CID:IQ – A New Image Quality Database

Xinwei Liu, Marius Pedersen, and Jon Yngve Hardeberg

The Norwegian Colour and Visual Computing Laboratory, Gjøvik University College,
P.O. Box 191, 2802 Gjøvik, Norway
http://www.colourlab.no

Abstract. A large number of Image Quality (IQ) metrics have been developed over the last decades and the number continues to grow. For development and evaluation of such metrics, IQ databases with reference images, distortions, and perceptual quality data, is very useful. However, existing IQ databases have some drawbacks, making them incapable of evaluating properly all aspects of IQ metrics. The lack of reference image design principles; limited distortion aspects; and uncontrolled viewing conditions. Furthermore, same sets of images are always used for evaluating IQ metrics, so more images are needed. These are some of the reasons why a newly developed IQ database is desired. In this study we propose a new IQ database, Colourlab Image Database: Image Quality (CID:IQ), for which we have proposed methods to design reference images, and different types of distortions have been applied. Another new feature with our database is that we have conducted the perceptual experiments at two viewing distances. The CID:IQ database is available at http://www.colourlab.no/cid.

Keywords: Image Quality Metric, Noise, Blur, Image Compression, Gamut Mapping, Viewing Distance, Perceptual Experiment.

1 Introduction

The quality evaluation of digital images is an important part in many image processing applications. In order to enhance Image Quality (IQ) while to reduce distortions, it is important to have an indicator which represents IQ. For this purpose, IQ metrics are commonly used to assess the quality of images. It is important to have a ground truth for the assessment and benchmarking of IQ metrics. Because of this, IQ databases are developed.

There are many existing IQ databases: e.g. TID2008 [1], LIVE [2], Toyama (MICT) [3] and so on. TID2008 [1] contains the largest number of observers, while Toyama (MICT) [3] only focuses on JPEG and JPEG2000 compression distortions. All databases have their own special purposes, but most of them have three main shortages. First, reference images are the core part of an IQ database, so they have to be selected very cautiously. Some of the databases, such as TID2008, LIVE, Toyama, IVC [4] etc. used images which were scanned from negatives as their reference. The quality of these images might be limited and incomparable to current digital images. Although some of databases used

A. Elmoataz et al. (Eds.): ICISP 2014, LNCS 8509, pp. 193–202, 2014.

high quality digital images, they have a lack of analysis for the reference images to demonstrate whether they cover different ranges of characteristics, such as spatial information or colorfulness. In the CSIQ database [5], spatial information and colorfulness are in a limited range. Second, the types of distortion used in most of the databases are similar, such as JPEG compression artifacts, Gaussian noise, and blurring are used in TID2008, LIVE, VCL@FER [6], and CSIQ databases. Nevertheless, these distortions are the most common distortions, but none of the databases focus on the color distortions (e.g. color shifts, gamut mapping artifacts). Third, viewing conditions are an important aspect in IQ experiments. Particularly the choice of surround illumination and the viewing distance. It is a challenge for IQ metrics to evaluate IQ by considering different viewing distances. The viewing conditions are rarely strictly controlled or described in detail in the existing databases. For example, in TID2008 and TID2013 some evaluations were carried out over the internet, without the possibility to control the viewing conditions. Ambient illumination is not mentioned in IVC and A57 databases, which could influence the quality. Furthermore, same sets of images are always used for evaluating IQ metrics, this is also a disadvantage. More images are needed for the evaluation of IQ metrics. In conclusion, the development of a new IQ database is required.

The current work aimed at developing a new IQ database. The Colourlab Image Database: Image Quality (CID:IQ) database has the following new characteristics: (1) Reference images are selected by proposed principles and analyzed by proposed approaches. (2) Distortions cover both normal distortion types and color related distortion types. (3) Subjective experiments were strictly controlled by taking into account the viewing conditions based on recommended standards. The experiments were also conducted at different viewing distances.

The paper is organized as follows. We state the aspects of reference images design in Section 2. In Section 3 we present the image distortion types, and the perceptual experiment is presented in Section 4. Further, the experimental results and analysis are given before the conclusion. Finally, conclusions and ideas for future work are presented.

2 Reference Image Design

It is obvious that reference images are one of the core aspects of an IQ database. So the selection or design of reference images is a significant issue needed to be taken into account. In this section we present aspects around the design of reference images as the number of images, types of image, and methods for analyzing reference images.

Typically there are two kinds of reference images in the field of IQ assessment: pictorial images and research images [7]. Pictorial images are usually the best choice for many evaluation experiments, since observers are more used to pictorial images compared to generated research images. However, the choice of pictorial images has to be very careful because the experimental result is highly depends on the test images. The main disadvantage of pictorial images is that they make measurements difficult to be consistently quantitative [7].

Keelan *et al.* [8] proposed that at least three images have to be used in subjective experiments so that the relative quality values of just noticeable difference can be included. ISO 20462-1 [9] states that three, or more than three images, should be used in subjective experiments, and that it is recommended with six or more images. Field [7] recommends that the number of test images should be between five to ten so that the full range of color and IQ factors can be assessed.

Analyzing reference images can help us to decide what kind of images that are suited for our database. Furthermore, the analysis provide information to other researchers, and help them to choose the most suitable database for their particular benchmarking or some additional requirements. Winkler [10] proposed to analyze two parameters: Spatial information (SI) and Colorfulness (CF) in order to characterize the reference images. Spatial information (SI) represents edge energy. If we define si_h as images where a horizontal Sobel kernel filter is applied, and si_v as images where a vertical Sobel kernel filter is applied. Then $em = \sqrt{si_h^2 + si_v^2}$ represents the edge magnitude at each pixel. Therefore, SI is the root mean square of the edge magnitude in image:

$$SI = \sqrt{vr/1080}\sqrt{\sum em^2/p}, \quad (1)$$

where p is the image pixel number and $\sqrt{vr/1080}$ (vr is vertical resolution of the image) is a normalization factor. Since SI is calculated in grayscale, the SI for RGB images (SI_c) are converted to grayscale with the following equation:

$$SI_c = 0.229R + 0.587G + 0.114B. \quad (2)$$

CF represents the intensity and assortment of colors in the image. First an opponent color space is defined, where $c1 = R - G$ and $c2 = 0.5(R + G) - B$. CF is then:

$$CF = \sqrt{\alpha_{c1}^2 + \alpha_{c2}^2} + 0.3\sqrt{\beta_{c1}^2 + \beta_{c2}^2}, \quad (3)$$

where α_{c1} is the trigonometric length of the standard deviation in $c1$ space and β_{c1} is the distance of the centre of gravity in ab space to the neutral axis. Orfanidou *et al.* [11] proposed to analyze the 'busyness' of the scene. Busyness is defined as the image property indicating the presence or absence of details in an image. Because the level of details is a very important characteristic of an image, we need to have reference images with a wide range of busyness values.

We concluded six categories with 27 attributes from [7] in order to check if our reference images contain all necessary content. These attributes are: hue, saturation, lightness, contrast, memory colors and others. In the hue category there are even attributes: red, green, blue, cyan, magenta, yellow, black and no specific hue. For saturation, lightness and contrast, three levels of intensity are stated: low, medium and high. Skin color, sky-blue and grass-green are three attributes in memory colors and skin colors include black, caucasian and asian. Some of the attributes cannot fall into any of these normal categories we defined them as attributes for other purposes: large area of the same color; neutral gray; color

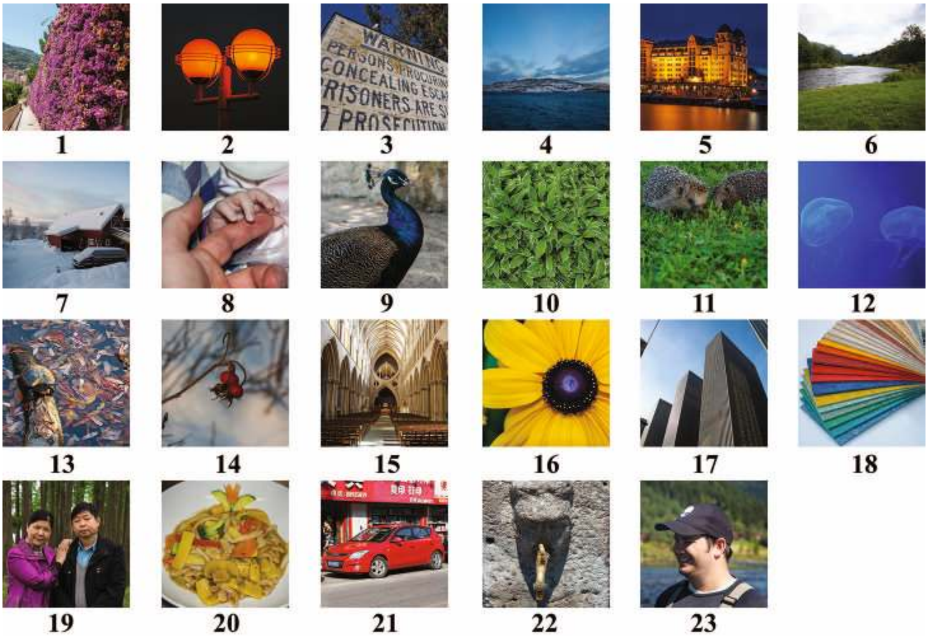
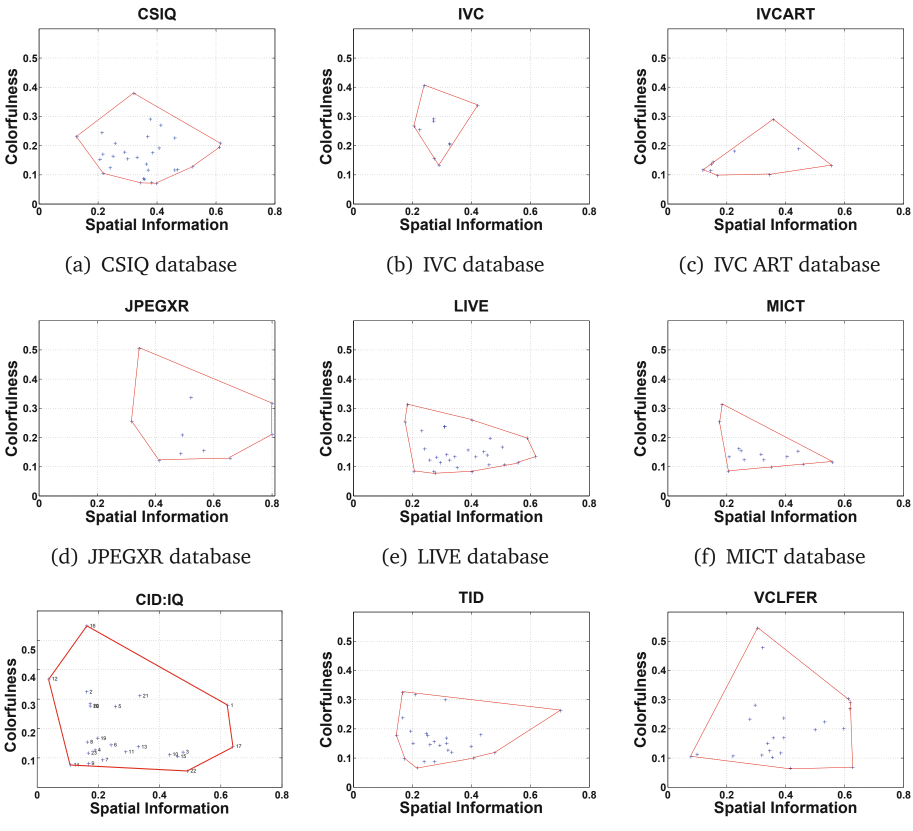


Fig. 1. Reference images in the new IQ database

transition; fine detail; and text. We started with a large set of reference images, and then reduced the set to match the criteria set forward. The results show that only 'black skin tone' is missing in CID:IQ database. At last, 23 pictorial images are selected as the reference images in the CID:IQ database (Figure 1). The resolution of all images is 800 pixels by 800 pixels. The reason of selecting this image resolution is that a resolution of 1920×1080 is a standard resolution, and in order to be able to display two images on the screen simultaneously, the maximum width of the image should be less than 960 pixels. By taking into account the area immediately surrounding the displayed image and its border,we set the resolution of reference image as 800 pixels by 800 pixels.

As introduced, we use SI, CF, and 'busyness' to evaluate the reference images in the CID:IQ database and to compare the results with the other existing databases. The SI versus CF and the 'busyness' of the CID:IQ database are compared to eight other most commonly used databases (CSIQ, IVC, IVC Art, JPEGXR [12], LIVE, Toyama, TID2013, and VCL@FER). The results by comparing SI and CF values from CID:IQ database and the other databases are plotted in Figure 2 and the comparison of 'busyness' values between the databases is shown in Figure 3. In Figure 2 the X axis represents the SI and the Y axis represents the CF. The X axis in Figure 3 shows the 'busyness' values in different range and the value on Y axis is the quantity of the images. From Figure 2 it can be seen that the CID:IQ database is covering a wider area than the others. The CID:IQ database has images with high/low SI versus low/high CF values. The images are scattered both near the edge and in the center of the plot. This means the reference images in CID:IQ database better represent real world scenes. As can be seen from Figure 3, the distribution of the busyness values in CID:IQ covers a wider range than the others. The CID:IQ database has therefore images with varying levels of details. In conclusion, CID:IQ covers a wider range of characteristics compared to others.



(a) CSIQ database (b) IVC database (c) IVC ART database

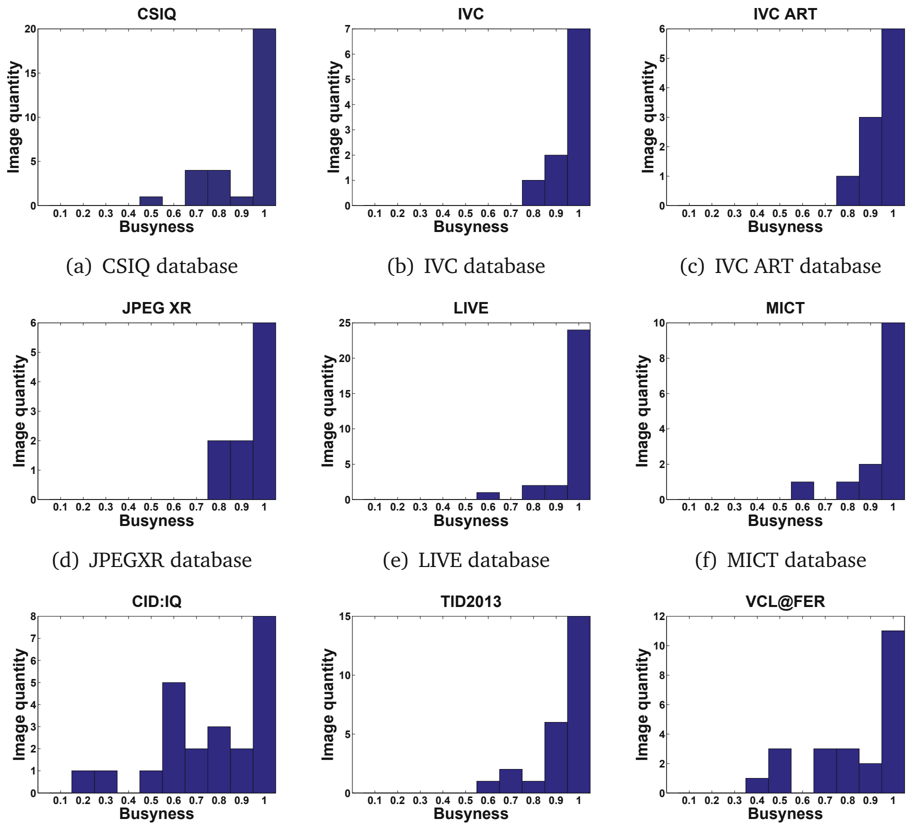
(d) JPEGXR database (e) LIVE database (f) MICT database

(g) CID:IQ database (h) TID2013 database (i) VCL@FER database

Fig. 2. Comparison of SI vs. CF results between image quality databases

3 Image Distortions

Four categories within six different distortions are used in the new IQ database. The categories are: compression artifacts, noise, blurring and gamut mapping artifacts. In the compression artifacts category, both JPEG and JPEG2000 compression standards are selected; we used Poisson noise in noise category because the photon Poisson noise is the dominant contributor to uncertainty in the raw data captured by high-performance sensors using in color digital cameras. Gaussian blur is selected as a type of blurring. In gamut mapping artifacts category, CIE [13] proposed that two gamut mapping algorithms are obligatory need to be included when evaluating gamut mapping algorithms: constant hue minimum ΔE and SGCK gamut mapping. We selected these two gamut mapping methods based on CIE's [13] recommendation.



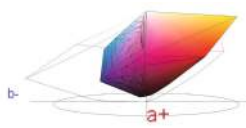
(a) CSIQ database (b) IVC database (c) IVC ART database

(d) JPEGXR database (e) LIVE database (f) MICT database

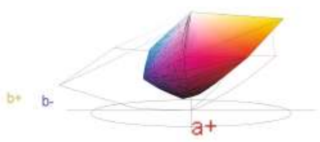
(g) CID:IQ database (h) TID2013 database (i) VCL@FER database

Fig. 3. Comparison of Busyness results between image quality databases

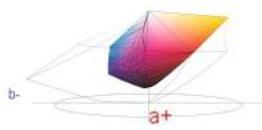
All reference images are applied these distortions in five levels from low degree of quality to high degree of quality degradation. For JPEG and JPEG2000 compression, we used different compression rates (0.7,0.5,0.3,0.2,0.1 for JPEG and 0.9,0.6,0.4,0.3,0.2 for JPEG2000) to conduct the five degradation levels. For Poisson noise and Gaussian blur we used separate magnitude (0.5,1,1.5,2,2.5 for noise and 0.5,0.7,0.9,1.1,1.3 for blur) to generate different levels in Matlab. We selected five ICC profiles with different volumes to represent the five levels: PSO Coated v2 300 Glossy laminate profile (volume=552537), PSO LWC (light weight coating) Standard profile (volume=457606), PSO MFC (machine finished coating) Paper profile (volume=359510), ISO uncoated yellowish profile (volume=204334), and ISO newspaper 26v4 profile (volume=141632). These gamuts represent different paper types which influence the perceived quality. A preview of these gamuts is given in Figure 4. ICC3D [14] software is used to generate gamut mapped images. The design of the distorted sequences is based on the levels in existing databases. However, the levels in existing databases (TID2013, LIVE etc.) are large, making it is quite easy for human observers to distinguish the difference between each level. In CID:IQ the difference between each degradation level is smaller, and should therefore also be more challenging for IQ metrics.



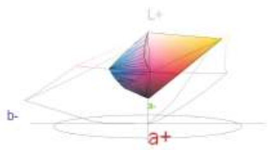
(a) PSO Coated v2 300 Glossy laminate



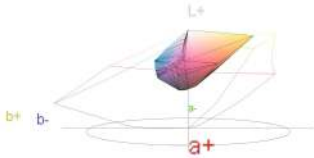
(b) PSO LWC Standard



(c) PSO MFC Paper



(d) ISO uncoated yellowish



(e) ISO newspaper 26v4

Fig. 4. Comparison between five ICC profile gamuts

4 Perceptual Experiment

The entire experiment took place in a laboratory. Our experiment viewing illumination setup followed the recommendations from CIE [13] and ITU [15]. The level of ambient illumination is approximately 4 lux. The chromaticity of the white displayed on the color monitor is D65 and luminance level of the monitor is 80 cd/m^2. All settings are suited for sRGB color space.

It is a challenge for IQ metrics to evaluate IQ at different viewing distances. So we decided to use a normal viewing distance and a longer distance to conduct our experiment. ITU [15] recommended the maximum observation angle relative to the normal is 30 degree. If we transfer it to viewing distance in our case, the shortest distance should be 38cm. As a result we use 50cm (viewing angle of 23 degrees) and 100cm (viewing angle of 12 degrees) as our viewing distances.

17 human observers participated the psychometric experiment. Category judgment is used as the scaling approach. Instead of using five categories as recommended from CIE [13] and ITU [15], we keep the original scale but extended it

to nine categories by simply add one extra category between the categories from the recommendation to allowing the observer to pick a mid-point. Because some observers are not inclined to select the extreme categories and there are five levels of degradation so by this reason it is better to have more than 5 categories. The 9 categories scale is: 1: Bad quality; 2; 3: Poor quality; 4; 5: Fair quality; 6; 7: Good quality; 8; 9: Excellent quality. Written experiment instruction is given to all observers, and a training sequence is included at the beginning of the first session in order to stabilize the observers' opinion.

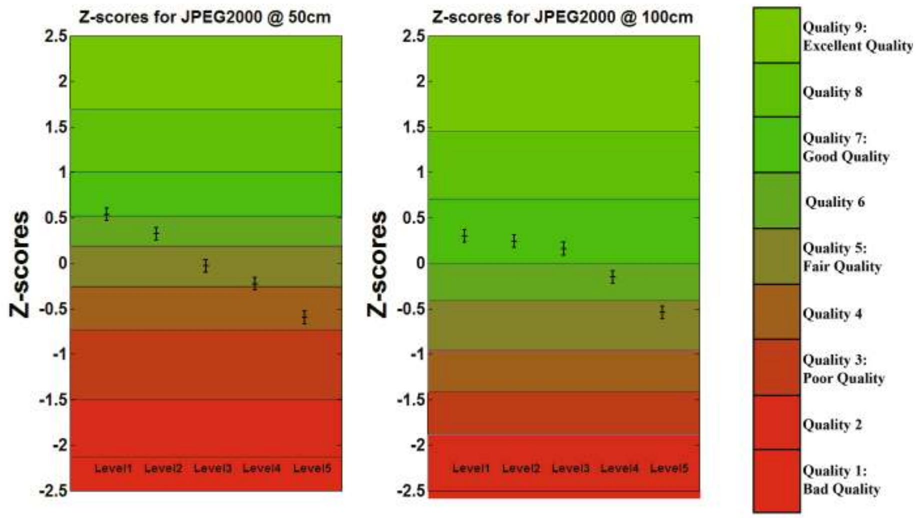


Fig. 5. Z-scores for JPEG2000 distortion. Data points on which the underlaying data is based on 17 observers.

5 Experimental Results and Analysis

We use Z-scores as given by Engeldrum [16] with a 95% confidence interval to present the experiment results. We will use the JPEG2000 compression distortion as an example for the data analysis. Additional details can be found in Liu [17]. Subjective results for JPEG2000 compression distortion are shown in Figure 5. The plot on the left is the Z-scores from the experiment conducted at 50cm viewing distance and on the right the Z-scores for 100cm. From the results we can see that when the observers were at 50cm, they can easily identify the difference between different degradation levels because there is no overlap in any of the confidence interval for the five levels. The first level has been assigned to category seven which means the IQ is good while the fifth level is located at category 4 means the IQ is better than poor but worse then fair. The plot on the right shows that it is difficult for the observers to see differences for the first three levels because the scale values are very similar, which is indicated by overlapping confidence intervals. This is different from the results at 50cm. In addition, the fifth level in the right plot is in category 5 which is better than the same level on the left plot (located at category 4). This is because the quality of compressed images are highly depended on the viewing distance, when the viewing distance increases the distortion is less perceptible. However, it is difficult for IQ metrics to consider human visual system in order to give the same results. So the subjective data in our new database have a significant advantage compared to other databases. Most likely IQ metrics that take into account human visual system and can simulate viewing distance will perform better on this new database. Additionally, gamut mapping provides changes in many quality attributes, and therefore they are difficult for metrics to evaluate [18].

6 Conclusions and Future Work

Through this study, a new IQ database has been successfully developed. Three key features are proposed in the CID:IQ database: first, it integrates state-of-the-art reference image design methods and evaluation approaches; second, new color related distortions are used; third, two viewing distances in the experiments and the experiment was conducted with controlled viewing conditions. The CID:IQ database is available at `http://www.colourlab.no/cid` for free downloading.

In further work, more types of distortions could be applied to the reference images. A concluding step of this work would be the methodology of creating a new IQ database provides the possibility and knowledge to develop other types of IQ database. The database should be used for testing IQ metrics, and that differences in metric performance between different databases would be interesting to analyze.

References

1. Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., Battisti, F.: Tampere image database 2008, TID 2008 (2008), `http://www.ponomarenko.info/tid2008.htm`
2. Sheikh, H., Wang, Z., Cormack, L., Bovik, A.: LIVE image quality assessment database release 2 (2006), `http://live.ece.utexas.edu/research/quality`
3. Tourancheau, S., Autrusseau, F., Sazzad, Z.M.P., Horitaa, Y.: MICT image quality evaluation database (2008), `http://mict.eng.u-toyama.ac.jp/mictdb.html`
4. Le Callet, P., Autrusseau, F.: Subjective quality assessment Irccyn/IVC database (2005), `http://www2.irccyn.ec-nantes.fr/ivcdb/`
5. Larson, E.C., Chandler, D.: Categorical subjective image quality CSIQ database (2010), `http://vision.okstate.edu/?loc=csiq`
6. Zaric, A., Tatalovic, N., Brajkovic, N., Hlevnjak, H., Loncaric, M., Dumic, E., Grgic, S.: VCLFER image quality assessment database (2011), `http://www.vcl.fer.hr/quality/`
7. Field, G.G.: Test image design guidelines for color quality evaluation. In: Color and Imaging Conference, Scottsdale, Arizona, USA, pp. 194–196 (November 1999)

8. Keelan, B.W., Urabe, H.: ISO 20462: A psychophysical image quality measurement standard. In: SPIE Proceedings, vol. 5294, pp. 181–189. Image Quality and System Performance (2003)
9. ISO. 2004. ISO 20462-1 photography - psychophysical experimental methods to estimate image quality - part 1: Overview of psychophysical elements
10. Winkler, S.: Analysis of public image and video databases for quality assessment. IEEE Journal of Selected Topics in Signal Processing 6(6), 616–625 (2012)
11. Orfanidou, M., Triantaphillidou, S., Allen, E.: Predicting image quality using a modular image difference model. In: SPIE Proceedings 2008, Image Quality and System Performance (January 2008)
12. De Simone, F., Goldmann, L., Baroncini, V., Ebrahimi, T.: JPEG core experiment for the evaluation of JPEG XR image (2009), http://mmspg.epfl.ch/iqa
13. CIE. 2004 Guidelines for the evaluation of gamut mapping algorithms. Technical Report. ISBN: 3-901-906-26-6, CIE TC8-03, 156
14. Farup, I., Hardeberg, J.Y., Bakke, A.M., Kopperud, S., Rindal, A.: Visualization and interactive manipulation of color gamuts. In: Color Imaging Conference, Scottsdale, Arizona, USA, pp. 250–255 (November 2002)
15. ITU. Recommendation BT.500 : Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, Geneva, Switzerland, 53-56 (2002)
16. Engeldrum, P.G.: Psychometric scaling: A toolkit for imaging systems development. Imcotek Press (2000)
17. Liu, X.: CID:IQ - a new image quality database. Master thesis, Gjøvik University College (2013)
18. Hardeberg, J.Y., Bando, E., Pedersen, M.: Evaluating colour image difference metrics for gamut-mapped images. Coloration Technology 124(4), 243–253 (2008)