

# Extending Video Quality Metrics to the Temporal Dimension with 2D-PCR

Christian Keimel, Martin Rothbucher and Klaus Diepold

Institute for Data Processing, Technische Universität München  
Arcisstr. 21, 80333 München, Germany

## ABSTRACT

The aim of any video quality metric is to deliver a quality prediction similar to the video quality perceived by human observers. One way to design such a model of human perception is by data analysis. In this contribution we intend to extend this approach to the temporal dimension. Even though video obviously consists of spatial and temporal dimensions, the temporal aspect is often not considered well enough. Instead of including this third dimension in the model itself, the metrics are usually only applied on a frame-by-frame basis and then temporally pooled, commonly by averaging. Therefore we propose to skip the temporal pooling step and use the additional temporal dimension in the model building step of the video quality metric. We propose to use the two dimensional extension of the PCR, the 2D-PCR, in order to obtain an improved model. We conducted extensive subjective tests with different HDTV video sequences at  $1920 \times 1080$  and 25 frames per seconds. For verification, we performed a cross validation to get a measure for the real-life performance of the acquired model. Finally, we will show that the direct inclusion of the temporal dimension of video into the model building improves the overall prediction accuracy of the visual quality significantly.

**Keywords:** visual quality, video quality metrics, PCR, 2D-PCR, temporal pooling, HDTV, no-reference metric

## 1. INTRODUCTION

The most common approach to the design of video quality metrics is to implement the characteristics of the human visual system (HVS) as well as possible. Unfortunately, many aspects of the HVS are not well understood. Another approach is therefore to consider the HVS as a black box: we can measure properties i.e. features of the video objectively at the input and can also obtain the output, the perceived visual quality, by subjective testing. We then use this dataset to generate a model, explaining the subjectively perceived quality with objectively measurable features.

In order to obtain such a model, principal component analysis (PCA) in combination with regression is often used, leading to the principal component regression (PCR). The advantage of such a data reduction before regression is, that for the model building only those components are considered that explain the variation in the data set best. This was first proposed by Miyahara<sup>1</sup>. We expanded this by using an extension of the PCR, the partial least squares regression (PLSR), and applied this approach successfully in<sup>2,3</sup> for no-reference video quality metrics.

Video is in many contributions considered to be just a straight forward extension of still images. Thus in most video quality metrics, the metric is applied to each frame of the video separately and then the metric's values are pooled temporally over all frames, usually by simple averaging. But this arbitrary pooling, especially averaging, obscures the influence of temporal distortions e.g. pumping effects on the human perception of quality. In<sup>4</sup> we have shown exemplarily for PSNR that even the use of more effective pooling methods alone can improve the accuracy of video quality metrics significantly. Also the importance of videos features temporal properties in the model building was shown recently in<sup>5</sup>.

---

Further author information: (Send correspondence to Christian Keimel)

Christian Keimel	E-mail: christian.keimel@tum.de	Telephone: +49 89 289 23629
Martin Rothbucher	E-mail: martin.rothbucher@tum.de	Telephone: +49 89 289 23632
Klaus Diepold	E-mail: kldi@tum.de	Telephone: +49 89 289 23602

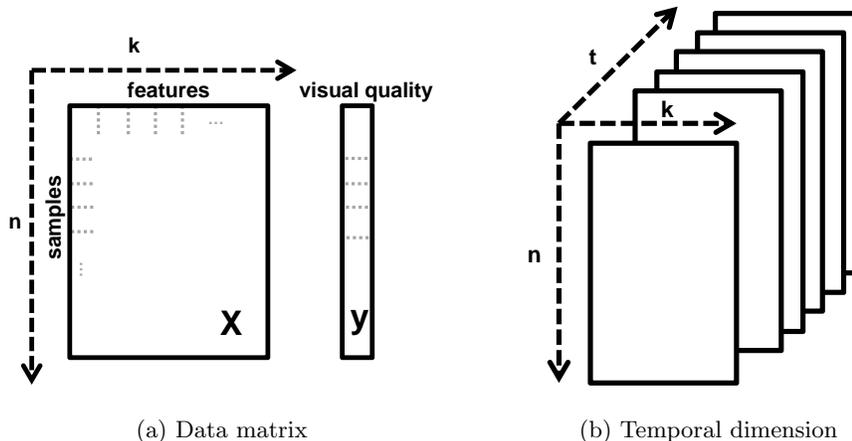


Figure 1: Model building with data analysis: structure of data matrix and temporal nature of video

Therefore, we propose to omit the temporal pooling step altogether and directly use the additional temporal dimension in the model building step of the video quality metric. This has already been done successfully in the audio domain by Rothbucher et.al.<sup>6</sup>. We no longer employ a a-priori temporal pooling step before the model building step, but use the whole video cube to generate the prediction model for the visual quality and thus consider the temporal dimension of video more appropriately.

In a first step, we propose to use the two dimensional extension of the PCR, the 2D-PCR, in order to obtain an improved model. Hence, we 2D-PCR with the results from subjective tests and extracted features to design a no-reference video quality metric for HDTV.

We conducted subjective tests with different HDTV video sequences at  $1920 \times 1080$  and 25 frames per seconds. For verification, we performed cross validation to get a measure for the real-life performance of the acquired model. We will show that the direct inclusion of the temporal dimension of video into the model building improves the overall prediction accuracy of the visual quality.

This contribution is organized as follows: Firstly, we discuss the design of video quality metrics with data analysis and present the one- and two-dimensional principal component analysis. Then we introduce the features used in our metric, before describing the subjective testing. Following the discussion and presentation of the results, we conclude with a short summary.

## 2. DESIGN OF VISUAL QUALITY METRICS

The traditional approach in designing visual quality metrics is to build a model based on the components of the HVS. The reasoning is, that if we are able to model the HVS well enough, we can easily determine the influence of artifacts on the visual quality perception of human observers. The HVS, however, is quite complex as not only its physical properties must be considered, but also the psycho-visual processing. Therefore the HVS is unfortunately not understood well enough to model it completely<sup>7</sup>. In another, more engineering minded approach, the mathematical methods of data analysis are utilized to determine the relationship between objectively measurable features and the perceived subjective quality. The subjective quality is gained in subjective tests and the features represent properties of the video that correspond to some properties of the HVS. Note that we do not assume a a-priori model between features and quality, but rather determine this model by applying data analysis methods. One method already used successfully in some of our previous contributions<sup>2,3,5</sup> is the partial least squares regression (PLSR)<sup>8</sup>.

We can represent each frame by a  $1 \times k$  vector of  $k$  features. Considering we have  $n$  different video sequences, we get a  $n \times k$  matrix  $\mathbf{X}$  where the  $n$  rows denote different samples of our training set and the  $k$  columns different features as shown in Fig. 1a. Also we have for each sample the actual visual quality as determined in subjective tests in the  $n \times 1$  vector  $\mathbf{y}$ . We can then apply methods like PLSR to  $\mathbf{X}$  in order to develop a model.

One problem in this context is the temporal nature of video: most methods for data analysis assume that the data can be represented in a  $n \times k$  matrix. Hence, the temporal dimension  $t$  as shown in Fig. 1b must be temporally pooled. Usually this is done by averaging, but also other statistical pooling methods e.g. percentiles, standard deviation etc. may be employed<sup>5</sup>. In doing so, however, we ignore some of the variation in the temporal dimension, that is not captured by the pooling functions.

Therefore, we propose to use data analysis methods that do not only use a simple two dimensional  $n \times k$  matrix for the model building, but methods that are able to use the complete three dimensional  $n \times k \times t$  cube in the process. In a first step, we will use a multi-dimensional extension of the well known principal component regression (PCR), the 2D-PCR. Even though PLSR in general delivers better predictions<sup>8</sup> for unknown samples not included in the training set, we decided to use the more commonly used PCR in order to get a first impression, if the inclusion of the temporal domain in the model building step delivers an improvement.

### 3. PRINCIPAL COMPONENT REGRESSION AND ITS 2D EXTENSION

PCR describes the regression with principal components (PCs) gained by the PCA. As PCA is a well known method for data analysis, we will just shortly repeat some basics. For further information we refer to Jolliffe<sup>9</sup>.

Assume  $\mathbf{X}$  is an (centered)  $n \times k$  matrix. Also let  $r = \min\{n, k\}$  Using a singular value decomposition (SVD) we get the following factorization:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{P}^T = \mathbf{T}\mathbf{P}^T \quad (1)$$

where  $\mathbf{U}$  is an  $n \times r$  orthogonal matrix and  $\mathbf{P}$  is a  $k \times r$  orthogonal matrix.  $\mathbf{D}$  is a diagonal matrix.  $\mathbf{P}$  is in this context called the *loadings* matrix and its columns  $\mathbf{p}_1, \dots, \mathbf{p}_r$  are called loadings. They represent the eigenvectors of  $\mathbf{X}^T\mathbf{X}$  and correspond to the non-zero eigenvalues.  $\mathbf{D}$  contains the non-zero eigenvalues  $\lambda_1, \dots, \lambda_r$  in decreasing order. Furthermore we define the *scores* Matrix

$$\mathbf{T} = \mathbf{U}\mathbf{D} = \mathbf{X}\mathbf{P}. \quad (2)$$

The basic idea behind PCR is now to approximate  $\mathbf{X}$  by only using the first  $g$  columns of  $\mathbf{T}$  and  $\mathbf{P}$ , representing the  $g$  largest eigenvalues of  $\mathbf{X}^T\mathbf{X}$  and also the first  $g$  PCs. We hereby assume, that the combination of the largest  $g$  eigenvalues describe the variance in our data matrix  $\mathbf{X}$  sufficiently and that we can therefore discard the smaller eigenvalues.

We then perform a multiple linear regression of the selected first  $g$  PCs on the visual quality  $n$  column vector  $\mathbf{y}$  acquired in subjective test:

$$\mathbf{y} = \mathbf{T}_g\mathbf{c} + \mathbf{e}, \quad (3)$$

where  $\mathbf{T}_g$  represents a matrix with the first  $g$  columns of  $\mathbf{T}$ ,  $\mathbf{c}$  the (unknown) weight vector and  $\mathbf{e}$  the estimation error. Hence we determine with the least squares method the weights  $\hat{\mathbf{c}}$  as

$$\hat{\mathbf{c}} = . \quad (4)$$

With  $\mathbf{P}_g$  representing the matrix with the first  $g$  columns of  $\mathbf{P}$ , we can then define the regression vector  $\mathbf{b}$

$$\mathbf{b} = , \quad (5)$$

and predict an unknown sample with the (centered) feature vector  $\tilde{\mathbf{x}}$  according to

$$\hat{\mathbf{y}} = \tilde{\mathbf{x}}_u\mathbf{P}_g(\mathbf{T}_g^T\mathbf{T}_g)^{-1}\mathbf{T}_g^T\mathbf{y} \quad (6)$$

Note that  $\mathbf{X}$  is a two dimensional matrix and therefore we have to temporally pool the features extracted from the video sequences, as they are extracted frame-by-frame.

In order to avoid this temporal pooling, Yang et.al. proposed a different method to handle the additional dimension  $t$ , the two-dimensional PCA (2D-PCA)<sup>10</sup>. For a video sequence with  $T$  frames, we can slice our  $n \times k \times t$

video cube  $V$  into  $T$   $n \times k$  matrices or slices, where each slice represents one frame as illustrated in Fig. 1. Using a MatLab like notation, the  $i$ -th slice of  $V$  can be represented as  $V(:, :, i)$ . Assuming that all values are centered,

$$\mathbf{X}_{Sct} = \frac{1}{T} \sum_{i=1}^T \mathbf{X}(:, :, i)^T \mathbf{X}(:, :, i) \quad (7)$$

represents the covariance or *scatter* matrix. It describes the average covariance over the temporal dimension  $t$ . The SVD is now performed on  $\mathbf{X}_{Sct}$  instead of  $\mathbf{X}$ ; the rest is similar to the one dimensional PCR (1D-PCR) described above. For further information on the 2D-PCR see Rothbucher et.al.<sup>6</sup>.

#### 4. FEATURES

In this contribution we used seven different no-reference features in the design of the video quality metric: blockiness, blurriness, activity, predictability, motion continuity, edge continuity and color continuity. The first three features are intra features, describing aspects of the video quality with respect to one frame, and the later four are inter features, describing aspects of video quality with respect to changes between frames. These are all based on the assumption that human observers prefer smooth transitions between neighboring frames: predictability describing how well one frame can be predicted using only the previous frame, motion continuity measuring the smoothness of the motion, color continuity describing the color changes between two successive frames and edge continuity describing the change of edge regions between two successive frames. These features have already been used in<sup>2,3,5</sup>.

The blur measurement used is described in<sup>11</sup>. 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 i.e. motion blur this measurement is adjusted by a piecewise linear correction if the video contains high amount of fast motion. Blockiness is measured using the algorithm introduced in<sup>12</sup>. It calculates the horizontal and vertical blockiness by applying a Fourier transform along each line or column. The blockiness can be easily detected in the spectra by comparing the measured spectrum with a smoothed version of the spectrum. Blockiness should appear as peaks at distinct frequencies while the frequency spectrum of an image without blocks should be more smooth and in generally decrease with higher frequencies. For both blockiness and blur measurement it is sufficient to only take into account the luminance channel. The activity is determined by the amount of details. Therefore the turning points along each line and each row are calculated. Horizontal and vertical activity are then averaged to obtain one single value. This measurement is part of the BTFR metric included in<sup>13</sup>. 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. For simplicity this measurement again is performed on the luminance channel only.

For predictability, a image is generated by motion compensation with a simple block matching algorithm. The current image and its prediction are then compared block by block. A  $8 \times 8$  block is considered to be noticeable different if the SAD exceeds a experimentally determined threshold. In order to avoid that single pixels dominate the SAD measurement both images are filtered using a Gaussian blur filter, followed by median filtering. The output of this process is the percentage of blocks that are not noticeable different. The current image and its motion compensated prediction are also compared using the Edge-PSNR algorithm<sup>14</sup> to generate the edge-continuity measurement. This measurement should reflect how much the main structure of the image changes. Most objects should follow a relatively smooth motion trajectory. Reasons for non-smooth motion trajectories may either be chaotic object movement or artifacts like jitter. To detect if motion is continuous for two successive frames, 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. We calculate a color histogram with 51 bins for each RGB channel for the actual image and its prediction. Color continuity is then the linear correlation between those two histograms. This allows gradient changes in color e.g. due to illumination changes, but would point out artifacts like color bleeding.

The motion compensated images were created by using a simple block matching algorithm with a spiral search pattern<sup>15</sup>.

## 5. SUBJECTIVE TESTING

In order to have a sufficient amount of training data available, we conducted subjective tests. This section describes the testing methodology, the encoder scenarios, used test sequences used and present the results of our subjective tests.

### 5.1 TESTING METHODOLOGY

We used a variation of the standard DSCQS test method as proposed by Baroncini<sup>16</sup>. This Double Stimulus Unknown Reference (DSUR) test method differs from the standard DSCQS test method, as it splits a single basic test cell in two parts: the first repetition of the reference and the processed video is intended to allow the test subjects to identify the reference video. Only the repetition is used by the viewers to judge the quality of the processed video in comparison to the reference. The structure of a basic test cell is shown in Fig. 2.

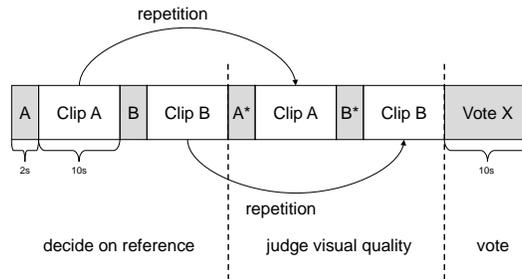


Figure 2: Basic test cell DSUR

A professional 24 inch LCD reference display with a native resolution of  $1920 \times 1080$  pixels was used (Cine-tal Cinemagé 2022). The decoded videos were converted to 4:2:2  $YC_B C_R$  for output to the reference display via a HD-SDI link. Only two viewers took part in the test at the same time due to the limited screen size. This was done in order to allow stable viewing conditions for all participants. All test subjects were screened for visual acuity and color blindness. The distance between the screen and the observers was three times the picture height (3H). The tests were performed in the video quality evaluation laboratory of the Institute for Data Processing at the Technische Universität München in a room compliant with recommendation ITU-R BT.500<sup>17</sup> as shown in Fig. 3.



Figure 3: Test room

To allow the test subjects to differentiate between relatively small quality differences, a discrete voting scale with eleven grades ranging from 0 to 10 was used. A short training was conducted before the test with three sequences of different content to the test, but similar rate points to the test. This resulted in a training session of ten sequences. During this training the test subjects had the opportunity to ask questions regarding the testing procedure. In order to verify if the test subjects were able to produce stable results, a small number of test cases were repeated during the test.

### 5.2 ENCODER SCENARIOS

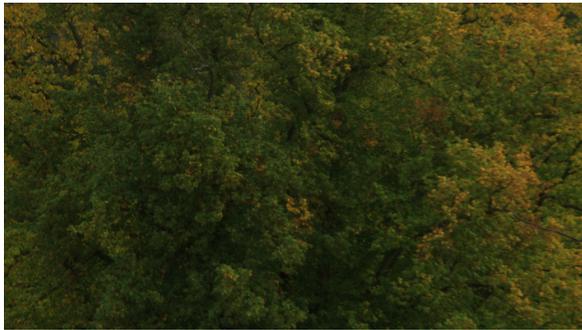
In order to take the performance of different coding technologies for HDTV with respect to the visual quality into account, we selected two different encoders:  $H.264/AVC$ <sup>18</sup> and  $Dirac$ <sup>19,20</sup>.



(a) CrowdRun



(b) ParkJoy



(c) InToTree



(d) OldTownCross

Figure 4: Test sequences from the SVT high definition multi format test set.

While H.264/AVC is the latest representative of the successful MPEG and ITU-T standards, Dirac is an alternative, wavelet based video codec. Its development was initiated by the British Broadcasting Cooperation (BBC) and was originally targeted at high definition resolution video material. Whereas it follows the common hybrid coding paradigm, it utilizes the wavelet transform instead of the usual block-based transforms e.g. DCT. Hence, it is not necessary for the transform step itself to divide the frame into separate blocks, but the complete frame can be mapped into the wavelet domain in one piece. This fundamental difference to the popular standards of the MPEG-family was also the main reason we chose Dirac as the representative of alternative coding technologies in this contribution. The Dirac reference software used in this contribution, version 0.7, offers a simplified selection of settings by only selecting the resolution and frame rate, instead of specific coding tools. Therefore, only the bitrate was varied.

For H.264/AVC, we used two significantly different encoder settings, each representing the complexity of various devices and services. The first setting is chosen to simulate a low complexity (LC) H.264/AVC encoder representative of standard devices: many tools that account for the high compression efficiency are disabled. In contrast to this, we also used a high complexity (HC) setting that aims at getting the maximum possible quality out of this coding technology, representing sophisticated broadcasting encoders. We used the reference software<sup>21</sup> of H.264/AVC, version 12.4. The difference in computational complexity is also shown by the average encoding time per frame: 34 and 201 seconds per frame for the LC and the HC H.264/AVC version, respectively. Selected encoding settings for H.264/AVC are given in Table 1.

The artifacts introduced into the videos by this encoding scheme include pumping effects i.e. periodically changing quality, a typical result of rate control problems, obviously visible blocking, blurring or ringing artifacts, flicker and similar effects. We selected four bitrates from 5.4 Mbit/s to 30 Mbit/s representing a range of real life bitrates. An overview of the sequences and bitrates is given in Table 2.

The test sequences were chosen from the SVT high definition multi format test set<sup>22</sup> with a spatial resolution of  $1920 \times 1080$  pixels and a frame rate of 25 fps. The used sequences are shown in Fig.4: *CrowdRun*, *ParkJoy*, *InToTree* and *OldTownCross*.

Table 1: Selected encoder settings for H.264/AVC

	LC	HC
Encoder	JM 12.4	
Profile&Level	Main, 4.0	High, 5.0
Reference Frames	2	5
R/D Optimization	Fast Mode	On
Search Range	32	128
B-Frames	2	5
Hierarchical Encoding	On	On
Temporal Levels	2	4
Intra Period	1 second	
Deblocking	On	On
8x8 Transform	Off	On

They represent different coding complexity: InToTree and OldTownCross are encoded rather easily, CrowdRun and ParkJoy are encoded more difficultly due to the large number of moving objects in them. Each of those videos was encoded at the selected four different bitrates. This results in a quality range from ‘not acceptable’ to ‘perfect’.

### 5.3 SUBJECTIVE TESTING RESULTS

A total of 19 subjects participated in the test, all students with no or little experience in video coding aged 20-30. Processing of outlier votes was done according to<sup>17</sup> and the votes of one test subject were removed based on this procedure. The 95% confidence intervals of the subjective votes are below 0.7 on a scale between 0 and 10 for all single test cases, the mean 95% confidence interval is 0.4. We determined the mean opinion score (MOS) by averaging all valid votes for each test case.

An overview of the results of the subjective results are given in Fig. 5. The results were also used partly in developing the visual quality metrics proposed in<sup>2,5</sup> and have also been presented more detailed in<sup>23</sup>.

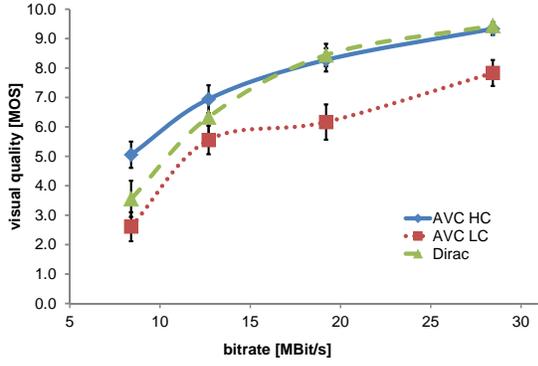
## 6. RESULTS

We applied the 2D-PCR to the subjective testing results presented in Section 5.3 and the extracted features according to Section 4. For comparison, we also applied 1D-PCR to the data. In order to get a realistic measure of the performance on unknown video sequences, we used a cross validation approach i.e. we excluded one test sequence during the model building and tested then the generated model on the left out sequence.

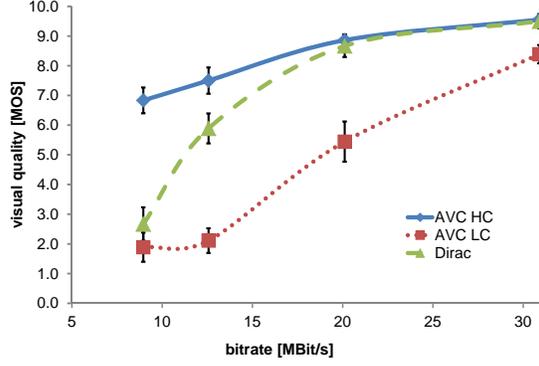
To compare the prediction quality, we used the Pearson correlation and the Spearman rank order correlation. The Pearson correlation (8) gives an indication about the prediction accuracy of the model and the Spearman rank order correlation (9) gives an indication how much the ranking between the sequences changes between the predicted quality and the real subjective quality. The Pearson correlation is defined as follows:

Table 2: Video sequences

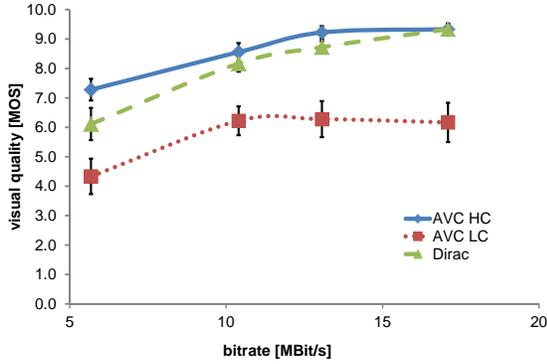
Sequence	Frame Rate	Bitrate [MBit/s]
CrowdRun	25 fps	8.4 / 12.7 / 19.2 / 28.5
IntoTree	25 fps	5.7 / 10.4 / 13.1 / 17.1
OldtownCross	25 fps	5.4 / 9.6 / 13.7 / 19.0
ParkJoy	25 fps	9.0 / 12.6 / 20.1 / 30.9



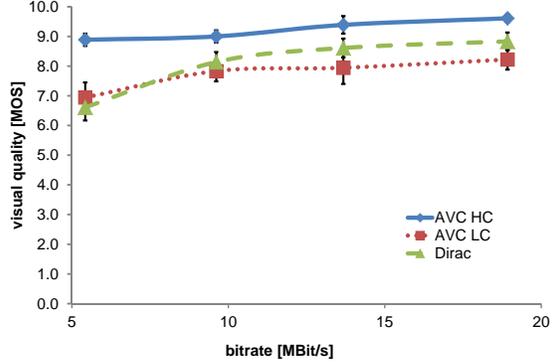
(a) CrowdRun



(b) ParkJoy



(c) InToTree



(d) OldTownCross

Figure 5: Results of the subjective tests: visual quality at bitrates from 5.4 Mbit/s to 30 Mbit/s for different video sequences and encoders. The whiskers represent the 95% confidence intervals of the subjective test results for the visual quality.

$$r^p = \frac{\sum_k (q_k - \bar{q})(MOS_k - \overline{MOS})}{\sqrt{\sum_k (q_k - \bar{q})^2} \sqrt{\sum_k (MOS_k - \overline{MOS})^2}}, \quad (8)$$

where  $q_k$  is the predicted quality and  $\bar{q}$  is the average over all predictions.  $MOS_k$  and  $\overline{MOS}$  are the respective subjective quality. For the Spearman rank order correlation  $r^s$ ,  $\chi_k$  is the rank of  $q_k$  and  $\gamma_k$  is the rank of the subjective value  $MOS_k$ .  $\bar{\chi}$  and  $\bar{\gamma}$  are the average of the respective ranks:

$$r^s = \frac{\sum_k (\chi_k - \bar{\chi})(\gamma_k - \bar{\gamma})}{\sqrt{\sum_k (\chi_k - \bar{\chi})^2} \sqrt{\sum_k (\gamma_k - \bar{\gamma})^2}}. \quad (9)$$

The results are shown in Table 3 and Fig. 6. For both 1D-PCR and 2D-PCR we selected the first two PCs as these already explained on average 93% and 99% of the variance in the training sets, respectively.

We can see that 2D-PCR delivers a better overall quality prediction compared to 1D-PCR with a Pearson correlation coefficient of 0.47 compared to 0.37. Also 2D-PCR is able to estimate unknown video content better as shown in Table 3 by the different correlation coefficients for validation with only the training set and with a proper cross validation. This seems to indicate that the inclusion of the temporal variance into the model

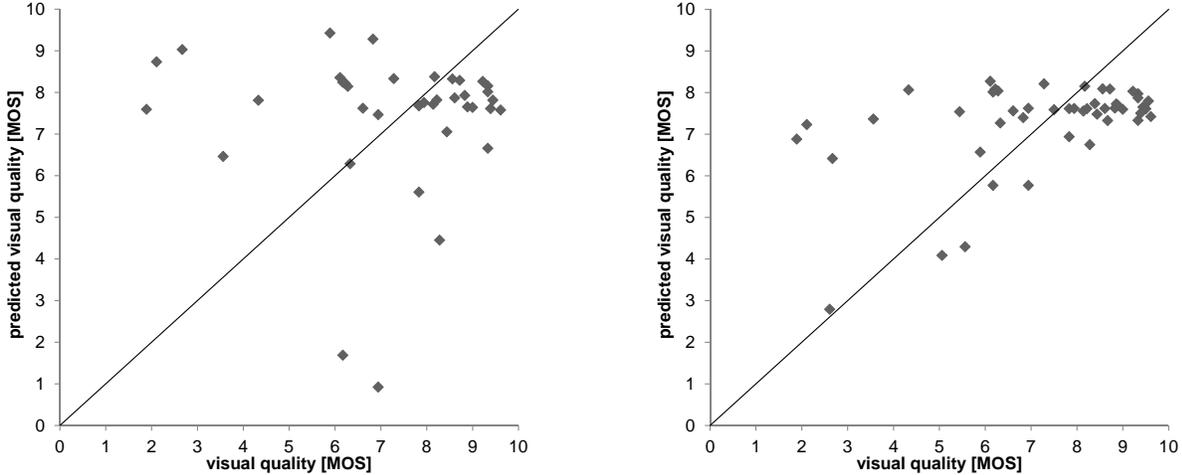


Figure 6: Cross validation prediction results for 1D-PCR (left) and 2D-PCR (right)

building step with 2D-PCR not only improves the prediction quality, but also the ability of the generated model to handle unknown content. In general, for higher visual quality our 2D-PCR model underestimates, for lower visual quality it overestimates the quality of the video sequences.

Nevertheless, we must note that the overall prediction quality is rather bad compared to other contributions<sup>2,3,5</sup>, where usually a correlation of higher than 0.80 has been achieved for no-reference video quality metrics with other data analysis methods, particularly PLSR. We should keep in mind, however, that in these contributions additional corrections e.g. sigmoid correction are applied to the quality prediction of the model in order to improve the prediction accuracy. Also 1D-PCR, and by extension therefore 2D-PCR, are known to be often inferior to PLSR, if the aim of the model is primarily to predict unknown samples, not included in the original training set<sup>8</sup>.

## 7. CONCLUSION

We presented the application of an extended version of the PCR, the 2D-PCR, to the design of a no-reference video quality metric in this contribution. In doing so, we exploited the temporal nature of video more appropriately, as the temporal variance of video was included in the model building step and no previous temporal pooling had to be applied. The results of the cross validation show, that 2D-PCR outperforms standard one-dimensional PCR with a higher prediction accuracy for unknown video sequences.

Although the overall prediction quality of 2D-PCR in this contribution does not achieve the same level as other no-reference video quality metrics, we believe that the direct inclusion of the temporal domain of video in the model building step will improve the quality prediction abilities of metrics. Also we should keep in mind, that the metric presented in this contribution is basically just a simple regression model, and many additional methods to improve the predication quality have not been included in the proposed metric.

Therefore, we intend to enhance this approach in future work with existing methods as presented in our other contributions and thus achieve an overall improvement of the prediction accuracy of no reference video quality metrics.

Table 3: Prediction results

	Pearson Linear Correlation		Spearman Rank Order Correlation	
	cross validation	no cross validation	cross validation	no cross validation
1D-PCR	0.37	0.63	0.19	0.56
2D-PCR	0.47	0.58	0.37	0.45

## REFERENCES

- [1] Miyahara, M., “Quality assessments for visual service,” *IEEE Communications Magazine* **26**(10), 51–60 (1988).
- [2] Keimel, C., Oelbaum, T., and Diepold, K., “No-reference video quality evaluation for high-definition video,” *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* , 1145–1148 (Apr. 2009).
- [3] Oelbaum, T., Keimel, C., and Diepold, K., “Rule-based no-reference video quality evaluation using additionally coded videos,” *Selected Topics in Signal Processing, IEEE Journal of* **3**, 294–303 (April 2009).
- [4] Keimel, C. and Diepold, K., “Improving the prediction accuracy of PSNR by simple temporal pooling,” in [*Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2010*], (Jan. 2010).
- [5] Keimel, C., Oelbaum, T., and Diepold, K., “Improving the prediction accuracy of video quality metrics.,” *Acoustics, Speech and Signal Processing, 2010. ICASSP 2010. IEEE International Conference on* , 2442–2445 (Mar. 2010).
- [6] Rothbucher, M., Durkovic, M., Shen, H., and Diepold, K., “HRTF customization using multiway array analysis,” in [*18th European Signal Processing Conference, EUSIPCO 2010*], 229–233 (Aug. 2010).
- [7] Winkler, S., [*Digital Video Quality: Vision Models and Metrics*], Wiley (2005).
- [8] Martens, H. and Martens, M., [*Multivariate Analysis of Quality*], Wiley & Sons (2001).
- [9] Jolliffe, I., [*Principal Component Analysis*], Springer (2002).
- [10] Yang, J., Zhang, D., Frangi, A., and yu Yang, J., “Two-dimensional pca: a new approach to appearance-based face representation and recognition,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **26**(1), 131–137 (2004).
- [11] Marziliano, P., Dufaux, F., Winkler, S., and Ebrahimi, T., “A no-reference perceptual blur metric,” in [*Proc. IEEE International Conference on Image Processing (ICIP)*], **3**, 57–60 (Sept. 2002).
- [12] Wang, Z., Bovik, A. C., and Evans, B. L., “Blind measurement of blocking artifacts in images,” in [*Proc. IEEE International Conference on Image Processing (ICIP)*], **3**, 981–984 (Oct. 2000).
- [13] “ITU-T J.144. objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference,” (Mar. 2004).
- [14] Lee, C., Cho, S., Choe, J., Jeong, T., Ahn, W., and Lee, E., “Objective video quality assessment,” *SPIE Optical Engineering* **45**, 7004 (Jan. 2006).
- [15] Zahariadis, T. and Kalivas, D., “Fast algorithms for the estimation of block motion vectors,” in [*Proceedings of the Third IEEE International Conference on Electronics, Circuits, and Systems, ICECS*], 716–719 (Oct. 1996).
- [16] Baroncini, V., “New tendencies in subjective video quality evaluation,” *IEICE Transaction Fundamentals* **E89-A**, 2933–2937 (Nov. 2006).
- [17] “ITU-R BT.500 methodology for the subjective assessment of the quality for television pictures,” (June 2002).
- [18] “ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG4-AVC), advanced video coding for generic audiovisual services,” (July 2005).
- [19] Borer, T., Davies, T., and Suraparaju, A., “Dirac video compression,” Tech. Rep. WHP 124, BBC Research & Development (Sept. 2005).
- [20] Borer, T. and Davies, T., “Dirac - video compression using open technology,” Tech. Rep. WHP 117, BBC Research & Development (July 2005).
- [21] Sühning, K., “H.264/AVC software coordination,” (2007).
- [22] SVT, “The SVT high definition multi format test set,” (Feb. 2006).
- [23] Keimel, C., Habigt, J., Habigt, T., Rothbucher, M., and Diepold, K., “Visual quality of current coding technologies at high definition IPTV bitrates,” in [*Multimedia Signal Processing 2010. IEEE International Workshop on*], 390–394 (Oct. 2010).