Analyzing the Variability of Subjective Image Quality Ratings for Different Distortions

Olga Cherepkova, Seyed Ali Amirshahi, Marius Pedersen
Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway
{olga.cherepkova,s.ali.amirshahi,marius.pedersen}@ntnu.no

***Abstract*—When it comes to evaluating the quality of images, individual observers show different opinions depending on the type of distortion affecting the quality of the image. While in the opinion of one observer a distortion could have a dramatic influence on the quality of the image, another observer could see the same distortion as not having an important effect on the quality of the same image. Using a subjective experiment, we aim to identify the distortions which show the largest variability among observers. For this, 22 observers evaluated the quality of 10 reference images and the 630 test images created from them (21 distortions at three levels). Our results show that the highest variability in subjective scores is linked to distortions like saturation, contrast, sharpness, quantization, some types of added noise, and radial lens distortion.**

***Index Terms*—image quality, subjective evaluation, image distortion, individual differences.**

I. Introduction

With the increase in the creation and use of digital images in our daily life, evaluating and enhancing their quality has turned into a critical matter. In general, evaluating the quality of images is a subjective matter, which ideally should be performed by observers judging the quality of an image based on their preferences, their visual acuity, and task context [1].

In general, subjective evaluation done by observers is the most accurate way to estimate image quality. This is usually done through different subjective experiments in which the Mean Opinion Score (MOS), which is calculated by averaging the subjective scores given by all observers to an image, represents the overall image quality. Since it is not always possible to conduct such experiments and to automatize the evaluation process, objective metrics which are commonly referred to Image Quality Metrics (IQMs) have been developed. Most, if not all, IQMs provide a single value, which represents the quality of the image. While IQMs show a high correlation with MOS [2], there still exists a high variability between the scores given by different observers to the same image [3].This issue demonstrates the need of an IQM, which is able to work for each individual observer. In spite of such a clear need, due to its complexity, a personalized IQM has rarely been investigated in the research community [4].

In this study we aim to find distortions, which show the largest variability in the subjective scores given by different observers. Finding such distortions is the first step in collecting a large-scale subjective dataset, which would be used as the ground truth data for training and testing a personalized IQM.

This work is structured as following. First, in Section II we introduce the relevant background, before presenting the experimental design in Section III. The experimental results are presented in Section IV, before concluding and proposing future work in Section V.

978-1-7281-8750-1/20/$31.00 ©2022 IEEE

II. Background

While most if not all IQMs aim to predict the quality score for an average observer [5]–[8], metrics which try to evaluate the aesthetic quality of images have paid more attention to an individual approach. Such metrics take advantage of different approaches such as psychological personality tests [9], convolutional neural network algorithms designed for personal preferences prediction [10], and utilizing datasets for a personal assessment learning model [11].

As highlighted in [6], [12] an image with an average MOS usually shows higher variations in the subjective scores given to it. This normally occurs due to the averaging of opposite individual subjective scores, which then results in an average score, while low or high values mostly indicate that people agree with each other. Ghadiyaram and Bovik [6] also found that observers show a high variability when evaluating the quality of images, which have been distorted by multiple distortions. This could be seen as evidence that observers are sensitive to different kinds of distortions and judge image quality from various perspectives. Virtanen et al. in CID2013 [12] provide standard deviation values for different image sets, which shows the influence of image content on the variability of ratings. They further investigate the role of image content on individual quality scores by analyzing the correlation between IQMs and subjective data for different image scenes. In their experiments, one of the lowest correlation values is seen in the case of an image with the most artificial scene, which was possibly harder for observers to judge. This further emphasizes the importance of image content on quality assessment. During the preparation of the MDID database [13], Sun et al. found that the standard deviation of MOS in some existing databases are much larger than in others. This further leads to the point that image content in combination with different ways to distort images influence peoples' preferences differently. The importance of image content was also pointed out by [14], highlighting that predicting MOS values for some images is much easier than others. For the JPEG2000 distortion, they mention that images with many textures and contours were evaluated to have higher quality than images with larger uniform areas. Jayaraman et al. in [5] demonstrated the masking effect of one distortion applied to an image on top of another. They conclude that a high level of one distortion would result in the other distortion (even if its intensity increases) to have a lower influence on the overall judgment of the observers and so the MOS. This means that people are more sensitive to some type of distortions, and some image artifacts may become crucial in their judgment.

Speaking in more detail about the variability in distortions, Ponomarenko et al. [7] analyzed the variability of MOS values for each distortion and level during the creation of the TID2013 database. They used mean RMSE to characterize variations in judgment and found that the smallest variations belonged to JPEG2000 and spatially correlated noise, while the highest variability in subjective scores among observers was seen in the case of contrast changes, especially a large contrast increase. The study finds that people tend to agree more in the case of images with a higher level of distortion.

Using subjective experiments, Leisti et al. [15] found hierarchical relations between high level attributes, such as realism, naturalness, clarity, and low level attributes such as brightness, sharpness, and gloss, which observers used to describe the high level attributes. They discovered that asking reasons for observers' judgments does not have a significant influence on the results. Leisti and Jukka [16] continued the work and studied the relationship between learning and reasoning and the process of decision making. They found that learning plays an important role in preference stability, it especially helps observers to stay consistent in weighting information, while starting to ignore less visible artifacts. Reasoning in turn helps to sustain and even improve stability results, while helping observers to keep in mind a wider variety of factors.

Extensive research has been done investigating the perception of discrete image attributes such as contrast. Calabria and Fairchild [17] proved how one image attribute is dependent on other attributes, finding that the perceived contrast is scalable and can be modeled with a function of image lightness, sharpness, and chroma information. Pedersen et al. [18] in turn, have shown the possibility of perceived contrast dependency on luminance and chroma variance. They also found a big difference between expert and non-expert observers.

III. Experiment design

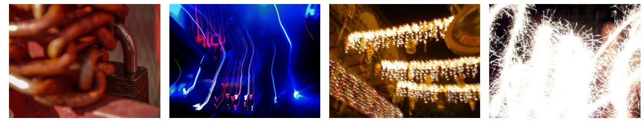
In this Section we discuss in detail the preparation stages of the subjective experiment, including data preparation, hardware, and software setup.

A. Dataset preparation

To understand which images tend to have larger variations between observers, we analyzed existing databases and found images with the largest standard deviation in the subjective ratings provided by the authors. More specifically, we analysed different distortions and images from Kadid10K [19], Koniq10K [20], TID2013 [7] and LIVE in the Wild [6]

(a) (b) (c) (d)

Fig. 1: Image samples of highest std from Kadid10K [19]. Quantization (a) and (b), blocking artifacts (c) and (d).

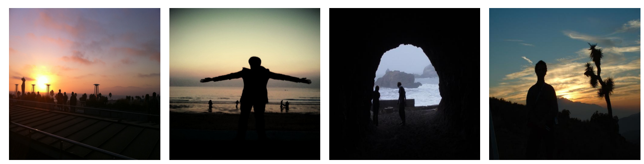


(a) (b) (c) (d)

Fig. 2: Image samples of highest std from Koniq10K [20]. Low aesthetic quality (a), artistic effects (b), blur overexposure (c), and overexposure (d).

(a) (b) (c) (d)

Fig. 3: Image samples of highest std from TID2013 [7]. Tone-mapping, increased saturation and contrast.

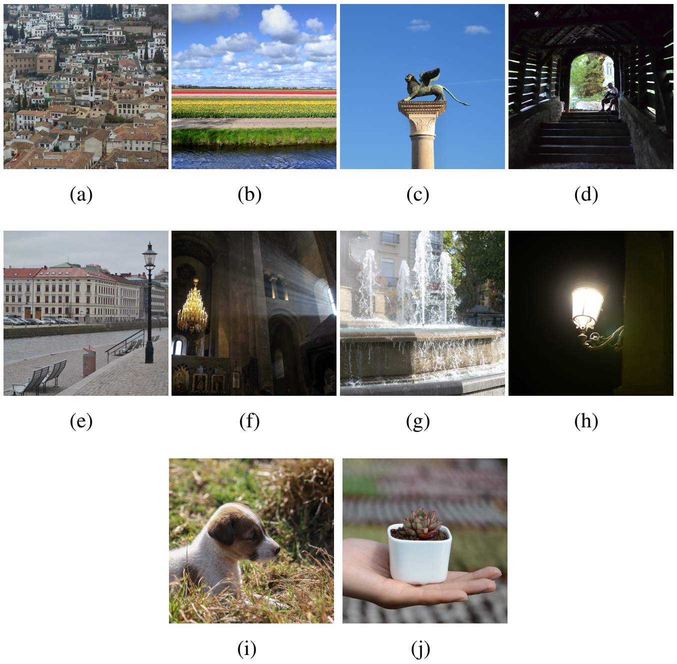


(a) (b) (c) (d)

Fig. 4: Image samples of highest std from LIVE in the wild [6]. Images of high dynamic range content and silhouettes.

databases, which show the largest standard deviation values among subjective scores (Figures 1 - 4). Analyzing the images, a few general trends can be observed. For example, in the case of Kadid10K images (Figure 1), images with blocking (Figures 1(c) and (d)) and quantization artifacts (Figures 1(a) and (b)) show a high standard deviation in the subjective scores. High standard deviation is seen in the images, which include artistic effects (Figure 2(b)), low aesthetic content (Figure 2(a)), overexposed or underexposed areas (Figure 2(d)), and blurred areas (Figure 2(c)) in the Koniq10K (Figure 2) dataset. TID2013 results (Figure 3) represent tone-mapped images, often with increased contrast and saturation (Figures 3(a) - (d)). Finally, in the case of the LIVE in the wild database (Figure 4), images with high dynamic range content and silhouettes, which are often difficult to judge (Figure 4(a) - (d)) show the highest variation among subjective scores. Analyzing these images, we can gain some insights about how to structure our dataset and which images and distortions should we include.

For our experiment, we chose ten images (Figure 5). The images were chosen to cover a range of image attributes, color volume, and content. The dataset includes busy and homogeneous images, architecture with straight lines, day and night images, high dynamic range images, sky, grass and skin



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 5: Images chosen for this work. Last image is from [21].

tones, people and animals, simple and more complex images, blurred background, and silhouette or back light. The ten original images will be distorted with various algorithms to simulate natural camera effects during acquisition.

To find the distortions with the largest variability between observers, we degraded the images with 21 different distortions. Most of the distortions have been chosen from the Kadid10K [19] database with publicly available distortion generating codes of Kadid700K [22]. These distortions are: gaussian blur (1), lens blur (2), motion blur (3), color shift (5), color quantization (6), color saturation 1 or saturation decrease (7), color saturation 2 or saturation increase (8), JPEG2000 (9), JPEG (10), white noise (11), multiplicative noise (14), denoise (15), brighten (16), darken (17), mean shift (18), jitter (19), quantization (22), high sharpen (24), contrast increase (25). Contrast increase (25) and decrease (27) were split into two separate distortions. Additionally, lens distortion (26), which imitates wide angle geometrical lens distortion was added. The distortions were chosen according to their authenticity and possibility of emergence in the process of image acquiring. We used the three first degradation levels from the Kadid700k distortion generator. The same perception level difference was applied during the level generation. For Kadid distortions, the original level changes were applied. For lens distortion, parameters -0.4, -0.6, -0.8 were chosen [23].

In total, we had 640 images (10 originals $\times$ 21 distortion $\times$ 3 levels plus 10 originals). 63 images were randomly selected and repeated as a control factor to check observer repeatability. We utilized 713 images. The images are 800×800 pixels to avoid rescaling on a display with a resolution of 1920×1080, which will allow us to display two images simultaneously.

B. Experimental setup

The experiment was set up in a laboratory under controlled environment. According to CIE [24] and ITU [25] standards, the ambient illumination at the location of the observer was fixed to 20 lux. We used an Eizo ColorEdge CS2420 monitor, which was calibrated with an i1-xrite spectrophotometer to sRGB. The viewing distance for the observers was fixed to 50 cm. The visual acuity of each observer was tested with Snellen visual acuity and color blindness Ishihara test.

Category judgment approach with 5 categories was used to evaluate the images using the QuickEval platform [26]. Categories were chosen in accordance with CIE and ITU standards, which were: very good, good, fair, bad, very bad. Each observer should decide for her/himself how to use and understand the scale. Instructions included general guidance: “Please rate the quality of the image.” The users had an opportunity to get acquainted with the scale during a brief training session, which included six images of different degradation levels. Images which were shown randomly to observers and included the original images. Observers did not have prior information about applied distortions.

C. Experiment procedure

The experiment consisted of two parts for each observer. The whole dataset of 713 images was split in half and observers had a choice if they want to do both parts together, take a break in between, or finish in different days. In total, each image was evaluated by 22 observers (16 males and 6 females with an average age of 30) with normal color vision. All observers had prior experience with image processing. Average observation time per image was around 3 seconds. The quantitative experiment was followed by a short interview in which they were asked the following questions:

- general comments,
- which distortions did you find most annoying,
- which distortions were less annoying for you, even when you noticed them,
- which original images did you rate higher or lower and why.

The answers were recorded along with observation remarks, noted while the observers carried out the experiment.

IV. Experimental results

A. Qualitative results

The following observations can be concluded from the subjective scores and remarks given by the participants:

1) The highest variability among observers are seen in the case of: saturation, contrast, lightness, lens distortion, blur, noise artifacts, blocking artifacts, and denoising artifacts.
2) Most observers agreed that a small shift in attributes such as contrast, lightness, and saturation can sometimes improve the image quality, while big changes are very disturbing. The same observation was seen in quantitative results (Figures 6(b), 6(f), and 7(b)) where in the case of saturation, contrast, and lightness an increase at level 1 or 2 has a positive impact for some observers, whereas level 3 was usually judged more negatively.

3) When it comes to the most disturbing distortions, observers generally agree that visual artifacts like "grainy type of noise", contouring, compression and quantization are mostly disturbing.
4) Observers judged distorted images, which are similar to gray-scale images (saturation decrease, level 3), as well as slightly saturated images and images with increased contrast, higher than the original images.
5) Some observers had completely opposite opinions about the quality of the original images.
6) Observers agreed that content plays an important role on the perception of distortions in the image. As an example, the same distortions were perceived on different images quite differently. This could be linked to the fact that depending on the content they were less visible or mattered less (Figures 8(a) and 8(b)). This phenomena was previously used in [27], where specific regions in the image where seen to be more sensitive to distortion than others.
7) While observers were instructed to evaluate the quality of the image, it was evident that aesthetic properties played an impact on their judgment.
8) Observers usually disagree not on the distortion itself, but on a combination of the distortion and image content.
9) Different observers are sensitive to different images, distortion types, and levels. If some accept one distortion, others completely reject an image, which has been slightly affected by the same distortion. This is similar to what was also pointed out in [5].
10) Throughout the experiment, observers were able to detect the original images and distortions used in the dataset. This naturally resulted in a higher consistency in their judgment, which is similar to the findings of Leisti et al. [16].

B. Quantitative results

To find the distortions with the largest variability between observers, we calculated the standard deviation of the observers' ratings for each distortion. The results are represented in Table I. Additionally, we analyzed distribution plots to find bimodal, multimodal, or equal distributions, which indicate the distortions with the largest variability (Figures 6, 7). We also analyzed consistency, using hidden duplicates for each distortion of randomly chosen originals (total 63 images per observer). Based on the ratings given to the duplicate images, the standard deviation for each observer was computed and fluctuated between 0.22 to 0.53.

The standard deviation in Table I is computed using Equation (1). Since individuals use the scale differently, instead of the original subjective quality scores we use the difference between the scores given to the original and the corresponding distorted images (Equation (2)).

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2} \tag{1}$$

where n is the number of observations, $\overline{d}$ is the mean value of observations and d_i is the difference between the quality ratings of distorted and corresponding original images, calculated as:

$$d_{i,j} = r_{i,ref}(j) - r_{i,j} \tag{2}$$

where $r_{i,j}$ is the rating of observer i for image j, while $r_{i,ref}(j)$ is the rating for the corresponding original image [28]. Standard deviation in column one is computed separately for each unique combination of distortion, level and image (Equation (3)) and averaged afterwards, while in second column we omit image information and in the third column we compute standard deviation for each distortion across all images regardless of level. The last column represents corresponded standard deviation from Kadid10K [19]. The table is sorted according to the first column, which takes the most information into account.

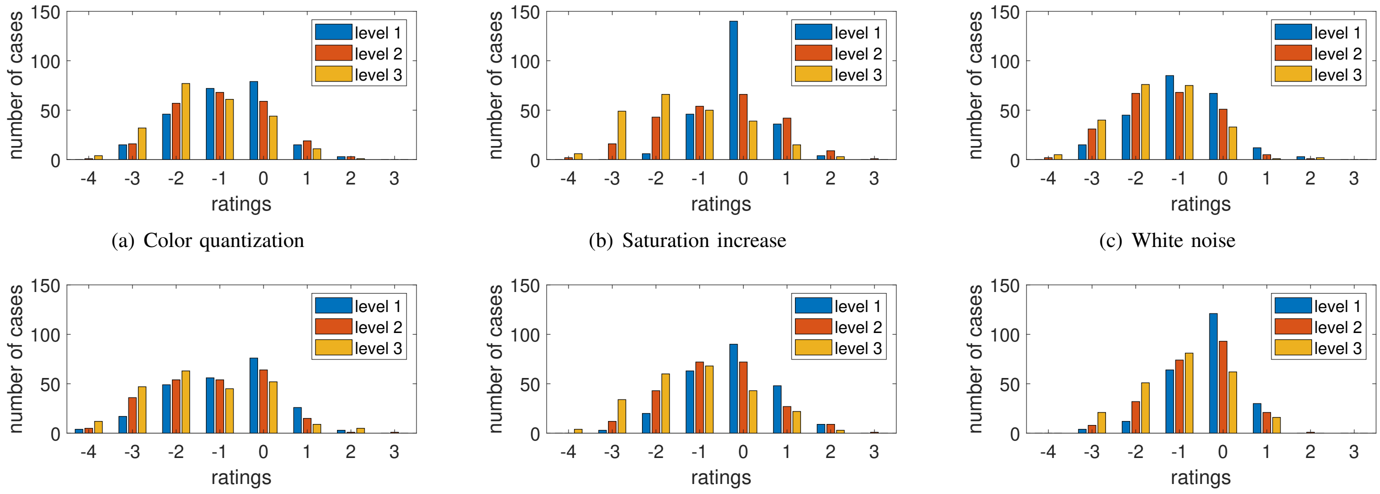
$$for\ image = 1:10,\ level = 1:3$$
$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_{i,image,level} - \overline{d}_{image,level})^2} \tag{3}$$

According to the results, we can see that the distortions with the largest variability include saturation, contrast, quantization, white noise, and radial lens distortion (Figure 6). On the contrary, the distortions with the lowest variability and so a degree of consistency among observers are mean shift, change in lightness, jitter, blurring effects, and JPEG2000 (Figure 7). Variability in the ratings for lens distortion might be not visible on general level, but more apparent on image level (Figures 6(f) and 8(a)). Variability in the ratings for the images with different types of added noise is on the contrary more visible on general level (Table I, column 3).

On image level it is clear that the image content plays an important role on the overall judgment of the observers. Figure 8(a)) shows that observers are more sensitive to radial lens

TABLE I: Averaged standard deviation for observers' ratings across all distortions sorted from highest to lowest variability.

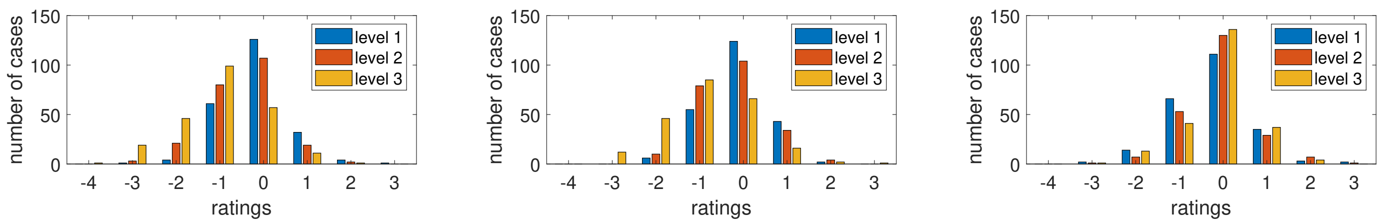
ID	Distortion name	Standard deviation level & image	Standard deviation by level	Standard deviation general	Standard deviation Kadid10K
7	color saturation 1	1.06	1.17	1.19	0.98
25	increase contrast	1.06	1.16	1.24	0.76
8	color saturation 2	1.05	1.12	1.30	0.76
6	color quantization	0.99	1.13	1.15	0.84
22	quantization	0.95	1.29	1.32	0.86
26	lens distortion	0.93	0.95	0.99	-
11	white noise	0.92	1.06	1.09	0.79
27	decrease contrast	0.92	1.04	1.23	0.76
15	denoise	0.92	1.02	1.07	0.70
10	JPEG	0.89	1.03	1.10	0.69
5	color shift	0.89	0.94	1.26	0.75
14	multiplicative noise	0.89	1.05	1.15	0.76
3	motion blur	0.88	0.92	1.07	0.67
24	high sharpen	0.88	0.92	1.01	0.73
2	lens blur	0.87	0.90	1.11	0.63
9	JPEG2000	0.86	0.91	0.93	0.67
1	gaussian blur	0.84	0.86	1.03	0.61
19	jitter	0.84	0.88	0.95	0.64
16	brighten	0.83	0.87	0.94	0.72
17	darken	0.82	0.86	0.86	0.69
18	mean shift	0.80	0.84	0.84	0.63



(a) Color quantization (b) Saturation increase (c) White noise

(d) Quantization (e) Contrast increase (f) Lens distortion

Fig. 6: Distribution of subjective scores for distortions with the largest variability among observers at different level. In the figure blue corresponds to level 1, red to level 2, and yellow to level 3.



(a) Jitter (b) Brighten (c) Mean shift

Fig. 7: Distribution of subjective scores for distortions with the lowest variability among observers. In the figure blue corresponds to level 1, red to level 2, and yellow to level 3.

distortion in the images that contain geometrical forms and lines, e.g., buildings (Figures 5(b),(d) and (e)) rather than the ones with circled or undefined shape objects (Figures 5(i) and (j)). In the case of multiplicative noise (Figure 8(b)), the noise is more visible in the images with larger homogeneous areas (Figure 5(b), (c), and (j)), comparing to more complex images (Figure 5(a) and (i)). This finding is similar to that of Ninassi et al. [14]. While including a diverse set of images could help in neutralizing content dependency, for accurate results, the differences on an image level and the role of content in subjective evaluation should be taken into account.

Combining quantitative and qualitative data together, we can make a conclusion about certain types of distortions. For example, image distortions such as changes in contrast, saturation, and lightness play a significant role in observers' preferences. Although image content plays an important role in distortion visibility, we still can conclude that distortions such as lens curvature, quantization, some types of noise, and sharpness/blur play a role in observers' preferences and are judged by observers differently.

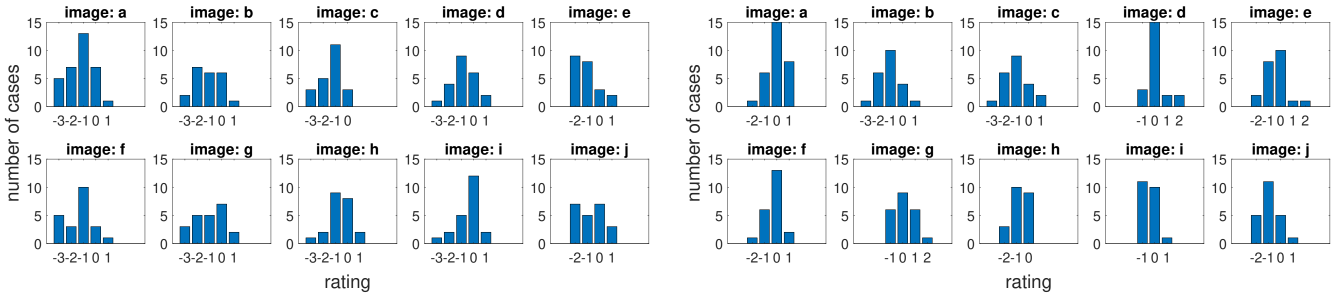
C. Results comparison

Ponomarenko et al. in TID2013 database [7] reported the results of the largest variability between observers utilizing RMSE values. They reported the smallest variability in JPEG2000 and spatially correlated noise. While we did not use exactly spatially correlated noise, the closest distortion to this could be multiplicative noise, which is rated as having an average variability. JPEG2000 was considered one of the distortions with less variability in our study as well. They found contrast increase to be one of the most difficult distortions for observers to judge, which has one of the largest variability in our study as well.

It is also interesting to compare our results to the Kadid10K database. The standard deviation results are provided for each image, so we can analyze their data to see the distortions that have the largest variability. As can be seen (Table I), color saturation, contrast change, quantization, color quantization, and white noise have higher standard deviations, which is similar to our results.

V. Conclusion and Future Work

In this work, we investigated the influence of different distortions on individual observers' judgment of image quality. We conducted a subjective study with 640 distorted images judged by 22 observers. In total, we investigated 21 distortions at three degradation levels each. We conducted a qualitative and quantitative subjective study under controlled lab environment and combined the results together. For analysis, we used combinations of distortion type and level, and additionally



(a) Lens distortion, level 3

(b) Multiplicative noise, level 1

Fig. 8: Distribution of subjective scores on an image level. Labels on the top of each plot correspond to the images in Figure 5.

investigated the results on the image level. Results show that the following distortions have the largest variability in observer ratings: saturation, contrast, quantization, lens distortion, blur/sharpness, and noise. We found a dependency between distortion perception and image content.

In future work, we are planning to use these results to create a larger dataset. This will allow us to create an IQM, able to predict individual observer preferences, and additionally, enhance images according to particular user preferences.

References

[1] Wang, Z., Bovik, A.C., Lu, L.: Why is image quality assessment so difficult? In: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing. Volume 4. (2002) IV–3313–IV–3316

[2] Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters **20**(3) (2012) 209–212

[3] Pedersen, M., Ali Amirshahi, S.: Colourlab image database: Geometric distortions. In: Color and Imaging Conference. Volume 2021., Society for Imaging Science and Technology (2021) 258–263

[4] Azimian, S., Torkamani Azar, F., Ali Amirshahi, S.: How good is too good? a subjective study on over enhancement of images. In: Color and Imaging Conference. Volume 2021., Society for Imaging Science and Technology (2021) 83–88

[5] Jayaraman, D., Mittal, A., Moorthy, A.K., Bovik, A.C.: Objective quality assessment of multiply distorted images. In: 2012 Conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR), IEEE (2012) 1693–1697

[6] Ghadiyaram, D., Bovik, A.C.: Massive online crowdsourced study of subjective and objective picture quality. IEEE Transactions on Image Processing **25**(1) (2016) 372–387

[7] Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., et al.: Image database tid2013: Peculiarities, results and perspectives. Signal processing: Image communication **30** (2015) 57–77

[8] Ribeiro, F., Florencio, D., Nascimento, V.: Crowdsourcing subjective image quality evaluation. In: 2011 18th IEEE International Conference on Image Processing. (2011) 3097–3100

[9] Li, L., Zhu, H., Zhao, S., Ding, G., Lin, W.: Personality-Assisted Multi-Task Learning for Generic and Personalized Image Aesthetics Assessment. IEEE Transactions on Image Processing **29** (2020) 3898–3910

[10] Park, K., Hong, S., Baek, M., Han, B.: Personalized image aesthetic quality assessment by joint regression & ranking. Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017 (2017) 1206–1214

[11] Ren, J., Shen, X., Lin, Z., Mech, R., Foran, D.J.: Personalized Image Aesthetics. Proceedings of the IEEE International Conference on Computer Vision **2017-Octob** (2017) 638–647

[12] Virtanen, T., Nuutinen, M., Vaahteranoksa, M., Oittinen, P., Häkkinen, J.: Cid2013: A database for evaluating no-reference image quality assessment algorithms. IEEE Transactions on Image Processing **24**(1) (2014) 390–402

[13] Sun, W., Zhou, F., Liao, Q.: Mdid: A multiply distorted image database for image quality assessment. Pattern Recognition **61** (2017) 153–168

[14] Ninassi, A., Le Callet, P., Autrusseau, F.: Pseudo no reference image quality metric using perceptual data hiding. In: Human vision and electronic imaging XI. Volume 6057., International Society for Optics and Photonics (2006) 60570G

[15] Leisti, T., Radun, J., Virtanen, T., Halonen, R., Nyman, G.: Subjective experience of image quality: attributes, definitions, and decision making of subjective image quality. In Farnand, S.P., Gaykema, F., eds.: Image Quality and System Performance VI. Volume 7242., International Society for Optics and Photonics, SPIE (2009) 130 – 138

[16] Leisti, T., Häkkinen, J.: Learning to decide with and without reasoning: How task experience affects attribute weighting and preference stability. Journal of Behavioral Decision Making **31**(3) (2018) 367–379

[17] Calabria, A.J., Fairchild, M.D.: Perceived image contrast and observer preference i. the effects of lightness, chroma, and sharpness manipulations on contrast perception. Journal of imaging Science and Technology **47**(6) (2003) 479–493

[18] Pedersen, M., Rizzi, A., Hardeberg, J.Y., Simone, G.: Evaluation of contrast measures in relation to observers perceived contrast. In: Conference on Colour in Graphics,Imaging, and Vision, IS&T (2008) 253–258

[19] Lin, H., Hosu, V., Saupe, D.: Kadid-10k: A large-scale artificially distorted iqa database. In: 2019 Tenth International Conference on Quality of Multimedia Experience (QoMEX), IEEE (2019) 1–3

[20] Hosu, V., Lin, H., Sziranyi, T., Saupe, D.: Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. IEEE Transactions on Image Processing **29** (2020) 4041–4056

[21] WorkingDOG: Pixabay image. https://pixabay.com/images/id-6679662/ (2021-09-05)

[22] Lin, H., Hosu, V., Saupe, D.: KADID-10k IQA Database. http://database.mmsp-kn.de/kadid-10k-database.html (2021)

[23] Jaap de Vries: Barrel and pincushion lens distortion correction. https://se.mathworks.com/matlabcentral/fileexchange/37980-barrel-and-pincushion-lens-distortion-correction (2012)

[24] BT RECOMMENDATION ITU-R: Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union (2002)

[25] Morovic, J. ITU. Recommendation BT.500: Guidelines for the evaluation of gamut mapping algorithms. PUBLICATIONS-COMMISSION INTERNATIONALE DE L ECLAIRAGE CIE **153** (2003) D8–6

[26] Van Ngo, K., Dokkeberg, C.A., Farup, I., Pedersen, M., et al.: Quickeval: a web application for psychometric scaling experiments. In: Image Quality and System Performance XII. Volume 9396., International Society for Optics and Photonics (2015) 93960O

[27] Amirshahi, S.A., Torkamani-Azar, F.: Human optic sensitivity computation based on singular value decomposition. Optica Applicata **42**(1) (2012) 137–146

[28] Mohammadi, P., Ebrahimi-Moghadam, A., Shirani, S.: Subjective and objective quality assessment of image: A survey. arXiv preprint arXiv:1406.7799 (2014)