CD^2 : Combined Distances of Contrast Distributions
for the Assessment of Perceptual Quality of Image Processing

Sascha Xu
X-Motive GmbH
sascha.xu@xmotive.de

Jan Bauer
Hochschule Karlsruhe
jan.bauer@hs-karlsruhe.de

Benjamin Axmann
Daimler AG
benjamin.axmann@daimler.com

Abstract

The quality of visual input is very important for both human and machine perception. Consequently many processing techniques exist that deal with different distortions. Usually image processing is applied freely and lacks redundancy regarding safety. We propose a novel image comparison method called the Combined Distances of Contrast Distributions (CD^2) to protect against errors that arise during processing. Based on the distribution of image contrasts a new reduced-reference image quality assessment (IQA) method is introduced. By combining various distance functions excellent performance on IQA benchmarks is achieved with only a small data and computation overhead.

1. Introduction

Once captured a raw image data flow undergoes a complex chain of processing. The dynamic range may be adjusted, the image may be sharpened or denoised, contrast may be enhanced, then the image may be compressed to meet bandwidth requirements. Many instances like cameras, apps, etc. offer and perform processing. A lot of resources are spent altering and improving images for different contexts. However nobody is perfect, so all image processing techniques have potential failure modes (examples in figure 1). Image sharpening can amplify noise, denoising on the other hand might erase relevant details. Lossy JPEG compression leaves behind blocking artifacts. Neural-network enhancers are little understood and unreliable.

For safety-critical situations unguarded processing is an issue. Consider a rear-view camera in a car which provides live video data to the driver. Wrong processing in wrong moments can have dangerous consequences by hindering/misleading the driver. Algorithms are affected by this as well: Dodge and Karam [6, 7] show that many popular networks such as ResNet [29] and VGG [27] are susceptible to noise and blur distortions, so much that their classification performance falls far below human level.

We propose in this paper a redundancy mechanism to oversee the results of image processing. The method examines an original image and its processed version to judge the quality of the result and helps ensure that no significant perceptual deterioration happened. Overall such a measure can be used to increase the safety integrity level of an image processing system.

This task fits very well into the domain of image quality assessment (IQA). The goal is to conduct a basic comparison between pairs of images: What kind of distortions have been introduced? How severe do they impact the image?

There exist three modi operandi for IQA: Full-reference IQA (FR-IQA) has access to the complete reference and result image. The structural similarity index (SSIM) by Wang et al. [32] is the most prominent IQA index. SSIM assesses image similarity with pixel statistics about luminosity, contrast and structure. Other FR-IQA approaches operate on gradient maps [38, 4], frequency domain coefficients [22] or in case of the FSIM index on both [40].

![Figure 1: Examples for image processing failures](image)

(a) Original scene  (b) White balanced
(c) JPEG compressed  (d) Noisy
Blind IQA aims to judge the quality of an individual image. There is no reference image, only the observed image is available. Traditionally this was approached similar to full-reference methods using features from the frequency and gradient domain [21, 37, 18]. With the rise of neural networks they have been applied for blind IQA as well [11, 2].

The middle ground between full-reference and blind IQA is occupied by reduced-reference IQA (RR-IQA). Here the use of natural scene statistics and distributions to represent images is widespread [14, 33, 30]. By applying statistical distances on the distributions of reference and processed image RR-IQA requires only limited data.

The goal of this paper is to develop a method to safeguard and control results of image processing and transmission. Therefore the reference image should be taken into account. A full-reference method is costly in bandwidth and usually infeasible. Consequently we opt for a reduced reference approach like seen in figure 2. The extracted features should be small enough to be attached as meta-data to an image stream. Then for a processed image based on the features of the reference a significant deterioration may be detected with a corresponding level of quality decrease to enact a warning or counter measures.

To achieve this, the natural scene statistics of contrast, a key factor in visual perception, are investigated. The distribution of image contrast, condensed into a histogram is derived from image gradients. Dalal et al. [5] display the power and utility of gradient histograms by using them for a both accurate and rapid object detection algorithm.

Our method does not look explicitly at the individual contrast histogram values. Instead the change on the histogram caused by various distortions is observed and measured with special distance functions. In total 11 statistical distances are computed between the contrast histogram features of a reference and processed image. They are combined using LightGBM [12] boosting models to predict distortion types and evaluate quality differences.

The topics and issues from this introduction are covered by the following chapters: First the feature extraction process for contrast histograms is discussed and described in section 2. Section 3 examines the impact of different image distortions on these histograms and designs a corresponding set of distance functions. Related work is covered in section 4, then in section 5 we evaluate the performance of our method on the LIVE [25], TID2013 [20] and CSIQ [13] datasets. The final section finishes with conclusions.

2. Contrast Distribution as Features

Contrast is a major stimulus for the human visual system. Shapley and Reid show in their experiments that the perception of an objects brightness is mainly determined by the average contrast around the object, not the absolute brightness of itself [23, 24]. Testing the effect of visual disturbances, Arden finds that contrast sensitivity is the key [1].

In this paper we consider contrasts exclusively in the brightness domain. They can be quantified in form of the difference in luminance. For color on the other hand there is no one-dimensional space. Differences are hard to quantify and compare in a universal context.

2.1. Gradient Distributions

Filtering an image with gradient operators is an effective way to measure luminance contrast. As result one possesses a pixelmap of gradients. This gradient map is aggregated into the distribution of gradients and much reduced in size.

\[ S_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, S_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \]

Figure 3: 3x3 Sobel Filters
3x3 Sobel filters (figure 3) are convolved with the luminance channel to extract the x-gradients $G_x$ and y-gradients $G_y$ of an image. RGB is transformed to Luma using the formula: $I = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B$. Assuming 8 bit grayscale resolution the image gradients $G_x, G_y$ are integers contained in the domain $[-1020, 1020]$.

This results two discrete gradient distributions $P(G_x = n)$ and $P(G_y = n)$. They describe the frequency of how often a gradient value occurs in the image. For natural images Huang et al. [10] find that the pixel differences are zero symmetric distributed. To simplify the distribution the gradient sign is dropped. Only the absolute gradients (denoted as $g = |G|$) and their distributions $P(g = n)$ will be considered from now on.

2.2. Magnitudes versus Individual Gradients

Gradients can be processed further by transforming them to their magnitude $g_{mag} = \sqrt{g_x^2 + g_y^2}$ and the respective magnitude distribution $P(g_{mag} = m)$. Many IQA models like GMSD [38] or FSIM [40] rely exclusively on gradient magnitudes. Substituting $g_x$ and $g_y$ by their magnitudes however comes at a loss of information. The gradient distributions $P(g_x = n)$ and $P(g_y = n)$ are not reconstructable from $P(g_{mag} = m)$ unless $g_x$ and $g_y$ in images in general are sampled from the same distribution.

Figure 4 shows a section of the empirical distribution of signed and absolute gradients from the reference image in figure 2. First we observe that the assumption of zero-symmetry seems to hold. But the two individual gradient distributions seem to be visually different.

\[ H_0 : \forall n \in [0, 1024] \cap N : P(g_x = n) = P(g_y = n) \quad (1) \]

We test the hypothesis of gradient distribution equality with the reference images of three IQA datasets [25, 20, 13] using the two sample Kolomogorov-Smirnov test. The test is conducted on the empirical distributions $\hat{P}(g_x)$ and $\hat{P}(g_y)$ of every image. Note that the distributions are allowed to vary between, but not within the images.

The resulting p-values for each dataset are grouped together and the Holm-Bonferroni adjustment is performed. Using this methodology $H_0$ is rejected for every single image. We conclude that in images $g_x$ and $g_y$ are sampled from different distributions $P(g_x = n)$ and $P(g_y = n)$ and that $P(g_{mag} = m)$ is not sufficient to represent them both. Hence the feature representation of an image comprises of both distributions.

2.3. Gradient Distributions to Contrast Histograms

Each discrete gradient distribution $P(g = n)$ consists of 1021 probabilities. To reduce this considerable amount of data we start with this intuition: A gradient with value $n$ or $n+1$ will be perceived very similarly. There is little gain in making a distinction so it lends itself to quantize the distribution into a histogram. The histogram for $P(g_x = n)$ is denoted as $Hist(g_x)$ (likewise $Hist(g_y)$). The main challenge lies in defining bins for similar gradient/contrast values. To divide the domain of gradients we rely on the Fechner law [9]. The Fechner law postulates that the actual and perceived change of a stimulus intensity scales logarithmic.

The domain of $[0,1020]$ is split into bins using powers of 2 as thresholds. Each bin edge represents a linear increase in the perceived strength of a gradient stimulus. 5 intermediate thresholds are added in the middle of the domain to increase resolution to a total of 16 bins.

For our reduced-reference features this means that the gradient distributions $P(g_x = n)$ and $P(g_y = n)$ are transformed into histograms $Hist(g_x)$ and $Hist(g_y)$. These histograms are linearly spaced for the actually perceived contrast of the gradients. The result of such a binning for the distributions in figure 4b can be seen in figure 5.
2.4. Interaction between Gradients

In section 2.2 gradient magnitudes were discussed. They offer information about the interaction of X-gradients and Y-gradients. The information contained in the joint distribution of gradients \( P(g_x = n, g_y = m) \) is examined more closely in the following.

It can be expected that there is a dependency between \( g_x \) and \( g_y \) since they are partly computed on the same pixels. Table 1 shows for three IQA-datasets the median Pearson and Spearman correlation coefficients between the gradients \( g_x \) and \( g_y \) as well as the Information Quality Ratio (IQR) [34] using the previously defined gradient contrast histograms \( \text{Hist}(g_x) \) and \( \text{Hist}(g_y) \) (Mutual Information = MI, Entropy = H).

\[
IQR(g_x, g_y) = \frac{MI(\text{Hist}(g_x), \text{Hist}(g_y))}{H(\text{Hist}(g_x), \text{Hist}(g_y))}
\] (2)

The Pearson and Spearman correlations are very low (with low std.). There appears to be no monotone relationship between the gradients. Figure 6 shows a scatter plot of \((g_x, g_y)\) from the reference image in figure 2. No obvious pattern exists and the Information Quality Ratio indicates that there is relatively little mutual information between \(\text{Hist}(g_x)\) and \(\text{Hist}(g_y)\).

Sending over a joint histogram \(\text{Hist}(g_x, g_y)\) with \(16^2 = 256\) bins causes a lot of extra overhead. It contains only a modest amount of mutual information. Therefore we restrict ourselves to transmitting the two marginal distribution histograms \(\text{Hist}(g_x)\) and \(\text{Hist}(g_y)\) for together \(2*16 = 32\) bins.

| Dataset     | Pearson | Spearman | IQR [34] |
|-------------|---------|----------|----------|
| LIVE [25]   | -0.0132 | -0.001063| 0.2713   |
| CSIQ [13]   | -0.0212 | -0.0077  | 0.2799   |
| TID2013 [20]| -0.0192 | -0.0207  | 0.2756   |

Table 1: Correlation measures between \(g_x\) and \(g_y\)

2.5. Feature Set

The complete feature set representing an image consists of two 16 bin histograms \(\text{Hist}(g_x)\) and \(\text{Hist}(g_y)\) for the Sobel X/Y gradient distributions (see figure 5) In total \(16*2 = 32\) numbers have to be transmitted. For lossless histogram transmission of an image with \(m\) pixels the bins must have \([\log_2(m)]\) bits. For a 512x512 image our features use \(32*18 = 576\) bits, compared to the full reference requirement of \(512*512*3*8 = 6291456\) bits. Such feature data can easily be transmitted or attached as meta data to an image stream. The histogram itself is a compact data structure that simultaneously acts as a logarithmic contrast response model for gradients.

3. Distances for Contrast Distributions

This feature set for images is the base for the assessment of perceptual quality and distortions due to processing. For a reference image the contrast histograms \(\text{Hist}_{ref}(g_x)\) and \(\text{Hist}_{ref}(g_y)\) are available. The same features are extracted from the processed image.

The problem may be approached with a variety of statistical distance functions. They are combined into a new feature vector describing the perceptual differences between the reference image and the processed image. All distance functions compare one histogram of the reference image to one of the processed image. They are used once with the X-contrast histograms and once more with the Y-contrast histograms. No distinction is made between X/Y-contrast. The histograms for reference and processed image are denoted as \(\text{Hist}_{ref}\) and \(\text{Hist}_{pro}\). We will look at general distortion types and how they change the contrast histograms. A selection of distance functions is designed to capture these changes.

3.1. Contrast Enhancement

Increasing the contrast in an image should normally increase the images gradients. Objects/structures are made better visible. Accordingly the strength of existing edges
and gradients increases.

Figure 7 shows an example for image enhancement and its effect on the contrast histograms. The histogram of the reference in yellow changes to the green histogram of the enhanced image. The distributions weight is shifted towards larger gradients but still the histograms remain relatively similar.

To quantify similarity between reference and result distribution we use the Kullback-Leibler divergence

\[ KL = \sum_{i=1}^{16} Hist_{ref}(i) \cdot \log \left( \frac{Hist_{ref}(i)}{Hist_{pro}(i)} \right) \] (3)

and the histogram intersection.

\[ Intersection = \sum_{i=1}^{16} \min(Hist_{ref}(i), Hist_{pro}(i)) \] (4)

Additionally the Earth-Movers-Distance (Wasserstein distance) is used. It describes the amount of contrast that needs to be shifted to transform the reference histogram to the result. (Earth-Movers Distance = Emd, Emd0 = 0)

\[ Emd_{i+1} = Hist_{ref}(i) + Emd_{i} - Hist_{pro}(i) \]

\[ TotalEmd = \sum_{i=1}^{16} |Emd_{i}| \] (5)

The set of histogram distances is rounded out with the Total-Variation distance, which simply describes the maximum bin difference between Hist_{ref} and Hist_{pro}.

\[ TotalVariation = \max( |Hist_{ref}(i) - Hist_{pro}(i)| ) \] (6)

Together these distances give different estimates of perceptual similarity or degradation.

3.2. Image Blur

Blur typically average filters the image to reduce noise but also contrast. The variety and range of contrasts can be represented by the Shannon Entropy of a histogram. Thus depending on how much contrast is lost the entropy decreases. The amount of blur is assessed by the difference in entropy between Hist_{ref} and Hist_{pro}. (Entropy = H)

\[ EntropyGap = H(Hist_{ref}) - H(Hist_{pro}) \] (7)

The worst case outcome for contrast reductions makes objects indistinguishable from their surroundings. Figure 1b shows a scene where due to processing the pedestrians on the right are barely visible.

For this we incorporate a simple model for the just noticeable difference (JND). When the gradients of an object are not noticeable anymore it vanishes from our perception. Globally a loss of image scene content results in more gradients below the JND threshold. We model contrast sensitivity with constant JND gradient thresholds. This corresponds to comparing the sum of the b lower bins of H_{ref} and H_{pro}. For robustness two JndChange values are calculated for b = 3 and b = 4.

\[ JndChange = \sum_{i=1}^{b} Hist_{ref}(i) - \sum_{i=1}^{b} Hist_{pro}(i) \] (8)

3.3. Compression Artifacts

Previously image processing from the gradient domain was examined. However lossy compression techniques operate in the frequency domain. The JPEG [31] standard uses the discrete cosinus transform (DCT) to encode 8x8 pixel blocks. DCT is widespread and also used by video coding schemes like MPEG [8] or H.261.
JPEG compression causes blocking: Small DCT coefficients are quantized to zero. When this happens to all coefficients of a block it results in a 8x8 block of equal pixel values with zero gradients. Figure 8 shows the histogram changes caused by JPEG compression for example 1c. The amount of blocking is strongly reflected in an increased proportion of zero gradients. They are contained in the first histogram bin.

\[ \text{Blocking} = \text{Hist}_{ref}(1) - \text{Hist}_{pro}(1) \quad (9) \]

### 3.4. Noise Distortions

Lastly the effect of noise is observed. Figure 9 shows the change after noise is introduced in image 1d. There is a significant increase in the very high contrast proportion of the histogram.

When random noise is inserted it produces strong, non-natural edges and gradients. We gauge this by thresholding the amount of large gradients. The upper \( t \) bins are summed up. Again two \( t \) values (10 and 12) are applied for the sake of robustness.

\[ \text{NoiseInc} = \sum_{i=t}^{16} \text{Hist}_{ref}(i) - \sum_{i=t}^{16} \text{Hist}_{pro}(i) \quad (10) \]

### 3.5. Combined Distances on Contrast Distributions

In total there are 10 unique distance functions that are applied on the two x-histograms and the y-histograms (for function 8 and 10 two thresholds are applied). They are computed one time each for the X- and Y-contrast distributions so there are in total 2*10 = 20 combined distances. This combined distance vector represent the perceptual changes between a reference image and a processed version.

There are two tasks left: detect/classify the kind of distortion and judge the difference in quality. Gradient tree boosting offers classification and regression models with excellent performance and fast inference times. In our implementation the LightGBM [12] boosting framework is employed. A classifier takes the combined distances vector and predicts the type of distortion. A regression boosting model uses the error type in addition to the combined distances to predict the difference in quality. To train these models we need suitable data which is acquired from IQA databases [25, 20, 13].

#### Table 2: Summary of Distances

| Distance Function | Formula |
|-------------------|---------|
| KL (3) | \( \sum^{16}_{i=1} \text{Hist}_{ref}(i) \ast \log(\frac{\text{Hist}_{ref}(i)}{\text{Hist}_{pro}(i)}) \) |
| Intersection (4) | \( \sum^{16}_{i=1} \min(\text{Hist}_{ref}(i), \text{Hist}_{pro}(i)) \) |
| Emd (5) | \( \sum^{16} |Emd_i| \) |
| TotalVariation (6) | \( \max(|H_{ref}(i) - H_{pro}(i)|) \) |
| EntropyGap (7) | \( H(\text{Hist}_{ref}) - H(\text{Hist}_{pro}) \) |
| JndChange (8) | \( \sum^{b=3}_{i=1} \text{Hist}_{ref}(i) - \sum^{b=3}_{i=1} \text{Hist}_{pro}(i) \) |
| JndChange (8) | \( \sum^{b=4}_{i=1} \text{Hist}_{ref}(i) - \sum^{b=4}_{i=1} \text{Hist}_{pro}(i) \) |
| Blocking (9) | \( \text{Hist}_{ref}(1) - \text{Hist}_{pro}(1) \) |
| NoiseInc (10) | \( \sum^{16}_{i=10} \text{Hist}_{ref}(i) - \sum^{16}_{i=10} \text{Hist}_{pro}(i) \) |
| NoiseInc (10) | \( \sum^{16}_{i=12} \text{Hist}_{ref}(i) - \sum^{16}_{i=12} \text{Hist}_{pro}(i) \) |

This concludes the methodology section of the proposed “Combined Distances of Contrast Distributions for the Assessment of Perceptual Quality of Image Processing” (short CD²). After related work the uses and performance of this new method will be put under inspection.

### 4. Related Work

Traditional FR-IQA methods operate in two steps: first pixelwise feature maps are computed and compared. Then
they are pooled together to a quality index. The peak signal-to-noise-ratio (PSNR) is a simple instantiation of this framework. The pixelwise squared differences are averaged together and the logarithm is taken. Wang et al. [32] have a more statistical approach by combining image mean, variance and pixel covariance similarity to the SSIM index. The visual saliency index (VSI) by Zhang et al. [39] is also a threefold index based on gradient magnitude, visual saliency and chrominance similarity. Frequency domain analysis of images is popular too. Sampat et al. [22] compare the similarity of wavelet coefficients for their CW-SSIM. An approach that combines both gradients and frequency is the FSIM index by Zhang et al. [40]. Gradient magnitude and phase congruency similarity maps are computed and pooled by congruency. LCSIM by Oszust [19] is a data-analytic approach that uses genetic algorithms to linearly combine FR-IQA indices (including the ones mentioned here).

RR-IQA is built on similar ideas and models, but must not use the complete reference image. Therefore most methods utilize so kind of image statistics for a reduced reference. Wang and Simoncelli [33] model wavelet-coefficient distributions as gaussians for different subbands. Li and Wang [14] further process the wavelet-coefficients with the divisive-normal-transform. As feature distance they both use the KL-divergence. Soundararajan and Boviks [28] entropic differencing RRED index is built on gaussian mixture models for subband coefficients. They measure image quality by comparing and summing up scaled block entropies for subbands.

Outside of the frequency domain Wu et al. [35] represent images with gradient orientation patterns aggregated into a histogram. Xu and Bauer [36] take the squared-sum of KL-divergences over a grid of gradient histograms as their distance index. Min et. al [16] base their reduced reference on visual saliency maps. Binary maps of pixel significance are compared to each other using SSIM and combined with features describing contrast change.

Neural networks are another option to provide compact embeddings for image. Bosse et al. [3] use convolutional nets for blind and FR-IQA. Mocanu et al. [17] restricted Boltzmann machines enable learned RR-IQA features and similarities. Liu et al. [15] use boosting for statistics of relative gradient orientation and magnitudes.

5. Performance Evaluation

The proposed method is tested on the LIVE (29 reference, 779 distorted images) [25], CSIQ (30 reference, 866 distorted images) [13] and TID2013 (25 reference, 3000 distorted images) [20] datasets. The reference images are subject to different types of compression, noise and other distortions. Human subjects give opinions about image quality that are transformed to difference in mean opinion score (DMOS) between reference and processed image (TID uses mean opinion scores).

We test two variations of our CD² method. CD²-A has access to the ground truth of distortion types for the DMOS regression. This corresponds to applications like quality aware image compression where the error type is given. When no specific information is available about the distortion, the alternative method CD²-B will rely on predicted labels provided by the trained boosting classifier. Both versions are compared with the FR-IQA metrics PSNR, SSIM [32], FSIM [40], VSI [39] and LCSIM [19] as well as the best performing RR-IQA indices SIR [16] and RRED [28].

Except for PSNR the original Matlab source code of the authors was used. For the mapping from index to DMOS/MOS score we follow the video quality experts [26] and fit a 5 parameter logistic function.

\[ f(x) = \beta_1 \left( \frac{1}{2} - \frac{1}{1 + \exp(\beta_2 \cdot (x - \beta_3))} \right) + \beta_4 \cdot x + \beta_5 \]

The experiments were run with Matlab R2019b on Ubuntu 18.04 using an AMD Ryzen 5 2600 processor and 16GB RAM. Our method is implemented in Python 3.7 with the Python version of LightGBM [12]. 1000 random train-test-splits (80% train, 20% test set) were performed. First the performance of the DMOS/MOS prediction is observed. Reported are the median root mean-square error (RMSE), mean absolute error (MAE) and coefficient of determination \( R^2 \).

5.1. Boosting Model Training

The weak learners for LightGBM were configured as follows: max_depth = 5, min_data_in_leaf = 5, num_leaves = 20. On the training set first a classification model for distortion types was trained with the softmax objective for 5 iterations. Then a DMOS/MOS regression model is optimized for 30 iterations with the ground truth distortion labels. CD²-A shows the performance on the test set using the original distortion labels. CD²-B is the performance when using the predicted labels from the classification model.

| Distortion      | Precision | Recall | F-Score \(_1\) |
|-----------------|-----------|--------|----------------|
| JPEG            | 0.8868    | 0.9462 | 0.9141         |
| White Noise     | 0.9777    | 0.9785 | 0.9777         |
| JPEG2000        | 0.6821    | 0.7293 | 0.6992         |
| Gaussian blur   | 0.6270    | 0.6822 | 0.6469         |
| Fastfading      | 0.5957    | 0.4415 | 0.5071         |

Table 3: Classification performance on LIVE [25]
5.2. Overall Performance

Table 4 shows the performance across different databases. The three best methods are highlighted in each row. Variant A is able to surpass most full-reference methods and is amid the best performing methods for each dataset. It is the only method that is a top performer across all datasets. The competing RR-IQA approaches are consistently outmatched and high levels of explained variance $R^2$ are reached.

CD$^2$-B uses predicted error labels and performs on a similar level to SIRR and RRED. Table 6 lists RMSE achieved on LIVE by distortion type. Here we see that CD$^2$-B is equivalent to CD$^2$-A on JPEG and noise distortions, but drops significantly in performance for JPEG2000, gaussian blurred and fastfading images. To further investigate this we focus on the performance of the distortion prediction on which CD$^2$-B relies.

Looking at the LIVE classification stats in Table 3 the prediction similarly is accurate for JPEG and noise distortions, but struggles when classifying JPEG2000, gaussian blurred and fastfading images. Fastfading distortions in LIVE are simulated by bit errors in a JPEG2000 stream. Because CD$^2$-B operates on distributions there needs to be a significant amount of bit errors to make an observable change for CD$^2$-B JPEG2000 leaves behind blurry images which CD$^2$-B can confuse as gaussian blurred and vice versa.

5.3. Complexity and Extensibility

The average CPU running time of CD$^2$ is in the middle along VSI (see Table 5). Most of the time is claimed by the feature extraction. For further optimization however the feature extraction can also be performed on a pixel stream (without frame buffer) making it a good fit for a possible implementation on integrated circuits.

In general the method operates on distributions and is thus applicable to all image sizes. It does not make restrictive assumptions about the contrast distribution like gaussian modelling. With 32 numbers of feature data CD$^2$ is on par with the best methods that have access to millions of pixels. The predicted quality and error types can be adapted to new applications/data, such as image improvements.

6. Conclusion

We presented a RR-IQA-method to safeguard and control image processing results where an anticipated application revolves around the signal integrity of an image stream in a car. Compact contrast distribution features are extracted by aggregating gradients using logarithmic stimulus strength modelling. Various distances were designed and selected with different aspects of quality and distortion in mind. They are combined with gradient boosted trees to the CD$^2$ method and make accurate predictions about image quality. Compared to state-of-the-art IQA methods CD$^2$ performs the best across IQA datasets.

References

[1] GB Arden. The importance of measuring contrast sensitivity in cases of visual disturbance. British Journal of Ophthalmology, 62(4):198–209, 1978.
[2] Simone Bianco, Luigi Celona, Paolo Napoletano, and Raimondo Schettini. On the use of deep learning for blind image quality assessment. *Signal, Image and Video Processing*, 12(2):355–362, 2018. 2

[3] Sebastian Bosse, Dominique Maniry, Thomas Wiegand, and Wojciech Samek. A deep neural network for image quality assessment. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3773–3777. IEEE, 2016. 7

[4] Guan-Hao Chen, Chun-Ling Yang, and Sheng-Li Xie. Gradient-based structural similarity for image quality assessment. In *2006 International Conference on Image Processing*, pages 2929–2932. IEEE, 2006. 1

[5] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. 2005. 2

[6] Samuel Dodge and Lina Karam. Understanding how image quality affects deep neural networks. In *2016 eighth international conference on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE, 2016. 1

[7] Samuel Dodge and Lina Karam. A study and comparison of human and deep learning recognition performance under visual distortions. In *2017 26th international conference on computer communication and networks (ICCCN)*, pages 1–7. IEEE, 2017. 1

[8] Chad Fogg, Didier J LeGall, Joan L Mitchell, and William B Pennebaker. *MPEG video compression standard*. Springer Science & Business Media, 2007. 5

[9] Selig Hecht. The visual discrimination of intensity and the weber-fechner law. *The Journal of general physiology*, 7(2):235–267, 1924. 3

[10] Jinggang Huang and David Mumford. Statistics of natural images and models. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR99149)*, volume 1, pages 541–547. IEEE, 1999. 7

[11] Le Kang, Peng Ye, Yi Li, and David Doermann. Convolutional neural networks for no-reference image quality assessment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1733–1740, 2014. 2

[12] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, pages 3146–3154, 2017. 2, 6, 7

[13] Eric Cooper Larson and Damon Michael Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006, 2010. 2, 3, 4, 6, 7, 8

[14] Qiang Li and Zhou Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE journal of selected topics in signal processing*, 3(2):202–211, 2009. 2, 7

[15] Lixiong Liu, Yi Hua, Qingjie Zhao, Hua Huang, and Alan Conrad Bovik. Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Processing: Image Communication*, 40:1–15, 2016. 7

[16] Xiongkuo Min, Ke Gu, Guangtao Zhai, Menghan Hu, and Xiaokang Yang. Saliency-induced reduced-reference quality index for natural scene and screen content images. *Signal Processing*, 145:127–136, 2018. 7

[17] Decebal Constantin Mocanu, Georgios Exarchakos, Haitham Bou Ammar, and Antonio Liotta. Reduced reference image quality assessment via boltzmann machines. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pages 1278–1281. IEEE, 2015. 7

[18] Anush Krishna Mourthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*, 20(12):3350–3364, 2011. 2

[19] Mariusz Oszust. Full-reference image quality assessment with linear combination of genetically selected quality measures. *PloS one*, 11(6):e0158333, 2016. 7

[20] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Image database tid2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 2015. 2, 3, 4, 6, 7, 8

[21] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE transactions on Image Processing*, 21(8):3339–3352, 2012. 2

[22] Mehul P Sampat, Zhou Wang, Shalini Gupta, Alan Conrad Bovik, and Mia K Markey. Complex wavelet structural similarity: A new image similarity index. *IEEE transactions on image processing*, 18(11):2385–2401, 2009. 1, 7

[23] Robert Shapley. The importance of contrast for the activity of single neurons, the vep and perception. *Vision research*, 26(1):45–61, 1986. 2

[24] Robert Shapley and R Clay Reid. Contrast and assimilation in the perception of brightness. *Proceedings of the National Academy of Sciences*, 82(17):5983–5986, 1985. 2

[25] HR Sheikh. Live image quality assessment database release 2. *http://live.ece.utexas.edu/research/quality*, 2005. 2, 3, 4, 6, 7, 8

[26] Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006. 7

[27] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[28] Rajiv Soundararajan and Alan C Bovik. Red indices: Reduced reference entropic differentiating for image quality assessment. *IEEE Transactions on Image Processing*, 21(2):517–526, 2011. 7

[29] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017. 1

[30] Dacheng Tao, Xuelong Li, Wen Lu, and Xinbo Gao. Reduced-reference iqa in contourlet domain. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(6):1623–1627, 2009. 2
[31] Gregory K Wallace. The jpeg still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992. 5

[32] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 1, 6, 7

[33] Zhou Wang and Eero P Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In *Human Vision and Electronic Imaging X*, volume 5666, pages 149–159. International Society for Optics and Photonics, 2005. 2, 7

[34] Dedy Rahman Wijaya, Riyanto Sarno, and Enny Zu-laika. Information quality ratio as a novel metric for mother wavelet selection. *Chemometrics and Intelligent Laboratory Systems*, 160:59–71, 2017. 4

[35] Jinjian Wu, Weisi Lin, Guangming Shi, Leida Li, and Yuming Fang. Orientation selectivity based visual pattern for reduced-reference image quality assessment. *Information Sciences*, 351:18–29, 2016. 7

[36] Sascha Xiaguang Xu and Jan Bauer. Verfahren und vorrichtung zur bertragung von bildern in einem fahrzeug, patent application, DE2019P01978, 2019. 7

[37] Wufeng Xue, Xuanqin Mou, Lei Zhang, Alan C Bovik, and Xiangchu Feng. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. *IEEE Transactions on Image Processing*, 23(11):4850–4862, 2014. 2

[38] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, 2013. 1, 3

[39] Lin Zhang, Ying Shen, and Hongyu Li. Vsi: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014. 6, 7

[40] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011. 1, 3, 7