AN EFFECTIVE AND EFFICIENT VISUAL QUALITY INDEX BASED ON LOCAL EDGE GRADIENTS

Heinz Hofbauer and Andreas Uhl

Department of Computer Sciences
University of Salzburg
{hhofbaue, uhl}@cosy.sbg.ac.at

ABSTRACT

The structural similarity index measure is a well known and widely used full reference visual quality index. In this paper we introduce a new full reference visual quality index based on local edges and edge gradients in the wavelet domain. The proposed metric corresponds better to human judgement and is more efficient, in terms of computational complexity, than the structural similarity index measure. Furthermore, the proposed metric is more efficient than other state of the art metrics and surpasses them for certain visual impairment classes.

***Index Terms*—** Image analysis, Quality control

1. INTRODUCTION

The assessment of image and video quality is important for transmission (assessment of transmission errors) and compression (compression vs. quality). Optimally an evaluation where human observers judge the perceived quality or impairment should be performed. However, the time and cost requirements to perform such tests are high. Thus, algorithms to assess image quality automatically are employed, which are referred to as visual quality indices (VQI). These VQIs are in turn compared to human assessment over a number of distortion and compression types via established databases. It is important for VQIs to strongly correlate to human judgement. Furthermore, ease of use and low computational complexity are desired traits.

Still widely used the peak signal-to-noise ratio (PSNR) is unrivaled in speed and ease of use. However, it is also well known that the correlation to human judgement is somewhat lacking [1]. With this problem in mind newer VQIs were developed which take the human visual system (HVS) into account in order to increase the correlation with human judgement.

These VQIs utilize the knowledge of the HVS to a lesser or higher extent. However, the trend over all VQIs is the more information about the HVS is included in the generation of a quality score the more complex and time consuming the VQI becomes. This ranges from the fast luminance and edge similarity score (LSS and ESS) as introduced by Mao and Wu [2] to the refined but slow visual information fidelity criterion (VIF) by Sheikh and Bovik [3] and CPA1 by Carosi et al. [4]. In terms of HVS LSS and ESS uses the basic knowledge that edge information reflects the assessment of humans regarding the shape or contour of objects and the luminance score reflects changes in the color space. The VIF on the other hand uses a more refined model which starts with the modeling of the reference image using natural scene statistics (NSS). Furthermore, the possible distortion is modeled as signal gain and additive noise in the wavelet domain and parts of the HVS which have not been covered by the NSS are modeled, i.e. internal neural noise is modeled by using a additive white Gaussian noise model.

However, the most widely used VQI is the structural similarity index measure (SSIM) by Wang et al. [5] because it offers an excellent tradeoff between performance and quality. The SSIM extracts three separate scores from the image and combines them into the final score. First the visual influence is calculated locally then luminance, contrast and structural scores are calculated globally. These separate scores are then combined with equal weight to form the SSIM score.

In this paper we propose a VQI which is more efficient as well as more effective than the SSIM and for certain cases is even better than VQIs which are heavily modelled on the HVS. The proposed metric is a full reference metric, i.e. both original and impaired image are used, based on local edge direction and gradient which exploits properties of the HVS to some extent. Based on the utilization of local edge gradients we will refer to this metric as LEG for the rest of the paper.

Local edge information is a good feature when it comes to image comparison, classification or retrieval. Local binary patterns (LBP) have been widely employed to assess local edge information. LBPs compare a center pixel to its neighborhood to gain an edge description, and have been successfully used from face recognition, e.g. [6], to medical image classification, e.g. [7]. Usually these LBPs are used in the form of histograms in the comparison of images. In our case image based histograms do not provide enough error localization, thus a different approach of direct comparison is used.

Furthermore, we use a one-step wavelet decomposition

which allows the utilization of HVS based knowledge to increase the accuracy of error assessment. The decomposition serves a number of purposes. First the peripheral vision is taken into account by using the low frequency subband. The eye of an observer focuses on one point but also takes in surrounding information. Furthermore, the HVS is sensitive to different spatial and temporal frequencies which are present in a stimulus and this is modelled by using the high frequency subbands. Additionally, error information is perceived stronger among large edge structures in the image and the high frequency bands can be used for structure detection. In the proposed algorithm this structural information is used, by means of local edge gradients, as a weighting function for the local error features generated form the low frequency bands. Furthermore, the difference of overall luminance will be used to model the light sensitivity of the eye allowing to detect impairments which do not change the structure of the image.

In order to assess how the LEG performs in comparison to other metrics based on edge features the local feature based visual security metric (LFBVS) by Tong et al. [8] as well as the natural image contour evaluation (NICE) quality index by Rouse et al. [9] will be evaluated and compared to LEG. The LFBVS uses basic color features as well as edge amplitude and direction on local blocks. The NICE uses gradient maps on different scales, adjusts for possible image shift by using a morphological dilation with a plus shaped structuring element. The actual score is computed by doing a thresholding on the image and calculating differences. For the implementation of the NICE we choose the version with only one scale, Sobel edge detector and morphological dilation since it is significantly faster than using steerable wavelet decomposition but shows similar performance, c.f. [9]. For weighting the hamming distance and normalization as given in [10] was used.

The implementations of NICE, LFBVS, LEG and SSIM which are used in the following sections are available online at `http://www.wavelab.at/sources/VQI/`.

In section 2 the algorithm for the LEG VQI will be given. In section 3 we give the evaluation and analysis of the VQI with respect to LIVE, MICT and IVC image database as well as a comparison to other metrics. Section 4 concludes the paper.

2. ALGORITHM

Let I and O denote the impaired and original (gray scale) image of size $W \times H$ with maximum pixel value $M = 2^b$ with b bits per pixel. The following steps are performed to calculate the LEG index.

Step 1: The following steps only use edge difference, consequently a change in the image which does not influence the structural influence would go unnoticed. To compensate for this the difference in luminance between I and O is calculated:

$$\mathrm{lum}(I,O) = 1 - \sqrt{\frac{|\mu(O) - \mu(I)|}{M}},$$

$$\mu(X) = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H} X(x,y),$$

where $X(x,y)$ is the pixel value of image X at position x, y.

Step 2: One step wavelet decomposition with Haar wavelets resulting in four sub images for each image X denoted as X_0 for the LL-subband, and X_1, X_2, X_3 for LH, HH and HL subband respectively. Figure 1 illustrates the decomposition of a sample image, upper left is X_0 and the numbering continues clockwise.

Fig. 1: Wavelet decomposition of the lighthouse2 image with the Haar wavelet.

Step 3:A local edge map is calculated for each position x, y in the image, reflecting the change in coarse structure of the image. Through the Haar wavelet decomposition and the comparison of each center pixel with its neighborhood the actual area of influence of the original image is a 6×6 window. The local edge map is used to prevent faulty information gained from gradients on parts of the image where the edges are off by a certain degree.

$$\mathrm{le}(I,O,x,y) = \begin{cases} 1 & \text{if } EDC(I,O,x,y) = 8, \\ 0.5 & \text{if } EDC(I,O,x,y) = 7, \\ 0 & \text{otherwise.} \end{cases}$$

$$EDC(I,O,x,y) = \sum_{p \in N(x,y)} ED(I,O,x,y,p)$$

$$ED(I,O,x,y,p) = \begin{cases} 1 & \text{if } I(x,y) < I(p) \text{ and } O(x,y) < O(p), \\ 1 & \text{if } I(x,y) > I(p) \text{ and } O(x,y) > O(p), \\ 0 & \text{otherwise.} \end{cases}$$

where $N(x,y)$ is the eight neighborhood of the pixel x, y. Edge extension is done by copying the last edge value, e.g. $I(-1,0) := I(0,0)$.

Step 4: In order to assess the contrast changes a difference of gradients in a neighborhood will be calculated. The effect of a contrast change will be more pronounced along large edges. Thus, this step serves as a contrast sensitivity function as well as a detector for large structures in the image.

$$\mathrm{led}(I,O,x,y) = \frac{1}{8}\sum_{p\in N(x,y)}\left(1-\sqrt{\frac{|LD(I,O,x,y,p)|}{M}}\right)^2$$

$$LD(I,O,x,y,p) = (O(x,y)-O(p)) - (I(x,y)-I(p))$$

Step 5: The edge score is calculated by using local edge conformity (le) and local edge difference (led).

$$\mathrm{es}(I,O) = \frac{4}{WH}\sum_{x=1}^{\frac{W}{2}}\sum_{y=1}^{\frac{H}{2}}\Big(\mathrm{le}(I_0,O_0,x,y) * \frac{1}{3}\sum_{i=1}^{3}\mathrm{led}(I_i,O_i,x,y)\Big).$$

Step 6: The LEG visual quality index is calculated by combining es and lum:

$$\mathrm{LEG}(I,O) = \mathrm{lum}(I,O)\,\mathrm{es}(I,O).$$

An example of the generated edge score and edge conformity is given in fig. 2, where the original image O (lower left) as well as an impaired image I (top right) is given in low and high quality. The top left illustrates the values of $\mathrm{le}(I,O)$ and the lower right gives the average of the led values as used in **step 5**, i.e. $\frac{1}{3}\sum_{i=1}^{3}\mathrm{led}(I_i,O_i,x,y)$. The values have been mapped onto a continuous gray scale, $[0,1]\mapsto[0,255]$, where black indicates a high amount of errors and white represents no distortion between I and O.

2.1. Analysis of Weight Functions

The features used in the metric are average luminance (μ), edge direction conformity (EDC) and contrast change difference (LD) which are combined with weight functions in order to model their impact on the HVS. For example, should a small difference in μ have a huge impact, super linear weight, or a small impact, sub linear weight. In order to find the best weight function for each feature the individual weight functions were tested on the IVC database.

It turned out that super linear weight performed better, indicating that the HVS is able to detect even small aberrations from the original. For this reason we tested various functions, as depicted in fig. 3, with varying degrees of superlinearity: different slope for linear with capping at the boundaries, as used in le for EDC; logarithmic or square root, again with different forms of superlinearity, e.g., normal square root as used in lum for μ or squared as used in led for LD.

(a) High Quality

(b) Low Quality

Fig. 2: Comparison of high and low quality error maps. The original image is in the lower left corner, the impaired version is in the top right corner, the edge error map is top left and the difference of gradient error map is lower right.

Table 1 shows results for this test, for reasons of brevity only some results are presented. The table gives the Spearman rank order correlation (SROC) for a given function on each testset of the IVC database as well as the summation of the SROC scores for overall comparison. The test uses the final weight functions for all features except for the feature indicated in the table for which the given weight function is evaluated. The weight functions indicated in the table are those illustrated in fig. 3.

The trends towards the selected weight functions as used in the final version of LEG is clearly visible.

3. EXPERIMENTAL RESULT AND ANALYSIS

In order to evaluate the LEG VQI we compare it to the widely used SSIM. Furthermore, a comparison to slow VQIs which highly exploit the HVS, such as VIF and CPA1, is included. In addition a comparison to fast VQIs, PSNR and LSS, which rely less on the HVS are performed. Furthermore, a com-

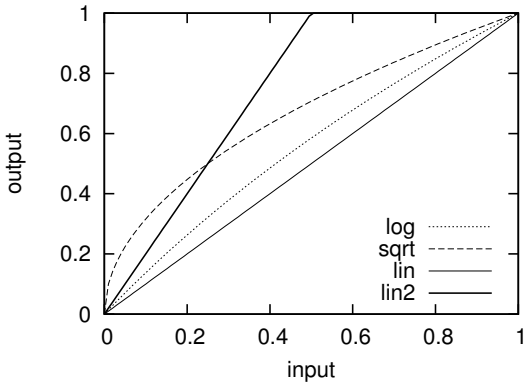

Fig. 3: Different forms of weight functions, log is actually calculated as $log = \log(1+x)/\log(2)$ to compensate for the singularity at 0 and lin2 is $\min(2x, 1)$.

		Testset					
feature	weight	lar	j2000	flou	jpeg	lumichr	sum
LEG		0.8504	0.8837	0.9692	0.8962	0.9392	4.5386
μ	lin	0.8535	0.8708	0.9669	0.8942	0.9473	4.5327
μ	log	0.8535	0.8708	0.9669	0.8942	0.9473	4.5327
LE	sqrt	0.7448	0.8361	0.9263	0.8015	0.6985	4.0072
LE	log	0.8038	0.8724	0.9429	0.8375	0.8073	4.2639
LE	lin	0.8087	0.8871	0.9519	0.8551	0.8615	4.3643
LE	lin2	0.8390	0.8918	0.9692	0.8909	0.9338	4.5248

Table 1: Result of the test of a selected number of weight functions and features on the IVC database.

parison with the NICE and LFBVS, which use similar image features as LEG, is included.

The comparison is done based of the LIVE (release 2) [11], MICT [12] and IVC [13] databases. Three different image databases were used because a single database can be biased due to different evaluation methodology, number of participants, demography, expertise of participants etc. An example of this are the JPEG and JPEG 2000 compression testsets which are contained in each of the three databases. Throughout this section the absolute value of the Spearman rank order correlation (SROC) is given. We also used the Kendall tau (τ) measure for evaluation and both measures coincide.

The SROC for LIVE is given in table 2a, for MICT table 2b lists the results and for IVC the results are given in table 2c, where each table shows separate entries per testset. Likewise the results for τ are given in table 4. The main focus of the comparison is SSIM versus LEG and the higher correlation per testset is given in bold for both SROC and τ for easier reference. Table 3 shows the SROC for NICE and LFBVS in comparison to LEG and SSIM, for brevity reasons no Kendall τ results are given for this comparison.

Overall the LEG outperforms the SSIM except for two

(a) LIVE

testset	LSS	PSNR	SSIM	LEG	VIF	CPA1
fastfading	0.843	0.891	0.942	**0.971**	0.965	0.881
gblur	0.916	0.782	0.903	**0.966**	0.972	0.927
jp2k	0.953	0.895	0.936	**0.945**	0.968	0.958
jpeg	0.970	0.881	0.946	**0.960**	0.984	0.962
wn	0.965	0.985	**0.962**	0.960	0.985	0.984

(b) MICT

testset	LSS	PSNR	SSIM	LEG	VIF	CPA1
jpeg	0.839	0.285	0.631	**0.938**	0.907	0.725
jpeg2000	0.908	0.860	**0.915**	0.914	0.956	0.923

(c) IVC

testset	LSS	PSNR	SSIM	LEG	VIF	CPA1
flou	0.889	0.805	0.869	**0.969**	0.973	0.902
j2000	0.938	0.850	0.851	**0.884**	0.936	0.927
jpeg	0.952	0.674	0.805	**0.896**	0.924	0.906
lumichr	0.914	0.563	0.749	**0.939**	0.878	0.834
lar	0.863	0.699	0.711	**0.850**	0.888	0.881

Table 2: Absolute Spearman rank order correlation (SROC) for LSS, PSNR, SSIM, LEG, VIF, CPA1 on the IVC, MICT and LIVE databases.

cases. For the white noise testset on the LIVE database the SSIM reaches a SROC of 0.96151 while the LEG only scores 0.96000, a negligible difference given that the performance over the other testsets is significantly higher. The other exception is the JPEG 2000 testset on the MICT database where again the SSIM outperforms the LEG by a small margin. However, for the JPEG 2000 testsets on both LIVE and IVC the LEG performs better. This fluctuation is most likely due to the, relatively low, number of human observations which are used as ground truth.

When it comes to class of similar metrics we see that LEG outperforms NICE on all three databases and all testsets. LFBVS on the other hand is slightly better than LEG on the IVC lar testset and significantly better than LEG, and all other metrics including VIF and CPA1, on the IVC lumichr testset.

For some testsets the LEG even surpasses the VIF and CPA1 VQIs. On average the CPA1 is surpassed by LEG although only by a small margin, the LEG is on average surpassed by the VIF with an equally small margin. For metrics which use similar features the LEG on average outperforms both LFBVS and NICE. The average difference of a given VQI to the LEG is shown in table 5, and table 6 for NICE and LFBVS, for the SROC on a per database basis. Positive results indicate a VQI which is on average better than the LEG on the given database.

(a) LIVE

testset	SSIM	LFBVS	NICE	LEG
fastfading	0.942	0.932	0.959	**0.971**
gblur	0.903	0.937	0.949	**0.966**
jp2k	0.936	0.853	0.925	**0.945**
jpeg	0.946	0.920	0.950	**0.960**
wn	**0.962**	0.891	0.910	0.960

(b) MICT

testset	SSIM	LFBVS	NICE	LEG
jpeg	0.631	0.814	0.861	**0.938**
jpeg2000	**0.915**	0.584	0.881	0.914

(c) IVC

testset	SSIM	LFBVS	NICE	LEG
flou	0.869	0.955	0.924	**0.969**
j2000	0.851	0.825	0.882	**0.884**
jpeg	0.805	0.889	0.860	**0.896**
lumichr	0.749	**0.969**	0.809	0.939
lar	0.711	**0.853**	0.795	0.850

Table 3: Absolute Spearman rank order correlation for SSIM, LFBVS, NICE, LEG on IVC, MICT and LIVE databases.

(a) LIVE

testset	LSS	PSNR	SSIM	LEG	VIF	CPA1
fastfading	0.668	0.707	0.784	**0.848**	0.841	0.698
gblur	0.751	0.584	0.726	**0.841**	0.858	0.761
jp2k	0.811	0.711	0.771	**0.785**	0.843	0.817
jpeg	0.849	0.691	0.795	**0.822**	0.892	0.827
wn	0.835	0.894	**0.832**	0.824	0.894	0.888

(b) MICT

testset	LSS	PSNR	SSIM	LEG	VIF	CPA1
jpeg	0.644	0.199	0.446	**0.791**	0.737	0.524
jpeg2000	0.742	0.682	**0.752**	0.747	0.821	0.764

(c) IVC

testset	LSS	PSNR	SSIM	LEG	VIF	CPA1
flou	0.741	0.667	0.709	**0.878**	0.899	0.741
j2000	0.803	0.726	0.693	**0.705**	0.790	0.784
jpeg	0.822	0.519	0.627	**0.719**	0.791	0.750
lumichr	0.752	0.442	0.563	**0.813**	0.718	0.631
lar	0.659	0.571	0.571	**0.656**	0.713	0.705

Table 4: Absolute Kendall tau (τ) for LSS, PSNR, SSIM, LEG, VIF, CPA1 on the IVC, MICT and LIVE databases.

database	LSS	PSNR	SSIM	VIF	CPA1
LIVE	-0.031	-0.073	-0.023	0.015	-0.018
MICT	-0.053	-0.354	-0.153	0.005	-0.102
IVC	0.003	-0.189	-0.111	0.012	-0.018

Table 5: Comparison of average SROC per database in relation to LEG.

database	SSIM	LFBVS	NICE
LIVE	-0.023	-0.053	-0.022
MICT	-0.153	-0.227	-0.055
IVC	-0.111	-0.010	-0.054

Table 6: Comparison of average SROC per database in relation to LEG.

3.1. Runtime Efficiency Analysis

For an efficiency analysis the LEG metric is compared to the SSIM, LSS, PSNR, NICE and LFBVS. The CPA1 and VIF are not included since they are only available as matlab implementations and the matlab overhead would unfairly influence the results. However, the CPA1 uses a full frame Fourier transformation which is of complexity $\mathcal{O}(N \log N)$ while the LEG is $\mathcal{O}(N)$. And for the VIF, Sheikh et al. [3] evaluated that the VIF is 6.5 times slower than the MS-SSIM. The MS-SSIM is a multi scale variant of the SSIM utilizing wavelet decompositions and repeat calculations of the SSIM for different subbands, thus the VIF is at least 6.5 times slower than the SSIM.

As testset the 779 impaired images from the LIVE database were used. The results of the comparison are given in table 7 where it is shown that LEG is faster than the SSIM for best, worst and average case but can not match the efficient simplicity of LSS or PSNR. For the average case LEG is 6 times faster than the SSIM.

For the edge based metrics, NICE, LFBVS and LEG, the performance is similar with the LEG outperforming LFBVS and NICE by a factor of 1.46 and 1.22 respectively.

metric	$\bar{t}$	t_{min}	t_{max}
SSIM	764 ms	588 ms	882 ms
LFBVS	186 ms	140 ms	247 ms
NICE	155 ms	107 ms	193 ms
LEG	127 ms	94 ms	198 ms
LSS	52 ms	40 ms	83 ms
PSNR	55 ms	43 ms	68 ms

Table 7: Runtime performance of the given metrics over 779 impaired images of the LIVE database.

4. CONCLUSION

We have proposed a novel visual quality index based on local edge features augmented by knowledge about the human visual system. We have given the algorithmic details of a new visual quality index, LEG, and have evaluated it in terms of runtime efficiency and correlation with human judgement.

It was shown that the proposed VQI is more effective, i.e., corresponds better to human judgement, than the SSIM and is superior in runtime efficiency. Furthermore, on average the proposed metric is more effective than the CPA1 and has a far superior runtime efficiency than either CPA1 or VIF.

The LEG also has a higher runtime efficiency than NICE and LFBVS, which use image features similar to those used in LEG. Furthermore, LEG is more effective than NICE and on average more effective than LFBVS.

5. REFERENCES

[1] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, June 2008.

[2] Y. Mao and M. Wu, "Security evaluation for communication-friendly encryption of multimedia," in *Proceedings of the IEEE International Conference on Image Processing (ICIP'04)*, Singapore, Oct. 2004, IEEE Signal Processing Society.

[3] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, May 2006.

[4] Maurizio Carosi, Vinod Pankajakshan, and Florent Autrusseau, "Towards a simplified perceptual quality metric for watermarking applications," in *Proceedings of SPIE, Multimedia on Mobile Devices*, San Jose, CA, USA, Jan. 2010, vol. 7542, SPIE.

[5] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[6] Shengcai Liao, Xiangxin Zhu, Zhen Lei, Lun Zhang, and Stan Li, "Learning multi-scale block local binary patterns for face recognition," in *Advances in Biometrics*, pp. 828–837. Springer, 2007.

[7] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Pit pattern classification using extended local binary patterns," in *Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB'09)*, Larnaca, Cyprus, Nov. 2009.

[8] Lingling Tong, Feng Dai, Yongdong Zhang, and Jintao Li, "Visual security evaluation for video encryption," in *Proceedings of the international conference on Multimedia*, New York, NY, USA, 2010, MM '10, pp. 835–838, ACM.

[9] S. S. Hemami D. Rouse, "Natural image utility assessment using image contours," in *IEEE International Conference on Image Processing (ICIP'09)*, Cairo, Egypt, Nov. 2009, pp. 2217–2220.

[10] S. S. Hemami D. Rouse, "The role of edge information to estimate the perceived utility of natural images," in *Western New York Image Processing Workshop (WNYIP)*, Rochester, NY, Sept. 2009, p. 4 pp.

[11] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2," http://live.ece.utexas.edu/research/quality.

[12] Yuukou Horita, Keiji Shibata, and Yoshikazu Kawayoke, "MICT image quality evaluation database," http://mict.eng.u-toyama.ac.jp/mictdb.html.

[13] Patrick Le Callet and Florent Autrusseau, "Subjective quality assessment IRCCyN/IVC database," 2005, http://www.irccyn.ec-nantes.fr/ivcdb/.