31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Practical Considerations for Optic Nerve Location in Telemedicine

T. P. Karnowski, *Member, IEEE,* D. Aykac, *Member, IEEE,* E. Chaum, M.D., Ph.D., *Member, IEEE,* L. Giancardo, *Member, IEEE,* Y. Li, Ph.D., *Member, IEEE,* K.W. Tobin, Jr.,Ph.D., *Senior Member, IEEE,* and M.D. Abramoff, M.D., Ph.D. *Member, IEEE,*

Abstract—**The projected increase in diabetes in the United States and worldwide has created a need for broad-based, inexpensive screening for diabetic retinopathy (DR), an eye disease which can lead to vision impairment. A telemedicine network with retina cameras and automated quality control, physiological feature location, and lesion / anomaly detection is a low-cost way of achieving broad-based screening. In this work we report on the effect of quality estimation on an optic nerve (ON) detection method with a confidence metric. We report on an improvement of the method using a data set from an ophthalmologist practice then show the results of the method as a function of image quality on a set of images from an on-line telemedicine network collected in Spring 2009 and another broad-based screening program. We show that the fusion method, combined with quality estimation processing, can improve detection performance and also provide a method for utilizing a physician-in-the-loop for images that may exceed the capabilities of automated processing.**

I. Introduction

A BROAD-BASED inexpensive screening technique for diabetic retinopathy and other diseases of the eye will be a vital component of the battle against blinding eye disease with the projected growth in diabetes in the next few decades. In the United States alone, more than 18 million Americans have diabetes, and the number of adults with the disease is projected to more than double by the year 2050 [1]. We are developing and deploying a telemedicine network [2] in the Mid-South region of the United States to provide retina screening to patients in walk-in clinics. All images taken are reviewed by an ophthalmologist, but our eventual goal is automatic review of the images. Even with full automation, however, the use of a physician-in-the-loop can be used to improve the performance of the system. A conceptual diagram of the system is shown in Figure 1. After image capture a quality assessment [7] is performed. If the image quality is sufficient, the system then locates anatomic structures, performs lesion detection and population description, and finally performs automatic diagnosis using an image database as shown by the lower right pathway [9,10]. Even after full database population and classification training, a level of physician oversight will be required. Thus, by estimating (and to some degree controlling) the quality of the image, the performance of the overall system should be improved because images that are more difficult for automatic diagnosis will be referred to the physician-in-the-loop. In such a practical system we would control attempts at automated processing by ensuring that some minimum quality level is achieved, then using the confidence measure of ON location to control the variability of images submitted for automated processing. Thus, in this paper we show how the quality assessment affects the estimate of the ON location using method that features a confidence measure.

The paper is organized as follows. In the Approach section we present an overview of the quality assessment method of [7] and the complementary optic detection methods used in [16]. We discuss the improvements to the complementary detection methods made for this paper. In

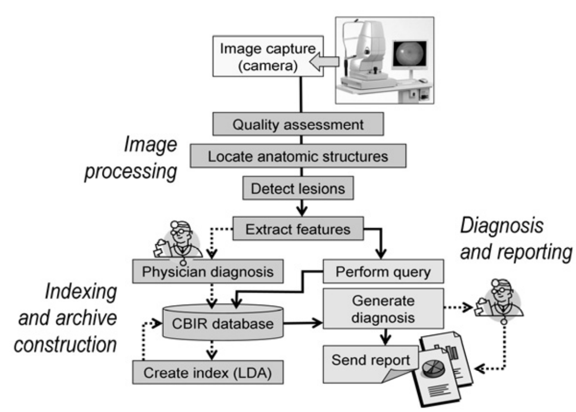


Fig. 1. Overview of system for retina screening. After image acquisition, a quality check is conducted followed by detection of physiological structures and the building of a database of labeled examples. After the database is sufficiently mature the disease states can be estimated automatically, but even then a level of physician review will be required.

Manuscript received April 20, 2009. This work was supported in part by the National Eye Institute of the National Institutes of Health (R01 EY017065 and R01 EY017066), the Research to Prevent Blindness, New York, NY, the Plough Foundation, Memphis TN, and the Delta Health Alliance, Stoneville MS. This paper was prepared by OAK RIDGE NATIONAL LABORATORY, Oak Ridge, TN, USA, 37831-6285, operated by UT-BATTELLE, LLC for the US DEPARTMENT OF ENERGY under contract DE-AC05-00OR22725.

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

E. Chaum and Y. Li are with the University of Tennessee Health Science Center, 930 Madison Avenue, Suite 731, Memphis TN 38163 where E. Chaum is the Plough Foundation Professor of Retinal Diseases.

M.D. Abramoff is with the Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, Iowa City, IA 52242 USA and also with the Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 55242 USA

the Experiments section, we show the results of a 2-fold validation test, then we test the classifier on a database from our telemedicine network. We show how the performance of the method changes with the image quality and conclude with observations and projections for future research.

II. APPROACH

A. Quality Assessment

The quality assessment method we use [7] is well-suited for a telemedicine network due to its speed and allows quick feedback to the camera operator. The method is summarized here. A segmentation of the vascular tree is conducted using the method of [3] and the image is divided into 8 spatial regions. The vessel density is measured in each region and an additional set of 2 features based on the saturation of the fundus are created. The training images are manually labeled as good or poor quality, then a classifier is trained to distinguish them. After creating the library, an additional set of images was classified and a threshold of 0.4 was set by a reviewing ophthalmologist (E. Chaum). Note that this initial threshold is set to provide images of sufficient quality to be diagnosed by a human; generally, this quality level may not be sufficient for automated diagnosis. In addition, the telemedicine network allows the operator to invoke a "best we can do" setting so that if the image quality cannot be improved the image can still be submitted. Finally, in addition to the speed of the method, another advantage is the production of the vasculature tree segmentation which can then be used in the optic nerve detection method.

B. Optic Nerve Detection Methods

Optic nerve detection is a mature field and there are many different methods that have been applied to a wide variety of data sets (see, for example, [11-15]). In [16] we discussed how two complementary methods (one focusing on the vasculature tree, the other on the ON appearance) could be used in tandem to produce a confidence measurement. For more background on these methods, we refer to [4,8,16]. Briefly, we denote these two methods as the feature-based likelihood ratio (FBLR) [8] and the Principle Component Analysis – Linear Discriminant Analysis (PCA-LDA) method [4].

In the FBLR method, the vasculature of the image is segmented and a set of four features are generated at each pixel: the brightness of the pixel region and the thickness, orientation and density of the vasculature. A training set of labeled ON centers is used to estimate the parameters of a Gaussian distribution describing the ON regions and the non-ON regions. We also use the hand-segmented training set to estimate the ON center probability density function (PDF). The Gaussian parameters and the PDF are used to compute a likelihood ratio using maximum a posteriori (MAP) estimation.

The complementary method, the PCA-LDA method, is based on the model-based method of [11] which uses principal component analysis (PCA) to capture the information content of a training set of optic nerve images. "Candidate regions" of a retina image are projected to PCA space and then used to reconstruct the regions. The pixel with the smallest residual error is chosen as the ON location. In [4], we extended the method to include LDA. After generating the PCA coefficients for the training set, the reconstruction distances were calculated for the entire image of all training images. Up to twenty different pixels were chosen to comprise a data set in PCA space, including the ON, the five smallest reconstruction distances (which comprise a set of training examples which are non-ON but score well on the reconstruction distance metric), and up to 14 randomly selected pixels in the candidate region. The vectors corresponding to these twenty pixels were projected back to PCA space. After repeating this process for all images in the training set, LDA was employed to compute a transform to a one-dimensional space. This feature was then used to formulate a MAP likelihood ratio modelling the feature as a Gaussian random process.

C. Confidence Measures in Optic Nerve Location

In [16] we showed empirically, on two different data sets, how limiting the number of automatically screened images by thresholding the distance between the estimates of complementary methods could improve accuracy. In this work, we extend this by training the PCA-LDA approach on specific failed images by the complementary vascular-based method. The data set of interest is divided randomly into a training set and a test set (2-fold validation). During the first of the two tests, the FBLR classifier is trained, then tested against its training set. We record the images where the ON is incorrectly identified (the classifier places the ON location greater than one ON radius from the true location). During the PCA-LDA training, those locations are chosen as an example of a "non-ON" location. We refer to this as "directing" the LDA towards the images that give the FBLR method trouble. This is an intuitive and straightforward modification to the method and essentially comes "for free" in that we simply choose different examples of the non-ON areas to train the LDA classifier; the main goal is to prevent the two methods from picking the same wrong location.

III. EXPERIMENTS

In this section we describe the experimental results using the directed complementary method on two different data sets. We first describe the data sets, then the improvements found from the directed LDA. We then discuss how changing the quality threshold improves performance.

A. Data Sets

Three data sets were used in this study. The first data set was used to build a classifier and study the effect of the improvements to the fusion method. This set, which we will call Chaum370, was composed of 370 retinal images representing 18 different retinal pathologies and normal (non-diseased) retina. These images were originally captured at a resolution of 12 microns per pixel. Note that this image set (supplied by E. Chaum) represents an actual population from an ophthalmology practice and has a large number of abnormal retinas (326 abnormal or 88%).

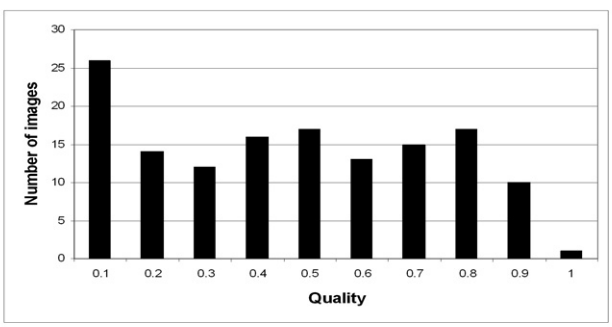
The second data set consisted of 141 images collected from the first clinic in our telemedicine network [2]. These were collected over the first month of operation and consequently include several images of sub-optimal quality. All images were referred to screening by a general practitioner in the clinic. All images were collected with the same camera, a non-mydriatic Zeiss VisuCam Pro, and indeed all clinics in the telemedicine network will use the same camera model. All images were collected with a 45 degree field of view, in color, with a resolution of roughly 6 microns per pixel. For the ON location processing, the green plane of all images was extracted and resized to a scale consistent with the first data set. Finally, images of the left eye were flipped so that the optic nerve was always on the right side of the image. Based on an ophthalmologist review, 97 (or 68%) of these were judged "normal" with a recommended "treatment" of a follow-up in 6 months.

Since we are still in the early stages of our telemedicine network, the final dataset, dubbed Abramoff748, is composed of 748 images from a much larger data set supplied by Dr. M. Abramoff [5,6]. We selected a subset captured at 45 degree field of view, a common resolution of roughly 200 microns per pixel and macula-centered. This set represents a broad-based screening effort and is more representative of the type of set automatic screening would encounter in practice.

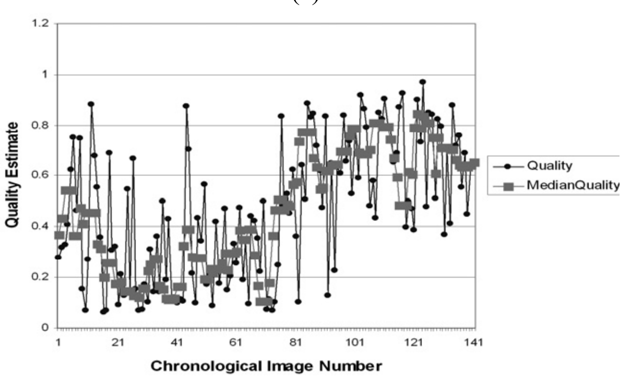
The quality of the first set was not assessed, but generally it is very high as these images were hand picked to form an initial library for our research. This is not the case for the other data sets and it is instructive to examine the quality of the images. In Figure 2a we show a histogram of the image quality from the telemedicine set. We see that there is a large number of relatively poor quality images in our data set. Figure 2b shows the quality value as a function of image number, which is proportional to time (larger image numbers means new images). Some smoothing (a median filter over six adjacent values) has been applied in the second curve and we see that generally the image quality improves as the camera operators in the clinic become more skilled at using the camera. The quality of the Abramoff748 set is shown in Figure 3 and is generally good overall although there are some lower quality images.

B. Improvements to Complementary Methods

To determine how the improved complementary method increases the ability to detect the ON, we performed a two-fold validation study using the Chaum370 data as discussed in the Approach section. During the testing phase, the ON is found with both methods and the distance is measured between them. A threshold is tested in increments of 1/10 the radius of the ON and images where the distance between methods exceeds the threshold are rejected as problematic. The non-problematic images are then evaluated by averaging the ON location estimates and the performance measured in terms of detections within one ON radius. In Figure 4 we show the improvement obtained in the complementary method when the LDA method is directed toward the problematic cases for the FBLR method. In this



(a)



(b)

Fig. 2. Image quality of the telemedicine set. (a) Histogram of image quality (b) Image quality over time. We see that the image quality improves over time as the operators become more skilled. Images scoring less than 0.4 were deemed not acceptable, but the operator could still submit these under a "Best We Can Do" qualifier.

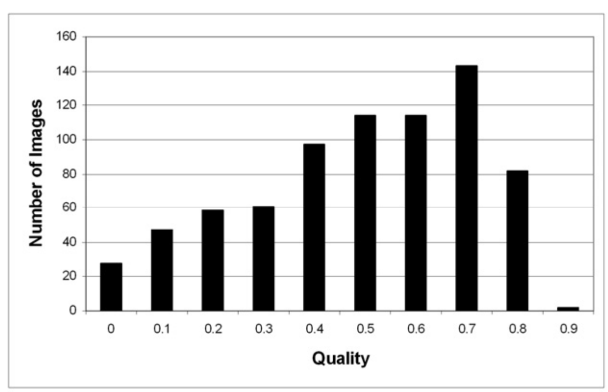


Fig. 3. Histogram of the image quality estimate of the Abramoff748 data set as measured by the quality estimator of [7] trained on an archive of images from the Zeiss VisuCam Pro camera.

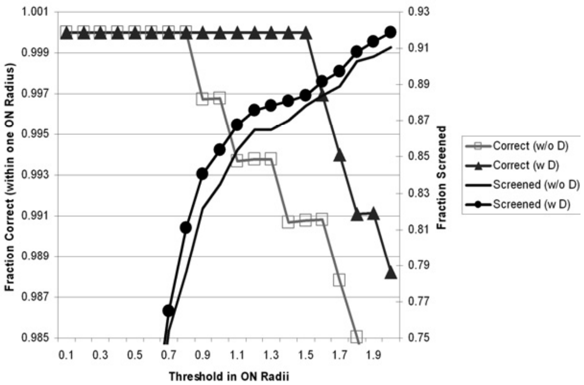


Figure 4. Improvements due to directing the LDA toward erroneous locations found in the FBLR method. We see the fraction screened and correct curves are both higher for the directed method over the undirected method.

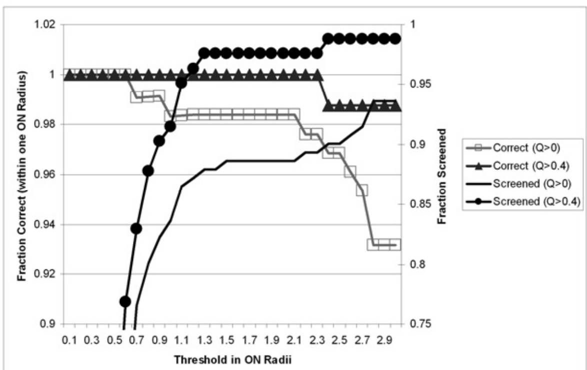


Figure 5. ON location estimates using the trained classifier from the Chaum370 set on the images from the telemedicine network. Two quality levels are assessed: 0 quality (meaning all images are submitted) and 0.4 quality, which is the level chosen manually as the threshold for image submission in the telemedicine network.

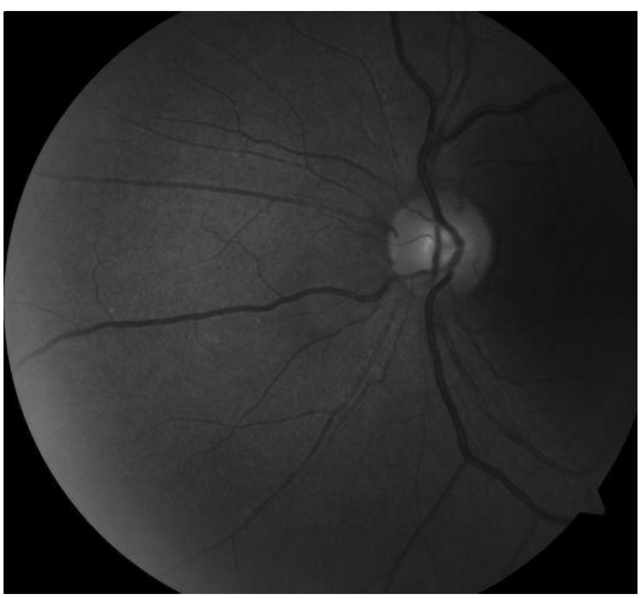


Figure 6. Example of image where ON detections were not consistent, resulting in a rejection. In this case, although the ON is clear, the use of the prior biased the results toward macula-centered images and this image is clearly out of the operating range (the macula is not visible).

figure we compare the results for non-direction with direction and we find that the former results in higher performance and more images successfully screened. In [16] we found empirically that a threshold of 1.5 ON radii gave good results and here we have 100% performance with increased fraction screened at this level. Note that the higher threshold is not significant, since it is the distance between the estimates of the two complementary methods, but the higher performance at the target 1.5 threshold (both in terms of fraction screened and fraction correct) is the key metric of improvement here.

C. Performance and Quality

In the second set of tests, the ON location classifier trained from the previous section was tested on the dataset of images from the telemedicine network and the Abramoff748 set. In Figures 5 and 7 we show the effect of screening for quality on ON detection. Images with quality below the tested threshold were removed from the dataset and then the ON location of the remaining images was estimated. We show results for quality level of 0 (meaning we attempt ON location on all images) and 0.4, which was the level set for acceptance of the image in the telemedicine network. We see in both data sets as the quality level increases, the success of the algorithm improves in that the fraction screened with confidence and successful estimates increase. In our case, setting the level to 0 resulted in 141 images screened, and of those, we rejected 16 as having too low confidence in their ON locations at the 1.5 ON radii threshold level. Of the remaining 125, we correctly estimated the ON location on 123 images. With the quality level set to 0.4, we screened 82 images and rejected 2 of those as having too low ON confidence at the 1.5 level; the positions of all 80 remaining were correctly estimated in that case. The rejected images, in fact, were clearly not macula-centered (one example is shown in Figure 6). We use this image to illustrate the point that while the ON location is clear, the use of the prior is intended to not only facilitate ON location but also to help ensure that the system operates with a suggested operational range, since clinically

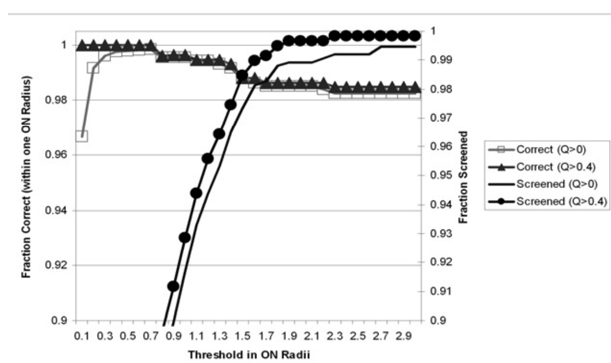


Figure 7. ON Location estimates using the trained classifier from the Chaum370 set on the Abramoff data set.

significant retinopathy requires observation of the macula. The results in the Abramoff748 case were comparable, with 748 images screened with quality threshold of 0 resulting in confident ON detections in 728 cases with 720 correct, and a threshold of 0.4 resulting in 552 screened with confident ON detections in 545 cases with 537 correct. As a final note, if the ON confidence threshold is moved back to 0.7 ON radii, the quality threshold of 0.4 gives 100% correct classifications with 83% and 84% screened in the telemedicine and Abramoff748 data sets.

D. Discussion of Practical Considerations

This work shows that there are three key thresholds of consideration in the estimate of the ON location. First, the quality level must exceed some basic threshold before it is deemed acceptable for submission to the network. As mentioned earlier we do allow for a "best we can do" mode where the operator can submit the image regardless of its quality. Many of the initial submissions fell into this category but later submissions improved and "best we can do" submissions become much less frequent. Second, once an image is received, another, higher quality level threshold may be necessary to deem the image of sufficient quality for automatic screening. Finally, the confidence of the ON detection can be judged by the complementary method which has another third threshold, although it is not a quality level but rather a "degree of agreement". Images which fail that final metric may still be evaluated but would best be passed to the reviewing physician. One can imagine other thresholds of interest; for example, since we assume the images are macula centered, images with estimates of the ON location that stray too far from the ideal or mean position may be flagged for manual review as well.

IV. CONCLUSION

In this work we discussed some practical issues in the implementation of a telemedicine network with a goal of automatic retina screening. We showed how a complementary method for optic nerve detection, which was initially described in [16], can be improved through directed training and through the use of quality estimates, which is important because it provides a practical threshold for manual image review. We would like to recognize that there may well be room for improvement in image screening so that (for example) the state of the retina can be assessed even when the image quality is sub-optimal, but in our case we seek to improve the initial image capture as much as possible and therefore we believe our approach is sound. For future work we will continue to implement these aspects of the processing in our telemedicine network and also incorporate lesion detection and disease stratification estimation, which has been shown [9] to be susceptible to image quality.

REFERENCES

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

[2] Li, Y., Karnowski, T., Tobin, K., Giancardo, L., Chaum, E. (2009) A network infrastructure for automated diagnosis of diabetic retinopathy. American Telemedicine Association Annual Meeting, April 2009.

[3] Zana, F., Klein, J., (2001) Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. IEEE Trans. Image Processing, 10:1010-1019

[4] Karnowski, T.P., Govindasamy, V.P., Tobin, K.W., Chaum, E., (2006) Locating the optic nerve in retinal images: comparing model-based and Bayesian decision methods. 28th Annual International Conf. of the IEEE EMBS New York, NY 2006

[5] Abramoff, M.A. and Suttorp-Schulten, M., (2005) Web-based screening for diabetic retinopathy in a primary care population: the eyecheck project. Telemedicine and e-Health 11:668-674

[6] Abramoff, M.A., Veirgever, M., Neimeijer, M., Russel, S., Suttorp-Schulten, M., and Van Ginneken, B., (2008) Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. Diabetes Care 31:193-198

[7] Giancardo, L., Abramoff, M.D., Chaum, E., Karnowski, T.P., Meriaudeau, F., and Tobin, K. W., (2008) Elliptical local vessel density: a fast and robust quality metric for fundus images. 30th Annual International Conf. of the IEEE EMBS Vancouver, Canada, 2008

[8] Tobin KW, Chaum E, Govindasamy VP, Karnowski TP. (2007) Detection of anatomic structures in human retinal imagery. IEEE Trans Med Imaging 26:1729-39

[9] Tobin, K.W., Abramoff, M.D., Chaum, E., Giancardo, L., Govindasamy, V.P., Karnowski, T.P., Tennant, M.T.S., (2008) Using a patient image archive to diagnose retinopathy. 30th Annual International Conf. of the IEEE EMBS, Vancouver, Canada, August 2008

[10] Chaum, E., Karnowski, T.P., Govindasamy, V.P., Abdelrahamen M., Tobin K.W., (2008) Automated diagnosis of retinopathy by content-based image retrieval. Retina 28:1463-1477

[11] Li, H., Chutatape, O., (2004) Automated feature extraction in color retinal images by a model based approach. IEEE Trans Biomed Eng. 2:246-54

[12] Niemeijer M, Abràmoff MD, van Ginneken B., (2007) Segmentation of the optic disc, macula and vascular arch in fundus photographs. IEEE Trans Med Imaging. 26:116-27

[13] Foracchia, M., Grisan, E. Ruggeri, A., (2004) Detection of the optic disc in retinal images by means of a geometrical model of vessel structure. IEEE Trans. on Medical Imaging 23:1189-1195

[14] Hoover, A., Goldbaum, M., (2003) Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. IEEE Tran. on Medical Imaging 22:951-958

[15] Abdel-Haleim, A., et. al., (2008) Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter. IEEE Trans. on Medical Imaging 27:11-18

[16] Karnowski, T.P., Aykac, D., Chaum, E., Giancardo, L., Li, Y., Tobin, K.W., Abràmoff MD, (2009) Evaluating the accuracy of optic nerve detections in retina imaging by using complementary methods, World Congress on Medical Physics 2009