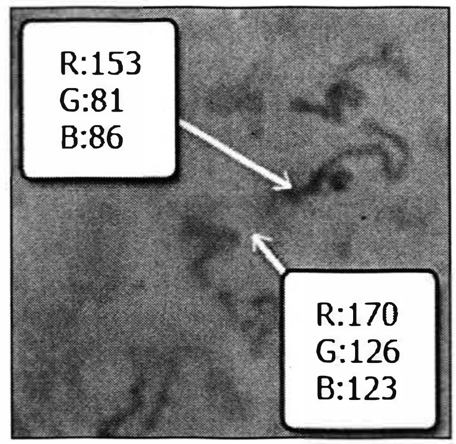

Figure 1: Color drops

These changes in color were taken into account and a heuristic algorithm that essentially detects color changes in the pixels surrounding a given center pixel was developed. For this specific application all input images are accompanied by a lesion mask, which is a TIFF (Tagged Image File Format) image that is 0 for pixels not in the lesion, 1 for the border of the lesion, and 2 for the interior of the lesion. This algorithm iterated through every single pixel inside the lesion, selecting a new center pixel with each iteration, it then moved outward a set number of pixels (`NumPix`) from a given center pixel in 8 directions (see Figure 2). It would start with the pixels with the number 1 (see Figure 2), then the second iteration would move on to the pixels with a number 2 shown (see Figure 2), and so on. Once a center pixel is found that matches the required color drops in two of the eight directions that are at least 135 degrees apart (i.e. North and Southwest directions are 135 degrees apart), then the center pixel is masked. When a pixel is masked, it is placed in an array that is the same size as the input image and recorded as a Boolean value. In the process of creating the initial mask (this initial mask created will be referred to as the OutputMask from this point on), a give center pixel must also not have a higher green value than `MaxG` and no less than `MinG`, there is also a minimum blue value required for a given center pixel (`MinB`).

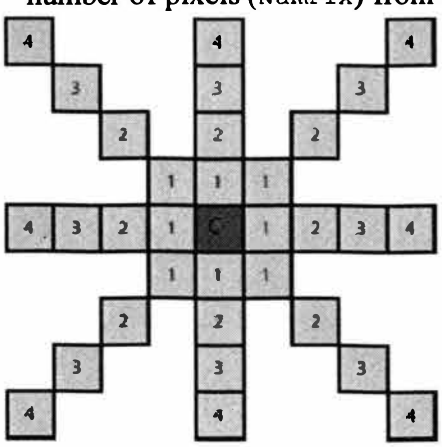

Figure 2: The pixel search directions

Values required for the changes in red, green, and blue were all represented by variables at the beginning of the Matlab script in order to allow easy modification. By observing the color changes of the blood vessels it was determined, with the help of Dr. Stoecker who is a local dermatologist, that red should not increase by more than `RedChng` while green and blue were required to increase by `GreenChng`, and `BlueChng` respectively.

The result of this method of searching for color changes provided an initial mask that was rather noisy and not particularly accurate, and obviously extensive post processing was required. The first post processing step take was to skeletonize the OutputMask and remove objects that were not long enough to be considered irregular linear vessels. The process of skeletonizing a binary image essentially reduces all objects to a single line (hence the name skeletonizing) by removing pixels on the edge of an object without breaking the object apart. A copy of the OutputMask was created and once this image was skeletonized all objects less than the value `MinObjectSize` removed from the skeletonized mask. Objects were removed from the skeletonized mask by connected component labeling the skeletonized image, and removing the object with areas less than `MinObjectSize`. Once the objects that were too small in the skeletonized image were removed, the objects in the OutputMask that correspond to those skeletonized objects are then also removed. After this is completed, a morphological closing operation is performed on the OutputMask, which essentially fills in all the holes in the OutputMask objects.

After the skeletonizing is complete, if the user of the code has specified, various other post processing operations can be performed, such as removing objects that are less than a certain area. This operation is fairly similar to the skeletonizing procedure, but relies on the absolute area of an object as opposed to the skeletonized length of an object. The ability to perform two-dimensional median filtering on the OutputMask is also given. Median filtering is a nonlinear operation that reduces noise in the image, and tends to smooth the edges of the detected objects.

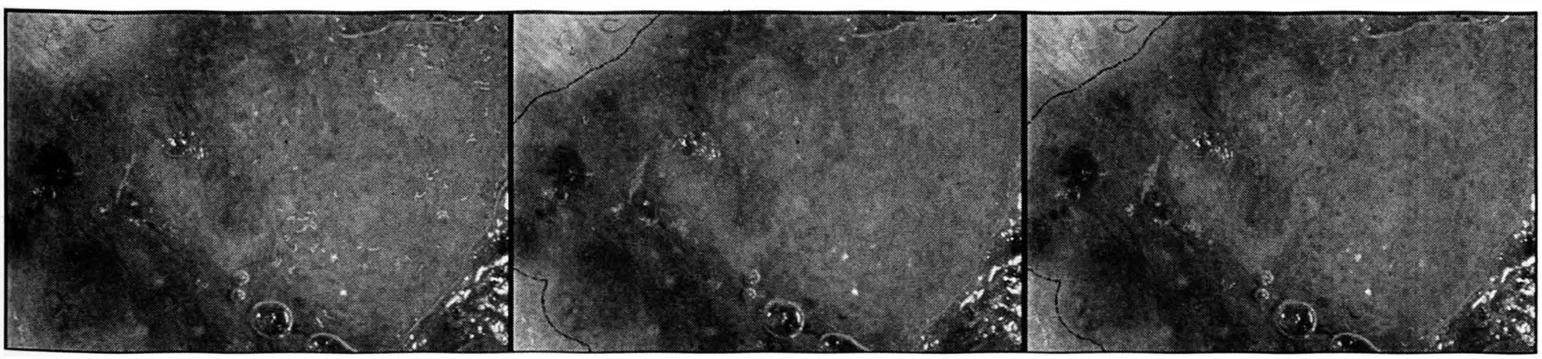

Figure 3: Manually masked vessels highlighted (left), automatically masked before skeletonized filtering (center), automatically masked after skeletonizing and minimum area filtering (right)

In order to achieve the best results from the color change algorithm, a heuristic approach was taken. Initial values for the required changes in color were obtained by observing the color changes using an image editing program. Once these initial values were obtained, the thresholds were tuned by running the algorithm on different images with the same values. This was a very time consuming task since the algorithm can take upwards of 5 minutes per image. Although after much experimentation acceptable results were obtained for the result of the OutputMask.

After all of these post processing features that were added, there were still too many objects found that were not real telangiectasia, these false positives were usually other dark regions of the lesion. To aid in the removal of false positives it was decided that the average (arithmetic mean) color values of the surrounding skin may be a good indicator of whether or not a certain object may be telangiectasia or not. The area of surrounding skin is determined by dilating the OutputMask by `SrndDisk` then subtracting a copy OutputMask that has been dilated by 1. This method of dilating the objects and the subtracting a smaller dilation of the same objects, provides an outline of the surrounding skin around a candidate telangiectasia. This mask is then used to obtain the average color values from the input image for each object's surrounding skin. If the ratio of the red to green and green to blue averages of a certain objects surrounding skin values are not within a certain range then that object is removed the OutputMask. The expected range that the ratios or red to green averages was determined by observing the colors in the skin surrounding various vessels, these values were tested on various images containing telangiectasia. Although it turned out that these values could easily be selected for a certain image, but when moving to another image, these values were essentially useless due the high variability of the lesions as far as the color of the lesion itself, and the color of the surrounding skin.

Due to the complex nature of this problem of finding telangiectasia in highly variable lesions, it was decided that a classifier should be used to aid in the post processing. In order to use most classifiers, a ground truth must be established. A total of 18 images were selected for the dataset, all containing some form of linear blood vessels, a mask of these vessels was created by manually 'painting' over the vessels in an image editing program. These manually masked vessels were reviewed by a dermatologist (Dr. Stoecker) to ensure that an accurate mask of the telangiectasia in the images was obtained.

The first attempt at better filtering the data points was by using a clustering method. A vector of all of the pixels in an image that were masked by the manual mask was created. Then another vector of all of the pixels from the OutputMask was generated. These data points were then clustered using the K-means algorithm. K-means is a clustering algorithm that clusters objects (pixels from the lesion in this case) into k partitions through an iterative process and obtains the centroid of each set. Each pixel can be represented in three-dimensional space by using the red, green, and blue values as separate axes. The intent was to cluster the manually masked pixels from the input image, and then cluster the OutputMask pixels and if the OutputMask pixels were not within

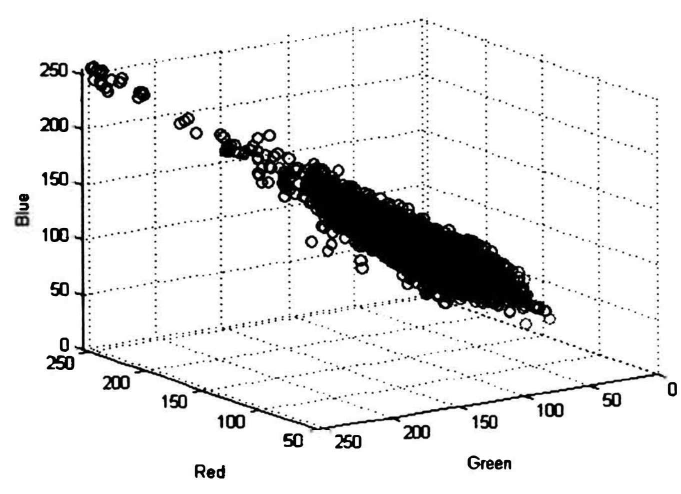

Figure 4: The pixel color data

a certain Euclidian distance from the cluster of manually masked pixels, then they were to be removed from the mask.

Although once the data points for the manually masked and automatically masked pixels were graphed in three-dimensions it was clear that there was not a clear distinction between the two types of pixels. Figure 4 shows the pixel color data graphed, the black circles are the pixel values from the OutputMask, while the red circles represent the color data from the manually masked pixels. Due to the extensive amount of overlap (which is not all visible from Figure 3 alone), this method of clustering the color values of the pixels did not provide a very accurate method of classifying the pixels into telangiectasia and non-telangiectasia points.

Another approach attempted was to use previous code already in the possession of the research group to obtain salient points from the image, and attempt to use that data in the algorithm. A salient point is a point at which two noncrossing branches of a curve meet with different tangents[5]. Salient points would normally be found on the edges of an object, where there is a significant change in color or direction. If a candidate pixel was not within a certain Euclidian distance from any other salient points then it would be removed from the mask. Although this in theory will work, there was a rather high number of points found and in all of the same areas as the OutputMask, therefore eliminating any possibility of using the salient points to identify regions as non-telangiectasia.

After discussing the issues with Dr. Stanley and Dr. Stoecker, it was determined that a more advanced classifier, such as an artificial neural network (usually referred to simply as a neural network), may be a better fit for this type of complex problem. A neural network attempts to model the way that biological neurons in the brain function. It consists of interconnected processing elements called nodes or neurons that work together to produce an output function, the processing of data in neural networks is performed in parallel. Neural networks have proven very useful in finding complex patterns in data and modeling complex relationships between inputs and outputs.

The first idea was to gather the data from the images on a pixel by pixel basis. An Excel spreadsheet was created with all of the color data from the pixels in the OutputMask, and the color data of 8 more pixels (four pixels out from the center, see Figure 2) in the 8 directions discussed earlier. Once this dataset was obtained, it was clear there was a huge amount of data, probably too much to process in a meaningful way. There were upwards of 42,000 data points for some images, processing such a large amount of data would be quite a difficult task, and might not provide meaningful results. It turned out that a pixel by pixel approach may not be the best method for a classification problem of this type.

After discussing it with Dr. Stanley and Dr. Stoecker, it was determined that an object-based approach may work better. In an object based approach, each object found in the OutputMask would be considered a candidate telangiectasia. Various features would then be calculated for these objects, then these feature vectors would be used as the input for a classifier to classify whether or not a certain object was telangiectasia or not. The ground truth would be obtained from the manually drawn masks of the vessels as mentioned previously. This approach would provide a much smaller and possibly more meaningful dataset.

There are 34 values included in the dataset for the object based results. The attempt was made to add as many meaningful values to the dataset as possible, in the hopes that the classifier will determine which values are actually meaningful or needed in deciding whether an object is telangiectasia or not. Some of the fields included in the dataset are: entropy (a scalar value representing the entropy of intensity image), the maximum RGB values of the object, the mean RGB values of the object, the mean RGB values of the surrounding skin, the mean red minus blue values (and several other combinations of colors), eccentricity (the eccentricity of the ellipse that has the same second-moments as the region), the perimeter of the object, the area of the object, and the perimeter squared divided by the area. There were two separate Excel spreadsheets created per image, one with the data from the manually masked vessels, the other with the data from the automatically detected vessels. The spreadsheet with the automatically detected vessels also contained whether or not an automatically found object overlaps one of the manually masked vessels, if it does, the percentage of overlap is provided. The percent overlap was used as the desired output, in other words, the percentage overlap was what a classifier was expected to determine based on the given values mentioned previously.

The neural network software package chosen to attempt to classify the data was Synapse by Peltarion (www.peltarion.com). Synapse is a powerful commercial neural network and adaptive systems software package. It allowed easy input of the data and had a preconfigured neural network designed for classification problems, it also allows these adaptive systems to be easily deployed using the Microsoft .NET framework. Within Synapse a neural network with three function layers and three weight layers was created and trained for 800 epochs with a batch length of 1000. A graph was created showing the sensitivity of the output to the values of various variables (see Figure 5).

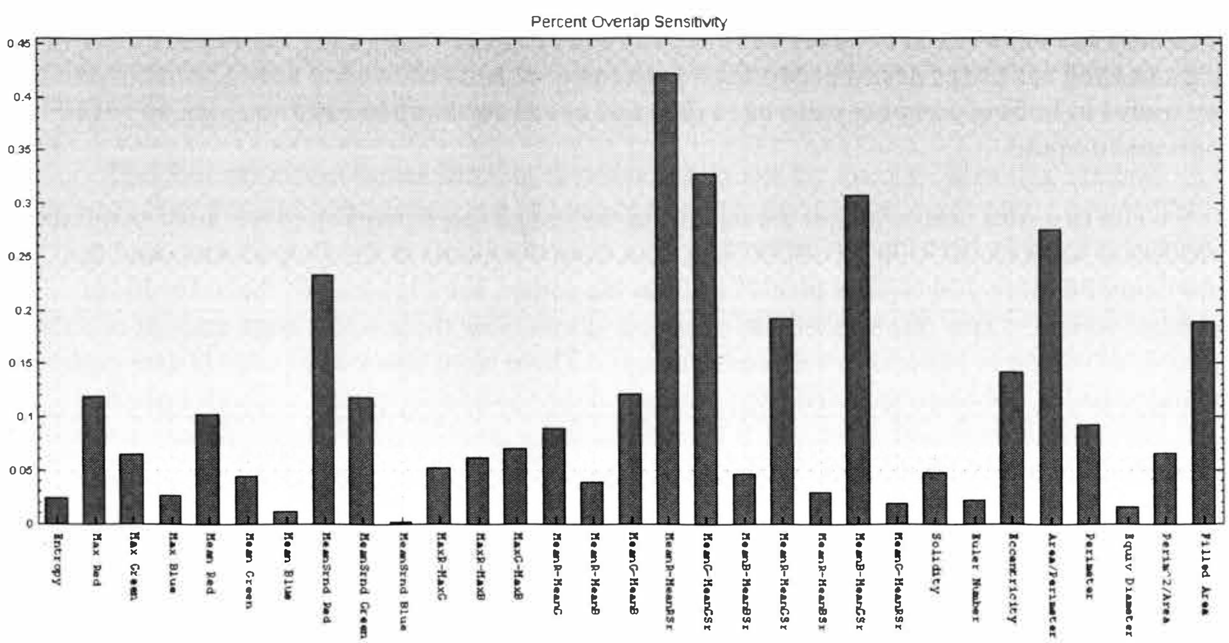

Figure 5: The sensitivity of the output to the input variables.

It turns out that the five most sensitive variables are: (mean red) – (mean read surrounding), (mean green) – (mean green surrounding), (mean blue) – (mean green surrounding), mean red surrounding, and the area divided by the perimeter. The importance of these values validates what was previously thought important in identifying telangiectasia. Although there was not enough time to obtain data about how accurately this neural network can identify telangiectasia, the results look promising. It also showed which variables were not meaningful in determining whether or not an object is telangiectasia.

Although these results are hardly conclusive they validated what was hypothesized, that the changes in red and green values from the vessel to the surrounding skin are a significant factor in identifying telangiectasia, as well as the shape and other morphological properties of the objects (vessels). It also showed that neural networks and the color change detection algorithm developed could be used to identify telangiectasia in dermoscopy images of skin lesions. Although it was not done in this dataset, ideally all of the color values input into the neural network should be normalized by dividing all color values by the corresponding relative color value of the lesion (i.e. mean read / relative red). This could significantly improve the accuracy of the neural network since it would eliminate a high degree if differentiability between the images.

Overall this method of detecting telangiectasia looks promising, although there is still quite a bit of work left before it works well enough to be used in a commercial product. With more time to test and verify the accuracy of using a neural network to detect telangiectasia, while ensuring the same images are not used for training and validation, a neural network used on an object based dataset appears to have the potential to provide the level of accuracy needed.

Nomenclature

Telangiectasia –irregular linear blood vessels found near the surface of the skin

OutputMask – refers to the mask generated by the color change detection method described

Neural network – a simplified software model of biological neurons that work together to produce an output function

Acknowledgements

Special thanks to:

Dr. Stanley, my research advisor, for his help with getting me started with image processing, and for giving me ideas about how to approach the problem.

Dr. Stoecker, for his help with defining the problem and verifying whether or not I was finding the right things.

And everybody else in the DERMVIS group who contributed to this project in any way.

References

[1] American Cancer Society's 2007 Cancer Facts & Figures. (www.cancer.org)

[2] Johnson, TM et al. "Multiple primary melanomas." Journal of the American Academy of Dermatology. 1998; 39(3):422-7

[3] Pfahlberg A, Kolmel KF, Gefeller O. Adult vs childhood susceptibility to melanoma: Is there a difference? Arch Dermatol. Sep 2002; 138: 1234 – 1235

[4] J Dtsch Dermatol Ges. 2003 May;1(5):369-73.Click here to read Links Amelanotic/Hypomelanotic melanoma--is dermatoscopy useful for diagnosis?

[5] Weisstein, Eric W. "Salient Point." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/SalientPoint.html