SCA-2023: A TWO-PART DATASET FOR BENCHMARKING THE METHODS OF IMAGE PRECOMPENSATION FOR USERS WITH REFRACTIVE ERRORS

Nafe B. Alkzir
HSE University
20 Myasnitskaya Ulitsa, Moscow 101000, Russia;
Institute for Information Transmission Problems, RAS
Bolshoy Karetny per. 19, Moscow, 127051, Russia
E-mail: nalkzir@hse.ru

Ilia P. Nikolaev
Institute for Information Transmission Problems, RAS
Bolshoy Karetny per. 19, Moscow, 127051, Russia
E-mail: i.p.nikolaev@visillect.com

Dmitry P. Nikolaev
Institute for Information Transmission Problems, RAS
Bolshoy Karetny per. 19, Moscow, 127051, Russia
E-mail: dimonstr@iitp.ru

KEYWORDS
benchmark dataset, visual accessibility, inverse filtering, image precompensation, ocular aberration, refractive error modeling, image deblurring

ABSTRACT

This paper considers the problem of precompensating images shown to users with various anomalies of refraction of the eyes (e.g. myopia or astigmatism) in situations where they are not equipped with glasses or other corrective devices. Researchers have proposed a considerable number of such precompensation methods, but to this day there has been no way to accurately compare their quality. We propose an original dataset, which we called "SCA-2023", of images specially designed for this purpose. Its most important feature is the fact that it includes not only a set of ground-truth images for implementing the precompensation transform, but also a separate set of images characterizing specific types and degrees of manifestation of the refractive errors. The second part of the dataset is used for computer simulation of the so-called retinal image (the distribution of light on the retina of an imaginary observer). We demonstrated the capabilities of our approach using three prior-art precompensation methods and found that not all the image comparison metrics provide adequate results when applied to precompensated retinal images.

INTRODUCTION

Most computer users with refractive errors, such as nearsightedness, farsightedness, and/or astigmatism, experience problems with deteriorating image clarity when these errors are not corrected with glasses or contact lenses. This situation can arise, for example, when using a virtual reality headset where there is no physical space for glasses. It is also quite common for users with minor refractive errors to not wear glasses or contact lenses at all, leading to discomfort when interacting with smartphone displays and other devices that communicate with the user through a self-luminous screen on which certain images are reproduced. Due to the refractive errors, the clarity of the perceived image is reduced to the state when certain parts of the image cannot be recognized, especially those with important fine details.

Researchers consider the human eye as an imaging system, similar to any other optical system composed of lenses. For the discussed set of problems, the object viewed by the eye can be represented as a two-dimensional array of points of a different brightness. An ideal eye would transfer each object point to a well-focused point on the retina. However, no human eye is ideal: its refracting surfaces (primarily the cornea's front surface and both surfaces of the lens) refract light rays differently from what the ideal lens would do. This is attributed to the refractive errors: the stronger the deviation of the ray path from the ideal one, the greater the values of nearsightedness, farsightedness, and/or astigmatism that characterize the eye. The refractive errors manifest themselves as follows: light rays emanating from some object point (e.g., a computer screen pixel) cannot be focused at a single point on the retina, forming a blurred spot instead of a bright point (Alonso et al. 2005; Wilson and McCreary 1995).

A review of the scientific literature reveals several approaches to improving image quality for individuals with refractive errors. Earlier works focused on methods for improving text legibility for users. In (Lawton 1992), it is shown that losses in contrast sensitivity in visually impaired subjects can be compensated for by using various normalizing filters. These filters increase the amplitude of spatial frequencies, which are perceived less by individuals with poor vision. In (Fine and Peli 1995), it was found that spatial filtering, increasing screen brightness, and enlarging text size affected reading speed for respondents. Later, images became the subject of research. Leat et al. (2005) aimed to experimentally compare the most common image filters and found, for

Communications of the ECMS, Volume 37, Issue 1,
Proceedings, ©ECMS Enrico Vicario, Romeo Bandinelli,
Virginia Fani, Michele Mastroianni (Editors) 2023
ISBN: 978-3-937436-80-7/978-3-937436-79-1 (CD) ISSN 2522-2414

example, that participants preferred contrast-enhancing filters when viewing portraits, while filters with HSB histogram equalization were preferred when observing environmental images.

The works described above considered the problem of improving the quality of images for people with refractive errors without taking into account the characteristics of specific eyes. This approach involves the same image processing for a large group of people, that is, it is not personalized, which leads to a low indicator of improving the perception of visual information by subjects. In parallel with it, another approach developed, based on explicit accounting of the individual parameters of the visual system of the subject. Ones of the first to suggest such an approach were Alonso and Barreto (2003). They conducted clinical trials on patients, who watched images of letters with personalized preliminary compensation, and came to the conclusion that such a technique overcomes the non-personalized approach in perceived image quality.

It is important to note that it was the work of Alonso and Barreto (2003) where the term “precompensation” was first used in relation to the considered circle of tasks, although without a clear definition of it. In this work, we will call precompensation of an image, when showing it to an observer with imperfect vision, who, as a result of this, is unable to perceive this image in its entirety, such a pure-software conversion of this image (without changing the physical characteristics of the image carrier), the purpose of which is to approach the perception of the image to that of an observer with perfect vision. Note that the not-personalized methods also fall under this definition. In Figure 1, you can find some images illustrating how such a precompensation works. The simulated retinal images (c, d) model the visual perception of a short-sighted observer looking at images a and b, Figure 1b corresponding to our reproduction of the algorithm of Montalto et al. (2015). Note that the precompensated image (Figure 1b) looks pretty weird itself but becomes much closer to the ground truth after being blurred by the observer’s eye (Figure 1d).

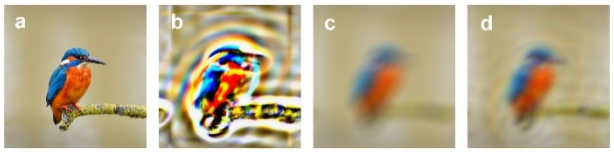

Figure 1: Image Precompensation Example: a Ground-Truth Image Taken from our Dataset (a); the Result of its Precompensation (b); the Simulated Retinal Image of the Ground Truth (c); the Simulated Retinal Image of the Precompensated Image (d)

The next significant step in the development of the personalized approach was taken as Huang et al. (2012) suggested their dynamic image precompensation, the essence of the technique being that the precompensated image was dynamically updated on the basis of the simultaneous measurement of the observer’s eye pupil size. Their idea follows from two well-known facts: refractive errors manifest themselves stronger as the pupil enlarges; and the pupil diameter, even at rest, experiences significant fluctuations in people (Fernández 2012). They (Huang et al. 2012) carried out their testing on binary icons of various sizes and compared their precompensation with the static one (when a single average pupil size was taken).

The paper (Montalto et al. 2015) was the first where color images were precompensated. Also the authors first used regularization via the total variation for controlling the artifacts arising in the course of optimization procedure. In this regard, it is important to note that the task of precompensation is an ill-posed inverse problem, which, in the general case, has no exact solution. The authors carried out their human studies on the public dataset “USC-SIPI Image Database”, however they did not provide the necessary data on the parameters of the eyes of the subjects, which makes it impossible to directly compare their results with that of other researchers.

Later, Ye et al. (2018) conducted a study inspired by the previously noted fact that the precompensated image has a significantly higher dynamic range compared to the original image, and that there are problems with displaying such images. For example, the “ideal” precompensated image may require that some pixels of this image have a brightness exceeding the maximum brightness of the monitor on which it is displayed. And some other pixels must have a negative brightness, which is unattainable even theoretically. As a result, the real precompensation is always non-ideal, which manifests itself primarily in a low contrast of the image perceived by the subject, as well as in artifacts present in it. The authors studied how tone mapping can affect the artifacts and the contrast. According to them, a linear display of colors restores the image without significant artifacts, but at the same time there is a low contrast, and a nonlinear transform can raise the contrast, but only due to the residual presence of artifacts. In this work, the authors proposed an algorithm for finding the optimal balance between the contrast and artifacts. They used images from the public dataset “BSDS500 Database”, however, just like their predecessors, they did not provide the necessary data on the eye parameters of the subjects.

In publications on image precompensation, various methods were used to assess the quality of problem solving. In (Lawton 1992), the quality metric was the speed of reading text, which required long-term experiments with subjects. Alonso and Barreto (2003) used as a metric the visual acuity score (LogMar) when viewing original and precompensated images, which also required human studies. The work (Huang et al. 2012) is interesting in that they simulated a myopic human eye with a defocused camera. This can help to partially automate benchmarking of various precompensation algorithms, but only for spherical refractive errors.

Simulating even an astigmatism is already causing serious problems, not to mention higher-order refractive errors. The paper (Ye et al. 2018) is interesting in that they are the first to compare their method with previous ones using the SSIM quality metric, which does not require any human studies. The images were taken from a public dataset, but the parameters of the eyes of the "virtual subjects" were generated in a way that, unfortunately, was not described in the paper. This prevents their results from being compared with others in the future.

From the foregoing, we can conclude that the problem of precompensation is still relevant. Its importance follows from the fact that the number of people with minor refractive errors continues to grow, and many of them do not want to wear glasses or lenses for vision correction. We see that today there are many different methods for solving the precompensation problem, however, none of them claims to be called optimal, because the problem does not have the exact solution, and the issue of comparing these methods has not yet been solved due to complexity and a high cost of clinical trials. It seems very appropriate to develop a relevant dataset for making such comparisons, as well as to introduce adequate computable quality metrics, which would automate this procedure. We believe that the solution of these problems will contribute to a significant acceleration of research in the field of image precompensation.

DESCRIPTION OF THE DATASET

In the works described above, the authors used various sets of ground-truth images to test their methods. For example, Huang et al. (2012) used binary images of single characters, as well as high-contrast monochrome icons, while Montalto et al. (2015) used more complex color images. But worst of all is the fact that the authors usually give only examples of eye refraction errors, without creating any coherent dataset of such errors that could be used in a bench to measure the quality of different methods. This approach has led to the fact that so far there is no way to check the validity of their results and to correctly compare the quality of the proposed precompensation methods with one another.

Note that there are special datasets designed for blind and non-blind deblurring methods (Sun et al. 2013). From the mathematical point of view, the non-blind deblurring problem has much in common with the precompensation problem, but the former implies a fundamentally different kind of the blur kernel: it describes hand shake or other physical phenomena that lead to image blurring, while in the precompensation problem the blur kernel should follow the refractive errors of the human eye. Therefore, these specific datasets cannot be used for the precompensation problem. Typical kernels for these two kinds of blur are shown in Figure 2. It is clearly seen that the two kernels have hardly anything in common.

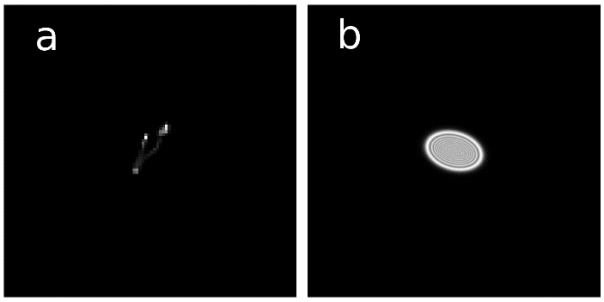

Figure 2: Typical Kernels for Motion Blur (a) and Refractive-Error Blur (b)

In this paper, we propose an original image dataset specially designed to compare different precompensation algorithms with one another. Its most important feature is the fact that it includes not only a set of ground-truth images for precompensating transformation, but also a separate set of images characterizing specific types and strength of refractive errors. This is the first dataset ever proposed, which has this feature.

Part I: Categorized Ground-Truth Images

The first part of the suggested dataset is a set of color ground-truth images, and it is deliberately divided into a few categories. The categorization was made to have a possibility of evaluating the performance of precompensation algorithms in different scenarios.

The first category is text images, which were generated using quotes from news feeds in the NLTK Corpus (Loper and Bird 2002). In addition, we took into account that in modern media the text is not always placed on a white background and added backgrounds of different shades to the test. Recall that the speed of reading text images was used as a metric for improving image quality in (Fine and Peli 1995).

The second category is icons. Such images make it possible to adequately compare algorithms that were tuned to work with binary images. In addition, precompensation artifacts show themselves brighter on binary images. This dataset portion was collected from flaticon.com, which allows one to use the icons for free for academic or educational purposes.

The third category is real images, which in turn are divided into four subcategories: animals, faces, urban and natural landscapes. Images of animals and human faces are presented mainly in close-ups, which makes it possible to determine how well different algorithms restore large objects. In turn, urban and natural landscapes usually have many fine details, which allows us to conclude whether the restoration of small objects is effective.

The categories of animals, urban and natural landscapes have been filled in using open sources, where the images are licensed to the public domain. The "face" category

consists of computer-generated non-existing faces taken from the https://this-person-does-not-exist.com website. Small-size samples of our ground-truth images are shown in Figure 3.

Figure 3: Sample Ground-Truth Images from our Categorized Dataset: Text, Icons, Animals, Faces, Urban and Natural Landscapes (from Top to Bottom).

All the pictures from the ground-truth image set (735 images in total) were resized to the same resolution of 512x512 pixels. Such a resolution, on the one hand, allows one to study rather complex images, and on the other hand, does not lead to unacceptable amounts of computation when crossing the dataset parts (when each ground-truth image is checked in pair with each blur kernel).

Part II: Characteristic PSFs of Human Eyes

According to the classical approach (Dai 2008), the human eye can be considered as a linear, shift-invariant optical system, whose action is characterized by the function $K(x,y)$, which in the general theory of linear systems is usually called the impulse response function, and specifically in optics – 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 superposition principle, the image $R(x,y)$ of an arbitrary input object, $I(x,y)$, produced by our optical system, can be obtained through the convolution operation (*):

$$R(x,y) = I(x,y) * K(x,y).$$

Note that in this paper the PSF is sometimes referred to as the blur kernel; these two being synonyms.

The PSF of the human eye is determined by the strength of refractive errors and the current pupil size (Dai 2008): the stronger the errors and the wider the pupil, the wider the PSF. The technique for calculating the PSF of the eye is not an original part of our work, so we will not focus on it in detail, referring the reader to the relevant literature, for example (Dai 2008). From a computational point of view, this calculation presents no difficulties: first, taking into account the size of the pupil and the characteristics of the refractive errors, a two-dimensional array of the so-called generalized pupil function is filled in, and then it is subjected to the fast Fourier transform (FFT). The square of the modulus of the resulting Fourier image is the required PSF. An important fact is that the characteristics of the refractive errors, necessary for the calculation, can be measured *in vivo* using various ophthalmic instruments. The classic instrument for such measurements is the aberrometer. Traditionally, in aberrometry, the refractive errors of the patient's eye are described by Zernike polynomials, or more precisely, by the amplitudes of these polynomials (Dai 2008).

The approach based on the Zernike polynomials makes it possible to describe PSFs of an arbitrarily complex shape: the more polynomials are taken into account, the more complex the PSF shape can be. But it should be borne in mind that the vast majority of people suffering from refractive errors have a refractive error which is either spherical (myopia and hypermetropia) and/or cylindrical (astigmatism); these errors corresponding to the Zernike polynomials of the second order. Therefore, for our task, we decided to restrict ourselves to these types of 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.

It is easy to show that in the geometrical optics approximation, a spherical error generates a PSF having the shape of a circle (Strasburger et al. 2018), which turns into an ellipse when a cylindrical error is added. In Figure 4, you can see sample PSFs from our dataset, which are shaped roughly like this. As we already mentioned, our calculation method is based on wave optics, not geometrical optics, which certainly affects the "fine structure"' of a PSF (generating characteristic diffraction rings at the boundaries of the ellipses), but does not affect its overall shape and size.

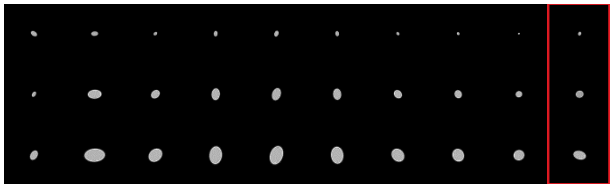

Figure 4: Sample PSFs from our Dataset: Top Row – "Narrow" Ones; Second Row – "Medium" Ones; Bottom Row – "Broad" Ones

We also divided the PSF set into three subcategories, this was done because different images require different blur

kernels. For example, if a text is blurred with a too broad kernel, it will hardly be possible to extract any useful information by precompensation; and if an icon is blurred with a too narrow kernel, we will get an image that is hardly distinguishable from the ground truth and there is no need for precompensation at all. Figure 5 shows examples of blurring images with a narrow, a medium, and a broad kernel.

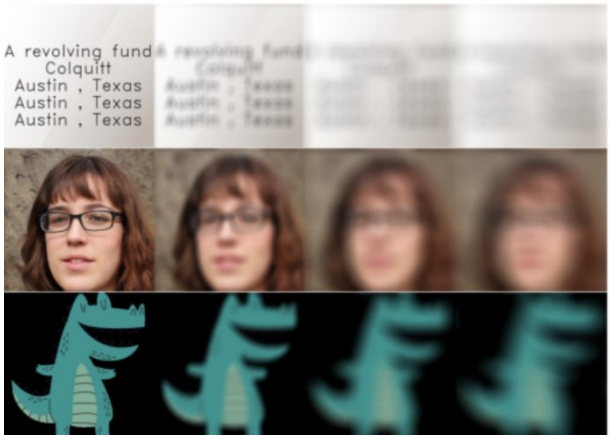

Figure 5: Examples of Blurring Three Images from our Dataset with Different PSFs: the Ground Truth (First Column); Blurring with a Narrow (Second Column), a Medium (Third Column), and a Broad PSF (Fourth Column)

It can be seen from Figure 5 that for the text (top row), the only candidate for checking the quality of precompensation is the image obtained with the narrow blur kernel; the other two look irretrievable. For the image from the "faces" category, such a candidate is, first of all, the image obtained with the medium blur kernel. For a large high-contrast "icon"', it is quite possible to work with the rightmost image, the other two blurry images are recognized even without precompensation. Note that these images were obtained with PSFs, which are enclosed in the red box in Figure 4.

Quantitatively, the following parameters were used for the "narrow PSFs" (256 pcs in total):

- *S* was randomly selected in the range of [-2.5..0] diopters;
- *C* was randomly selected in the range of [-1.25..1.25] diopters;
- *A* was randomly chosen in the range of [0..360] degrees.

For the "medium PSFs" (256 pcs in total):

- *S* was randomly selected in the range of [-5.0..-2.5] diopters;
- *C* was randomly selected in the range of [-2.5..2.5] diopters;
- *A* was randomly chosen in the range of [0..360] degrees.

For the "broad PSFs" (256 pcs, as well):

- *S* was randomly selected in the range of [-5.0..-7.5] diopters;
- *C* was randomly selected in the range of [-3.75..3.75] diopters;
- *A* was randomly chosen in the range of [0..360] degrees.

It can be seen from these data that we considered only myopic eyes (the PSF width depends only on the modulus of *S*, but does not depend on its sign, so there was no loss of generality) and took the cylinder error, on average, twice as small in amplitude as compared to the sphere error, which is similar to what is seen in the human population. The eye pupil diameter (on which the PSF size also depends) was set at a typical value of 3.5 mm; and the angular width of the ground-truth image was set at 6°. With these parameters and with the maximum sphere and no cylinder (*S*= -7.5 diopters, *C*=0, and *A*=0), the PSF diameter was about 20% of the ground-truth image width. Conducting calculations with a wider PSF hardly makes sense when working with the ground-truth images from SCA-2023.

Note that in our dataset PSFs are digitized with the same resolution as the ground-truth images: 512x512 pixels. This makes the convolution between the image and the PSF extremely convenient: it is calculated through any standard FFT routine. It should also be borne in mind that, if necessary, all our results are applicable to any actual observation conditions. In this case, the linear size of the observed image and the observation distance generate an angular grid on the corresponding ground-truth image, and the PSF is also recalculated into the angle arguments (Dai 2008).

The SCA-2023 dataset can be downloaded at https://doi.org/10.5281/zenodo.7848576 as a ZIP archive. There are two folders there: the "Images" folder contains our ground-truth images, while the "PSFs" folder contains the simulated PSFs along with the parameters used in the simulations. The former has three subfolders: "Texts", "Icons" and "Real_images". The "Texts" folder contains 225 text images. The "Icons" folder contains 104 icon images. In turn, the "Real_images" folder has 4 subfolders: "Animals" with 100 images, "Faces" with 102 images, "Natural" with 103 images, and "Urban" with 101 images. The "Faces" folder also contains a link to the site where the image data were generated, it is contained in the link.txt file. All the images are saved in the JPG format and have the same size of 512x512.

The "PSFs" folder contains three subfolders: "Broad", "Medium", and "Narrow", each of them containing 256 PSFs saved in the CSV file format. The "PSFs" folder also contains the parameters.csv file, which describes the parameters corresponding to each PSF. All the PSFs are also sized 512x512.

DATASET-BASED COMPARISON OF THREE PRIOR-ART PRECOMPENSATION METHODS

To check the capabilities of our approach, we reproduced 3 known image precompensation algorithms. As a starting point, we chose one of the classic non-personalized precompensation algorithms proposed by Peli and Peli (1984). The idea of this algorithm implies amplifying high-frequency image components while shifting the local brightness to its medium level. The first allows you to enhance fine details, the second expands the dynamic range for visualization of these details. In Figure 6, column 3 shows some examples of applying this algorithm to images from our ground-truth set.

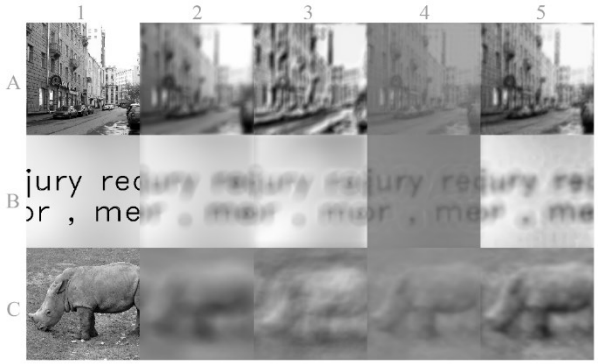

Figure 6: Simulated Retinal Images for 3 Ground-Truth Images (A, B, C). Column 1– Ground Truth; Column 2 – Non-Precompensated Retinal Images ("natural blur"); Column 3 – Peli-Peli Algorithm; Column 4 – Huang Algorithm; Column 5 – Montalto Algorithm. Rows A, B, and C Use the Three Red-Framed PSFs from Figure 4 (Top to Bottom, Respectively).

The retinal images were simulated on the basis of PSFs taken from the second part of our dataset. Note that our approach is suitable both for color and grayscale images, but we deliberately converted our ground-truth images to grayscale ones. The reason was to reproduce exactly the original Peli-Peli algorithm, which had been proposed for black-and-white images and tested by the authors on such images. It can be seen from Figure 6 that this algorithm shows perhaps the worst retinal image quality of the three (columns 3, 4, and 5). This is not surprising, since it is the only non-personalized algorithm among the three reproduced.

The second algorithm that we reproduced is, in a certain sense, an extension of the first one for the case of personalized precompensation. While Peli and Peli (1984) emphasize high spatial frequencies always in the same way, Huang (2013) suggests pre-amplifying precisely those spatial frequencies that have suffered the most due to blurring. Knowing the blur kernel, Huang builds a precompensated image through Wiener filtering. In Figure 6, column 4 shows examples of applying this algorithm to images from our ground-truth image dataset. The Huang's algorithm was also proposed for grayscale images, and we did not change this. It can be seen from Figure 6 that Huang's retinal images are much clearer than those of Peli-Peli, but have much lower contrast. This is because, mathematically, the Wiener filter produces the optimal solution to the precompensation problem in the form of such pixel brightness values, not all of which can be reproduced on the image carrier. Many pixels must be brighter than the maximum achievable brightness, and the brightness of some other pixels must be even negative. To overcome these difficulties, Huang proposes to linearly map all the pixel brightness values onto the actual dynamic range of the imaging device. As a result of this mapping, the retinal image contrast dramatically decreases.

The third algorithm that we reproduced in the framework of our approach is one of the modern precompensation algorithms that uses gradient optimization of images instead of their Fourier processing. It was proposed by Montalto et al. (2015). In this paper, the authors consider several modifications of their approach, and we would like to fix the setting that we have chosen. We will minimize the following functional:

$$p(\theta, t) = \operatorname{argmin}_{0 \le p \le 1}(\|k * p - t\|_2 + \theta\|\nabla p\|_1),$$

where t and p are the ground-truth and the precompensated image, respectively, and k is the blur kernel (PSF). In this functional, the first term in round brackets requires the proximity of the precompensated image convolved with the PSF to the original image in the L_2-sense. The second term is regularizing, with the choice of the total variation of p as the regularizer, $\theta > 0$ being the regularization weight. The total variation is a measure of the variability in the brightness of pixels in an image and is traditionally used to minimize the blur and sharpen the image.

The algorithm takes two images as the input: the ground truth and the PSF, both from SCA-2023. Functional minimization is carried out using the gradient descent method, which makes this algorithm the slowest of the three, but the temporal performance is not the subject of this study. In Figure 6, column 5 shows examples of applying this algorithm to images from our ground-truth image dataset. It can be seen from Figure 6 that this algorithm shows, perhaps, the best retinal image quality of the three. This could have been expected, as it is the most recent algorithm reproduced.

To complete our benchmarking, we need to choose an appropriate quality metric for measuring the error between the ground-truth and the precompensated image. In previous works, the SSIM and PSNR metrics (Tanaka and Kawano 2021) have already been used for comparison, which is quite common for works on image restoration. However, we have already mentioned above that precompensation algorithms can produce both low-contrast but artifact-free images and high-contrast but significantly distorted ones. This trade-off seems to be a fundamental feature of the precompensation problem. Therefore, it makes sense to use metrics that react

differently to the loss of contrast. The correlation coefficient (CORR) does not change with either contrasting or shifting the grayscale of any of the compared images. High CORR values, obtained when other quality scores are low, signal a contrast change.

Loss of contrast is accompanied by brightness distortions at least at one end of the grayscale. It is known that humans have mechanisms to adapt to lowering the maximum brightness of an image, but there are no mechanisms to adapt to "bleaching the image", when the grayscale shifts up from zero. The standardized residual sum of squares (STRESS) metric does not penalize a linear transform of the image intensity (Garcia et al. 2007), but it is sensitive to brightness shifts. Thus, combining the CORR and the STRESS metrics makes it possible to distinguish between different causes of contrast loss. When using the CORR and the STRESS metrics, it is natural to take as the basic metric not PSNR but the normalized root mean square error (NRMSE), the two latter differing only in scale.

In addition to the previously used SSIM metric, we consider it reasonable to use the more modern MS-SSIM as well. To make all our metrics measuring image similarity, we subtracted NRMSE and STRESS from unity. Thus, we obtained 5 similarity metrics: 1-NRMSE, 1-STRESS, CORR, SSIM and MS-SSIM. All the chosen metrics have a range of values from 0 to 1, where 1 means the exact similarity to the ground-truth image.

Table 1 presents the results of comparing the Peli-Peli, Huang, and Montalto algorithms using the 5 image similarity metrics we chose. Corresponding values of the metrics obtained for non-precompensated retinal images are given for the reference. The leading results are bolded. We used all the three subsets of PSFs (each consisting of 256 PSFs) and applied each PSF to each of the 735 images from our ground-truth dataset in order to obtain the reference. Then we repeated this procedure with the precompensated images. It is important to recall that the Huang and Montalto algorithms used PSF in their precompensation calculations, but the Peli-Peli algorithm did not. Next, we compared each retinal image with the ground truth according to the above metrics, averaged the results over all the images, and summarized everything in the table below.

Table 1: Comparison of Three Prior-Art Precompensation Algorithms with Five Image Similarity Metrics

Methods	Metrics	Broad PSF	Medium PSF	Narrow PSF
No precom-pensation	1-NRMSE	0.785	0.847	0.863
	1-STRESS	0.736	0.775	0.849
	CORR	0.741	0.799	0.894
	SSIM	0.620	0.687	0.809
	MS SSIM	0.596	0.656	0.806
Peli-Peli	1-NRMSE	0.749	0.751	0.729
	1-STRESS	0.734	0.733	0.821
	CORR	0.595	0.620	0.770
	SSIM	0.504	0.502	0.591
	MS SSIM	0.451	0.506	0.647
Huang	1-NRMSE	0.654	0.688	0.701
	1-STRESS	0.568	0.588	0.655
	CORR	**0.839**	**0.877**	**0.933**
	SSIM	0.426	0.451	0.557
	MS SSIM	0.536	0.596	0.748
Montalto	1-NRMSE	**0.835**	**0.897**	**0.921**
	1-STRESS	**0.766**	**0.803**	**0.871**
	CORR	0.796	0.843	0.919
	SSIM	**0.667**	**0.723**	**0.828**
	MS SSIM	**0.653**	**0.718**	**0.858**

Comparing the results we obtained for the algorithms chosen, we came to the following conclusions. First, the use of broader PSFs leads to a stronger deterioration in the numerically predicted quality of the precompensated retinal image, especially when using the SSIM and MS-SSIM metrics. This confirms the “intuitive” idea that the precompensation technique cannot compensate for a too strong image blur caused by eye refraction errors. Second, the Montalto approach, on most of the metrics, shows better results than the Huang and the Peli-Peli algorithms do. However, at all the three PSF levels, the CORR metric announces the Huang method as the winner. This is quite natural, because the Huang method, by design, minimizes any deviations from the exact solution of the deconvolution problem, but at the cost of an arbitrarily large loss of contrast.

Thus, the Montalto algorithm showed the best results in most of the metrics that give an idea of the quality of the precompensated retinal images from the point of view of a computer. However, it is not guaranteed that the metrics fully reflect the quality of the images in terms of human perception. For example, it looks strange that some of the metrics prefer the non-precompensated image to the precompensated one. Therefore, for a more complete understanding of the problem of correct automatic comparison of such images, it is quite promising to conduct additional research aimed at studying the perception of these images by human observers.

CONCLUSION

In this work, we have proposed a special dataset for benchmarking and comparison of image precompensation methods for displaying images to users with refractive errors. The most important feature of this dataset is the fact that it includes not only a set of ground-truth images for precompensating transformation, but also a separate set of images characterizing specific types and strength of refractive errors. This is the first dataset ever proposed, which has this feature.

The dataset is called “SCA-2023” and can be downloaded at https://doi.org/10.5281/zenodo.7848576 as a ZIP archive. Our dataset contains both a categorized set of ground-truth images and a three-scale set of simulated kernels, aimed at blurring these images, along with the parameters used in the simulations. The blur kernels represent low-order errors of human eye refraction, such as spherical and/or cylindrical ones. All the ground-truth images are saved in the JPG format and have the same size of 512x512. The blur kernels are saved in the CSV file format and are also sized 512x512.

While using the SCA-2023 dataset for benchmarking three prior-art precompensation methods, we have found that not all the image similarity metrics widely used in image processing are suitable for this task. In the future, we plan to conduct human studies to select the metric that would be most correlated with the assessments of people observing images such as those shown in Figure 6. Moreover, we believe that the very best precompensation method has not been yet found, and we are going to look for it, using the proper metrics as guides.

REFERENCES

Alonso, M. and A.B. Barreto. 2003. “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, volume 1, 556-559.

Alonso, Jr, M.; A. Barreto; J.G. Cremades; J.A. Jacko; and M. Adjouadi. 2005. “Image pre-compensation to facilitate computer access for users with refractive errors.” *Behaviour & Information Technology* 24, No. 3, 161-173.

Dai, G.M. 2008. *Wavefront optics for vision correction*. SPIE press, Bellingham, WA.

Fernández, E.J. 2012. “Adaptive optics for visual simulation.” *International Scholarly Research Notices* 2012.

Fine, E.M. and E. Peli. 1995. “Enhancement of text for the visually impaired.” *JOSA A* 12, No. 7, 1439-1447.

Garcia, P.A.; R. Huertas; M. Melgosa; and G. Cui. 2007. “Measurement of the relationship between perceived and computed color differences.” *JOSA A* 24, No. 7, 1823-1829.

Huang, J. 2013. “Dynamic image precompensation for improving visual performance of computer users with ocular aberrations.” *FIU Electronic Theses and Dissertations* 902.

Huang, J.; A. Barreto; and M. Adjouadi. 2012. “Evaluation of dynamic image pre-compensation for computer users with severe refractive error.” In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*, 175-182.

Lawton, T. B. 1992. “Image enhancement filters significantly improve reading performance for low vision observers.” *Ophthalmic and Physiological Optics* 12, No. 2, 193-200.

Leat, S.J.; G. Omoruyi; A. Kennedy; and E. Jernigan. 2005. “Generic and customised digital image enhancement filters for the visually impaired.” *Vision research* 45, No. 15, 1991-2007.

Loper, E. and S. Bird. 2002. “NLTK: The natural language toolkit.” *arXiv preprint arXiv:cs/0205028*.

Montalto, C.; I. Garcia-Dorado; D. Aliaga; M.M. Oliveira; and F. Meng. 2015. “A total variation approach for customizing imagery to improve visual acuity.” *ACM Transactions on Graphics* 34, No. 3, 1-16.

Peli, E. and T. Peli. 1984. “Image enhancement for the visually impaired.” *Optical engineering* 23, No. 1, 47-51.

Strasburger, H.; M. Bach; and S.P. Heinrich. 2018. “Blur unblurred – a mini tutorial.” *i-Perception* 9, No. 2, 1-15.

Sun, L.; S. Cho; J. Wang; and J. Hays. 2013. “Edge-based blur kernel estimation using patch priors.” In *IEEE international conference on computational photography (ICCP)*. IEEE, 1-8.

Tanaka, H. and H. Kawano. 2021. “Image correction for improving visual acuity using Zernike-based vision simulation.” In *20th International Symposium on Communications and Information Technologies (ISCIT)*, 32-36.

Wilson, R.G. 1996. “Fourier series and optical transform techniques in contemporary optics: an introduction.” *Optics & Photonics News* 7, No. 2, 59-60.

Ye, J.; Y. Ji; M. Zhou; S.B. Kang; and J. Yu. 2018. “Content aware image pre-compensation.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, No. 7, 1545-1558.

AUTHOR BIOGRAPHIES

Nafe B. Alkzir was born in Donetsk, Ukraine. He studied computer sciense, obtained his master’s degree in 2021 at the HSE Universty. From 2021 he has been developing image pre-processing frameworks with the Vision Systems Lab at the Institute for Information Transmission Problems. His research interests are image processing and enhancement in the areas of image precompensation and deblurring. His e-mail address is nafekzir@gmail.com.

ILYA P. NIKOLAEV was born in Moscow, USSR. He studied physics, obtained his master’s degree in 1994 and his Ph.D. degree in 1997, all at the Moscow State University. After some years of working in the field of adaptive optics, in 2020 he joined the Vision Systems Lab at the Institute for Information Transmission Problems RAS. His current research activities are image processing and visual perception. His e-mail address is i.p.nikolaev@visillect.com.

DMITRY P. NIKOLAEV was born in Moscow, USSR. He studied physics and computer science, obtained his master’s degree in 2000 and his Ph.D. degree in 2004, all at the Moscow State University. Since 2007 he is a head of the Vision Systems Lab at the Institute for Information Transmission Problems RAS. His research activities are in the areas of computer vision and image processing with focus on the computationally effective algorithms. His e-mail address is dimonstr@iitp.ru.