Evaluation of contrast measures in relation to observers perceived contrast

Marius Pedersen[1]*, Alessandro Rizzi*[2]*, Jon Yngve Hardeberg*[1] *and Gabriele Simone*[1]*.*

1: Gjøvik University College; Gjøvik, Norway. 2: University of Milano; Milano, Italy.

Abstract

We have carried out a psychophysical experiment to register perceived contrast. 17 observers viewed 15 images, each image was shown for 40 seconds where the observer stated the perceived contrast of the image. The results from the observers indicate that the consensus of contrast among experts decreases as the perceived contrast decreases. Experts also rate the contrast higher then non-experts. A number of contrast algorithms, developed to predict perceived contrast was evaluated against the perceived contrast from the observers.

Introduction

Contrast can be defined as the difference between the light and dark parts of a photograph. Where less contrast gives a flatter picture, and more a deeper picture. This is only one of the defintions of contrast, others are the difference in visual properties that makes an object distinguishable or just the difference in color from point to point. Because various definitions of contrast is used in different situation, measuring contrast is very difficult. Just measuring the difference between dark and light points of the image does not predicted perceived contrast because perceived contrast is influenced by the surround. Parameters as resolution, viewing distance, enlightment, memory color etc. will all effect how observers perceive contrast. It is clear that contrast is local, but how local does it have to be?

A measure of perceived contrast in images is not clearly defined, several measures to predict perceived contrast have been proposed. It is very important in many fields to predict perceived contrast correctly, in image quality assessment and for displays where correct contrast is important. The goal of this research is to evaluate contrast measures, the predicted contrast by the measures are compared against perceived contrast from a psychophysical experiment.

Wandell and Zhang [1] found out that S-CIELAB had problems with images with negative contrast (i.e. the luminance value of a point is higher than the local mean of the that point), and they concluded that the ability to predict perceived contrast is very important for image difference models [1]. This is also noted by Wang et al. [2] where the SSIM model incorporate a comparison of contrast to predict image quality. This method has further been incorporated by many researchers [3, 4, 5]. Taylor et al. [6] incorporated a contrast measure in their image fidelity measure. McCann [7] also states the importance of a metric that predict contrast in order to calculate the best image appearance.

Peli [8] proposed a local contrast measure (PELI). This important characteristic makes it suitable for the use on natural images. To obtain an efficient measure of contrast, it is necessary to apply the following steps:

- The use of a pyramidal structure of band-pass filters (with a width equal to one octave of the bandwidth) centered on different frequencies and distanced one octave from each other. The image is then filtered from the pyramid to obtain a further series of images, each one representing a portion of the image at a prefixed frequency.
- The average luminance is calculated at each level frequency.
- Every pixel of the image is divided by the average luminance, obtaining a local contrast measure at each level, on a limited bandwidth for every frequency.

Tadmor and Tolhurst [9] analysis of contrast (TT) is based on the difference of gaussian model (*DOG*), modified and adapted to natural images.

In the conventional model, the spatial sensitivity in the center of receptive-fields is described by a bi-dimensional Gaussian with amplitude 1.0: $Center(x,y) = exp[-(x/r_c)^2(y/r_c)^2]$ where the radius r_c represents the distance beyond which the sensitivity decreases following $1/e$ with respect to the peak level. The surround component is represented by another Gaussian curve, with a larger radius, r_s: $Surround(x,y) = 0.85(r_c/r_s)^2 exp[-(x/r_c)^2(y/r_s)^2]$. When the central point of the receptive-field is placed at the (x,y), the output of central component is calculated as: $R_c(x,y) = \sum i \sum jCentre(i-x, j-y)Picture(i,j)$, while the output of the surround component is: $R_s(x,y) = \sum i \sum jCentre(i-x, j-y)Picture(i,j)$. The result of the DOG model is obtained as: $DOG(x,y) = R_c(x,y)/R_s(x,y)$. 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 gangliar 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. They propose the following three criteria for the measure of contrast:

$$C(x,y) = [R_c(x,y) - R_s(x,y)]/R_c(x,Y)$$
$$C(x,y) = [R_c(x,y) - R_s(x,y)]/R_s(x,Y)$$
$$C(x,y) = [R_c(x,y) - R_s(x,y)]/[R_c(x,y) + R_s(x,y)]$$

Rizzi et al. [10] proposed a contrast measure (RAMMG) in 2004. This algorithm subsample the image to various levels in the CIELAB colorspace, the undersampling is simple where the image is halfed without pre-filtering. Then calculating local contrast by taking the difference between one pixel and the surrounding 8 pixels, obtaining transition maps of each level. A recombination of the of the averages for each level results in the global measure. This measure was evaluated by changing contrast in different softwares, and comparing the predicted contrast against this.

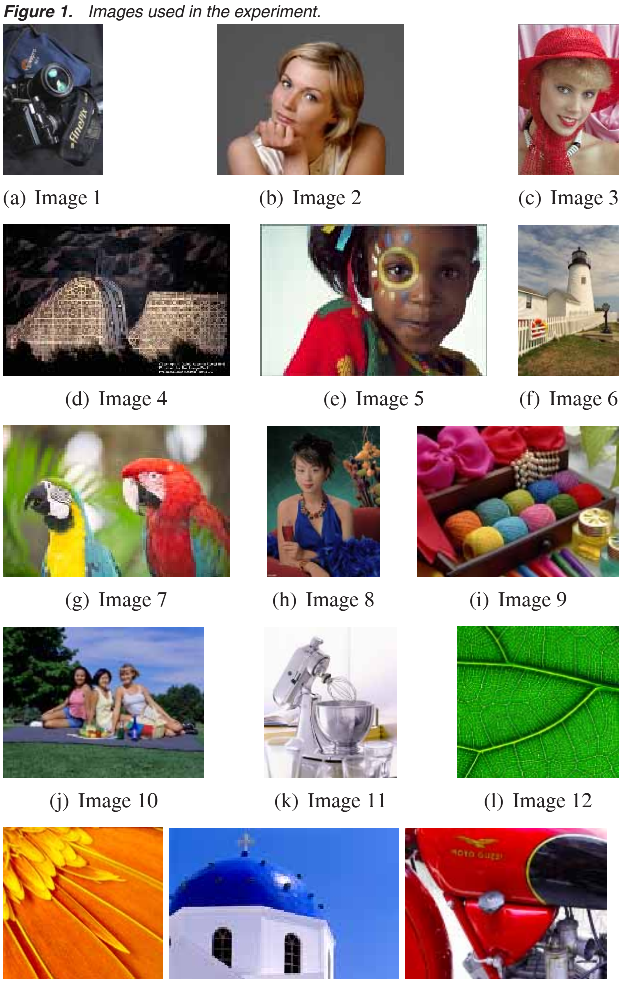
Rizzi et al. propose the RSC algorithm [11], it combines Rizzi et al's [10] multilevel approach with Tadmor and Tolhurst's [9] evaluation of a color stimulus. After computing all subsampled images creating a pyramidal image structure starting from the given image, it executes a neighborhood contrast calculation for every pixel in each level using *DOG* on the lightness and on the chromatic channels separately. Unlike all other algorithms, in order to consider also isoluminant color contrast configurations, also chromaticity planes of the CIELab space are used, weighted differently than L. This algorithm derives from RAMMG in which the *DOG*'s substitute the simple neighboorod differences. The attempt is to investigate mainly two directions: first checking if the use of *DOG*'s on the multilevel pyramid have a better performance in considering more extended edges and gradients and second if the use of the chromatic channels in the computation of the perceived contrast lead to more accurate measures. As well as the previous presented measure, only one number of contrast is produced at the end.

Calabria and Fairchild [12] carried out an experiment on a set of images changed with different lightness, chroma and sharpness levels. No large differences between experts and non-experts when it came to rating contrast was found, but there were a larger variability among the non-experts than for experts. It was also identified that observers rated contrast in grayscale images different than for color images, perceived contrast in achromatic images are higher than perceived contrast of very low-chroma images.

Experiment Setup

15 different images have been used in this experiment (Figure 1), representing different characteristics. Images 1 and 2 are provided by Ole Jakob Bøe Skattum, image 10 is provided by CIE, images 8 and 9 from ISO 12640-2 standard, images 3, 5, 6 and 7 from Kodak PhotoCD, images 4, 11, 12, 13, 14 and 15 from ECI Visual Print Reference. 17 observers were asked to rate the contrast in the 15 images. 9 of the observers were experts, i.e. had experience in color science, image processing, photography or similar and 8 non-experts non or little experience in these fields. All observers were recruited from Gjøvik University College, both students and employees. 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 made by the researchers before commencing the experiment. All observers had normal or corrected to normal vision. Each image was shown for 40 seconds with the surrounding screen black, and the observers stated the perceived contrast within this timelimit. 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 [13] from the monitor, and the lights were dimmed and measured to approximately 17 lux.

Figure 1. *Images used in the experiment.*



(a) Image 1 (b) Image 2 (c) Image 3
(d) Image 4 (e) Image 5 (f) Image 6
(g) Image 7 (h) Image 8 (i) Image 9
(j) Image 10 (k) Image 11 (l) Image 12
(m) Image 13 (n) Image 14 (o) Image 15

Results

This section contains results from both the psychophysical experiment and from the different algorithms, a comparison between these is also carried out.

Perceived contrast

Figure 2 shows the perceived contrast stated by the observers with a 95% confidence interval and Table 1 shows mean values and standard deviation for each image. The image rated with the highest mean by the observers is image 15, but it can not be differeniated 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 differenated from a number of other images.

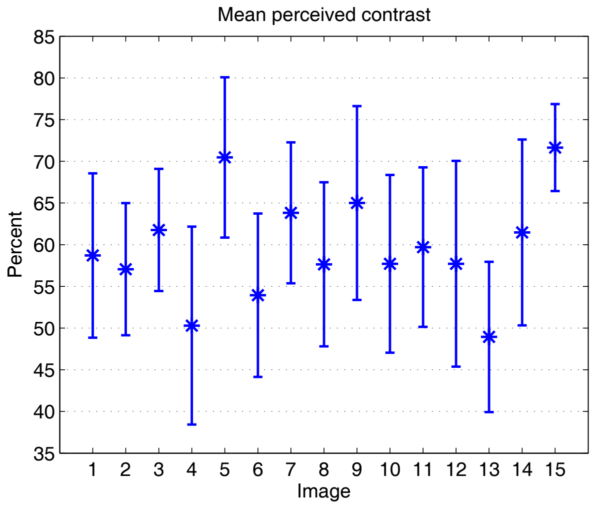
Table 1. Perceived contrast results for all observers. Image 15 has the highest mean value, while image 13 has the lowest mean value. Image 15 also has the lowest mean standard deviation, indicating that observer's concensus about the high contrast in this image.

Image	Mean value	Mean std
1	58,71	19,16
2	57,06	15,42
3	61,76	14,25
4	50,29	23,08
5	70,47	18,69
6	53,94	19,06
7	63,82	16,44
8	57,65	19,13
9	65,00	22,61
10	57,71	20,72
11	59,71	18,61
12	57,71	24,00
13	48,94	17,53
14	61,47	21,67
15	71,65	10,15

Image 12 has the highest standard deviation value, indicating the biggest difference between the answers from the observers. While for image 15 the observers agree more upon the rating.

We have also analyzed the results for expert and non-expert oberservers. Figure 3 and Table 2 show the mean contrast and the standard deviation for the experts. Image 13 has been rated as

Figure 2. *Mean perceived contrast score for all observers.*



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.

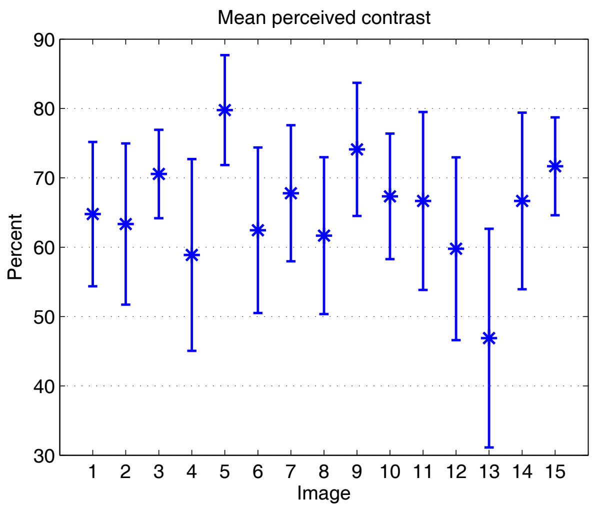


Figure 3. *Mean values for experts. Image 5 has the highest mean value, but cannot be differeniated from many of the other images. Image 13 has the lowest mean value, this image also has the highest mean standard deviation, indicating high deviation in the contrast score for this image.*

Table 3 and Figure 4 show the mean contrast value for each image and standard deviation for the non-experts. 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 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 20,46. Image 13 is the only image where the non-experts have a higher mean than the experts, but the difference is only 4,36.

The non-experts have used more of the scale than experts,

Table 2. Perceived contrast results for experts. Image 5 has the highest mean value, while image 13 has the lowest mean value. Image 3 has the lowest mean standard deviation, while image 13 has the highest.

Image	Mean value experts	Mean std
1	64,78	13,80
2	63,33	15,41
3	70,56	8,46
4	58,89	18,33
5	79,78	10,51
6	62,44	15,84
7	67,78	13,02
8	61,67	15,00
9	74,11	12,73
10	67,33	12,00
11	66,67	17,02
12	59,78	17,48
13	46,89	20,91
14	66,67	16,87
15	71,67	9,35

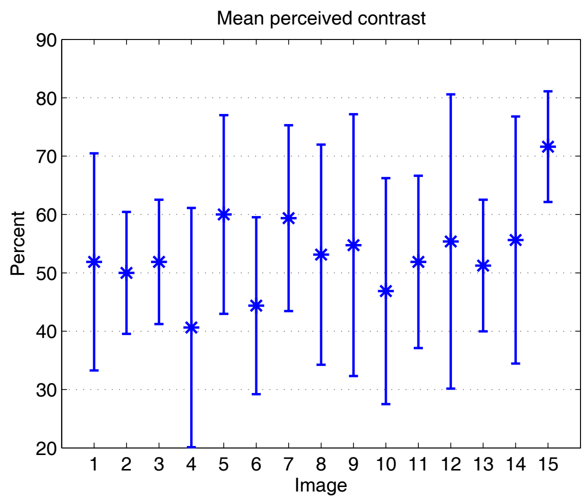


Figure 4. *Mean values for non-experts. Image 15 has the highest mean value, and also the lowest mean standard deviation. Here the non-expert agree upon the contrast score without strong deviations in the contrast value. Image 4 has the lowest mean value.*

Table 3. Perceived contrast results for non-experts. Image 15 has the highest mean value, and also the lowest mean standard deviation. Image 4 has the lowest mean value.

Image	Mean value non-experts	Mean std
1	51,88	22,83
2	50,00	12,82
3	51,88	13,08
4	40,63	25,13
5	60,00	20,87
6	44,38	18,60
7	59,38	19,54
8	53,13	23,14
9	54,75	27,50
10	46,88	23,75
11	51,88	18,11
12	55,38	30,91
13	51,25	13,82
14	55,63	25,97
15	71,63	11,64

i.e. they have a larger mean difference between the maximum value and minumum value.

We have also investigated the connection between mean perceived contrast and mean standard deviation. For all observers there is a correlation of 0.41, while for non-experts only 0.23. For the expert observers we have a correlation of 0.83, indicating a high concensus among the experts for the images with higher perceived contrast, and as the perceived contrast decreases the standard deviation increases (Figure 5).

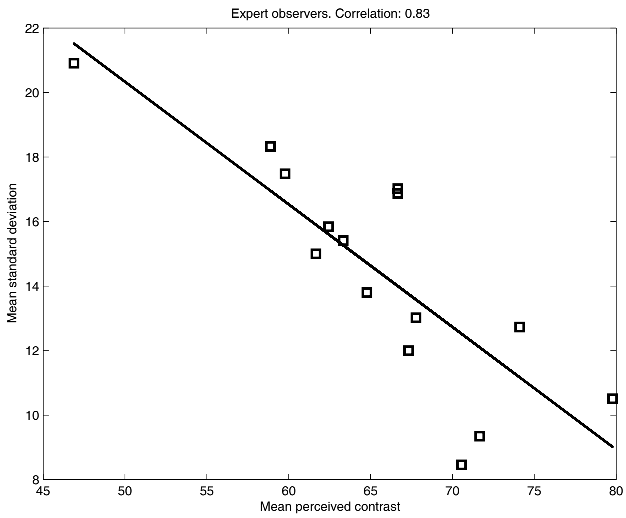


Figure 5. Correlation between mean perceived contrast and mean standard deviation for experts. A high correlation is found, indicating that when the mean perceived contrast decreases the mean standard deviation increases.

Contrast Algorithms

We have tested 5 different contrast algorithms, PELI [8], TT [9], RAMMG [10], RSC [11] and *Lab* variance.

For the PELI algorithm the images must have the power of 2, because of this the images have been resized to 512x512. The PELI algorithm rates the image with the highest and lowest 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 4).

The TT algorithm has a low correlation between the perceived contrast and predicted contrast (Table 4). 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.

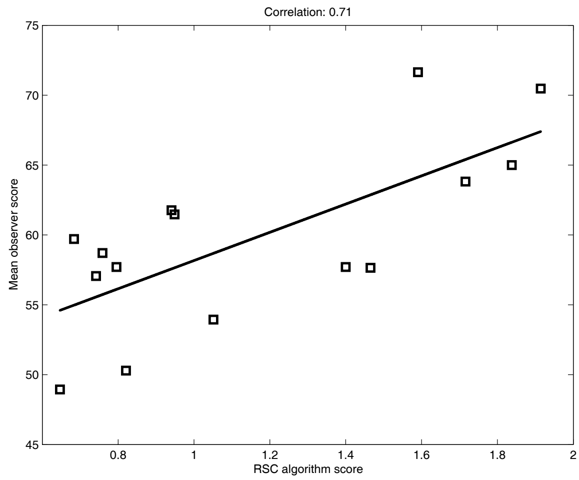


Figure 6. Correlation between observer mean score and RSC algorithm score. The results give a Pearson correlation of 0,71, and a Spearman correlation of 0,69. This algorithm has a higher Spearman correlation than the other measures when the contrast scores from all observers are taken into account.

For the RAMMG score compared against mean perceived contrast score the Pearson correlation is 0.57, indicating a relationship between the predicted contrast and perceived contrast. Even though the RAMMG rates images with a difference of over 20 on the observer scale almost similar, the Spearman correlation is 0.49.

The RSC algorithm has been weighting with different parameters, the L^*, a^* and b^* has been weighting according to the variance in each channel, for the center and surround the standard values of 1 and 2 have been used. The Pearson correlation between perceived and predicted contrast is 0.71 as seen on Figure 6, and the Spearman correlation is 0.64 indicating a relationship between the ranking of the two scores.

Lab variance is calculated as the geometrical mean of the variance in each channel in the CIELAB colorspace. The results are plotted against mean observer scores give a Pearson correlation of 0.81 as seen on Figure 7, and the highest Spearman correlation of 0.68 and also the lowest RMSE for all observers. Variance only in *L* channel has been tested, but with a lower performance.

Experts

The PELI algorithm 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 4). For the Spearman correlation the PELI algorithm shows an increased correlation, where the correlation is 0.50 (Table 4). This indicate that the PELI algorithm has a more correct ranking while the frequency distribution is not correct.

The TT algorithm compared against the expert observers has the same results as for all observers. Images rated in the similar by the observers receive different TT scores, resulting in a Pearson and Spearman correlation close to 0 and a high RMSE.

The RAMMG has a Pearson correlation of 0.54 between the

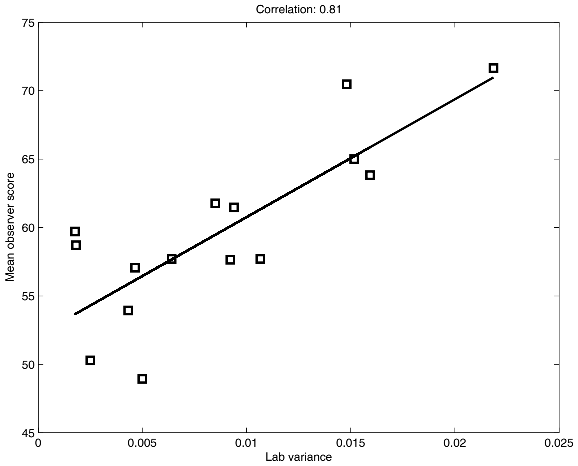


Figure 7. Correlation between observer mean score and Lab variance.

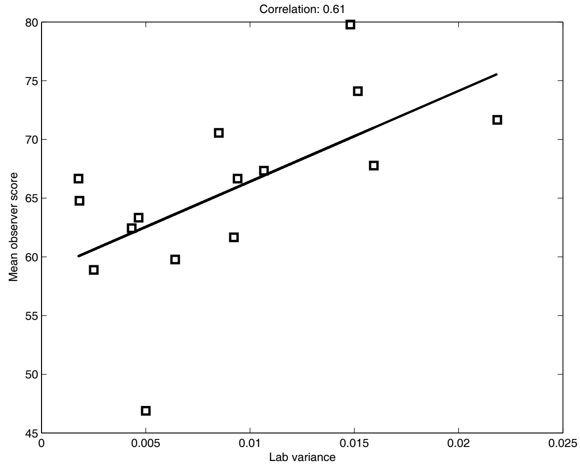


Figure 9. Correlation between observer mean score and Lab variance for expert observers.

RAMMG score and expert observers mean contrast score, almost the same as for all observers. Image 13 is the image furthest away from the linear regression line, indicating that it should have been given an even lower RAMMG score.

The RSC algorithm has a Pearson correlation of 0.69 as seen on Figure 8. The image furthest from the regression line is the image rated with the lowest contrast by the obervers, even though this has been ranked as the image with the lowest contrast by RSC. By looking at the Spearman correlation we can see that the RSC has the highest correlation with 0.69, this measure therefore has a more correct ranking than the other metrics.

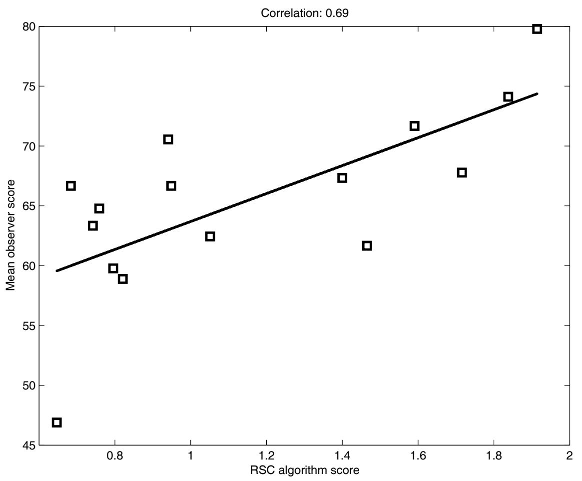


Figure 8. Correlation between observer mean score and RSC algorithm score for expert observers.

Lab variance plotted against mean observers scores for experts gives a Pearson correlation of 0.61 as seen on Figure 9. The image rated with the lowest contrast by the observers is located furthest away from the regression line. This image is miscalculated by all algorithms. The Spearman correlation is a bit higher than the Pearson correlation, indicating a better ranking performance.

Non-experts

For the non-experts the PELI algorithm has a scattering of the data points, where the image with the highest perceived contrast is rated as second to worst by the PELI algorithm. This results in a negative Pearson correlation. The TT algorithm for the non-experts has the same low performance as for all and expert observers.

The Pearson correlation between RAMMG and non-expert observers mean contrast score is 0.43, indicating some ability to predict perceived contrast. The image rated third highest by RAMMG is rated as the image with the lowest contrast by the non-experts. The Spearman correlation indicate a low performance in ranking. The RSC algorithm has a Pearson correlation of 0.52 (Table 4) for the non-experts, lower than for both the expert and for all observers. This is also the situation for algorithms except the *Lab* variance.

Lab variance plotted against mean observers scores for non-experts gives a Pearson correlation of 0.79, with a similar distribution as seen for all observers in Figure 7. Most of the data points are located around the regression line, and the correct prediction of the image with the highest perceived contrast. The Spearman correlation is 0.70 and we also have the lowest RMSE of the tested measures.

Overall observervations

From Table 4 *Lab* variance shows the highest correlation for all observers and non-experts, while RSC has the highest correlation for the experts. The RAMMG algortihm shows a medium correlation for all groups, but not as high as the *Lab* variance. With the high correlation for *Lab* variance we can say that contrast is connected with the luminance and chroma variance in the image. For the correlation calculated with the Spearman approach the *Lab* variance shows the highest correlation for all observers and non-experts, while RSC has the highest for experts. The same for RMSE where the *Lab* variance has the lowest RMSE for all observers and non-experts, while RSC for the experts.

Conclusions and perspectives

The perceived contrast in an image is different when it comes to the background of the observer. Experts agree more upon the contrast in the image than non-experts. Experts also rate the contrast to be higher then non-experts in most scenes. As the perceived contrast for experts decreases also does the concensus.

For the contrast measures the *Lab* variance shows the best correlation between predicted contrast and perceived contrast, indicating that variance in luminance and chroma can be connected to perceived contrast. For expert observers the RSC algortihms has the highest performance. The results here indicate that the

Table 4. Pearson correlation, Spearman correlation and RMSE for the different contrast algorithms for all observers, experts and non-experts. Gray cell indicate best performance.

Algorithm	Pearson correlation			Spearman correlation			RMSE		
	All	Experts	Non-experts	All	Experts	Non-experts	All	Experts	Non-experts
PELI	0.32	-0.03	-0.16	0.43	0.5	0.23	6.28	7.54	7.26
TT	-0.09	0.02	-0.19	-0.19	-0.01	-0.37	6.59	7.87	7.37
RAMMG	0.57	0.54	0.43	0.49	0.54	0.31	5.45	6.63	6.79
RSC	0.71	0.69	0.52	0.64	0.69	0.54	4.63	5.66	6.41
Lab Variance	0.81	0.61	0.79	0.68	0.66	0.70	3.92	6.25	4.58

work done in RSC is improving the state of the art contrast measures, but also that the *Lab* variance is an important element that should be taken into account in contrast measures.

Another way to improve contrast measures could be by using gaze information, saliency maps or region-of-interest algorithms.

References

[1] X. Zhang and B. A. Wandell. Color image fidelity metrics evaluated using image distortion maps. *Signal Processing - Special issue on image and video quality metrics*, 70:201 – 214, 1998.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.

[3] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. In *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers*, Nov 2003.

[4] N. Bonnier, F. Schmitt, H. Brettel, and S. Berche. Evaluation of spatial gamut mapping algorithms. In *Proceedings of IS&T and SID's 14th Color Imaging Conference*, volume 14, pages 56–61, Nov 2006.

[5] X. Gao, T. Wang, and J. Li. A content-based image quality metric. *Lecture notes in computer science*, 3642:231–240, 2005.

[6] Christopher C. Taylor, Jan P. Allebach, and Zygmunt Pizlo. The image fidelity assessor. In *IS&Ts 1998 PICS Conference*, Portland, Oregon, USA, 1998. IS&T.

[7] J. J. McCann. A comparison of color metrics. In *4th IS&T/SID Color Imaging Conference*, 1995.

[8] E. Peli. Contrast in complex images. *Journal of the Optical Society of America*, 7:2032–2040, 1990.

[9] Y. Tadmor and D.J. Tolhurst. Calculating the contrasts that retinal ganglion cells and lgn neurones encounter in natural scenes. *Vision Research*, 40:3145–3157, 2000.

[10] A. Rizzi, T. Algeri, G. Medeghini, and D. Marini. A proposal for contrast measure in digital images. In *CGIV 2004 – Second European Conference on Color in Graphics, Imaging and Vision*, 2004.

[11] A. Rizzi, G. Simone, and R. Cordone. A modified algorithm for perceived contrast measure in digital images. In *CGIV*, 2008.

[12] A. J. Calabria and M. D. Fairchild. Perceived image contrast and observer preference: I. the effects of lightness, chroma, and sharpness manipulations on contrast perception. *The Journal of imaging science and technology*, 47:479–493, 2003.

[13] D. R. Ankrum. Viewing angle and distance in computer workstations. 13/11/06: www.combo.com/ergo/vangle.htm.

Author Biography

__Marius Pedersen__ received his BsC in Computer Engineering in 2006, and MiT in Media Technology in 2007, both from Gjøvik University College, Norway. He is currently pursuing a PhD in Color Imaging. He is also a member of the Norwegian Color Research Laboratory at Gjøvik University College. His work is centered on image quality metrics for color prints.

__Alessandro Rizzi__ received his degree in Computer Science at University of Milano and PhD in Information Engineering at University of Brescia. Now he is an assistant professor, and senior research fellow at the Department of Information Technologies at University of Milano. Since 1990 he is researching in the field of digital imaging and vision. His main research topic is color information with particular attention to color perception mechanisms. He is the coordinator of the Italian Color Group.

__Jon Yngve Hardeberg__ is a Professor of Color Imaging at Gjøvik University College. He received his Ph.D from Ecole Nationale Suprieure des Télécommunications in Paris, France in 1999,with a dissertation on colour image acquisition and reproduction, using both colorimetric and multispectral approaches. He has more than 10 years experience with industrial and academic colour imaging research and development, and has co-authored over 100 research papers within the field. His research interests include various topics of colour imaging science and technology, such as device characterisation, gamut visualisation and mapping, image quality, and multispectral image acquisition and reproduction. He is a member of IS&T, SPIE, and the Norwegian representative to CIE Division 8. He has been with Gjøvik University College since 2001 and is currently head of the Norwegian Color Research Laboratory.

__Gabriele Simone__ received his BiT in 2005, and his MSIT in 2007 both at University of Milan - Department of Information Technology, Italy. Hes currently pursuing a PhD at Gjøvik University College, Norway. His main research topic is contrast measure and tone mapping algorithm in HDR images.