Retinal Image Quality Assessment by Mean-Subtracted Contrast-Normalized Coefficients

Adrian Galdran, Teresa Araújo, Ana Maria Mendonça, and Aurélio Campilho

Abstract The automatic assessment of visual quality on images of the eye fundus is an important task in retinal image analysis. A novel quality assessment technique is proposed in this paper. We propose to compute Mean-Subtracted Contrast-Normalized (MSCN) coefficients on local spatial neighborhoods of a given image and analyze their distribution. It is known that for natural images, such distribution behaves normally, while distortions of different kinds perturb this regularity. The combination of MSCN coefficients with a simple measure of local contrast allows us to design a simple but effective retinal image quality assessment algorithm that successfully discriminates between good and low-quality images, while delivering a meaningful quality score. The proposed technique is validated on a recent database of quality-labeled retinal images, obtaining results aligned with state-of-the-art approaches at a low computational cost.

1 Introduction

Objective quality evaluation of eye fundus images plays a critical role in the field of retinal image analysis. It has been shown that between 10% and 20% of the

Adrian Galdran
INESC TEC Porto, Portugal, e-mail: adrian.galdran@inesctec.pt

Teresa Araújo
INESC TEC Porto, Portugal, and Faculdade de Engenharia, Universidade do Porto, Portugal, e-mail: tfaraujo@inesctec.pt

Ana Maria Mendonça
INESC TEC Porto, Portugal, and Faculdade de Engenharia, Universidade do Porto, Portugal, e-mail: amendon@fe.up.pt

Aurélio Campilho
INESC TEC Porto, Portugal, and Faculdade de Engenharia, Universidade do Porto, Portugal, e-mail: campilho@fe.up.pt

acquired images in large-scale screening Diabetic Retinopathy programs can be of insufficient quality for clinical examination [3]. With the increasing availability of portable eye fundus cameras [15], these figures could easily increase.

The goal of automatic retinal image quality assessment algorithms is to model the subjective quality evaluation performed by experienced ophthalmologists. In this context, specialized humans observers grade the quality of fundus images in comparison with their notion of what represents a good quality image. However, this task can easily become tedious and error-prone, considering the amount of retinal images an ophthalmologist needs to review. Thus, the design of an automatic system for quality assessment capable of modeling this knowledge and emulate the decision of professional ophthalmologists is a relevant computer vision challenge.

Several approaches have been proposed in the past to automatically assess the quality of images of the fundus of the eye [1]. They can be broadly divided into structural and generic methods. Structural methods rely on the assumption that a low-quality retinal image will present distorted anatomical landmarks, e.g. low visibility of the vascular trees. These techniques usually proceed in following steps: first the relevant structures are located; then, some measure of quality is computed locally on these structures. Examples of this approach are presented in [3] and [10]. Generic retinal image quality assessment algorithms ignore the location step, and attempt to model the quality of retinal images by adapting visual features that are known to characterize correctly generic image quality, e.g. [11]. Hybrid methods that attempt to combine both approaches have also been proposed [14].

In this work, we propose to apply Mean-Subtracted Contrast-Normalized (MSCN) coefficients of retinal image intensities to objectively measure retinal image quality. MSCN coefficients are grounded on a well-established theory of human vision science [5], and have long been considered useful to assess image quality of natural scene [12]. Specifically, the statistical distribution of these coefficients has been observed to be highly regular for undistorted generic natural images [9]. However, when an images is corrupted by certain type of degradations, this regularity is modified. Interestingly, different classes of degradation impact the distribution of MSCN coefficients in different ways. This information can be exploited to identify image degradation of different natures, and allows to easily characterize the class of undistorted images with respect to corrupted samples.

Interestingly, MSCN coefficients have been employed before in relationship with retinal fundus imaging [4]. However, in [4], regularity of MSCN coefficients is imposed on degraded images in order to improve their quality. In this paper, we propose to apply those coefficients not to improve the quality of retinal images, but to measure it. Thus, the main contributions of this work are the following:

- We provide an exploratory analysis on the ability of MSCN coefficients to discriminate between low and good quality retinal images.
- Based on our finding, we develop a simple retinal image quality scoring system to assess the discriminative capability of MSCN coefficients.
- The proposed technique is validated on a publicly available database of retinal images, yielding promising results in terms of accuracy and other performance metrics.

The remaining of this paper is organized as follows. First, we give a formal definition of the MSCN coefficients. Then, we apply these coefficients in a series of controlled cases of retinal image degradation, demonstrating that the regularity of their statistical distribution is influenced in a predictable manner by different kind of perturbations. In the following section we build a straightforward retinal image quality scoring system based on the MSCN coefficients and a supplementary simple measure of local image contrast. Then, we validate our approach by means of experimental results on a public database of expert quality-labelled retinal images. To finalize, we present a discussion of potential extensions of the presented technique.

2 Mean-Subtracted Contrast-Normalized Image Coefficients

It is well-known that natural images possess some regularity in terms of statistical distribution of their intensity [5]. This regularity is modified under the presence of image degradations. The ability to measure such deviations allows to build image quality assessment techniques able to identify these degradations.

A straightforward mechanism to quantify local irregularities in image intensity statistics is through the analysis of the distribution of Mean-Subtracted Contrast-Normalized Coefficients on a neighborhood of a given pixel $\mathbf{x}$ of an image $I(\mathbf{x})$. These coefficients are computed via local mean subtraction and contrast divisive normalization as follows:

$$MSCN(I)(\mathbf{x}) = \frac{I(\mathbf{x}) - \mu(\mathbf{x})}{\sigma(\mathbf{x}) + \varepsilon}, \tag{1}$$

where $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are local estimates of the mean and the variance of the image intensities in a neighborhood of $\mathbf{x}$, and ε is a small constant to avoid division by zero. Following [2], we compute these local estimates after weighting the importance of surrounding pixels with a circular-symmetric Gaussian function ω sampled out to e standard deviations and rescaled to unit volume:

$$\mu(\mathbf{x}) = \sum_{\mathbf{y}} \omega(\mathbf{x},\mathbf{y}) I(\mathbf{y}), \tag{2}$$

$$\sigma(\mathbf{x}) = \sum_{\mathbf{y}} \omega(\mathbf{x},\mathbf{y}) (I(\mathbf{y}) - \mu(\mathbf{x}))^2. \tag{3}$$

For undegraded images, Ruderman *et al.* [12] first observed that MSCN coefficients tend to follow a normal distribution, while for distorted images coefficients in Equation (1) exhibit different local distributions, which allows to quantify the degree of distortion. An easy way to measure the deviation of MSCN coefficients from a normal distribution is through the estimation of their variance. Below we provide an exploratory analysis on the behavior of the variance of MSCN coefficients in the context of retinal fundus images and the typical degradations that affect them, such as blur, uneven illumination, or low contrast.

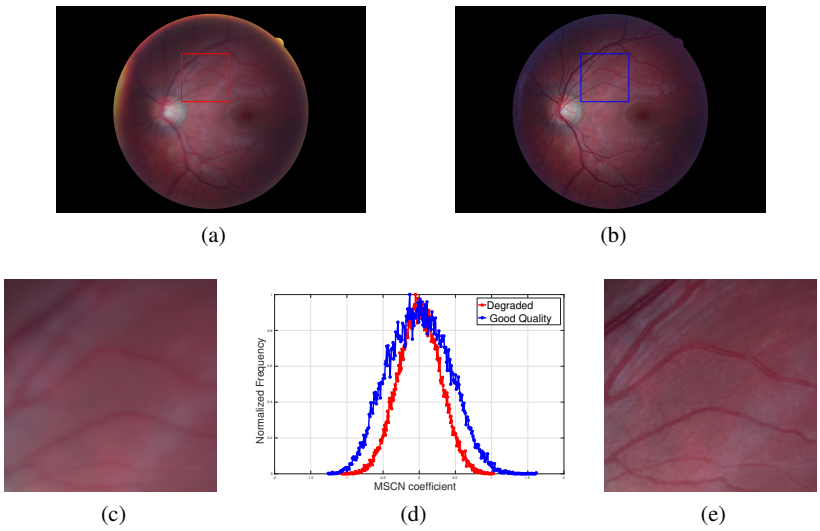


Fig. 1 MSCN coefficients computed on a local neighborhood for a distorted (blurred), (1a), and a good quality image, (1b), from the HRF database. Figures (1c) and (1e) show examples of patches extracted from these images, respectively. The MSCN coefficients of each of these patches are depicted in (1d). In the case of the degraded image, the MSCN coefficients (in red) show a variance of 0.372, while the MSCN coefficients of the undegraded image (in blue) have a variance of 0.695.

2.1 MSCN for Retinal Image Quality Assessment

To illustrate the ability of MSCN coefficient distributions to locally discriminate between good and low-quality visual content in the setting of retinal images, we employ the HRF database [7]. This public database consists of pairs of images from the same eye, captured with a Canon CR-1 fundus camera with a 45° field of view (FOV). For each image pair, there is a poor-quality and a corresponding good-quality image, achieved by repeting the acquisition. The HRF is an excellent test set to evaluate MSCN coefficient distributions, since both images share the same FOV, showing only small displacements due to eye movements between the acquisitions.

Low-quality images from the HRF database suffer from a loss of sharpness due to camera defocus at the acquisition time. This produces a noticeable blur and contrast loss, which is not present on their good-quality counterparts. Figure (1b) shows an example of a good-quality image in the database. From this image, we randomly extract a patch, displayed in Figure (1e), and compute the MSCN coefficients on it according to Equation (1). It can be appreciated how MSCN coefficients, shown in blue in Figure (1d), approximately follow a normal distribution.

In Figure (1a), we show the defocused version of the initial image in Figure (1b). The contrast loss can be clearly noted in the local patch shown in Figure (1c), which has been extracted from the same spatial coordinates as Figure (1e). Local MSCN coefficient distribution is shown in red for this image patch in Figure (1d). In the case

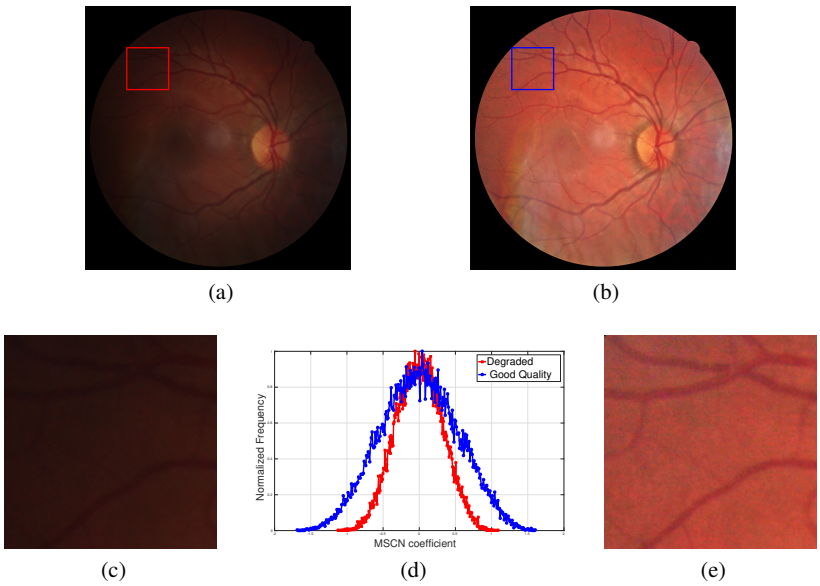


Fig. 2 MSCN coefficients computed on a local neighborhood for a wrongly illuminated, (2a), and a good quality image, (2b), from the DRILL database. The well-illuminated image is the result of processing image (2a) with the method in [13]. Figures (2c) and (2e) show examples of patches extracted from these images, respectively. The MSCN coefficients of each of these patches are depicted in (2d). In the case of the degraded image, the MSCN coefficients (in red) show a variance of 0.413, while the MSCN coefficients of the undegraded image (in blue) have a variance of 0.911.

of the blurred image patch, we can observe how its MSCN distribution contains a decreased variance, due to the underlying blur.

To verify the applicability of MSCN coefficients to characterize other distortions different than defocusing, we show a further example in Figure (2a), in which a retinal image affected by a lack of contrast due to poor illumination is displayed. The image was extracted from the DRIIL (Digital Retinal Images for ILlumination correction) database [8]. We show in Figure (2b) a corrected version of the same image, applying the method in [13]. In this case, we can also appreciate, by observing Figure (2d), how the distribution of MSCN coefficients on a randomly extracted patch ((Figure (2c)), is markedly different than the distribution corresponding to a patch extracted from the same location on the better-quality image (Figure (2e)).

3 MSCN-based Retinal Image Quality Scoring System

To experimentally evaluate the applicability of MSCN coefficients to discriminate bad and good-quality images, we have built a simple quality scoring system for reti-

nal images. Our technique is based on the local computation of MSCN coefficients, coupled with a local measure of contrast, image entropy at a patch level.

The first step of the method consists of partitioning the image into non-overlapping $M \times M$ pixel patches, herein denoted by $\mathscr{P}$. For each of these patches, we compute the variance of the distribution of MSCN coefficients and the local entropy:

$$\varepsilon(\mathscr{P}) = -\sum_{x \in \mathscr{P}} p(I(x)) \log(p(I(x)), \tag{4}$$

where $p(I(x))$ denotes the probability of occurrence of the pixel intensity in the image $I(x)$, which can be estimated from a normalized histogram of patch intensities.

Every patch extracted from a good-quality image is labeled as good-quality, and conversely every patch from a low-quality image is considered as low-quality. Therefore, from a database of quality-labeled images, we extract two metrics from each patch $\mathscr{P}(x)$. For the MSCN coefficients, we apply formula (1) and compute the variance of the resulting coefficients from $\mathscr{P}$, while for the local entropy each patch has a single associated value.

3.1 Quality Classification

After computing this set of feature pairs, denoted $\mathbf{F} = (f_1, f_2)$, on every training image, we characterize the good and low-quality classes by a simple multivariate Gaussian model, i.e., we compute the mean and covariance of $\mathbf{F}_{good}$ and $\mathbf{F}_{low}$, which we denote $(\mu_{good}, \Sigma_{good})$ and $(\mu_{low}, \Sigma_{low})$ respectively.

This simple model can be applied for quality prediction on a retinal test image I^{test} by extracting the mean of the same pair of features from local patches:

$$\mathbf{F}^{test} = \left\{ (f_1^{test}(\mathscr{P}), f_2^{test}(\mathscr{P})), \quad \forall \mathscr{P} \in I^{test} \right\}. \tag{5}$$

We then estimate the bi-dimensional mean from both sets of values:

$$\mu^{test} = \frac{1}{p} \sum_{i=1}^{p} \mathbf{F}^{test}, \tag{6}$$

where p is the number of elements in $\mathbf{F}^{test}$, and finally compute the Mahalanobis distance from μ^{test} to μ_{good} and μ_{low}:

$$d_{good} = \frac{1}{2\pi |\Sigma_{good}|^{1/2}} \exp\left[\frac{1}{2} (\mu^{test} - \mu_{good})^t \cdot \Sigma_{good}^{-1} \cdot (\mu^{test} - \mu_{good})) \right], \tag{7}$$

$$d_{low} = \frac{1}{2\pi |\Sigma_{low}|^{1/2}} \exp\left[\frac{1}{2} (\mu^{test} - \mu_{low})^t \cdot \Sigma_{low}^{-1} \cdot (\mu^{test} - \mu_{low})) \right]. \tag{8}$$

After this process, I^{test} is assigned to one of both classes according to which class μ^{test} is closest to: I^{test} is predicted as good-quality if $d_{good} < d_{low}$ and vice-versa.

3.2 Quality Scoring

To also estimate a continuous quality score for a given test image, we compare both distances by considering the above procedure as a two-class Bayesian classification problem, where the good-quality class is denoted by **G**, and the low-quality class as **L**. Hence, we can define a quality score based on the posterior probability of belonging to **G**. We can then use Bayes' formula to re-write that probability in terms of the likelihood of x belonging to **G**, **L**, and the prior probabilities of both classes:

$$P(\mathbf{G} \mid x) = \frac{P(x \mid \mathbf{G}) \cdot P(\mathbf{G})}{P(x \mid \mathbf{G}) \cdot P(\mathbf{G}) + P(x \mid \mathbf{L}) \cdot P(\mathbf{L})}. \tag{9}$$

If we model the likelihood of a sample x being in **G** by its distance to **L**, and we assume that the prior probabilities of an image being in **G** or **L** are the same, eq. (9) reduces to dividing the distance of x to **L** by the sum of both distances:

$$s(x, d_{low}, d_{good}) = \frac{d_{low}}{d_{low} + d_{good}} \tag{10}$$

The probabilistic formulation of the score function in eq. (10) enjoys several desirable properties. First, for samples with pairs of distances such that $d_{low} = d_{good}$, the score always takes the value $s = 1/2$. It is also independent of the distance magnitudes, i.e.:

$$s(x_1, \alpha \cdot d_{low}, \alpha \cdot d_{good}) = \alpha \cdot s(x_2, d_{low}, d_{good}), \tag{11}$$

for $\alpha > 0$. Finally, it allows us to perform ROC analysis on the resulting predictions, see Section 4 below.

However, the output of the above score is not well-calibrated. Calibration of this kind of procedures can be achieved with specialized techniques, but it is beyond the scope of this work. In our case, we empirically observed that formula (10) requires a large difference in distances to generate predictions close to 1 or 0. In practice, the alternative definition below retrieves more evenly distributed scores while preserving all the desirable outlined features:

$$\widetilde{s}(x, d_{low}, d_{good}) = \frac{d_{low}^2}{d_{low}^2 + d_{good}^2}. \tag{12}$$

In all our experiments we employ formula (12).

4 Experimental Results

To validate our approach, we conducted a simple experiment on the DRIMDB database [14]. Images in DRIMDB were acquired with a Canon CF-60UVi Fundus

Table 1 Average Model Performance for a patch size of 8×8.

Performance	AUC	Accuracy[1]	Sensitivity[1]	Specificity[1]	F-score[1]
Our method	0.997	0.984	0.995	0.971	0.980
[14] (good vs low+outliers)	-	-	-	-	0.996
[14] (low vs good+outliers)	-	-	-	-	0.965
[6]	0.998	-	-	-	
[15]*	-	-	0.780	0.850	0.900

[1] Computed for our method at an optimal threshold $t = 0.520$ maximizing F-score.
* In the original work, this method was trained on a different dataset, and tested on DRIMDB.

Camera using 60° FOV and stored at 570×760 pixels resolution. DRIMDB contains three quality classes: good, bad, and outlier. In this work we focus on the correct discrimination of the good/bad classes. For the good quality class, the database contains 125 examples, while for the bad quality class it contains 69 examples.

To assess the performance of the proposed technique, we randomly divided the database into two disjoint subsets. We learned the model in Equation (6) using a patch size of 8×8 pixel on our training set, and estimated its performance in the hold out set. Afterwards, we re-train the same model but interchanging the roles of the train and test set, and averaging the performance on both test sets.

Apart of computing raw accuracy in terms of well-classified images, we take advantage of the continuous distribution of the quality score defined in Equation (10) to perform ROC analysis of our method. In Table 1 we show the performance of the method in terms of accuracy, specificity, sensitivity, and F score, for the optimal operating point. It can be seen that the proposed technique correctly classified every bad-quality image in all cases, while some good-quality images were misclassified. Providing a fair comparison to other recent methods is a hard task, for two reasons. First, while here we focus on the limited task of discriminating between good and bad quality images, other techniques were tested on the three-class problem, considering also outlier images, which is beyond the scope of this work. Second, the DRIMDB database is relatively recent and few methods have been tested on it.

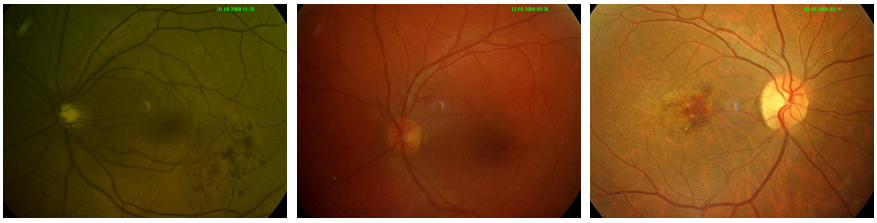
For reference, we provide also in Table 1 the results of the methods in [14, 6, 15]. It can be seen that almost all techniques also provided a high performance, which could be due to the severe degradations present in the low-quality images of DRIMDB. This may also be the reason why our method shows a slightly higher performance in detecting low-quality images, with a sensitivity of 0.992 and a specificity of 0.971. It should also be noted that the complexity of the other considered approaches is substantially higher than the technique proposed in this paper.

A relevant parameter of our technique is the considered patch size. To test its robustness, we re-trained model (6) at a varying patch size ranging from 8×8 to 128×128. This influences two aspects of the proposed technique: the training speed and the performance of the method. In Table 2 we show the evolution of training time and performance in terms of AUC as a function of the patch size. Reported times include the training and prediction stages. The proposed model is implemented in Matlab, and executed in a Intel® Xeon® CPU (2.40GHz). We see

Table 2 Training time and AUC as a function of the considered patch size.

Patch Size	8×8	16×16	24×24	32×32	40×40	48×48	96×96	128×128
AUC	0.997	0.996	0.995	0.994	0.993	0.993	0.988	0.986
Exec. Time [1]	119.96	35.66	20.08	14.49	11.56	10.20	7.44	7.00

[1] Execution times are given in seconds and include training and testing processes.



(a) (b) (c)

Fig. 3 Examples of scores (s) produced by our method. (a) $s = 0.408$, (b) $s = 0.515$, (c) $s = 0.919$.

that a reduced spatial neighborhood increases model performance, although its general behavior is satisfying at all resolutions. The computational cost of extracting features from a larger quantity of patches is considerable. However, the simplicity of the technique is an important advantage. Training the model and testing it on the separate test set, even at the finer patch resolution, takes approximately 2 minutes.

Finally, to verify if the score in formula (12) is consistent with human perception, we show in Figures 3a and 3b the good-quality images that were assigned the lowest scores. Although the model assigned them medium scores, a certain degree of uneven spatially-variying illumination affects these images. We also show the image with the highest assigned score in Figure (3c). In this case, the image contains more visible vessels and has good contrast, which qualitatively validates our approach.

5 Conclusions

We have investigated the application of MSCN image coefficients for objectively measuring retinal image quality. Good-quality images produce MSCN coefficients with approximately normal distribution, while MSCN coefficients extracted from low-quality images generate distributions with a reduced variance. With the help of MSCN coefficients and an auxiliary simple feature, local entropy, we have proposed a retinal image quality assessment system providing both a quality label and a meaningful quality score. Results are in line with the reported for other recent techniques, validating our approach for retinal image quality assessment. Future work will involve testing our model on high resolution images and studying the impact of different regions (e.g. optic disc, background) on statistics computed from eq. (1).

Acknowledgements This work is financed by the ERDF European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme, and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia within project CMUP-ERI/TIC/0028/2014. Teresa Araújo is funded by the FCT grant contract SFRH/BD/122365/2016.

References

1. Abdel-Hamid, L., El-Rafei, A., El-Ramly, S., Michelson, G., Hornegger, J.: Retinal image quality assessment based on image clarity and content. Journal of Biomedical Optics **21**(9) (2016)
2. Choi, L.K., You, J., Bovik, A.C.: Referenceless Prediction of Perceptual Fog Density and Perceptual Image Defogging. IEEE Transactions on Image Processing **24**(11), 3888–3901 (2015). DOI 10.1109/TIP.2015.2456502
3. Fleming, A.D., Philip, S., Goatman, K.A., Olson, J.A., Sharp, P.F.: Automated Assessment of Diabetic Retinal Image Quality Based on Clarity and Field Definition. Investigative Opthalmology & Visual Science **47**(3), 1120 (2006)
4. Foracchia, M., Grisan, E., Ruggeri, A.: Luminosity and contrast normalization in retinal images. Medical Image Analysis **9**(3), 179–190 (2005)
5. Geisler, W.S.: Visual Perception and the Statistical Properties of Natural Scenes. Annual Review of Psychology **59**(1), 167–192 (2008)
6. Hamid, L.A., El-Rafei, A., El-Ramly, S., Michelson, G., Hornegger, J.: No-reference wavelet based Retinal Image Quality Assessment. In: Computational Vision and Medical Image Processing V: Proceedings of the 5th Eccomas Thematic Conference on Computational Vision and Medical Image Processing (VipIMAGE), Spain, 2015), p. 123. CRC Press (2015)
7. Kohler, T., Budai, A., Kraus, M.F., Odstrilik, J., Michelson, G., Hornegger, J.: Automatic no-reference quality assessment for retinal fundus images using vessel segmentation. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, pp. 95–100 (2013)
8. Kolar, R., Odstrcilik, J., Jan, J., Harabis, V.: Illumination correction and contrast equalization in colour fundus images. In: European Signal Processing Conference, pp. 298–302 (2011)
9. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-Reference Image Quality Assessment in the Spatial Domain. IEEE Transactions on Image Processing **21**(12), 4695–4708 (2012)
10. Niemeijer, M., Abrmoff, M.D., Ginneken, B.v.: Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. Medical Image Analysis **10**(6), 888–898 (2006)
11. Pires Dias, J.M., Oliveira, C.M., da Silva Cruz, L.A.: Retinal image quality assessment using generic image quality indicators. Information Fusion **19**, 73–90 (2014)
12. Ruderman, D., Bialek, W.: Statistics of natural images: Scaling in the woods. Physical Review Letters **73**(6), 814–817 (1994)
13. Savelli, B., Bria, A., Marrocco, C., Molinara, M., Tortorella, F., Galdran, A., Campilho, A.: Illumination Correction by Dehazing for Retinal Vessel Segmentation. In: the 30th IEEE International Symposium on Computer-Based Medical Systems (CBMS) (2017)
14. Sevik, U., Kose, C., Berber, T., Erdol, H.: Identification of suitable fundus images using automated quality assessment methods. Journal of Biomedical Optics **19**(4) (2014)
15. Wang, S., Jin, K., Lu, H., Cheng, C., Ye, J., Qian, D.: Human Visual System-Based Fundus Image Quality Assessment of Portable Fundus Camera Photographs. IEEE Transactions on Medical Imaging **35**(4), 1046–1055 (2016)