

# Colour Normalisation to Reduce Inter-Patient and Intra-Patient Variability in Microaneurysm Detection in Colour Retinal Images

M. J. Cree<sup>1</sup>, E. Gamble<sup>1</sup> and D. Cornforth<sup>2</sup>

<sup>1</sup>Dept. Physics and Electronic Engineering  
University of Waikato  
Hamilton, New Zealand

<sup>2</sup>School of Information and Environmental Sciences  
Charles Sturt University  
Albury, Australia  
E-mail: m.cree@ieee.org

## Abstract

*Images of the human retina vary considerably in their appearance depending on the skin pigmentation (amount of melanin) of the subject. Some form of normalisation of colour in retinal images is required for automated analysis of images if good sensitivity and specificity at detecting lesions is to be achieved in populations involving diverse races. Here we describe an approach to colour normalisation by shade-correction intra-image and histogram normalisation inter-image. The colour normalisation is assessed by its effect on the automated detection of microaneurysms in retinal images. It is shown that the Naïve Bayes classifier used in microaneurysm detection benefits from the use of features measured over colour normalised images.*

## 1 introduction

Indigenous populations such as the Australian Aborigine, the New Zealand Māori and the Canadian Inuit all have 4–5 times the incidence of diabetes compared to the Caucasian population resident in these countries [4]. This large percentage of the population and their geographical distribution necessitate special diabetes screening models to optimise screening, detection and treatment. One such model is to undertake a mobile population screening programme of diabetic retinopathy [13]. Although the cost is reduced when compared to current costs associated with visits to the general practitioner for a referral followed by a visit to the ophthalmologist, this endeavour is still prohibitive due to the cost and lack of specialists [3, 15]. These shortcomings can be addressed by utilising automated procedures (that are easily implemented by diabetes technicians), and which

identify pre-proliferative diabetic retinopathy that is often already present before diabetes is detected by clinical symptoms [9].

Automated assessment of pre-proliferative diabetic retinopathy has been possible for some time using fluorescein-labelled images [18, 6, 7]. Results for colour fundus analysis identifying microaneurysms, exudates and cotton-wool spots have only been reported more recently [10, 17, 20, 8, 16, 19].

To optimise automated processing of colour images one has to consider intra-image variation such as light diffusion, the presence of abnormalities, variation in fundus reflectivity and fundus thickness and inter-image variation (being the result of using different cameras, illumination, acquisition angle and retinal pigmentation). Osareh selected a retinal image as a reference and histogram specification followed by a global and local contrast enhancement step [16]. A comparison between methods was recently undertaken by Goatman *et al.* [11], who compared grey world normalisation, histogram equalisation and histogram specification to that of a standard image. In their study histogram specification performed best. The problem with histogram specification is that certain lesions are reflected in the shape of the histogram and by reshaping the histogram to that of a standard image, which does not necessarily contain the lesion, the evidence for the lesion can be masked in the resultant histogram. An example is that exudates, which have a yellow appearance and occur only in the occasional retinal image, result in a long tail in the histogram of the green plane. This tail is removed if histogram specification to a retinal image not containing exudates is used.

We anticipate that a form of normalising image colour, both intra-image and inter-image, that preserves the shape of individual colour component histograms is likely to bet-

ter preserve evidence for certain lesions. We therefore propose a new approach to colour normalisation of colour retinal images, and test it for its efficacy to increase the discrimination in certain features useful for the automated detection of microaneurysms—a lesion that often occurs as one of the first signs of diabetic retinopathy.

## 2 Colour Normalisation



**Figure 1. Two retinal images; one of a Caucasian and one of a Polynesian.**

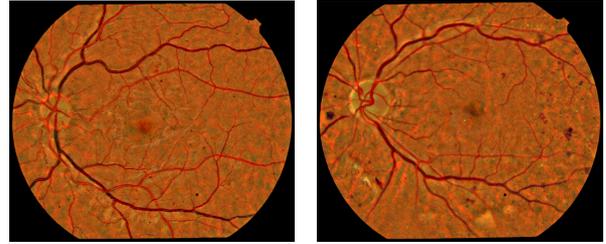
Two examples of retinal images, one of a Caucasian and one of a New Zealand Māori, can be seen in Figure 1. The difference in colouration between the two is quite noticeable as is the variation in colour within a single retinal image. Since we normalise colour both for intra and inter-image variation, and because the intra-image variation is partially due to misillumination of the retina and thus should be corrected for first, we separate the normalisation process into two stages, the first for intra-image correction and the second to normalise between images.

The well known technique of divisive shade-correction is first applied to each colour plane (in RGB colour space) of the retinal image to correct for intra-image variation. This is achieved by dividing each colour plane of the retinal image by the background approximated by gross median filtering of the respective colour plane.

The histogram of each colour plane of the shade-corrected image is then adjusted to have a specified mean and standard deviation within a region-of-interest delineating the camera aperture. This process retains the overall shape of the histogram, but shifts the hue (which is roughly dominated by the ratio of green to red in retinal images) to be consistent between images. The colour normalised images for Figure 1 are shown in Figure 2.

## 3 Microaneurysm Detection

To test the colour normalisation we examine its effect on the automated microaneurysm detector of Streeter and



**Figure 2. Colour normalised images of Figure 1.**

Cree [19]. Microaneurysms appear as small round red objects, usually separated from the vasculature, in colour retinal images. The automated microaneurysm detector used follows a similar process to that established by Spencer *et al.* [18] and Cree *et al.* [5, 6]. Candidates (i.e. objects that bear some similarity to microaneurysms) are segmented from the green plane of the retinal image by shade-correcting the green plane, removing the blood vessels, then match-filtering with a standard microaneurysm template to detect candidates. This process is not specific enough so a number of shape and colour features are measured on the candidates, which are used as inputs to a classifier, to better distinguish between the microaneurysms and other segmented spurious objects. We ask whether the colour normalisation process described in this paper provides better features for classification of the candidates, than features derived without normalisation.

## 4 Testing Methodology

Sixty retinal images of patients with diabetic retinopathy at 50° field-of-view were obtained from a Topcon fundus camera with a Nikon D1X 6 megapixel digital camera. The automated microaneurysm detector was run on each of the 60 images to the stage of segmenting candidates bearing similarity to microaneurysms. A number of shape and colour features were measured on each segmented candidate. The mean, standard deviation and second moment, about the axis perpendicular to the candidate through the centroid, normalised to area (which we refer to as ‘rotational inertia’) were measured on each colour plane (red, green and blue) and on hue (calculated as red/green) using the original images and using the colour normalised images. This gives a total of 24 colour features. In addition a number of other features, including those based on shape, were extracted for an overall total of 52 features.

Each candidate was labelled as a microaneurysm or as a spurious object by an expert in the field. The feature dataset formed from the 60 retinal images contained 2623 candidates of which 2222 were marked as spurious objects and

401 as microaneurysms by the expert. The expert identified another 14 microaneurysms in the images that were not segmented by the automated procedure generating the candidates. As it is the improvement in classification achieved with the colour normalisation that is under question, we report sensitivities as out of the 401 microaneurysms in the feature dataset.

Two analyses were used to quantify the improvement due to the use of colour normalisation. Exploratory statistical analysis was applied to the individual features to measure class means and standard deviations for each feature. Signal-to-noise ratios (SNR) were derived from these measures based on the assumption of a Gaussian probability distribution function for each of the two classes, according to

$$SNR = \frac{|\bar{x}_{obj} - \bar{x}_{ma}|}{\sqrt{\frac{1}{2}(\sigma_{obj}^2 + \sigma_{ma}^2)}} \quad (1)$$

where  $\bar{x}$  represents the mean and  $\sigma$  the standard deviation of the two classes when measured over one feature. It is to be noted, that with the assumption of underlying Gaussian probability distributions, the SNR is a monotonically increasing function of the area under the curve of the receiver operating characteristic curve [1].

As a more rigorous test, a cross-validation of training and testing using a Naïve Bayes classifier was applied using the Weka package [21]. The Naïve Bayes algorithm [2] assumes that features are independent. Knowing how these features have been derived would lead one to suspect this of being a rather flimsy assumption, but the algorithm is known to perform surprisingly well in some domains, and is very fast to run [12]. It estimates prior probabilities by calculating simple frequencies of the occurrence of each feature value given each class, then returns a probability of each class, given an unclassified set of features. These probabilities were used to derive ROC curves in the results section.

The feature dataset was split up into 60 separate training data sets containing the candidates for 59 images; each training set missing out the candidates for each image in turn. In addition 60 testing data sets were made to accompany the training data sets by including those candidates that are not in the respective training set. The reason to split up the training/testing datasets based on images rather than taking a naïve random selection of candidates is to ensure that test datasets actually simulate a truly new image for classification. The Naïve Bayes classifier was used to quantify the relative success of different feature sets.

It is well known that using too many features can actually degrade accuracy of the prediction, so optimising the accuracy of such methods involves a choice not only of classifier algorithm, but also of the appropriate features. Kohavi [14] has studied the automatic selection of features and

concluded:

- The optimum feature set will depend on the classifier model chosen
- Therefore the feature set may be considered a parameter of the model
- The evaluation of feature sets will be biased in a favourable direction unless it uses independent data.

Kohavi suggests a wrapper approach, where the actual classifier algorithm is used to evaluate the features selected.

Included in the Weka toolbox is a Wrapper Subset Evaluator. This takes as a parameter the name of the classifier being used for the discriminant function. The wrapper does a search in feature space for the set that gives the lowest error on the given classifier.

To implement the wrapper process, we took the 60 training datasets and applied wrapper subset evaluation to each one to find the best feature set for each dataset, that is, the feature set that maximised the classification accuracy using the Naïve Bayes classifier. The results of these 60 trials were then combined to provide counts for the number of times each feature was indicated. Following this, all 52 features were ranked according to how often they were selected, and the ten most frequent were selected. A new collection of 60 training and testing datasets were prepared as described before, but contained only these 10 features.

The results for the 60 images were combined and used to generate an ROC curve. As the Naïve Bayes classifier is completely deterministic, there was no variation observed over multiple runs, so only one run was necessary to evaluate.

Subsequently, we prepared a control by taking the same datasets, and scrambled the class labels, so that the same records were randomly labelled as spurious objects or as microaneurysms. We repeated the Naïve Bayes classifier test as described in the first evaluation above, and prepared an ROC curve.

To provide some measure of the benefits afforded by selecting the correct images, we prepared a further 10 datasets, where the features selected were chosen at random. We performed the Naïve Bayes classifier test as described in the first evaluation above, but using these non-optimal feature sets, and reported the results.

## 5 Results

Table 1 lists the SNRs for the various colour features measured over each segmented candidate. Four colour variables are used, where red, blue and green are from the RGB colour space, and hue is calculated as red/green. The label ‘Mean’ refers to the mean colour measured over the extent of the candidate, likewise ‘Std. Dev’ to the standard

Colour	Measurement	SNR (original)	SNR (normalised)
Red	Mean	0.29	0.11
	Std. Dev.	0.05	0.32
	Rot. Inert.	0.24	0.29
Green	Mean	0.06	0.85
	Std. Dev.	0.67	1.23
	Rot. Inert.	0.40	0.94
Blue	Mean	0.15	0.02
	Std. Dev.	0.16	0.50
	Rot. Inert.	0.25	0.25
Hue	Mean	0.22	0.44
	Std. Dev.	0.01	0.44
	Rot. Inert.	0.23	0.27

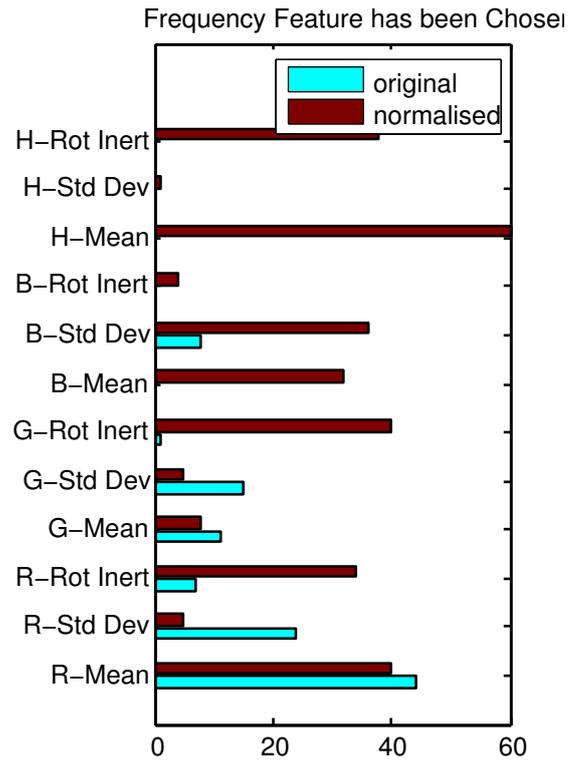
**Table 1. SNR as a measure of discrimination for the features measured.**

deviation and ‘Rot. Inert.’ to the second moment calculated along the radial direction from the centroid of the candidate (equivalent to the rotational inertia), divided by the area of the candidate.

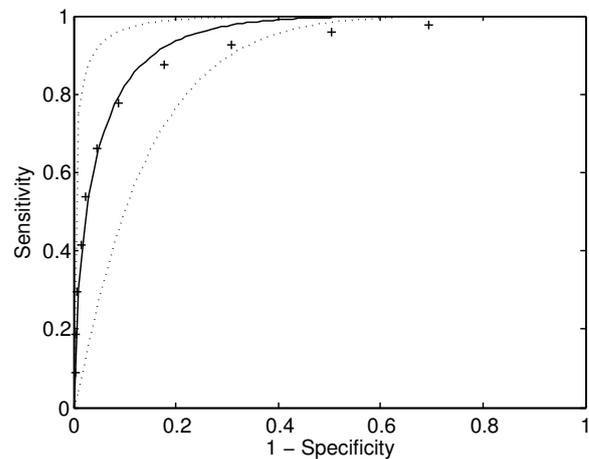
As can be seen in the table, the SNRs calculated over the colour normalised images are for the most part equal to or greater than those calculated over the original images. The two exceptions are for the means over the red and blue colour planes, for which the colour normalisation has reduced the SNR.

Fig. 3 shows the results of the wrapper process on the original dataset. The vertical axis shows labels for each of the 12 colour features available measured over the original images and the colour normalised images. The length of each bar shows the relative frequency with which that features was selected. As there were 60 applications of the wrapper method, the maximum any feature could be selected was 60 times. One of the features, H-Mean (mean of the hue) measured over the colour normalised image, was indeed selected in every application of the wrapper method. As can be seen in Fig. 3, the colour normalised features were, in general, preferentially chosen over those calculated over the original images. It should be noted that for the above trial the feature database included some shape features in addition to the colour features reported herein; for the purposes of this paper we are only interested in the results pertaining to the colour features.

The results of the evaluation of the original dataset are shown in Fig. 4. The area under this ROC curve is indicative of the discriminant ability of the classifier, and in this case indicates a good performance. The results of the evaluation using randomly labelled data are shown in Fig. 5. In complete contrast, the control shows zero discriminant ability. The results of the ten evaluations using randomly selected features are shown in Fig. 6. In this case, the classifier is capable of making a reasonable performance, but

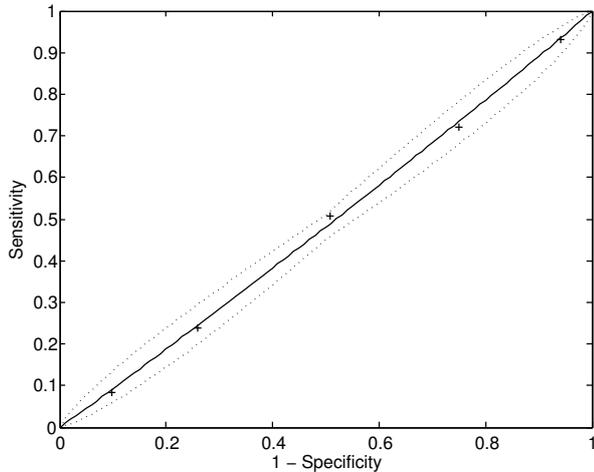


**Figure 3. Bar graph showing how often a particular feature was chosen in the forward feature selection process.**

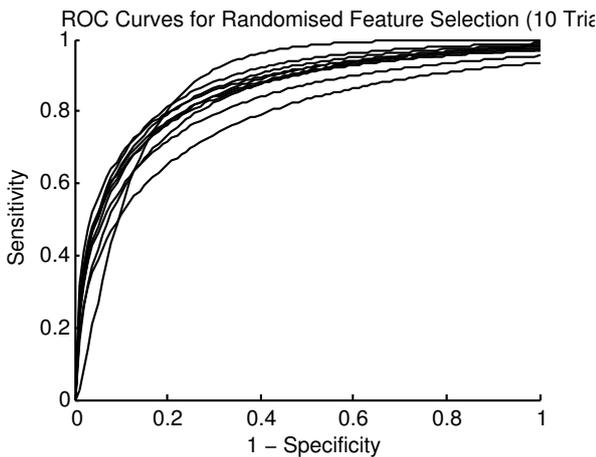


**Figure 4. ROC graph for the classifier using the 10 best features. The solid line is the fitted ROC curve to the data points (plus signs). The dotted curve indicates the 95% confidence intervals.**

lacks the performance of the feature set chosen by the wrapper method.



**Figure 5. ROC graph for randomly labelled data.**



**Figure 6. 10 ROC curves for the classifier using 10 randomly selected features.**

## 6 Discussion

The SNR results (table 1) demonstrate that the colour normalisation process increases the discrimination in almost all of the colour features treated individually. To quantify the relative predictive power of the combined features a forward selection process was run with a Naïve Bayes classifier. This preferentially chose the colour normalised features over the features that were measured over the original

images.

Previous studies have tended to focus on histogram equalisation or histogram specification however we argue that the distortion that can occur in the histogram with these methods can mask certain lesions. An example of such a lesion is exudate, which appears in the green histogram as a long extended tail. This tail can be masked if histogram specification is used. We therefore prefer to use colour normalisation that preserves the shape of the histogram.

Our results demonstrate that for detecting microaneurysms in colour retinal images, colour normalisation is beneficial. It is still to be established whether the colour normalisation process described herein will be beneficial for such tasks as the automated segmentation of the vasculature or of other lesions such as exudate and cotton wool spots. However, we have demonstrated a reasonable approach to enable diabetic retinopathy screening for indigenous populations.

Acknowledgements: MJC gratefully acknowledges the financial assistance of the Waikato Medical Research Foundation.

## References

- [1] H. H. Barrett, C. K. Abbey, and E. Clarkson. Objective assessment of image quality. III. ROC metrics, ideal observers and likelihood-generating functions. *J. Opt. Soc. Am. A*, 15:1520–1535, 1998.
- [2] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Royal Soc. London*, 53:370–418, 1763.
- [3] J. Betz-Brown, K. L. Pedula, and K. H. Summers. Diabetic retinopathy—contemporary prevalence in a well-controlled population. *Diabetes Care*, 26(2637–2643), 2003.
- [4] S. Colagiuri, R. Colagiuri, and J. Ward. *National Diabetes Strategy and Implementation Plan*. Diabetes Australia, Paragon Printers, Canberra, Australia, 1998.
- [5] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. Automated microaneurysm detection. In *IEEE International Conference on Image Processing*, volume 3, pages 699–702, Lausanne, Switzerland, September 1996.
- [6] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. A fully automated comparative microaneurysm digital detection system. *Eye*, 11:622–628, 1997.

- [7] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. The preprocessing of retinal images for the detection of fluorescein leakage. *Phys. Med. Biol.*, 44:293–308, 1999.
- [8] B. M. Ege, O. K. Hejlesen, O. V. Larsen, K. Møller, B. Jennings, D. Kerr, and D. A. Cavan. Screening for diabetic retinopathy using computer based image analysis and statistical classification. *Comput. Meth. Prog. Biomed.*, 62:165–175, 2000.
- [9] M. M. Engelgau, K. M. V. Narayan, and W. H. Herman. Screening for type 2 diabetes. *Diabetes Care*, 23:1563–1580, 2000.
- [10] G.G. Gardner, D. Keating, T. H. Williamson, and A. T. Elliott. Automatic detection of diabetic retinopathy using an artificial neural network: A screening tool. *Br. J. Ophthalmol.*, 80:940–944, 1996.
- [11] K. A. Goatman, A. D. Whitwam, A. Manivannana, J. A. Olson, and P. F. Sharp. Colour normalisation of retinal images. In *Proceedings of Medical Image Understanding and Analysis*, 2003.
- [12] J. Han and M. Kamber. *Data Mining Concepts and Techniques*. Morgan Kaufmann, 2001.
- [13] H. F. Jelinek, D. Cornforth, and et al. M. Cree. Automated characterisation of diabetic retinopathy using mathematical morphology: A pilot study for community health. In *2nd Annual NSW Primary Health Care Research and Evaluation Conference*, page 48, Sydney, Australia, 2003.
- [14] R. Kohavi and G. John. Wrappers for feature subset selection. *Artificial Intelligence Journal*, 97:273–274, 1996.
- [15] S. J. Lee, C. A. McCarty, H. R. Taylor, and J. E. Keefe. Costs of mobile screening for diabetic retinopathy: A practical framework for rural populations. *Aust. J. Rural Health*, 9:186–192, 2001.
- [16] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham. Classification and localisation of diabetic-related eye disease. In *7th European Conference on Computer Vision (ECCV)*, pages 502–516, May 2002.
- [17] C. Sinthanayothin, J. F. Boyce, H. Cook, and T. Williamson. Automated localisation of the optic disc, fovea and retinal blood vessels from digital colour fundus images. *Br. J. Ophthalmol.*, 83:902–912, 1999.
- [18] T. Spencer, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput. Biomed. Res.*, 29:284–302, 1996.
- [19] L. Streeter and M. J. Cree. Microaneurysm detection in colour fundus images. In *Image and Vision Computing New Zealand 2003*, pages 280–284, Palmerston North, New Zealand, November 2003.
- [20] H. Wang, W. Hsu, K.G. Goh, and M. L. Lee. An effective approach to detect lesions in colour retinal images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 181–187, 2000.
- [21] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools with Java implementations*. Morgan Kaufmann, San Francisco, CA., 2000.