J. Vis. Commun. Image R. 23 (2012) 491–506

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

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

Measuring perceptual contrast in digital images

Gabriele Simone *, Marius Pedersen, Jon Yngve Hardeberg

The Norwegian Color Research Laboratory, Faculty of Computer Science and Media Technology, Gjøvik University College, P.O. Box 191, N-2802 Gjøvik, Norway

ARTICLE INFO

Article history:
Received 30 July 2009
Accepted 29 December 2011
Available online 20 January 2012

Keywords:
Perceived contrast
Local measure
Multilevel analysis
Difference of Gaussians
Regions-of-Interest
Saliency maps
Psychophysical experiment
Observers correlation

ABSTRACT

In this paper we present a novel method to measure perceptual contrast in digital images. We start from a previous measure of contrast developed by Rizzi et al. [26], which presents a multilevel analysis. In the first part of the work the study is aimed mainly at investigating the contribution of the chromatic channels and whether a more complex neighborhood calculation can improve this previous measure of contrast. Following this, we analyze in detail the contribution of each level developing a weighted multilevel framework. Finally, we perform an investigation of Regions-of-Interest in combination with our measure of contrast. In order to evaluate the performance of our approach, we have carried out a psychophysical experiment in a controlled environment and performed extensive statistical tests. Results show an improvement in correlation between measured contrast and observers perceived contrast when the variance of the three color channels separately is used as weighting parameters for local contrast maps. Using Regions-of-Interest as weighting maps does not improve the ability of contrast measures to predict perceived contrast in digital images. This suggests that Regions-of-Interest cannot be used to improve contrast measures, as contrast is an intrinsic factor and it is judged by the global impression of the image. This indicates that further work on contrast measures should account for the global impression of the image while preserving the local information.

1. Introduction

Since the initial research on image contrast, it has become clear how difficult it is to define perceptual contrast in images. One reason for this is the contextual influence on the observer task and observer experience.

A possible definition of contrast is the difference between the light and dark parts of a photograph, where less contrast gives a "flatter" picture, and more contrast gives a "deeper" picture. Many other definitions of contrast are also given, such as the difference in visual properties that makes an object distinguishable or simply the difference in color from point to point. As various definitions of contrast are given, measuring contrast is surely not a trivial task. Measuring the difference between the darkest and lightest point in an image does not predict perceived contrast since perceived contrast is influenced by the surround (viewing conditions) and the spatial arrangement of the image. Parameters as resolution, viewing distance, lighting conditions, image content, memory color etc. will affect how observers perceive contrast.

For this reason, the first approaches to this topic have confined themselves to study the phenomenon at a global level, operating in controlled situations and under the same constraints, the so-called "void conditions". After these very first experiments more complex measures have been devised, but a universal measure of contrast in images is still not clearly defined.

After introducing the state of the art of global and local contrast measures in Section 2, we present our proposal in Section 3. We start from a previous measure of contrast developed by Rizzi et al. [26], which presents a multilevel analysis. In the first part of the work the study is aimed mainly at investigating the contribution of the chromatic channels and whether a more complex neighborhood calculation can improve this previous measure of contrast. Following this, we analyze in detail the contribution of each level developing a weighted multilevel framework. Finally we perform an investigation of Regions-of-Interest in combination with our measure of contrast. Section 4 describes the psychophysical experiment carried out while Section 5 presents the experimental results and it discusses how our proposal reflects perceptual contrast estimation performance. Finally in Section 6 conclusions are drawn.

2. Background

2.1. Global contrast measures

Several contrast measures have been proposed so far. The classic approaches consist of global measures, which implicitly or explicitly claim that the response of the human visual system (HVS) depends on the absolute luminance and not on the relation

* Corresponding author.
E-mail address: gabriele.simone@hig.no (G. Simone).

1047-3203/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2012.01.008

of its local variations. Global measures have been mainly developed during the second half of the 20th century and most of them take into account only the luminance in the calculation of contrast.

The very first measure of global contrast, in the case of sinusoids or other periodic patterns of symmetrical deviations ranging from the maximum luminance, L_{max}, to minimum luminance, L_{min}, is the Michelson [20] formula:

$$C^M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}. \quad (1)$$

King-Smith and Kulikowski [19], Burkhardt et al. [7], and Whittle [33] follow a similar concept replacing L_{max} or L_{min} with L_{avg}, which is the mean luminance in the image.

An alternative global measure, *single image perceived contrast*, has been developed recently by Calabria and Fairchild [8], which modeled the following equation to measure the perceived contrast:

$$SIP_k = -1.505 + 0.131 \cdot k_C + 0.151 \cdot k_L + 666.216 \cdot k_S \quad (2)$$

where k_C, k_L, and, k_S are respectively the standard deviation of image chroma, lightness and high-passed lightness.

2.2. Local contrast measures

Global measures have some disadvantages, in fact, the study of contrast in an image at a global level provides only a measure related to the maximum global difference in lightness and in some cases chromaticity, which have been shown to be inadequate for measuring real visual configurations. Two single points of extreme brightness or darkness can determine the measure of contrast of the whole image, while the perceived contrast is clearly affected as illustrated in Fig. 1.

To overcome the limits of global measures, local measures have been developed. We present here a selection of these measures focusing on those of particular relevance for our work. This selection with the key features of each measure is summarized in Table 1 at the end of this section.

2.2.1. Hess et al.

The issue of contrast of complex scenes at different spatial frequencies in the context of image processing and perception was addressed explicitly by Hess et al. [16] who defined contrast in the Fourier domain as:

$$C(u, v) = \frac{2A(u, v)}{DC}, \quad (3)$$

where $A(u, v)$ is the amplitude of the Fourier transform of the image, u and v are the horizontal and vertical spatial frequency coordinates, respectively, and DC is the zero-frequency component.

2.2.2. Peli

Frankle and McCann [14], Adelson et al. [1] proposed the use of the multilevel representation as an important implementation feature to mimic HVS. This consists of a set of lowpass or bandpass copies of an image, each representing pattern information at a different scale. This data structure used to represent image information is referred as "pyramid".

Peli [23] proposed the idea of a pyramidal image-contrast structure where for each frequency band, the contrast is defined as the ratio of the bandpass-filtered image at that frequency to the lowpass image filtered to an octave below the same frequency (local luminance mean).

Fig. 1. Weakness of global measures. In all four pictures, global contrast measures would typically define contrast by the pixel values of the eyes (see arrows) as these have the highest and lowest luminance pixel values (L_{max} = 100, L_{min} = 0). The visual contrast of these four images are clearly different, however indicating that global measures cannot adequately predict perceived contrast.

Table 1
Discussed local contrast measure.

Author	Key features
Local contrast measures	
Hess et al. [16]	Fourier domain
Peli [23]	Multilevel analysis (pyramid representation), band-pass filtering
Ahumada and Beard [2]	Gaussian filtering
Tremeau [31]	Adjacency graph representation, Fisher distance
Boccignone et al. [5]	Thermodynamical representation, anisotropic diffusion
Tadmor and Tolhurst [30]	Difference of Gaussians (DOG) model
Rizzi et al. [26]	Multilevel analysis (pyramid representation), 8-neighborhood

To define local band-limited contrast for a complex image, he obtains a band-limited version of the image in the frequency domain $A(u,v)$:

$$A(u,v) \equiv A(r,\theta) \equiv F(r,\theta)G(r), \tag{4}$$

where u and v are the respective horizontal and vertical spatial frequency coordinates, r and θ represent the respective polar spatial frequency coordinates: $r = \sqrt{u^2+v^2}, \theta = tan^{-1}\left(\frac{u}{v}\right), F(r,\theta)$ is the Fourier transform of the image $I(x,y)$, and $G(r)$ is the band pass filter.

In the spatial domain the filtered image $a(x,y)$ can be represented similarly, that is, as:

$$a(x,y) = I(x,y) * g(x,y), \tag{5}$$

where $g(x,y)$ is the inverse Fourier transform of the band-pass filter $G(r)$.

In Peli's approach of measuring local contrast, the pyramid is obtained as follows:

$$A_i(u,v) \equiv A_i(r,\theta) \equiv F(r,\theta)G_i(r), \tag{6}$$

where $G_i(r)$ is a *cosine log* filter centered at frequency of 2^i*cycles/picture*, expressed as:

$$G_i(r) = \frac{1}{2}(1 + cos(\pi log_2 r - \pi i)). \tag{7}$$

The resulting contrast at the band of spatial frequencies can be represented as a two-dimensional array $c_i(x,y)$:

$$c_i(x,y) = \frac{a_i(x,y)}{l_i(x,y)}, \tag{8}$$

where $a_i(x,y)$ is the corresponding local luminance mean image and $l_i(x,y)$ is a low-pass-filtered version of the image containing all energy below the band.

2.2.3. Ahumada and Beard

Ahumada and Beard [2] proposed a method for measuring contrast which can be described as follows: the image I is first convolved with a Gaussian low pass filter F_B:

$$B(x,y) = I(x,y) * F_B(x,y), \tag{9}$$

and the blurred image B is convolved with a second Gaussian low pass filter F_L:

$$L(x,y) = B(x,y) * F_L(x,y). \tag{10}$$

Then, for every pixel of the image, the local contrast is defined as:

$$C(x,y) = \frac{B(x,y)}{L(x,y)} - 1. \tag{11}$$

2.2.4. Tremeau

Tremeau [31] proposed a formula for measuring contrast based on a region adjacency graph defined as follows:

$$C_{ij} = \frac{4P_{ij}^2}{min(P_i,P_j) \times max(P_{\{i,j\}})}, \tag{12}$$

where P_i and P_j are the perimeters of the regions i and j, P_{ij} is the length of the shared boundary between regions i and j, $max(P_{\{i,j\}}) = max\{P_{ik}/k \neq j, P_{kj}/k \neq i\}$ is the longest shared boundary of a region adjacent either to region i, or to region j.

The resulting weighted color contrast is given by the following equation:

$$WC_{ij} = d^2(i,j) \times C_{ij}, \tag{13}$$

where $d^2(i,j)$ is, commonly the Fisher distance, used to quantify the color distance between two regions.

2.2.5. Boccignone et al.

An alternative approach of measuring contrast is presented by Boccignone et al. [5] and Ferraro and Boccignone [13]. They state that any image can be considered as an isolated thermodynamical system by identifying the image intensity with some thermodynamical variable. For measuring contrast they use the following formula:

$$\tilde{c}(x,y) = \int_0^\infty I(x,y,t)\sigma_{an}(x,y,t)dt, \tag{14}$$

with:

$$\begin{aligned}\sigma_{an}(x,y,t) &= \sigma'_{an}(x,y,t) - \sigma''_{an}(x,y,t) \\ &= \chi(x,y)\frac{\nabla I(x,y,t)\cdot\nabla I(x,y,t)}{I(x,y,t)^2} - \chi(x,y) \\ &\quad \times \frac{\nabla I(x,y,t)\cdot\nabla I^*(x,y,t)}{I(x,y,t)I^*(x,y,t)},\end{aligned} \tag{15}$$

where $\sigma'_{an}(x,y,t)$ represents the density of entropy production during anisotropic diffusion, whereas $\sigma''_{an}(x,y,t)$ tends to prevent entropy production, and χ is a non-negative decreasing function of the magnitude of the local image gradient, which forces convergence of the diffusion process toward some desired image representation and $*$ denotes the stationary point of the transformation.

2.2.6. Tadmor and Tolhurst

Tadmor and Tolhurst [30] based their analysis of contrast on the difference of Gaussians (DOG) model, which is modified and adapted to natural images.

In the conventional DOG model, the spatial sensitivity in the center component of the receptive-field is described by a bi-dimensional Gaussian with unit amplitude:

$$Center(x,y) = exp\left[-\left(\frac{x}{r_c}\right)^2 - \left(\frac{y}{r_c}\right)^2\right], \tag{16}$$

where the radius r_c represents the distance at which the sensitivity decreases to $1/e$ with respect to the peak level and (x,y) are the

spatial coordinates of the receptive-field. The surround component is represented by another Gaussian curve, with a larger radius, r_s:

$$Surround(x,y) = 0.85\left(\frac{r_c}{r_s}\right)^2 exp\left[-\left(\frac{x}{r_s}\right)^2 - \left(\frac{y}{r_s}\right)^2\right]. \quad (17)$$

When the central point of the receptive-field is placed at the location (x,y), the output of the central component is calculated as:

$$R_c(x,y) = \sum_{i=x-3r_c}^{i=x+3r_c}\sum_{j=y-3r_c}^{j=y+3r_c} Center(i-x,j-y)I(i,j), \quad (18)$$

while the output of the surround component is:

$$R_s(x,y) = \sum_{i=x-3r_s}^{i=x+3r_s}\sum_{j=y-3r_s}^{j=y+3r_s} Surround(i-x,j-y)I(i,j), \quad (19)$$

where in both cases $I(i,j)$ is the image pixel value at position (i,j). The simplest case, with $r_c = 1$ and $r_s = 2$, results in a 7×7 center mask and a 13×13 surround mask.

The result of the DOG model is obtained as:

$$DOG(x,y) = R_c(x,y) - R_s(x,y). \quad (20)$$

The conventional DOG model assumes that the response of a neuron depends uniquely on the local luminance difference (ΔI) between the center and the surround. After the light adaptation process, the gain of the ganglion cells of the retina and the Lateral Geniculate Nucleus (LGN) neurons depends on the average local luminance I. Thus the model response depends on the contrast stimulus, and the DOG model must be modified by a division by the local mean luminance. They propose the following criterium for the measure of contrast:

$$c^{TT}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_c(x,y) + R_s(x,y)} \quad (21)$$

In their experiments, using 256×256 images, the overall image contrast is calculated as the average local contrast of 1000 pixel locations taken randomly while assuring that the center and surround masks do not exceed the edges of the image:

$$C^{TT} = \frac{1}{1000}\sum_{n=1}^{1000} c_n^{TT} \quad (22)$$

The number of pixels to consider in the calculation should change according to the image size.

2.2.7. Rizzi et al.

Rizzi et al. [26] have developed a very simple and efficient measure, able to estimate global and local components of contrast. It is based on two principles: to collect a simplified measure of difference among neighboring pixels and to do it on various frequency levels. We include this measure because there is evidence, that the use of multilevel is an important implementation feature to mimic the HVS [14,1].

In the rest of the paper we will refer to this contrast measure as RAMM. It performs a pyramid subsampling of the image to various levels in the CIELAB color space [10]. A pyramidal structure is created by halving the image at each iteration. For each level, it calculates the local contrast in each pixel by taking the average difference between the lightness channel value of the pixel and the surrounding eight pixels, thus obtaining a contrast map of each level. The final overall measure is a recombination of the average contrast for each level:

$$C^{RAMM} = \frac{1}{N_l}\sum_{l=1}^{N_l} \bar{c}_l, \quad (23)$$

where N_l is the number of levels and, $\bar{c}_l$ is the mean contrast in the level l:

$$\bar{c}_l = \sum_{i=1}^{i_{max}}\sum_{j=1}^{j_{max}} c_{i,j}, \quad (24)$$

where i_{max} and j_{max} indicate respectively the height and the width of the image, and $c_{i,j}$ is the contrast of each pixel calculated as:

$$c_{i,j} = \sum_{n \in N_8} \alpha_n |Pixel_{i,j} - Pixel_n|.. \quad (25)$$

The following values are used to define the weights of the neighboring pixels:

$$\alpha_n = \frac{1}{4+2\sqrt{2}}\begin{bmatrix} \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \\ 1 & & 1 \\ \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \end{bmatrix} \quad (26)$$

This measure has a computational complexity of $\Theta(NlogN)$, where N is the number of pixels, which is lower than alternative local methods, keeping a comparable level of correctness in the contrast estimate [26]. The steps of the measure are illustrated in Fig. 3(a).

2.3. How many numbers for measuring contrast

As well as concerning the notion of contrast itself, for which a clear and common definition is not found in the literature, there is an ongoing debate on how to reduce the concept of contrast from local values at each pixel location to a single number representing the whole contrast of the image.

It is possible to measure contrast in several ways, with different degrees of summarization. Four of the possible hypothesis are: one number per image, one number per color channel, one number per level using a multilevel approach or one number for each pixel. Here we want to address some of the arguments for and against each hypothesis without exhaustively discussing each detail.

Using one number per image is not sufficiently informative and it easily loses the ability to distinguish among various type of images that usually give rise to different contrast perceptions: e.g. geometric vs natural images. Furthermore several weights, such as the importance of each level and each color channel, must be chosen a priori to aggregate all the components into a single number. On the other hand it has the advantage of being a very compact measure, allowing it for instance to be used as a trigger for image dependent measures. A further advantage is that it can be easily used in comparison with perceptual experiments where observers usually provide a single number for each image.

Using one number per channel keeps chromatic information separated; depending on the color space used (e.g. luminance/chromaticity). For this reason this way of measuring it is color space dependent. Regarding observers, it is more difficult for them to express their judgment keeping channels separated.

Using one number per level allows linking the measure of contrast with the frequency scales in the image, and a posteriori recombine them with different weights. A drawback is that it is image-size dependent and it requires a vector to be stored rather than a single number. Observers can also find it extremely difficult to express a judgment separated per frequency content.

Using one number per pixel has the advantage of keeping the full locality of the information which can be further aggregated into various measures, e.g. for segmented areas or on image subsets. Moreover it can be used for pixelwise measure modification.

As to disadvantages, it is obviously image-size dependent and it requires heavy memory use. Furthermore it needs to take into account masking phenomena and it is completely unsuitable for comparison with observer ratings.

As briefly discussed above, all approaches have positive and negative aspects which make none of them the clear winner.

When the desired number of values for measuring contrast is decided, the method on how to reduce them must be determined. These methods are commonly referred to as pooling. The main motivation for performing pooling is that one value is easier to relate to than hundreds of values.

The simplest pooling strategies are the arithmetic and geometrical mean but Minkowski pooling [11] is perhaps the most common and most popular for HVS based measures.

3. Proposed approaches

As presented in Section 2.2.7, the RAMM measure takes into account only the lightness channel. For this reason our investigation has focused mainly in two directions: first checking whether the use of the DOG model on the multilevel pyramid yields a better performance in considering more extended edges and gradients in the image and, second, whether the use of the chromatic channels in the computation of the perceived contrast leads to more accurate measures. Fig. 2 shows an original image, an image filtered with an 8-neighborhood defined by Eq. (25), and images filtered with the DOG model defined by Eq. (21) with different center and surround radiuses. The superior performance of the DOG model over the 8-neighborhood for the edge detection is outstanding.

We will start proposing an extension of the RAMM measure, the so called "retinal-like subsampling contrast (RSC)" measure, and then will build the so called "weighted-level framework (WLF)" for measuring perceptual contrast in digital images.

3.1. Retinal-like subsampling contrast (RSC)

We have combined RAMM multilevel approach with Tadmor and Tolhurstï£¡s evaluation of a color stimulus [27]. It works with the same pyramid subsampling as RAMM with the following differences: it computes in each pixel of each level the DOG contrast instead of the 8-neighborhood local contrast and it computes the DOG contrast not only for the lightness but also for the two chromatic channels. The three independent measures of each channel are then merged by a weighted linear combination. The final overall measure can be expressed by the formula:

$$C^{RSC} = \alpha \cdot C^{RSC}_{L^*} + \beta \cdot C^{RSC}_{a^*} + \gamma \cdot C^{RSC}_{b^*}, \tag{27}$$

where α, β, and γ represent the weights of each color channel. The steps of the measure are described in Fig. 3(b).

The computational complexity of the RSC measure is the same as for RAMM:

$$\Theta(NlogN), \tag{28}$$

where N is the number of pixels, but with a slightly heavier multiplication constant due to the DOGs instead of the 8-neighbor difference computation.

As well as the previous presented measure, only one number of contrast is produced at the end, the averages of all the levels are averaged again among them with uniform weights.

This measure takes the name of Retinal-like subsampling contrast (RSC) and it derives from the fact that the DOG model has been used successfully in many studies to describe the receptive fields and responses of mammalian retinal ganglion cells and LGN neurones [30] and from the way of building the pyramid structure [26].

3.2. The weighted-level framework (WLF)

In this section we introduce the weighted-level framework (WLF) and we address mainly two aspects: the subsampling method and the weights in the level recombination [29].

An antialiasing filter is introduced in the subsampling in order to minimize distortion artifacts at low resolutions.

As demonstrated by Frankle and McCann [14], Adelson et al. [1] and Peli [23], each level has a different contribution to contrast so we redefine Eq. (23) as follows:

$$C_i = \frac{1}{N_l} \sum_{l=1}^{N_l} \lambda_l \cdot \bar{c}_l, \tag{29}$$

where N_l is the number of levels, $\bar{c}_l$ is the mean contrast in the level l and i indicates the applied color channel as before and the new parameter λ_l is the weight assigned to each level l.

The overall final measure is given by:

$$C^{WLF} = \alpha \cdot C_1 + \beta \cdot C_2 + \gamma \cdot C_3, \tag{30}$$

where α, β, and γ are the weights of each color channel.

The measure can be extended to different color spaces such as XYZ and RGB and it is not limited to CIELAB.

The general structure of our proposed measure can be seen in Fig. 3(c) where the most important novelties are shown in red: the antialiasing filter in the pyramid and weighted recombination of local contrast maps.

In this framework the previously developed RAMM and RSC can be considered as just special cases with uniform weighting levels in the CIELAB color space.

3.2.1. Finding appropriate weights

In this work we have also chosen to summarize the contrast measure in a single number per image. We leave to the reader the possibility to use the proposed measure in a different way, keeping some information separated. This choice allowed us to easily calculate statistics for comparing the results of our framework with the observers perceived contrast.

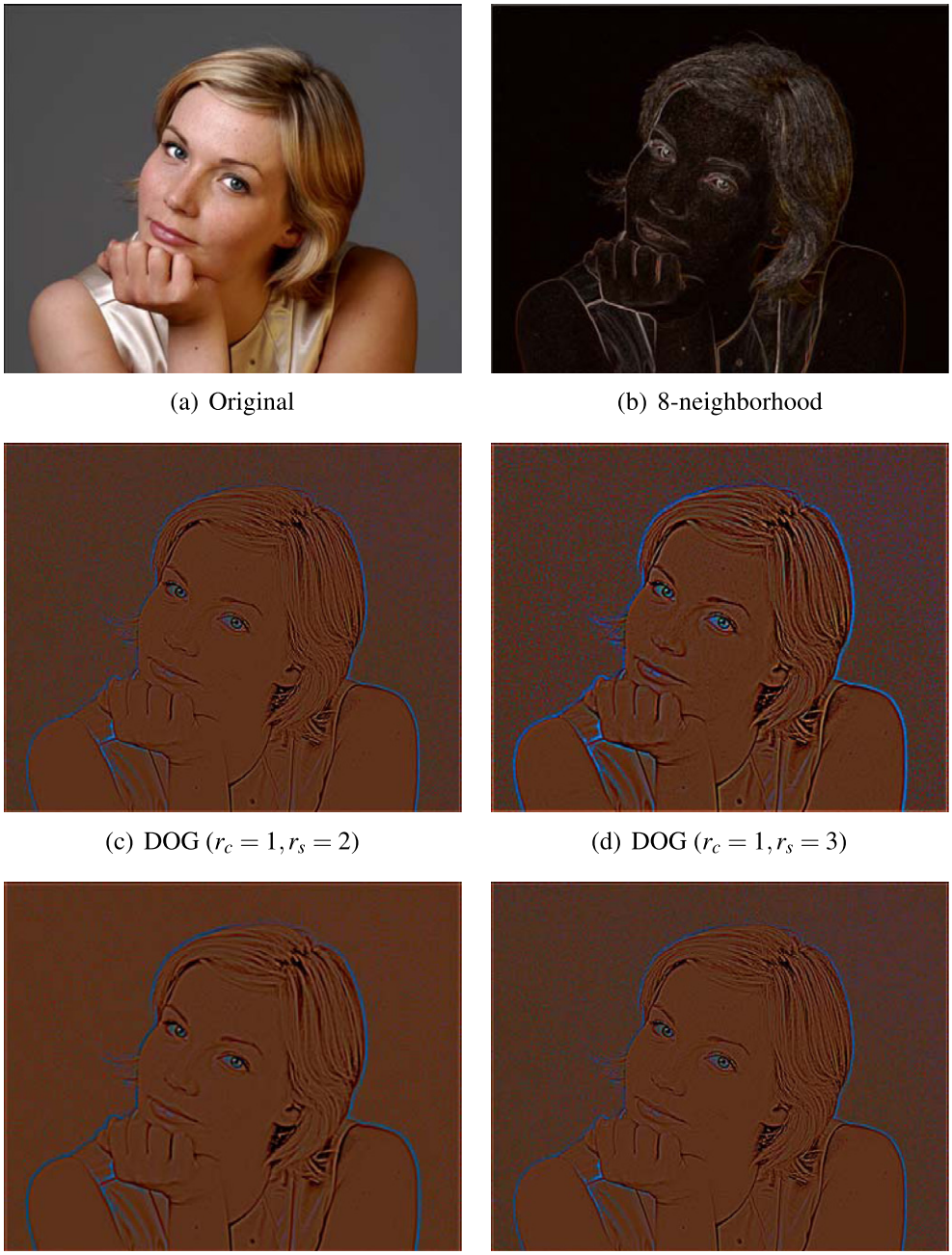
In order to obtain a single contrast number we had to properly tune α, β, γ, and λ. The proposed idea is to give them values from particular measures taken from the image itself. We have tested several alternatives for these parameters. For the parameter λ we have chosen:

- 1
- $\frac{1}{l}$, where l is the corresponding level of the pyramid
- $\frac{1}{m}$, where m is the mean of pixel values in each channel separately at each level of the pyramid
- τ as the variance of pixel values in each channel separately at each level of the pyramid

For the parameters α, β, and γ we have chosen respectively[1]:

- 1, 0, 0
- 0.333, 0.333, 0.333
- 0.5, 0.25, 0.25
- 0, 0.5, 0.5
- $\frac{1}{v}$, where v is the mean of pixel values in each channel separately
- ω as variance of pixel values in each channel separately

[1] Note that the weights in 1, 3, and 4 above only applies to lightness-chromaticity color space such as in our case CIELAB.



(a) Original (b) 8-neighborhood

(c) DOG ($r_c = 1, r_s = 2$) (d) DOG ($r_c = 1, r_s = 3$)

(e) DOG ($r_c = 2, r_s = 3$) (f) DOG ($r_c = 2, r_s = 4$)

Fig. 2. 8-neighborhood against DOGs. All images have been overboosted in luminosity to make them suitable for a printable version of the article.

3.3. Applying Regions-of-Interest weighting to the WLF

Eye tracking has been used in a number of different color imaging research projects with great success, allowing researchers to obtain information on where observers gaze. Gaze information has been used by researches to improve image quality metrics [22,21], using region-of-interest as a weighting map for the metrics. The motivation for incorporating eye tracking as a region-of-interest weighting is to take advantage of the fact that some areas in an image are more important than others.

We use a similar approach and apply gaze information as a weighting map for the contrast measures. Regions that draw attention will be weighted higher than regions that observers do not look at or pay attention to.

From the eye tracking data a number of different maps have been calculated, among them the time used at one pixel multiplied with the number of times the observer fixated on this pixel, the number of fixations at the same pixel, the mean time at each pixel, and the total time for each pixel. All of these have been normalized by the maximum value in the map, and a Gaussian filter corresponding to the 2-degree visual field of the human eye was applied to the map to even out differences [4,3], and to simulate that we look at an area rather than one particular pixel [15].

Gathering gaze information is time consuming and resource demanding, and because of this we have investigated other ways to obtain similar information. One possibility is saliency map, which is a map that represents visual saliency of a corresponding visual scene. One proposed model was introduced by Walther

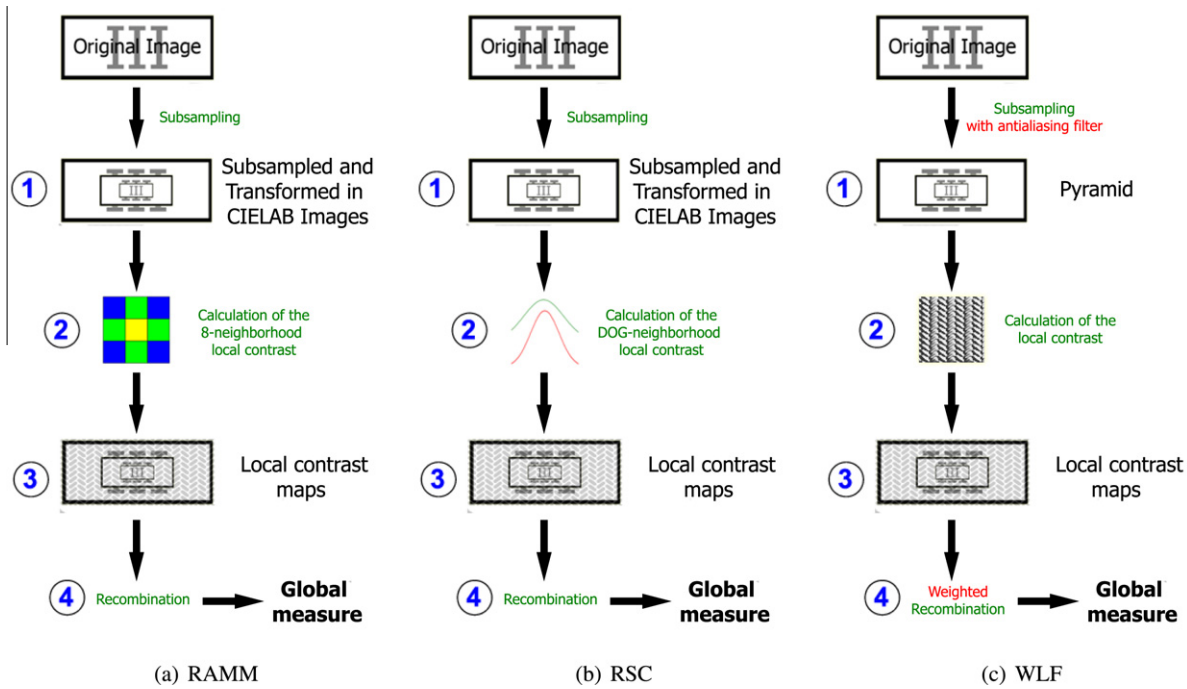


Fig. 3. Workflow for RAMM, RSC, and WLF. The difference between RAMM and RSC is found in the neighborhood calculation of local contrast, where RAMM uses a 8-neighborhood while RSC uses DOG-neighborhood. RAMM and RSC are just special cases with uniform weighting levels in CIELAB color space of WLF, which implements an antialiasing filter in the subsampling, a weighted recombination of the local contrast maps and it is extended also to RGB and XYZ color space.

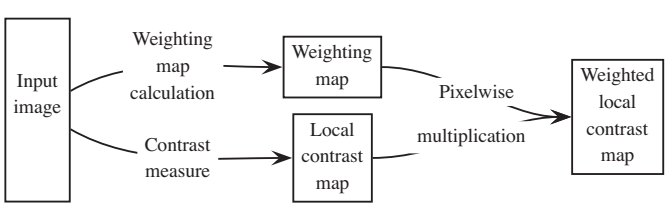


Fig. 4. Framework for using weighting maps with contrast measures. A weighting map is multiplied pixelwise with the relative local contrast map in order to generate a weighted contrast map. As weighting maps we have used gaze maps, saliency maps, and GAFFE maps.

and Koch [32] for bottom-up attention to salient objects, and this has been adopted for the saliency maps used in this study. The saliency map has been computed at level 1 (i.e. the size of the saliency map is equal to original images) and 7 fixations (i.e. giving the seven most salient regions in the image), for the other parameters standard values in the SaliencyToolbox [32] have been used.

Rajashekar et al. [25] proposed gaze-attentive fixation finding engine (GAFFE) based on statistical analysis of image features for fixation selection in natural scenes. The GAFFE uses four foveated low-level image features: luminance, contrast, luminance-bandpass and contrast-bandpass to compute the simulated fixations of a human observer. The GAFFE maps have been computed for 10, 15, and 20 fixations, where the first fixation has been removed since this always is placed in the center resulting in a total of 9, 14, and 19 fixations. A Gaussian filter corresponding to the 2-degree visual field of the human eye was applied to simulate that we look at an area rather than at one single point, and a larger filter (approximately 7-degree visual field) was also tested.

In this work each contrast map of each level is weighted pixelwise with its relative gaze information, or saliency map, or gaze-attentive fixation finding engine as shown in Fig. 4.

We will test the following maps: fixations only, time only, mean time, fixations multiplied with time, saliency, 10 fixations GAFFE map, 15 fixations GAFFE map, 10 fixations big Gaussian GAFFE map, 15 fixations big Gaussian GAFFE map, 20 fixations GAFFE map 20 fixations big Gaussian GAFFE map and six combinations of gaze maps and GAFFE maps.

4. Experimental methodology

In order to evaluate the performance of the measures in relation to perceived contrast a psychophysical experiment has been carried out. In this paper we quantify performance by estimating the correlation between the measure score and the observer score.

Several guidelines have been given in the literature for the selection of images for psychophysical experiment, in the context of investigating image quality issues. Holm et al. [17] recommend the use of a broad range of natural images as well as test charts to reveal the quality issues. The Commission Internationale de l' Èclairage (CIE) [9] suggests to include images with the following characteristics: high-key, low-key, low lightness contrast, leaves and sky, no neutrals, no white point, no black point, heavy cast, few hues, business graphic, and flesh tones. Büring et al. [6] propose to use natural images, as well as saturated colors.

We have chosen 15 different images (Fig. 5), each representing different characteristics, such as different levels of lightness and colorfulness, following the recommendations from above.

17 observers were asked to rate the contrast of these 15 images. Nine of the observers were considered experts, i.e. had experience in color science, image processing, photography or similar and eight were non-experts or with little experience in these fields. All observers were recruited from Gjøvik University College, both students and staff in the ages between 18–45. Observers rated contrast from 1 to 100, where 1 was the lowest contrast and 100 maximum contrast. The observers were told to rate the contrast as they comprehended contrast, i.e. no definition of contrast was given by the researchers before commencing the experiment in order not to influence the observers. All observers had normal or corrected to normal vision. Each image was shown for 40 seconds with the

Fig. 5. Images used in the experiment. Images 1 and 2 are provided by Ole Jakob Bøe Skattum, image 10 is provided by CIE, images 8 and 9 from the ISO 12640-2 standard, images 3, 5, 6, and 7 from Kodak PhotoCD, images 4, 11, 12, 13, 14, and 15 from the ECI Visual Print Reference.

surrounding screen black,[2] and the observers stated the perceived contrast within this time limit. The experiment was carried out on a calibrated CRT monitor, LaCIE electron 22 blue II, in a gray room. The observers were seated approximately 80 cm from the monitor, and the lights were dimmed and measured to approximately 17 lux.

4.1. Statistical analysis

In order to compare the scores of each measure, the perceived contrast score was recorded for each observer for each image and then two types of correlation coefficients (CC) were computed [18]: the Pearson-product-moment CC and the Spearman-rank CC. The Pearson CC assumes that the variables are ordinal and evaluate the linear relationship between two variables. The Spearman CC is a non-parametric measure of correlation and it is used as a measure of linear relationship between two sets of ranked data, instead of the actual values. This describes the relationship between variables with no assumptions on the frequency distribution of the variables and on how tightly the ranked data clusters are around a straight line.

For both correlation coefficients the relative p-value has been calculated. With a p-value of 5% (or 0.05) there is only a 5% chance that results you are seeing would have come up in a random distribution.

4.1.1. Statistical methods for weighting maps

After a short investigation of the results we found that the data cannot be assumed to be normally distributed, and therefore a special care must be given to the statistical analysis. One common method for statistical analysis is the Z-score [12], this requires the data to be normally distributed, and in our case this analysis would not give valid results. Simply using the mean opinion score would also result in problems, since the dataset cannot be assumed to be normally distributed. Because of this we use the rank from each observer to carry out a Wilcoxon signed rank test, a non-parametric statistical hypothesis test. This test does not make any assumption on the distribution, and it is therefore an appropriate statistical tool for analyzing this data set.

In order to test the performance of the contrast measures with different weighting maps and parameters an extensive statistical

[2] The black surround was provided to the observer in order to have as reference the black point (darkest color reproducible).

Table 2
Perceived contrast results for all observers, experts and non-experts. Considering all the observers, image 15 has the highest mean value, while image 13 has the lowest mean value. Image 15 also has the lowest standard deviation. Considering the experts, image 5 has the highest mean value while image 13 has the lowest mean value. Considering the non-experts, image 15 has the highest mean value, and also the lowest standard deviation while image 4 has the lowest mean value. Highest values are highlighted in cyan while lowest values are highlighted in red. (For interpretation of the references to colour in this table legend, the reader is referred to the web version of this article.)

	All observers		Experts		Non-experts	
Image	Mean value	Standard deviation	Mean value	Standard deviation	Mean value	Standard deviation
1	58.7	19.2	65.0	13.8	51.9	22.8
2	57.1	15.4	63.3	15.4	50.0	12.8
3	61.8	14.3	70.6	8.5	51.9	13.1
4	50.3	23.1	58.9	18.3	40.6	25.1
5	70.5	18.7	79.8	10.5	60.0	20.9
6	53.9	19.1	62.4	15.8	44.4	18.6
7	63.8	16.4	67.8	13.0	59.4	19.5
8	57.7	19.1	61.7	15.0	53.1	23.1
9	65.0	22.6	74.1	12.7	54.8	27.5
10	57.7	20.7	67.3	12.0	46.9	23.8
11	59.7	18.6	66.7	17.0	51.9	18.1
12	57.7	24.0	59.8	17.5	55.4	30.9
13	49.0	17.5	46.9	20.9	51.3	13.8
14	61.5	21.7	66.7	16.9	55.6	26.0
15	71.7	10.2	71.7	9.4	71.6	11.6

analysis has been carried out. First the images with high and low perceived contrast has gone through a rank sum test for all 15 images, the p values from this test indicates the ability of the contrast measure to differentiate between the images with perceived low and high contrast. These p values have been used as a basis for a sign test for all parameters for each map and the same for each parameter for all maps. The results from this analysis indicate whether a weighting map is significantly different from no map (standard method), or if a set of parameters is significantly different from other sets of parameters. In the case of a significant difference additional analysis is carried out to indicate whether the performance is better or worse for the tested weighting map or set of parameters.

5. Results and analysis

5.1. Observer rating

Table 2 shows the mean perceived contrast and standard deviations for all observers, experts and non-experts for each image.

Fig. 6 shows the perceived contrast stated by all the observers with a 95% confidence interval. The image rated with the highest mean by the observers is image 15, but it can not be differentiated from many of the other images due to the confidence intervals. The image with the lowest rated contrast is image 13, but this cannot be differentiated from a number of other images as well. This indicates that the perceived contrast of many of the images are statistically similar. Image 12 has the highest standard deviation value, indicating the largest difference between the answers from the observers.

Fig. 7 shows the perceived contrast stated by the experts with a 95% confidence interval. Image 13 has been rated as the image with the lowest contrast, while image 5 has the highest contrast according to the experts. The experts agree most upon image 3, while the highest standard deviation is found in image 13, indicating high deviation in the contrast score for this image.

Fig. 8 shows the perceived contrast stated by the non-experts with a 95% confidence interval. Image 4 is given the lowest contrast, this is also the darkest image i.e. having the lowest mean L^* value. Image 15 is given the highest contrast, this image also has the lowest standard deviation. Image 12 has the highest standard

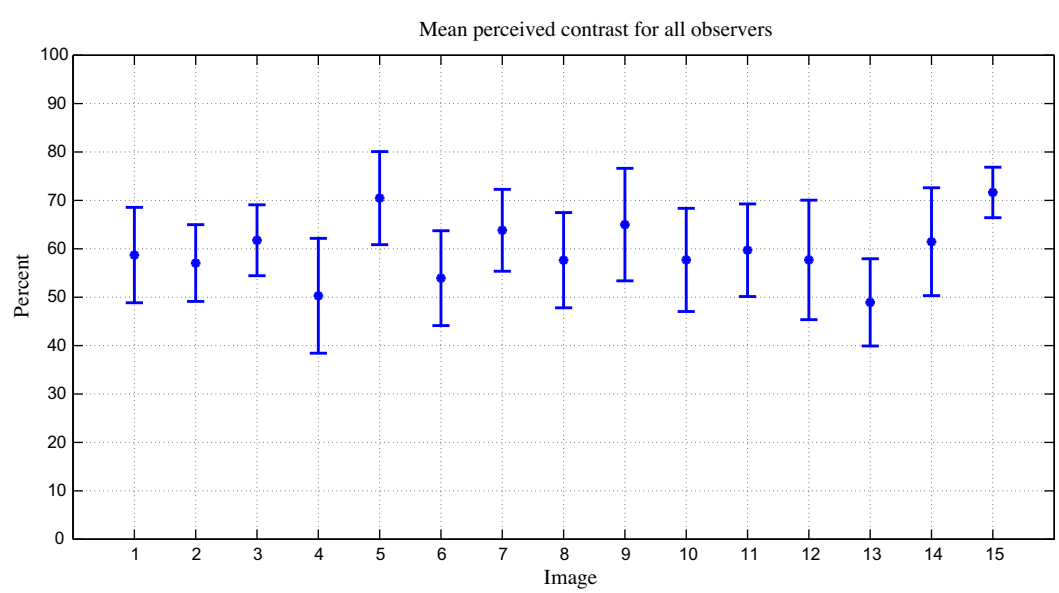


Fig. 6. Mean perceived contrast score for all observers. Image 15 has the highest mean value, while image 13 has the lowest mean value. Image 15 also has the lowest standard deviation, indicating that observers agree about the high contrast in this image.

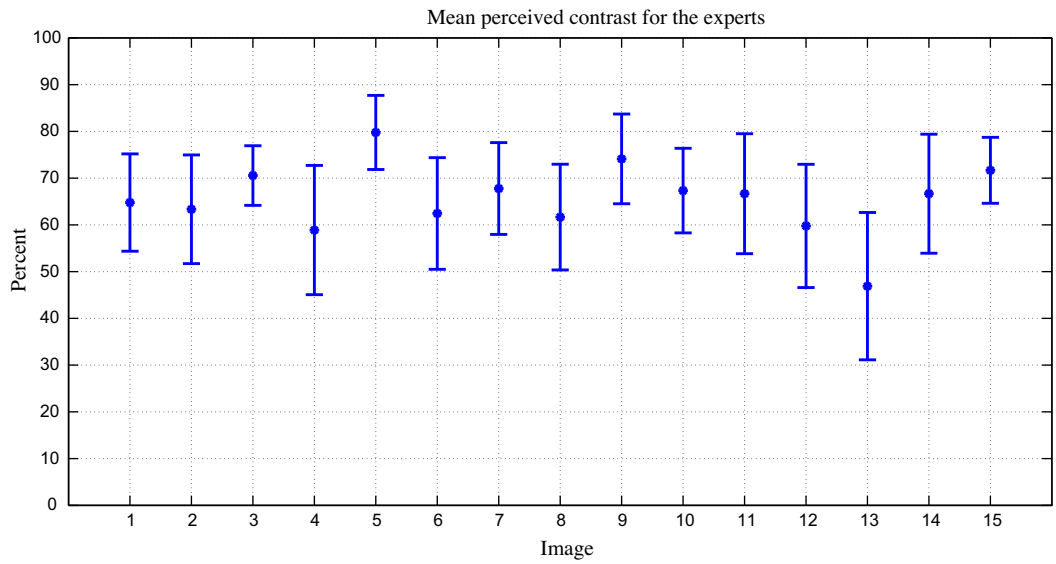


Fig. 7. Mean perceived contrast score for the experts. Image 5 has the highest mean value. Image 13 has the lowest mean value, and also has the highest standard deviation.

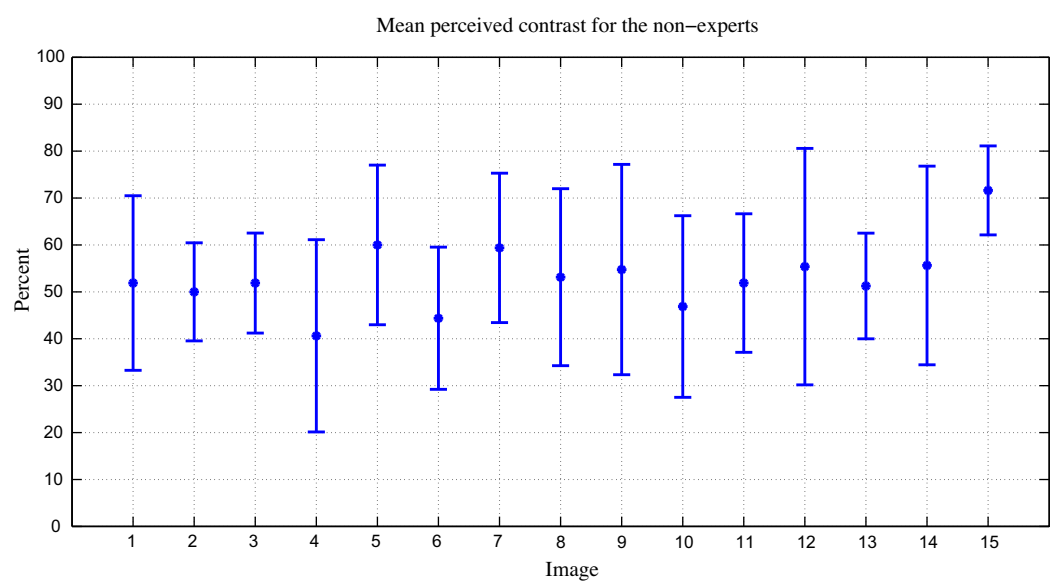


Fig. 8. Mean perceived contrast score for the non-experts. Image 15 has the highest mean value, and also the lowest standard deviation. Image 4 has the lowest mean value.

deviation, indicating a high degree of disagreement among the observers.

There is a clear difference between the experts and non-experts, in 14 of the 15 images the experts have a higher mean value than the non-experts. For image 10 the difference in mean perceived contrast is 20.5 (experts = 67.3, non-experts = 46.9), resulting in a large difference in perceived contrast between the two groups. Image 13 is the only image where the non-experts have a higher mean than the experts, but the difference is low, only 4.4 (experts = 46.9, non-experts = 51.3). The non-experts have used more of the scale than experts, i.e. they have a larger mean difference between the maximum value and minimum value. Experts rate the contrast to be higher then non-experts in most scenes. Experts also agree more upon the contrast in the image than non-experts. We can thus conclude that the perceived contrast in an image depends on the background of the observer.

We have also investigated the connection between mean perceived contrast and standard deviation. For all observers there is a correlation of −0.40 (Fig. 9(a)), while for non-experts only −0.23 (Fig. 9(c)). For the expert observers we have a correlation of −0.83, indicating a high consensus among the experts for the images with higher perceived contrast, and as the perceived contrast decreases the standard deviation increases (Fig. 9(b)).

5.2. Performance of contrast measures

We have tested four different contrast measures with the following abbreviations, PELI [23], TT [30], WLF (RAMM) and WLF (RSC) [29]. We remind that in the framework RAMM and RSC previously developed can be considered just special cases with uniform weighting levels.

Around 10,000 variations have been calculated in the framework developed by Simone et al. [29], so given the huge amount of data we show here only results with Pearson correlation greater than 0.75 for WLF (RAMM) and greater than 0.8 for WLF (RSC), in addition to default parameters.

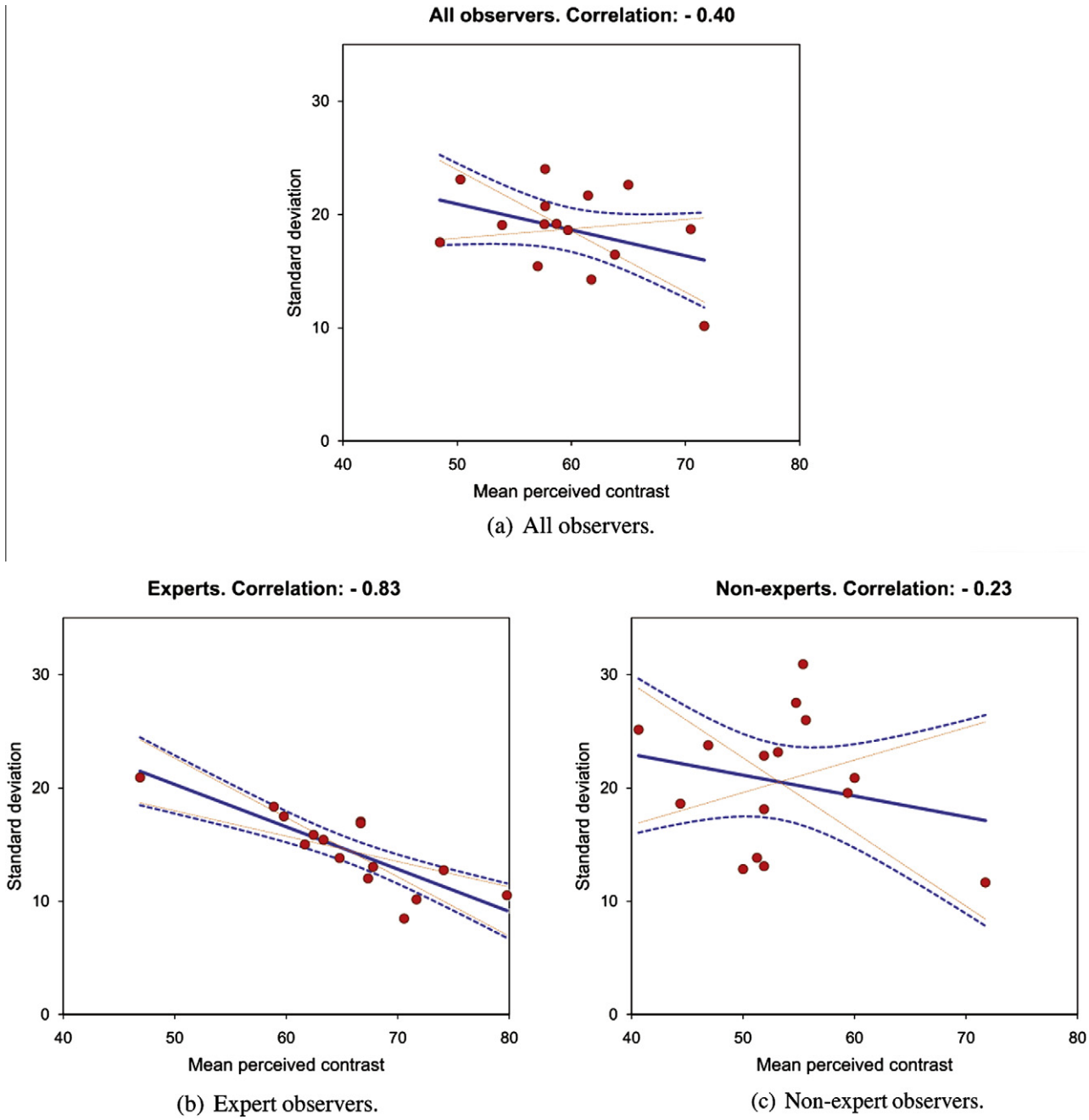


Fig. 9. Correlation between the mean perceived contrast and standard deviation for all, expert, and non-expert observers. For experts high correlation is found, indicating that when the mean perceived contrast decreases the standard deviation increases. This indicates that expert observers agree on the rating of images with high contrast, but when the contrast decreases the observers rate the images more differently. The solid blue line is the linear regression line, the dashed red lines indicate the 95% confidence interval (CI) of the slope of the regression line, and the dotted blue curves indicate the total 95% CI of the linear regression estimate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Tables 3 and 4 show respectively the results using RAMM neighborhood and RSC neighborhood. First we can notice from both tables the absence of results with XYZ color space, which seems to be inappropriate for contrast measure. RGB and CIELAB appear to be equal as results show a difference not greater than ±0.03. We can point out that the parameter λ, which gives a weight to each level of the pyramid, is the key of improving contrast measure especially if λ is equal to τ or, to be more precise equal, to the variance of each channel separately (setting 4 in the list above). Furthermore, unlike previous studies, α, β, and γ lose their importance as we can see that results do not differ so much. It has also to be underlined that the RSC measure works better with greater values of r_c and r_s compared to the standard values given by Tadmor and Tolhurst [30] in their previous studies. The same discussion can be done for the Spearman correlation which is always lower than Pearson correlation. For both Pearson and Spearman correlation very low p-values confirm the usefulness of the measures.

The weighted-level framework using the RSC neighborhood (Fig. 10(d)) reaches higher performance than using RAMM neighborhood (Fig. 10(c)), stating that a DOG neighborhood perform a better evaluation of perceived contrast than a simple 8-neighborhood.

For the PELI measure the size of image must be a power of two, because of this the images have been resized to 512×512. The PELI measure rates the image with the highest and lowest

Table 3
Pearson and Spearman correlation for WLF using RAMM neighborhood. Yellow cells indicate the standard values. Grey cells indicate the highest Pearson and Spearman correlation. All numbers are rounded to two decimals. λ equal to τ indicates that each level of the pyramid is weighted with the variance of the three channels separately. $\frac{1}{v}$ indicates that the considered channel is weighted with the reciprocal of its mean, while ω indicates that it is weighted with its variance. (For interpretation of the references to colour in this table legend, the reader is referred to the web version of this article.)

WLF (RAMM)								
Color space	λ	α	β	γ	Pearson correlation	Pearson *p*-value	Spearman correlation	Spearman *p*-value
CIELAB	1	1	0	0	0.41	0.13	0.40	0.14
CIELAB	1	0.33	0.33	0.33	0.54	0.04	0.47	0.08
RGB	τ	0.33	0.33	0.33	0.78	0.01	0.71	0.00
RGB	τ	ω	ω	ω	0.78	0.00	0.71	0.00
RGB	τ	$\frac{1}{v}$	$\frac{1}{v}$	$\frac{1}{v}$	0.78	0.00	0.69	0.00
CIELAB	τ	0.33	0.33	0.33	0.78	0.00	0.58	0.03
CIELAB	τ	0.50	0.25	0.25	0.79	0.00	0.65	0.02
CIELAB	τ	ω	ω	ω	0.80	0.00	0.68	0.03

Table 4
Pearson and Spearman correlation for WLF using RSC neighborhood. Yellow cells indicate the standard values. Grey cells indicate the highest Pearson and Spearman correlation. Cyan row indicates the best result achieved by the framework. All numbers are rounded to two decimals. λ equal to τ indicates that each level of the pyramid is weighted with the variance of the three channels separately. $\frac{1}{v}$ indicates that the considered channel is weighted with the reciprocal of its mean, while ω indicates that it is weighted with its variance. (For interpretation of the references to colour in this table legend, the reader is referred to the web version of this article.)

WLF (RSC)										
Color space	r_c	r_s	λ	α	β	γ	Pearson correlation	Pearson *p*-value	Spearman correlation	Spearman *p*-value
CIELAB	1	2	1	0.333	0.333	0.333	−0.23	0.42	−0.21	0.44
CIELAB	1	2	1	0.50	0.25	0.25	−0.16	0.57	−0.20	0.48
CIELAB	1	2	1	1	0	0	0.02	0.95	−0.21	0.45
RGB	1	2	τ	0.33	0.33	0.33	0.82	0.00	0.68	0.01
RGB	2	4	τ	$\frac{1}{v}$	$\frac{1}{v}$	$\frac{1}{v}$	0.82	0.00	0.72	0.00
CIELAB	3	4	τ	0.50	0.25	0.25	0.82	0.00	0.72	0.00
RGB	2	3	τ	$\frac{1}{v}$	$\frac{1}{v}$	$\frac{1}{v}$	0.83	0.00	0.77	0.00
CIELAB	2	3	τ	0.50	0.25	0.25	0.83	0.00	0.73	0.00
CIELAB	2	4	τ	0.50	0.25	0.25	0.83	0.00	0.82	0.00
RGB	2	4	τ	0.33	0.33	0.33	0.84	0.00	0.79	0.00
RGB	2	4	τ	ω	ω	ω	0.84	0.00	0.80	0.00

perceived contrast to have approximately the same score, an image with a medium perceived contrast is rated as the image with the lowest contrast by PELI. This results in a Pearson correlation of 0.32 (Table 5 and Fig. 10(a)). The PELI measure has a scattering of the data points, where the image rated with the highest contrast and lowest contrast by the observers receive similar PELI scores. This results in a low Pearson correlation (Table 5). For the Spearman correlation the PELI measure shows an increased correlation, where the correlation is 0.43 (Table 5). This is due by the fact that the Spearman CC looks at the rank of the results, and even though PELI has a spread scatter plot, the order (rank) of the results is more similar to the observers. In this sense the distribution of Peli contributes to a low Pearson correlation, but when we discard the distribution and only look at the rank it performs better. However, it has lower performance than WLF.

The TT measure has a low correlation between the perceived contrast and predicted contrast (Table 5 and Fig. 10(b)). Four of the images with the highest perceived contrast have been rated in the lower half of the TT scale, resulting in a low performance.

The analysis in this section needs to be put in perspective with the variations in the observer results.

5.3. Results with gaze information

The perceived contrast for the 15 images (Fig. 5) from 17 observers were gathered. The 15 images have been grouped into three groups based on the Wilcoxon signed rank test: high, medium and low contrast. From the signed rank test observers can differentiate between the images with high and low contrast, but not between high/low and medium contrast. Images 5, 9, and 15 have high contrast while images 4, 6, 8, and 13 have low contrast. This is further used to analyze the performance of the different contrast measures and weighting maps.

We have tested many different weighting maps, and due to page limitations we cannot show all results. We will show results for fixations only, fixations multiplied with time, saliency, 10 fixation GAFFE map (GAFFE10), 20 fixations big Gaussian GAFFE map (GAFFEBG20) and no map (normally unweighted contrast measure). The maps that were excluded are time only, mean time, 15 fixations GAFFE map, 20 fixations GAFFE map, 10 fixations big Gaussian GAFFE map, 15 fixations big Gaussian GAFFE map, and 6 combinations of gaze maps and GAFFE maps. The maps that have been excluded show no significant difference from no map, or have a lower performance than no map.

As we can see from Tables 6 and 7, using maps is not significantly different from not using them as they have the same performance at a 5% significance level. We can only see a difference between saliency maps and gaze maps (fixation only and fixations ×time), but since these are not significantly different from no map they do not increase the contrast measure's ability to predict perceived contrast. The contrast measures with the use of maps have been tested in the framework developed by Simone et al. [28] with different settings shown in Tables 8 and 9. For RAMM the standard parameters (LAB-1–1-0–0–0 and LAB-1–0.33–0.33–0.33) perform significantly worse than the other parameters in the table, having a *p*-value greater than 0.05. For RSC we noticed that three parameters are significantly different from the standard parameters (LAB-1–2–1–0.33–0.33–0.33 and LAB-1–2–1–0.50–0.25–0.25), but after further analysis of the underlying data these perform worse than the standard parameters.

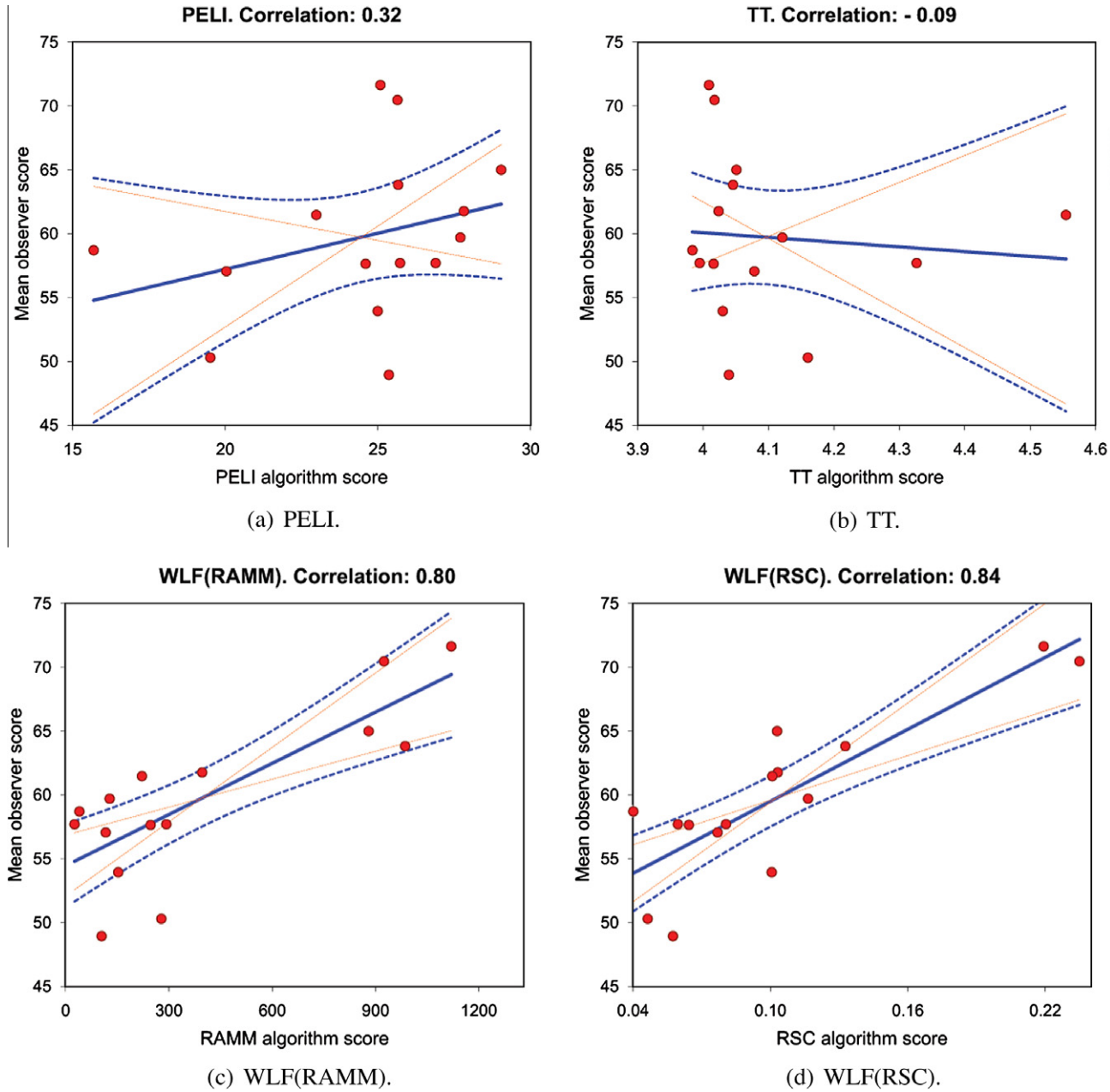


Fig. 10. Pearson correlation between observer mean score and PELI, TT, WLF (RAMM), and WLF (RSC) measure score respectively. The solid blue line is the linear regression line, the dashed red lines indicate the 95% confidence interval (CI) of the slope of the regression line, and the dotted blue curves indicate the total 95% CI of the linear regression estimate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Pearson and Spearman correlation of the four tested measures. The TT measure has the lowest correlation between the perceived contrast and predicted contrast while WLF (RSC) has the highest.

Measure	Pearson correlation	Pearson *p*-value	Spearman correlation	Spearman *p*-value
PELI	0.32	0.01	0.43	0.01
TT	−0.09	0.73	−0.19	0.51
WLF (RAMM)	0.80	0.001	0.68	0.03
WLF (RSC)	0.84	0.00	0.80	0.00

Table 6
Statistical significance of using different maps with WLF (RAMM). If the weighting maps are significantly different than using no weighting map then the value in the table is below 0.05. With these results we can see that the different weighting maps have the same performance as no map at a 5% significance level, indicating that weighting RAMM with maps does not improve predicted contrast.

Map	fixation only	Fixation *x* time	Saliency	GAFFE10	GAFFEBG20	No map
Fixation only	–					
Fixation *x* time	1.00	–				
Saliency	0.63	1.00	–			
GAFFE10	0.25	0.25	0.25	–		
GAFFEBG20	0.13	0.38	1.00	0.06	–	
No map	0.50	0.50	0.63	1.00	0.06	–

Table 7
Statistical significance of using different maps with WLF (RSC). If the weighting maps are significantly different than using no weighting map then the value in the table is below 0.05. We can see that no maps are significantly different than no map, indicating that the have the same performance at a 5% significance level. We can see a difference between saliency maps and gaze maps (fixation only and fixation x time), but since these are not significantly different from no map they do not increase the contrast measure's ability to predict perceived contrast. Gray cells indicate significant difference at a 5% significance level.

Map	Fixation only	Fixation *x* time	Saliency	GAFFE10	GAFFEBG20	No map
Fixation only	–					
Fixation *x* time	1.00	–				
Saliency	0.02	0.03	–			
GAFFE10	0.29	0.51	1.00	–		
GAFFEBG20	0.23	0.227	0.73	0.69	–	
No map	0.50	1.00	0.13	0.73	0.34	–

Table 8
Statistical significance in terms of *p*-values of using different parameters with WLF (RAMM). Gray cells indicate significant difference at a 5% (0.05) significance level between the results of the two configurations. RAMM parameters are the following: color space (CIELAB or RGB), pyramid weight, and the three last parameters are channel weights. "var" indicates the variance.

Parameters	LAB-1–	LAB-1–	RGB-4–	LAB-4–	LAB-4–	LAB-4–
	1-0–0	0.33–0.33–0.33	var1-var2-var3	0.33–0.33–0.33	0.50–0.25–0.25	var1-var2-var3
LAB-1–1–0–0	–					
LAB-1–0.33–0.33–0.33	0.09	–				
RGB-4-var1-var2-var3	0.00	0.01	–			
LAB-4–0.33–0.33–0.33	0.00	0.01	1.00	–		
LAB-4–0.50–0.25–0.25	0.00	0.00	0.50	1.00	–	
LAB-4-var1-var2-var3	0.00	0.00	0.50	1.00	1.00	–

Table 9
Statistical significance in terms of *p*-values of using different parameters with WLF (RSC). Gray cells indicate significant difference at a 5% (0.05) significance level between the results of the two configurations. RSC parameters are the following: color space (CIELAB or RGB), radius of the center Gaussian, radius of the surround Gaussian, pyramid weight, and the three last parameters are channel weights. "*m*" indicates the mean.

Parameters	LAB-1–2–1–	LAB-1–2–1–	LAB-1–2–1–	RGB-1–2–4–	RGB-2–4–4–	RGB-2–3–4–	LAB-2–3–4–
	0.33–0.33–0.33	0.5–0.25–0.25	1-0–0	0.33–0.33–0.33	*m*1–*m*2–*m*3	*m*1–*m*2–*m*3	0.50–0.25–0.25
LAB-1–2–1–0.33–0.33–0.33	–						
LAB-1–2–1–0.5–0.25–0.25	1.00	–					
LAB-1–2–1–1-0–0	0.00	0.00	–				
RGB-1–2–4–0.33–0.33–0.33	0.45	0.45	0.00	–			
RGB-2–4–4–*m*1–*m*2–*m*3	0.00	0.00	0.58	0.00	–		
RGB-2–3–4–*m*1–*m*2–*m*3	0.00	0.00	0.77	0.00	0.22	–	
LAB-2–3-4–0.50–0.25–0.25	0.29	0.29	0.00	0.00	0.00	0.00	–

We can see from Fig. 11 that using a saliency map for weighting discards relevant information used by the observer to judge perceived contrast since contrast is a complex feature and it is judged by the global impression of the image.

Furthermore, in order to validate the results from the other dataset we have carried out the same analysis for 25 images each with four contrast levels from the TID2008 database [24]. The score from the two contrast measures have been computed for all 100 images, and a similar statistical analysis is carried out as above but for four groups (very low contrast, low, high and very high contrast). The results from this analysis supports the findings from the first dataset.

In conclusion using weighting maps from gaze information, saliency maps, and GAFFE maps do not improve the WLF to predict perceived contrast in digital images.

6. Conclusions

We have developed a weighted-level framework (WLF) to measure perceptual contrast. It performs for each channel a pyramid subsampling of the image to various levels. A pyramidal structure is created halving the image at each iteration with prefiltering in order to avoid aliasing at low resolutions. For each level, it calculates the local contrast of each pixel using the Difference of Gaussians, thus obtaining a contrast map of each level. The overall measure of each channel is a weighted recombination of the average contrast for each level. The final global measure is given by a weighted sum of the contrast of each channel.

In conclusion, after testing the framework, we propose a measure of contrast, based on RSC neighborhood computation, that is called $C^{WLF(RSC)}$ and defined as follows:

$$C^{WLF(RSC)} = \omega_R \cdot C_R^{RSC} + \omega_G \cdot C_G^{RSC} + \omega_B \cdot C_B^{RSC}, \tag{31}$$

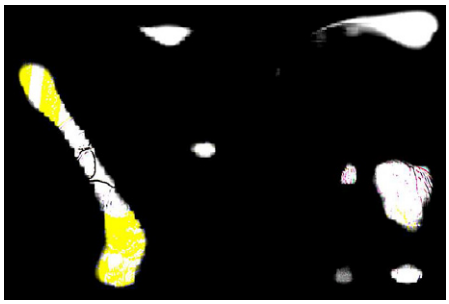
where

$$\begin{aligned} C_R^{RSC} &= \frac{1}{N_l}\sum_{l}^{N_l} \tau_{R,l} \cdot \bar{c}_{R,l}, \\ C_G^{RSC} &= \frac{1}{N_l}\sum_{l}^{N_l} \tau_{G,l} \cdot \bar{c}_{G,l}, \\ C_B^{RSC} &= \frac{1}{N_l}\sum_{l}^{N_l} \tau_{B,l} \cdot \bar{c}_{B,l}, \end{aligned} \tag{32}$$

where N_l is the number of levels, $\bar{c}_{R,l}$, $\bar{c}_{G,l}$, and $\bar{c}_{B,l}$ are respectively the mean contrast in the level l for R, G and B channel, $\tau_{R,l}$,$\tau_{G,l}$, and $\tau_{B,l}$ are respectively the variance of pixel values in the level l for R, G, and B channel, ω_R, ω_G, and ω_B are respectively the variance of pixel values for R, G, and B channel used for recombining the overall contrast of each channel C_R^{RSC}, C_G^{RSC} and C_B^{RSC}. A performance of 0.84 in Pearson correlation and 0.80 in Spearman correlation is achieved in relation to observers perceived contrast.

(a) Original

(b) Relative local contrast map



(c) Saliency weighted local contrast map

Fig. 11. The original, the relative local contrast map and saliency weighted local contrast map. Fig. 11(c) clearly shows how saliency maps discard relevant information used by the observer to judge perceived contrast.

We have tried to improve the WLF using weighting maps, from gaze information, saliency maps, and GAFFE maps. None of them improve the WLF to predict perceived contrast in digital images as relevant information used by the observer to judge perceived contrast are mainly discarded. This suggests that Regions-of-Interest cannot be used to improve contrast measures as contrast is an intrinsic factor and is judged by a global impression of the image. This indicates that further work on contrast measures should be carried out accounting for the global impression of the image while preserving the local information. Future work can also follow different experimental methodologies to record perceived contrast by the observers: one opportunity is to reperform exactly the same experiment in uncontrolled environments while a second one is to use pairwise comparison to rank the the images.

Acknowledgment

The authors thank Alessandro Rizzi (University of Milano, Italy), Ivar Farup (Gjøvik University College, Norway) and Fritz Albregtsen (University of Oslo, Norway) for their interesting discussions and feedback.

References

[1] E.H. Adelson, C.H. Anderson, J.R. Bergen, P.J. Burt, J.M. Ogden, Pyramid methods in image processing, RCA Engineer 29 (6) (1984).
[2] A.J. Ahumada, B.L. Beard, A simple vision model for inhomogeneous image quality assessment, in: SID International Digest of Technical Papers, 1998.
[3] J.S. Babcock, Eye tracking observers during color image evaluation tasks, Master's thesis, Rochester Institute of Techonology, 2000
[4] J. Bai, T. Nakaguchi, N. Tsumura, Y. Miyake, Evaluation of image corrected by retinex method based on *S*-CIELAB and gazing information, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E89-A (11) (2006) 2955–2961.
[5] G. Boccignone, M. Ferraro, T. Caelli, Encoding visual information using anisotropic transformations, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 207–211. URL <http://www.computer.org/tpami/tp2001/i0207abs.htm>.
[6] H. Büring, P.G. Herzog, E. Jung, Evaluation of current color management tools: Image quality assessments, in: 2nd European Conference on Color in Graphics, Imaging and Vision (CGIV), Aachen, Germany, 2004, pp. 459–462.
[7] D.A. Burkhardt, J. Gottesman, D. Kersten, G.E. Legge, Symmetry and constancy in the perception of negative and positive luminance contrast, Journal of Optimizations Society of American A 1 (3) (1984) 309. URL <http://josaa.osa.org/abstract.cfm?URI=josaa-1-3-309>.
[8] A.J. Calabria, M.D. Fairchild, Perceived image contrast and observer preference II. Empirical modeling of perceived image contrast and observer preference data, The Journal of Imaging Science and Technology 47 (2003) 494–508.
[9] CIE, 156, Guidelines for the evaluation of gamut mapping algorithms, Technical Report, ISBN: 3-901-906-26-6, 2004, CIE TC8-08.
[10] CIE, Colorimetry, Technical Report, 15, 2004.
[11] H. de Ridder, Minkowski-metric as a combination rule for digital-image-coding imparaments, in: B. Rogowitz (Ed.), Human Vision, Visual Processing, and Digital Display III, vol. 1666, SPIE/ IS& T, San Jose, CA, 1992, pp. 16–26.
[12] P.G. Engeldrum, Psychometric Scaling, Imcotek Press, Winchester, MA, USA, 2000.
[13] M. Ferraro, G. Boccignone, Image contrast enhancement via entropy production, Real-Time Imaging 10 (4) (2004) 229–238.
[14] J. Frankle, J. McCann, Method and apparatus for lightness imaging United States Patent, no. 4, 1983, p. 336–384.
[15] J.M. Henderson, C.C. Williams, M.S. Castelhano, R.J. Falk, Eye movements and picture processing during recognition, Perception & Psychophysics 65 (2003) 725–734.
[16] R.F. Hess, A. Bradley, L. Piotrowski, Contrast-coding in amblyopia. I. Differences in the neural basis of human amblyopia, Proceedings of Royal Society of London Series B (217) (1983) 309–330.
[17] J. Holm, I. Tastl, T. Johnson, Definition & use of the iso 12640-3 reference color gamut, in: Fourteenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications, IS&T/SID, Scottsdale, AZ, 2006, pp. 62–68.
[18] M.G. Kendall, A. Stuart, J.K. Ord, Kendall's Advanced Theory of Statistics: Classical inference and relationship, 5th ed., vol. 2, A Hodder Arnold Publication, 1991.
[19] P.E. King-Smith, J.J. Kulikowski, Pattern and flicker detection analysed by subthreshold summation, Journal of Physiology 249 (3) (1975) 519–548. URL <http://view.ncbi.nlm.nih.gov/pubmed/1177103>.
[20] A. Michelson, Studies in Optics, University of Chicago Press, 1927.
[21] M. Pedersen, Importance of region-of-interest on image difference metrics. Master thesis, Gjøvik University College, 2007. <http://www.hig.no/content/download/9052/122117/file/Pedersen%20-%20Importance%20of%20region-of-interest%20on%20image%20difference%20metrics%20-%20SCREEN.pdf>.
[22] M. Pedersen, J.Y. Hardeberg, Rank order and image difference metrics, in: CGIV 2008 Fourth European Conference on Color in Graphics, Imaging and Vision, IS&T, Terrassa, Spain, pp. 120–125.
[23] E. Peli, Contrast in complex images, Journal of the Optical Society of America 7 (1990) 2032–2040.

[24] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, F. Battisti, Color image database for evaluation of image quality metrics, 2008, pp. 403–408 <http://www.ponomarenko.info/tid2008.htm>.
[25] U. Rajashekar, I. van der Linde, A.C. Bovik, L.K. Cormack, Gaffe: a gaze-attentive fixation finding engine, in: IEEE transactions on image processing, vol. 17, April 2008, pp. 564–573, <http://live.ece.utexas.edu/research/gaffe>.
[26] A. Rizzi, T. Algeri, G. Medeghini, D. Marini, A proposal for contrast measure in digital images, in: CGIV 2004 – Second European Conference on Color in Graphics, Imaging and Vision, Aachen, April 2004, Germany.
[27] A. Rizzi, G. Simone, R. Cordone, A modified algorithm for perceived contrast in digital images, in: CGIV 2008 – Fourth European Conference on Color in Graphics, Imaging and Vision, IS&T, Terrassa, Spain, June 2008, pp. 249–252.
[28] G. Simone, M. Pedersen, J.Y. Hardeberg, I. Farup, On the use of gaze information and saliency maps for measuring perceptual contrast, in: Computer Science, L.N. (Ed.), 16th Scandinavian Conference on Image Analysis, vol. 5575, Oslo, Norway, 2009, pp. 597–606.
[29] G. Simone, M. Pedersen, J.Y. Hardeberg, A. Rizzi, Measuring perceptual contrast in a multilevel framework, in: B. Rogowitz, T. Pappas (Eds.), Human Vision and Electronic Imaging XIV, vol. 7240, SPIE, 2009. pp. 72400Q–72400Q–9.
[30] Y. Tadmor, D. Tolhurst, Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes, Vision Research 40 (2000) 3145–3157.
[31] A. Tremeau, Color contrast parameters for analysing image differences, in: Colour, Institute, I. (Eds.), Proceedings of the University of Derby Color Image Science 2000 Conference, Derby, 2000, pp. 11–23.
[32] D. Walther, C. Koch, Modeling attention to salient proto-objects, Neural Networks 19 (2006) 1395–1407.
[33] P. Whittle, Increments and decrements: luminance discrimination, Vision Research (26) (1986) 1677–1691.