SVD-based methods for image quality assessment

SoVividDay

December 19, 2020

1 Team

- **Emmanuel Akeweje**: HAAR and PSNR implementation
- **Yatipa Chaleenutthawut**: Structural-based approaches
- **Mohammed Deifallah**: Statistical-based approaches
- **Nikita Koritskiy**: SSVD implementation, dataset preprocessing
- **Suparat Srifa**: HVS-based approaches and presentation

2 Introduction

Image Quality Assessment or IQA is an important measurement to evaluate the quality of the modified/distorted visual signal especially in nowadays rapidly growth of digital communication. It is obviously that the best measurement of the quality is the opinion of human observers. However, it is well-known that such evaluations are complicated, expensive and time consuming.

This project aims to propose a full reference image quality metric based on the image's structural similarity called "Structural SVD-based image quality assessment (SSVD)" according to [1]. The proposed algorithm will be explained and the effectiveness with other IQA metrics comparison will be evaluated. The resource coding to reproduce the result is uploaded to GitHub repository.

3 Problem statement

Designing an IQA metric which predicts human judgments is a challenging issue. Various novel and modern methods are proposed to solve the challenges for an appropriate and a better quality metric to be applied in visual processing applications such as image/video coding, information hiding and visual enhancement. The purposed method will be compared to existing and well-known IQA metrics that can be classified into 3 groups idea which are structural, HVS and statistical-based IQA metrics.

3.1 Structural-based approaches

This approach will emphasise on structures and irregularities at the pixel level. Many difference type of distance transforms will be considered.[2]

3.1.1 FSIM

Feature-similarity (FSIM) index [3] has two importance feature to consider which are Phase congruency (PC) and Gradient magnitude (GC). First is the phase congruency (PC), which is a dimensionless quantity of the significance of a local structure map. We could consider PC_i in terms of vector at position x on scale n. Considering that PC is contrast invariant but no effect of contrast information the contrast information on HVS perception of image quality, Second feature is the image gradient magnitude (GM) which can be expressed by different convolution masks. PC and GM are complementary of each other in characterizing the image local quality.

The computation of FSIM contain two stages. The first stage, the local similarity map is computed, and then in the second stage, we pool the similarity map into a single similarity score. In order to compute similarity, We separate the feature similarity measurement between two images into two components f_i. We can compute FSIM by following equation where S_{PC} and S_G are similarity measure of PC and GC, respectively. We used $PC_m(\mathbf{x}) = \max(PC_1(\mathbf{x}), PC_2(\mathbf{x}))$ to weight the importance of similarity measure.

$$S_L(\mathbf{x}) = [S_{PC}(\mathbf{x})]^{\alpha} \cdot [S_G(\mathbf{x})]^{\beta} \tag{1}$$

$$\text{FSIM} = \frac{\sum_{\mathbf{x}\in\Omega} S_L(\mathbf{x}) \cdot PC_m(\mathbf{x})}{\sum_{\mathbf{x}\in\Omega} PC_m(\mathbf{x})} \tag{2}$$

where Ω is the whole image spatial domain.

The main advantage of FSIM is this method gives good prediction capability, higher accuracy, consistent and stable performance. In addition, FSIM are capable well in dealing with the distortions of denoising, quantization noise, and mean shift. In contrast, this method cannot work well with the distortions of masked noise and impulse noise.

3.1.2 MS-SSIM

Multi-scale structural similarity method (MS-SSIM) [4] is a perceptual metric that quantifies image quality degradation caused by data compression or by losses in data transmission in multi dimension. The overall SSIM evaluation is obtained by combining the measurement at different scales using below equation.

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l_M(\mathbf{x}, \mathbf{y})]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(\mathbf{x}, \mathbf{y})]^{\beta_j} [s_j(\mathbf{x}, \mathbf{y})]^{\gamma_j} \tag{3}$$

Low-pass filtering and down sampling are the two main operations used in this multi scale structure similarity method. The original and the distorted or noisy images are iteratively low-pass filtered and then down sampling will be done on that by factor of 2. For this multi scale operation, the original image is taken as scale1. The highest scale is for example scale M so a total of M-1iterations are taken place. In the SSIM method, three comparisons have been done i.e., contrast comparison $(c_j(x, y))$, luminance comparison and the structure comparison$(s_j(x, y))$, similar to that multi scale structure similarity also has three comparisons. The one comparison is performed on scale M is luminance comparison $(l_M(x, y))$. This process will be shown below. Other two comparisons are performed from scale 1 to M and after all these the final quality measurement metrics is the combination of these three comparisons.

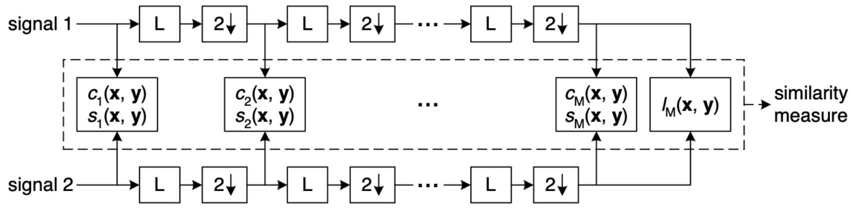

Figure 1: Multi-scale structural similarity measurement system. L: low-pass filtering; $2\downarrow$: downsampling by 2.

In terms of advantages, this method capable to deal with highly complex picture and non linear operation due to the fact that this method based on top-down approach. In addition, this method could generate a texture with statistics matching an original texture, and a human subject. This method is much better than normal single-scale approach (SSIM) owing to the fact that MS-SSIM provides more flexibility than SSIM in incorporating the variations of image resolution and viewing conditions.

3.2 HVS-based

Human Visual System or HVS, is a well-known method that references how human eyes would do. However, HVS measurement is very complex to be understood with psycho-physical means[5].

3.2.1 PSNR

Peak signal-to-noise ratio (PSNR) [5] is considered as the simplest and widely used full reference image quality measurement. The method is defined via mean squared error (MSE) and usually expressed in terms of the logarithmic decibel scale between two images.

Given a noise-free m×n monochrome image I and its noisy approximation K, MSE is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{4}$$

And the PSNR is defined as:

$$\begin{aligned} PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) &= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\ &= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE) \end{aligned} \tag{5}$$

where MAX_I is the maximum possible pixel value of the image and equals to 255 for an 8-bit image.

Although PSNR is useful in many task involving image processing, the main disadvantage is that it is poorly measurement compared to other quality metrics since it does not consider image structure.

3.2.2 HaarPSI

Haar wavelet-based Perceptual Similarity Index [6] is a method that measures similarity of images by perceptual similarity assessment between two images based on human observer. The basic idea of HaarPSI is that it expresses the perceptual similarity of two images in the interval [0,1]:

$$\ell^2\left(Z^2\right) \times \ell^2\left(Z^2\right) \rightarrow [0,1] \tag{6}$$

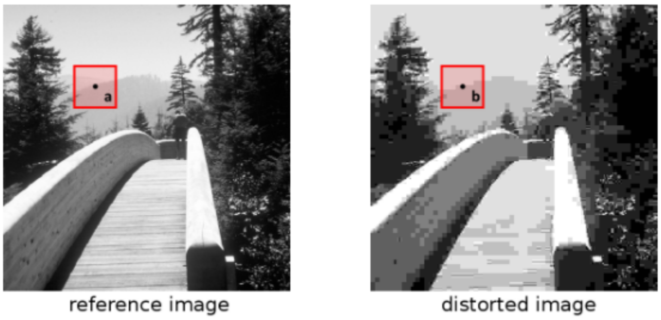

Figure 2: The HaarPSI of two images is based on the similarity of local features a and b

According to reference [7], it claims that HaarPSI could achieve higher correlations with human opinion scores on large benchmark databases in almost every distortion types and the computational is very efficient and quite faster that most other metrics. However, the reference also indicates that HaarPSI could perform lower efficiency if the original image consists of Gaussian blur because the method is mostly rely on on high-frequency information and maybe too sensitive when the distortions purely based on low-pass filtering.

3.3 Statistical-based

These metrics are based on natural scene statistics (NSS) purely based on information-theoretic adjustments. Images are treated as natural scenes, distinguished from sound, text cartoon, and anime, that have certain statistical properties. The following subsections shows a brief description of the two most well-known metrics from this category.

3.3.1 IFC

Information Fidelity Criterion (or IFC) is a Full-Reference (FR) QA methods that tries to remedy some of the limitations of other approaches by treating the images as signals with statistical properties. The figure below shows a high-level abstraction of the model:

Figure 3: The main components of IFC architecture

- The source model forms the wavelet coefficients of the image decomposition, which are used to conduct the Gaussian Scale Mixture (GSM) to model those coefficients.
- The distortion model makes use of the wavelet domain to generate a combination of additive noise and signal attenuations, that is independent of the source.
- The receiver model measures the amount of information extracted from the previously mentioned models.

The most prominent advantages of such a model is that:

- it's parameterless, which means it doesn't require any parameters or viewing configurations, training data, or stabilizing constants.
- signal gains are handled differently from additive noise components, so that signal attenuations aren't ignored unlike HVS-based methods.

However, it has some pitfalls which need more research in the future, such as:

- It lacks color statistics.
- Inter-subband correlations aren't efficiently utilized.
- It uses either GSM or generalized Gaussian density to predict the non-Gaussian marginal distribution of wavelets.
- The same similarity measures are used for different features although these features have different properties.

3.3.2 VIF

Visual Informatio Fidelity (or VIF) was developed in 2006 at the University of Texas at Austin. As the figure below shows, it has several common similarities with the previously mentioned IFC:

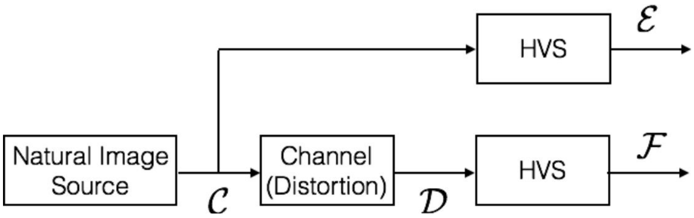

Figure 4: The main components of VIF architecture

It extends the architecture of the 1^{st} method (IFC) to make use of it. However, it's more numerically stable because it adds a normalizing denominator to bound the result from 0 to 1 instead of infinity. Moreover, it supports contrasting images by producing values more than 1.

3.4 Proposed Method

The basic idea of the proposed method is to create an algorithm that would rate images quality based on input image using SVD approach. SVD approach can be used toward IQA by utilized the difference between the singular values of the reference and the distorted images [8]. An image (A) can be decomposed into three matrices which are a diagonal singular value (S) and two orthonormal eigenvector matrices (U & V).

$$A_{m\times n} = \sum_{i=1}^{k} S_i U_i V_i^T, \quad k = \min(m,n) \tag{7}$$

The proposed method will be measured using TID2013 dataset[9] that contains the same fixed size 512×384 pixel images, and 3000 distorted images (25 test images with 24 types and 5 levels of distortions). Moreover, the method will be evaluated via Spearman rank order correlation coefficient (SROCC) and Kendall rank order correlation coefficient (KROCC) comparing with other IQA metrics mentioned above.

4 Algorithm

SSVD

Structured Singular value decomposition (SSVD) is proposed in the original paper[1]. The key idea is to use not only information about luminance, decoded in singular values, but also structural information that can be extracted from the singular vectors. However, it's computationally expensive to perform SVD for big number of image matrices. To face this problem, R-SVD[10] approach is used. In essence, singular values $\hat{S}_i$ and singular vectors $\hat{\mathbf{U}}_i$ of the distorted images are estimated using right singular vectors of the reference image $\mathbf{V}_i$ using formulas 8 and 9 correspondingly.

$$\hat{S}_i = \|\hat{\mathbf{A}}\mathbf{V}_i\| \tag{8}$$

$$\hat{\mathbf{U}}_i = \begin{cases} 0, & \textit{if } \hat{S}_i = 0 \\ \hat{\mathbf{A}}.\mathbf{V}_i/\hat{S}_i, & \textit{otherwise} \end{cases} \tag{9}$$

Algorithm in the proposed paper exploit this approach the following way. To get luminance factor, estimated singular values compared with the ones from the reference image S_i using formulas 10.

$$F^{LU} = \sqrt{\sum_{i=1}^{CPF} \left| \left(S_i - \hat{S}_{Ui} \right) \right| w_i}, \quad F^{LV} = \sqrt{\sum_{i=1}^{CPF} \left| \left(S_i - \hat{S}_{Vi} \right) \right| w_i} \tag{10}$$

LU and LV corresponds to the left and right SVD estimations, CPF is the threshold parameter since small singular values contribute insignificantly and $w = S_i / \sum_i S_i$. Final result is obtained by averaging: $F^L = \frac{F^{LU} + F^{LV}}{2}$

Structural factor utilizes estimated singular vectors. The main idea is that the more matrix $U\hat{U}^T$ differs from I, the higher structural distortion and hence the formulas 11.

$$F^{SU} = \sqrt{\sum_{i=1}^{CPF} \left[(SU_i - 1)\, w_i \right]^2}, \quad F^{SV} = \sqrt{\sum_{i=1}^{CPF} \left[(SV_i - 1)\, w_i \right]^2} \tag{11}$$

Authors also take into consideration the distortions that do not change structural information such as DC shift and contrast change. For this purpose the third, regulating factor is used. Basically, it's just comparison of estimated and honestly computed singular values of distorted image.

$$Q_{SU} = \sqrt{\sum_{i=1}^{CPF} \left(Sd_i - \hat{S}_{Ui} \right)^2}, \quad Q_{SV} = \sqrt{\sum_{i=1}^{CPF} \left(Sd_i - \hat{S}_{Vi} \right)^2} \tag{12}$$

Then this values are averaged using harmonic mean to work around large values using $Q_{RH} = \frac{2Q_{SU}Q_{SV}}{Q_{SU} + Q_{SV} + \varepsilon}$. Final result is obtained by $F^R = 1 - e^{(-Q_{RH}/h)}$

In algorithm distorted and reference images are divided into 9x9 squares and compared block by block. Resulting factors are multiplied and the product is averaged over all the blocks.

$$SSVD = \frac{\sum_{\text{i=1}}^{\text{\#blocks}} F_i^L F_i^S F_i^R}{\text{\#blocks}}$$

The closer the total factor to 0, the closer a distorted image to a referenced.

5 Experiments description and results

We implemented[11] an algorithm described in section 4 using Python3. Then we applied it to the images from TID2013[9] dataset, converting 3-channel RGB pictures to the grey scale. Obtained $F^l, F^R, F^S, SSVD$ factors were compared with MOC (Mean Opinion Score) – averaged subjective metric from the dataset. The results are presented on the Fig 5

The bottom right image is a major result one should consider. Monotonic dependence is a good sign that the proposed method is valid for applications. To measure the monotonicity Spearman Rank-Order Correlation Coefficient (SROCC) and Kendall Rank-Order Correlation Coefficient (KROCC) are used from `scipy.stats` module.

We calculated SROCC for each of 24 distortion types and compared it with the numbers from the paper. As seen from the Fig 6, our implementation is rather close to the author's.

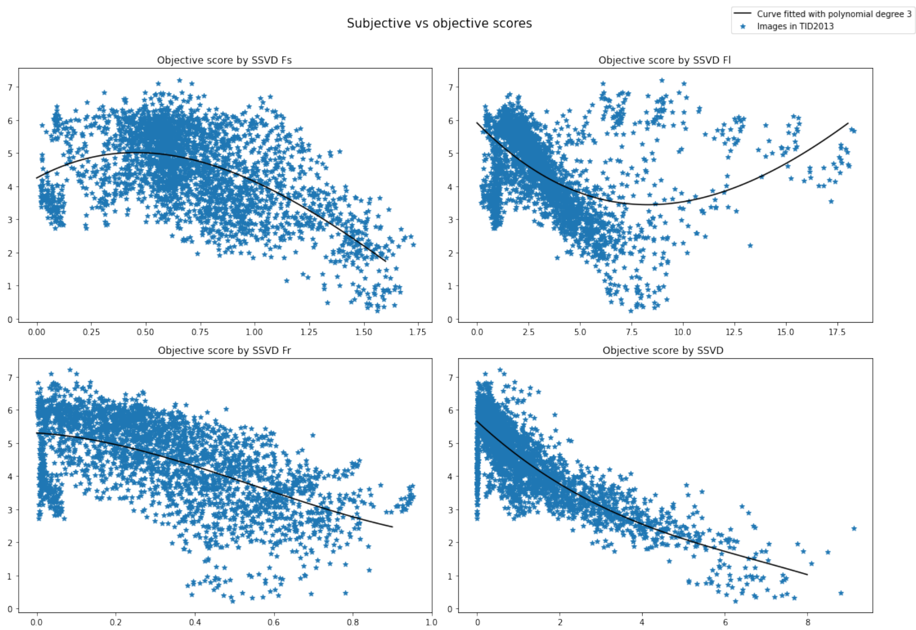

Figure 5: Comparison of the subjective MOC factor and objective factors from SSVD algorithm

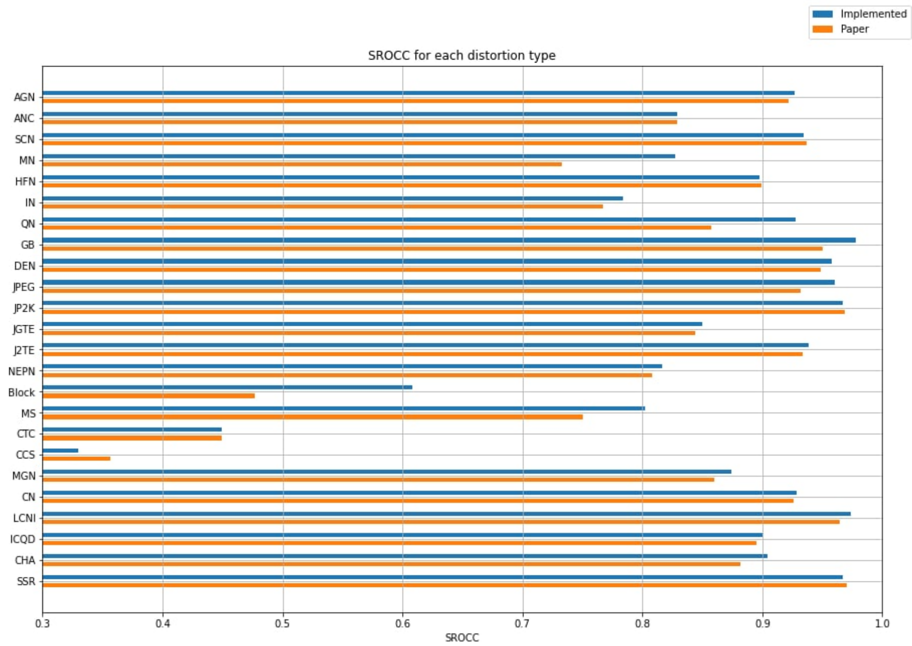

Figure 6: SROCC for each type of distortion for implemented algorithm and results from the paper

5.1 Performance Comparison

While we sought to create a basis for comparing the performance of SSVD on the TID2013 dataset, we implemented two of the IQA metrics already descibed above: Peak Signal Noise Ratio (PSNR) and Haar wavelet-based Perceptual Similarity Index (HaarPSI). Since we are majorly interested in the result produced by these metrics, we took advantage of the implementations already available and adapted them for our purpose. For PSNR, we used the specialized library `skimage.metrics` while for HaarPSI, an implementation was derived from the work of Reisenhofer et al. [7] The SROCC obtained for both metrics aligned almost exactly with the base paper as shown in Figures 7 and 8

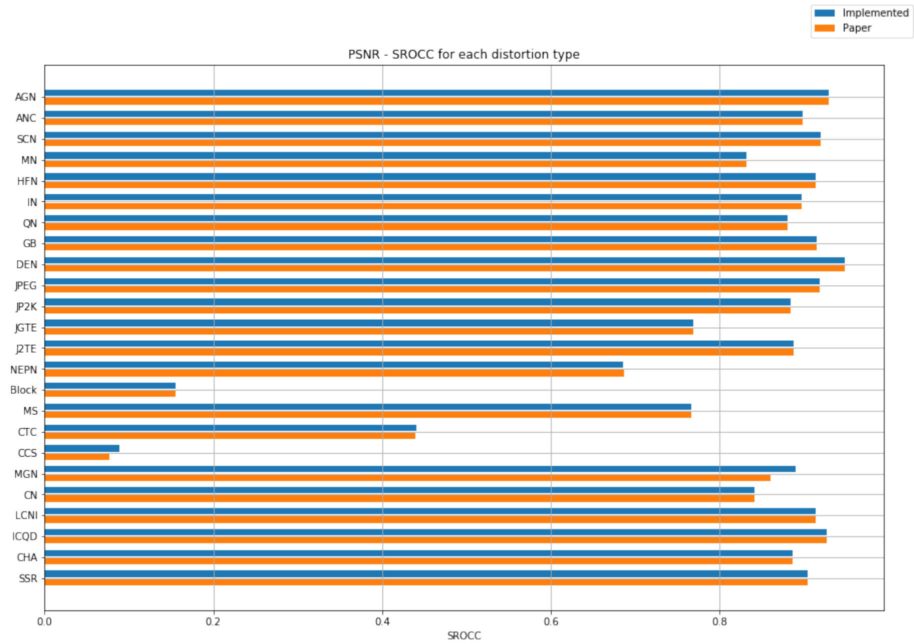

Figure 7: SROCC for each type of distortion for implemented PSNR and results from the paper

Tables 1 and 2 clearly display the difference in the performance of the three IQA metric showing that the Structural SVD has higher SROCC scores in 15 of 24 distortions available in the TID2013 dataset. The distortion type are AGN (Additive Gaussian noise), ANC (Additive noise in color components), SCN (Spatially correlated noise), MN (Masked noise), HFN (High frequency noise), IN (Impulse noise), QN (Quantization noise), GB (Gaussian blur), DEN (Image denoising), JPEG (JPEG compression), JP2K (JPEG2000 compression), JGTE (JPEG transmission errors), J2TE (JPEG2000 transmission errors), NEPN (Non eccentricity pattern noise), BLOCK (Local block-wise distortions of different intensity), MS (Mean shift (intensity shift), CTC (Contrast change), CCS (Change of color saturation), MGN (Multiplicative Gaussian noise), CN (Comfort noise), LCNI (Lossy compression of noisy images), ICQD (Image color quantization with dither), CHA (Chromatic aberrations), SSR (Sparse sampling and reconstruction), FF (Fat Fading), Lar (Locally Adaptive Resolution),

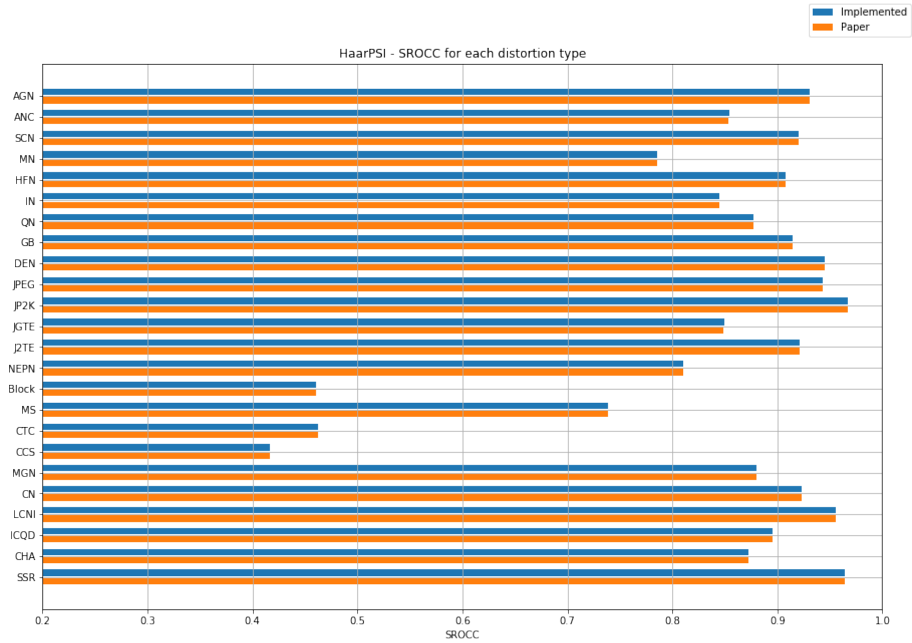

Figure 8: SROCC for each type of distortion for implemented HaarPSI and results from the paper

GCD (Global Contrast Decrements).

6 Conclusion

In this work multiple types of IQA metrics were considered. Structure, HVS and statistical-based are among them. 3 particular algorithms: PSNR, HaarPSI and SSVD were implemented. Their performance were tested on TID2013 dataset using SROCC and KROCC metrics. It was observed that SSVD showed better results for most distortion types.

However, due to limited time frame, there are gaps in this study that can be fulfilled in further works including: (1) providing more datasets into the analysis of SSVD method in order to attain more observation of the accuracy in difference datasets and distortion types; (2) compare the algorithm effectiveness with more IQA metircs.

Distortion Type	HaarPSI	PSNR	SSVD
AGN	**0.9304**	0.9292	0.9270
ANC	0.8540	**0.8983**	0.8291
SCN	0.9199	0.9199	**0.9348**
MN	0.7851	**0.8322**	0.8277
HFN	0.9075	**0.9141**	0.8979
IN	0.8447	**0.8968**	0.7836
QN	0.8773	0.8808	**0.9276**
GB	0.9149	0.9149	**0.9782**
DEN	0.9451	0.9480	**0.9579**
JPEG	0.9432	0.9189	**0.9606**
JP2K	**0.9674**	0.8840	0.9672
JGTE	0.8490	0.7685	**0.8498**
J2TE	0.9215	0.8883	**0.9388**
NEPN	0.8104	0.6860	**0.8166**
Block	0.4602	0.1552	**0.6079**
MS	0.7391	0.7671	**0.8025**
CTC	**0.4622**	0.4403	0.4494
CCS	**0.4166**	0.0885	0.3296
MGN	0.8804	**0.8905**	0.8742
CN	0.9230	0.8411	**0.9288**
LCNI	0.9560	0.9145	**0.9733**
ICQD	0.8958	**0.9269**	0.9004
CHA	0.8723	0.8873	**0.9046**
SSR	0.9643	0.9042	**0.9666**
Total hits	4	6	**14**

Table 1: Comparison of SROCC of IQA metrics for each distortion type.

	HaarPSI	PSNR	SSVD
SROCC	0.8093	0.6394	0.7916
KROCC	0.6372	0.4696	0.6211

Table 2: Performance of IQA metrics over the whole dataset.

References

[1] Azadeh Mansouri and Ahmad Mahmoudi-Aznaveh. SSVD: Structural SVD-based image quality assessment. *Signal Processing: Image Communication*, 74:54–63, May 2019.

[2] Jean-Luc Bouchot. Structures and irregularities in image processing and analysis. October 2012.

[3] Xuanqin Mou Lin Zhang, Lei Zhang and David Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20:2378 – 2386, August 2011.

[4] E.P. Simoncelli Z. Wang and A.C. Bovik. FSIM: Multiscale structural similarity for image quality assessment. *IEEE Transactions on Image Processing*, November 2003.

[5] Yusra Al-Najjar and Soong Der Chen. Comparison of image quality assessment: Psnr, hvs, ssim, uiqi. *International Journal of Scientific Engineering Research*, 3:1–5, 01 2012.

[6] Reisenhofer, Rafael and Bosse, Sebastian and Kutyniok, Gitta and Wiegand, Thomas. A haar wavelet-based perceptual similarity index for image quality assessment. https://http://www.haarpsi.org/, 2018. Retrieved 17-December-2020.

[7] Rafael Reisenhofer, Sebastian Bosse, Gitta Kutyniok, and Thomas Wiegand. A haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication*, 61:33–43, 2018.

[8] Aleksandr Shnayderman, Alexander Gusev, and Ahmet Eskicioglu. An svd-based grayscale image quality measure for local and global assessment. *Image Processing, IEEE Transactions on*, 15:422 – 429, 03 2006.

[9] Nikolay Ponomarenko, O. Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, J. Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. Color image database tid2013: Peculiarities and preliminary results. pages 106–111, 06 2013.

[10] Azadeh Mansouri, Ahmad Mahmoudi Aznaveh, Farah Torkamani-Azar, and J. Afshar Jahanshahi. Image quality assessment using the singular value decomposition theorem. *Optical Review*, 16(2):49–53, March 2009.

[11] *https://github.com/koritsky/ssvd*.