Predicting blur visual discomfort for natural scenes by the loss of positional information

Elio D. Di Claudio, Paolo Giannitrapani, Giovanni Jacovitti

A Dept. of Information Engineering, Electronics and Telecommunications (DIET), University of Rome “La Sapienza”, via Eudossiana 18, I-00184 Rome, Italy
B Retired, formerly with DIET, University of Rome “La Sapienza”, via Eudossiana 18, I-00184 Rome, Italy

1. Introduction

Among the various sources of non-clinical visual discomfort (Conlon, Lovegrove, Chekaluk, & Pattison, 1999; Ingrid, 2009; O’Hare, Zhang, Nefs, & Hibbard, 2013), the blur caused by refractive errors is perhaps the most common one.

The discomfort associated with blur is often explained as a consequence of the concentration of the spatial energy spectrum of the perceived image into some bands, or as a byproduct of the discrepancy of this spectrum from the expected spectrum of natural images (O’Hare & Hibbard, 2011; Wilkins, 2016).

Alternative explanations addressed the mismatch of the spatial patterns with the expected ones (Kayargadde & Martens, 1996; Wang & Simoncelli, 2004; Baroncini, Capodiferro, Di Claudio, & Jacovitti, 2009). An in-depth account of previous studies and mathematical models about the blur phenomenon is provided in (Watson & Ahumada, 2011).

Looking at a possible physical mechanism of blur discomfort, three hypotheses are examined in (O’Hare & Hibbard, 2013). The first hypothesis is that discomfort is stimulated by the weak response of the accommodation system. A second hypothesis, somewhat related to the first one, is that discomfort arises because the “micro-fluctuations” observed in the accommodation feedback signal become ineffective (Charman & Heron, 2015; Metlapally, Tong, Tahir, & Schor, 2016; Marin-Franch et al., 2017; Cholewiak, Love, & Banks, 2018). A third hypothesis maintains that, when an image is correctly projected onto the retina, the receptive fields produce a parsimonious, sparse representation of this image (McIlhagga & May, 2012). The spatial spread caused by blur excites more receptors, producing a metabolic overload (Juricevic, Land, Wilkins, & Webster, 2010).

In the present approach, the discomfort is quantified in terms of localizability measured by Fisher Information losses. It is argued that, among the basic functions of the human visual system (HVS), detection, recognition, and coarse localization functions are strongly conditioned by the individual experience. Conversely, it seems plausible that the fine localization role is committed to stabler and inter-subjective functions of the HVS.

Based on the above consideration, the present approach starts from postulating that, under normal conditions, the HVS performs the fine localization of the observed patterns with the best accuracy allowed by...
its physical macro-structure, given the characteristics of the environment and man’s interaction. This assumption is fundamental, because it is known from the estimation theory that the maximum accuracy attainable when measuring the fine position of patterns in background noise is deduced from the Fisher Information about positional parameters. In fact, the Fisher Information inverse yields the minimum estimation variance (Van Trees, 1992).

Therefore, the focus here is on how the discomfort of blurring is related to the unwanted losses of Positional Fisher Information (PFI) on observed patterns, viewed as cognitive losses. This is a cognitive approach because it is inspired to concepts of cognitive sciences and, specifically, the classical Marr’s scheme (Marr, 2010). The PFI losses are calculated at the computational and representational levels, irrespective of the underlying physical mechanisms characterizing the implementation level. See (Peebles & Cooper, 2015) and the introduction of (Di Claudio & Jacovitti, 2018).

Therefore, the present approach is agnostic as to whether the discomfort is related to accommodation fatigue or metabolic imbalances. From a more general viewpoint, the cognitive approach may account for the fact that blur discomfort concerns the regions of visual interest (Taylor, 2015) and that the blur is not always undesired or detrimental (Sprague, Cooper, Reissier, Yellapragada, & Banks, 2016). Sometime, the blur is a wanted effect in photography (Datta, Joshi, Li, & Wang, 2006) and microscopy.

Previous analyses of the blur perception phenomenon were mainly oriented to the study of the visual acuity, employing specific stimuli localized either in space, such as edges, lines, crosses, or in the spatial frequency domain, such as sinusoidal gratings, or even in both domains, such as Gabor wavelets (Watson & Ahumada, 2005).

The model presented here is oriented to the evaluation of the visual impact of blur in the vision of natural scenes. To this purpose, a generic image projected on the retina is viewed as an element of the random set of natural images, characterized by stable statistical features.

The proposed approach is based on an abstract, functional model of the receptive fields (RF) of the HVS, allowing for a direct computation of the PFI. The model accounts, at the same time, for both the contrast sensitivity function (spectral response) and the local directional selectivity of the eye.

For analytical convenience, the blur is modelled as a Gaussian shaped isotropic blur. Then, the analysis is generalized to other types of blur, invoking a criterion of informational equivalence with respect to isotropic Gaussian blur under the PFI paradigm.

To verify the limits of the present approach, the model-based discomfort predictions were first compared to empirical data about the subjective quality loss of blurred images, which is arguably related to blur visual discomfort. These data are available in organized databases containing the results of experimental sessions finalized to Image Quality Assessment (IQA), conducted for multimedia industry purposes (Wang, Bovik, Sheikh, & Simoncelli, 2004; ITU, 2008; Bosse, Maniry, Muller, Wiegand, & Samek, 2018). Subsequently, the model is applied to blurred images annotated with ratings of visual discomfort. The results of these experiments confirm the validity of the hypothesis.

The paper is organized as follows. Section 2 describes the features of the functional RF model in the space and in the spatial frequency domains. Section 3 provides the definition of the PFI. In Section 4, the expected Fisher Information acquired during the visual exploration of natural images is computed, and the informational equivalence of a generic blur with a Gaussian isotropic blur is stated. In Section 5, the measure of the visual discomfort is defined. Section 6 presents the comparison of the predicted discomfort with subjective ratings. Some remarks are provided in Section 7. Conclusions are finally drawn in Section 8.

2. The virtual receptive field model

As the luminance plays a dominant role for the localizability of patterns, for the sake of simplicity only the luminance component of the images is accounted for. Chrominance is neglected.
The radial frequency response magnitude of the VTNF for $s_0 = 2.5$ arcmin normalized with respect to its maximum value.

The retina in the foveal vision is abstractly modeled as a rectangular grid of receptors, whose position is individuated by the coordinate pair $p = (x_1, x_2)$. Receptors are regularly spaced one arcmin apart. The density of 60 receptors/degree assures that all the image information within the 30 (-60/2) cycles/degree spatial bandwidth is captured by the retina, according to the Nyquist sampling rule.

For any $p$, the RF calculates a weighted sum of the luminance $I(p)$ in a neighborhood of $p$, yielding a visual map $y(p)$. This operation corresponds mathematically to a spatial convolution, indicated by the symbol $\ast$, between $I(p)$ and the visual map of a single lighting point in the dark, indicated by $h(p)$ and referred to as the Point Spread Function (PSF) of the RF:

$$y(p) = I(p) \ast h(p).$$  \hspace{1cm} (1)

The RF model considered here is a harmonic angular filter (HAF). HAF filters are complex valued functions, i.e., they represent pairs of real filters (Jacovitti & Cusani, 1990).

With reference to the spatial frequency domain, defined by the horizontal and vertical frequencies $f_1, f_2$, and to the polar coordinates $\rho = \sqrt{f_1^2 + f_2^2}$ and $\theta = \arctan\left(\frac{f_1}{f_2}\right)$, where $\rho$ (Daugman, 1980) is referred to as the radial frequency and $\theta$ is the azimuth, the Fourier spectrum $H(\rho, \theta)$ of the RF model is defined as

$$H(\rho, \theta) = j2\pi \left(\rho e^{-j\phi} \cdot e^{j\theta}\right)$$  \hspace{1cm} (2)

where $j$ is the imaginary unit, and $s_0$ is a parameter. This spectrum is polar separable, i.e., it is the product of a function of the radial frequency and a function of the azimuth. One outstanding feature of this RF model is that it is also polar separable in the space domain (Daugman, 1983). In fact, taking the inverse Fourier transform of $H(\rho, \theta)$, with reference to the space polar coordinates $r = \sqrt{x_1^2 + x_2^2}$ and $\varphi = \arctan\left(\frac{x_1}{x_2}\right)$, the PSF of the RF is:

$$h(r, \varphi) = -2 \frac{\pi^2}{s_0^2} \left(\cdot e^{\frac{r^2}{2s_0^2}} \cdot e^{j\varphi}\right)$$  \hspace{1cm} (3)

which has the same shape of the VNTF, except for a scale factor.

This RF is used herein as a functional spatial vision model and, for this reason, it is referred to as Virtual Receptive Field (VRF). For short, the term “VRF” will be used in the following to indicate also its PSF and its HAF shape, whereas its Fourier transform $H(\rho, \theta)$ will be referred to as Virtual Neural Transfer Function (VNTF), because it represents the spatial frequency response of the VRF.

The magnitude and the real components of the VRF are displayed in the upper row of Fig. 1, where the ideal retinal grid is shown in the background. In the same figure, the magnitude, the real and the imaginary parts of the VNTF are displayed in the lower row.

The magnitude of the VNTF frequency response versus the radial frequency is the same in any orientation. It is displayed in Fig. 2. The VNTF is a simple model of the eye contrast sensitivity function.

Here, and from now on, $s_0$ is assumed equal to $s_0 = 2.5$ arcmin, unless otherwise noted. This choice sets the maximum of the radial frequency response magnitude at about 8.5 cycles/degree in the radial frequency, according to the experimental data provided in (Campbell & Green, 1965 & Williams, 1985).

In the lowest spatial frequency range, the magnitude of the VNTF increases linearly. At higher spatial frequencies, the VNTF exhibits a soft-decaying low-pass behavior, reaching an attenuation of about 40 dB at 30 cycles/degree at the Nyquist frequency.

This behavior can be interpreted by regarding the VNTF as the cascade of two basic operators:

- an orientation selective complex gradient operator $\partial_x + j\partial_y$ (Reisert & Burkhardt, 2008), whose frequency response is obtained by the Fourier transform derivation rule:

$$j2\pi f_1 + j(2\pi f_2) = j(2\pi \rho \cos \theta + 2\pi \rho \sin \theta) = j2\pi \rho e^{j\theta};$$  \hspace{1cm} (4)

Fig. 3. The luminance of an image (left) of the database LIVE (Sheikh et al., 2006) compared to the corresponding visual map $y(p)$ (right), where the magnitude of the edges is coded into luminance, and their direction, in the interval $[0, \pi)$, into hue. The hue/direction code is read in the edges of the upper-left superimposed polygonal.
• a radial frequency selective Gaussian smoothing operator, represented by the frequency response:

\[ G(\rho, \theta) = e^{-j\rho^2} \]  

which is responsible of a neural blur.

Therefore, the visual map \( y(p) \) of the VRF is globally interpreted as a complex, Gaussian-smoothed gradient field associated to the retinal image. The parameter \( s_0 \) will be referred to as the spread of the VRF, or as the neural spread.

Different from the complex Gabor functions, whose paired real components are aligned each other (Daugman, 1993; Watson & Solomon, 1997), the paired real components of the VRF are geometrically orthogonal in the image plane (Fig. 1). As such, the VRF is steerable (Simoncelli, Freeman, Adelson, & Heeger, 1992) and more specifically scalar steerable, i.e., it rotates in azimuth by multiplication by a complex number as follows

\[ h(r, \varphi - \alpha) = h(r, \varphi)e^{i\alpha} \]

\[ H(\rho, \theta - \alpha) = H(\rho, \theta)e^{i\alpha} \]  

This implies that rotated version of the VRFs by a generic azimuth \( \alpha \) are obtained by linear combinations of the real components \( \text{Re}(h(r, \varphi)) \) and \( \text{Im}(h(r, \varphi)) \):

\[ \text{Re}(h(r, \varphi - \alpha)) = \text{Re}(h(r, \varphi)) \cos \alpha - \text{Im}(h(r, \varphi)) \sin \alpha \]

\[ \text{Im}(h(r, \varphi - \alpha)) = \text{Re}(h(r, \varphi)) \sin \alpha + \text{Im}(h(r, \varphi)) \cos \alpha \]  

The VRF model materializes into the visual map \( y(p) \). In the correspondence of an edge of \( I(p) \), the magnitude of \( y(p) \) measures the edge strength, while the phase \( e^{-\frac{1}{\psi(p)}} \) indicates the orientation orthogonal to the edge (Jacovitti, 1991; Jacovitti & Neri, 1995; Mellihaia & May, 2012). In the example of Fig. 3, the luminance of a retinal image (left) and its visual map in false color (right) are displayed. In the visual map the gradient strength is indicated by the luminance component whereas, for visual immediateness, only the direction of the gradient in the interval \([0, \pi]\) is indicated with the hue color component.

The visual map \( y(p) \) is a near-complete, sparse representation of \( I(p) \). In fact, except for its mean value, \( I(p) \) can be fully recovered from \( y(p) \) by spectral inversion (i.e., by division by \( H(p, \theta) \) in the frequency domain).

Standard multichannel spatial vision models follow a tomographic-like approach. Around any point, they analyze the image from a limited number of azimuthal views. For each view, they apply co-ordinated filters tuned to different bands. Their outputs are then combined in different ways. See (Schutt & Wichmann, 2017) for a historical account.

In comparison, around any point, the VRF performs a full-band radial tomographic analysis in every orientation, as described in (Cassani & Jacovitti, 1989; Jacoviti & Cassani, 1990). For the scope of this work, the outstanding advantage of the VRF is that it outputs the visual map) allows straightforward computation of the PFI (Neri & Jacovitti, 2004) as described in the next section.

The VRF is the simplest functional spatial vision model based on HAFs. It coincides with the first order component of the orthogonal family of the Laguerre Gauss (LG) functions (Jacoviti & Neri, 2000; Victor & Knight, 2003; Massey & Regefrier, 2005) or, equivalently, of the 2D Hermite functions, which span the same signal space (Martens, 1990; Di Claudio, Jacovitti, & Laurenti, 2011). Higher order LG analysis provides functional spatial vision models oriented to structures more complex that simple edges (Neri & Jacovitti, 2004; Di Claudio, Jacovitti, & Laurenti, 2010). HAF based wavelets can be also used for multi-resolution analysis (Jacoviti & Neri, 2000).

### 3. The Positional Fisher Information

A detail \( d_p(q) \) of a visual map \( y(p) \) around \( p \) is defined as

\[ d_p(q) = w_p(q) \cdot y(q - p) \]  

where \( w_p(q) \) is a sampling window.

A comprehensive calculus of the Fisher Information of a detail about its position, orientation, and scale, in the presence of a background Gaussian white noise, was provided in (Neri & Jacovitti, 2004). As specified in (Di Claudio et al., 2010), the total PFI of a detail is calculated as:

\[ \psi(p) = \frac{\lambda'(p)}{\sigma^2} \]  

where \( \lambda'(p) \) is the smoothed gradient energy of the detail, computed as

\[ \lambda'(p) = \sum_q w_p(q) |y(q - p)|^2 \]  

and \( \sigma^2 \) is the variance of the background noise.\(^1\) The inverse square root of the PFI

\[ \sqrt{\psi(p)} = \frac{1}{e_{\text{MIN}}(p)} \]  

represents the minimum standard deviation of the detail position error \( e_{\text{MIN}}(p) \), achievable with an unbiased estimator, irrespective of the employed estimation method (Van Trees, 1992). Thus, the quantity

\[ \sqrt{\psi(p)} = \frac{1}{e_{\text{MIN}}(p)} \]  

measures the certainty of the detail position in the visual plane. For a given amount of background noise, the higher the smoothed gradient energy, the higher the PFI, the greater the certainty about the detail position.

The smoothed gradient energy \( \lambda'(p) \) is significantly expressed by the (two-dimensional) Fourier transform of the detail. Applying the Parseval theorem (which equates the energy calculated in the space and in the spatial frequency domains) the PFI of the detail represented by its Fourier transform \( D_p(\rho, \theta) \) is

\[ \psi(p) = \frac{1}{\sigma^2} \int_0^{2\pi} \int_0^{\infty} p^2 |G(\rho, \theta)|^2 |D_p(\rho, \theta)|^2 |B(\rho, \theta)|^2 p \rho \rho d\theta \]  

where \( B(\rho, \theta) \) is the Optical Transfer Function (OTF), i.e., the spatial frequency response of the optical system, from the observed object to the retina (Watson, 2013). Its inverse 2D transform is the Optical PSF \( b(r, \varphi) \).

The overall OTF of a vision system, including the human eye, is a function of the OTFs of cascaded subsystems including the OTF of correcting lenses, the OTF of an imaging system, the OTF of a display system, etc.

Under the hypothesis of linearity, the overall OTF is the product of the single OTFs. In other terms, the overall Optical PSF is the cascaded 2D convolution of the single PSFs.

\(^1\) Strictly speaking, the PFI of a detail does not coincide with the PFI of the pattern “contained in” the detail. In fact, the window itself carries its own PFI. In the following, the latter contribution will be neglected, assuming that the window is so smooth that the information carried by its shape is small with respect to the information carried by the captured pattern.
4. The natural scene spectrum and the PFI equivalence

4.1. The PFI of natural scenes

For a generic image, the average PFI calculated on a group of N details visited during the visual exploration is:

\[
\frac{\sum_i \mathcal{J}(p)}{N\sigma_v^2} = \frac{1}{\sigma_v^2} \int_0^{\pi^2} \int_0^{\pi^2} \rho^2 |G(p, \theta)|^2 |D_N(p, \theta)|^2 |B(p, \theta)|^2 \rho d\rho d\theta
\]

where

\[
|D_N(p, \theta)|^2 = \frac{1}{N} \sum_\rho |D_p(p, \theta)|^2
\]

is the average energy spectrum of the \(N\) visited details. The expected value \(\Psi\) of the PFI is defined as:

\[
\Psi = \mathbb{E}\left\{ \sum_i \mathcal{J}(p) \right\} = \frac{1}{\sigma_v^2} \int_0^{2\pi} \int_0^{2\pi} \rho^2 |G(p, \theta)|^2 \mathbb{E}\{|D_N(p, \theta)|^2 \} |B(p, \theta)|^2 \rho d\rho d\theta
\]

where \(\mathbb{E}\{|D_N(p, \theta)|^2 \}\) denotes the expected value of the energy spectrum of the visited details over the random set of natural images.

It is well known that the expected value of the energy spectrum of natural images is proportional to the inverse of the square of the radial frequency, i.e. to \(\frac{1}{\rho^2}\). The generality of this spectral distribution is supported by theoretical arguments (Van Der Schaaf & Van Hateren, 1996; Bell & Sejnowski, 1997; Field & Brady, 1997; Simoncelli & Olshausen, 2001; Graham, Chandler, & Field, 2006; Kuang, Poletti, Victor, & Rucci, 2012).

Here, this property is attributed to the average spectrum of the visited details. Posing, for the sake of generality (Torralba & Oliva, 2003)

\[
\mathbb{E}\{|D_N(p, \theta)|^2 \} = f(\theta) \frac{1}{\rho^2}
\]

it follows that

\[
\Psi = \frac{1}{\sigma_v^2} \int_0^{\pi^2} f(\theta) \int_0^{\pi^2} \rho^2 |G(p, \theta)|^2 |B(p, \theta)|^2 \rho d\rho d\theta
\]

and, in the absence of blur,

\[
\Psi_0 = \frac{1}{\sigma_v^2} \int_0^{\pi^2} f(\theta) \int_0^{\pi^2} |G(p, \theta)|^2 \rho d\rho d\theta.
\]

In the case of isotropic Gaussian blur, the OTF is

\[
B(p, \theta) = e^{-2\rho^2}
\]

where \(s_0\) is referred to as the optical spread. Therefore,

\[
\Psi = \frac{F}{\sigma_v^2} \int_0^{\pi^2} e^{-2\rho^2} \rho d\rho
\]

\[
\Psi_0 = \frac{F}{\sigma_v^2} \int_0^{\pi^2} e^{-2\rho^2} \rho d\rho
\]

where the coefficient \(F\) is

\[
F = \int_0^{\pi^2} f(\theta) d\theta.
\]

Finally, from the equality

\[
\int_0^{\pi^2} e^{-2\rho^2} \rho d\rho = \frac{1}{4(s_0^2 + s_1^2)}
\]

Fig. 4. The radius \(R\) of the sinc blur versus the spread \(s_0\) of its PFI equivalent Gaussian blur. The red circles indicate some average results of experiments conducted with synthetic images emulating natural images (Murray & Bex, 2010).

it follows that

\[
\Psi = \frac{F}{\sigma_v^2} \int_0^{\pi^2} e^{-2\rho^2} \rho d\rho = \frac{F}{\sigma_v^2} \int_0^{\pi^2} e^{-2\rho^2} \rho d\rho.
\]

4.2. The PFI equivalence

The above result is so simple owing to the Gaussian shape of the blur. However, having assumed that blur discomfort depends only on the PFI loss, the discomfort due to different types of blur could be predicted by the same formula applying the concept of PFI equivalence.

Definition 4.1. (PFI Equivalence) A blur characterized by the generic OTF \(B(p, \theta)\) is said to be PFI equivalent to an isotropic Gaussian blur with standard deviation (spread) \(s_0\) if it yields the same expected PFI for natural images, i.e.:

\[
\int_0^{2\pi} \int_0^{\pi^2} e^{-2\rho^2} |B(p, \theta)|^2 \rho d\rho d\theta = \int_0^{2\pi} \int_0^{\pi^2} e^{-2\rho^2} \rho d\rho d\theta.
\]

This equivalence criterion is intuitive. It equals the energies of the actual OTF and of an isotropic Gaussian OTF, both weighted by the squared magnitude of the VNTF. In particular, the PFI equivalence does not depend on the phase of the OTF.

An important example of PFI equivalence is the one of the out-of-focus blur, whose PSF is modeled as a cylinder of unitary volume and radius \(R\). It is referred also to as the disc blur, or the sinc blur, and is characterized by the following OTF:

\[
B(p, \theta) = \frac{2}{\pi R} J_1^2(2\pi R p) .
\]

Equating the PFI of the sinc blur and of the Gaussian isotropic blur yields

\[
\frac{8}{\pi R} \int_0^{\pi^2} e^{-2\rho^2} \rho d\rho = \int_0^{\pi^2} e^{-2\rho^2} \rho d\rho = \frac{1}{4(s_0^2 + s_1^2)}.
\]

The first integral is not available in closed form. A careful numerical integration provides a value of the optical spread \(s_0\) of the isotropic Gaussian blur as a function of its PFI equivalent sinc blur of radius \(R\), as plotted in Fig. 4 and is roughly expressed by the rule \(\frac{s_0}{s_1} \approx \frac{2}{3}\). In the same figure some subjective equivalence judgments averaged over a pool of six observers are also reported. These empirical data refer to synthetic...
images whose contours emulate the ones of natural images (Murray & Bex, 2010).

From a perceptual viewpoint, it appears that this equivalence works generally well. Here, to provide the reader with visual examples, the most critical cases in the Tampere image database (TID2013) (Ponomarenko et al., 2015), are reported in Fig. 5, including the non-natural image i25 as a benchmark. Notice that some grating patterns are cancelled out or amplified by the sidelobes of the OTF of the sinc blur in comparison to the Gaussian blur. Notice also that these effects are not present in the natural out-of-focus blur, owing to the apodization of the pupil, which attenuates the sidelobes (Zhang, Ye, Bradley, & Thibos, 1999).

A second example of PFI equivalence regards the non-isotropic Gaussian blur, referred to as astigmatic Gaussian blur. For the sake of simplicity, this equivalence is calculated here only for the case of isotropic image spectral energy distribution. Using for convenience the Cartesian coordinates, the OTF of this blur is:

\[ B(f_1, f_2) = e^{-2\pi i (f_1 s_H + f_2 s_V)} \]

where \( s_H \) and \( s_V \) are the horizontal and vertical optical spreads. A straightforward algebraic analysis shows that the astigmatic Gaussian blur is PFI equivalent to the isotropic Gaussian blur with spread

\[ s_B = \sqrt{s_H^2 + s_V^2 (s_H^2 + s_V^2) + s_H^2 s_V^2 - s_0^2} \]

An example is provided in Fig. 6. Two versions of an original image, respectively affected by an astigmatic Gaussian blur with \( s_H = 4, s_V = 1 \) and with \( s_H = 1, s_V = 4 \) are compared to the same image affected by their PFI equivalent isotropic Gaussian blur. Notice that the sea waves are better localized in presence of horizontal blur, while masts are better localized in presence of vertical blur. The average localizability loss is visually balanced in the case of the isotropic Gaussian blur.
5. Prediction of blur discomfort

Provisionally, the amount of blur discomfort is assumed proportional to the relative certainty loss about the details position defined as:

\[
\varepsilon = \frac{\sqrt{\Psi_0} - \sqrt{\Psi}}{\sqrt{\Psi_0}} = 1 - \frac{\Psi}{\Psi_0}. \tag{31}
\]

This assumption is suggested by the belief that the perceived cost of wrong localization is proportional to the uncertainty of the Euclidean distances, at least for small errors.

Then, for the random set of natural images and for isotropic Gaussian blur, using (25) the relative certainty loss takes the form of the following non-dimensional discomfort index:

\[
\varepsilon = 1 - \sqrt[2]{1 + \left(\frac{s_B}{s_G}\right)^2}.
\]  \tag{32}

This index ranges between 0 (in the absence of blur) to 1 (for diverging blur). It depends only on the normalized blur, i.e., the ratio between the optical spread on the retinal image and the neural spread of the VRF.

The optical blur spread \(s_B\) is subject to change by the action of the natural accommodation system and, in the context of a composite optical system, by the action of technical devices. The neural spread \(s_G\) plays instead the role of an inner reference. At glance, it could be argued that it is a stable parameter. However, some experiments indicate that the \(s_G\) value is adaptive (Webster, 2002). It appears that the visual adaption to a blurred image causes a dilation of the spread \(s_G\), leading to a reduction of the normalized blur, so that the spectrum of the image looks wider and the image sharper.

The discomfort formula is now applied to the blur caused by out-of-focus condition of the eye optics, i.e., in natural vision. Using the geometrical arguments of (Strasburger, Bach, & Heinrich, 2018), it is deduced that the radius \(R\) of the out-of-focus optical PSF in arcmin is

\[
R = 1.7\pi p |D| \tag{33}
\]

where \(p\) is the pupil diameter (in mm) and the \(D\) is the out-of-focus...
measure in diopter units (m⁻¹).

Applying the \( \frac{1}{s} \approx \frac{1}{s_p} \) rule it follows that the spread of the PFI equivalent blur of the out-of-focus blur measured in diopters is approximated as

\[
s_b \approx 0.64\rho|D|
\]

(34)

so that a coarse estimate of blur discomfort in natural vision is:

\[
\varepsilon = 1 - \frac{1}{1 + \left(0.64\frac{\rho}{D}\right)^2}.
\]

(35)

In the chart of Fig. 7 this theoretical discomfort index \( \varepsilon \) is plotted versus \( D \) for different pupil diameters. The scale of \( \varepsilon \) is expressed in centesimal units.

This chart essentially says that, if the out-of-focus discomfort is proportional to the relative certainty loss, as assumed, it is not linear with the blur spread. Moreover, since the sensitivity of the discomfort index with respect to the normalized spread \( \xi = \frac{s_b}{s} \) is calculated as

\[
\frac{d\varepsilon}{d\xi} = \xi \left[1 + \frac{1}{1 + \xi^2}\right]^{\frac{1}{2}}
\]

(36)

the increment \( \Delta \xi \) necessary to produce a given increment \( \Delta \varepsilon \) is approximated as:

\[
\Delta \xi = \frac{1}{\xi^2 + 1} \frac{1}{\xi^2 + 1} \Delta \varepsilon.
\]

(37)

This increment exhibits a typical “dipper shape” (Solomon, 2009; O’Hare & Hibbard, 2011) as shown in Fig. 8. The theoretical minimum occurs in the correspondence of the normalized spread value \( \xi = \frac{s_b}{s} = \frac{1}{\sqrt{2}} \).

In the above formulas, diffraction and aberration contributions to blur discomfort are considered negligible. Non-negligible aberrations could be accounted for by adopting parametric models for characterizing their PSFs (Goodman, 2005; Watson, 2015), and calculating the PFI equivalent blur. A simple example is the aberration due to astigmatic Gaussian blur provided in Section 4, where the blur is characterized by the two parameters \( s_h \) and \( s_v \). However, broad generalizations would be quite an undertaking beyond the scope of this paper.

6. Comparison with empirical data

The above results about the phenomenon of blur discomfort for natural images are derived by principles and assumptions. To assess their effectiveness in the reality, the theoretical model was first verified against empirical ratings of the subjective quality loss of images caused by blur, which are argued to be strictly related to blur discomfort. These data were released in response to the growing demand by the media industry for reliable automatic image quality assessment (IQA) through objective metrics. Four independent IQA databases were employed, based on different methodologies and protocols, and following different strategies to prevent biases and side effects.

IQA databases include images affected by Gaussian blur, which is considered sufficiently representative of the perceptual effect of the actual blur in many technical applications.

Subsequently, the predictions of the model were compared to experimental data where subjects were literally asked to rate “the visual discomfort” due to blur (O’Hare & Hibbard, 2013).

All these experiments do not account for blur discomfort secondary effects, such as psychological consequences that lie behind the scope of the paper. Furthermore, in these experiments the eye optical blur of the observers is corrected if present. So, the blur applied to the observed images emulates an undesired natural optical blur on the retina plane. In fact, under the hypothesis of linearity, the natural optical blur and the artificial blur applied to the observed images are interchangeable, because of the well-known property of cascaded convolutions.

Before presenting the experiments and discussing the results, it is essential to illustrate how the discomfort index must be scaled to fit the experimental settings of a database.

6.1. The Scaled Blur Discomfort Index

To this purpose, the following parametric “Scaled Blur Discomfort Index” (SBDI) is defined from (32):

\[
SBDI=\alpha \left[1 - \frac{1}{1 + \left(\frac{\rho}{D}\right)^{\gamma}}\right]
\]

(38)

where the gain \( \alpha \) fixes the scoring scale (Sheikh & Bovik, 2006), and \( \gamma \) is a distance parameter defined as

\[
\gamma = \frac{\delta_0}{\delta}
\]

(39)

where \( \delta_0 \) is nominal viewing distance, i.e., the distance from which the density of pixels projected on the retina matches the previously assumed density of the receptors (60/degree) and \( \delta \) is the viewing distance adopted in the experiment.

Differently stated, \( \gamma \) equals the ratio between the number of pixels viewed within one degree at distance \( \delta_0 \) (60 pixels) and the number of pixels viewed within one degree at distance \( \delta \).

The role of the distance parameter is understood considering that the VRF spread on the image projected on the retina is \( \frac{\omega}{\theta} \), i.e., it increases proportionally with the viewing distance. Conversely, the projection of the spread of the blur applied to the observed image is \( \omega_{\text{blur}} \), i.e., it is inversely proportional to the viewing distance.

If not specified, the parameters \( a \) and \( \gamma \) can be determined from data by regression.

6.2. The “subjective quality loss” experiments

The essential features of the employed IQA databases are illustrated below.

The LIVE Image Quality Assessment Database Release 2 (DBR2) (Sheikh, Sabir, & Bovik, 2006) reports the quality ratings of 779 distorted versions of 29 reference images (included 145 blurred images) from about 23 subjects. Ratings of subjective quality loss with respect to
reference images were expressed on a DMOS (Difference of Mean Opinion Score) scale ranging from 0 (perfect quality) to 100 (absymal quality) using a double stimulus strategy.

The Tampere Image Database 2013 (TID2013) (Ponomarenko et al., 2015) contains 3000 distorted images, including 125 blurred images. Quality ratings were collected in five independent labs and on the internet using more than 300 subjects. They were asked to select the best image between two distorted images in direct comparison to the reference image. The average quality scores were expressed on a Mean Opinion Score (MOS) scale ranging from 0 (bad quality) to 9 (perfect quality).

The Computational and Subjective Image Quality Database (CSIQ) (Larson & Chandler, 2010) contains 30 reference images and 866 distorted versions, including 150 blurred images. The database includes 5000 ratings of 25 subjects, obtained by comparative ratings between different images, reported in DMOS units.

The LIVE Multiply Distorted Image Quality Database (LIVE MD) (Jayaraman, Mittal, Moorthy, & Bovik, 2012) contains 15 reference images and 405 distorted images, including 45 blurred images, whose quality was rated by 37 subjects. The study was conducted using a single stimulus with hidden reference strategy, using DMOS scores.

In the TID2013 database, ratings are available as MOS values. The DMOS values were inferred considering that the best MOS ratings do not exceed 7.5, as seen from Fig. 20 of (Ponomarenko et al., 2015). Therefore, posing $\text{MOS} = 7.5$ in correspondence to $\text{DMOS} = 0$ and $\text{MOS} = 0$ in correspondence to $\text{DMOS} = 100$ yields

$$\text{DMOS} = \left(\frac{100}{7.5}\right) \times (7.5 - \text{MOS}).$$

In the CISQ database, the DMOS was normalized between its minimum and maximum empirical value.

The blur values of the LIVE MD database were not available. They were estimated through a regularized spectral division of the blurred images with the unblurred ones.

The fitting of the empirical data with the theoretical prediction of the subjective quality loss is substantially linear, as evidenced by the scatterplots of the DMOS empirical ratings versus the predicted ones for all the images contained in the different databases (Fig. 10). In the TID2013 and LIVE MD scatterplots, the averages with respect to the sample images are indicated by filled circles.

In Table 1 the most relevant data about the experimental validation of the model are resumed. The Pearson Linear Correlation Coefficient (PLCC) and the Root Mean Square Error (RMSE) between theoretical and empirical data are also provided. The “claimed” normalized viewing distances were calculated with the information about the experimental settings provided by the respective authors. The sign “–” stands for “not available.”

3 The viewing distances of CISQ and LIVE MD are slightly (10%) underestimated. The maximum absolute difference is within ten centimeters. However, the “physical” viewing distance does also depend on the physical dimension of the pixels of the screen.

4 The Authors are grateful to Prof. Ponomarenko for providing details about the blur settings in the TID2013 database in a personal communication.
6.3. The "blur discomfort" experimental data

In (O’Hare & Hibbard, 2013) the results of some experiments aimed to investigate the relationship between visual discomfort judgments and image manipulations causing blur are reported. In particular, “Experiment 3” regards natural images.

In these experiments sixty natural images were taken from a database whose images pertain to two general categories: distant natural scenes and closeups natural scenes. In particular, ten images from the first category whose spectral energy has a mean energy spectral radial decay of the kind $\hat{\beta} = 1.39$, and ten images from the second category, characterized by $\hat{\beta} = 0.95$, were selected. The scope of this diversity was to see if visual discomfort judgments would depend on deviation from the ideal spectral slope for natural scenes ($\hat{\beta} = 1$).

During experimental sessions, thirteen subjects were asked to formulate discomfort judgments following a pairwise comparison strategy. Differently from the preceding experiments, ratings are averaged not only over the pool of observers, but also over groups of images. One average regards the distant natural scenes, and the other one regards the closeups natural scenes. As in the TID2013 and in the LIVE MD experiments, few (three) Gaussian blur values were employed, with standard deviations 8, 16, and 32 cycles/degree in the spatial frequency domain, corresponding respectively to $s_B = 3.75$, $s_B = 1.875$, and $s_B = 0.9375$ arcmin in the spatial domain. In the Fig. 11 the results of these experiments are reported, along with the discomfort predicted by the SBDI index. The results were expressed in the Thurstone scale (Tsukida & Gupta, 2011), according to the method employed for calculating the discomfort scores, and then converted into the discomfort scale of the SBDI by the following affine transformation:

$$SBDI = 50(\text{Thurstone}+1).$$ (41)

The normalized viewing distance, determined by regression, is $\frac{\hat{\delta}}{\delta_0} = 0.6$, corresponding to $\gamma = 1.66$.

The scatterplots of the discomfort ratings versus the normalized

---

Table 1

<table>
<thead>
<tr>
<th>DATABASE</th>
<th>$a$ estimated</th>
<th>$\frac{\delta}{\delta_0}$ claimed</th>
<th>$\frac{\delta}{\delta_0}$ estimated</th>
<th>PLCC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIVE DBR2</td>
<td>93</td>
<td>0.46/0.57</td>
<td>0.53</td>
<td>0.96</td>
<td>5.44</td>
</tr>
<tr>
<td>TID2013</td>
<td>80</td>
<td></td>
<td>0.44</td>
<td>0.92</td>
<td>6.82</td>
</tr>
<tr>
<td>CSIQ</td>
<td>98</td>
<td>0.65</td>
<td>0.60</td>
<td>0.97</td>
<td>7.47</td>
</tr>
<tr>
<td>LIVE MD</td>
<td>107</td>
<td>0.84</td>
<td>0.76</td>
<td>0.83</td>
<td>8.57</td>
</tr>
</tbody>
</table>

Fig. 10. The scatterplots of the DMOS ratings versus the corresponding SBDI values predicted by the model for each blurred image in the different databases. Averages with respect to images are indicated by filled circles in the TID2013 and LIVE MD scatterplots.
Gaussian blur spread are displayed in Fig. 11 for the groups of images contained in the database, along with the superimposed theoretical prediction curve. The triangles refer to the distant natural scene category, and the circles to the closeups scene category.

The fitting of the empirical data with the theoretical prediction of the subjective quality loss is substantially linear, as evidenced by the scatterplots of the empirical ratings versus the predicted ones for all the images contained in the database (Fig. 12).

The Pearson Linear Correlation Coefficient (PLCC) between the average ratings and the predicted SBDI is 0.98. These results support the argued strict relationship between the subjective quality loss ratings and the visual discomfort ratings caused by blur.

7. Remarks

From the IQA experiments, the judgments expressed by pool of observers for the different natural images appear moderately scattered around the SBDI curve, as indicated by the Root Mean Square Error values of Table 1. This fact is relevant, considering the large variety of the images contained in the databases. This relatively small dispersion can be explained by the adaptive nature of the image exploration, which tends to focus on the details of prominent interest (Baddeley & Tatler, 2006).

This appears even more evident looking at the results of the “blur discomfort” experiments, whose purpose was to verify if the discomfort judgments depend on the deviations from ideal statistics of natural images. The average ratings for selected categories of images characterized by markedly different spectral slopes are so close each other (see Fig. 12), that in (O’Hare & Hibbard, 2013) the authors concluded that the difference among these spectral slopes is not important to blur discomfort judgments.

This evidence underlines the validity of the attribution of the ideal spectrum of natural images to the subset of the visited details. This assumption, made in Section 4, supports the approach followed herein.

However, the image content still influences blur discomfort, as shown by the said moderate, but not negligible, scattering of the subjective ratings in the IQA databases, and as suggested by the common experience. The identification of essential quantitative attributes of images producing different blur discomfort in the presence of the same amount of optical blur is a matter of current investigation.

8. Conclusion

The adoption of a polar separable, complex-valued receptive field model, and of a visual information loss criterion, led to compact theoretical formulas for the prediction of blur visual discomfort for natural scenes, exhibiting a good predictive power faced to several independent experimental data.

- From a general scientific viewpoint, the paper demonstrates that blur discomfort is well correlated with the loss of fine localizability of the observed patterns and with the perceived loss of quality.
- From a system theoretical viewpoint, the approach provides a quantitative prediction of the average psycho-physical findings irrespective of the underlying mechanisms.
- From a technical viewpoint, the results presented here may provide coarse estimates of the discomfort caused by undesired blur, for optical correction in natural vision, and for calibration of image reproduction apparatus.

Finally, these results could help in developing IQA methods.5

CRediT authorship contribution statement

Elio D. Di Claudio: Validation, Writing - review & editing, Supervision. Paolo Giannitrapani: Software, Formal analysis, Visualization. Giovanni Jacovitti: Conceptualization, Methodology, Writing - original draft.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the

---

5 For instance, some existing IQA methods model the local smearing of images originated from coding for compression as local blur (Kayargadde & Martens, 1996; Sheikh & Bovik, 2006; Li, Zhang, Ma, & Ngan, 2011; Di Claudio & Jacovitti, 2018). The model developed here may be employed to estimate the contribution of the blur of single details to the overall perceived quality loss.


