Retinal image quality assessment through a visual similarity index
Jorge Pérez, Julián Espinosa*, Carmen Vázquez and David Mas

Instituto de Física Aplicada a las Ciencias y la Tecnologías, Universidad de Alicante, Carretera San Vicente del Raspeig s/n. 03690, San Vicente del Raspeig (Alicante), Spain

(Received 7 March 2013; final version received 4 April 2013)

Retinal image quality is commonly analyzed through parameters inherited from instrumental optics. These parameters are defined for ‘good optics’ so they are hard to translate into visual quality metrics. Instead of using point or artificial functions, we propose a quality index that takes into account properties of natural images. These images usually show strong local correlations that help to interpret the image. Our aim is to derive an objective index that quantifies the quality of vision by taking into account the local structure of the scene, instead of focusing on a particular aberration. As we show, this index highly correlates with visual acuity and allows inter-comparison of natural images around the retina. The usefulness of the index is proven through the analysis of real eyes before and after undergoing corneal surgery, which usually are hard to analyze with standard metrics.

Keywords: vision model; retinal image quality; structural similarity index

1. Introduction

The objective determination of retinal image quality and its correlation with human perception has been an important topic over the past two decades. The retinal image is affected by aberrations, scattering, and diffraction; so optical properties of the eye are the most important physical limits to vision. It has been investigated by using both objective and psychophysical subjective techniques [1–6] and by analyzing the separated contribution of the different components of the optical part of the human visual system (HVS) [7–9]. The interest in the measurement of image quality has favored the development of different quality parameters, all of them determined from the point spread function (PSF), the modulation transfer function (MTF), or the wave aberration, \( W(x, y) \) [10]. Although these quantities strongly correlate with each other if the studied eyes show small aberrations [11], this correspondence is lost for highly aberrated systems [12].

Image quality prediction under image degradation is not an exclusive issue of visual optics. Other fields, like image and video processing, have explored this topic in order to analyze the human tolerance to image degradation due to ‘lossy’ compression, transmission, and other factors [13–15]. There, the image quality is usually determined through opinion tests carried by trained observers [16–18]. Connected with these tests, Wang et al. proposed a mean structural similarity index (MSSIM) [19], which measures the similarity between one image and its undistorted version by considering local correlations and dependences.

Natural images generally show a strong dependence between one point and its local neighborhood. This dependence carries important information about the inner structure of the scene and helps the HVS to perform the recognition and interpretation of the objects in it. We make the hypothesis that the visual quality of the HVS (usually measured through the decimal visual acuity, VA) can be evaluated through metrics that take into account the structure of the scenes. Thus, we adapt the MSSIM to the visual system and derive a visual mean structural similarity index (VMSSIM). This index takes into account both the imaging system and the structure of the scene and is correlated with quality variation in the retinal image and the VA. The novelty of our proposal is the translation of a metric typically used in the field of video and image processing to the field of the visual optics.

The manuscript is structured as follows. In Sections 2 and 3, we present the procedure followed to adapt and apply the metrics and the used subjects, respectively. Next, in Section 4, we demonstrate the performance of the metric on real eyes. In Section 5, we compare results between subjective visual acuity and objective values of the VMSSIM. Finally, we detail the main conclusions reached in this work.

*Corresponding author. Email: julian.espinosa@ua.es

© 2013 Taylor & Francis
2. Metrics and method

The MSSIM proposed by Wang et al. [19] is defined by the product of three factors: correlation between images, luminance distortion, and contrast distortion. They try to emulate the HVS’s ability to extract structural information from natural scenes in order to analyze and compare digital images. The metric is defined on local kernels that move over the reference and the target images, X and Y, respectively, as:

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^{M} \left( \frac{2\mu_{ij}(y_j) + C_1}{\mu_{ij}^2(y_j) + \mu_{ij}^2(y_j) + C_1} \times \frac{2\sigma_{ij}(y_j) + C_2}{\sigma_{ij}^2(y_j) + \sigma_{ij}^2(y_j) + C_2} \right),$$

(1)

where $M$ is the total number of fitting windows, $\mu(y_j)$ and $\mu(y_j)$ stand for mean values, $\sigma(x_i)$ and $\sigma(y_j)$ represent the standard deviations and $\sigma(x_i, y_j)$ is the covariance between the corresponding local window of the $X$ and $Y$ images, respectively. The higher the distortions of the target image $Y$ compared to the reference $X$, the more different the MSSIM from the unit. Constants in the formula are added to avoid the denominator becoming equal to zero and are defined according to the dynamic range of the image $L$ so that $C_2 = C_3 = (0.03L)^2$ and $C_1 = (0.01L)^2$. In [20], the authors show that Equation (1) can be simplified by dropping the first factor, with no appreciable change in the final numerical result.

We propose to adapt the MSSIM to visual metrics. In the absence of aberrations, the image quality of the image on the retina is only affected by diffraction. An eye free of aberrations with a 4 mm pupil diameter and a wavelength of 555 nm theoretically provides a light spot diameter at the retinal plane of around 6.0 μm, which is equivalent to a maximum decimal visual acuity of 2.4. According to this, we consider a renormalization of the MSSIM so that its maximum value corresponds to the best possible image on the retina and redefine the metrics to obtain the visual mean structural similarity index as:

$$VMSSIM(X, Y) = \frac{2}{4} \frac{\sum_{j=1}^{M} \left( \frac{2\sigma_{ij}(y_j) + C_2}{\sigma_{ij}^2(y_j) + \sigma_{ij}^2(y_j) + C_2} \right)}{\sum_{j=1}^{M} \left( \frac{2\sigma(x_i, y_j) + C_2}{\sigma(x_i)^2 + \sigma(y_j)^2 + C_2} \right)},$$

(2)

where the super index $N$ stands for the initial scene convolved with the diffraction mask. In our application, the reference image is a known and recognizable scene; and the distorted image is that obtained on the retina. Accurate calculation of the retinal image is described in [21–25]. Briefly, it starts with the determination of the corneal topographic map, the corneal thickness, and the size of the anterior chamber using a Pentacam system [26] and an IOL Master. The effect of crystalline lens is calculated according to the Kooijman’s model [27]. From these data, one can obtain the optical path for a bundle of rays entering the eye and arriving to the plane immediately posterior to the lens and use it to construct the emerging wavefront, $W_0(x, y)$, which is finally propagated through Fresnel algorithms up to the plane of interest inside the eye.

The main drawback of the above process is the use of a lens model, which usually does not correspond to a real one. Therefore, the objective refraction, which is computed from the Zernike polynomial expansion of the wavefront function, $W_0(x, y)$ [28], may not match up with the subjectively measured refraction. What we propose to overcome this problem is to modify the wavefront by changing the Zernike coefficients related to astigmatism and defocus in order to fit the subjective test. The differences between measured and calculated refraction are assigned to the crystalline lens, which somehow is re-modeled and customized to better suit the real one.

The patterns, $U_i(x, y)$, propagated up to any plane $z$ around the subject’s retina are obtained through the algorithm detailed in [25]. Let us consider the light pattern leaving the lens as:

$$U_0(x, y) = \exp\left[-\frac{p}{2}(x^2 + y^2)\right] \exp\left[-\frac{2\pi}{\lambda}W_0(x, y)\right],$$

(3)

where we have incorporated the Stiles–Crawford effect [29] modeled as an apodizing filter [30] with the parameter $p = 0.12 \text{ mm}^2$ for a pupil of 4 mm diameter. If we remove the convergent factor term $\exp[-i\pi(x^2 + y^2)/(\lambda z)]$ to the pattern in Equation (3), i.e. we add a divergent beam to the phase factor, we are relaxing the Nyquist condition so we can obtain the PSF at different planes around the retinal one as:

$$|U_z(x, y)|^2 = \left|DFT \left\{ U_0(x, y) \exp\left[-\frac{\pi}{\lambda} \left( 1 - \frac{1}{z^2} \right)(x^2 + y^2) \right] \right\} \right|^2.$$

(4)

Using this algorithm, we can accurately compute the PSF on the retina or any other plane inside the eye. The convolution of the ocular PSF with the reference scene provides a good approximation of the effect of the optical surfaces of the eye over the incident light. Next, each distorted image is compared with the original input image and an axial VMSSIM can be evaluated following 

Notice that neither the PSF nor the MTF, its counterpart, are taken as absolute metrics. The image obtained after convolution with the PSF is locally...
analyzed and compared to a realistic reference so that local information is taken into account.

3. Subjects

The performance and utility of the proposed metric is shown through the study of the eyes of two subjects 47 and 51 years old, both suffering from high ametropia. These subjects were surgically treated using the PresbyLasik technique [24]. Note that, the selection of these subjects allows us to analyze both emmetropic (post-surgery) and ametropic (pre-surgery) eyes. Moreover, their quality cannot be adequately assessed through standard metrics. The PresbyLasik technique consists of increasing the spherical aberration in order to extend the focal depth and, therefore, to create an effect on vision known as pseudo-accommodation.

The biometric magnitudes needed to obtain the light patterns inside the eye and the visual acuity for near and far vision (Rosenbaum card at 40 cm and Snellen chart, respectively) are presented in Table 1. We distinguished between considering the ophthalmic correction (near best corrected visual acuity, NBCVA; best corrected visual acuity, BCVA) and without correction (near visual acuity, NVA; visual acuity), and the residual accommodation amplitude (RA).

4. Simulations and results

We have determined the light patterns inside the eye and computed the VMSSIMs defined in Equation (2) for both subjects before and after undergone surgery. We would like to point here that the change in the terms of the expansion of $W_0(x,y)$ in Equation (3) in order to suit the subjective refraction is only performed for the pre-surgery eye. This change is assigned to an inaccurate lens model and it is characterized with a convergent and an astigmatic term. Those terms are introduced without further correction in the lens model when we analyze the post-surgery case.

The VMSSIMs obtained together with the axial profiles of the propagated beams are shown for both subjects in Figures 1 and 2. Each plane is normalized to the maximum in order to improve visualization and facilitate the comparison. The hyperopic subject shows, before the refractive surgery (Figure 1(b)), a spherical equivalent refractive error (SE) of 4.13 D. This value almost coincides with 4.07 D, which corresponds to the defocusing value between the plane of maximum VMSSIM ($z = 23.72$ mm) and the retinal plane ($z = 22.12$ mm). Moreover, the VMSSIM at the retinal plane (0.03) is close to the uncorrected VA (0.05). Regarding near vision, if one takes into account the subjective spherocylindrical correction, the addition and part of the residual accommodation, when the scene is 40 cm away from the subject, the image at the retinal plane is the one corresponding to $z = 23.71$ mm, where the VMSSIM is 1.09. This value is slightly higher than the visual acuity subjectively assigned when the eye is corrected (NBCVA = 1).

Post-surgery results for subject A (Figure 1(c)) are optimum because the correction is 4.24 D. We define the correction as the difference in diopeters between the planes that present a maximum VMSSIM in the pre- and post-surgery states. In this case, the correction is the distance in diopeters between 23.72 mm and 22.05 mm. Post-surgery VA is 1, while the objective VMSSIM value results 1.12, with no need of sphero-cylindrical correction. Regarding the post-surgery NVA and taking into account the residual accommodation, the VMSSIM value, which is computed from the pattern at $z = 21.94$ mm, is 0.84, also quite close to the subjective value. Furthermore, if we consider that the lens is able to provide a residual addition of 1 D, the subject is able to focus the scene on the retina with a SSVI of 1.12, similar to the NBCVA value.

Values for the subject B obtained with the proposed metric also coincide with the VA. Due to the high refractive error of the subject (~6.25 D), the VA before the surgery (Figure 2(b)) is very low (0.05) and also the

| Table 1. Pre- and post-surgical optometric data and biometric data from subjects under study. |
|--------------------------------------------------|------------------|------------------|------------------|------------------|
| Subject A (hyperopic)                             | Subject B (myopic) |
| Pre-surgical                                     | Post-surgical    | Pre-surgical     | Post-surgical    |
| Sphere (D)                                        | 4.50             | -5.50            | 0                |
| Cylinder (D)                                      | 0               | -1.50            | 0                |
| Axis (deg)                                        | 85              | 180              | 0                |
| Addition (D)                                      | 1.00            | 1.50             | 0                |
| VA                                               | 0.05            | 0.05             | 0.9              |
| NVA                                              | 0.05            | 0.1              | 0.8              |
| BCVA                                             | 2.00            | 2.00             | 2.00             |
| NBCVA                                            | 0.604           | 0.542            | 0.497            |
| RA (D)                                            | 2.36            | 22.12            | 25.26            |
| Ax. L. (mm)                                       | 0.599           | 0.599            | 0.497            |
| Central corneal thk. (mm)                         | 25.26           | 25.26            | 25.26            |
| Ant. chamber L. (mm)                              | 2.36            | 3.37             | 3.37             |
VMSSIM, which is even slightly lower (0.02). Note that the maximum value for the metric is 0.66 at $z = 22.81$ mm; which means an objective spherical equivalent refractive error of $-5.7$ D. Such low VMSSIM is due to the high astigmatism of the eye. The measurement of the NVA at 40 cm can be simulated moving the patterns 2.5 D in the direction of the retinal plane. Thus, we admit that the subject would observe an image in near vision without correction that corresponds with the one provided at the plane $z = 24.12$ mm, with a VMSSIM equal to 0.05. This result agrees with the subjective NVA (0.1).

The VMSSIMs obtained for this subject after surgery (Figure 2(c)) are also consistent with the VA. The maximum VMSSIM (0.88) results at the plane $z = 25.21$ mm. The surgery has displaced the best image plane to the retinal plane, around 5.6 D. This means that there is still a low myopia. At the retinal plane, the VMSSIM value (0.84) presents an excellent agreement with the subjective visual acuity (0.9). Regarding near vision, the VMSSIM is similar to subjective NVA.

5. Discussion

The proposed VMSSIM correlates with the VA and seems to allow quantification of the quality of vision, although a statistical analysis involving more subjects must be performed. A metric that determines the retinal image quality should have a monotonous behavior and be simple [31]. We expect the VMSSIM obeys this statement, so we have studied its evolution with the distance to the retinal plane and looked for a fitting function. The fitting function must fulfill a minimum set of physico-mathematical requirements: it must be continuous, positive, tend to zero when defocus tends to a very high value, present no symmetries, and have no vertical asymptotes. All these conditions lead us to choose an asymmetric variant of pseudo-Voigt function. This function is a linear combination of a Lorentian and a Gaussian function as follows:

$$\text{VMSSIM}(D) = A \frac{\omega_L}{\omega_L^2 + (D - D_{0L})^2} + B \exp \left(-\ln2 \frac{(D - D_{0G})^2}{\omega_G^2}\right),$$

where $D$ represents the distance from any plane to the retinal one expressed in diopters, $D_{0L}$ and $D_{0G}$, respectively, are the positions of the maxima of the Lorentzian and the Gaussian functions, and $2\omega_L$ and $2\omega_G$ are the full width at half maximum of each function.
The computed VMSSIM values have been fitted to Equation (5) and are presented, together with the 95% confidence bounds of the fitting in Figure 3. Positive and negative diopters denote anterior and posterior planes to the retinal one whereas 0 diopters stands for the retinal plane.

Figure 2. (a) Axial VMSSIM and propagated beam section inside the eye of subject B, (b) pre- and (c) post-surgery, together with natural images computed at the retinal plane.

Figure 3. VMSSIM (triangles and squares) at different planes from the retinal one in diopters, computed for both subject A (a) and subject B (b). Solid and dashed lines correspond to the fittings to Equation (5) and their confidence bounds, respectively. (The colour version of this figure is included in the online version of the journal.)
The slope of the fitted curves decreases as defocus increases, i.e. far away from the peaks in Figure 3. Holladay et al. [32] measured the subjective VA in subjects of radial keratotomy and experimentally found a similar decay. Results found by Dehnert et al. [33], whom related the VA with a simulated defocus and obtained mean values of 0.33 and 0.18 for a defocus of 1D and 2D, respectively, support those provided by our method. We respectively obtain VMSSIM values of 0.29 ± 0.08 and 0.13 ± 0.08 for the subject A, and 0.39 ± 0.03 and 0.11 ± 0.04 for subject B.

Parameters from the fitting of VMSSIM values together with the residual accommodation allow characterizing and predicting the visual quality. In Figure 4, we present a scheme to summarize the determination of the VMSSIM at the retinal plane for any object distance, \( z_{obj} \). It is divided, depending on the position of the maximum VMSSIM from retinal plane, \( D_0 \), measured in diopters. Moreover, we distinguish between different sections of the object space where total, partial, or null residual accommodation is available.

6. Conclusions

In this paper, we have presented a metric that takes into account the distortion on a natural scene caused by the human optical system and, together with a customized eye model and a light pattern propagation technique, provides similar results to those obtained using tests to measure visual acuity, both for far and near distances.

The emergence in recent years of new metrics based on SSIM (SSIM Gradient, three-component multi-scale SSIM, etc.) [34,35] allows continuation of these studies and improvements in the precision and accuracy of results; however, the complexity of calculations will inevitably increase. Furthermore, as van der Linde and Doe show in [36], the visual quality that a subject assigns to a distorted natural image depends on the affective content of the image. This fact increases the difficulty in the search for algorithms that relate the optical quality of the eye with the visual quality of the subject.

Despite the need of carrying out a statistical study of the posed technique to accurately value its performance, the method and the metric have shown objectively predict the subject’s visual capacity. We expect it to be really useful to better strategize and evaluate corrective optical systems and ophthalmic surgery. Note that, we have clearly separated the optical effects of cornea and crystalline lens so they can be independently modified if the refractive status of the eye changes due to a contact lens, corneal ablation or intraocular lens implant. The method can also be used to design new corneal profiles in order to obtain extended depth of focus and large pseudo-accommodation ranges [37].

Acknowledgements

The authors would like to acknowledge the financial support of the Generalitat Valenciana through the projects PROMETEO/2011/021 and ISIC/2012/013 and the University of Alicante through the project GRE10-09.

References


[35] Li, C.; Bovik, A.C. Proc. SPIE 2009, 7242, 72420Q.
