AUTOMATED QUANTIFICATION OF FUNDUS IMAGE QUALITY

J. Lowell*, A. Hunter*, M. Habib ** and D. Steel **

* University of Lincoln, Computer Science, Lincoln, UK
** Sunderland Eye Infirmary, Sunderland, UK

j.a.lowell@durham.ac.uk

Abstract: This paper presents an algorithm for the automatic assessment of retinal image quality for lesion detection. The paper focuses on the technical aspects of quantifying image quality into levels of gradability. Image quality is attained by summing a weighted contrast between vascular centre pixels and the background, multiplying the resultant by the contrast between the fovea and retina. It is assumed that blood vessel centrelines and the fovea have been previously located. Images are subdivided into 5 categories of gradability. The algorithm is evaluated against three alternative approaches using a set of 200 fundus images from a normal screening population. The results demonstrate a significant performance improvement over other previously published approaches.

Introduction

The suitability of fundus images for diagnosis is an important issue in automated detection of diabetic retinopathy. Inadequate image quality can affect image classification as subtle visual differences between diabetic and non-diabetic lesions may become hazy leading to misclassification. In more extreme cases, loss of definition or obscurity may cause diabetic lesions to blend completely into the retinal background and be left unidentified and unclassified. To prevent unsound classifications, ungradable images require automatic exclusion from automated analysis and should be flagged for either a repeat screening or ophthalmic review.

Defining image quality is a highly subjective abstract process based on the varying experiences, visual perception and the judgement of observers. In automated retinal analysis, a standardized so-called "typical experienced observer" is required to filter images into *achievable*, *minimum* and *ungradable* categories to avoid erroneous classifications from substandard images.

The general guidelines of image quality suggest using two 45° field images per eye. A macula centred field is essential and is referred to as the defined position. Optic disc and nasal fields are a bonus. Image quality is based on the macula centred images. The three defined levels are:

(1) **Achievable standard:** Optic disc less than or equal to one disc diameter from the defined position. Small vessels clearly visible within one disc diameter of the fovea and optic disc and visible across more than 90% of remaining image(s); see figure 1.

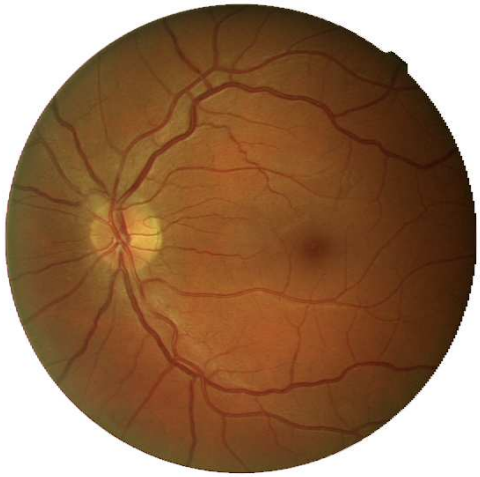

Figure 1: Achievable standard

(2) **Minimum standard:** Optic disc less than or equal to two disc diameters from the defined position. Small vessels clearly visible within one disc diameter of fovea and optic disc and visible across more than 66% of remaining image(s); see figure 2.
(3) **Inadequate (ungradable):** Optic disc less than or equal to two disc diameters from the defined position. Small vessels not clearly visible within one disc diameter of fovea and optic disc and visible across more than 33% of remaining image(s); see figure 3.

In order to comply with national screening guidelines, the image assessment algorithm should be capable of correctly partitioning images into the aforementioned three categories. To achieve this, a measure of quality is required; although image quality measures are well known in the domain of image restoration, diagnostic suitability is a relativity new research area with only limited publications. Usher *et al* [1] presented an algorithm to determine image quality using a quality metric based on the area of automatically detected blood vessel. As blood vessels should be present in all retinal images regardless of ethnic origin or retinopathy, Usher measured vessel frequency from gradable and ungradable images to determine an im-

ISSN: 1727-1983

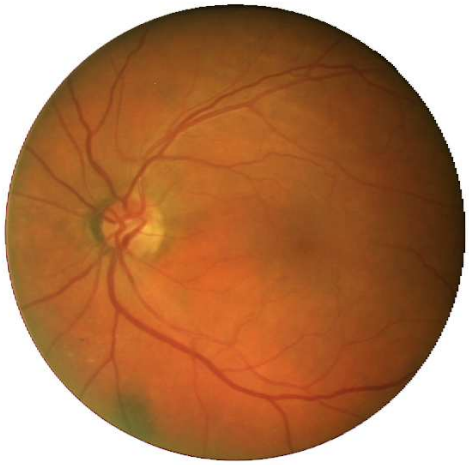

Figure 2: Minimum standard

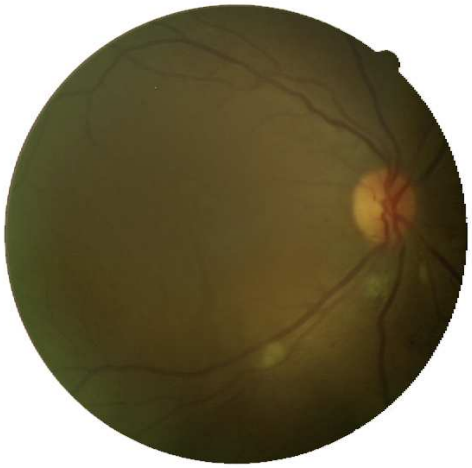

Figure 3: Inadequate (ungradable) a) poor illumination b) cataracts

age quality metric. Within each image an image quality metric score V was set from the total count of pixels classified as vessels. Images with blood vessel metrics above a threshold t_v were classified as gradable while images with metrics below t_v were classified as ungradable. The algorithm reportedly achieved 100% sensitivity and 94% specificity in detecting patients with at least one ungradable image.

Lee *et al* [2] studied 360 retinal images from the Oklahoma native Americans and concluded that image quality could be defined by three parameters: brightness, contrast and signal-to-noise ratio (SNR). From the sample set, twenty images with excellent quality were selected. Quality parameters were obtained from these images together with an average intensity histogram - referred to as the desired values and the template intensity histogram respectively. Lee observed that the brightness, contrast and signal-to-noise values of an image were close to their respective desired values when the image's intensity histogram was close to the template intensity histogram and that these values could be derived from the histogram. An image quality measure was therefore proposed using the convolution of the template histogram with the image histogram and computing a quality index.

In an evaluation of Lee's work, Lalonde *et al* [3] examined the interdependency between image quality and histogram similarity in 40 retinal images of varying quality. During this study, histograms from several poor quality images were found to closely resemble the template histogram. In addition, histograms from several good quality images were notably different from the template histogram, signifying a weak connection between image quality and histogram similarity. Lalonde *et al* [3] experimented with distribution of edge magnitudes and the local distribution (as opposed to the global histogram of Lee) of pixel intensity as quality indicators. In a similar approach to Lee, a typical edge magnitude histogram was formed using the edge maps from a set of good quality images. The difference between the typical and current image edge magnitude histogram formed a quality indicator. A second qualify indicator was derived by comparing local intensity distributions. This approach differs from Lee *et al*, by defining a set of local histogram templates instead of one global histogram template. Lalonde concluded that both quality indicators could help discriminate between good and bad images, although a larger image set was required to evaluate the performance of the approach.

Materials and Methods

The measurement of retinal image quality is concentrated in the macula region - two disc diameters around the centre of the fovea. This method assumes previous fovea localisation. It is also assumed that the vascular network has been previously segmented leaving a vascular map of segment centrelines. The image quality assessment is performed on the green intensity channel of the colour RGB fundus images and is then normalised to increase its dynamic range.

The appearance of small blood vessels within one disc disc diameter of the fovea (macula) is the primary indicator of fundal image quality. There are three aspects of macula vessels that indicate quality - distance from fovea, contrast and quantity. The more blood vessels that are visible within the macula, the closer they are to the fovea centre and the higher the contrast between the vessels and the background retina, the better the image quality.

For each vascular segment, $i = \{1..S\}$, the average pixel distance away from the fovea and the contrast with the background retina is measured. The distance of a vascular segment i from the fovea is determined by averaging the distance between each pixel $p_0..p_{\eta_i}$ from centreline α_i and the fovea centre c; as expressed by equation 1. Let η_i represent the number of pixels contained within

ISSN: 1727-1983

centreline α_i.

$$\overline{\gamma_i} = \frac{1}{\eta_i} \cdot \sum_{j=1}^{\eta_i} \sqrt{(p_{x_j} - c_x)^2 + (p_{y_j} - c_y)^2} \quad (1)$$

The contrast of vascular segment i is measured by taking the average intensity of centreline pixels α_i and subtracting it from the average intensity of boundary pixels β_i. To ensure that pixels measured by a segment's boundary are of the surrounding background retina and not the edge of its own blood vessel, boundary diameters are set to twice the average macula vessel diameter of 7 pixels. A segment boundary β_i is formed by dilating a segment centreline α_i by two structuring elements A and B; structuring element B is subsequently subtracted from A, where element $A = ones(6,4)$ and $B = ones(4,2)$. The mean pixel intensities of centreline α_i and boundary β_i are expressed by equation 2 and 3 respectively. Let $b_0..b_{\tau_i}$ represent the pixels within boundary β_i and I denote the normalised green channel of the fundus image.

$$\overline{\alpha_i} = \frac{1}{\eta_i} \cdot \sum_{j=1}^{\eta_i} I_{p_{x_j}, p_{y_j}} \quad (2)$$

$$\overline{\beta_i} = \frac{1}{\tau_i} \cdot \sum_{j=1}^{\tau_i} I_{b_{x_j}, b_{y_j}} \quad (3)$$

where τ_i represents the pixel count of boundary β_i. The intensity contrast for vessel segment i is calculated by simply subtracting $\overline{\alpha_i}$ from $\overline{\beta_i}$, as expressed by equation 4.

$$\omega_i = \overline{\beta_i} - \overline{\alpha_i} \quad (4)$$

In general, the greater the length of a visible macula blood vessel, the greater the quality. This quality improves further if the vessel has good contrast with the retina. The vascular metric for a vessel is therefore the product of the segment pixel length by its contrast measure. The final aspect of vascular quality is the distance away from the fovea; the closer a vessel is to the fovea centre the higher the quality and the smaller the segment's average distance. The vascular metric penalises distant blood vessels by dividing through by the average distance. The vascular metric φ_i for vessel segment i is expressed by equation 5. The overall vascular metric υ is simply the sum of individual vessel metrics; as expressed by equation 6.

$$\varphi_i = \frac{\eta_i \cdot \omega_i}{\overline{\gamma_i}} \quad (5)$$

$$\upsilon = \sum_{i=1}^{S} \varphi_i \quad (6)$$

The foveal contrast within the macula is the secondary indicator of fundal image quality, because its also indicative of lesion contrast. Macula regions with limited contrast between the fovea and background retina can cause lesion concealment, with lesions appearing either washed out or shaded depending upon the macula exposure. The foveal contrast is defined by comparing the core intensity of the fovea to the background retina of the macula. To compensate for any minor disparities between the real fovea centre and its estimated location, the foveal core is defined as a circular region with radius $r = 10$ (half that of an average fovea radius), originating at centre of the fovea. This reduced size ensures that intensity measurements are kept within the fovea. The fovea intensity $\overline{\rho}$ is determined by taking the average intensity of each pixel $f_0..f_\varsigma$ from the foveal core, where ς is the core's pixel count; as expressed by 7.

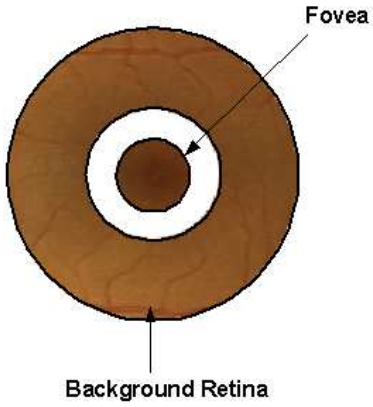

Figure 4: Fovea Quality Measure

$$\overline{\rho} = \frac{1}{\varsigma} \cdot \sum_{i=1}^{\varsigma} I_{f_{x_i}, f_{y_i}} \quad (7)$$

To maintain maximum intensity differences between the fovea and background retina, a buffering zone is used to separate intensity regions; see Figure 4. The background intensity of the macula is measured in a region between two concentric circles originating at centre of the fovea with radius $r = 30$ and $r = 60$ respectively. The average intensity $\overline{\kappa}$ of each pixel $m_0..m_\varepsilon$ within this region is defined as the macula intensity, as expressed by equation 8, where ε represents the macula pixel count.

$$\overline{\kappa} = \frac{1}{\varepsilon} \cdot \sum_{i=1}^{\varepsilon} I_{m_{x_i}, m_{y_i}} \quad (8)$$

The overall fovea contrast measure μ is simply the difference between the average intensity of the fovea core $\overline{\rho}$ and the macula retina $\overline{\kappa}$; as depicted by equation 9.

$$\mu = \overline{\rho} - \overline{\kappa} \quad (9)$$

An overall image quality metric χ, is defined by the product of the primary and secondary quality indicators. The vascular quality metric υ, incorporating the foveal distance, contrast and pixel quantity of each macula vessel segment is multiplied by the macula contrast

ISSN: 1727-1983

μ between fovea and background retina; as expressed by equation 10:-

$$\chi = \upsilon \cdot \mu \tag{10}$$

The national screening guidelines for image quality divides images into three categories: *achievable*, *minimum* and *ungradable*. Deciding which category an image should join is a subjective process, and is especially so if an image is of a mixed quality lying between two categories. A wider categorisation of image quality was therefore deemed necessary to accommodate borderline images. A further two categories were added for images between *achievable* and *minimum*, and *minimum* and *ungradable* quality.

For the overall macula quality metric χ to be able to differentiate between image quality categories, metric boundaries are required for each quality division. This was achieved using a training set of 100 images that were assessed by an ophthalmologist, who categorised the images into 5 groups of ascending quality. Category metric boundaries were subsequently extracted; as shown in table 1.

Category	From	To
1	331	
2	101	330
3	36	100
4	6	35
5	0	5

Table 1: Quality Metric Boundary

Results

The performance of the macula image quality assessment algorithm is evaluated against two alternate techniques. Results from 200 screening images showed that the macula model is 6% more accurate than the compared approaches and is capable of categorising image quality into 5 groups, matching the clinician's classification with an accuracy of 91% and detecting all clinically ungradable images.

Fundus image quality guidelines specify that macula clarity must be such that blood vessel are clearly visible. With this in mind, 200 standard 760×570 fundus images were presented to an ophthalmologist, with the macula region of interest emphasized by an overlayed circle.

Images were graded by the ophthalmologist on a scale of $1-5$. Images with small blood vessels visible around the fovea and with good foveal contrast with the background macula area were graded as 1, images with similar vascular detail but with reduced foveal contrast were labelled as 2. Images that retained foveal contrast but only included macula periphery blood vessels where graded as 3, and were labelled as 4 with reduced foveal contrast. Where no vessels were visible, the image was labelled as 5. It is against this benchmark that the accuracy and precision of the algorithm is measured.

A comparison is made between the presented algorithm and two alternate fundal image assessment approaches - Lalonde *et al*'s [3] template intensity histogram and Usher *et al*'s [1] vascular metric. In Usher's algorithm, the vascular metric consisted of the sum of all pixels contained within the blood vessel network. In this study it was found more reliable to morphologically thin the segmented blood vessels to a centreline and sum the vascular centreline pixels. This reduces metric variability due to blood vessel width and treats all blood vessels with equal importance, whereas Usher's algorithm is biased by the pixels contained within the major temporal retinal vascular arcades and macula vessels having little influence on the overall vascular metric. It is against this modified algorithm that the macula model is evaluated.

Lalonde stated that images could be crudely categorised into three groups: "good", "fair" and "bad". However, in testing Lalondes algorithm on 200 screening images, it was deemed impractical to perform a $3-$way split due to the overlapping metrics within each category. Therefore, as the alternate algorithms can only distinguish between gradable and ungradable images, comparisons were made with benchmark classification $1-4$ deemed as gradable, and 5 ungradable.

In this application it is important that the algorithm identifies 100% of ungradable images thus avoiding any potential misclassifications due to poor image quality. For this reason, performance is sensitivity biased, with metric thresholds dividing gradable and ungradable images selected from Receiver Operating Characteristic (ROC) curves where the sensitivity achieves 100%. The thresholds used to distinguish gradable and ungradable images for the macula model, Usher and Lalonde algorithms are 5, 4586.31 and 11419 respectively. Table 2 shows the results of the presented and alternate algorithms.

Usher [1] reported a sensitivity of 100% and specificity of 94%; however, in this evaluation, the specificity was 87%. Lalonde's algorithm achieves a poor specificity of 20% with 100% sensitivity, but results dramatically improve by allowing one false negative classification, giving 95% sensitivity and 81% specificity respectively. It is worth noting that the macula model algorithm has an accuracy of 94%, which is 6% more than Usher's algorithm and has almost half the false positive classifications.

The observed relationship between compared algorithms was tested using an ANOVA statistical significance test, testing the null hypothesis that there are no mean differences between groups. The null hypothesis was rejected, implying that there is a significant difference between groups.

To test the null hypothesis that the error mean of the "macula model" is equal to the next best model, the "Usher's" model, a t-test was performed. The t-test value is 5.3355×10^{-025}. With 99 degrees of freedom in each sample, the critical value is 4.2207×10^{3} at 99% con-

ISSN: 1727-1983

Table 2: Image Quality Assessment Algorithm Performance

Model	Macula model	Usher	Lalonde
Sensitivity	100	100	100
Specificity	93	87	19.5
Accuracy	94	88	28
True Positive	21	21	21
False Negative	0	0	0
True Negative	167	156	35
False Positive	12	23	144

fidence. The null hypothesis is therefore rejected, signifying that the Macula model is better than the Usher's image quality algorithm. Similar pairwise comparisons indicate the Combined vascular model's superiority over Lalonde's Template intensity histogram metric.

The purpose of this algorithm is to detect ungradable images. The evaluation above considers clinically ungradable images. There is, however, a distinction between clinical and computerised gradability. To achieve an optimum classification it may be necessary to use a more stringent gradability criteria, thus ensuring only the sharpest images proceed to neural assessment. Although this ensures an idyllic classification environment, there is an increase in ungradable images which require manual review.

Table 3: Macula Model Image Quality Categorisation Performance

Grade	Correct	Incorrect	%
1	25	1	96
2	33	0	100
3	20	3	85
4	15	4	73
5	7	0	100
Total	92	8	92

The categorisation performance of the algorithms is evaluated against 100 of the 200 clinically assessed screening images. Table 3 shows that 91% of the automated image quality assessments matched the clinician. The remaining 9% are all within one grade of the clinician's classification. It is worth noting the system detected 100% of clinically ungradable images (grade 5). The high categorisation accuracy of this approach means that automated image quality assessment can not only exclude clinically ungradable images, but also exclude borderline cases leaving only the highest quality images for automated classification.

Conclusions

Automated assessment of fundal image quality is an important stage of automated retinal analysis, with inadequate image quality leading to unsound diabetic lesion and image classifications.

General guidelines for fundal image quality provide three definitions of image clarity; all referring to the visibility of small vessels within one disc diameter of the fovea. Previous image quality methods have concentrated on template histogram comparisons and global vascular pixel counting to identify ungradable images. The presented technique has focused on measuring the clarity of small macula blood vessels. Comparing the quality rating returned by the presented algorithm, Lalonde, and Usher, it was shown that the presented model is more accurate by 6%.

Using only macula vascular measurements, images could be correctly identified as gradable or non-gradable. By adding a foveal contrast measure, images can be further subclassified into 5 degrees of gradability. In addition to the further investigations into foveal template models and amalgamated vascular pixel metrics, further work is required to include optic disc clarity and image alignment as described in the National Screening Committee's (NSC) [4]) image quality guidelines.

References

[1] D.B. USHER, M. HIMAGA, M.J. DUMSKYJ, J.F. BOYCE, A. SABATE-CEQUIER, and WILLIAMSON T.H. Automated assessment of digital fundus image quality using detected vessel area. *Procedings of medical image understanding and analysis the university of sheffield*, pages 81–84, 2003.

[2] S.C. LEE and Y. WANG. Automatic retinal image quality assesment and enhancements. In *Proc SPIE Vol 3661 Med Imaging:Img Processing*, pages 1581–90, 1999.

[3] M. LALONDE, M. BEAULIEU, and L. GAGNON. Automated visual quality assessment in optical fundus images. In *Vision Interface*, pages 259–264, 2001.

[4] NHS. A model for a photographic diabetic retinopathy screening programme. Internet, 10 2004.

ISSN: 1727-1983