Retina Lesion and Microaneurysm Segmentation using Morphological Reconstruction Methods with Ground-Truth Data

Thomas P. Karnowski, *Member, IEEE*, V. Priya Govindasamy, *Member, IEEE*, Kenneth W. Tobin, Jr., *Senior Member, IEEE*, Edward Chaum, M.D., *Member, IEEE*

***Abstract*—In this work we report on a method for lesion segmentation based on the morphological reconstruction methods of Sbeh et. al. We adapt the method to include segmentation of dark lesions with a given vasculature segmentation. The segmentation is performed at a variety of scales determined using ground-truth data. Since the method tends to over-segment imagery, ground-truth data was used to create post-processing filters to separate nuisance blobs from true lesions. The segmentation results are used to characterize images as "normal" or "abnormal" and results are shown on a data set of 86 images.**

I. Introduction

The World Health Organization estimates that 135 million people have diabetes mellitus worldwide and that this number will increase to 300 million by the year 2025 [1]. More than 18 million Americans currently have diabetes and the number of adults with the disease is projected to more than double by the year 2050 [2]. Visual disability and blindness have a profound socioeconomic impact upon the diabetic population and diabetic retinopathy (DR) is the leading cause of new blindness in working-age adults in the industrialized world [3]. Thus, there is a significant need to develop inexpensive, broad-based screening programs for DR.

An excellent overview of computer aided diagnosis of retina images is provided in [4], which summarizes the image segmentation as a two-step process: identifying the expected physiological features of the retina, and identifying the pathology of the retina, including micro-aneurysm segmentation, and exudates and hemorrhage segmentation. In this paper we focus on the detection of the presence and characteristics of lesions, particularly in the region near the center of the retina (the fovea). Our objective is to develop algorithms to automatically characterize the retina, particularly its lesions and the overall population of lesions, for use in content-based image retrieval applications. We begin by discussing our algorithm approach which is based on the morphological reconstruction method of [5] and discussing our modifications and feature measurements for post-processing including the use of ground-truth data. In the results section we discuss the data sets tested and the performance of the segmentation algorithms and characterization measurements. We conclude with some observations about the performance and expected future development.

Manuscript received April 7, 2008. This work was supported in part by the National Eye Institute of the National Institutes of Health (RO1-EY017065). This paper was prepared by the OAK RIDGE NATIONAL LABORATORY, Oak Ridge, Tennessee, 37831-6285, operated by UT-BATTELLE, LLC for the U.S. DEPARTMENT OF ENERGY under contract DE-AC05-00OR22725.

T.P. Karnowski, V.P. Govindasamy and K.W. Tobin, Jr are with the Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (e-mail: karnowskitp@ ornl.gov).

E. Chaum is the Plough Foundation Professor of Retinal Diseases with the University of Tennessee Health Science Center, 930 Madison Avenue, Suite 731, Memphis TN 38163 (e-mail: echaum@utmem.edu)

II. Approach

A. Morphological Reconstruction Overview

In [5] the morphological reconstruction method described was used to segment drusen. The algorithm worked particularly well for this type of lesion due to its high adaptability to local contrast changes. We were interested in applying this algorithm to a variety of lesion types including microaneurysms, hemorrhages, and exudates. We summarize the approach taken by [5] as follows. First an image is pre-processed by determining the region of interest, usually the fovea area (the center of the eye, where most vision receptors are located) and cropping this section. Noise filtering is performed by median filtering and smoothing filtering. The regional maxima of order λ and regional minima of order γ are determined (they generally assume $\lambda \geq \gamma$). These are used to create the *relative support* of each regional maximum, essentially a neighborhood surrounding each regional maximum up to nearby regional minima. A synthetic image h(x) is created where the relative support is "filled in" by the *relative dynamics*, in essence the difference between the intensity of the regional maximum and the intensity of the closest regional minimum. The regional maxima of h(x) of order β are found next, and are taken to be the candidate contours of the drusen.

B. Algorithm Modifications

For our implementation we used a vascular segmentation based on [7] to mask out regional minima in the bright segmentation step that happened to fall within the vasculature. The idea behind this move was to allow better segmentation of lesions that may have vessels running through them.

We adapted the method to segment dark lesions by inverting the image and applying the algorithm as described. In the dark segmentation step, after inversion we remove the regional maxima that fall within the vasculature. However, a post-processing step is introduced where blobs are removed that are close to the vasculature by masking with a dilated version of the vasculature map. This step was introduced because a large number of small dark blobs were created around the vasculature due to missed pixels in the vasculature segmentation.

C. Parameter selection

Some efforts to automate the parameter selection were desired. We selected a variety of settings for (λ,β,γ) which we refer to as different scales. These scale settings were compared to our ground-truthed image set with the hope of finding optimal scales for different lesion types. The mutual area overlap was chosen as the metric for good segmentation, then for each lesion type, the median mutual overlap was computed at each scale. Note that cases where the lesion was missed or oversegmented were not included in the calculation; only those machine-segmented lesions that intersected at least one pixel of a ground-truth lesion were used.

D. Post-processing

Processing is performed after the segmentation step to remove nuisance blobs. In [5] some simple rule-based methods were used based on features of the blobs. In our case, we attempted to use ground-truth data to design classifiers to separate lesions into nuisance and actual lesion classes. However, ground-truthing is a laborious manual task and inevitably did not always correlate well with machine segmentations. For example, confluent drusen that were close together were often grouped together as a single lesion during ground-truthing, but machine segmentation typically separated them into smaller, yet still distinct and well-defined groups (see Figure 1 as an example). As another example, it was not uncommon for some lesions to be inadvertently omitted during the ground-truth process. This led to issues where the post-processing classifier could become confused because, for example, a machine-segmented microaneurysm that was not identified as a ground-truth lesion would present the classifier training set with an ambiguous example. To mitigate these issues, two steps were taken. First, only the images that were classified as “normal” with no lesions were used for examples of nuisance blobs. Second, only machine segmented lesions that had significant mutual area overlap with ground-truth lesions were chosen as examples of actual lesions. The machine segmented lesions that closely matched the ground-truth lesions were selected by hand as good “positive” examples of actual lesions by reviewing the machine segmentation results on abnormal images. Finally, note that it is possible that even with these steps, ambiguity could be introduced in the ground truthing such as failing to label lesions on an image classified as “normal”.

To characterize each lesion, a set of 43 gray-scale features were measured on each machine-segmented lesion. The extension to color is straightforward but our initial data set was gray-scale only so we restricted our development to intensity levels. The data was projected using principle component analysis to retain 90% of the information, then a simple Bayesian classifier using a Gaussian probability density function was created to distinguish between the two classes of “nuisance blob” (NB) and “actual lesion” (AL), with different classifiers trained for the dark and bright blobs.

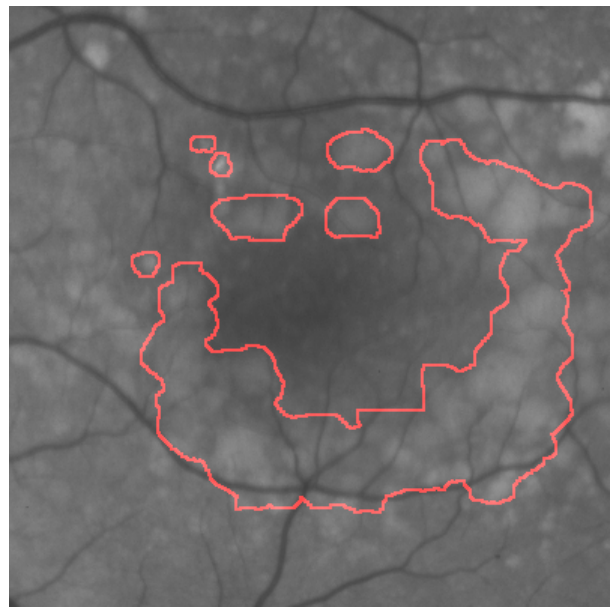

Fig. 1. Example of ground truth image with discrete drusen grouped together. The machine segmentation algorithm tends to break these up into discrete lesions.

III. EXPERIMENTS

The data set for this work was a subset of those images used in [6, 8] and was composed of 86 retinal images. These images were ground-truthed with lesions hand-drawn in the macular region. The entire set of 370 images were ground-truthed by one of us (E.C.), but for this work we restricted the set to those 86 that were either normal with no lesions, or abnormal and contained only lesion types of interest (soft and hard drusen, exudates, cotton wool spots, hemorrhages, flame hemorrhages, and microaneurysms). The images were captured at a resolution of 12 microns per pixel. Note that this image set represents an actual population from an ophthalmology practice, so we should point out that these represent individuals who have sought medical attention and are likely more advanced with respect to their DR and AMD variability than real broad-based screening data may encompass. Of the 86, 27 were normal, 35 had AMD in various stages, and the remaining 24 were NPDR or PDR in various stages.

The initial parameter selection process was performed on the set of 86. The chosen parameters were then used for the remaining experiments. However, the actual tests of post-processing were performed by using hold-one-out analysis whereby a single image is removed from the data set, the classifiers are retrained on the remaining data, then the held-out image is tested. All experiments were performed using software created in the MATLAB programming environment on a standard desktop PC.

A. Parameter Selection using Mutual Overlap

The parameters of (λ,β,γ) used shown in Table I below. To save processing time, the larger scales were created by resizing the image itself so that an additional parameter, s for image resize factor is included as well. This effectively increases the size of the reconstruction kernels without requiring a corresponding increase in the processing time.

TABLE I
PARAMETERS FOR MORPHOLOGICAL RECONSTRUCTION AND SCALES

Scale	λ,β,γ	Size
1	29	0.25
2	15	0.25
3	5	0.25
4	1	0.25
5	1	0.5
6	1	1

The results of the mutual area overlap on the 86 data set are shown in Figures 2 for the bright lesions and dark lesions. The bright scale 5 is effective at capturing the content of the bright lesions, although hard drusen seem to work better at scale 4. Bright scale 6 seems to do the best job of segmenting microaneurysms, but we also see that scale 5 is effective as well at hemorrhages. Flame hemorrhages worked best at scale 3, but they were regarded as somewhat minor to the other types so we focused on scale 5 bright, and scale 5 and 6 for the dark lesions. We elected to create a single machine-segmented image by using a combination of parameters: scale 5 bright blobs, the scale 5 dark blobs, and then scale 6 dark blobs that do not overlap with scale 5 dark blobs. This combination was arrived through a somewhat heuristic process based on the overlap between ground-truth and machine-segmented lesions.

B. Post Processing Performance

The post-processing performance in terms of removing nuisance blobs had a maximum sensitivity/specificity of 90% / 90% for both the bright and dark lesions. This was measured by performing the hold-one-out analysis on an image-by-image basis. The PCA projection reduced the feature dimensionality from 43 to 6 and 7 for bright and dark blob classifiers respectively. Receiver operating characteristics curves are shown in Figure 3 for the bright and dark blob cases. Some example images are shown with the ground truth, machine segmented, and post-processed results from left to right.

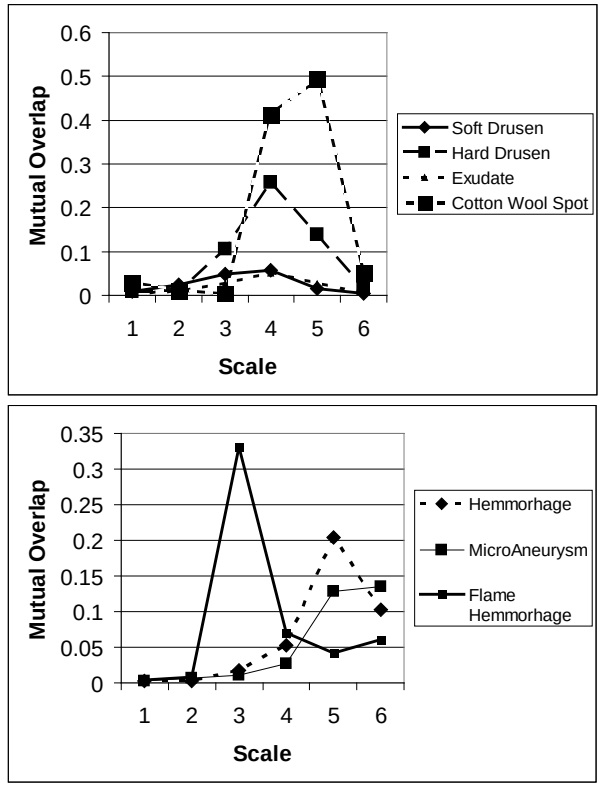

Fig. 2. Mutual overlap with different scales. Top: Bright lesions Bottom: Dark lesions (including microaneurysms)

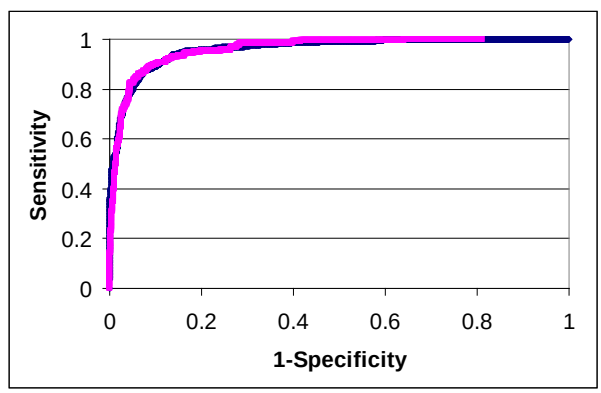

Fig. 3. ROC curve for classification of nuisance blob / automatic lesion for bright (blue) and dark (magenta) blob types. Both curves showed best performance at 90% sensitivity and 90% specificity (or 10% false alarm).

C. Classification of Images into Abnormal/Normal

We were interested in how the performance of the lesion segmentation post-processing would affect the overall classification of images into normal / abnormal classes (where AMD and DR are combined into the abnormal state). The image retrieval context of [8] was used in this experiment with a value of σ=0 meaning no data is rejected due to ambiguity between classifications. Each image was characterized using a "lesion population" feature vector comprised of 34 values which were essentially histograms of

the lesions characteristics such as area, intensity, contrast. In addition, 6 features were measured as well on non-lesion background. Clearly, if the post-processing leaves too many nuisance blobs, normal example images will appear to have many lesions but the hope is that their overall population as measured by the lesion population features will allow some discrimination between normal and abnormal.

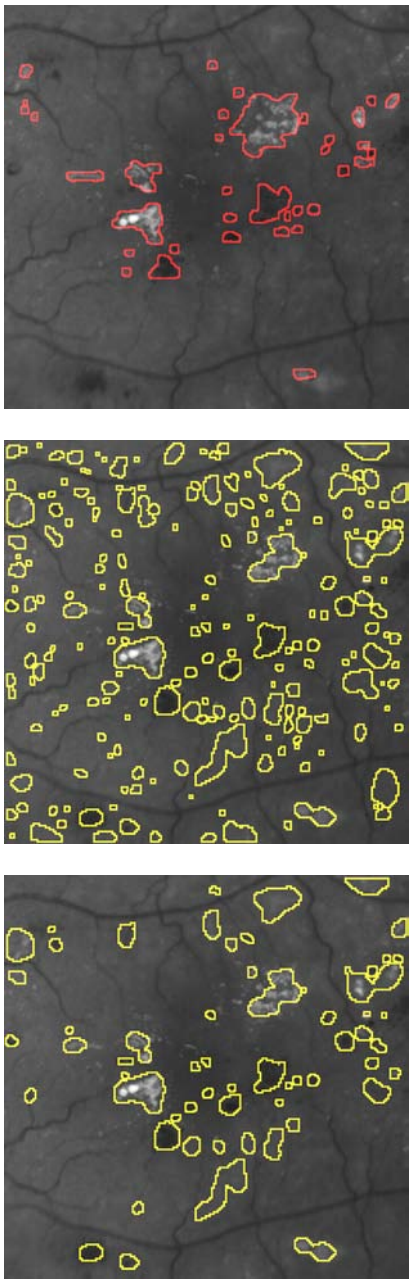

Fig. 4. Example of ground truth, machine segmented, and postprocessed image. (Top) Ground truth lesions (Middle) Segmentation without post-processing. Some blobs are clearly oversegmented but some may be legitimate lesions (Bottom) Postprocessed result. Many oversegmented blobs are gone but some false positives remain, such as the elongated blob at the middle / bottom of the image.

Eight different combinations of NB/AL sensitivity / specificity were tested. The best achieved sensitivity / specificity in the two-class problem of normal and abnormal was 88% / 67%. When the features were simply the number of bright and dark lesions, the best performance was slightly better, at 88% / 70%. Both results suggest that some level of robustness to segmentation inaccuracies can be attained. We are optimistic this performance will be improved with more realistic data sets (such as those from a more general population).

IV. CONCLUSIONS

In this work we used the morphological reconstruction method of [5] to perform segmentation of a variety of bright and dark lesion types on retina images. We used ground-truth data to design post-processing classifiers that separated the machine-segmented results into nuisance and actual lesion classes. Our results were encouraging with respect to removing nuisance blobs, but additional efforts will be needed to determine how effective this approach will be overall in predicting retina health in our content-based image retrieval context. Future work includes extension to color imagery, better classification features for the post-processing steps and application to a larger, more realistic screening data set [10,11].

REFERENCES

[1] A.F. Amos, D.J. McCarty, P. Zimmet, "The rising global burden of diabetes and its complications: estimates and projections to the year 2010," Diabetic Med 1997;14: S57-85.

[2] Centers for Disease Control and Prevention (2003) National Diabetes Fact Sheet. (http://www.cdc.gov).

[3] C. Stellingwerf, P. Hardus, J. Hooymans, "Two-field photography can identify patients with vision-threatening diabetic retinopathy: a screening approach in the primary care setting," Diabetes Care. 2001, 24:2086-90.

[4] T. Teng, M. Lefley, D. Claremont, "Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy", Medical & Biological Engineering & Computation, v 40, pp. 2-13.

[5] Ben Sbeh Z, Cohen LD, Mimoun G, Coscas G, "A New Approach of Geodesic Reconstruction for Drusen Segmentation in Eye Fundus Images", IEEE Trans Med Imaging. v. 20, pp. 1321-33.

[6] Tobin, K.W., Chaum, E., Govindasamy, V.P., Karnowski, T.P., Sezer, O., "Characterization of the Optic Disk in Retinal Imagery using a Probabilistic Approach", SPIE International Symposium on Medical Imaging, San Diego, California, Proceedings of SPIE, Vol. 6144, February 2006.

[7] F. Zana, J. Klein, "Segmentation of Vessel-Like Patterns Using Mathematical Morphology and Curvature Evaluation," IEEE Trans. On Image Processing, v10, pp. 1010-1019.

[8] Karnowski, T.P., Govindasamy, V.P., Tobin, K.W., Chaum, E., "Locating the Optic Nerve in Retinal Images: Comparing Model-Based and Bayesian Decision Methods", Conf Proc IEEE Eng Med Biol Soc. August 2006

[9] Tobin, K.W., Abdelrahman M, Chaum E, Govindasamy V, Karnowski TP. "A probabilistic framework for content-based diagnosis of retinal disease", Conf Proc IEEE Eng Med Biol Soc. August 2007

[10] Abramoff, M.A. and Suttorp-Schulten, M., "Web-based Screening for Diabetic Retinopathy in a Primary Care Population: The EyeCheck Project", Telemedicine and e-Health, v 11, No. 6, pp. 668-674

[11] Abramoff, M.A., Veirgever, M., Neimeijer, M., Russel, S., Suttorp-Schulten, M., and Van Ginneken, B., "Evaluation of a System for Automatic Detection of Diabetic Retinopathy From Color Fundus Photographs in a Large Population of Patients with Diabetes", Diabetes Care, V. 31, pp 193-198.