Algorithms evaluation for fundus images enhancement

V Braem¹, M Marcos¹, G Bizai¹, B Drozdowicz¹² y A Salvatelli¹

¹Grupo de Inteligencia Artificial, Facultad de Ingeniería, Universidad Nacional de Entre Ríos. ²Facultad de Ciencia y Tecnología, Universidad Autónoma de Entre Ríos.
E-mail: gbizai@bioingenieria.edu.ar

Abstract. Color images of the retina inherently involve noise and illumination artifacts. In order to improve the diagnostic quality of the images, it is desirable to homogenize the non-uniform illumination and increase contrast while preserving color characteristics. The visual result of different pre-processing techniques can be very dissimilar and it is necessary to make an objective assessment of the techniques in order to select the most suitable. In this article the performance of eight algorithms to correct the non-uniform illumination, contrast modification and color preservation was evaluated. In order to choose the most suitable a general score was proposed. The results got good impression from experts, although some differences suggest that not necessarily the best statistical quality of image is the one of best diagnostic quality to the trained doctor eye. This means that the best pre-processing algorithm for an automatic classification may be different to the most suitable one for visual diagnosis. However, both should result in the same final diagnosis.

1. Introduction

In general, image pre-processing aims to improve the visual quality and basically prepare it for a subsequent automatic processing. Fundus images acquired with a fundus camera have artifacts such as noise and a characteristic effect of non-uniform illumination [1]. The illumination heterogeneity is due to the combined effects of the eyeball geometry, the dilation of the pupil, the alignment of the eye with the optical axis of the camera and the direction and shape of the lighting system.

A large number of algorithms that reduce the noise in the image, have been widely explored and developed. In contrast, the non-uniform illumination correction is a more difficult task that requires the combination of more complex techniques.

Both noise and illumination artifacts can hide information of diagnostic interest. In particular, diabetic retinopathy is a retinal complication that may lead to blindness. A fundus examination may provide an early diagnosis when the expert is able to detect the first signs, consisting of small aneurysms and exudates [2].

The visual results can be very dissimilar depending on the pre-processing technique applied. A pre-processed image can be seen as an improvement by an inexperienced observer, while an expert may consider it an image of a low diagnosis quality. The aim at this stage is, to standardize the images for later stages of unsupervised classification. It is also expected to provide the specialist with an image of better quality than the original for medical evaluation.

The performance of some pre-processing techniques was tested based on the results of two segmentation schemes: Fuzzy C-means and Fractal Dimension [1]. The diversity of the pre-process
results (visual results) leads to the necessity of proposing a quantitative method to judge, according to the change of some parameters, which is the performance of the tested techniques.

2. Materials and methods

2.1. Materials
Images from the “DIARETB1-Standard Database Diabetic Retinopathy” database, collected as part of the "ImageRet" project, from Lappeenranta University of Technology Kuopio, Finland, were used. They are available online at no cost [3]. The database consists of 89 images which were acquired using a digital fundus camera in 50 degrees field, with different image settings (flash intensity, shutter speed, aperture, gain), controlled by the system itself. Images contain noise. Optical aberrations and photometric accuracy are the same for all.

Thirty images chosen randomly from the database were used in this work. In order to reduce the Gaussian noise present in the images, a Wiener adaptive filter was applied to all images. The window was selected as 3x3, so as not to affect the small structures of the retina.

2.2. Methods
From the literature, eight algorithms were selected and implemented. The algorithms deal separately with the illumination problems, contrast enhancement and color preservation. The main purpose of the chosen techniques is the correction of non-uniform illumination, but it was also considered the contrast improvement and color preservation.

It should be noted that the implemented algorithms use a mask that defines the region of interest (ROI), as proposed by Frank ter Haar [4]. The red channel is binarized and the mask with morphological operators is delimited. The mask eliminates undesirable border effects.

An overview of the proposed algorithms follows.

Algorithm 1
The first algorithm is based on the Equalization of Illumination technique [5]. Each pixel is adjusted according to the equation:

$$I_{eq}(x,y)=I(x,y)+m-A(x,y)$$

where $m$ is the desired average intensity (128 in a grayscale image of 8 bits) and $A(x,y)$ is the average value of pixels in an $N \times N$ window $W$, centered on the pixel to be set.

Algorithm 2
It uses the Homomorphic Filtering technique, which considers the image as the product of an illumination component $I(x,y)$ and the reflectance component $R(x,y)$. $I(x,y)$ has smooth variations and its spectrum is concentrated on the low frequencies, while $R(x,y)$ shows sharp variations with its spectrum concentrated on the high frequencies. The idea is to isolate the reflectance $R$ and remove the illumination $I$ in order to obtain an image without illumination variations.

Algorithm 3
This algorithm applies an averaging filter to each color channel $C$ in order to obtain an illumination estimate $LC$. The mean value $mLC$ is subtracted from $LC$ and the corrected channel $CC$ is:

$$CC=C-(LC-mLC)$$

Later, the Local Contrast Enhancement (LCE) technique proposed by Xiaohui and Chutatape [6] is applied. Considering a sub-size image $W N \times N$ pixels size, centered on the pixel to be set, the mean $\langle f \rangle$ and standard deviation $\sigma$ of the intensity is calculated. These local values are the parameters for the following transformation function:
\[ \Psi_w(f) = \left[ 1 + e^{-\frac{f - f_{\text{ave}}}{\sigma_w}} \right]^{-1} \]

**Algorithm 4**

The mean value \( \mu_{\text{channel}} \) of the logarithm of each channel is calculated and a high pass filter \( PA \) is applied. From the result of the high pass filter \( PA_{\text{channel}} \) the mean value \( \mu_{PA_{\text{channel}}} \) is obtained, so that:

\[ \text{channel} = \mu_{\text{channel}} + A.(PA_{\text{channel}} - \mu_{PA_{\text{channel}}}) \]

where \( A \) is a constant equal for the three channels. The aim is to obtain the reflectance alone, without illumination, while preserving the original image colors. Finally the LCE algorithm is applied.

**Algorithm 5**

Both a high pass and a low pass filter are applied to the logarithm of each color channel to obtain \( PA_{\text{channel}} \) and \( PB_{\text{channel}} \). Then the mean values \( \mu_{PA_{\text{channel}}} \) and \( \mu_{PB_{\text{channel}}} \) are calculated and the following formula is applied (\( A \) and \( B \) are weighting factors):

\[ \text{channel} = \text{channel} + A.(PA_{\text{channel}} - \mu_{PA_{\text{channel}}}) - B.(PB_{\text{channel}} - \mu_{PB_{\text{channel}}}) \]

By subtracting the low pass results, the luminosity variations are decreased. By adding the high pass results, the contrast of the retinal structures is increased, the original colors of the image are preserved by removing the value of the media. Finally the LCE algorithm is applied.

**Algorithm 6**

This one is similar to algorithm 5 except for the fact that instead of adding and subtracting the filtered images to the original channel, it is made to its mean \( \mu_{\text{channel}} \), and LCE is applied to the result.

**Algorithm 7**

This algorithm is based on the iterative Retinex technique developed by Frankle and McCann [7, 8] and applied to fundus images [9]. The original image in RGB color space is converted to YUV color model and the V channel only is used.

**Algorithm 8**

This technique was proposed by Jobson et al [10]. It is based on multiscale Retinex technique with some modifications [9] so as to keep color fidelity in the image.

### 3. Results

The gray level mean and standard deviation of the images before and after the Wiener filtering was evaluated. The results showed a variation average percentage of 0.0971% while the standard deviation decreased by an average of 10.522%. These results indicate that the image brightness was kept while the noise was decreased.

The 30 sample images were processed with each of the algorithms above, previously applying 3x3 window size Wiener filtering.

To assess the result of the algorithms on the illumination homogeneity and the contrast, each original image was processed and transformed into grayscale images. Then one hundred 20 x 20 pixels ROI randomly located in the grayscale image were selected. The mean value \( \mu \) and standard deviation \( \sigma \) was calculated for every ROI.

#### 3.1. Algorithms performance to correct uneven illumination.

The average values of the original and processed ROI were used as illumination estimators in that area of the image. The standard deviation of the 100 samples was calculated to get an idea of their
dispersion, which gives an idea of the uniformity of illumination (greater deviation is related to less uniform illumination). Figure 1 shows the result for Algorithm 2 (the one of worst performance in homogenizing illumination) and Algorithm 3 (the one of best performance in this task.) Table 1 summarizes the obtained values for an image.

The 30 images were processed and the mean and standard deviation was obtained for each algorithm. Considering these results, the number of images for which the algorithms decreased or increased their illumination homogeneity with respect to the original was obtained. Algorithms 3, 4 and 7 increase the homogeneity 100% every time (the homogeneity increased in all 30 analyzed images), algorithms 5, 6 and 8 show values above 80%; Algorithm 1 gets close to 50%. Finally, Algorithm 2 yields poor results because the 74% of the times worsens the homogeneity.

Table 1. Average illumination and standard deviation for each algorithm, for an image.

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Alg. 1</th>
<th>Alg. 2</th>
<th>Alg. 3</th>
<th>Alg. 4</th>
<th>Alg. 5</th>
<th>Alg. 6</th>
<th>Alg. 7</th>
<th>Alg. 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Illumination</td>
<td>49.82</td>
<td>87.46</td>
<td>130.95</td>
<td>109.36</td>
<td>102.43</td>
<td>72.98</td>
<td>92.85</td>
<td>89.00</td>
<td>88.75</td>
</tr>
<tr>
<td>Standard D</td>
<td>10.82</td>
<td>12.84</td>
<td>14.25</td>
<td>3.61</td>
<td>4.20</td>
<td>7.22</td>
<td>10.42</td>
<td>8.74</td>
<td>9.84</td>
</tr>
</tbody>
</table>

Finally, the average value measured in which each algorithm improves or worsens the illumination homogeneity in regards to the original was quantified. The average value of the brightness deviation for each algorithm was calculated for all images and it was compared with the one obtained for the image without any processing. The algorithm showing better performance is Algorithm 3, which increased the homogeneity value in 78.74% regarding the original value, followed by Algorithm 4.
with a 77.86%, Algorithm 8, 7, 5 and finally Algorithm 6. Algorithm 1 almost did not change and Algorithm 2 worsened the value 23%. These results are shown in Figure 2.

![Figure 2. Average value of improvement or worsening of illumination homogeneity regarding the original.](image)

Figure 3 compares an image and the result of applying the algorithm 3, the one with best performance for this case.

![Figure 3. An image of the sample with light distribution: at the top shows the image without any processing and at the bottom the one processed with Algorithm 3](image)

3.2. Contrast evaluation.
In order to analyze the change in contrast induced for each algorithm, a similar analysis was performed. In this case the 100 standard deviation data was used, which gave an idea of the image contrast.
Figure 4 shows an example of algorithms 5 and 8 applied to one image of the database. The figure shows the increasing or decreasing of the contrast of the processed image with respect to the original one.

Figure 4. The original deviation in blue and the deviation of algorithms 5 and 8 in red, as example, for an image. The blue and red lines correspond to the average values of the deviations of the original image and the processed one respectively.

Then the average value of the ROI deviations for all original and processed images was calculated and the number of images in which the algorithms decreased and increased the contrast with respect to the original was calculated. Algorithm 3 and 5 always increased the contrast.

Finally the average value of deviations was compared with the average of the original images, in order to obtain the percentage of general contrast enhancement. This can be seen in Figure 5, which also shows that Algorithm 5 is the one that most increases the contrast, followed by Algorithm 4 and 3. Figure 6 shows an example of contrast increasing in an image of the sample.

Figure 5. Increasing percentage of contrast obtained by each algorithm, regarding the average contrast of the original images.
3.3. Color preservation evaluation.

With the purpose of assessing if the chromatic features between original and processed images were kept, the R, G and B channels from both the original and the processed images were obtained.

A color normalization of each channel pixel by pixel is performed for all images as follows (for the red channel $R(x, y)$):

$$ r(x, y) = \frac{R(x, y)}{D(x, y)} $$

with

$$ D(x, y) = R(x, y) + G(y, y) + B(x, y) $$

Finally, with these normalized values the Euclidean distance between the processed channel and the original one for each color is calculated, giving a general estimation about the preservation of the color ratio. In order to obtain a single general value for each algorithm, the results are averaged in the three channels:

$$ P_{\text{dist}} = \frac{R_{\text{dist}} + G_{\text{dist}} + B_{\text{dist}}}{3} $$

The results of each algorithm can be seen in Figure 7. The smaller the distance, the more the color characteristics between the original and the processed image are kept. It can be noticed that the algorithm that best preserves color is 8 (multiscale Retinex), while the worst is 7 (iterative Retinex).
3.4. General Performance.

The results obtained in illumination homogenization, in contrast and in color preservation were normalized between 0 and 1. For illumination and contrast it was performed as follows:

\[ V_N = \left( \frac{V - V_{\text{min}}}{V_{\text{max}} - V_{\text{min}}} \right) \]

A value of 1 represents the best performance and 0 represents the worst. In color preservation the best result is the smaller, so the scores are calculated as follows:

\[ V_N = 1 - \left( \frac{V - V_{\text{min}}}{V_{\text{max}} - V_{\text{min}}} \right) \]

As shown in Table 2 and Figure 8, Algorithm 4 obtains the highest value, together with Algorithms 3 and 5. Then, with a slightly lower value are placed Algorithms 6, 8 and 1. Finally, the one of lowest score is Algorithm 2. It should be mentioned that the algorithms with better scores were proposed by the authors, that is to say, those containing the major modifications to standard processes known.

<table>
<thead>
<tr>
<th>Table 2. Performance Scores for each algorithm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alg1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Illum. Homogeneity</td>
</tr>
<tr>
<td>Contrast</td>
</tr>
<tr>
<td>Color Preservation</td>
</tr>
<tr>
<td><strong>Final Score</strong></td>
</tr>
</tbody>
</table>

3.5. Expert opinion.

The proposed algorithms were pre-selected based on expert opinion. For this, a series of processed images was presented to an expert and he was asked to choose the one that he considered as the best for diagnosis. Based on this classification, algorithms 1, 3 and 6 were chose.

Another expert was provided with a graphic interface so as to evaluate which image was of his preference (original or processed) for diagnosis and which structure he considered as the best in the processed image: optic nerve, macula, vessels, none or another (to be specified by him). As a first result, in the 95.56% of the cases (86 of 90 images), the ophthalmologist chose the processed image instead of the original.
Regarding the performance of the algorithms, Algorithm 3 is the best in the opinion of the expert. Among the 30 images processed with this algorithm, 29 were chosen, generally standing out the three main features of the images (optic nerve, macula and arterial tree). Algorithms 1 and 6 also show good results in improving visualization of retinal vessels. In contrast, regarding optic disc visualization improvement, these algorithms were chosen a 50% of the cases.

The expert remarked the importance of maintaining the optic disc original color, as it is relevant while making a diagnosis and patient monitoring, as well as the characteristic orange color of the arteries compared to wine-red veins, because facilitates differentiation.

\[\text{Figure 8. An original image and processed images by the algorithms in order of performance.}\]

4. Conclusions and discussion

The proposed algorithms get different scores when evaluated for improvement of homogeneity of illumination, contrast enhancement or color preservation. In order to choose the one that produces the best result in general improvement of the image (which includes the three properties mentioned), Algorithm 4 is the one that achieves the best score. In turn, algorithms 3 and 5 get scores very similar to each other (2.46 and 2.45) and very close to the first (2.50). Figure 8 shows the results of the algorithms applied to the sample.

It should be noted that Algorithms 3 to 6 which are supported by the model where the image is the product of the illumination and reflectance, do not use standard methods, but combine results of high
pass and low-pass filters with different weights given by the authors. The aim was to emphasize the reflectance component and attenuate the illumination.

When asked to choose the best algorithm for diagnoses, the expert did not choose the best qualified algorithms according to our score method. This supports the hypothesis that not necessarily the statistically best image is the best for diagnosis as evaluated by the experts. This means that the best pre-processing algorithm for an automatic classification may be different to the most suitable one for visual diagnosis. However, both should result in the same final diagnosis. To conclude, it is considered that the Algorithm 3 is the one that provides the best pre-processing performance obtaining a high overall score in the comparative scale. Hence, it is considered appropriated for the pre-processing of images that must be subjected to unsupervised processing. Besides, the experts considered the images pre-processed with Algorithm 3 as useful, indicating that it provides appropriate enhancement of the image and preserves the usual conditions expected by the expert for visual diagnosis.

References