Characterising glaucoma using texture

Characterising Glaucoma Using Texture

Tim Morris
http://staff.cs.manchester.ac.uk/~tmorris/
Suraya Mohammed

School of Computer Science
The University Of Manchester
Oxford Road
Manchester
M13 9PL

Abstract

In this paper, we present our ongoing work on glaucoma classification using fundus images. The approach makes use of texture analysis based on Binary Robust Independent Elementary Features (BRIEF). This texture measurement is chosen because it can address the illumination issues of the retinal images and has a lower degree of computational complexity than most of the existing texture measurement methods. Unlike other approaches, the texture measures are extracted from the whole retina without targeting any specific region. The method was tested on a set of 196 images composed of 110 healthy retina images and 86 glaucomatous images and achieved an area under curve (AUC) of 84%. A comparison performance with other texture measurements is also included, which shows our method to be superior.

1 Introduction

Glaucoma is the second leading cause of blindness worldwide. Its effects are irreversible but early detection can prevent vision loss. Unfortunately it is known that only half of the cases are identified, which may be due to glaucoma being asymptomatic in the early stages and there being no single test to diagnose it [1]. These have hampered the establishment of screening programs. With screening it is hoped that more cases can be detected at an earlier stage and treatment can be provided. With current techniques, screening is not cost effective; improved methods of detection could reduce the costs of screening and lead to its adoption.

Three tests are used in diagnosing glaucoma: tonometry, optic disc/nerve layer examination and visual field testing. Tonometry is a process of measuring intraocular pressure (IOP). But IOP alone has limited effectiveness as a screening tool as many studies have shown that there is no threshold value that discriminates between normal eyes and those with glaucoma [1]. Visual field testing has been shown to have relatively effective when used as a screening test [2]. However, it is time consuming, requires sophisticated equipment and trained operators. It also requires that patients understand the instructions, cooperate and complete the test; older patients sometimes fail to complete the test [1]. Optic disc assessment can be the method of choice for glaucoma screening. It involves examination of the optic disc for signs of glaucoma either directly or through a fundus camera or 3D imaging instruments. 3D instruments provide accurate results but are expensive and thus not widely available. A cheaper option is to use a fundus camera which is also easier to use and allows data to be captured quickly. In this paper, we present the latest results of our ongoing work on glaucoma classification using fundus images. The

approach makes use of texture analysis based on Binary Robust Independent Elementary Features (BRIEF) [3].

(a) Normal retina (b) Glaucoma

(c) Normal retina, CDR = 0.4 (d) Glaucoma, CDR = 0.7

Figure 1. Anatomy of (a) a normal retina and (b) a retina affected by glaucoma. CDR values for (c) a normal retina and (d) a retina affected by glaucoma.

2 Related Work

Examples of retinal images are shown in Figure 1. Figure 1(a) shows a normal retina and its main features. Contrast this with the appearance often found in a glaucoma-affected retina, Figure 1(b). The signs of glaucoma are enlargement of the optic cup, the presence of peripapillary atrophy, retinal and optic disc haemorrhages and Retinal Nerve Fibre Layer (RNFL) defects.

Previous work on glaucoma detection has either detected retinal anatomy or visual features. The first approach has segmented the main structures, derived some parameters from them and hence classified the image. Measurements based on cup to disc ratio (CDR), blood vessel area and RNFL defect detection [4, 5] have been used. Of these, CDR is the most common; it is defined as the ratio between the vertical diameter of the cup to the vertical diameter of the disc (Figures 1(c) and 1(d)). To measure it, segmentation of optic disc and optic cup is required.

One of the advantages of using CDR for glaucoma assessment is that it is sensitive to glaucomatous changes in the optic disc since it measures cup deformation [6]. However some studies found out that CDR alone cannot separate normal and glaucoma cases [7]. Like IOP, it is widely distributed with no single separating value [1]. Measuring CDR also requires accurate segmentation of the optic disc and cup. Errors in segmentation may lead to misdiagnosis of the disease. The reported accuracy of these methods is yet to achieve the required levels for large scale screening.

To avoid the difficulties associated with segmentation, we may extract features from the retina image to be used for classification. These features can be extracted from the whole image or from a specific region of interest i.e. the optic disc region. Many features have been used, e.g. image intensities, Discrete Fourier Transform (DFT) and B-spline coefficients [8], textural features [9, 10, 11, 12] and a combination of textural and structural features [6]. The advantage of this approach is it does not require segmentation as it performs a statistical data mining technique on image patterns themselves. With

careful design, this approach is capable of achieving robustness against inter and intra image variations [6].

3 Method

Our approach to classifying the retinal images uses image features. It follows the standard machine learning pipeline of feature extraction followed by classification.

3.1 Feature Extraction

Many authors pre-process retinal images to correct uneven illumination, remove distracting image structures [8] or resample the images to normalise the optic disc size [6]. We do not do this, we do not even smooth the image beforehand to suppress noise. Firstly, BRIEF is invariant to image illumination. Secondly, we believe that vascular change is one of glaucoma's indicators, so removing the vasculature is inappropriate. Thirdly, whilst it is believed that resampling the image can lead to robustness against size variations, the process may adversely affect the image's texture.

We process the green channel of the retinal image because it provides better contrast than the other channels. BRIEF has been used as a feature for image matching and recently as a texture measurement for optic disc segmentation [13]. It was chosen primarily because it addresses the illumination issues and is cheap to compute. BRIEF uses a binary string to encode the appearance of image patches. It forms a descriptor of a patch by comparing the intensities of n predefined pixel pairs. It uses a test τ on a patch p of size SxS:

$$\tau(p:\boldsymbol{x},\boldsymbol{y}) = \begin{cases} 1 \; if \; \big(p(\boldsymbol{x}) - p(\boldsymbol{y})\big) > Threshold \\ 0 \; otherwise \end{cases} \quad (1)$$

where $p(\boldsymbol{x})$ and $p(\boldsymbol{y})$ are the pixel intensities at locations $\boldsymbol{x}$ and $\boldsymbol{y}$, which are randomly selected (cf method I in [3]). The BRIEF descriptor is then defined as the n bit vector:

$$f_n = \sum_{1 \le i \le n} 2^{i-1} \tau(p:\boldsymbol{x}_i, \boldsymbol{y}_i) \quad (2)$$

In the current implementation, the descriptor is calculated using S = 27 and n = 16, as in [13], where we showed these parameters gave the best results. The threshold is based on the image noise, as estimated by the noise variance (σ). We set the threshold value to 3σ. Finally, the image is described by the histogram of BRIEF values.

3.2 Classification

In this step images are labelled using the computed features. A Support Vector Machine (SVM) is used as the classifier. We used the SVM from the LIBSVM package [14] and a non-linear radial basis kernel.

3.3 Comparison of Performance

Other than evaluating BRIEF's performance in classifying retinal images, we also compared BRIEF and other texture measurements: Grey Level Co-Occurrence (GLCM) [15], Grey Level Difference (GLD) [16], Local Binary Pattern (LBP) [17], Rank transform [18] and Completed Modelling of LBP (CLBP) [11]. These were chosen because they characterise texture in a similar manner to BRIEF. CLBP was also used for glaucoma classification [12].

In our implementation of GLCM the window size is 3x3, displacement is 1 pixel and the directions were 0, 45, 90 and 135 degrees. We used the GLCM features directly, similar to

[19], rather than compute the Haralick features. In the implementation of GLD, similar parameters were used as for GLCM.

In this experiment we used the original LBP methodology. CLBP is an extension of LBP. In CLBP, a local region is represented by its centre pixels and a local difference. The centre pixels represent the image grey level and are converted into binary code by global thresholding (CLBP C). The local difference is then decomposed into sign (CLBP S) and magnitude (CLBP M) components. The sign component is equivalent to normal LBP. The CLBP feature is the combination of the CLBM C, CLBP M and CLBP S. In the implementation of CLBP, similar parameters were used as for LBP.

4 Results and Discussion

We tested the approach using 86 glaucoma and 110 normal images, provided by the Manchester Royal Eye Hospital (from a data set used in training optometrists). The images were originally 2000 by 1312 pixels and were manually cropped to exclude the vignette, samples are shown in figure 2. During the classification experiment, we employed 10 fold cross validation.

4.1 Classification Results

The receiver operating characteristic (ROC) curve and balance accuracy (Bac) were used as the performance measures. Bac was used because our classes have different numbers of samples. It is defined as the arithmetic mean of sensitivity and specificity, or the average accuracy obtained on the classes. In the case of balanced classes, this parameter corresponds to the classification accuracy. We obtained a Bac of 78% and an area under the ROC (AUC) of 84%.

Samples of images that have been classified correctly and incorrectly are shown in Figure 2.The first column shows correctly classified normal images. The second column shows images correctly classified as glaucoma. These images are characterised by bigger cups and in some cases the presence of atrophy (the pale regions outside the disk).

The images shown in the third column are misclassified normal images, it can be seen that they show similar characteristics to the glaucoma retina. A similar observation can be made for misclassified glaucoma images in the fourth column, they have characteristics that resemble normal images.

4.2 Comparison of Performance Results

The ROC of BRIEF, GLCM, GLD, LBP, CLBP and Rank transform is shown in Figure 3. BRIEF achieved the highest AUC of 84%. LBP achieved the second highest with 77% and CLBP was in third place with 76%.

Our results seem to disagree with [12] who found CLBP to be effective in classifying glaucoma. There are two possible reasons for this. Firstly, they used a small dataset (41 images), so their results may not be entirely reliable. Secondly, they reduced the image resolution; it can be argued that the scaling process biases the final classification.

[8 and 6] tested with larger datasets. [8] proposed a two stage glaucoma classification system using features based on intensities, DFT coefficients and B-spline coefficients. It was tested with 575 images and achieved an AUC of 88%. [6] took a hybrid approach,

combining features derived from the optic disc region based on: CDR, cup to disc area ratio (CDA), RNFL defect detection probabilities and atrophy defect detection probabilities. The approach was evaluated using 1962 images. The AUCs obtained were 64% using CDR, 61% using CDA, 68% using features derived from optic disc regions and 73% using a combination of CDR, CDA and defect detection probabilities from RFNL and atrophy. For this comparison, we will consider the result obtained using only image features. Altogether, [6] extracted 84 image features and the AUC was 68%.

Based on the AUC, the results in [8] are better than ours. However, their approach involved normalisation of optic discs which may, as they stated, bias the glaucoma variation and affect the classification. To counter this effect, they evaluated their approach using a dataset with limited optic disc size variation. This restriction was not applied in our dataset or in [6]. For the record our dataset has an average vertical optic disc diameter of 444±53 pixels. Further evaluation by [6] revealed that performance of the method in [8] drops to AUC of 61% when tested using a wider variation of optic disc sizes.

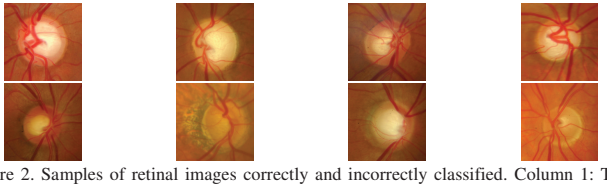

Figure 2. Samples of retinal images correctly and incorrectly classified. Column 1: True negatives, Column 2: true positives, Column 3: false negatives, Column 4: false positives.

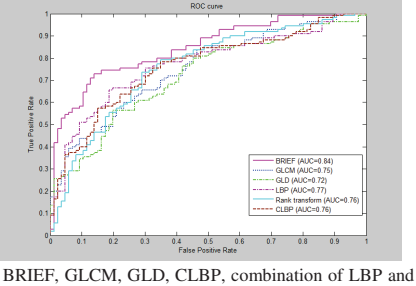

Figure 3. ROC of BRIEF, GLCM, GLD, CLBP, combination of LBP and CLBP and rank transform.

5 Conclusions and Further Work

In this paper, we described an approach to glaucoma classification using a texture-based description: BRIEF. The method was able to achieve an AUC of 84% for glaucoma classification. Comparison of performance between BRIEF and other texture measurements (GLCM, CLD LBP, CLBP and Rank Transform) was also conducted and we showed that our method performed best.

The key differences of our approach compared to others are: firstly, the use of a texture feature for the classification. In most of the methods more types of features are used, which means feature selection must be applied to find the most relevant features for classification. Second, in our approach the image features are derived from the whole image rather than just the optic disc region. This way, we take account of both deformation found within the optic disc and from other regions of the retina.

In future we plan an evaluation using a larger dataset to evaluate the feasibility of the approach for a glaucoma screening program.

References

[1] A. Rotchford, "What is practical in glaucoma management?" Eye, vol. 19, no. 10, 2005.

[2] J. Katz, A. Sommer, D. E. Gaasterland, and D. R. Anderson, "Comparison of analytic algorithms for detecting glaucomatous visual field loss," Archives of ophthalmology, vol. 109, no. 12, p. 1684, 1991.

[3] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in Computer Vision–ECCV 2010. Springer, 2010, pp. 778–792.

[4] C. Muramatsu, T. Nakagawa, A. Sawada, Y. Hatanaka, T. Yamamoto, and H. Fujita, "Automated determination of cup-to-disc ratio for classification of glaucomatous and normal eyes on stereo retinal fundus images," Journal of biomedical optics, vol. 16, no. 9, pp. 096 009– 096 009, 2011.

[5] J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T.-C. Lim, "Automated diagnosis of glaucoma using digital fundus images," Journal of medical systems, vol. 33, no. 5, pp. 337–346, 2009.

[6] G. D. Joshi, "Automatic retinal image analysis for the detection of glaucoma," Ph.D. dissertation, International Institute of Information Technology, Hyderabad, 2014.

[7] J. S. Lai, C. C. Tham, and J. C. Chan, "The clinical outcomes of cataract extraction by phacoemulsification in eyes with primary angle-closure glaucoma (pacg) and co-existing cataract: a prospective case series," Journal of glaucoma, vol. 15, no. 1, pp. 47–52, 2006.

[8] R. Bock, J. Meier, L. G. Ny´ul, J. Hornegger, and G. Michelson, "Glaucoma risk index: Automated glaucoma detection from color fundus images," Medical image analysis, vol. 14, no. 3, pp. 471–481, 2010.

[9] U. R. Acharya, S. Dua, X. Du, S. Vinitha Sree, and C. K. Chua, "Automated diagnosis of glaucoma using texture and higher order spectra features," Information Technology in Biomedicine, IEEE Transactions on, vol. 15, no. 3, pp. 449–455, 2011.

[10] M. R. K. Mookiah, U. Rajendra Acharya, C. M. Lim, A. Petznick, and J. S. Suri, "Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features," Knowledge-Based Systems, vol. 33, pp. 73–82, 2012.

[11] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," Image Processing, IEEE Transactions on, vol. 19, no. 6, pp. 1657–1663, 2010.

[12] M. Alsheh Ali, T. Hurtut, T. Faucon, and F. Cheriet, "Glaucoma detection based on local binary patterns in fundus photographs," Proc. SPIE, Medical Imaging, Computer Aided Diagnosis, vol. 9035, pp. 903 531–903 531–7, 2014.

[13] S. Mohammad, D. T. Morris, and N. A. Thacker, "Segmentation of optic disc in retina images using texture," in VISAPP 2014 – Proceedings of the 9th International Conference on Computer Vision Theory and Applications, Volume 1, Lisbon, Portugal, 5-8 January, 2014, 2014, pp. 293–300.

[14] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu. edu.tw/~cjlin/libsvm.

[15] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," Systems, Man and Cybernetics, IEEE Transactions on, vol. 6, pp. 610–621, 1973.

[16] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," Systems, Man and Cybernetics, IEEE Transactions on, vol. 6, no. 4, pp. 269–285, 1976.

[17] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on, vol. 1. IEEE, 1994, pp. 582–585.

[18] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in Computer Vision ECCV'94. Springer, 1994, pp. 151–158.

[19] P. Pacl´ık, S. Verzakov, and R. P. Duin, "Improving the maximum likelihood co-occurrence classifier: a study on classification of inhomogeneous rock images," in Image Analysis. Springer, 2005, pp. 998–1008.