Building a Reduced Reference Video Quality Metric with Very Low Overhead using Multivariate Data Analysis

Tobias OELBAUM and Klaus DIEPOLD
Institute for Data Processing
Technische Universität München
Munich, Germany

***Abstract*—In this contribution a reduced reference video quality metric for AVC/H.264 is proposed that needs only a very low overhead (not more than two bytes per sequence). This reduced reference metric uses well established algorithms to measure objective features of the video such as 'blur' or 'blocking'. Those measurements are then combined into a single measurement for the overall video quality. The weights of the single features and the combination of those are determined using methods provided by multivariate data analysis. The proposed metric is verified using a data set of AVC/H.264 encoded videos and the corresponding results of a carefully designed and conducted subjective evaluation. Results show that the proposed reduced reference metric not only outperforms standard PSNR but also two well known full reference metrics.**

***Index Terms*—video quality metric, reduced reference, multivariate data analysis, AVC/H.264.**

1. INTRODUCTION

Knowing the visual quality of an encoded video is essential for next to all applications dealing with digital video. This task can be accomplished either by a very time consuming (but accurate) subjective test or by an objective video quality metric. A good objective video quality metric would produce results highly correlated to those obtained by a subjective test. Four years after the first version of the upcoming video coding standard AVC/H.264 [1] was released, next to no results exist to demonstrate the prediction capabilities of video quality metrics for AVC/H.264 encoded video data. Up to now most video quality metrics have been verified using MPEG-2 encoded videos, but as AVC/H.264 encoded video has significant different characteristics (e.g. no fixed block sizes, filtering in the decoder loop), those results do not necessarily apply for this new generation of encoded video.

Being the de-facto standard for objective video quality metrics PSNR is still used for comparing AVC/H.264 encoded video with other video codecs or for comparing different encoder implementations or coding settings for AVC/H.264. This is in spite of the knowledge, that PSNR values may be heavily misleading [2].

The rest of the contribution is organized as follows: In section 2 a short overview about related works is presented. The method used to develop the proposed reduced reference metric is described in section 3 and the model itself is described in detail in section 4. Section 5 presents the results for the proposed method and finally section 6 concludes this contribution.

2. RELATED WORKS

Full Reference (FR) Quality Metrics

The most popular video quality metric is the Peak Signal to Noise Ratio (PSNR). This simple metric just calculates the mathematical difference between every pixel of the encoded video and the original video. In fact up to now PSNR is the only video quality metric that is widely accepted and therefore PSNR is the de-facto standard for measuring video quality.

In 2004 the ITU released a recommendation which included four different full reference (not only the coded video but also the original video is needed for the evaluation) metrics which outperformed PSNR in terms of correlation to results of extensive subjective tests [3]. Among those is the Edge PSNR [4] method developed by Lee et al. which was chosen as a comparison point to the metric presented in this contribution. This metric is based on the observation that human observers are especially sensitive to degradations in regions around edges. Therefore this metric evaluates the PSNR only at those pixels that have been classified to belong to an edge region. One FR image metric which has gained a high popularity since it was introduced in 2002 is the so called SSIM (Structural SIMilarity index) as presented in [5] and [6]. This metric was the second metric chosen for comparison. The SSIM performs a separate comparison on luminance, contrast and structure in the original and the coded image and uses this information to calculate one overall quality index. Among the quite extensive list of FR image and video quality metrics four recently proposed metrics are [7], [8], [9] and [10]. Shnayderman et al propose to use a singular value decomposition representation of the original and the coded image for the comparison [7]. In [8] Zhai et al. propose to decompose the images using a fiterbank of 2D Garbor filters and then measures the correlation between the decompositions of the original and the coded image. This attempt is inspired by the fact, that simple cell in the visual cortex can be modeled as 2D Garbor functions. The same authors extend the SSIM by a previous multiscale decomposition of an image in [9]. Finally Ong et al. propose a full reference metric that combines a measurement for block fidelity and content richness with some visibility masking functions [10].

Reduced Reference and No Reference Quality Metrics

Comparably few approaches were presented for reduced reference (RR) quality evaluation and even less for no reference (NR) quality evaluation. Compared to full reference metrics for a RR metric only parts of the original video or some extracted properties of the original video are needed for evaluation. 2003 Carnec et al. presented a RR image quality metric that is based on properties of the human visual system (HVS) such as color perception and masking effects. For the RR approach only high level visual information is needed, however an additional data load of 1056 real numbers for every frame is reported [11]. Kusuma et al. presented a RR approach for still images using a linear combination of blocking, blur, ringing, masking and lost blocks in 2005 [12]. Wang and Simoncelli showed that natural images have a certain frequency distribution and therefore the frequency distribution of a coded image can be used to predict the visual quality [13]. To transmit a representation of the frequency distribution of the original image they require 162 bits.

For a NR metric no information about the original video is needed. One popular approach for a NR image and video quality metric is the inclusion of watermarks in the original image and then measuring the amount to which these watermarks can be recovered at the receiver [14]. An approach similar to the metric proposed here was presented just recently by Callet et al in [15]. While the set of objective features used in both metrics is quite similar, the main differences can be found in the way this objective features are then combined to an overall video quality metric. For this task Callet et al use a neural network approach to simulate the human visual system (HVS) and its reaction to a set of features.

In addition to complete quality metrics there exist several measurements that concentrate on one single image property or a special artifact. Prominent candidates from this field are the blocking measurement by Bovik and Wang [16], or the blur measurement proposed by Winkler [17].

3. Multivariate Data Analysis for Objective Video Quality Assessment

Video quality metrics that try to model the HVS face the problem, that what they want to model is very complicated and up to the moment not well understood. Measuring the strength of a certain artifact (e.g. blocking, blur) and trying to predict the quality by a linear combination of the measured artifacts introduces the problem, that it is not known to which extend a certain artifact affects the perceived video quality. In addition this method ignores the possibility that there may be interference between certain types of artifacts.

For these two reasons it is proposed to build new video quality models using methods provided by multivariate data analysis. Multivariate data analysis is a tool that is widely used in chemo metrics and food science where the aim is to find the value of a latent variable (e.g. taste) by measuring some fixed variables (e.g. sugar, milk, cocoa). For the field of video quality assessment this translates to measure the latent variable video quality by measuring fixed variables such as blocking, blur, activity, continuity or noise. The main advantages of this approach are twofold:

- The HVS is regarded as black box model and no attempt to model any aspect of the HVS is needed. The 'black box' has some inputs (the fixed variables) and hopefully the output (visual quality) can be somehow predicted using these input values.
- The approach requires no previous assumptions about the influence of the fixed variables.

Feature Selection

A set of simple no reference feature measurements was selected representing the most common kind of distortions namely blocking, blurriness and noise. One feature measurement was added to measure the amount of detail present in the encoded video. To take into account the time dimension of

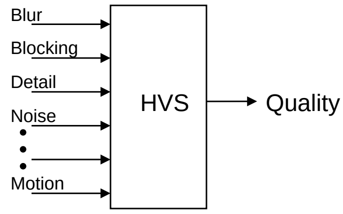

Fig. 1. Black Box HVS Model

video four different continuity measurements were performed: predictability (shows how good one frame can be predicted using the previous frame only), motion continuity (measurement for the smoothness of the motion), color continuity (shows how much color changes between two successive images) and edge continuity (shows how much edge regions are changing between two successive images).

- Blur: the blur measurement used is described in [17]. The algorithm measures the width of an edge and then calculates the blur by assuming that blur is reflected by wide edges. As blur is something natural in a fast moving sequence this measurement is adjusted if the video contains a high amount of fast motion.
- Blocking: for measuring the blockiness the algorithm introduced in [16] is used. This algorithm calculates the blockiness by applying a FFT along each line or column. The unwanted blockiness can be easily detected by the location in the spectra.
- Noise: to detect the noise present in the video a very simple noise detector was designed. First a prediction of the actual image is built by motion compensation using a simple block matching algorithm. Second a difference image between the actual image and its prediction is calculated and a low pass version of this difference image is produced by first applying a median filter and a Gaussian low pass filter afterward. A pixel is classified to contain noise if the difference value between the original difference image and the low pass difference image exceeds a threshold of 25 (assuming 8 bit values ranging from 0 to 255) for one of the three color planes.
- Details: to measure the amount of details that are present in a video the percentage of turning points along each line and each row are calculated. This measurement is part of a BTFR metric included in [3]. As the amount of details that are noticed by an observer decreases with increasing motion the activity measurement is adjusted if high motion is detected in the video.
- Predictability: A predicted image is built by motion compensation using a simple block matching algorithm. The actual image and its prediction are then compared block by block. A 8×8 block is considered to be noticeable different if the SAD exceeds 384. To avoid that single pixels dominate the SAD measurement both images are filtered using first a Gaussian blur filter and a median filtering afterward.
- Edge Continuity: The actual image and its motion compensated prediction are compared using the Edge-PSNR algorithm as described in [4].
- Motion Continuity: Two motion vector fields are calculated: between the current and the previous frame and between

- the following and the current frame. The percentage of motion vectors where the difference between the two corresponding motion vectors exceeds 5 pixels (either in x- or y-direction) determines the motion continuity.
- Color continuity: A color histogram with 51 bins for each RGB channel is calculated for the actual image and its prediction. Color continuity is then given as the linear correlation between those two histograms.

All feature measurements are done for each frame of the video separately and the mean value of all frames is then used for further processing.

The above selected measurements are just one example for a set of variables that are used for building such a model. The presented variables were used for their simplicity, using more complex measurements for artifacts like noise or blur may result in even more accurate models as well as adding measurements for artifacts not considered here (e.g. ringing). For this case only no reference feature measurements are considered, including some feature measurements that require the original video a RR or FR metric could be built. The nature of the multivariate calibration allows including an unrestricted number of fixed variables in the calibration step. If the calibration phase is done properly, fixed variables that do not contribute to the latent variable 'video quality' do not spoil the calibration process. The regression model will contain these useless fixed variables with zero (or very close to zero) weight and those variables then can be removed from the model.

The calibration step

Multivariate calibration is the method of learning to interpret a number of k input sensory signals that contribute to a common output y. For the presented metric the input signals are the above mentioned feature measurements while the output would be the visual quality of the video. The data set used for calibrating (training) the model consisted of four different standard video test sequences (Bus, Football, Harbour, Mobile) at CIF resolution that were encoded according to AVC/H.264 at three (Bus, Harbour) and seven (Football, Mobile) different bit rates ranging from 96 kbps to 1024 kbps and with a frame rate of 15 or 30 fps. Different encoder settings concerning the number of B-Frames that were inserted (zero to two B-Frames), or the I-Frame periodicity (only one I-Frame or periodic I-Frames) were used. For each of the l calibration sequences the selected feature values f_{mi} ($m \in \{1 \cdots k\}, i \in \{1 \cdots l\}$) were computed, for reference the $l \times k$ matrix containing the feature values is denoted as $\mathbf{F}$.

Correcting the Features using MSC: As it is expected that the measured features are not free from multiplicative or additive effects (e.g. the measurement for noise may be correlated and affected by the amount of details present in the video) a multiplicative signal correction (MSC) step is performed before starting the multivariate regression. MSC was originally developed to correct measurements in reflectance spectroscopy, but can also help in this context to remove multiplicative and additive effects between different objective features. The MSC corrected value of one feature m for one sequence i is calculated as following:

$$f'_{mi} = c + f_{mi} * d$$

The two variables c and d are obtained by simple linear regression of the feature values of the sequence i compared to the average of the feature values of all calibration sequences. For a detailed description of MSC see chapter 7.4 in [18]. Consequently the matrix $\mathbf{F}$ becomes $\mathbf{F}'$ after MSC treatment.

Multivariate Regression with PLS: The obtained feature values f'_{mi} are then used together with the corresponding subjective ratings y_i that form the column vector $\mathbf{y}$ to built a regression model using the method of Partial Least Squares Regression (PLSR). PLSR is an extension of the Principal Component Regression (PCR), that tries to find the principal components (PC) that are most relevant not only for the interpretation of the variation in the input values in $\mathbf{F}$ but also for the variation in the output values $\mathbf{y}$. So while the PCR is a bilinear regression method that consists of a Principal Component Analysis (PCA) of $\mathbf{F}'$ into the matrix $\mathbf{T}$ that contains the PCs of $\mathbf{F}'$ followed by a regression of $\mathbf{y}$ on $\mathbf{T}$, for the PLSR the modeling of $\mathbf{F}'$ and $\mathbf{y}$ is done simultaneously to ensure that the PCs gained from $\mathbf{F}'$ are relevant for $\mathbf{y}$.

$\mathbf{F}'$ can be modeled as:

$$\mathbf{F}' = \overline{\mathbf{f}} + \mathbf{T} \bullet \mathbf{P}^{\mathbf{T}} + \mathbf{E_f}.$$

With $\mathbf{P}$ being the loadings of the k input features, $\mathbf{T}$ being the scores of the l input sequences. $\overline{\mathbf{f}}$ represents the row vector of the mean values of the features and $\mathbf{E_f}$ is the error in $\mathbf{F}'$ that can not be modeled.

Likewise $\mathbf{y}$ can be modeled as:

$$\mathbf{y} = \overline{y} + \mathbf{T} \bullet \mathbf{Q}^{\mathbf{T}} + \mathbf{E_y}.$$

The prediction $\widehat{y}$ can then be modeled as:

$$\widehat{y_i} = b_0 + \mathbf{f_i'} * \mathbf{b}$$

$\mathbf{b}$ is the column vector of the single estimation weights b_m, b_0 is the model offset. A detailed description of PLSR can be found in chapter 3.5 of [18].

Prediction Correction using Additional Quality Information

The NR quality metric gained by the previous steps faces the problem that even the original video may contain a certain amount of blur or blocking and different sequences do not only have a different amount of details but also do have different motion properties. For this reason the overall prediction accuracy of the so far described model is low. But plotting the predicted quality against the quality measured in subjective tests reveals that the prediction accuracy for each single sequence is very high: the data points for one single sequence lie on one straight line only with unknown slope s and unknown offset o. Figure 3 shows the prediction without the correction step: the dashed diagonal shows the required regression line. The regression lines for the three sequences Crew, Foreman and Husky have a significant different slope and offset. The overall prediction accuracy therefore can be improved by estimating the slope and the offset of these lines by calculating the predicted quality of the original video ($\widehat{y}_{orig}$) and of a low quality version of the video ($\widehat{y}_{low}$) using the same quality predictor.

While the original video is available and the subjective visual quality of this original is inherently given to be 1 on a 0 to 1 scale with a comparably small error only, an estimation of a low quality video can be produced by e.g. encoding the original with a low bit rate. Obviously the subjective visual quality of this low quality video can only be guessed (here set to 0.25). Including the predicted quality of the original video and the

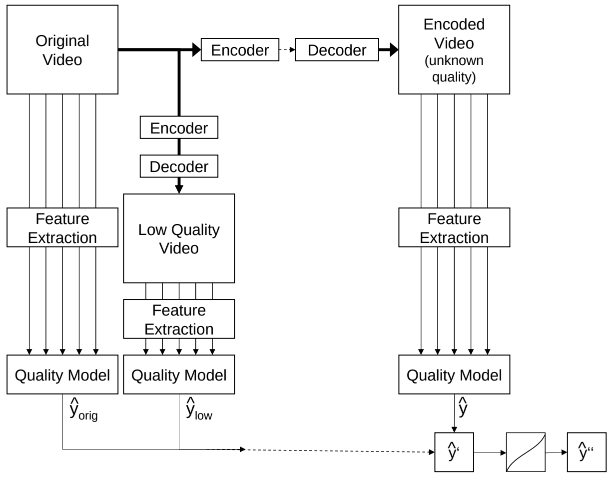

Fig. 2. Complete Prediction Model

predicted quality of the low quality video, the NR model will become a RR model, even if the additional data that has to be send is very low (only two values per sequence). The corrected prediction $\widehat{y}_i'$ is calculated as

$$\widehat{y}_i' = \frac{\widehat{y}_i - o}{s} \text{ with } s = \frac{\widehat{y_{orig}} - \widehat{y_{low}}}{1.0 - 0.25} \text{ and } o = \widehat{y}_{low} - 0.25 * s \ .$$

Correcting Nonlinearities of Subjective Ratings

It is known, that at the extremes of the test range (very good or very bad quality) subjective testing does have a nonlinear quality rating and ratings do not reach the very extremes of the scale but are saturated before. For this reason it is proposed to slightly correct the prediction values $\widehat{y}'$ using a sigmoid nonlinear correction. The general sigmoid function is given as

$$\widehat{y}'' = a/(1 + e^{(-(\widehat{y'} - b)/c)})$$

For the correction the following values were chosen: $a = 1.0$, $b = 0.5$, $c = 0.2$. The nonlinear sigmoid correction function is shown in figure 4, it has to be noted, that the applied correction function is very close to be linear over a wide quality range. Figure 2 gives an overview over the presented prediction model.

4. A Reduced Reference Model for AVC/H.264 Encoded Video

Subjective Testing for Generating Ground Truth Data

A reduced reference metric using the above described method was built using data from two subjective tests that included AVC/H.264 encoded video. Tests were done on video encoded at CIF resolution and were performed according to the rules given in ITU-R BT-500 [19]. This especially includes:

- Room setup compliant to ITU-R BT-500
- To maintain a fixed viewing distance the video were displayed using a DLP projector, the viewing distance was set to four times the picture height
- SSIS (Single Stimulus Impairment Scale) evaluation using a discrete impairment scale ranging from 0 to 10 (later rescaled to 0 to 1)
- All test sequences were evaluated by at least 20 naive viewers (students who were not familiar with video coding or video quality evaluation), all screened for visual accuracy and color blindness

TABLE I
Test Sequences

Sequence	Bit Rate (kbps)	fps	Conditions
Sequences used for calibration			
Bus	128,256 512	15fps 30fps	IBP, only one I-Frame IBBP, only one I-Frame
Football	256, 512 1024	15fps 30fps	IBP, only one I-Frame IBBP, only one I-Frame
Mobile	96, 192, 384, 768	15fps	IBP, one I-Frame every 2 seconds
Harbour	192 384, 750	15fps 30fps	IBP, one I-Frame every 1.2 seconds IBBP, one I-Frame every 1.2 seconds
Sequences used for validation			
City Crew Ice	192 384, 750	15fps 30fps	IBP, one I-Frame every 1.2 seconds IBBP, one I-Frame every 1.2 seconds
Deadline Paris Zoom	96, 192, 384, 768	15fps	IPPP, only one I-Frame
Foreman	256, 512 1024 96, 192, 384, 768	15fps 30fps 15fps	IBP, only one I-Frame IBBP, only one I-Frame IBP, one I-Frame every 2 seconds
Husky Tempete	96, 192, 384, 768	15fps	IBP, one I-Frame every 2 seconds

- To minimize the contextual effect, which is known to affect results in a single stimulus environment, every encoded sequence was shown twice in the test and rated twice by each test subject. Like this also the ability of the viewers to rate one video could be tested and outliers (indicated by two votes for the same sequence that differ to much) could be removed.
- Each test was preceded by an extensive training session to train the subjects on the task of evaluating the video
- Each single test session did not last longer than 25 minutes and an adaptation phase of five sequences was set at the start of each test session (this was not undisclosed to the subjects).

The 95% confidence intervals were below 0.04 on a 0 to 1 scale, which shows, that the results from the tests are very reliable. Table I shows the sequences and bit rates used for these two tests. Before building the model the data from those tests was split into two parts: only four out of 13 sequences were used for calibration of the metric, while the other nine sequences were used for the verification phase.

The Regression Model

After applying a MSC on the calibration data, a very simple regression model with only one PC can be built by applying a PLSR. The resulting weights b_m of the objective features and the model offset b_0 are given in Table II. The PLSR on the matrix $\mathbf{F}'$ revealed that the feature 'noise' does not have an influence on the model, therefore this feature was removed and only the remaining seven features were taken into account.

Correcting the Results of the Model

The low quality video needed for the correction step was constructed by encoding the video using the AVC/H.264 standard with a high (fixed) quantization parameter (resulting in low quality). It has to be noted, that not only the coding parameters

TABLE II
WEIGHTS OF OBJECTIVE FEATURES

Feature	Weight b
Activity	0.036
Blocking	-0.120
Blur	-0.109
Color Continuity	0.054
Edge Continuity	-0.072
Motion Continuity	0.090
Predictability	0.095
b_0	4.073

for producing this low quality video differ quite significant from those used to encode the videos under test, but also a different encoder has been used for this task.

5. RESULTS

Verification was done on the sequences not used for the development of the presented RR quality metric. In total 36 data points were available spanning a quality range from 0.098 to 0.913. Beside PSNR two other FR metrics were calculated for the presented data. The Edge-PSNR metric [4] was chosen as one representative of the methods standardized in ITU-T J144 [3]. The second FR metric chosen for comparison is the popular SSIM (Structural SIMilarity index) as presented by Wang in [5]. In addition to the classical Pearson Linear Correlation and Spearman Rank Oder Correlation to express the ability of the model to correctly predict the visual quality, also an outlier ratio is given. A data point is considered to be an outlier if the difference between measured and predicted quality is more than 0.05 on a 0 to 1 scale (remember, that the 95% confidence intervals from the tests were below 0.04). Note that for the outlier ratio no data fitting was applied for the proposed method while linear fitting was applied for the other three models. Linear data fitting has been chosen to fit the predicted values to the actual given data. While nonlinear fitting (logistic, sigmoid) is sometimes proposed for this purpose, higher order fitting always carries the danger of fitting the model too much to the actual data and possibly jeopardizing the ability to predict unknown data.

In addition the slope and offset of the regression line before linear data fitting are given in Table IV. This shows how much the model relies on a final fitting stage (an information, that is not given by the correlation measurement) and the ability to finally provide a correct and meaningful quality measurement as without the knowledge of that line no prediction can be made. For a perfect model the slope of this regression line would be 1.0 with 0 offset. The problems of a regression far different compared to the one of an optimal model can be shown easily for the SSIM metric: if the regression line that comes with the model is only a little bit different to the real data (e.g. assume the slope to be 0.16 with 0.8 offset instead of 0.173 and 0.821), the outlier ratio increases dramatically from 0.667 to 0.861. Detailed results of each metric are given in figures 5 to 8.

6. CONCLUSION

A reduced reference quality metric for AVC/H.264 was built using methods provided by multivariate data analysis. The metric was validated using results from careful conducted subjective tests and no sequence used for calibration of the model was used

TABLE III
COMPARISON OF OBJECTIVE METRICS

Model	Pearson Linear Correlation	Spearman Rank Oder Correlation	Outlier Ratio
Proposed	0.851	0.782	0.500
PSNR	0.690	0.623	0.833
Edge-PSNR	0.802	0.745	0.833
SSIM	0.763	0.623	0.667

TABLE IV
REGRESSION LINE BEFORE DATA FITTING

Model	Slope	Offset
Proposed	0.969	0.042
PSNR*	0.625	0.183
Edge-PSNR	0.353	0.348
SSIM	0.173	0.821

For easier comparison PSNR was rescaled to PSNR* = (PSNR-15)/30

during the verification phase. Beside providing a high prediction accuracy, the gained model allows a quality prediction by transmitting only two additional values, while most other reduced reference metrics need a much higher amount of additional data to be transmitted. Compared to FR metrics the presented RR metric also carries the advantage, that no temporal or spatial alignment between the coded sequence and the original sequence has to be made. In fact the correction step delivers equally good results if the quality prediction for the original video and the low quality video is performed on a part of the video only.

REFERENCES

[1] *ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG4-AVC), Advanced Video Coding for Generic Audiovisual Services* v1, May, 2003; v2, Jan. 2004; v3 (with FRExt),Sept. 2004; v4, Jul. 2005

[2] U. Benzler, M. Wien, *Results of SVC CE3 (Quality Evaluation).* ISO/IEC JTC1/SC29/WG11 M10931, Redmond, Jul. 2004

[3] ITU-T J.144. *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.* 2004

[4] C. Lee, S. Cho, J. Choe, T. Jeong, W. Ahn and E. Lee *Objective video quality assessment* SPIE Journal of Optical Engineering, Volume 45, Issue 1, Jan. 2006

[5] Z. Wang, L. Lu, A. C.Bovik. *Video quality assessment using structural distortion measurement.* ICIP 2002. International Conference on Image Processing, IEEE Proceedings on, 2002

[6] Z. Wang, A. C.Bovic, E. P.Simoncelli, *Structural Approaches to Image Quality Assessment.* Chapter 8.3 in Handbook of Image and Video Processing, 2nd Edition Academic Press, 2005

[7] A. Shnayderman, A. Gusev, A. M.Eskicioglu, *An SVD-based grayscale image quality measure for local and global assessment* Image Processing, IEEE Transactions on Volume 15, Issue 2, Page(s):422 - 429, Feb. 2006

[8] G. Zhai, W. Zhang, X. Yang, S. Yao, Y. Xu, *GES: a new image quality assessment metric based on energy features in Gabor transform domain* ISCAS 2006, Proceedings IEEE International Symposium on, Page(s):4 pp., 21-24 May 2006

[9] G. Zhai, W. Zhang, X. Yang, Y. Xu, *Image quality assessment metrics based on multi-scale edge presentation* Signal Processing Systems Design and Implementation, 2005. IEEE Workshop on, Page(s):331 - 336, 2-4 Nov. 2005

[10] E. Ong, W. Lin, Z. Lu, S. Yao, *Colour perceptual video quality metric* ICIP 2005. IEEE International Conference on Image Processing, Volume 3, Page(s):III - 1172-5, 11-14 Sept. 2005

[11] M. Carnec, P. Le Callet, D. Barba, *Full reference and reduced reference metrics for image quality assessment* Seventh International Symposium on Signal Processing and Its Applications, Proceedings, Volume 1, Pages:477 - 480, 1-4 July 2003

[12] T. M.Kusuma, H.-J. Zepernick, M. Caldera, *On the development of a reduced-reference perceptual image quality metric* Systems Communications, Proceedings, Pages:178 - 184, 14-17 Aug. 2005

[13] Z. Wang; E. P.Simoncelli *Reduced-Reference Image Quality Assessment using a Walet-Domain Natural Image Statistic Model* Human Vision and Electronic Imaging X, Proceedings, Spie, vol 5666, San Jose, CA, Jan. 2005

[14] Y. Fu-zheng, W. Xin-dai, C. Yi-lin, W. Shuai, *A no-reference video quality assessment method based on digital watermark* Personal, Indoor and Mobile Radio Communications, 2003. 14th IEEE Proceedings on Volume 3, Page(s):2707 - 2710 vol.3, 7-10 Sept. 2003

[15] P. Le Callet, C. Viard-Gaudin, D. Barba, *A Convolutional Neural Network Approach for Objective Video Quality Assessment.* IEEE Transactions on Neural Networks, Volume 17, Issue 5, Page(s):1316 - 1327, Sept. 2006

[16] A. C. Bovik, S. Liu, *DCT-domain blind measurement of blocking artifacts in DCT-coded images.* International Conference on Acoustics, Speech, and Signal Processing. IEEE Proceedings on, Volume 3, Page(s): 1725 -1728, 07-11 May 2001

[17] S. Winkler, T. Ebrahimi, *A no-reference perceptual blur metric.* Proceedings of the International Conference on Image Processing, vol. 3, pp. 57–60, Rochester, NY, Sep. 22-25, 2002

[18] H. Martens, T. Naes, *Mutltivariate Calibration.* Wiley & Sons, 1992.

[19] *ITU-R Rec. BT. 500-11, Methodology for the Subjective Assessment of Quality for Television Pictures. 2002*

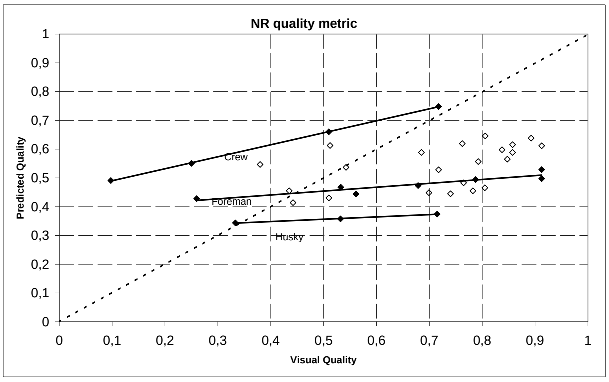

Fig. 3. Predicted vs. measured quality before slope and offset correction

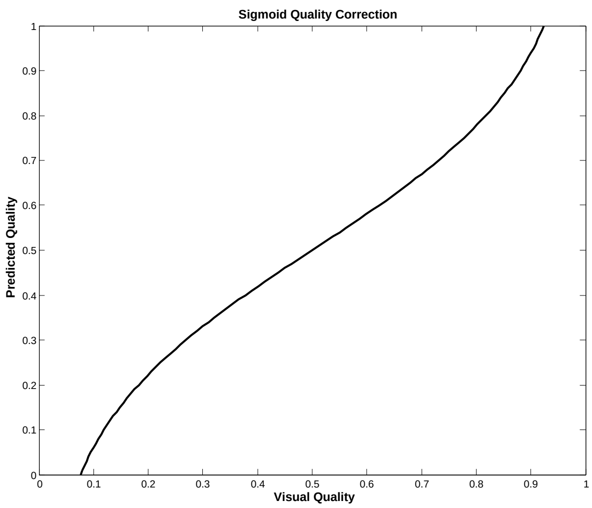

Fig. 4. Sigmoid correction function

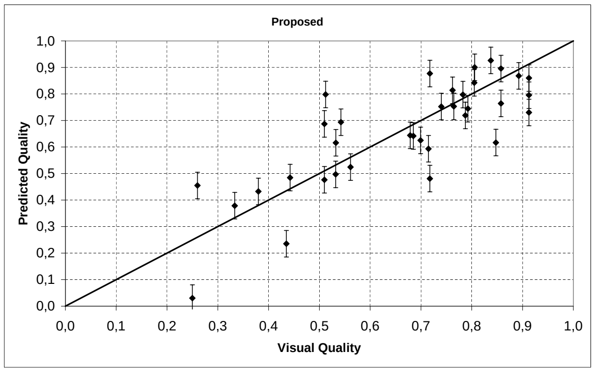

Fig. 5. Proposed RR metric - no data fitting

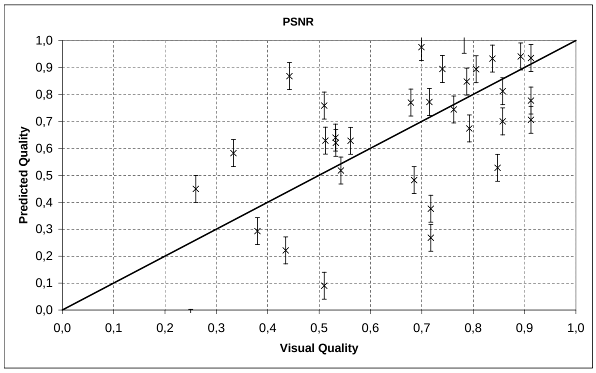

Fig. 6. PSNR - linear data fitting

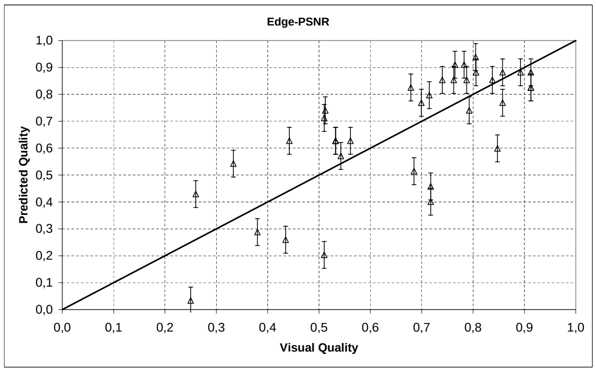

Fig. 7. Edge-PSNR - linear data fitting

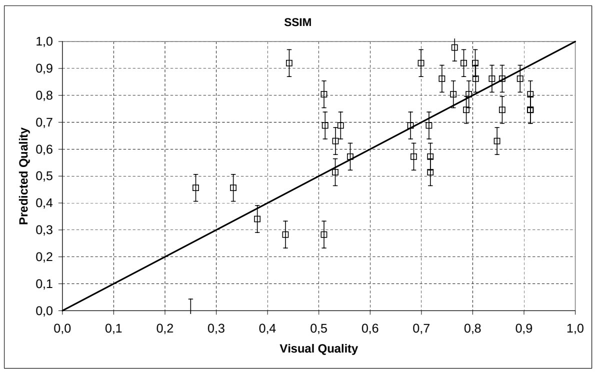

Fig. 8. SSIM - linear data fitting