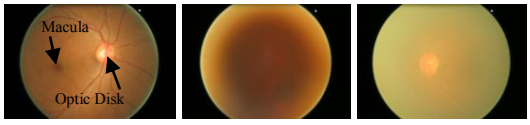
Automatic Retinal Interest Evaluation System (ARIES)

Fengshou Yin, Damon Wing Kee Wong, Ai Ping Yow, Beng Hai Lee, Ying Quan, Zhuo Zhang, Kavitha Gopalakrishnan, Ruoying Li, and Jiang Liu

***Abstract*— In recent years, there has been increasing interest in the use of automatic computer-based systems for the detection of eye diseases such as glaucoma, age-related macular degeneration and diabetic retinopathy. However, in practice, retinal image quality is a big concern as automatic systems without consideration of degraded image quality will likely generate unreliable results. In this paper, an automatic retinal image quality assessment system (ARIES) is introduced to assess both image quality of the whole image and focal regions of interest. ARIES achieves 99.54% accuracy in distinguishing fundus images from other types of images through a retinal image identification step in a dataset of 35342 images. The system employs high level image quality measures (HIQM) to perform image quality assessment, and achieves areas under curve (AUCs) of 0.958 and 0.987 for whole image and optic disk region respectively in a testing dataset of 370 images. ARIES acts as a form of automatic quality control which ensures good quality images are used for processing, and can also be used to alert operators of poor quality images at the time of acquisition.**

I. Introduction

Cataracts, glaucoma, age-related macular degeneration, diabetic retinopathy and pathological myopia are the major causes of blindness. Of these, vision lost in eye diseases such as glaucoma and AMD tends to be permanent and irrecoverable. Such permanent visual loss is caused by the damage and degeneration of retinal layer. Hence, the critical need to detect blinding eye diseases early and effectively in large populations has spurred the development of computer-aided tools in recent years.



(a) *(b)* *(c)*

Fig. 1: Fundus of eye (*a*); Poor-quality fundus images (*b*, *c*)

Many detection methods are developed based on datasets of clean, filtered images. Due to the necessity of assessing the performance of the algorithms and techniques on disease and lesion detection, clean, sharp and clear images are often selected for technique development, training and testing. However, in practice, such clean high-quality images are often difficult to acquire due to imaging artefacts during image capturing. When such poor-quality images (Fig. 1b and 1c) are encountered by the automated analysis systems, the diagnosis results generated may not be reliable and leads to inaccuracy.

F. Yin, D.W.K. Wong, A.P. Yow, B.H. Lee, Y. Quan, Z. Zhang, K. Gopalakrishnan, R. Li and J. Liu are with the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. (Contact e-mail: fyin@ i2r.a-star.edu.sg).

Moreover, it is difficult to restore such poor images to an improved state without artificial interpolations as they are obtained during the acquisition process as part of the imaging optics or disease pathologies rather than as a result of compression. Furthermore, many computer aided diagnosis (CAD) systems assess a specific region of interest in a fundus image, such as optic disk and macula (Fig. 1*a*). In addition, a poor quality image may not be necessarily made up of all poor quality focal regions. Thus, there remains a need to detect both poor quality images and poor quality focal regions at a pre-processing level for filtering out before use in any computer-aided detection method.

There are a number of methods for image quality assessment and can be divided into full-reference and no-reference methods. Since full-reference methods are mainly used to compare the effects of image compression, we will focus on no-reference methods, in particular of ocular images. Examples of no-reference methods include distribution-based method [1], segmentation-based method [2], bag of words-based method [3] and, vessel and color-based method [4, 5]. However, these methods do not analyse the focal regions of interest.

Hence, we proposed ARIES, Automated Retinal Interest Estimator System to automatically assess and control the quality of both full input images and focal region of interest (optic disk) as a pre-processing step before passing processed images for subsequent analysis.

II. Methods

In ARIES, the quality of an input image is assessed in three steps, viz. retinal image identification, confirmation of non-retinal images and images quality assessment of retinal images, as illustrated in Fig. 2.

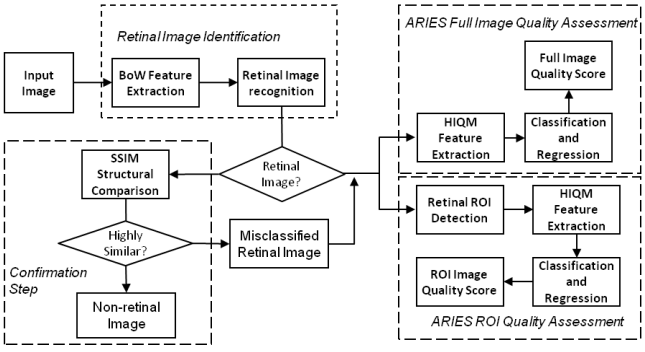


Fig. 2: ARIES flowchart

At the end of the processing, both full image quality score and focal region (optic disk) quality score will be given, to filter out low quality images or provide a warning of low confidence for further steps of processing.

A. Retinal Image Identification

It is important to filter out irrelevant images in the computer aided diagnosis of ocular diseases. The proposed retinal image classification method utilizes a weakly-supervised learning approach by deciding whether an image exhibits certain characteristics with no segmentation of objects or manual selection of features. In the first step, the training images are represented using a bag of visual words representation. Next, a Support Vector Machine (SVM) classifier is trained using the extracted features and labels. Finally, the trained classifier can be applied to test on new images.

B. Confirmation of Non-retinal Images

In fundus images, artefacts, especially overexposure can affect the accuracy of the image classification in the previous step and often lead to wrong classification of fundus images into non-fundus types. Thus, a confirmation step is desirable to distinguish real non-fundus images and fundus images with artefacts. The identified fundus image with artefacts can be classified as poor quality fundus image directly or be fed to ARIES quality assessment step to assess focal regions.

In order to reduce the misclassification rate, a confirmation method based on the structural similarity index (SSIM) [6] is introduced. SSIM is a full reference method for measuring the similarity between two images. However, SSIM is used in a different way in our proposed method. Instead of the full reference approach, a mean fundus image is generated from a set of fundus images as the reference image. Intuitively, non-fundus images such as scenery images and other ocular images should have low SSIMs. On the other hand, fundus images with artefacts should have high SSIMs. Therefore, overexposed fundus images can be identified and separated from non-fundus images.

C. Image Quality Assessment of Retinal Images

The image quality assessment step distinguishes between the high and poor quality retinal images as well as focal regions of the image. The first step for focal region image quality assessment is to detect the focal region of interest (ROI). Subsequently, high level image quality measures (HIQM) are extracted from the ROI image to form the feature space. Finally, a SVM classification is performed. The full image quality assessment process follows a similar flow except that the full image rather than ROI image is used to extract the HIQM features. In the followings, we will illustrate the process using the optic disk (OD) image.

- *Optic Disk ROI Detection*

  We use the method described in [7] for OD localization. To detect the OD centre accurately based on intensity values, bright fringes are identified and removed. The fringes are extracted by locating a circle slightly smaller than the eyeball in the greyscale image and thresholded for high intensity pixels outside the circle. The centre of the OD is approximated by the centroid of the remaining bright pixels. The ROI is then defined as an image that is about twice the diameter of the normal OD.

- *HIQM Feature Extraction*

  Instead of using low level image features that have extremely high dimensions, we propose to use high level image quality measures (HIQM) for its high relevance to the problem, low dimensionality and hence fast speed. There are three categories of HIQM features and they are summarized as follows:

  Contrast and Blur Features

  High level contrast and blurriness measures include the contrast ratio measurements, the blur measures, the intensity ranges, and saturation metrics. Contrast ratio is calculated as:

$$CR_j = \bar{p}_j / s_j \tag{1}$$

  where $\bar{p}_j = \text{mean}(I_j)$ is the mean intensity of all the pixels in channel j of the image I in RGB color space, $s_j = \text{std}(I_j)$ is the standard deviation of all the pixel intensities, and j can be red channel r, green channel g, blue channel b and greyscale gs.

  Higher contrast ratios correspond to higher blurriness. It is intuitively true as blur images usually have small variance in intensity, which leads to a high contrast ratio. Similar to contrast ratio, we also use the local contrast ratio on non-overlapping sub-windows of the image, which is defined as:

$$LCR = \left(\sum_{i=1}^{n} \frac{\bar{p}_{w,i}}{s_{w,i}}\right)/n \tag{2}$$

  where w is an N x N pixel window, and n is the total number of sub-windows.

  Blur metric (*BM*) [8] is based on the discrimination between different levels of blur perceptible on the same picture. This measure is robust in measuring focal blur and motion blur. Mathematically, it is obtained by comparing the intensity variations of the original image and its blurred version by a low-pass filter.

  Intensity ranges, including full intensity range (*R*), relative intensity range (*RR*) and interquartile range (*IQR*), are important metrics to measure the greyscale spread of images. Image with high quality or contrast usually has a larger intensity range compared to one with low quality.

$$R = \max(I) - \min(I) \tag{3}$$

$$RR = \frac{\max(I) - \min(I)}{\text{mean}(I)} \tag{4}$$

$$IQR = \mathrm{Q}_3(I) - \mathrm{Q}_1(I) \tag{5}$$

  where I is the array of all pixel intensities of a greyscale image, Q_1 and Q_3 are the 1st and 3rd quartile values.

  Saturation metrics include percentage of maximal (*Pmax*) and percentage of minimal (*Pmin*), measuring the proportions of pixels at the highest and the lowest intensity respectively. The former is useful to identify overexposed images; while the latter is able to identify underexposed images.

  Entropy Features

Entropy of an image can be used to represent the amount of information in it. It is calculated as:

$$E = -\sum_i p_i log_2(p_i) \quad (6)$$

where p_i is the probability that the difference between two adjacent pixels is equal to i.

If an image is perfectly histogram equalized, the spread of greyscale values is at maximum i.e. image has maximum entropy. Conversely, the entropy for a binary image is very low as it has only two states. Entropy is zero if an image has flat-value pixels.

For optic disk region of interest, a high quality image should contain clear structure of optic disk, optic cup and blood vessels, which corresponds to high entropy. On the other hand, a poor quality image without fine structural features has low entropy.

Image Structure Features

The optic disk region has a high density of blood vessels. Thus, blood vessel density (*BVD*) can be used as an important feature to distinguish low and high quality optic disk image. Fig. 3 illustrates the difference of detected vasculatures of two different images. Another important structure feature is the maximum edge length or edge spread (*ES*). The edges of blood vessels and optic disk boundary are usually continuous for a high quality image, as is the maximum edge spread. To compute *BVD* and *ES*, we first detect blood vessels in the image using a fast bottom-hat filtering method. The bottom-hat filter is applied to the histogram equalized green channel image (g) to obtain $\tilde{g}$. Subsequently, the blood vessel map *M* is determined by

$$M(i,j) = \begin{cases} 1 & \tilde{g}(i,j) > T \\ 0 & otherwise \end{cases} \quad (7)$$

After obtaining the vessel map, *BVD* can be obtained through

$$BVD = \frac{\sum_{i=1,j=1}^{m,n} M(i,j)}{m \times n} \quad (8)$$

where *m* and *n* represent the width and height of the image respectively.

The edge spread is calculated as the maximum major axis length of all connected components in *M* divided by the diagonal length of the image.

$$ES = \frac{\max(l_1, l_2, \ldots, l_k)}{\sqrt{m^2+n^2}} \quad (9)$$

where l represents the major axis length of each connected component in *M*.

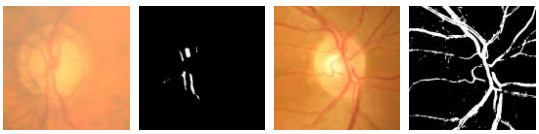


Fig. 3: Blood vessel detection for a low quality image and a high quality image

- *Image Quality Classification and Assessment*

In the final step, support vector machine (SVM) is used to train the classifier. The SVM-based classification can be formulated as a quadratic programming problem with linear constraint:

$$\min_{w,b,\xi} \quad \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N}\xi_i \quad (10)$$

$$subject\ to \quad y_i(\omega^T\phi(x_i) + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, \ldots, N$$

where x_i is the training vector, y_i is the training label, ω is the normal vector, b is the offset, ξ_i is the slack variable to implement soft margin, C is the penalty term and ϕ is the kernel function.

Instead of using binary classification results only from SVM, the outputs of the SVM decision function (decision value) are also used. We normalized the decision values to produce a retinal image quality score (*RQS*):

$$RQS = wd + b \quad (11)$$

where *d* represents the decision value, *w* and *b* are constant coefficients. *RQS* have values from 0 to 1, with higher value represents better image quality.

III. Results and Discussions

A. Datasets

The proposed system is studied on images from a range of fundus image databases such as the Singapore Malay Eye Study (SiMES) [9], Singapore Chinese Eye Study (SCES) [10] and Blue Mountains Eye Study (BMES). A number of non-fundus image databases such as slit-lamp images, OCT images, Retcam images and scenery images are also used to test the retinal image identification. Images for subsequent quality assessments are from SiMES database.

B. Fundus Image Identification

The fundus image identification algorithm was trained on 6200 images, including 2700 fundus images and 3500 non-fundus images. It was tested on a batch of 35342 images, consisting of 23441 fundus images and 11902 non-funds images. The system achieves 99.54% accuracy in fundus and non-fundus image classification in the testing set.

In the training stage, 2700 fundus images, 500 OCT images, 500 Retcam images, 500 slit-lamp images and 2000 scenery images are randomly chosen as the training dataset. The rest of the images are used as the testing dataset. The summary of training and testing datasets, and experimental results is shown in Table 1.

Table 1: Summary of experimental results for fundus image identification

Database Name	Image Type	# Training Images	Accuracy on Training Set	# Testing Images	Accuracy on Testing Set
SiMES	Fundus	500	100%	5428	98.80%
SCES	Fundus	500	100%	1176	100%
BMES	Fundus	500	100%	5722	99.76%
Other Fundus	Fundus	1200	100%	11114	99.26%
ACHIKO-NC	Slit-lamp	500	100%	5030	100%
AGAR Database	OCT	500	100%	1164	100%
AGATE Database	Retcam	500	100%	3199	100%
Scenery	Scenery	2000	100%	2509	100%
Total		**6200**	**100%**	**35342**	**99.54%**

C. *Fundus Image Quality Assessment*

- Full Image Assessment

  The algorithm was trained and tested on a database of 740 images, in which 600 images of good quality and 140 images are of poor quality. The main causes for poor quality images are existence of cataracts, uneven illumination and overexposure during taking. All images are of the dimension of 3072 x 2048 pixels. In the experiment setting, half of images are randomly chosen for training (300 good-quality images and 70 poor-quality images) and the other half are used for testing.

  The algorithm can achieve a classification accuracy of 94.3% and 91.6% for the training set and testing set respectively. In terms of full image quality score (fullRQS), the area under curve (AUC) of the receiver operating characteristic curve achieves 0.967 and 0.958 respectively.

  For comparison purpose, we implemented a no reference quality metric (Q_V) based on blood vessel trees in the fundus image [5]. Tested on the same testing data set of 370 images, the Q_V metric achieves an AUC of 0.796. The ROC curves from both methods are shown in Fig. 5*a*.

- Optic Disk ROI Assessment

  Focal region (ROI) quality assessment is the most unique feature of this work. For optic disk ROI quality assessment, the ROI images are extracted with a dimension of 800 x 800 pixels. The dataset used for this part is optic ROI images from the full image dataset, as it was selected according to the quality of optic disk region as well as the full image. The system achieves 95. 4% and 96.0% accuracy in the testing and training set respectively.

  The ARIES system also produces an OD quality score (odRQS) in the range of 0 to 1. A score that is close to 1 represents a high quality image. On the other hand, a score that is close to 0 indicates a very low quality image. Fig. 4 shows images with different levels of image quality score.

  The Q_V metric is also implemented on the optic disk ROI images of the testing dataset, which achieves an AUC of 0.532. However, the proposed odRQS can achieve a better AUC of 0.987 (Fig. 5*b*).

IV. Conclusion

In this paper, we propose an image quality assessment system which can automatically distinguish retinal fundus images from non-fundus images and assess both full retinal image and the region of interest for key structures. Experimental results show that the system achieves a high accuracy in the retinal image identification step. In addition, the produced *RQS* measurement can effectively quantify the quality of both full image and focal region of interest. The proposed system will improve the reliability of existing retinal image processing systems, provide a way to control the input quality, as well as be a potential acquisition tool, through considering useful retinal interest regions or the whole image.

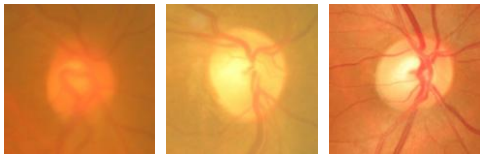
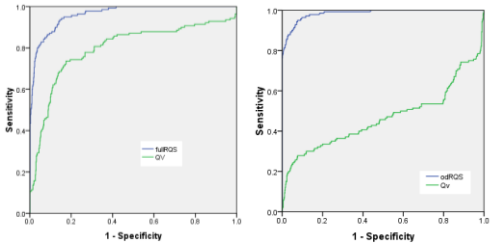


Fig. 4: Optic disk interest image with odRQS of 0.30 (left), 0.66 (middle) and 0.92(right)



(a) *(b)*

Fig. 5: *(a)* ROC curves of fullRQS (blue) and Q_V metric (green); *(b)* ROC curves of odRQS (blue) and Q_V metric (green)

References

[1] M. Lalonde, L. Gagnon, and M. Boucher, "Automatic visual quality assessment in optical fundus images", *Proceedings of Vision Interface*, 2001, pp. 259–264.

[2] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson and P. F. Sharp, "Automated assessment of diabetic retinal image qualiy based on clarity and field definition", *Investigative Ophthalmology and Visual Science*, 2006 Mar, 47(3): 1120–1125.

[3] M. Niemeijer, M. D. Abramoff and B. van Ginneken, "Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening", *Medical Image Analysis*, 2006 Dec, 10(6): 888–898

[4] L. Giancardo, M. D. Abramoff, E. Chaum, T. P. Karnowski, F. Meriaudeau and K. W. Tobin, "Elliptical local vessel density: a fast and robust quality metric for retinal images", *Conf Proc IEEE Eng Med Biol Soc.* 2008, pp. 3534-7

[5] T. Kohler, A. Budai, M. F. Kraus, J. Odstrcilik, G. Michelson, J. Hornegger, "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation", *IEEE 26th International Symposium on Computer-Based Medical Systems (CBMS)*, 2013, pp.95-100

[6] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[7] Z. Zhang, B. H. Lee, J. Liu, D. W. K. Wong, N. M. Tan, J. H. Lim, F. Yin, W. Huang, H. Li, T. Y. Wong, "Optic Disc Region of Interest Localization in Fundus Image for Glaucoma Detection in ARGALI," *5th International Conference on Industrial Electronics & Applications (ICIEA)*, pp. 1686-1689, 15-17 Jun 2010.

[8] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric", Proc. *SPIE 6492, Human Vision and Electronic Imaging XI*I, 64920I (February 12, 2007); doi:10.1117/12.702790.

[9] A. W. Foong, S. M. Saw, J. L. Loo, S. Shen, S. C. Loon et al., "Rationale and methodology for a population-based study of eye diseases in malay people: The singapore malay eye study (SiMES)", *Ophthalmic Epidemiol* 14: 25-35, 2007.

[10] C. C. Sng, J. C. Allen, M. E. Nongpiur, L. L. Foo, Y. Zheng, C. Y. Cheung, M. He, D. S. Friedman, T. Y. Wong and T. Aung, "Associations of Iris Structural Measurements in a Chinese Population: The Singapore Chinese Eye Study", *Invest. Ophthalmol. Vis. Sci*. 23 April 2013 vol. 54 no. 4, 2829-2835.