ENHANCED FULL-REFERENCE QUALITY ASSESSMENT METHOD FOR IMAGE AND VIDEO

Hagen Kaprykowsky, Martin Köppel, and Patrick Ndjiki-Nya

Image Communication Group, Image Processing Department
Fraunhofer Institute for Telecommunications - Heinrich-Hertz-Institut, Berlin, Germany
{hagen.kaprykowsky/patrick.ndjiki-nya}@hhi.fraunhofer.de

ABSTRACT

Evaluation and optimization of visual quality is of fundamental importance to numerous image and video processing applications including broadcasting, archiving, and restoration. As quality assessment using subjective tests is very costly and time-consuming, objective measures are required that can predict the perceptual judgment of human viewers. In this paper, a top-down full-reference quality assessment tool that mimics a selection of prominent human visual system properties is presented. Perceptually salient feature points are thereby used. The proposed approach is based on an enhanced structure detection method using spatial segmentation. Experimental evaluations are conducted on two test sets. It is shown that compared to the best quality measure by the Video Quality Experts Group, the proposed method yields the most robust prediction of subjective quality given a wide range of video content.

1. INTRODUCTION

Automatic quality assessment of digital pictures is a crucial issue in several image and video processing applications [1]. However, quality analysis is one of the most challenging tasks since the properties of the very complex human vision system (HVS) determine the visibility of distortions and thus the perceived quality. Low complexity objective measures are required that can predict the perceptual judgment of the HVS. Hence, major efforts have been made by the Video Quality Experts Group (VQEG) to establish standardized quality models [2].

The evaluation of quality may be divided into two classes, subjective and objective methods. In subjective evaluations human subjects (real end-users) have to vote for the quality of a medium in a controlled test environment. Subjective assessment is the most reliable way to determine the actual quality of the material but the main disadvantage of these methods is that they are very expensive and time-consuming. Objective methods however attempt to predict the quality automatically as it would be perceived by a human observer. They have been traditionally performed by measuring physical characteristics of a signal and are therefore repeatable and can be standardized. The main difficulty is to find an appropriate trade-off between very low complexity and most likely low perceptual correlation on the one hand and very high complexity and most likely high perceptual correlation on the other hand.

Furthermore, image and video quality estimation algorithms can be classified based on the required knowledge of the source material. Three categories are typically defined in the literature: full-reference, reduced-reference and no-reference approaches. Full-reference approaches require that the original and the distorted signal are known in order to determine the corresponding image or video quality measure. They are usually impractical in applications where the original signal is not available. Reduced-reference approaches assume that a compact description of the original signal is given. The same description is extracted from the distorted signal and compared to the reference metadata. No-reference measures (also known as blind measures) do not require any information on the original signal. They are thus applied only on the distorted signal. These methods are usually not generic and require a priori knowledge of the type of quality problems that can occur in the given framework. The correlation of full-reference methods with subjective perceived quality is in general higher than for no-reference methods. This is to be expected, since full-reference methods have more information to evaluate the quality. In most practical applications, on the other hand, neither the original signal nor a compact description of the original signal is available.

2. PRELIMINARY CONSIDERATIONS

The proposed objective full-reference video quality measure (VQM) is based on the metrics described in [3], [4], [5]. These methods are top-down approaches that integrate some salient properties of the HVS. This type of approach is preferred in this work in order to avoid detailed formulation of assumptions on sparsely understood functional properties of early vision stages. The approach by Ong et al. [3], [4] consists of three features: the block-fidelity, the content richness fidelity and the distortion-invisibility. The first feature aims to measure the amount of distortion at block-boundaries and the latter two model measure properties of the HVS (e.g. masking). The formulation of the different masking properties results in a very complex model with a lot of degrees of freedom, which appears hardly manageable.

The VQM defined by Ong et al. [3], [4] was revisited in [5] with the aims of simplification and significant performance improvement. To overcome the problem of complexity, Ndjiki-Nya et al. [5] mainly focus on the optimization of the block-fidelity feature to increase performance. The approach is based on feature points and has only two degrees of freedom (macro block size and γ, cp. Section 3.1.). A sampling of the edge masks orthogonally to the gradient's direction is proposed, which yields better feature point selection, since major structural information may be ignored by the measure proposed by Ong et al. [3], [4] if the former does not match the macro block grid, which will assumingly be the case in most natural images. In addition a filter like Sobel is suggested for contour detection to achieve more robustness against spurious edges. Due to the purposeful inclusion of boundary samples into the set of feature points used for quality assessment, not only can tiling effects be detected by the measure by Ndjiki-Nya et al. but also distortions affecting object boundaries. Hence, the initial block fidelity measure has been generalized to achieve a simple, generic impairment detection tool. Experimental results showed important gains compared to [3], [4] in particular and other state-of-the-art quality measures [6] in general.

3. ENHANCED QUALITY MEASURE

3.1. Feature point-based quality analysis

In this work, an improved contour detection method is proposed and integrated in the approach of [5]. Due to the enhanced detection of the contour mask, more accurate quality predictions compared to [3], [4], [5] and [6] can be achieved. The mathematical formulation as defined in [3], [4] of the VQM for single pictures, $Q(t)$, is given as

$$Q(t) = \frac{e^{\frac{\gamma}{1+\delta(t)}}}{e^{\gamma}} \tag{1}$$

where the term γ can be freely selected and steers the interval of $\delta(t)$ for which the contrast is enhanced or reduced. The denominator is a normalization factor. The variable $\delta(t)$ represents a differential term that assesses the distance between a given reference and a corresponding distorted signal. The difference term $\delta(t)$ is defined as follows

$$\delta(t) = \frac{|E_o(t) - E_d(t)|}{E_o(t)} \tag{2}$$

where $E_o(t)$ and $E_d(t)$ correspond to mean absolute spatial gradients in the original and the distorted signal respectively. As the contrast of the pictures plays an important role in quality perception, it was suggested by Ong et al. to build this feature into the quality measures. Hence, for the calculation of $\delta(t)$, a contour matrix $C(x,y,t)$ is used that emphasizes regions of large spatial contrast as object boundaries in the edge masks, as the former are assumed to be of salient relevance for subjective quality perception [7]. The global quality score for an entire video sequence is determined as the mean of the single picture qualities.

The proposed feature point selection in [5] can be formalized as follows

$$E_o(t) = \frac{1}{\left(\frac{X}{\kappa}-1\right)Y}\sum_{x=1}^{\frac{X}{\kappa}-1}\sum_{y=1}^{Y}|o_{0^\circ}(\kappa x,y,t)|m(x,y,t) + \frac{1}{X\left(\frac{Y}{\kappa}-1\right)}\sum_{x=1}^{X}\sum_{y=1}^{\frac{Y}{\kappa}-1}|o_{90^\circ}(x,\kappa y,t)|m(x,y,t) \tag{3}$$

$$E_d(t) = \frac{1}{\left(\frac{X}{\kappa}-1\right)Y}\sum_{x=1}^{\frac{X}{\kappa}-1}\sum_{y=1}^{Y}|d_{0^\circ}(\kappa x,y,t)|m(x,y,t) + \frac{1}{X\left(\frac{Y}{\kappa}-1\right)}\sum_{x=1}^{X}\sum_{y=1}^{\frac{Y}{\kappa}-1}|d_{90^\circ}(x,\kappa y,t)|m(x,y,t) \tag{4}$$

where a $\kappa \times \kappa$ macro block size is assumed and $m(x,y,t)$ is a binary mask that defines the regions of interest and (X,Y) represent the width and the height of a video frame. $o(x,y,t)$ represents the original signal, while $d(x,y,t)$ constitutes the distorted signal. $o_\beta(x,y,t)$ and $d_\beta(x,y,t)$ ($\beta \in \{0°, 90°\}$) represent the high pass filtered original and distorted signals respectively and are defined as follows

$$o_\beta(x,y,t) = \left[o(x,y,t) * f_\beta(x,y)\right]C(x,y,t) \tag{5}$$

$$d_\beta(x,y,t) = \left[d(x,y,t) * f_\beta(x,y)\right]C(x,y,t) \tag{6}$$

where $f_\beta(x,y)$ ($\beta \in \{0°, 90°\}$) is a linear, anisotropic gradient filter of orientation β used for edge detection. "*" thereby represents the convolution operation. $C(x,y,t)$ is determined from the original signal and operates object contour enhancement in the original and distorted edge masks. The assumption thereby is that objects and particularly their boundaries are of salient relevance for subjective quality perception [13]. Hence, locations featuring spatial discontinuities are assigned a high weight, while other locations are assigned low weights. $C(x,y,t)$ is defined in [3], [4], [5] as follows

$$C(x,y,t) = \frac{|o_{0^\circ}(x,y,t)| + |o_{45^\circ}(x,y,t)| + |o_{90^\circ}(x,y,t)| + |o_{135^\circ}(x,y,t)|}{4} \tag{7}$$

It must, however, be noted that the impairment detector fails if a noisy source signal is given, where object contours cannot be properly identified. Fortunately, improved structure detection algorithms to improve contour detection can be easily integrated into the given approach VQM.

3.2. General approach

The general approach for the proposed enhanced quality measure using advanced structure detection is illustrated in Figure 1. Contour detection of [3], [4], [5] based on the mean of spatial gradients (cp. (7)) has been replaced by an enhanced contour detection approach using spatial segmentation (see Figure 1, dashed blocks, left). The detection of vertical and horizontal edges (see Figure 1, right block) has not been modified within the new framework. A detailed description of the processing steps is given in the following sections.

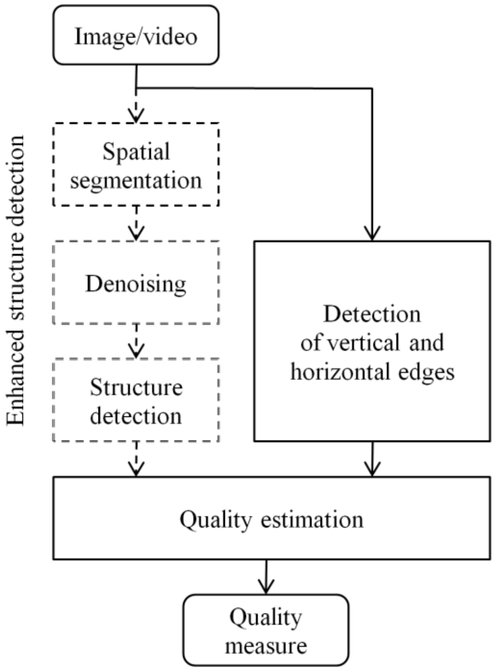

Figure 1 – General approach of the proposed feature point-based quality measure including enhanced structure detection (dashed blocks, left).

4. ENHANCED STRUCTURE DETECTION

4.1. Spatial segmentation

The segmentation algorithm is based on a method by Spann and Wilson [8]. They propose a multi resolution analysis approach with the fundamental assumption of object property invariance across spatial scales. By applying a quad-tree smoothing operation on the original image, a multi resolution image pyramid is generated. Homogeneous regions are extracted at the highest level of the quad-tree, i.e. at the lowest spatial resolution, via statistical classification or histogram clustering. The latter is based on a local centroid algorithm [9]. The classification step is followed by a coarse to fine boundary estimation based on the partition obtained at the top level of the pyramid. No a priori information, such as the number of segments, is required in this framework. Histogram thresholding approaches have been first introduced for monochrome images and widely used for segmentation [10]. Hence, they do not take advantage of the information available in the typically three color channels. Several attempts to extend histogram clustering approaches to color images have been made in the past. The fundamental difference between monochrome and color images is that the latter are represented by a tristimulus R, G, and B or their linear, respectively non-linear, transformations. The method used in this paper is an enhanced version of [8] including color information. By integrating a redundancy elimination unit, the proposed framework by Spann and Wilson has been extended to take color information into account [11]. The redundancy elimination approach is similar to the algorithm by Ohta et al. [12]. Color features with high discrimination power are extracted using principal component analysis (PCA) [13].

Figure 2 – Key picture of "Tempete" sequence.

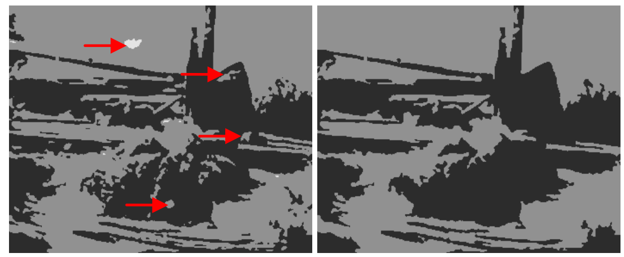

Figure 3 – Key picture of "Tempete" sequence: a) segmentation result including little blobs (left) and b) segmentation result after denoising (right).

PCA is an approach that is typically used for linear dimensionality reduction and that inherently allows for explicit control of the error introduced by dimensionality reduction. Consider a region A in a D-dimensional color space. RGB will be considered in the following ($D=3$), since it is the color space used in this work, but other color spaces (e.g. HSV) may also be used. Let the distributions of R, G and B in A be h_R, h_G, and h_B, respectively. Let Σ represent the covariance matrix of the set of row vectors of the matrix $[h_R\ h_G\ h_B]$ of dimension $N \times 3$, where N is the histogram resolution. Further assume that $\lambda_1 \geq \lambda_2 \geq \lambda_3$, where the λ_i are the eigenvalues of Σ. Then the color features F_1, F_2, and F_3 defined as

$$F_i = f_{Ri} \times R + f_{Gi} \times G + f_{Bi} \times B \tag{8}$$

where the eigenvectors of Σ that are given by $f_i = (f_{Ri}, f_{Gi}, f_{Bi})$ (RGB color space) can be shown to be uncorrelated with F_1 having the largest variance equal to λ_1, and thus the largest discriminant power. F_2 has the largest discriminant power among the vectors orthogonal to F_1 [13]. In this work, F_1 is used for multi resolution segmentation. As an example, a segmentation result of a key picture of the "Tempete" sequence (see Figure 2) is shown in Figure 3 a). The red arrows highlight little blobs that will be removed in the denoising step.

4.2. Denoising and structure detection

Structure detection is done based on a coarse segmentation mask. It is assumed that the segmentation result highlights the most important dominant gradients in the considered picture. Denoising of the segmentation mask is required before conducting proper structure detection. This encompasses the elimination of little blobs (cp. Figure 3 a), red arrows). Denoising results are illustrated in Figure 3 b). Afterwards, the segmentation mask is processed in scan line order to determine segment borders.

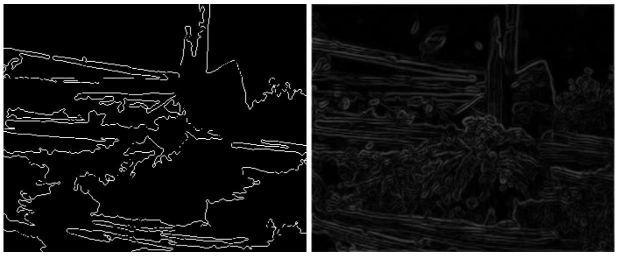

Figure 4 – Key picture of "Tempete" sequence: a) binary mask generated by structure detection using enhanced method (left) and b) gray-level mask generated by structure detection using Sobel filter (right).

This can be formulized as follows

$$C(x,y,t)=\begin{cases}1 \text{ if } \mathrm{S}(x+1,y,t)\neq \mathrm{S}(\mathrm{x},y,t) \text{ or} \\ \quad \mathrm{S}(\mathrm{x},y+1,t)\neq \mathrm{S}(\mathrm{x},y,t) \\ 0 \text{ otherwise}\end{cases} \qquad (9)$$

where $C(x,y,t)$ corresponds to the contour mask (structure mask) and $S(x,y,t)$ represents the segmentation mask after denoising. Finally, in a post-processing step, the determined structures are thinned to obtain one sample thick curves. The generated structure mask is a binary image (see Figure 4 a)).

This approach typically gives more reliable structure masks than simple gradient operators (e.g. Sobel, Prewitt, etc). The structure mask generated by these operators is a gray-level image (cp. Figure 4 b). The gray-level contour mask $C(x,y,t)$ (cp. (7)) is replaced in (5), (6) by the improved binary contour mask and used for quality analysis within the feature point-based approach (cp. (3), (4)).

5. EXPERIMENTAL EVALUATION

An important application scenario for quality assessment is block-based hybrid video coding. Standardized algorithms such as H.264/AVC process an input picture macro-block-wise [14]. The artifacts that typically occur in the given framework are blocking (tiling), blurring as well as unnatural, jerky motion. The proposed quality measure mainly aims to detect blockiness, blurriness and contrast loss and can be applied within a video coding scenario.

Experimental evaluations are conducted to validate the proposed full-reference quality measure for block-based hybrid video coding. Unfortunately, the official VQEG test data (Phase II) [6] were not accessible to the authors. Hence, two different ground truth sets have been used in the present work to evaluate the video quality measure. The data sets are provided by MPEG [15] and Fraunhofer Heinrich-Hertz-Institut (HHI). All quality measures are evaluated by matching the corresponding predicted differential mean opinion scores (DMOSp) with the subjective differential mean onion scores (DMOS). The data sets are presented in the following section.

5.1. Data sets

The TV data set provided by HHI corresponds to five video clips obtained from several German television channels. The clips feature various contents as news, sports, cartoon, monochrome and color movies that are MPEG-2 coded (cp. Table 1).

Property Designation	**Property**
Subj. eval. appr.	DSCQS
Video codec	MPEG-2
Resolution	QVGA (320x240)
Frame rate	12.5 Hz
Duration	10s
Bit rate	Variable
Number of test sequences	5

Table 1 – Properties of TV ground truth provided by HHI.

The data set was used for evaluation in [5]. The data set is unfortunately not publicly available due to copyright issues. Subjective evaluations were conducted using the double stimulus continuous quality scale (DSCQS) approach [16] yielding subjective differential mean opinion scores.

The MPEG data set consists of four video sequences and was formerly used to benchmark the performance of MPEG-4 and H.26L anchors [15]. Details on the video sequences selected by MPEG can be found in Table 2.

Property Designation	**Property**
Subj. eval. appr.	DSIS
Video codec	MPEG-4 and H.26L
Resolution	QCIF (176x144)
Frame rate	10 Hz and 15 Hz
Duration	10s
Bit rate	32 kbps and 64 kbps
Number of test sequences	4

Table 2 – Properties of ground truths provided by MPEG.

This data set was subjectively evaluated yielding subjective differential mean opinion scores using the double stimulus impairment scale (DSIS) approach [16]. The test sequences provided by MPEG are “Container Ship”, “Foreman”, “News”, and “Tempete”. The sequences are MPEG-4 and H.26L coded.

5.2. Experimental results

The proposed quality measure is evaluated based on the Pearson’s correlation coefficient (r_p) and Spearman’s rank correlation coefficient (r_s). r_p and r_s are two major benchmark criteria recommended by VQEG [6]. Pearson’s correlation coefficient relates to the prediction accuracy, i.e., the ability of the model to reliably predict the subjective video perception and Spearman’s rank correlation coefficient aims to measures the model performance with respect to the relative magnitudes of subjective quality ratings. Comparative evaluations to VQEG’s best quality measure proposed by the National Telecommunications and Information Administration (NTIA) [17], the quality measure proposed by Ndjiki-Nya et al. [5], the quality metric proposed by Ong et al. [3], [4], and peak signal-to-noise ratio (PSNR) are conducted. The correlation coefficients were determined by the entire data within a data set. Hence, no cross-validation for evaluation was used.

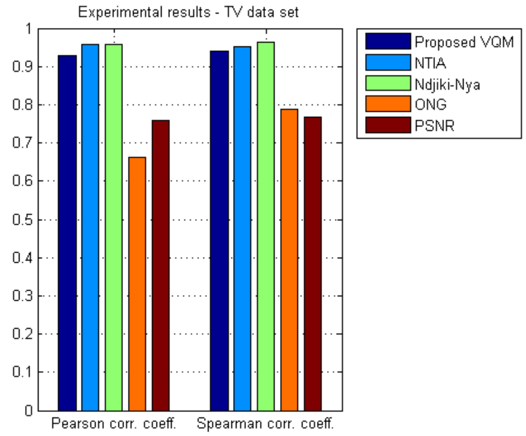

Figure 5 – Evaluation results on TV data set.

In the following, the gain of the proposed method with respect to the measure to be compared is defined as Δr_p and Δr_s and corresponds to the difference between the correlation coefficients (r_p and r_s respectively) of the proposed method and of another measure.

Evaluation results on the TV data are shown in Figure 5. The proposed method features comparable but slightly worse results than the measures by NTIA and Ndjiki-Nya et al. while the new measure outperforms PSNR and the metric proposed by Ong et al.. Δr_p and Δr_s are listed in Table 3.

	NTIA	Ndjiki-Nya	ONG	PSNR
Δr_p	-0,03	-0,03	0,27	0,17
Δr_s	-0,01	-0,02	0,15	0,17

Table 3 – Evaluation results on TV data set: difference between the correlation coefficients of the proposed method and the other measures under test.

The MPEG data set contains a video sequence called “Tempete”. A key picture of this video is depicted in Figure 2. The experimental results in [5] were determined based on the “inlier” sequences within a cross-validation framework. That is, the sequences for which the hypothesis that the relationship between DMOS and DMOSp is linear (F-test) holds. Notice that 95% confidence intervals were used for linear regression. The “Tempete” sequence was removed from the data set for all evaluations conducted in [5]. This relates to the fact that for all the considered VQMs, the “Tempete” sequence showed a severe outlier behavior such that regression computations led to the conclusion that no linear relationship exists between DMOS and DMOSp in all cases (F-test). In Figure 6 the evaluation results based on the MPEG data excluding the sequence “Tempete” are shown. As can be seen, the proposed VQM yields similar results to NTIA’s VQM. Δr_p and Δr_s are listed in Table 4.

	NTIA	Ndjiki-Nya	ONG	PSNR
Δr_p	-0,03	0,00	-0,01	-0,02
Δr_s	0,01	0,01	0,02	0,02

Table 4 – Evaluation results on MPEG data set excluding "Tempete": difference between the correlation coefficients of the proposed method and the other measures under test.

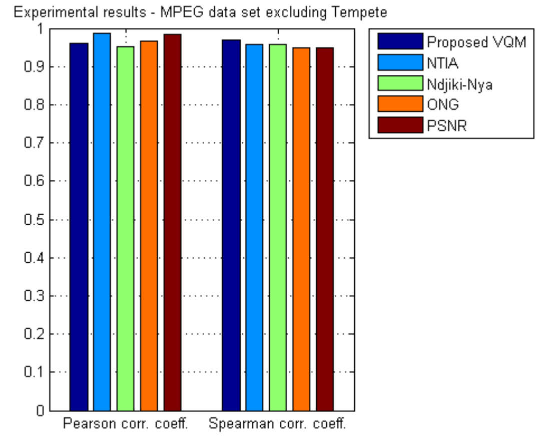

Figure 6 – Evaluation results on MPEG data set excluding "Tempete".

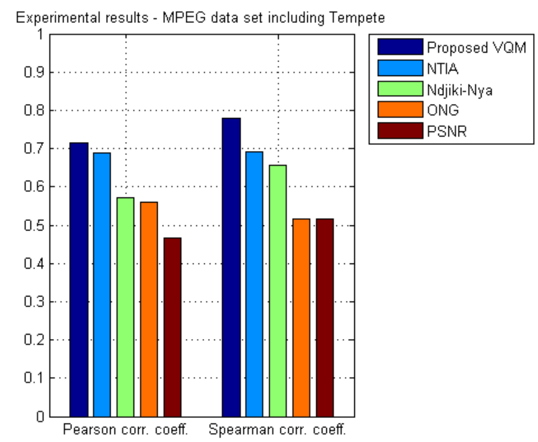

Figure 7 – Evaluation results on MPEG data set including "Tempete".

The proposed method was further evaluated on the MPEG data including the “Tempete” sequence and compared to the metrics presented above. In Figure 7 the evaluation results based on the MPEG data including “Tempete” are shown.

It can be seen that the proposed method outperforms all methods and achieves significantly higher correlations with quality prediction by humans (cf. r_s). The failure of most of the examined quality metrics can be explained by the extreme characteristics of the “Tempete” sequence. In this sequence, leaves of relevant resolution are continuously falling in the foreground of the scene. The falling of the leaves is chaotic and hardly predictable by motion estimation algorithms, which yields motion prediction failures. That implies lots of localized errors and low objective quality scores. These leaves are however most probably not the objects of interest in the scene. The former nevertheless contribute important distortions that yield to bad quality predictions although the background of the scene that might be subjectively more relevant may not feature annoying artifacts. As can be seen in Figure 4 b), the leaves yield high responses in the gradients which imply that these particular locations will be assigned high weights in the overall distortion evaluation. Using the proposed enhanced contour detection method, more accurate quality predictions can be achieved on the MPEG data set including the "Tempete" sequence. Significant improvement compared to NTIA's measure can especially be achieved for Spearman's correlation coefficient, which indicates that the proposed measure can predict the rank order of qualities more precisely. However, the overall performance of the proposed VQM is affected. Δr_p and Δr_s are listed in Table 5.

	NTIA	Ndjiki-Nya	ONG	PSNR
Δr_p	0,03	0,14	0,16	0,25
Δr_s	0,09	0,12	0,26	0,26

Table 5 – Evaluation results on MPEG data set including "Tempete": difference between the correlation coefficients of the proposed method and the other measures under test.

6. CONCLUSIONS AND FUTURE WORK

In this paper an objective full-reference quality metric was proposed that uses an enhanced structure detection method based on spatial segmentation. The enhanced structure mask was integrated into a feature point-based approach. Experimental evaluations have been conducted on two data sets. The most accurate quality predictions of all tested metrics have been achieved on a standard data set provided by MPEG using the new proposed measure. Evaluations on the TV data set showed that the proposed method features comparable but slightly worse results as the proposed methods by NTIA and Ndjiki-Nya et al. while it outperforms PSNR and the metric by Ong et al.. Hence, the proposed metric appears to be the only measure that can reliably explain the observed data. However, the correlation coefficients may be improved in future work by further optimization of the structure detection.

7. ACKNOWLEDGMENT

This work is supported by the German Federal Ministry of Economics and Technology within the THESEUS program.

8. REFERENCES

[1] H.-J. Engelke, and U. Zepernick, "Perceptual-based quality metrics for image and video services: A survey," Next Generation Internet Networks, 3rd EuroNGI Conference on, p. 190–197, May 2007.

[2] ITU-R WG6Q 6/39-E, "Objective Perceptual Video Quality Measurement Techniques for Standard Definition Digital Broadcast Television in the Presence of a Full Reference," October 2003.

[3] E. P. Ong, W. Lin, Z. Lu, and S. Yao, "Colour Perceptual Video Quality Metric," Proc. ICIP 2005, Vol. 3, p. 1172-1175, Genova, Italy, 2005.

[4] E. P. Ong, X. Yang, W. Lin, Z. Lu, and S. Yao, "Video Quality Metric for Low Bitrate Compressed Video," Proc. ICIP 2004, p. 3531-3534, Singapore 2004.

[5] P. Ndjiki-Nya, M. Barrado and T. Wiegand, "Efficient Full-Reference Assessment of Image and Video Quality Image Processing," Proc. ICIP 2007, Vol. 2, p. 125-128, September 2007.

[6] ITU-R WG6Q 6/39-E, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II," August 2003.

[7] Z. Wang, A. Conrad 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, April 2004.

[8] M. Spann, and R. Wilson, "A quad-tree approach to image segmentation which combines statistical and spatial information," Pattern Recognition, p. 257-269, 1985.

[9] R. Wilson, H. Knutsson, and G. H. Granlund, "The Operational Definition of the,Position of Line and Edge," Proc. ICPR 1982, Munich, Germany, September 1982.

[10] E. Littmann, and H. Ritter, "Adaptive Color Segmentation – A Comparison of Neural and Statistical Methods", IEEE Transactions on Neural Networks, Vol. 8, p. 175-185, January 1997.

[11] P. Ndjiki-Nya, G. Simo, and T. Wiegand, "Evaluation of Color Image Segmentation Algorithms Based on Histogram Thresholding," VLBV, p. 214-222, 2005.

[12] Ohta, Y., Kanade, T., and Sakai, T., "Color Information for Region Segmentation.,Computer Graphics and Image Processing," Vol. 13, No. 3, p. 222-241, 1980.

[13] C. M., Bishop,"Neural Networks for Pattern Recognition," Oxford University Press, 1995.

[14] J.-R. Ohm, "Multimedia Communication Technology," ISBN 3-540-01249-4, Springer, Berlin Heidelberg New York, 2004.

[15] ISO/IEC JTC 1/SC 29/WG 11 N3671, "Call for Proposals for New Tools to further Improve Video Coding Efficiency," La Baule, France, October 2000.

[16] ITU-R BT-500-11, "Methodology for Subjective Assessment of the Quality of Television Pictures", ITU-R Recommendations.

[17] M. H. Pinson and S. Wolf, "A New Standardized Method for Objectively,Measuring Video Quality," IEEE Transactions on Broadcasting, Vol. 50, No. 3, p. 312-322, September 2004.