Signal Processing: Image Communication 75 (2019) 64–75

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Contrast sensitivity in images of natural scenes☆

Sophie Triantaphillidou*, John Jarvis, Alexandra Psarrou, Gaurav Gupta

Computational Vision and Imaging Technology, Dept. Computer Science, University of Westminster, W1W 6UW, London, UK

ARTICLE INFO

Keywords:
Contrast sensitivity function
Image quality
Contrast detection
Image quality modeling
Visual modeling

ABSTRACT

The *contrast sensitivity function* (CSF) characterizes spatial detection in the human visual system and is typically measured from simple, synthetic stimuli. We used spatial frequency decomposition, RMS contrast modulation, a yes/no paradigm and an adaptive staircase to measure *isolated* and *contextual* CSFs (iCSFs and cCSFs) from natural images. We employed Barten's mechanistic model and adapted it for contextual modeling purposes by postulating that, signal detection in a given frequency band, when presented amongst other broadband signals, can be modeled as if amongst noise. We found that the iCSF varies with pictorial content, but that the standard CSF model and the image's contrast spectrums are sufficient to predict with relative success the cCSF for any given image. We finally discuss the suitability of cCSF models in image quality modeling.

1. Introduction

In spatial vision the *contrast detection threshold* indicates the minimum contrast required to detect a target and varies with spatial frequency. Contrast detection thresholds are important in the perception of the image fidelity and image quality. The inverse of contrast threshold is a measure of visual spatial *sensitivity*. The *contrast sensitivity function* (CSF) is obtained by plotting sensitivity as a function of spatial frequency. The measurement and modeling of the CSF, as well as its implementation in image fidelity and quality models have been researched for many decades [1–6]. The majority of relevant work is focused on CSFs derived from simple test charts employing sine-wave gratings. Whilst the resulting functions reveal contrast sensitivity to individual spatial frequencies, they do not describe the limits in contrast perception when free-viewing complex natural images. In this paper we present the measurement of CSFs from displayed images of natural scenes and the use of Barten's detection model [5] as the basis for modeling the observers' responses.

We measure directly from images of complex scenes two CSFs: (i) the *isolated contrast sensitivity function* (iCSF), which describes the ability of the *human visual system* (HVS) to detect contrast in a given spatial frequency band in isolation; and (ii) the *contextual contrast sensitivity function* (cCSF), which describes the ability of the HVS to detect contrast in a given frequency band, when the band contrast is embedded within the remaining visible contrasts of the image. One purpose of such measurements is to provide insight into the relationship between physical contrasts and just perceptible contrasts available in image viewing. Another purpose is to link threshold contrast perception of isolated band-limited signals to that of band-limited signals interspersed with other visible signals, and therefore to determine the role the former have in the perception of the latter. The functions should provide a basis for models of spatial vision more suitable for imaging applications.

In Section 2 of the paper we review definitions of contrast, threshold contrast perception, contrast masking and the role of the CSF in image fidelity and quality models. Barten's mechanistic model of the CSF, and the modeling of the iCSF and the cCSF are presented in Section 3. Section 4 discusses visual stimulus specification, capture, processing and display. In Section 5 we introduce the experimental paradigm and describe the visual experiments. Section 6 presents results from measuring and modeling the iCSF and the cCSF from a number of image stimuli, and in Section 7 we discuss our findings. Appendices A and B include additional cCSF modeling information and illustrate our test images.

2. Background

2.1. Contrast definition and metrics

Perceptual contrast is defined as a perceived luminance variation. In practice, it is not the absolute luminance variation that is important in contrast perception, but the relative difference. This can be expressed as a ratio between two luminances (contrast ratio), or as a difference between the luminance divided by their sum (contrast). In simple stimuli, such as sine-wave gratings and other periodic patterns, contrast is measured by Michelson's formula, with values ranging between 0 and

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.image.2019.03.002.

* Corresponding author.
E-mail address: triants@westminster.ac.uk (S. Triantaphillidou).

https://doi.org/10.1016/j.image.2019.03.002
Received 8 May 2018; Received in revised form 12 November 2018; Accepted 4 March 2019
Available online 21 March 2019

1.0 [7]. In uniform luminance stimuli against a uniform background, it is measured using the Weber fraction [8]. Michelson's method assumes that the observer is adapted to the sum of the background and foreground, whereas Weber's method assumes that the observer is adapted to the background luminance. This assumption does not hold for complex images. Peli discussed how these two most common metrics of perceptual contrast of simple test stimuli do not coincide and how related types of metrics share analogous problems [9].

Perceptual contrast is probably the most important attribute in natural image viewing [10]. The subjective evaluation of contrast in images is much more complex than that for simple patterns, such as sine-wave gratings. In natural images, contrasts vary significantly with spatial location and are visually *masked* by other contrasts (Section 2.3). There are a number of perceived image contrast metrics proposed in the literature, but none is without pitfalls. A relevant review can be found in [11]. *Root Mean Squared* (RMS) contrast is the prevalent measure and has been used extensively, both in visual and imaging studies [12–17]. RMS contrast has been used differently in various works. Normalization to 0,1 range [9] versus normalization by the mean [12,14] has different implications. Eq. (1) shows the C_{RMS} expression we employed in this work,

$$C_{RMS} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\frac{(L_j - \bar{L})^2}{\bar{L}^2}} \quad (1)$$

where N is the number of pixels, L_j is the displayed luminance of the jth pixel and $\bar{L}$ is the mean luminance. C_{RMS}, when calculated globally, does not account for the spatial distribution of contrast within the image, or for the contrast distribution across different spatial frequencies. It has been employed to measure band-limited contrast after spatial frequency decomposition [13,15,18], and local contrast after image segmentation into local regions, or blocks [12,19].

2.2. Threshold contrast perception and the CSF

Spatial vision relies mostly on the ability to sense luminance variations over space. Much of our understanding of basic, low level visual processing of spatial information is based on visual contrast thresholds, with relevant experiments typically employing test charts with sine-wave gratings, or Gabor patches. Michelson contrast is adjusted until a given frequency grating is at the threshold of contrast detection.

Schade was the first to measure luminance contrast sensitivity as a function of spatial frequency [1] using monochromatic sine-wave gratings. His results suggested that sensitivity varied with the frequency of the grating. A decade later Campbell and Robson revealed the multi-channel neural presence in vision, each channel selective to a different frequency range [2]. The CSF has been researched extensively since — see [5] for an all-encompassing study, and [6,20–22] for reviews.

The typical form of the CSF, measured with sine-wave gratings, displays band-pass characteristics, with peak sensitivity at around 4 *cycles per degree* (cpd). The high frequency decay is shown to be due to the optical limitations of the eye, the spacing of the photoreceptors and noise [5]. The low frequency reduction mechanisms are not universally agreed. They have been attributed to limited receptive field sizes, effects of masking by the DC component of the test chart [23] and lateral inhibition in the retina [5]. The CSF profile varies with luminance level, [5], field size [5], orientation [24,25], eccentricity [26] and chromatic channel [27]. Generally, contrast sensitivity reduces with luminance level, eccentricity, and orientations away from the horizontal and vertical. The latter is known as the *oblique effect* [28]. Variations in the experimental paradigm, stimulus type, or the spatial and temporal presentation of stimuli result in considerable variations in the CSF profile. When the CSF is for example measured with Gabor functions the profile tends to be more low-pass [29]. Gabor patches are of theoretical interest because their structure describes the spatial profile of simple cell receptive fields in the visual cortex (V1) [30]. Contrast sensitivity is reduced with old age, ophthalmic conditions and a number of diseases [31].

2.3. Contrast masking

Spatial sine-wave detection is masked when the stimulus is embedded within other *suprathreshold* (i.e. clearly visible) contrast information. Interference of the contrast of a stimulus by another spatial (or temporal) signal is generally referred to as *contrast masking*. It is a key feature of our visual system. The stimulus contrast effectively increases, or decreases due to the presence of the masking signal.

The impact of masking noise on threshold contrast has been extensively studied. Contrast masking due to noise generally leads to a suppression of the CSF [5,32–34]. Normally, masked detection thresholds increase as the contrast of the mask increases, but there is a point where the opposite effect occurs (facilitation, or *dipper effect*) [35]. Legge and Foley [36] performed a classical study of contrast masking with a 2.0 cpd sine-wave test signal, spatially superimposed upon a suprathreshold masking sine-wave. They derived so-called *dipper functions* (contrast discrimination threshold vs. pedestal contrast, or *TvC* functions [35]). The masking impact on the contrast discrimination of a 2.0 cpd test grating was seen to be maximum when the mask spatial frequency was close to the test frequency, with the effective masking frequency range being about 2 octaves around the test frequency. Their results were described in terms of the *linear amplification model* (LAM) that accounts for contrast detection (i.e. changes in threshold contrasts), contrast discrimination (i.e. changes in suprathreshold contrasts), and masking phenomena [5,37–39]. Cortical mechanisms were evoked in the model, which include a linear spatial frequency filter, a nonlinear transducer and a process of spatial pooling which acts at low contrasts only. Liu and Allebach more recently characterized contrast masking using the contrast in the adjacent visual channels, with their model parameters fitted to visual data obtained for a range of natural texture masks [40].

Contrast detection in natural images is a masked detection, and measured contrast thresholds derived from image stimuli represent masked detection thresholds. Thus, in relevant detection tasks detection, discrimination and masking mechanisms are in play. Two of the most cited psychophysical models of contrast masking in image viewing are Teo and Heeger's [41] and Watson and Solomon's [42]. In our work we infer that *contextual detection* (i.e. band limited contrast detection in the context of other suprathreshold image contrasts) can be modeled with sufficient success for image quality modeling via the LAM implementation. We discuss implementation of the LAM in the modeling of cCSF in later sections.

2.4. Contrast perception determination from natural images

Many have challenged the relevance of traditionally measured threshold CSFs to our understanding of visual processing of natural image stimuli. Unlike sine-wave gratings, the spatial frequency content of complex stimuli force neurons with different frequency response characteristics to respond to the same image location simultaneously, and in a fashion that is not predictable from their responses to sine-wave stimuli [43,44].

Relatively few studies have directly derived visual response functions from images. Peli [45] simulated successfully the appearance of displayed images for three viewing distances, by setting contrast thresholds in a pyramidal vision model of band-limited local contrast [9] using the individual observer's CSF. Although the study did not derive contrast responses directly from images, it showed that the testing method could be used to determine the type of CSF that best represents observer performance in a particular task relating to image viewing. He later repeated the study with 5 contrast versions of the original images, and verified the CSF relevance in image discrimination tasks for a wide range of spatial frequencies [21].

Bex, et al. [13] employed a derivative of the Legge and Foley technique [36] to determine contrast discrimination in images of four natural scenes. In their work, spatially filtered images were presented in

a number of different ways. One modification involved the presentation of a 1-octave wide, spatially band-filtered image of a given scene, where three peak frequencies were chosen in the filtration process. A second experimental modification was made, similar to the band-pass condition, but in that spatial frequencies outside of the band were not discarded, i.e. the band-image was presented in the context of the remaining frequency bands. Dipper functions were obtained. The latter condition revealed less discrimination sensitivity compared with the former condition that was attributed to masking. The filtration technique in this study formed the basis of further experiments with a wider range of frequency bands.

Later, Bex et al. [46] examined how threshold contrast sensitivity varied within the context it was measured. For the purpose they used band-pass filtered noise stimuli, presented on a uniform background of mean luminance, also the same stimuli but with the observer adapted to dynamic natural images (movies). They showed that adaptation to dynamic images causes selective attenuation in contrast sensitivity to low spatial frequencies, a finding that was in agreement with previous studies by Webster and Miyahara [47]. They attributed this loss to tuning to the *1/f* profile of the amplitudes of natural scene spectra, which varied with the slope of the spectra. From a third experimental condition with noise stimuli presented within a natural movie they reported that, the effects of masking within natural images are also low-pass, having little, or no effect at high frequencies.

In a more recent study, Haun and Peli [18] investigated how different spatial frequencies contributed to the perceived contrast of complex broadband images. They experimented with eight spatial frequency bands, centered from 1 to 64 cpd, and a large number of natural scenes. Although the main focus in this study was not determination of visual sensitivity, it did derive empirical threshold contrast functions from four humans and one model observer, using band-pass versions of the test images. Sensitivity measurements covered a frequency range from 0.1 to 16 cpd. They revealed minimum threshold contrasts at 2 and 4 cpd and CSFs having similar profiles to the standard CSF.

2.5. The CSF in image fidelity and quality modeling

The relevance and implementation of threshold contrast sensitivity in image quality and fidelity modeling has been extensively discussed in the literature. Amongst others, Haun and Peli [48], Triantaphillidou et al. [49] and Chandler [23,50] have provided relatively recent reviews on the subject. The most common implementations are, either to weight the signal by applying the CSF as a spatial filter, or to weight the channel sensitivity after channel decomposition so that the sum approaches the target CSF.

Schade and colleagues [1] first considered the human CSF as a component in the imaging chain. The fact that the CSFs can be modeled using signal transfer theory has provided important insights into the related neural mechanisms [5,51]. The same fact gave rise to the evolution of a number of metrics that involve the CSF — mainly sharpness [52–56] and signal-to-noise metrics [57–59], upon which modern sharpness and visual noise measurements are based [60,61]. More advanced visible difference algorithms [62,63], color difference models [64–66] and computational metrics using the CSF [67–70] have also been proposed. Some of the latter methods account for sensitivity according to orientation and also involve masking models.

But limitations in using the CSF in image quality modeling have been observed as long ago as 1976 [71]. Both the linear systems approach and the relevance of threshold contrasts in image quality have been questioned. More than two decades ago Ahamuda and colleagues [72,73] stated that, in image viewing the CSF is largely outweighed by contrast masking, and showed that a model with only within channel masking was greatly improved with a simple contrast energy masking term. Around the same period, Silverstein and Farrell declared that suprathreshold judgments are unrelated to contrast threshold judgments [74]. Since, many more have queried the role of CSFs to metrics evaluating quality [75–78].

In early 2000's Peli [21] suggested that low pass functions that are flat at low frequencies are better suited for quality modeling. During the same period the Modelfest project [6] attempted to derive a standard spatial observer from threshold measurements for application to image fidelity and quality modeling. Watson and his colleagues successfully applied the derived spatial observer in a number of applications. More recently, Haun and Peli [79] argued that estimating the visual quality of an image, contrast thresholds are of principle importance, whilst perceived suprathreshold magnitudes are relatively less important. They also suggest that the specific sensitivity functions commonly employed may be misapplied, or inappropriate for predicting visible differences, and in particular in reference-free quality predictions. Several contemporary models of suprathreshold contrast sensitivity used in image fidelity and quality employ the CSF to set thresholds for nonlinear contrast transducers that converge at high contrasts [67,80], implying that the standard CSF has a role in the viewing of natural images.

Although questions have been asked regarding the role of the CSF in preferential image quality modeling, fewer concerns have been raised regarding its applicability to modeling fidelity, i.e. image differences and the visibility of image artifacts. In fidelity studies, the common assumption is that the artifact is the detection target, which is typically masked by local image contrasts. Our opinion is that contrast threshold models are relevant to both image fidelity and quality, but questions remain on the CSF implementation, also on how the traditional CSF derived from simple stimuli relate to the viewing of complex pictorial imagery.

3. Modeling the CSF

There are numerous models of the CSF that have been employed in quality and fidelity modeling. Some of the most predominant are: Mannos and Sakrison [81], Kelly [4], Daly [62], Movshon and Kiorpes [82], Barten [5,83] and ModelFest [6,84].

For our modeling purposes, we have chosen Barten's model of the luminance CSF [5], not only because it relates directly to neurophysiological mechanisms in the visual system, but also because parameters related to these mechanisms are known and remain stable for any given experimental conditions (see Section 3.2). Barten's mechanistic model takes into account the adaptive nature of the HVS and incorporates information about the viewing distance, the angular display size and the effective display luminance, but not the orientation. Its core assumption is that contrast sensitivity is restricted by internal noise in the HVS. The model considers the CSF as a product of optical and neural factors [85]. It is based on Fourier analysis and signal transfer theory, which strictly cannot be used in non-linear systems. Barten adopts that, although the eye's response is not linear, at the threshold level of detection it can be considered as linear.

3.1. Barten's mechanistic model of the CSF

In the original Barten model [5], sensitivity, *S(u),* the inverse of the contrast threshold $M_t(u)$, is expressed as a function of spatial frequency, u in cpd by:

$$S(u) = \frac{1}{M_t(u)} = \frac{M_{opt}(u)/k}{\sqrt{\frac{2}{T}\left(\frac{1}{X^2} + \frac{1}{X^2_{max}} + \frac{u^2}{N^2_{max}}\right)}\left(\frac{1}{\eta p E} + \frac{\Phi_0}{1-e^{-(u/u_0)^2}}\right)} \quad (2)$$

where $M_{opt}(u)$ is the MTF of the eye, k is the signal-to-noise ratio required for detection , T is the integration time of the eye in seconds. X is the angular size of the object and X_{max} the maximum angular size of the integration area, with all measurements in degrees. N_{max} is the maximum number of integration cycles possible, η the quantum efficiency of the eye, E the retinal illuminance in Troland and p the photon conversion factor, which is dependent upon the wavelength and intensity of the light source. Φ_0 is the spectral density of neural noise,

and u_0 is the spatial frequency threshold that causes lateral inhibition to cease.

The optical MTF of the eye, $M_{opt}(u)$, is calculated according to Eqs. (3) and (4), in which σ is the standard deviation of the eye's line spread function and generally depends on d, the pupil's diameter, in mm, given in Eq. (5),

$$M_{opt}(u) = e^{-2\pi^2 u^2 \sigma^2} \tag{3}$$

$$\sigma = \left[{\sigma_0}^2 + \left(C_{ab} d\right)^2\right]^{1/2} \tag{4}$$

where C_{ab} and σ_0 are constants that describe the increase of σ with increasing pupil size and are set to 0.08 and 0.5 arcmin/mm respectively for observers with normal spatial vision. These values have been derived from several evaluations of the CSF [5]. For a given luminance, L, in cd/m^2, the pupil size is approximated by:

$$d = 5 - 3 \tanh\left[0.4 \log(L)\right] \tag{5}$$

3.2. iCSF and cCSF modeling

The general model in Eq. (2) was implemented and extended, when necessary, to account for the complexities in the visual signals resulting from viewing natural images. For modeling the isolated contrast sensitivity function, iCSF, i.e. the contrast sensitivity to band-limited signals presented in neutral gray backgrounds, the general model in Eq. (2) was implemented without alterations,

$$iCSF(u) = S(u) \tag{6}$$

The values of the parameters in Eq. (2) were set as follows:

X = 12 deg, the minimum angular size of the displayed stimuli (Section 4.3),

$E = L(\pi d^2)/4$ Troland, where L = mean individual stimulus luminance, in cd/m^2 (Section 4.3),

p: 1.1×10^6 photon/s/deg^2/Troland, from table 3.2 in [5, p.63] ,

T: 0.1 s, [5, p.39],

$\Phi_0 = 3 \times 10^{-8}$ s deg^2, [5, p.39]

$k = 2.8$ (Section 5.3)

Values for all other constants are as listed by Barten [5, p.39].

For modeling the contextual contrast sensitivity function, cCSF, i.e. the contrast sensitivity to band-limited signals presented within the context of the remaining image contrasts, the LAM model was implemented. It enabled us to account for the suppression effect caused by masking from the flanking bands, around a given frequency band. According to this interpretation, signal and noise information in the flanking bands are considered simply as masking noise.

The relationship between iCSF and cCSF in Eq. (7) was first revealed from measuring these functions, using the paradigms described in Section 5,

$$cCSF(u) = c_c(u)^{-1} = \left[K c_s(u)^2 + c_i(u)^2\right]^{-1/2} \tag{7}$$

where K is a scene-dependent constant, measured empirically and ranging between 0.025 and 0.015 (Section 6), $c_c(u)$ is the threshold value for contextual contrast detection for a frequency band centered at frequency u, $c_s(u)$ is the contrast spectrum of the image, and $c_i(u)$ is the threshold value for isolated contrast detection — the reciprocal of the iCSF.

In our implementation, masking in the band-limited image is assumed to arise from a frequency range of +1 and −1 octaves from the stimulus (see Appendix A for the LAM implementation, also the paper's discussion section on this point). With this assumption, the LAM, which normally relates the power spectral density of the masking noise to contrast detection, theoretically predicts Eq. (7), as demonstrated in Appendix A [see also reference [34]]. Eq. (7) enables the cCSF to be modeled directly from the iCSF and the contrast spectrum of the image. Eqs. (2) to (7) fully define our modeling framework for describing contrast sensitivities to images of natural scenes.

4. Stimulus processing, specification and display

A number of image processes were employed for developing and displaying a range of natural image stimuli. Computational processes were carried out at double precision floating-point. Image stimuli were saved at 16-bits per channel bit-depth. Below we provide information on the most important processes and any limitations these presented to the accurate representation of the stimuli. We further discuss the acquisition, selection, calibration and display of the stimuli and the characteristics of the relevant imaging devices.

4.1. Spatial frequency decomposition

Image filtering operations were performed in the spatial frequency domain of the luminance image. We employed Peli's cosine log filters [9] of 1-octave bandwidth, centered at frequency 2^i cycles per picture (cpp), to decompose the original image's amplitude spectrum to a number of frequency bands. The filters satisfied our requirements of symmetrical shape on a log frequency axis and the condition of additive reconstruction, i.e. all image information is reconstructed by summing all individual frequency bands, the mean (low) and high residuals. They have a number of convenient properties that make them suitable for representing visual system decomposition and have been employed in a variety of digital imaging applications, examples in [9,18,86–88].

The ith order filter is expressed by:

$$G_i(r) = \begin{cases} \frac{1}{2}\left[1 + \cos(\pi \log_2 r - \pi i)\right], & 2^{i-1} \le r \le 2^{i+1} \\ 0 \quad \text{elsewhere} \end{cases} \tag{8}$$

where r is the radial frequency, $r = \left({f_x}^2 + {f_y}^2\right)^{1/2}$.

We generated ten filters, centered at the corresponding retinal frequencies of 0.125, 0.25, 0.50, 1, 2, 4, 8, 16, 24 and 32 cpd, nine of which covered fully the frequency domain stimulus size at the experimental viewing distance. The 10th, corresponding to 32 cpd, was larger than the frequency range of the display, due to limitations in the pixel size. Although the resulting band represented less than a full 1-octave, it was modulated so as to produce equivalent C_{RMS} steps as the remaining bands.

4.2. Contrast modulation

Band limited images (test signal) were presented to the observers as visual stimuli, either in isolation, or contextually amidst all other available image frequencies. The C_{RMS} (in Eq. (1)) of the test signal was modulated using a sufficiently large number of contrast steps (1000), where the contrast ranged between 0 and 100% C_{RMS} band-limited contrast. The limits were imposed by the psychometric testing method presented in Section 5. The step size determined the lower and upper measurement boundaries.

Digital images have a limited range of luminance values per pixel. Even when the luminance does not consume the full available bandwidth of a digital image, the tonal curve cannot be readily readjusted for luminance accuracy to be maintained. This leaves limited headroom of luminance values available outside of the range contained naturally in the image. An attempt to lower, or boost luminance beyond the range allowed by the bit-depth of the digital image file, can result in clipping. This presents a problem when manipulating frequency bands in an image, where the signal and the context waveforms constructively and destructively interfere with each other. For the full range of amplitude steps to be presented to observers without any band clipping (or band overflow), the contrast spectrum of the original image stimuli had to be adjusted. The contrast was lowered to a fraction of the original image contrast, while maintaining the relative band amplitudes, thus retaining as close as possible the natural spectral characteristics of the source image. This operation resulted in increasing the available headroom for frequency band manipulation and ensured no luminance clipping. Two suitable contrast levels were used in the experimentation, denoted as "normal" and "low" in the following sections.

4.3. Image capture and selection

Careful thought was put in collecting suitable imagery for use as visual stimuli. We captured a large number of scenes and made a careful selection so that the experimental image stimuli varied notably in scene content. A high quality DSLR camera, equipped with a full frame sensor and a professional 50 mm f/1.4 lens was used for a 16-bit per channel RAW image capture, during which the aperture was kept constant at f/8 to ensure relatively consistent camera spatial frequency response. The ISO setting ranged between 200–600, resulting in images with virtually no visible noise. Original scene luminances were carefully recorded for each capture, along with other camera metadata. Captured images were saved as 16-bit per channel, sRGB uncompressed tiff files. The central 1794 by 1196 pixel area was cropped to match the pixel dimensions of the displays used in the visual experiments. Pixel values were converted to display luminance values using a 16-bit per channel LUTs. Our final test set comprised of seven original image stimuli (image names: *gallery, park, people, lines, subway, uni, wharf*), all of which were employed for the derivation of iCSFs, and three for the derivation of cCSFs (*gallery, park, people*). The three selected stimuli are radically different in terms of scene content (i.e. spatial structure, distribution of the content). Their displayed mean luminance varied between 8 and 22 cd/m^2. Two different contrast versions of each image were used in the experiments, as detailed in Section 4.2. Illustrations of the image stimuli at normal contrast are in Appendix B.

Fig. 1 illustrates normal and low contrast versions of one example image stimulus. Fig. 2 shows the C_{RMS} squared contrast spectra ($c_s(u)^2$ in Eq. (7)) of the normal and low contrast versions of the three stimuli used in the cCSF experiments. $1/f^n$ functions were fitted to the measured spectra of all image stimuli except for *people*, where a logarithmic function offered the best fit. These functions were consequently used in the modeling of the cCSF.

For comparative purposes, we show in Fig. 3. the C_{RMS} squared contrast spectra of the normal and low contrast versions of the three image stimuli selected for cCSF modeling, along with the C_{RMS} squared contrast spectra of all 65 scenes originally captured before stimulus selection. We notice that, the six selected stimuli are diverse enough for the purpose of our experimentation; they cover a very wide range of contrast spectral magnitudes and have typical natural scene spectral profiles.

4.4. Stimulus display

Two identical, very high quality wide gamut 24″ EIZO LCDs with built in hardware calibration were used for displaying the stimuli in two identically set up experimental rooms. The choice of employing LCDs in visual work is controversial. Recent work has shown that visual work on modern displays can match that on CRTs [89,90]. The displays incorporated digital uniformity equalizers, ensuring luminance uniformity across the screen. They were set to a mean luminance of 55 cd/m^2. They displayed 10 bits per channel output at 60 Hz from 16-bit per channel, linearized via LUTs images, so that each color channel was displayed with 1024 equally spaced in luminance steps. The display pixel pitch size was 0.270 mm square, corresponding to a theoretical maximum display spatial frequency of 1.85 mm^{-1}, or a maximum retinal frequency of 58 cpd, calculated for our set viewing distance of 1800 mm. The horizontal field size from this distance was 16.5 degrees.

The displays' optoelectronic conversion (gamma) functions were evaluated with a Konica Minolta CS200 luminance and color meter and were used to build LUTs for pixel value-to-luminance conversions [91]. The displays' spatial frequency response was measured using the method described by Triantaphillidou et al. [92] and was found relatively constant for the majority of spatial frequencies of interest, i.e. 88% modulation transfer at 32 cpd – the maximum spatial frequency investigated – and 96% at 16 cpd. We chose not to compensate for this loss of modulation, since it falls within the measurement error and is smaller than the typical error bar in our results. Identical high-spec 64-bit Windows 7 workstations were used for running the display interface during the visual tests. Each was driven by 1 GB NVIDIA Quadro 2000 graphics card, set to display 10-bit per pixel resolution.

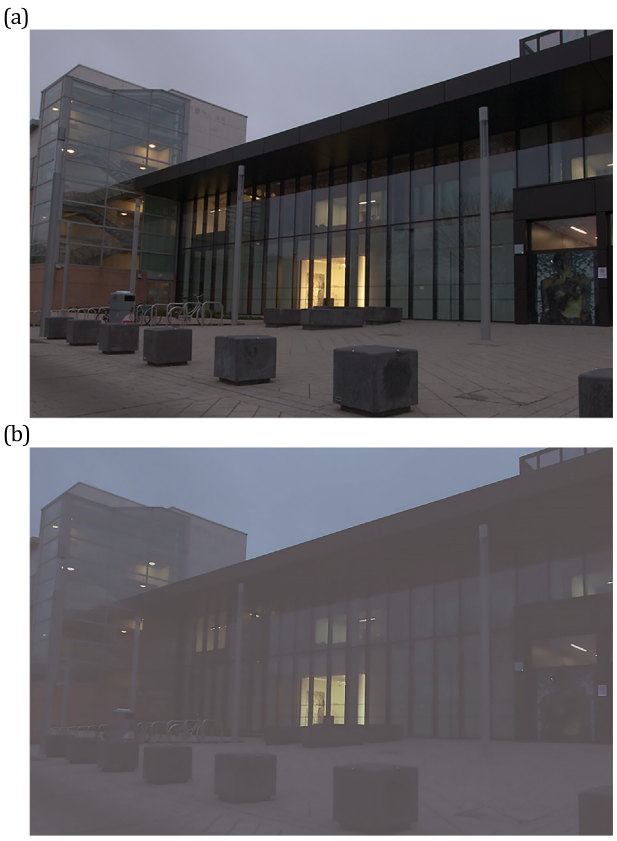

Fig. 1. (a) Normal contrast and (b) low contrast versions of the test image *gallery*.

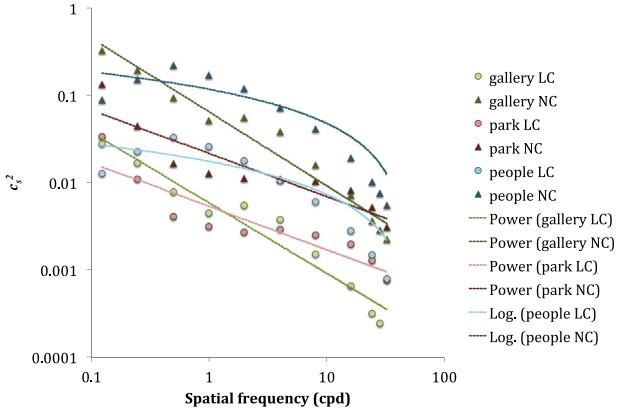

Fig. 2. C_{RMS} squared band contrasts of normal and low contrast versions of three stimuli, used in cCSF experiments. The dashed lines indicate regression curves, used in the cCSF modeling.

5. Experimental set-up

5.1. Interface and observers

A paired comparison yes/no interface was designed to present standard and test stimuli superimposed in random order, with a temporal separation of 300 msc, and a mid-gray screen displayed between stimuli

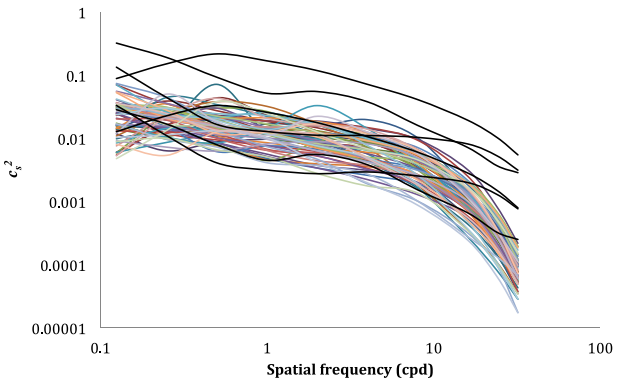

Fig. 3. C_{RMS} squared band contrasts of normal and low contrast versions of three stimuli selected for use in cCSF experiments (black curves), along with the C_{RMS} squared band contrasts of all captured scenes (colored curves).

presentation. Standard and test stimuli were displayed in alteration for 3 s each, and continued being displayed until the observer choose to answer yes or no to the question "are the two images different?". Observers sat 1800 mm from the display faceplate, with their head positioned on a chin rest and were asked to free-view the displayed images by only moving their eyes. They operated blindly two keys on a nearby keyboard, each corresponding to each answer. We found that the temporal stimulus separation arrangement provided a restricted eye-scanning movement compared to two side-by-side monitors displaying standard and test, resulting in less observer fatigue and ensuring no involuntary head movement. The presentation of the two stimuli with a gray screen in between was revealed to be the best arrangement for our purposes, but relied on short-term observer memory. The observers were given initial training with each image stimulus and all frequency bands, so that they (i) fully understood the task, (ii) got used to the free-viewing mode.

A large number of observers initially participated in pilot studies. After initial training, each observer experimented with a random stimulus at 3 different frequency bands (0.25 cpd, 2 cpd, 16 cpd), and took 6 repeated trials per frequency band. The contextual detection experimental set-up was used for the purpose (Section 5.3). Examination of the intraclass correlation coefficient (ICC) in software R led to the selection of ten observers with the highest ICC scores, for participation in the experiments. They were students or post-doctorate researchers in imaging, photography or visual arts, a mixture of males and females, with ages ranging from 18 to 35 years old. They all had normal, or corrected to normal visual acutance and were free from color deficiencies. Each data point in our measured functions is the result of minimum 3 observations.

5.2. Experimental method

We employed a non-parametric adaptive staircase method for the contrast threshold determination, exploiting a semi-stochastic approach. The method is based on the classical non-parametric up-down stochastic staircase presentations [93,94]. However, the step size δ_{n+1} for adding/removing contrast in the test frequency band of interest in trial $n+1$, instead of having a constant value δ, is dependent on the previous step, δ_n, and a measure α related to the count of successive identical responses, and is bounded at the maximum value of δ_n=1. The resulting δ_{n+1} is adapted by a stochastic factor $S\phi$. Following a positive response, Z_n=1 to the stimulus presented in trial n, $S\phi$ takes a random value between 0.51 and 0.75 that is independent of the value of past responses. In the case of a negative response, Z_n=0, $S\phi$ takes the value 0.51.

The stopping criterion depends on a minimum contrast threshold level, identified empirically after extensive pilot studies, and the number of iterations. The staircase typically converges within 20–25 iterations, and the target probability on the psychometric curve, ϕ, is minimum 50%. The full algorithm is included in Appendix C. The code can be obtained via [95].

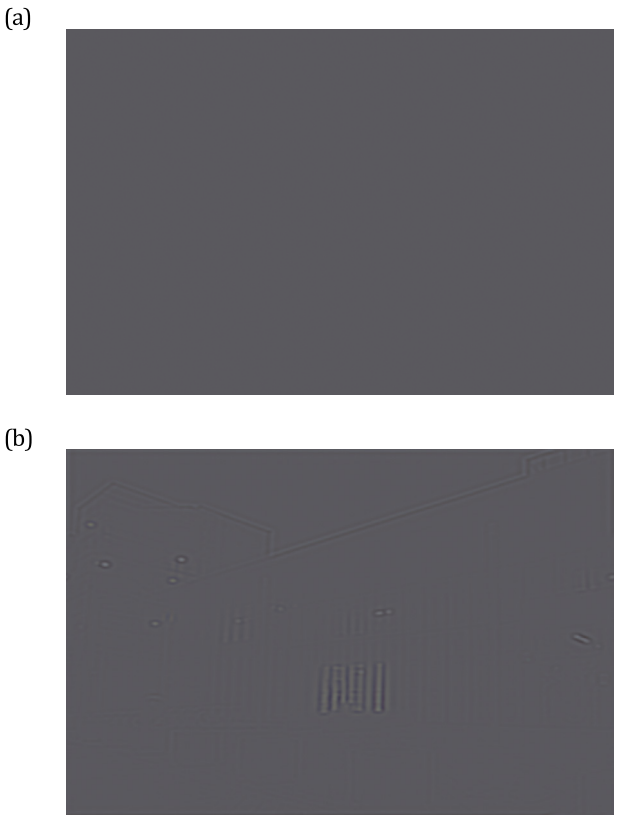

Fig. 4. (a) Standard and (b) example test stimuli for the iCSF experiment, for a given frequency band of the image *gallery*.

5.3. iCSF and cCSF measurement

The iCSF was measured by presenting the standard, comprising a uniform field of mean luminance equal to the mean luminance of the image (Fig. 4(a)), against the test, comprising a variable contrast portion of the band-limited image contrast (example in Fig. 4(b)), until the contrast threshold was identified (as described in Section 5.2). The inverse of the C_{RMS} detection threshold gave the isolated contrast sensitivity value for the given frequency band.

The cCSF was measured in the same fashion as the iCSF, but initially, both standard (Fig. 5(a)) and test contained all pictorial information outside the frequency band of interest. Again, the test band-limited contrast varied (example in Fig. 5(b)), until the contrast threshold was identified. Here, for any given frequency band, the C_{RMS} detection threshold was identified in the presence of the contrasts in the remaining frequency bands, and its inverse gave the contextual contrast sensitivity value for the given band. Figs. 4 and 5 present illustrative examples of standard and test stimuli for measuring iCSF and cCSF respectively.

6. Results

6.1. iCSF measurement and model predictions

In Fig. 6 measurements of isolated contrast sensitivity are plotted for all image stimuli. Since the original image contrast did not affect the measured iCSF (i.e. test band-limited image contrasts and threshold are

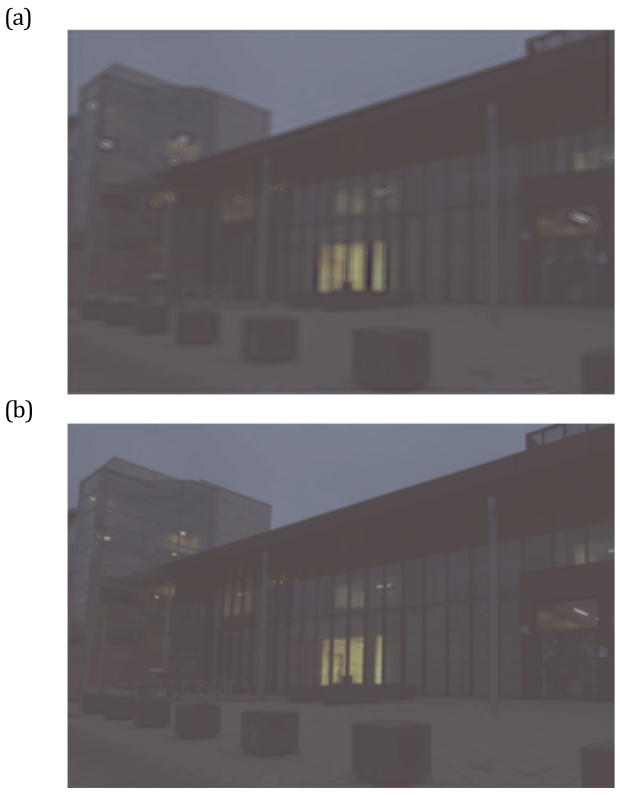

Fig. 5. (a) Example standard and (b) test stimuli for the cCSF experiment, for the low contrast version of image *gallery,* and for the same frequency band as in Fig. 4.

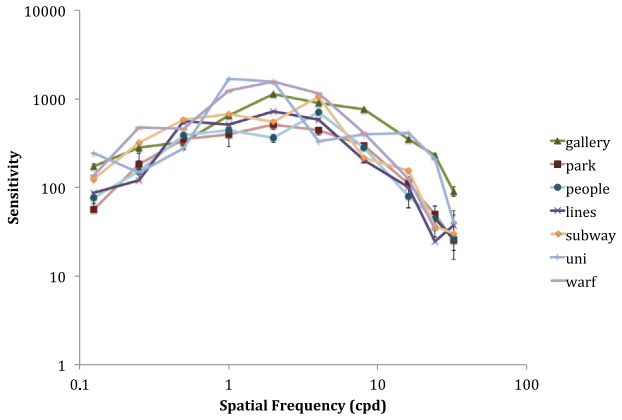

Fig. 6. iCSF measured for 7 image stimuli.

only small fractions of the original band-limited image contrasts), data for both contrast versions of the image stimuli were averaged for each scene. Lines connecting the points are included to give an indication of the measured iCSF profiles. These are typically band-pass in shape, with peaks centered between 1.0 and 4.0 cpd, dropping gradually due to the optical limitations of the eye. They show iCSF variation with scene content. The error bars indicate the standard error, calculated for a number of trained observers that varied between 3 and 6. Where the error bars are not visible, the size of the data point covers the error (see Fig. 6).

In Fig. 7 iCSF measurements are plotted for the three images used in our study of the cCSF (Section 6.2). The model iCSF Eqs. (2)–(6) accounting for the viewing conditions is also plotted, with the k value in Eq. (2) set to 2.8. Barten varies$k$ between 2.5 and 4.5 to account for variations in experimental set-ups [5, p.20]. It appears that a value

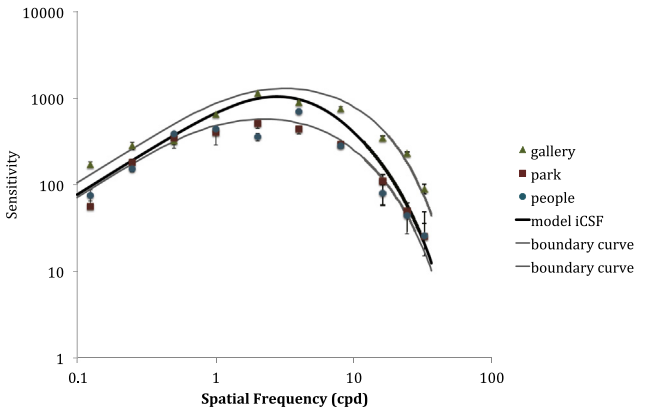

Fig. 7. iCSF measurements for 3 image stimuli and the model iCSF, upper and lower boundary curves.

Table 1
K values used for cCSF modeling.

	Gallery	*Park*	*People*
Norm. contrast	0.02	0.02	0.02
Low contrast	0.025	0.02	0.015

near nominal 3.0 [5, p.39] reflects our constant set-up and paradigm. The upper and lower curves are simply in place to visually indicate the relative spread of the data. We draw no mechanistic conclusions from these boundary curves. As will be shown in Section 6.2, their function is to simply give an estimate of the errors incurred by using a single iCSF (essentially a standard Barden CSF) in the calculation of cCSF. The boundary curves were calculated by arbitrarily adjusting Eq. (2) parameters (k and the internal visual noise).

6.2. cCSF measurement and model predictions

Three image stimuli were selected to investigate the contextual sensitivity. Figs. 8–10 show predictions of the cCSF (solid curves) when Eq. (6), with the standard Barten parameters values, was implemented in Eq. (7) for the *gallery*, *park* and *people* image stimuli respectively, each at two different contrast levels. The model is seen to offer good predictions of cCSF, but for one image (*gallery,* normal contrast version) there is a small undershoot at low frequencies. The model cCSF and boundary curves (dotted curves in Figs. 8–10) are obtained using the iCSF model and boundary curves (upper or lower — whichever gave the closest match to the data) respectively, and the estimated contrast spectra of the images. The K values in the cCSF model were derived empirically and are listed in Table 1. Errors incurred by not using the boundary iCSF curves, which may be more representative to individual image iCSF data, are relatively small and confined to high frequencies.

It is important to note that, this genre of model is not expected to give a precise fit to measured data [5], but to offer realistic predictions which have, unlike black box good fitting models, a full mechanistic basis.

7. Discussion

Through spatial frequency decomposition and band-limited image C_{RMS} contrast modulation we have measured isolated contrast sensitivity (iCSF) and contextual contrast sensitivity (cCSF) from seven and three considerably different natural images respectively, using a yes/no paradigm and a variable step-size staircase converging method.

Overall, the measured iCSF matches closely standard CSF measurement from sine-waves for the same viewing conditions, re-enforcing the view that a simple sine-wave grating CSF can sensibly describe

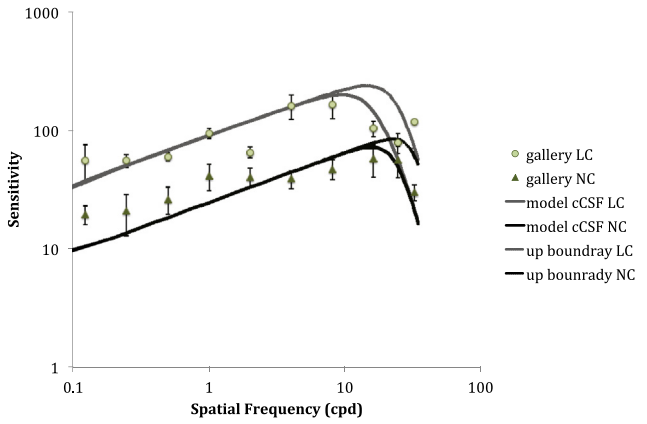

Fig. 8. cCSF measured and modeled for *gallery* low contrast (LC) and normal contrast (NC) image versions, and upper boundary curves.

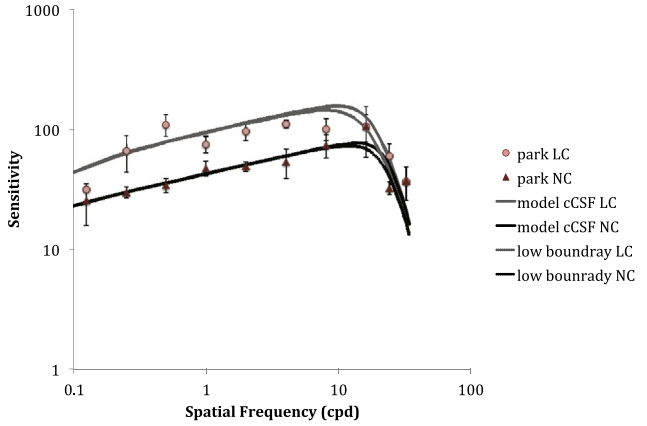

Fig. 9. cCSF measured and modeled for *park* low contrast (LC) and normal contrast (NC) image versions, and lower boundary curves.

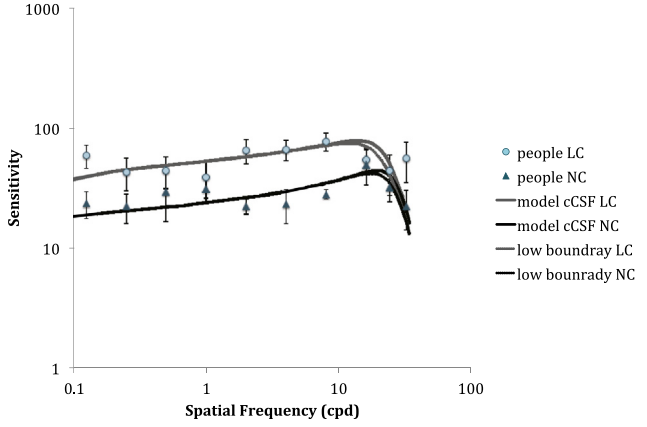

Fig. 10. cCSF measured and modeled for *people* low contrast (LC) and normal contrast (NC) image versions, and lower boundary curves.

sensitivity to more complex, isolated (2D) band signals. This follows recent findings from similar experiments, which have not necessarily used the exact same paradigms [18]. And it is not surprising, considering that any given spatial frequency band is alike a compound sine-wave grating, with frequencies ranging around a central frequency by ±0.5 octave. The impact of the iCSF matching closely the standard CSF measurement is that the latter may be used to represent isolated band-pass image sensitivity in our modeling of contextual sensitivity (cCSF).

Implementation of Eqs. (2)–(6) has enabled good predictions of a standard iCSF, appropriate for the size and luminance of the pictorial stimuli used in our study. There is no change in the measured iCSF with changes in original image contrast, as reflected by the model. There is a relative variation in the measured iCSF with scene content, which is reflected by the boundary curves (but which are mechanistically not meaningful), but, given the distinctive structural and color variations in the images we used, this is insignificant. Especially when considering that the real purpose of the iCSF is to facilitate a model prediction of cCSF. The latter, in our view, describes the real human contrast sensitivity to displayed natural images and is therefore more relevant to imaging and applied vision applications.

The modeling of individual pictorial stimulus cCSFs using a unique standard iCSF curve has shown that the discrepancies between the measured data and model iCSF have very little effect on cCSF prediction. This is a consequence of the non-linear power mechanism contained within the cCSF modeling, which followed the measured data closely. Individual cCSFs varied mainly because of contrast spectrum variations. Thus, if as pictorial content varies the iCSF remains relatively invariant and may be represented with the standard sine-wave CSF, the cCSF can be readily predicted, given the image's contrast spectrum.

Further, results indicate that band contextual detection is lowered as image contrast is raised. This is expected, it is due to the increase in effective noise generated from signal outside the band sensitivity being measured. The overall profile of the cCSF reveals a slow increase in contextual detection until a turning point is reached around 16 cpd, due to optical limitations of the eye. As spatial frequency increases above this figure, sensitivity decreases due to optical factors. Such low-pass-type CSFs are most commonly utilized in image quality algorithms [23]. cCSF profiles reveal why.

The modeling of the cCSF is based on the relatively simple hypothesis that pictorial masking from ±1 octaves from a given frequency band is primarily responsible for the decrease in sensitivity from the isolated case. This has already been established from previous determinations of sine-wave contrast sensitivity in the presence of a wide range of both static and dynamic noise patterns (see Appendix A for references). The majority of these studies have successfully predicted sine-wave contrast sensitivity through application of the LAM. Isolated band measurements (iCSF) show that, when viewed by an observer, these signals are essentially being processed in the visual system in a similar way to sine-wave gratings. Thus, the assumption that the contextual cCSF is essentially equivalent to sine-wave detection in the presence of noise seems reasonable. Other picture elements, such as spatial distribution of contrasts, edge density or color terms, are also likely to affect final contextual sensitivity at a given frequency [18,46]. Nevertheless, the model is shown to satisfactorily predict measurements, for the three distinctly different natural scenes we used in this work at two different levels of contrast.

Although the cCSF model's predictions would require subsequent validation using measurements from more images and different set-ups, the currently study of the nature of contextual contrast sensitivity provides a very solid foundation for further investigations. The remarkable stability of the LAM based constant, K, in the formulation of cCSF (in the order of 0.02, ±0.005) is very encouraging. It is worth noting that, even if a scene-related variation in K occurred, then it would simply shift the theoretical cCSF curves up, or down (change in function's magnitude, not its profile). The profile of the cCSF is mainly based on the contrast spectra of the stimuli Eq. (7). This follows from our theoretical analysis of the LAM (see Eq. (A.13), Appendix A). In Eq. (7), the isolated band-limited contrast $(c_i)^2$ term becomes significantly smaller than the contrast spectrum term $(c_s)^2$ over most of the visually significant spatial frequency range. Over this range, the calculated cCSF becomes directly proportional to $(c_s)^2$ whilst it is relatively insensitive to variations in iCSF. This theoretical prediction is supported well by

our measured data. Further, for the practical application of the cCSF in image quality models, a change in K, and thus in the magnitude of cCSF, is not important, because in such a case the cCSF is typically used as visual weighting function in a normalized form.

The RMS metric we used has been seen in contrast normalization models of cortical cell responses, and it is shown to predict contrast detection thresholds for patches as well as natural scenes [12,13]. Although most widely used, it is still debatable whether it is the best measure to describe visual contrast in natural scenes, where local contrasts usually vary importantly. One weakness of the metric is that in very dark regions, only some small bright pixels are enough to bring the RMS high, overestimating the perceived contrast. This weakness has been previously discussed by Frazor and Geisler [12], who also measured RMS in the fashion we did (C_{RMS} Eq. (1)). They further slightly modified the metric to account for this rare case failing, but with very little effect in their data and no real impact to the global trends of their results. Different work is necessary for providing a new contrast measure that takes into account the level of spatial spread of contrast fluctuations in a stimulus. Fluctuations that are spatially close together provide more of the effective contrast that reaches the human eye by contributing to the local average at similar spatial locations.

Overall, the model iCSF/cCSF fits are not perfect, but, given that these are mechanistic models, this is expected. They are, nevertheless, representative trends of low-level vision when viewing images. They are relatively simple to apply and hold well, while accounting for the relevant viewing condition. The majority of recent research into the HVS models relating to image quality is done using computational approaches instead, which attempt to combine lower level and higher level processes relevant to image quality judgment, typically in black-box modeling techniques. On this point, Chandler and colleagues have emphasized how research in image quality modeling has shifted from previous objectives of gaining a better understanding of human vision to the narrow objective of better fitting the available quality scores [23, 50].

Incorporating visual models in image quality models has, generally, led to increased correlations with perceived quality (subjective scores). This is not entirely true for all types of models. Some computational, or back-box image quality models, which derive values for model parameters on curve fitting operations with observers' scores, do not necessarily require visual models incorporated in them. On the other hand, signal-transfer based image quality models, which account for signal degradation in each component of the imaging system, benefit from visual models that describe accurately visual facility and degradation. So theoretically, for the latter genre, a "perfect" visual model would provide better correlations with subjective scores.

Further research into cCSFs and the equivalent contrast discrimination functions derived from natural scenes can help to re-establish bottom-up approaches, which offer straightforward and measurable pathways to visual and to quality modeling, and combine them with higher level processes. The latter image quality metrics are generally comprehensive, modular and therefore flexible and capable of incorporating imaging system performance measures and accounting for viewing conditions [96]. They are still widely desirable in the optical and digital systems engineering world.

Acknowledgments

We thank the UK Defence Science and Technology Laboratory (DSTL) for funding this work and all our observers for their participation to the visual experiments. Thanks to Ed Fry, PhD researcher in the Computational Vision and Imaging Technology research group labs, for assistance with image capturing and useful discussions on the role of the CSF in image quality modeling.

We confirm that the funding source had no involvement in the study design, in the collection, analysis and interpretation of data, in the writing of the report; and in the decision to submit the article for publication. We also confirm that the funding source has not imposed any restrictions with respect to the publication of the funded research.

Appendix A

Theoretical relationship between contextual detection (cCSF) and isolated detection (iCSF) through application of the linear amplification model (LAM)

The impact of spatiotemporal noise on signal contrast detection can be quantified through the LAM [5,37]. The model has been verified experimentally through a number of studies employing sine-wave gratings embedded in both static and dynamic noise [3,37–39,97]. If the signal and noise cover the same stimulus area, the LAM maybe expressed in the form:

$$(\Phi)_c - (\Phi)_i = const(\Phi)_n \tag{A.1}$$

where $(\Phi)_c$ and $(\Phi)_i$ denote the power spectral density of the signal at the threshold of detection, either with or without spatiotemporal noise respectively. $(\Phi)_n$ is the power spectral density of the noise.

The constant in Eq. (A.1) is a dimensionless number, independent of spatial or temporal frequency. It relates to the reciprocal of sampling efficiency, which describes the ability of an observer to make use of stimulus information relative to an ideal observer [97]. Barten indicates that this constant can also be described in terms of the Crozier coefficient [5]. In other words, this is a basic visual constant.

Our analysis treats individual band-limited signals acting as sources of noise when operating on a given band-limited signal. The general relationship between the two-sided version of Φ and RMS contrast c for static images is then given by:

$$(c)^2 = \iint \Phi(u,v)\,2du2dv \tag{A.2}$$

where the limits of each integration are 0 and ∞, and the terms u, and v denote spatial frequency.

For each band-limited image in our experiments we define c as:

$$(c)^2 = \frac{1}{N}\sum\left(\left(L_j - L_{mean}\right)/L_{mean}\right)^2 \tag{A.3}$$

where L_j is the luminance of the jth pixel, L_{mean} is the mean luminance and N is the number of pixels. The summation shown in Eq. (A.3) is conducted over the range $j = 1$ to N. Note that for all bands considered in our examination, N is constant and equal to the total number of pixels in the image.

Our particular band-limited signals were produced using rotationally symmetric log cosine filters [9], which define $u = v$. In this analysis, each band-limited signal is mathematically approximated by an idealized rectangular spectrum of 1 octave width, defined by a frequency difference of $(a - b)$ cpd. A constant spectral density is assumed within a given band. For such a band we have from Eq. (A.2):

$$(c)^2 = (2(a-b))^2\,\Phi \tag{A.4}$$

Noise masking effects on grating detection at a given frequency u cpd emanate mainly from spatial frequencies occurring within +1 and −1 octave of this frequency [34]. We assume that the same occurs in the masking of individual band-limited signals. In our experimentation, we initially measure band-limited contrast sensitivity when presented in isolation to the observer. The measured sensitivity represents a response to a combination of all frequency components defined within the given octave. We then measure the band-limited sensitivity with all the other signals present. Any change in contextual detection sensitivity compared with the isolated sensitivity will reflect the impact of spatial frequencies outside of the given octave of interest.

The impact of masking from a noise source centered at frequency u_n, on a signal centered at frequency u, can be quantified through:

$$\Phi_n(u) = \int_0^\infty \psi\left(u_n, u\right)\Phi\left(u_n\right)/u\,du_n \tag{A.5}$$

where $\Phi\left(u_n\right)$ represents the spectral density of the noise source at frequency u_n, remote from the given band-limited signal. $\Phi_n(u)$ is the effective spectral density of this noise operating at frequency u.

The function ψ, found empirically by Barten [5] from a study of the Stromeyer and Julesz [34] data, is given by:

$$\psi\left(u_n,u\right)=0.747\exp\left(-2.2\ln^2\left(u_n/u\right)\right) \tag{A.6}$$

Eq. (A.6) represents a log normal function, described by a Gaussian distribution of $\ln(u_n/u)$, with a half-height width of two octaves, consistent with the premise that noise masking on a given band-limited signal will be generated from spatial frequencies within the two flanking bands.

Consider now, a band-limited signal centered on frequency u, (with squared RMS contrast denoted by c_s^2). This represents one point on the contrast spectrum. If it assumed that the noise at u emanates from the two flanking bands, centered at +1 and −1 octaves from u, then for each of these Eq. (A.5) can be written as:

$$\left[\Phi_n(u)\right]_{-1}=\Phi_{-1}\int_a^b\psi\left(u_n,u\right)/u\,du_n \tag{A.7}$$

for frequencies centered at −1 octave below the band-limited signal centered at u, and:

$$\left[\Phi_n(u)\right]_{+1}=\Phi_{+1}\int_a^b\psi\left(u_n,u\right)/u\,du_n \tag{A.8}$$

for frequencies +1 octave above.

The integration limits (a, b) are $(1/\sqrt{2}u,\ 1/2\sqrt{2}u)$ for the −1 octave flanking band and $(2\sqrt{2}u,\ \sqrt{2}u)$ for the +1 octave flanking band. Using Eqs. (A.4), (A.7) and (A.8) can be re-written as Eqs. (A.9) and (A.10):

$$\left[\Phi_n(u)\right]_{-1}=\left(c_s^2\right)_{-1}(2(a-b))^{-2}\int_a^b\psi\left(u_n,u\right)/u\,du_n \tag{A.9}$$

$$\left[\Phi_n(u)\right]_{+1}=\left(c_s^2\right)_{+1}(2(a-b))^{-2}\int_a^b\psi\left(u_n,u\right)/u\,du_n \tag{A.10}$$

The total noise spectral density at u is then given by:

$$\left[\Phi_n(u)\right]=\left[\Phi_n(u)\right]_{-1}+\left[\Phi_n(u)\right]_{+1} \tag{A.11}$$

Eq. (A.11) was evaluated for each band-limited signal, as its central frequency u varied, for the gallery, park and people images. The integrals were calculated using Simpsons rule. It was found that, at any given u:

$$\Phi_n(u)=const\Phi_s(u) \tag{A.12}$$

where $\Phi_s(u)$ denotes the spectral density of the band-limited signal centered at u.

Combining Eq. (A.1) (the LAM) with Eqs. (A.4) and (A.12) predicts that, at a given value of u:

$$\left[c_c^2\right]-\left[c_i^2\right]=const\left[c_s{}^2\right] \tag{A.13}$$

The terms c_c^2 and c_i^2 denote the contextual and isolated detection threshold RMS values (squared) for the band-limited stimulus centered at u. Thus, the algebraic difference between these two quantities is directly proportional to the square of the contrast spectrum at u. At any given frequency, ${c_c}^{-1}$ and ${c_i}^{-1}$ represent values of the cCSF and iCSF respectively. Therefore, if both c_s and c_i are known, the cCSF can be evaluated through Eq. (A.13) (this is also expressed in Eq. (7) in the main text – initially derived empirically – where $const = K$). If the exact numerical value of the constant in Eq. (A.13) is known, then absolute values for the cCSF are obtained.

Appendix B

Test images

Fig. B.1 (a) to (g) illustrate the normal contrast versions of the natural image stimuli used in this study.

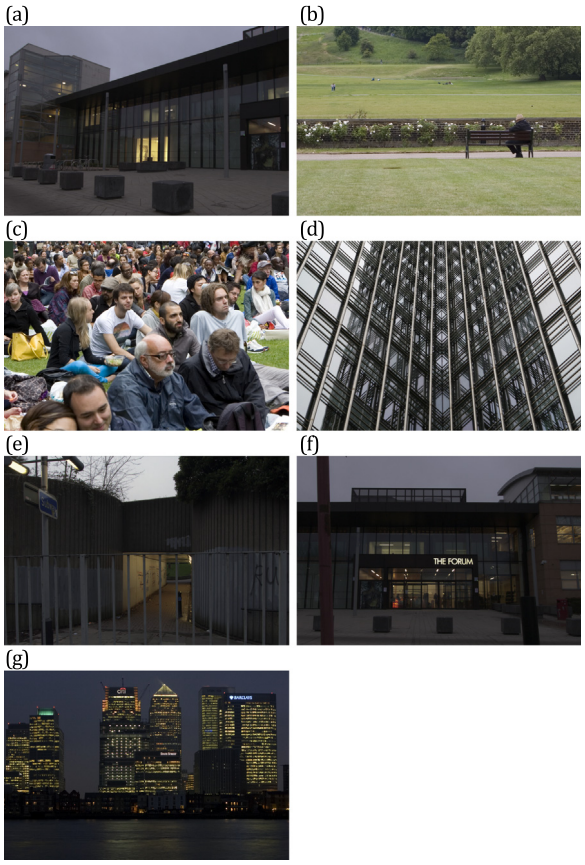

Fig. B.1. (a) *gallery*, (b) *park*, (c) *people*, (d) *lines*, (e) *subway*, (f) *uni*, and (g) *wharf*.

Appendix C

Adaptive staircase

A non-parametric adaptive staircase method, with a variable step size was used for the definition of the contrast thresholds. It exploits a semi-stochastic approach [93,94] to define the stepping rule. This is given by:

$For\ Z_n=1$

$$X_{n+1}=X_n-\delta_n S_\phi\left[1+Z_{n-1}(a-1)\right]$$

$For\ Z_n=0$

$$X_{n+1}=X_n-0.51\delta_n\left[1+(1-Z_{n-1})(a-1)\right]$$

where,

X_n is the stimulus level at trial n

X_{n+1} is the stimulus level at next trial $n+1$

Z_n is the response at the current trial n:

$Z_n=1$ if stimulus is detected (yes response)

$Z_n=0$ if stimulus is not detected (no response)

δ_n is the step used to determine stimulus value X_{n+1}

$$S_\phi=random\ (0.51:0.75)$$

$$\alpha=\frac{\log(mCount)}{\log(2)}1.5$$

mCount counts the number of successive unchanged "yes" or "no" responses

If $\delta_n\alpha>1$ then $\delta_n=1$

The stopping criterion depends on a minimum contrast threshold level, identified empirically, and the number of iterations. The target probability on the psychometric curve, ϕ, is minimum 50%.

The code for the method is accessible via [95].

References

[1] O.H. Schade, Optical and photoelectric analog of the eye, J. Opt. Soc. Amer. A 46 (1956) 721–739.
[2] F.W. Campbell, J.G. Robson, Application of fourier analysis to the visibility of gratings, J. Physiol. 197 (1968) 551–566.
[3] A. Watanabe, T. Mori, S. Nagata, K. Hiwatashi, Spatial sine-wave responses of the human visual system, Vision Res. 8 (1968) 1245–1263.
[4] D.H. Kelly, Visual contrast sensitivity, Opt. Acta 24 (1977) 107–129.
[5] P.J.G. Barten, Contrast Sensitivity of the Human Eye and its Effects on Image Quality, SPIE Press, 1999.
[6] A.B. Watson, A.J. Ahumada, A standard model for foveal detection of spatial contrast, J. Vision 5 (2005) 717–740.
[7] A.A. Michelson, Studies in Optics, Dover Publications Inc., 1995.
[8] H.E. Ross, D.J. Murray, transl (Eds.), E. H. Weber on the Tactile Senses, second ed., Taylor & Francis, 1996.
[9] E. Peli, Contrast in complex images, J. Opt. Soc. Amer. A 7 (1990) 2032–2040.
[10] S. Triantaphillidou, Introduction to image quality and system performance, in: E. Allen, S. Triantaphillidou (Eds.), The Manual of Photography 10th ed, Focal Press, Elsevier, 2010, pp. 345–363.
[11] G. Simone, M. Pedersen, J.Y. Hardeberg, Measuring perceptual contrast in digital images, J. Vis. Commun. Image R. 23 (2012) 491–506.
[12] R.A. Frazor, W.S. Geisler, Local luminance and contrast in natural scenes, Vision Res. 46 (2006) 1585–1598.
[13] P.J. Bex, I. Mareschal, S.C. Dakin, Contrast gain control in natural scenes, J. Vision 7 (2007) 1–12.
[14] O. Ukkonen, J. Rovamo, R. Näsänen, Effect of location and orientation uncertainty on r.m.s. contrast sensitivity with and without spatial noise in peripheral and foveal vision, Optom. Vis. Sci. 72 (1995) 387–395.
[15] P.J. Bex, W. Makous, Spatial frequency, phase, and the contrast of natural images, J. Opt. Soc. Amer. A 19 (2002) 1096–1106.
[16] S. Triantaphillidou, E. Allen, R. Jacobson, Image compression schemes, Part 2: scene dependency, scene analysis and classification, J. Imaging Sci. Tech. 51 (2007) (2000) 259–270.
[17] A.D. Hwang, E. Peli, New contrast metric for realistic display performance measure, SID Symp. Digest Tech. Papers 47 (1) (2016) 982–985, http://dx.doi.org/10.1002/sdtp.10893.
[18] A.M. Haun, E. Peli, Perceived contrast in complex images, J. Vision 13 (2013) 1–21.
[19] E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imag. 19 (2010) 011006-011006.
[20] D.G. Pelli, P. Bex, Measuring contrast sensitivity, Vision Res. 90 (2013) 10–14.
[21] E. Peli, Contrast sensitivity function and image discrimination, J. Opt. Soc. Amer. A 18 (2001) 283–293.
[22] G.M. Johnson, M.D. Fairchild, On contrast sensitivity in an image difference model, in: Proceedings of PICS 2002: Image Processing, Image Quality, Image Capture, Systems Conference, Society for Imaging Science and Technology, 2002, pp. 18–23.
[23] D.M. Chandler, Seven challenges in image quality assessment: past, present, and future research, ISRN Signal Process. 2013 (2013) 53.
[24] F.W. Campbell, J.J. Kulikowski, J.Z. Levinson, The effect of orientation on the visual resolution of gratings, J. Physiol. 187 (1966) 427–433.
[25] M.A. Berkley, F. Kitterle, D.W. Watkins, Grating visibility as a function of orientation and retinal eccentricity, Vision Res. 15 (1975) 239–244.
[26] J.G. Robson, N. Graham, Probability summation and regional variation in contrast sensitivity across the visual field, Vision Res. 21 (1981) 409–418.
[27] A.J. Ahumada, S.M. Wuerger, A.B. Watson, Estimation of chromatic channel spatial frequency responses, J. Vision 3 (2003) 43.
[28] S. Appelle, Perception and discrimination as a function of stimulus orientation: the oblique effect in man and animals, Psychol. Bull. 78 (1972) 266–278.
[29] E. Peli, L.E. Arend, G.M. Young, R.B. Goldstein, Contrast sensitivity to patch stimuli: Effects of spatial bandwidth and temporal presentation, Spatial Vis. 7 (1993) 1–14.
[30] D.L. Ringach, Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex, J. Neurophysiol. 88 (2002) 455–463.
[31] S.T.L. Chung, G.E. Legge, Comparing the shape of contrast sensitivity functions for normal and low vision, Invest. Ophthalmol. Vis. Sci. 57 (2016) 198–207, http://dx.doi.org/10.1167/iovs.15-18084.
[32] J. Rovamo, R. Fransilla, R. Näsänen, Contrast sensitivity as a function of spatial frequency, viewing distance, and eccentricity with and without spatial noise, Vision Res. 32 (1992) 632–637.
[33] J. Rovamo, H. Kukkonen, K. Tiipana, R. Näsänen, Effects of luminance and exposure time on contrast sensitivity in spatial noise, Vision Res. 33 (1993) 1123–1129.
[34] C.F. Stromeyer, B. Julesz, Spatial frequency masking in vision: Critical bands and spread of masking, J. Opt. Soc. Amer. A 62 (1972) 1221–1232.
[35] J.A. Solomom, The history of dipper functions, Atten. Percept. Psychophys. 71 (2009) 435–443.
[36] G.E. Legge, J.M. Foley, Contrast masking in human vision, J. Opt. Soc. Amer. A 70 (1980) 1458–1471.
[37] G.E. Legge, D. Kersten, A.E. Burgess, Contrast discrimination in noise, J. Opt. Soc. Amer. A 4 (1987) 391–404.
[38] A. Van Meeteren, J. Valeton, Effects of pictorial noise interfering with visual detection, J. Opt. Soc. Amer. A 5 (1988) 438–444.
[39] K.T. Blackwell, The effect of white and filtered noise on contrast detection thresholds, Vision Res. 38 (1998) 267–280.
[40] Y. Liu, J.P. Allebach, A computational texture masking model for natural images based on adjacent visual channel inhibition, Proc. SPIE 9016 (2014) 90160D.
[41] P.C. Teo, D.J. Heeger, Perceptual image distortion, in: Proceedings of 1st International Conference on Image Processing, Austin, TX, Vol. 2, 1994, pp. 982–986, http://dx.doi.org/10.1109/ICIP.1994.413502.
[42] A.B. Watson, J.A. Solomon, Model of visual contrast gain control and pattern masking, J. Opt. Soc. Amer. A 14 (1997) 2379–2391.
[43] S.V. David, W.E. Vinje, J.L. Gallant, Natural stimulus statistics alter the receptive field structure of the v1 neurons, J. Neurosci. 24 (2004) 6991–7006.
[44] J.L. Gallant, C.E. Connor, D.C. Van Essen, Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing, Neuroreport 9 (1998) 2153–2158.
[45] E. Peli, Test of a model of foveal vision by using simulations, J. Opt. Soc. Amer. A 13 (1996) 1131–1138.
[46] P.J. Bex, S.G. Solomon, S.C. Dakin, Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure, J. Vision 9 (2009) 1–19.
[47] M.A. Webster, E, Miyahara contrast adaptation and the spatial structure of natural images, J. Opt. Soc. Amer. A 14 (1997) 2355–2366.
[48] A. Haun, E. Peli, Is image quality a function of contrast perception?, Proc. SPIE 8651 (2013) 86510C.
[49] S. Triantaphillidou, J. Jarvis, G. Gupta, Spatial contrast sensitivity and discrimination in pictorial images, Proc. SPIE 9016 (2014) 901604.
[50] D.M. Chandler, M.M. Alam, T.D. Phan, Seven challenges for image quality research, Proc. SPIE 9014 (2014) 901402.
[51] N.S. Nagaraga, Effect of luminance noise on contrast thresholds, J. Opt. Soc. Amer. 54 (1964) 950–955.
[52] E.M. Crane, An objective method for rating picture sharpness: SMT acutance, J. SMPTE 73 (1964) 643–647.
[53] C.N. Nelson, G.C. Higgins, Image sharpness, in: Proceedings of the SPSE Technical Section Conference on Advances in the Psychophysical and Visual Aspects of Image Evaluation, 1977, pp. 1-1.
[54] K. Biedermann, Y. Feng, Lens performance assessment by image quality criteria, Proc. SPIE 0549 (1985) 36–43.
[55] E.M. Granger, K.N. Cupery, Optical merit function (SQF), which correlates with subjective image judgments, Photograph. Sci. Eng. 16 (1972) 221–230.
[56] H.L. Snyder, Modulation transfer function area as a measure of image quality, in: Visual Search Symposium, (National Academy of Sciences, 1973, pp. 93–105.
[57] P.G.J. Barten, Evaluation of subjective image quality with the square-root integral method, J. Opt. Soc. Amer. A 7 (1990) 2024–2031.
[58] K. Töpfer, R.E. Jacobson, The relationship between objective and subjective image quality criteria, J. Inf. Rec. Mats. 21 (1993) 5–27.
[59] R.B. Jenkin, M.A. Richardson, Comparison between the effective pictorial information capacities of JPEG 6b and 2000, Proc. SPIE 5823 (2005) 13–19.
[60] ISO 15739:2017 Photography – Electronic still-picture imaging – Noise measurements, International Organization of Standardization (2017).
[61] E.W. Jin, Elaine, J.B. Phillips, S. Farnand, M. Belska, V. Tran, E. Chang, Y. Wang, B. Tseng, Towards the development of the IEEE P1858 CPIQ Standard – A validation study, in: Proceedings of the IS & T Electronic Imaging, Image Quality and System Performance XIV, 2017, pp. 88–94, http://dx.doi.org/10.2352/ISSN.2470-1173.2017.12.IQSP-249.
[62] S. Daly, The visible difference predictor: an algorithm for the assessment of image fidelity, in: A.B. Watson (Ed.), Digital Images and Human Vision, MIT Press, 1993, pp. 179–206.
[63] J. Lubin, The use of psychophysical data and models in the analysis of display system performance, in: A.B. Watson (Ed.), Digital Images and Human Vision, MIT Press, 1993, pp. 163–178.
[64] X. Zhang, B.A. Wandell, A spatial extension of CIELAB for digital color-image reproduction, J. Soc. Inf. Disp. 5 (1997) 61–63.
[65] M.D. Fairchild, G.M. Johnson, Icam framework for image appearance, differences, and quality, J. Electron. Imaging 13 (2004) 126–138.
[66] S. Chen, A. Beghdadi, A. Chetouani, Color image assessment using spatial extension to CIE DE2000, in: Proceedings of IEEE Conference on 2008 Digest of Technical Papers - International Conference on Consumer Electronics, IEEE, 2008, pp. 1–2.
[67] D.M. Chandler, S.S. Hemami, Vsnr: a wavelet-based visual signal-to-noise ratio for natural images, IEEE Trans. Image Process. 16 (2007) 2284–2298.
[68] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, A. C.770, Image quality assessment based on a degradation model, IEEE Trans. Image Process. 9 (2000) 636–650.
[69] M. Pedersen, J.Y. Hardeberg, A new spatial hue angle metric for perceptual image difference, in: International Workshop on Computational Color Imaging, Springer Berlin Heidelberg, 2009, pp. 81–90.
[70] V. Laparra, J. Muñoz Marí, J. Malo, Divisive normalization image quality metric revisited, J. Opt. Soc. Amer. A 27 (2010) 852–864.

[71] M. Kriss, J. O'Toole, J. Kinard, Information capacity as a measure of image quality, in: Proceedings of the SPSE conference on Image Analysis and Evaluation, 1976, pp. 122–133.
[72] A.J. Ahumada Jr, Computational image quality metrics a review, Proc. SID 24 (1993) 305–308.
[73] A.J. Ahumada Jr, A.B. Watson, A.M. Rohaly, Models of human image discrimination predict object detection in natural backgrounds, Proc. SPIE 2411 (1995) 355-352.
[74] D.A. Silverstein, E.J. Farrell, The relationship between image fidelity and image quality, in: Proceedings of IEEE International Conference on Image Processing I, IEEE, 1996, pp. 881–884.
[75] R.E. Jacobson, An evaluation of image quality metrics, J. Photogr. Sci. 43 (1995) 7–16.
[76] B.E. Rogowitz, T.N. Pappas, J.P. Allebach, Human vision and electronic imaging, J. Electronic Image. 10 (2001) 10–19.
[77] S. Bouzit, L.W. MacDonald, Sharpness enhancement through spatial frequency decomposition, in: Proceedings of PICS 2001: Image Processing, Image Quality, Image Capture, Systems Conference, Society for Imaging Science and Technology, 2001, pp. 377–381.
[78] Z. Wang, A. Bovik, H. Sheikh, E, Simoncelli image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.
[79] A.M. Haun, E. Peli, The complexities of complex contrast, Proc. SPIE 8292 (2012) 82920E.
[80] M.P.S. To, P.G. Lovell, T. Troscianko, D.J. Tolhurst, Perception of suprathreshold naturalistic changes in colored natural images, J. Vision 10 (2010) 1–22.
[81] J.L. Mannos, D.J. Sakrison, The effects of a visual fidelity criterion on the encoding of images, IEEE Trans. Inform. Theory 20 (1974) 525–535.
[82] T. Movshon, L. Kiorpes, Analysis of the development of spatial sensitivity in monkey and human infants, J. Opt. Soc. Amer. A 5 (1988) 2166–2172.
[83] P.G.J. Barten, Physical model for the contrast sensitivity of the human eye, Proc. SPIE 1666, Human Vision, Visual Processing, and Digital Display III, 57 12.135956 1992.
[84] A.B. Watson, Visual detection of spatial contrast patterns: Evaluation of five simple models, Opt. Express 6 (2000) 12–33.
[85] F.W. Campbell, D.G. Green, Optical and retinal factors affecting visual resolution, J. Physiol. 181 (1965) 576–593.
[86] E. Peli, Feature detection algorithm based on a visual system model, Proc. IEEE 90 (2002) 78–93.
[87] M. Nezamabadi, E.D. Montag, R.S. Berns, An investigation of the effect of image size on the color appearance of softcopy reproductions using a contrast matching technique, Proc. SPIE 6493 (2007) 649309.
[88] M. Pedersen, G. Simone, M. Gong, I. Farup, A total variation based color image quality metric with perceptual contrast filtering, in: International conference on Pervasive Computing, Signal Processing and Applications, 2011.
[89] M. Dorr, L.A. Lesmes, Z.L. Lu, P.J. Bex, Rapid and reliable assessment of contrast sensitivity function on an iPad, Invest. Ophthalmol. Vis. Sci. 12 (2013) 7266–7273.
[90] L. To, R.L. Woods, R.B. Goldstein, E. Peli, Psychophysical contrast calibration, Vision Res. 90 (2013) 15–24.
[91] E.A. Day, L. Taplin, R.S. Berns, Colorimetric characterization of a computer-controlled liquid crystal display, Color Res. Appl. 29 (2004) 365–373.
[92] S. Triantaphillidou, R.E. Jacobson, Measurements of the modulation transfer function of image displays, J. Imaging Sci. Techn. 48 (2004) 58–65.
[93] C. Kaernbach, Simple adaptive testing with a weighted up-down method, Percept. Psychophys. 49 (1991) 227–229.
[94] H. Robbins, S. Monro, A stochastic approximation method, Ann. Math. Stat 29 (1951) 400–407.
[95] www.westminster.ac.uk/CVIT/.
[96] E. Fry, S. Triantaphillidou, R.E. Jacobson, J. Jarvis, R. Jenkin, Bridging the gap between imaging performance and image quality measures, in: Proc. IS & T Electronic Imaging Symposium: Image Quality System Performance XV, 2018.
[97] J.J. McAnany, K.R. Alexander, Spatial contrast sensitivity in dynamic and static additive noise, Vision Res. 50 (2010) 1957–1965.