Search for Image Quality Metrics Suitable for Assessing Images Specially Precompensated for Users with Refractive Errors

Nafe B. Alkzir[1,2], Ilya P. Nikolaev[1], Dmitry P. Nikolaev[1]
[1]Institute for Information Transmission Problems, RAS, Bolshoy Karetny per. 19, build. 1, Moscow, 127051, Russia;
[2]HSE University, 20 Myasnitskaya Ulitsa, Moscow 101000, Russia

ABSTRACT

Recently, we presented the SCA-2023 dataset, which had been developed specifically to evaluate the quality of various image precompensation algorithms for observers with imperfect vision. Such precompensation makes it possible to bring their image perception closer to that of an observer with the ideal vision. While experimenting with various image quality metrics, we realized that it was not so easy to evaluate the quality provided by different algorithms, since the metrics "voted" for different things, and their choice often seemed to contradict the human perception. This is a key motivation for our study, in which we set out to select the metric best correlated with the human perception of precompensated images. We selected a suitable subdataset from our SCA-2023 dataset and, based on it, created 90 grayscale images, which were shown to our colleagues in a pairwise comparison way. More than 2,000 pairwise comparison results were collected from 24 study participants. Further, according to our original methodology, these results were compared with the "opinion" of some popular quality metrics, which made it possible to rank these metrics according to their adequacy within the framework of this task. Finally, we showed how to use these results in optimization procedures aimed at improving the quality of precompensation.

Keywords: image quality metrics, Wiener filtering, image precompensation, ocular aberration, refractive errors

1. INTRODUCTION

Graphical user interfaces (GUI) are widely used in computer programs and applications to present information using visual elements such as images, icons, and menus [1]. However, many users experience difficulty in recognizing graphic elements due to refractive errors of vision, such as nearsightedness, farsightedness, or astigmatism. Such refractive errors can lead to a reduction in the effectiveness of the GUI for these users in cases where they do not use corrective lenses for one reason or another.

One of the modern methods of image processing aimed at solving this problem is image precompensation. The purpose of precompensation is to create a software-transformed image for an observer with non-ideal vision, which makes it possible to bring his perception of the image closer to that of an observer with ideal vision purely by software, without changing the physical characteristics of the image carrier. Today, there are several fundamentally different approaches to solving the problem of image precompensation. The first and most common approach is the Wiener filtering. The idea of the Wiener filtering is to use information about the spectral properties of the distorted image and the spectral characteristics of the distortion source to create the undistorted image. The method is based on applying a filter that optimally (in terms of the signal-to-noise ratio) compensates for distortion, minimizing the difference between the distorted and the restored images [2, 3, 4].

Another approach used in image precompensation problems is gradient optimization of some quality functional. Within this approach, the precompensation problem is formulated as an optimization problem, where the chosen quality functional determines the target image characteristics. This approach usually requires a fairly large number of iterations, so for sufficient flexibility of settings (the ability to choose different quality functionals, including different regularizers), one has to sacrifice the temporal performance (as compared to the Wiener filtering) [5].

Neural network models that have been rapidly developing in the last decade also represent a promising approach to the problem of image precompensation. The use of neural networks potentially allows you to speed up the optimization process without losing the accuracy of restoring the undistorted image. Neural network models can be trained on large datasets and adapted to different types of refractive errors [6]. The main difficulty in this context is that the neural network model must

Sixteenth International Conference on Machine Vision (ICMV 2023), edited by Wolfgang Osten,
Proc. of SPIE Vol. 13072, 130720S · © 2024 SPIE · 0277-786X
doi: 10.1117/12.3023509

be trained simultaneously for a large number of ground truth images and a large number of distorting kernels, which makes the model very “heavy”, not allowing realizing the potential performance advantage.

Obviously, the second and the third approaches require a suitable image quality evaluation metric. In fact, such a metric, in the framework of solving the precompensation problem, is also needed for the Wiener filtering. This is because the Wiener filter, by itself, is not able to solve the precompensation problem, since, for some image pixels, it gives brightness values that go far beyond the dynamic range of the image carrier (including negative ones). Following the Wiener filtering, one has to use a procedure known as tone mapping. The tone mapping parameters need to be optimized and this requires an appropriate image quality metric. In addition, it should be noted that the Wiener filter includes a regularization parameter, which also requires optimization. Thus, today there is no approach to solving the precompensation problem that does not require a suitable image quality metric.

In our previous study SCA-2023 [7] we tested several metrics and found that most of them are not able to adequately assess the quality of precompensation. Fig. 1 shows examples of situations where a precompensated image receives lower scores from various quality metrics (compared to a blurry image without precompensation), but it is subjectively perceived by a person as a higher quality one. The mathematical expressions for the metrics used will be given below in Section 4. For now, it is important to realize that these metrics have to compare images, each of them being much “worse” than the corresponding ground truth. This makes the discussed problem somewhat similar to what happens in medical imaging, where raw images are way far from what diagnosticians would like [8].

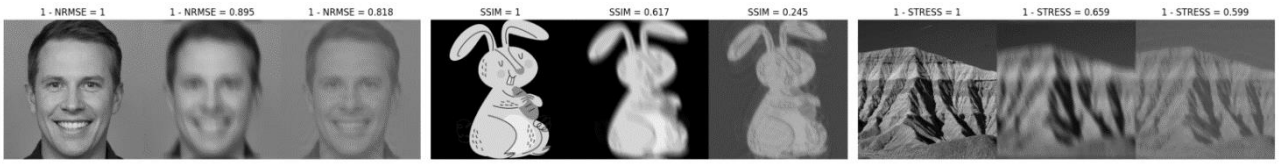

Figure 1. An example of comparing images according to three well-known quality metrics: ground truth (GT) images (left image of each trio); simulated retinal images of GTs as perceived by a person with refractive errors (center image of each trio); corresponding simulated retinal images precompensated beforehand (right image of each trio). All the images are compared with the corresponding GT using the three image similarity metrics: 1-NRMSE, SSIM, and 1-STRESS.

2. RETINAL IMAGE SIMULATION

Recall that the purpose of image precompensation is to improve the perception of this image by an observer with refractive errors. Therefore, the quality of precompensation should be assessed not by the image that is presented to the observer, but by the image that is formed by the optical system of his eye on the retina. Such an image is called retinal. Thus, computer assessment of the precompensation quality impossible without a subroutine that simulates retinal images.

The human eye functions as an imaging system and behaves similarly to other systems using lenses [9]. In the case of the eye, the refractive (light focusing) surfaces are primarily the outer surface of the cornea and both surfaces of the crystalline lens. Any object that the human eye images on its retina can be considered as a two-dimensional array of dots of varying brightness [10]. For example, when a person looks at a computer monitor screen, each pixel of the monitor can be considered as a point source of light, and a corresponding image of that pixel is created on the retina. However, the eye is not a perfect thin lens and therefore cannot focus the light rays emanating from that pixel into a retinal region of the same angular size as that of the pixel. The image is always wider (in angular terms) than the original pixel. The distribution of light on the retina corresponding to the image of a point light source is described by the point spread function (PSF). The wider the PSF of a particular eye, the more the image of each monitor pixel is blurred on the retina, which leads to blurring of the entire image perceived by the observer.

To calculate the PSF, we use a standard approach based on modeling the wavefront of optical radiation passed through the optical system of the eye [11]. To model the wavefront that characterizes a particular human eye, it is necessary to somehow measure the amount of refractive error that occurs in the eye. The classic instrument for such measurements is the aberrometer. Traditionally, in aberrometry, refractive errors of the patient's eye are described by Zernike polynomials, or rather, by the amplitudes of these polynomials [12]. The approach based on Zernike polynomials makes it possible to describe PSFs of an arbitrary shape: the more polynomials are taken into account, the more complex the PSF shape can be. However, it should be borne in mind that the vast majority of people suffering from refractive errors have a refractive error

either in the sphere (myopia and hypermetropia) or in the cylinder (astigmatism); these anomalies correspond to Zernike polynomials of the second order. Therefore, for our modeling, we decided to limit ourselves to these refractive errors, not taking into account the so-called "higher aberrations"', characterized by higher-order Zernike polynomials. It is important to note that for our calculations there is no need to use aberrometry data, a standard optometrist's prescription will be enough for us, containing only three numbers for each eye: the sphere error, *S*; the cylinder error, *C*; and the cylinder axis, *A*. The first two values are measured in diopters, and the last – in angular degrees.

In the case of the prescription-based data input, the wavefront aberration function is straightforwardly constructed on the basis of the three parameters (*S*, *C*, and *A*) [12]:

$$W(x,y) = c_3 Z_3(x,y) + c_4 Z_4(x,y) + c_5 Z_5(x,y)\,, \tag{1}$$

where the three Zernike polynomials,

$$Z_3(x,y) = \sqrt{6}\,2xy, \qquad Z_4(x,y) = \sqrt{3}(2x^2 + 2y^2 - 1), \qquad Z_5(x,y) = \sqrt{6}\,(x^2 - y^2), \tag{2}$$

represent the two orthogonal astigmatisms (Z_3 and Z_5) and the defocus (Z_4), while the amplitudes of these polynomials are simply derived from the prescription parameters, as follows:

$$c_3 = D^2 sin2A \,/\, 16\sqrt{6}, \qquad c_4 = D^2(S + C\,/\,2)\,/\,16\sqrt{3}\,, \qquad c_5 = D^2 C\, cos2A \,/\, 16\sqrt{6}. \tag{3}$$

Here *D* denotes the pupil diameter of the eye. It can be seen from these expressions that the wider the pupil, the stronger the refractive errors of the eye manifest themselves. Many of us can observe this on ourselves: in bright light (when the pupil is extremely narrowed) we usually see much clearer than at dusk (when the pupil can become 2-3 times wider).

Further, the so-called generalized pupil function $P(x,y)$ [11, 13, 14] is constructed from the wavefront $W(x,y)$ distorted by refractive errors:

$$P(x,y) = A(x,y)e^{-2\pi\, i\, W(x,y)/\lambda}, \tag{4}$$

where $A(x,y)$ is the amplitude attenuation across the pupil plane, usually set equal to one, and λ is the wavelength of light in vacuum. The generalized eye pupil function defined by Eq. (4) describes the complex amplitude of the light wave passed through the eye pupil. It is well-known [14] that when a lens focuses a light wave, it performs the 2D Fourier transform of the complex amplitude distribution of the incident wave and outputs it in its focal plane, which in our case coincides with the eye retina. This can be expressed as follows:

$$K(x,y) = \left|\hat{F}\{P\,\pi Dx/\lambda\,, \pi Dy\,/\lambda\}\right|^2, \tag{5}$$

where $\hat{F}$ is the Fourier transform operator and the resultant real-valued function $K(x, y)$ is the desired PSF of the eye. The magnitude square operation reflects the fact that the human retina senses only the intensity of the light wave but not its phase. Note that in expressions (1)-(5) the transverse coordinates, *x* and *y*, are not linear but angular.

According to the classical approach [12], the human eye can be considered as a linear and shift-invariant optical system. The operation of this system is characterized by a function known in the general theory of linear systems as the impulse response function, and in optics – as the point spread function (PSF). From both names it follows that this function contains the response of the optical system to a pulsed input, that is, the image of a point light source produced by this optical system. According to the principle of superposition, the image $R(x,y)$ of an arbitrary input object, $I(x,y)$, created by our optical system, can be obtained using the convolution operation ($\otimes$) of this object with the PSF $K(x,y)$:

$$R(x,y) = I(x,y) \otimes K(x,y). \tag{6}$$

So, expressions (1)-(6) exhaustively describe the procedure of modeling a retinal image: having at the input the image $I(x,y)$ presented to the observer and the parameters of refractive errors of the eye (*S*, *C*, and *A*), we get the output image $R(x,y)$, formed by the optical system of the eye on its retina. Of course, this image is not completely identical to the visual sensation of the observer, formed in the visual cortex of the brain, but today it is all that can be modeled accurately and reliably. Note that if we wanted to take into account not only defocusing and astigmatism, but also higher-order aberrations, we would simply have to modify expressions (1)-(3), involving more Zernike polynomials.

3. HUMAN STUDIES

In our previous work, we presented a dataset, SCA-2023, which we used to evaluate the effectiveness of three different image precompensation algorithms [7]. During the evaluation process, we encountered ambiguous results in the preferences of different metrics: it turned out to be very difficult to evaluate the performance of the algorithms, since the metrics "voted"' for different things and their choice often seemed to contradict human perception. These results highlight one of the main problems in the evaluation of image precompensation algorithms: metrics based on mathematical principles do not always adequately reflect the human perception of an image. This is a key motivation for our study, in which we plan to choose the metric that most closely correlates with the human perception of precompensated images.

A characteristic feature of the SCA-2023 dataset is the fact that it contains not only ground truth images (divided into 6 subcategories), but also specially selected PSFs that are characteristic of human eyes with refractive errors such as defocus and astigmatism. Since the large size of the SCA-2023 dataset does not allow us to use it for our human studies as a whole, we created a subset of 18 ground truth (GT) images and three PSFs. This made it possible to make the study reasonable in terms of data collection time, while maintaining data diversity. Our subset includes three key PSF categories (narrow, medium, and broad), and three images each for six categories (text, icons, faces, animals, natural scenery, cityscapes). From each GT image, five processed versions were generated, using the above method of computer simulation of retinal images. Four of them were obtained using the Wiener image precompensation taken with different settings (see below), and the fifth version was a blurred GT image, without precompensation. The participants of the human studies were asked to compare these images in pairs, and only images obtained from the same GT were compared with each other. Since there were 18 GT images, and 5 processed versions were obtained from each GT, the number of possible pairs for comparison was 180 (10 pairs can be made out of five images). We believe that such a dataset for human studies, on the one hand, contains significantly different images, among which it is easy to carry out pairwise comparison (see Fig. 5), and on the other hand, provides the number of pairs sufficient to carry out statistical analysis and extracting reliable results.

The choice of the Wiener filter as the basis for precompensation was due to several key factors. First, this filter is a widely used and accepted method, as confirmed by [2, 3]. Second, using the Wiener filter makes it possible to vary the processed images using different tone mapping algorithms [4], which adds some flexibility to our study. The application of the Wiener filter for image precompensation can be expressed by the following formula:

$$G(u,v) = \frac{F(u,v)\,H(u,v)}{|H(u,v)|^2 + K}, \tag{7}$$

where $G(u,v)$ is the Fourier transform of the precompensated image; $F(u,v)$ is the Fourier transform of the GT image; $H(u,v)$ is the Fourier transform of the PSF; K is a regularization constant.

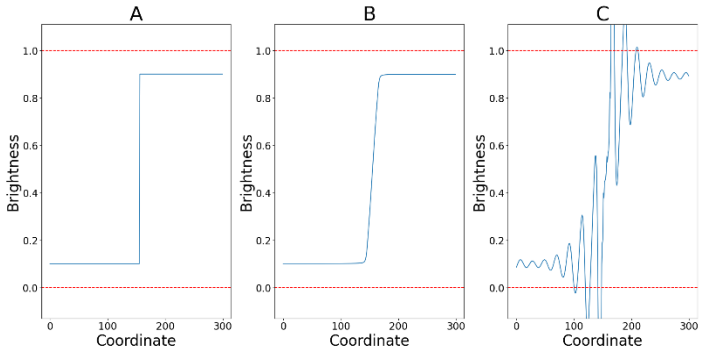

Figure 2. The main problem of Wiener precompensation. A: the original step signal; B: this signal, convolved with a certain PSF; C: the original signal after Wiener filtering. The red dotted lines indicate the range of brightness that can be displayed on the carrier.

The Wiener filter operates in the spatial frequency domain, and in the process of amplifying the high frequencies of the image spectrum, a significant increase in their amplitude can occur. As a result, after converting the image back to the spatial domain, the brightness of many pixels falls outside the brightness range of the original image. If the original image occupies the entire dynamic range of the carrier (for example, a computer monitor) in terms of brightness, then direct visualization of the Wiener image becomes impossible. This problem is considered in Fig. 2, using a one-dimensional step signal as an example. Thus, after applying the Wiener filter, additional processing is almost always required, aimed at returning the brightness of all pixels to the range of possible brightness of the carrier. This procedure is called tone mapping.

Recently, Ye et al. [4] proposed S-shaped tone mapping, which allows varying the parameters of visualization in such a way as to choose between high-contrast images with a lot of artifacts and low-contrast images with finer details. Using this approach, we obtained 4 different precompensated retinal images from each GT image:

- two images with high contrast, but with artifacts, differing from each other both in contrast and in the strength of the manifestation of artifacts;
- two images with low contrast but good edges, differing in average brightness.

Examples of such images are shown in Fig. 3.

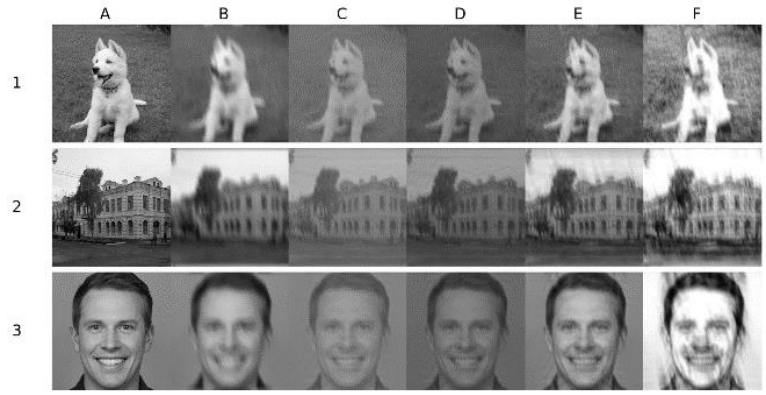

Figure 3. Examples of simulated retinal images obtained by selecting various tone mapping parameters: column A contains GT images (rows 1, 2 and 3); column B shows simulations of the corresponding retinal images obtained for human eyes with refractive errors; columns C-F contain precompensated retinal images obtained with different tone mapping settings.

For the pairwise comparison procedure, all possible pairs of images, obtained from the same GT image, were compiled. Considering that there were 5 different versions of simulated images in the set, each image was compared with 4 other versions to assess their relative preference. Thus, for the 18 GT images, 90 different simulations and 180 pairs were obtained, where each pair involved two simulations obtained from the same GT image. The GT images themselves were not included in the pairwise comparison set. The human studies were attended by volunteers from among the staff of our laboratory (24 participants in total). Each of them could evaluate a desired number of randomly selected pairs of images. For one cycle of the procedure, the participant was presented with a randomly selected (one out of 180) pair of images, and he/she was asked to choose the more preferred image from this pair. If the participant found it difficult to choose between the two images due to insufficient difference in their quality or other factors, he/she was given the opportunity not to select the better image and move on to the next pair by clicking the "Submit" button. Fig. 4 shows the user interface of our application that was used in the pairwise comparison procedure.

Figure 4. The main window of our pairwise comparison application for the human studies.

For each of the 180 pairs of images, the result of their comparison by participants was accumulated on an accrual basis:

1. If the participant preferred one of the images in the pair, then this image received 1 point, and the other image received nothing.
2. If the participant did not prefer any of the images, each image from the pair received 0.5 points.
3. For each pair, we summarized the points received from all the participants for each image.
4. To normalize the scores of the images in a pair, we divided the final score of each image by the total number of times the pair met during the evaluation process.
5. As a result, in each pair, not only the winner and the loser were revealed, but also the credibility of this victory was calculated. The score could vary from 1:0 ("crushing victory") to 0.5:0.5 ("draw").

In total, 2097 pairwise comparisons were collected from the participants, and only in 202 cases a participant did not choose a preferred image. We also analyzed the distribution of voting results in pairs. It turned out that in 82 pairs of images the winner received the absolute majority of votes (1:0). In 56 pairs, the winner scored between 0.76 and 0.99, indicating a relatively high preference for one of the images. In 39 pairs, the winner had a score ranging from 0.51 to 0.75, which indicates some preference for one of the images, but with less confidence. A tie (0.5:0.5) was recorded in only 3 out of 180 pairs, indicating that the remaining 177 pairs produced a score that can be used to further rank various image quality metrics. The distribution of the human study results by the final score of the winner in the pair is shown in Table. 1. Thus, the results of the human studies allow us to determine the preferences and perceptions of the participants regarding the image quality within the framework of the pairwise comparison procedure. Further, we will use these quantitative estimates to find the best metric(s) for assessing the image quality in the precompensation problem.

Table 1. Distribution of voting results in pairs according to the final score of the winner.

Credibility of victory	1.0	0.76-0.99	0.51-0.75	0.5
Number of Such Pairs	82	56	39	3

4. IMAGE QUALITY METRICS VS HUMAN STUDY RESULTS

For our study, we selected 5 widely used image quality metrics.

1. Normalized root mean square error (NRMSE). This metric was chosen because it is basic, extremely simple in its meaning:

$$NRMSE(x,y) = \frac{\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}}{\sqrt{\sum_{i=1}^{n} x_i^2}}, \tag{8}$$

where x is the GT image; y is the image to be evaluated; x_i, y_i are the brightness values of pixels at position i in images x and y, respectively; n is the total number of pixels in each image.

2. Standardized residual sum of squares (STRESS) [15]. This metric is interesting to us because it is invariant to any linear transform of brightness, which is fundamentally different from the previous one:

$$STRESS(x,y) = \sqrt{1 - \frac{\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}}{\sqrt{\sum_{i=1}^{n} x_i^2}\ \sqrt{\sum_{i=1}^{n} y_i^2}}}, \tag{9}$$

where the notations have the same meaning as in (8).

3. Correlation coefficient (CORR), whose distinguishing feature is the invariance to any affine transform of brightness:

$$CORR(x,y) = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{n\,\sigma_x\,\sigma_y}, \tag{10}$$

where μ_x, μ_y are the average brightness values of the pixels in images x and y, respectively; σ_x, σ_y are the standard deviations of brightness of pixels in x and y, respectively.

4. Structural similarity (SSIM) [16], a widely used metric that evaluates similarity in texture, contrast, and image structure:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (11)$$

where σ_{xy} is the covariance of x and y; $c_1 = (K_1 L)^2$ and $c_2 = (K_2 L)^2$ are two variables needed to stabilize the result of division in the case of a small denominator; L is the maximum brightness of pixels in x and y; K_1=0.01 and K_2=0.03.

5. Multiscale structural similarity (MS-SSIM) [17], which is a modification of the SSIM metric that considers the images at several spatial scales at once:

$$MSSSIM(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \prod_j^{m-1} \frac{2\sigma_{xy_j} + c_2}{\sigma_{x_j}^2 + \sigma_{y_j}^2 + c_2}, \quad (12)$$

where j is an index running through the considered scales. From this metric, we expect that it will allow us to more accurately assess the similarity between the original and the modified image at different levels of detail.

Since the NRMSE and STRESS metrics, unlike other metrics, are difference metrics, we will use the values 1–NRMSE and 1–STRESS to bring all the metrics to the same sense: now we have 5 metrics that can evaluate the similarity of images on a scale from 0 to 1.

We compared the results of the metrics with the assessments of the human study participants as follows:

- According to the results of the human studies, in each of the 180 pairs, we determined the winner and the loser, and their normalized scores.
- For each image from the pair and for each of the metrics, by comparing with the GT image, we got the metric score. Comparing these scores as a pair, we also got the winner and the loser (in terms of this metric).
- The metric added to its score the human study result of the image that it considered the winner in the pair. That is, the increment of the metric was large if its choice coincided with that of the participants and small if it did not.
- Finally, we summarized the increments of all the metrics for all 180 pairs of images and normalized the results by the sum of the scores of the winners according to the participants of the human studies. Thus, the final quality of the metrics was scaled from 0 to 1, where 1 corresponds to the full compliance with people's assessments.

Note that in our procedure we deliberately did not use Kendall or Spearman rank correlation, which is often employed to compare ranked lists [18, 19]. This is because our procedure did not simply fix the same or opposite choice of the winner and the loser in the pair, but explicitly used the fact that in some pairs people were absolutely confident in the winner (and the metric that agreed with them received an increment of 1, and the one that disagreed received 0), and in some other pairs they gave a result close to a draw, and on such pairs no metrics could get a significant advantage over each other, which could happen when using the Kendall/Spearman rank correlation. The final results of the comparison procedure are presented in Table 2.

Table 2. The degree to which the scores given by the image similarity metrics correspond to the human study results.

Metrics	CORR	MSSSIM	SSIM	1-STRESS	1-NRMSE
Compliance with human studies	0.88	0.72	0.67	0.54	0.43

Based on the data presented in Table 2, we can conclude that among the metrics we have chosen, there is a clear leader, the CORR metric. The points scored by the other metrics do not allow us to consider them adequate for assessing the quality of precompensated images. Let us now show how to use these results in optimization procedures aimed at improving the quality of precompensation.

5. EXAMPLES OF USING DIFFERENT IMAGE QUALITY METRICS FOR PRECOMPENSATED IMAGE OPTIMIZATION

As we noted earlier, in the process of image precompensation based on the Wiener filtering, the following parameters play a critical role and require careful optimization: the regularization parameter K (see Eq. 7) and a set of parameters that specify the tone mapping law. It is worth emphasizing that the processes of adjusting the tone mapping parameters and the regularization parameter are interconnected, therefore all these parameters require simultaneous optimization to achieve the best quality of the precompensated image. For the optimization, we used the gradient descent method implemented through the Adam algorithm. It is interesting to note that for the above two groups of parameters, we used different learning rates (lr): lr=0.001 for the K parameter and lr=0.01 for the tone mapping parameters. This is due to the fact that with the same absolute value change, the first parameter affects the resulting image stronger than the latter. As maximizing objective functionals, we chose the two metrics that showed the highest and the lowest degree of compliance with the human ratings, CORR and 1–NRMSE, expecting much better results for the former than for the latter. However, contrary to these "naive expectations"', we got the optimization results shown in Fig. 5. As we can see from the figure, despite the high correlation of the CORR metric with the human ratings, this metric does not attach any importance to the contrast of the image, and only values structural compliance. As a result, the optimization misses the optimum expected by a human and goes into the realm of clear, but low-contrast solutions. In turn, optimization using the 1–NRMSE metric performed above expectations, generating a high-contrast image, albeit with artifacts. Thus, we see that for using it in the optimization procedure, our leader of similarity to the opinion of people, the CORR metric, needs to be improved. In order to make the target functional sensitive to the contrast of the resulting image, a linear combination of the CORR and 1–NRMSE metrics can be used, with the former preferably dominating. The results of optimization using such a combined metric are shown in Fig. 6. As it can be seen from the figure, by varying the ratio between the two parts of the target functional, we can move either towards increasing the contrast of the precompensated retinal image and, at the same time, artifacts on it, or in the opposite direction.

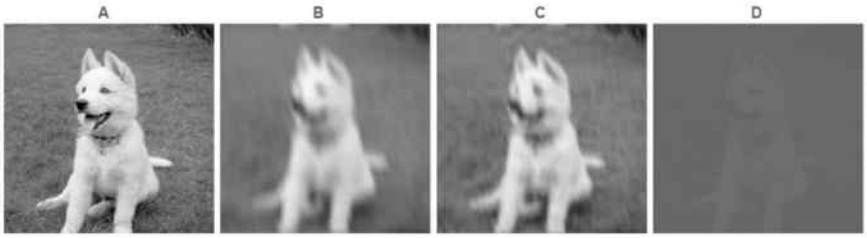

Figure 5. GT image (A); a retinal image obtained from GT for a human eye with refractive errors (B); a precompensated retinal image obtained by 1–NRMSE optimization (C); a precompensated retinal image obtained by optimization using the CORR metric (D).

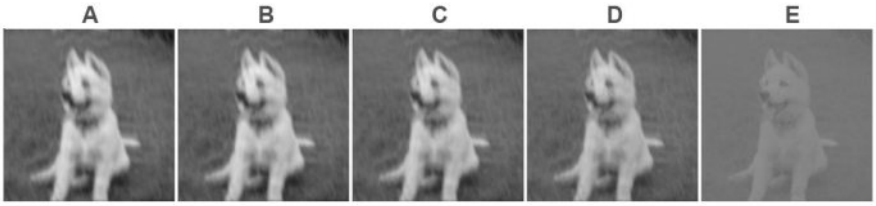

Figure 6. Precompensated retinal images obtained by optimization using various linear combinations of the CORR and 1–NRMSE metrics: (A) 0.75(CORR) + 0.25(1–NRMSE); (B) 0.8(CORR) + 0.2(1–NRMSE); (C) 0.85(CORR) + 0.15(1–NRMSE); (D) 0.88(CORR) + 0.12(1–NRMSE); (E) 0.9(CORR) + 0.1(1–NRMSE). Note the tradeoff between the contrast and the clarity.

6. CONCLUSION

We consider the main result of this work to be the elaboration of an original procedure for evaluation of image quality metrics suitable for assessing images specially precompensated for users with refractive errors. The procedure is based on an original dataset containing aggregated user preferences for images of the discussed nature. This dataset serves as the foundation for looking at various image quality evaluation metrics to figure out if they better or worse match people's

ratings. As a test of the technique, 5 commonly used image similarity metrics were tested and the most appropriate one was identified. It turned out to be the correlation coefficient. Based on the same methodology, one can further analyze any other metrics that seem promising for evaluating the quality of precompensated images. We also showed that the results of comparing the metrics could be effectively used in optimization procedures aimed at improving the perceived quality of precompensated images.

REFERENCES

[1] B. A. Myers, "A brief history of human-computer interaction technology," ACM Interactions, 5(2), 44–54 (1998).

[2] M. Alonso and A. B. Barreto, "Pre-compensation for high-order aberrations of the human eye using on-screen image deconvolution," in Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439), 1, 556–559, IEEE (2003).

[3] J. Huang, A. Barreto, M. Alonso, and M. Adjouadi, "Vision correction for computer users based on image precompensation with changing pupil size," in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society , 4868–4871, IEEE (2011).

[4] J. Ye, Y. Ji, M. Zhou, S. B. Kang, and J. Yu, "Content aware image pre-compensation," IEEE Transactions on Pattern Analysis and Machine Intelligence 41(7), 1545–1558 (2018).

[5] C. Montalto, I. Garcia-Dorado, D. Aliaga, M. M. Oliveira, and F. Meng, "A total variation approach for customizing imagery to improve visual acuity," ACM Transactions on Graphics (TOG) 34(3), 1–16 (2015).

[6] A. H. G ̈uzel, J. Beyazian, P. Chakravarthula, and K. Aks, it, "Chromacorrect: prescription correction in virtual reality headsets through perceptual guidance," Biomedical Optics Express 14(5), 2166–2180 (2023).

[7] N. B. Alkzir, I. P. Nikolaev, D. P. Nikolaev. "SCA-2023: A Two-Part Dataset For Benchmarking The Methods Of Image Precompensation For Users With Refractive Errors." Communications of the ECMS, 37 (1), 298–305 (2023).

[8] S. Ebrahimi and V. Y. Mariano, "Image Quality Improvement in Kidney Stone Detection on Computed Tomography Images", Journal of Image and Graphics 3(1), 40-46 (2015).

[9] L. N. Thibos, "Formation and sampling of the retinal image," in Seeing , 1–54, Elsevier (2000).

[10] R. G. Wilson and S. M. McCreary, Fourier series and optical transform techniques in contemporary optics: an introduction , Wiley-interscience (1995).

[11] T. O. Salmon, Corneal contribution to the Wavefront aberration of the eye , Indiana University (1999).

[12] G.-M. Dai, Wavefront optics for vision correction , SPIE Press (2008).

[13] M. Alonso, Jr, A. Barreto, J. G. Cremades, J. A. Jacko, and M. Adjouadi, "Image pre-compensation to facilitate computer access for users with refractive errors," Behaviour & Information Technology 24(3), 161–173 (2005).

[14] J. W. Goodman, Introduction to Fourier optics , Roberts and Company publishers (2005).

[15] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," Psychometrika 29(1), 1–27 (1964).

[16] 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 13(4), 600–612 (2004).

[17] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, 2, 1398–1402, Ieee (2003).

[18] D. S. Sidorchuk, "wEscore: quality assessment method of multichannel image visualization with regard to angular resolution," 46(1), 113–120 (2022).

[19] K. De and V. Masilamani, "A new no-reference image quality measure for blurred images in spatial domain", Journal of Image and Graphics, 1(1), 39-42 (2013).