A regression method for real-time video quality evaluation

Citation for published version (APA):

Torres Vega, M., Mocanu, D. C., & Liotta, A. (2016). A regression method for real-time video quality evaluation. In B. Abdulrazak, E. Pardede, M. Steinbauer, I. Khalil, & G. Anderst-Kotsis (Eds.), *MoMM '16 : Proceedings of the 14th International Conference on Advances in Mobile Computing and Multimedia, 28-30 November 2016, Singapore, Singapore* (pp. 217-224). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3007120.3007125>

DOI:

[10.1145/3007120.3007125](https://doi.org/10.1145/3007120.3007125)

Document status and date:

Published: 29/11/2016

Document Version:

Typeset version in publisher's lay-out, without final page, issue and volume numbers

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

A Regression Method for real-time video quality evaluation

Maria Torres Vega
Eindhoven University of
Technology
The Netherlands
m.torres.vega@tue.nl

Decebal Constantin
Mocanu
Eindhoven University of
Technology
The Netherlands
d.c.mocanu@tue.nl

Antonio Liotta
Eindhoven University of
Technology
The Netherlands
a.liotta@tue.nl

ABSTRACT

No-Reference (NR) metrics provide a mechanism to assess video quality in an ever-growing wireless network. Their low computational complexity and functional characteristics make them the primary choice when it comes to real-time content management and mobile streaming control. Unfortunately, common NR metrics suffer from poor accuracy, particularly in network-impaired video streams. In this work, we introduce a regression-based video quality metric that is simple enough for real-time computation on thin clients, and comparably as accurate as state-of-the-art Full-Reference (FR) metrics, which are functionally and computationally infeasible in real-time streaming. We benchmark our metric against the FR metric VQM (Video Quality Metric), finding a very strong correlation factor.

Keywords

Quality of Experience, No-Reference Video Quality, Regression Processes, Cognitive QoE Assessment Metrics

1. INTRODUCTION

In the context of Quality of Experience (QoE) [14], No-Reference (NR) metrics provide a new venue for video quality assessment in situations when either real-time measurement (e.g., for real time network management), or low-computation (e.g., in the mobile-streaming context) is required. In these cases, all other conventional approaches (i.e., subjective studies and full-reference, FR, methods) are unsuitable, due to complexity-, computation- and functional constraints.

On the other hand, NR metrics still present unresolved issues including lack of general validity and poor correlation to human perception. These are particularly evident when trying to assess video streams that have been distorted by real network impairments, as we have argued in our previous work [37, 38]. Thus, deriving a method capable of achieving the same accuracy as FR metrics (also in the presence of network impairments) while keeping the NR characteristics (i.e., functional features and computational speed), makes a breakthrough in video quality assessment. In this direction, we propose a regression-based video quality metric that

is simple enough for real-time computation on thin clients, and comparably as accurate as the state-of-the-art FR VQM (Video Quality Metric).

Video Quality Assessment (VQA) is nowadays in most of the cases done by means of a variety of QoE methods [14]. Being linked to human perception, QoE is intrinsically a subjective matter. Thus, it would only be natural to employ subjective analysis (through subjective methods) [32]. In this case, human subjects rate presented stimuli (for example impaired video sequences) under controlled conditions [2, 1], obtaining what is typically described as the Mean Opinion Score (MOS). However, due to the time-consuming nature and bias of subjective experiments [22], in the last years, great effort has been placed onto developing objective quality metrics, with the aim to automate QoE assessment through objective QoE (oQoE) algorithms [34]. The ultimate goal of oQoE metrics is to provide the best possible correlation to subjective studies and to the human vision system (HVS), by comparing the reference (i.e., the original video) to the received material (i.e. the video distorted by the streaming and transmission processes).

Thus, the first method of choice to automate and complement the subjective assessment has been based on FR metrics. Typical FR metrics include Peak Signal to Noise Ratio (PSNR), based on the frame's Mean Square Error (MSE) value, and Structural Similarity (SSIM) [40], that combines luminance, contrast and structure measurements. Specifically because of its good correlation with subjective values, the Video Quality Metric (VQM) [26] is commonly used as the FR benchmark metric [5].

However, while subjective studies and FR objective metrics succeed in providing relatively accurate assessments, their time, complexity and computational requirements make them unfeasible when it comes to assessing real-time streaming systems and, particularly, when mobile devices are scrutinized. In this case, there are just too many conditions to test out (which rules out subjective methods) and the original, undistorted signal is not available (which rules out FR methods).

To this end, NR metrics aim to estimate quality degradation by measuring specific features on the received distorted signal, without making reference to the original signal. Simple NR metrics belong to two classes, operating at bitstream level [4] or at the pixel (frame) level [7, 3], respectively. NR metrics have, however, a limited scope as they usually focus on specific artifacts. In order to improve their generality, several state-of-the-art approaches have looked at machine learning (ML) techniques. Examples of that are the bitstream based artificial neural network approach of Shahid

Pre-printed version.

Please cite as:

M. Torres Vega, D. C. Mocanu and A. Liotta:
A Regression Method for real-time video quality evaluation.
in *proc. of the 14th International Conference on Advances in Mobile Computing and Multimedia (MoMM2016)*,
28-30 November 2016, Singapore (Singapore).

et al. [31] or the LASSO algorithm of Pandremmenou et al. [24]. While showing promising results, these metrics are evaluated on synthetically impaired videos and there is little evidence that these would be effective in typical video streaming scenarios. In fact, our experimental survey of NR metrics unveiled how existing methods are only partly effective in detecting quality degradation due to compression, but fail to fully assess other distortions derived from network impairments [38, 37]. In this way, while most of the features studied appear to have certain operational boundaries, it was not possible to pin-point one metric which could be used as a general substitute to the FR benchmark.

In this work, we introduce a regression-based video quality metric that is simple enough for real-time computation on thin clients, and comparably as accurate as state-of-the-art FR metrics, which are functionally and computationally unviable in real-time streaming. We benchmark our metric against the FR metric VQM (Video Quality Metric), finding a very strong correlation factor.

The remainder of this paper is organized as follows. Section 2, provides a description of the state-of-the-art on NR metrics and machine learning. In Section 3, the proposed regression quality metric is presented. Evaluation set-up and results are presented in Section 4. Finally, Section 5 draws conclusions, highlighting key contributions and suggesting directions for future work.

2. BACKGROUND

Reibman et al. [29] classified NR approaches as either stemming from statistics derived from pixel-based features, NR pixel (NR-P), or computed directly from the coded bitstream, NR bitstream (NR-B). In a more recent classification, Shahid et al. added to this classification a third category in which approaches combining both pixel and bitstream assessments are included, i.e. Hybrid NR-P-B metrics [32].

NR-P methods have focused their attention on certain artifacts, related to a specific type of visual quality degradation. Blur, noise, blockiness or temporal impairments have been quantified for measuring the end-user's quality. Blur, measured frame by frame, appears as a loss of spatial detail and a reduction of edge sharpness [39]. Examples of blur based NR video quality assessment can be found in [7] and [8]. Noise has also been used to assess quality, like in the block-based approach of Rank et al. [28]. Blockiness manifests itself as a discontinuity between adjacent blocks in images and video frames [11]. Several research lines have focused on blockiness. Examples of that are the Block-Edge metric of Wu et al. [41] or the HVS-based blocking method of Liu et al. [17]. Finally, temporal impairments incur by the network through delay or packet loss. These result in a degradation of the video in the form of jerkiness (non-fluent and non-smooth presentation of frames) [3], frame freezes or jitter. Due to the fact that videos can be affected by more than one artifact at a time, methods combining contributions of different artifacts appeared, such as the linear combination of noise and blur components of Choi et al. [6].

NR-B methods are relatively simpler to compute and quality scores can be obtained in the absence of the full decoder. But, they tend to have a limited scope of application as

they are usually designed for specific coding techniques, for example H.264/AVC [4]. Another issue is to find the correlation between the bitlayer parameters and quality, without increasing the complexity.

The performance of NR-B metrics can be enhanced by adding some input form of NR-P based quality assessment. These methods are called hybrid methods. Shanableh et al. [33] suggest a multipass prediction system based on step-wise regression using features included in the coding information of a Macro Block (MB), some relative measures of motion vector of neighboring MBs, and some numerical values related to textures of the MB. This method has shown good correlation to SSIM but its complexity makes it not quite fit for online analysis. Another interesting approach is the one provided by Keimel et al. [13]. They measure quality by linearly combining bitstream and pixel related features. Despite substantially reducing complexity, their approach does not, however, correlate well with either PSNR (Peak Signal to Noise Ratio) or SSIM, two of the state-of-the-art objective metrics in current use.

Improving the performance of the NR metrics has also been done by exploring the possibilities of machine learning techniques [20]. Already in 2002, Gastaldo et al. introduced one of the first methods to estimate the video quality using artificial neural networks [10]. They proposed the use of circular back propagation networks (based on bitstream layer parameters) in order to mimic the user's perception of compressed MPEG2 videos. Their approach showed promising results on a 12-video dataset from the motion picture expert group (MPEG). Their study focused on video distortions deriving merely from compression and explored a specific machine learning method. Also working on compressed videos, Le Callet et al. [15] employed an interesting convolutional neural network as a Reduced Reference (RR) method to allow a continuous-time quality estimation and scoring of the video. Unlike our NR approach, in which the server transmits the machine learning model updates only on service launch and in the case that an update is due, their method (as any RR metric) requires the transmission of features extracted from the original video together with the video under scrutiny. Staelens et al. [35] presented an NR video quality estimation method which uses a symbolic regression framework trained on a large set of parameters extracted from the codec. While obtaining good correlation with subjective tests, their approach is suited only to H.264 compressed streams, thus loosing on generality. Shahid et al. [31] proposed a model combining different bitstream-layer features using an Artificial Neural Network to estimate the quality. They tested their method on compressed videos but focused on correlations with PSNR. Pandremmenou et al. [24] employed the Least Absolute Shrinkage and Selection Operator (LASSO) regression method for assessing the accuracy of bitstream parameters to FR metrics and subjective analysis in videos affected by compression and synthetic impairments. In our previous research we developed a lightweight algorithm combining bitstream parameters (video bitrate, complexity and motion) with pixel artifacts (blur and noise) [36]. We presented the machine learning-based algorithm showing high correlation with SSIM.

3. REGRESSION-BASED VIDEO QUALITY ASSESSMENT METHOD

In this section, we present our regression-based NR video quality method. Figure 1, shows the block diagrams for the processes running, respectively, on the server side and in the clients.

Figure 1: Regression based video quality method diagram.

The performance of a predictive model will substantially depend on the characteristics of the dataset used for training. In the case of a video service provider, the training set is composed by a number of video type samples stored in the server. In our method, each sample in the training set includes eight NR features (both in the pixel and the bitstream layers), two network condition parameters (packet loss rate and bitrate) and the ground truth quality index. Without loss of generality, we used VQM in the present study, which has shown to perform well even in the case of video streams distorted by the network transmission. This training set is used (in the server) to maintain the quality regression function, which is then employed on the client side to compute our predictive NR video quality assessment metric in real-time.

At service launch, the service provider will already have a representative video types set (e.g., sport, action movies, cartoons, and so forth); thus an initial prediction regression model is constructed (and made available to the client side). When a completely new video type is added, the prediction model may become less accurate. Yet, over the time the model will be updated based on new types and, what is more important, the chances of getting new video types will rapidly diminish. In this way, the server runs a process in the background in which the regression model is trained with the available video samples and new models (R) are uploaded to the clients (on a continuous or periodic basis).

On the other end of the transmission link, the video client employs the regression model trained by the server, to generate its prediction-based quality metric (Q_{reg}). During a streaming session, the client characterizes the incoming video in terms of NR features and real-time network conditions, matching this information against the prediction model, to generate the quality estimation.

Selecting features that will characterize a video stream most accurately, furthermore providing with a decision suited

to the degradation of the quality, and conforming an adequate set of characteristics that would better describe the video, is a very sensitive task. Thanks to the study performed in our previous work [38], we found that NR metrics are good at picking different features, under specific conditions. This motivated us to develop a hybrid, predictive metric starting from the eight common metrics (Figure 1).

A video stream can be characterized by several parameters, each one affecting different video types in diverse ways. Parameters regarding the video scene composition have been demonstrated to affect quality to a large extent [16]. From these parameters, the scene complexity and the video motion have proven to give high correlations with video quality [12]. The scene complexity is defined as the number of objects or elements present in the frame and the video motion as the amount of movement in the video [12]. Both these features can be empirically obtained from the encoding [16].

On the pixel level, noise and blur components (mean and ratio per feature) have been demonstrated to provide a good measure of degradations in a frame-by-frame assessment [6]. In the same way, blockiness [25, 41], described as a discontinuity between adjacent blocks in images and video frames [11], was shown in our earlier study to provide promising results [38]. Finally, measuring the inter-frame degradations becomes fundamental in the presence of network impaired video. In this way, temporal features such as the Jerkiness (non-fluent and non-smooth presentation of frames) become fundamental [3].

These eight NR features are averaged for the whole video and normalized between 0 and 1. More information on how to compute these metrics can be found in [37, 38].

Table 1: Video dataset parameters range in terms of video types, compression and network packet loss ratio (960 samples in total).

Video type	Compression	Packet loss
Blue Sky (bs1)	64kbps	PL0%
Mobile Calendar (mc1)	640kbps	PL0.5%
Pedestrian Area (pa1)	768kbps	PL1%
Park Run (pr1)	1024kbps	PL1.5%
River Bed (rb1)	2048kbps	PL2%
Rush Hour (rh1)	3072kbps	PL2.5%
Sunflower (sf1)	4096kbps	PL3%
Shields (sh1)	5120kbps	PL3.5%
Station (st1)		PL4%
Tractor (tr1)		PL4.5%
		PL5%
		PL10%

4. EVALUATION

In this section we present the results of the evaluation of our regression-based approach. First, Section 4.1 presents the evaluation set-up and video dataset. It provides details on the regression models used to assess the performance of our approach. Then, in Section 4.2 the results of our analysis are presented.

4.1 Evaluation methodology, video dataset and learning conditions

The dataset used is characterized in Table 1. It consists of 10 original raw, 10 seconds, 25fps video types (bs1, mc1, pa1, pr1, rb1, rh1, sf1, sh1, st1, tr1) from the Live Quality Video Database [30], encoded at MPEG4 part 10/H.264

Table 2: Parameters used for the different machine learning techniques.

Technique	Acronym	Parameters
Multiple Linear Regression	LR	Added bias
Standard Regression Tree	RT	type:binary N. Branches>15
Gaussian Process Regression	GPR	Method: exact Basis: constant Kernel: squaredexponential

to 8 bitrates levels (64, 640, 768, 1024, 2048, 3042, 4096, and 5120 kpbs) and with a resolution of 768x432. The selection of the encoding bitrates has been done in a way as to obtain the most diverse variety of video qualities. These 80 (10 videos at 8 bitrates) videos were then streamed in a controlled real network environment (given by a network emulator) and subjected to 12 levels of packet loss (0, 0.5%, 1%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5%, and 10%), which is the most impairing condition in networks [23]. This makes a total of 960 different videos that we use to benchmark our quality metric. In [37, 38] we showed the broad range of characteristics and behaviours that this video set presents, both in terms of the simple NR metrics and the FR benchmark. For example, while some videos, such as the Sunflower (sf1) or the Rush hour (rh1), reach maximum benchmark quality (VQM) at bitrates of 3 or 4 Mbps, others like the Park run (pr1), the River bed (rb1) or the Shields (sh1) are still far from the maximum quality index (1) encoded at 5Mbps. This variety makes the evaluation of our method stronger and our conclusions more general and applicable to an immense range of videos and conditions.

Given the broad variety of machine learning approaches in the literature, an important element of our work was to explore different algorithms and find suitable avenues. To this end, our experimental framework is sufficiently generic to perform tests on any type of regression-based supervised-learning algorithms (we have not included unsupervised learning methods in our study). For our metric, we selected three possible regression models broadly used in machine learning.

One of the most known and simplest regression models is linear regression [9] (dubbed LR in our analysis), which attempts to model the relationship between a scalar (output) and one or more independent variables by means of a linear multidimensional model of the input data.

Decision trees learning uses a decision tree as a predictive model, which maps observations about an item to conclusions about the item’s target value [27]. Within decision trees, in regression trees (dubbed RT) the target variable takes continuous values (typically real numbers).

The Gaussian Process Regression (or Kriging) [18] provides an example of higher complexity regression models. The basic idea of Kriging is to predict values by means of interpolation in which the interpolated values are modeled by a Gaussian process governed by prior covariances of the training data.

These three regression methods have been selected both for their simplicity and for their demonstrated good performance in quality measurements [6, 21]. We implemented

these methods based on the ML Matlab toolbox [19]. Each algorithm requires the tuning of certain parameters in order to optimize their performance. The values included in Table 2 (third column), have been found to perform the best. In order to perform the Multiple Linear Regression, we added a bias vector (a vector of all ones) to the input data.

Another important choice in performing machine learning experiments consists on the way the training set is picked out of the whole dataset. The method used is bound to have a sensitive effect on the performance of the prediction models and, ultimately, on the accuracy of the NR metric. In order to evaluate our algorithm we divided the data-set on 10 parts corresponding to each of the video types and we evaluated the performance of the algorithm via a 10-folded cross-validation.

4.2 Results

When a new video is made available in the content provider’s server, it is possible that the server administrator does not yet have video traces of it and, thus, cannot or does not want to re-train the ML model. In this analysis we explored this case which corresponds to a blind prediction approach, i.e. predicting the quality of one full video type based on a model trained with the other 9 types.

For each of the variants of our algorithm (LR, RT and GPR), we trained the model with 9 videos and tested on the remaining one; this adds up to 10 iterations per variant. Table 3 (columns 9, 10 and 11) shows the results achieved by our regression-based approach, including the LR, RT and GPR regression models, respectively. To have a qualitative comparison between our approach, and the eight NR metrics whose accuracy we aimed to improve, in Table 3 we also show the accuracy of the simple NR features (columns 1 to 8). Furthermore, we include SSIM in the comparison, to provide the lightweight FR counterpart accuracy (column 12). The accuracy of each of the metrics is computed as a Pearson Correlation to VQM, which we demonstrated to provide an accurate analysis for this dataset in [38]. Each of the rows shows the results for each of the video types; whereas the last row shows the overall correlation of the whole dataset to VQM.

As it can be seen in Table 3, for all the video cases, the regression-based models achieve an overall average accuracy of roughly 80% (LR achieves 76% and the RT achieves an 83%). This is a significant improvement compared to the classic NR metrics, which achieve correlations to VQM lower than 30%. What is even more striking, our NR metrics achieve an overall performance that is in one of its variants comparable to the FR metric SSIM (LR-based metric) and in the other two cases is over a 5% better (RT and GPR variants).

The reason why standard deviations tend to be relatively high, reflects a broad range of video types (row 1 to 10), which was crucial to providing a comprehensive study. Interesting though, one of our regression methods (RT) gives the smallest deviation of all the metrics (0.084). The reason for this comes from the characteristics of the regression trees, which even in their simplest variants are able to classify data patterns in cases where other more sophisticated techniques fail (such as the GPR). This indicates that the RT approach combined with the NR metrics conform a very powerful NR tool.

Table 3: PCC correlations to VQM of the eight NR original metrics, our regression approach using three variants (LR, RT and GPR) and SSIM. Cell colors give qualitative correlation levels: green (best), yellow (median), and red (worst).

Type	CX	MO	NM	NR	BM	BR	BL	JE	LR-NR	RT-NR	GPR-NR	SSIM
bs1	0.16831	0.01102	-0.48789	0.118084	-0.013234	-0.637055	0.439352	-0.701313	0.812965	0.85936	0.831816	0.73489
mc1	0.6632	0.017724	-0.6442	0.538208	-0.065169	-0.813207	0.084655	0.36792	0.883728	0.842941	0.8668	0.903
pal	0.2915	-0.02809	0.645745	0.456716	0.110438	0.464662	0.441793	0.05716	0.871275	0.95522	0.934233	0.8829
pr1	0.30362	-0.164135	-0.70394	-0.121612	0.009715	-0.200413	0.496992	0.6075	0.89029	0.690704	0.868404	0.68772
rb1	0.5331	0.570433	0.4320174	0.51387	0.546452	0.44004	0.200401	-0.594002	0.886588	0.706741	0.808595	0.25549
rh1	0.39129	-0.474693	0.165556	0.320134	0.351533	0.369301	-0.686387	-0.670956	-0.315646	0.78308	-0.117229	0.9068
sfl	0.41274	-0.41412	-0.7285	0.13638	0.516196	0.41971	0.552045	-0.41496	0.943491	0.929498	0.954198	0.83992
sh1	0.41275	-0.0925	-0.35164	0.468456	0.215916	-0.7191	0.469199	0.53221	0.848927	0.828483	0.0.926687	0.8731
st1	0.4689	-0.328997	-0.646985	0.350274	0.437243	-0.208797	0.632241	-0.267012	0.935718	0.858198	0.963359	0.7554
tr1	0.53072	-0.17782	0.087274	0.73765	0.511014	0.157158	0.307524	0.581363	0.923678	0.859136	0.975721	0.88523
All	0.41761	-0.108116	-0.22325	0.351816	0.262	-0.0732666	0.2937816	-0.0502088	0.768101	0.830436	0.801259	0.7725013
	± 0.1345	± 0.27962	± 0.4859734	± 0.23678	± 0.227552	± 0.48674	± 0.36191	± 0.51401	± 0.382840	± 0.084724	± 0.327775	± 0.1873

(a) Correlation maps for video bs1 with the 8 NR features, our regression method with three variations and SSIM. (b) Correlation maps for video mc1 with the 8 NR features, our regression method with three variations and SSIM.

(c) Correlation maps for video pal with the 8 NR features, our regression method with three variations and SSIM. (d) Correlation maps for video pr1 with the 8 NR features, our regression method with three variations and SSIM.

Figure 2: Pearson correlation to VQM of the 8 NR features, SSIM and our regression model in its three variations (LR, RT and GPR), considering bitrates between 64 and 5,120 Kbps and packet losses between 0 and 10%. Video types: (a) Blue sky (bs1); (b) Mobile calendar (bs1); (c) Pedestrian area (pa1); and (d) Park run (pr1).

(a) Correlation maps for video rb1 with the 8 NR features, our regression method with three variations and SSIM. (b) Correlation maps for video rh1 with the 8 NR features, our regression method with three variations and SSIM.

(c) Correlation maps for video sf1 with the 8 NR features, our regression method with three variations and SSIM. (d) Correlation maps for video sh1 with the 8 NR features, our regression method with three variations and SSIM.

(e) Correlation maps for video st1 with the 8 NR features, our regression method with three variations and SSIM. (f) Correlation maps for video tr1 with the 8 NR features, our regression method with three variations and SSIM.

Figure 3: Pearson correlation to VQM of the 8 NR features, SSIM and our regression model in its three variations (LR, RT and GPR), considering bitrates between 64 and 5,120 Kbps and packet losses between 0 and 10%. Video types: (a) River bed (rb1); and (b) Rush hour (rh1); (c) Sun flower (sf1); (d) Shields (sh1); (e) Station (st1); and (f) Tractor (tr1).

Next, we studied in greater detail the working range of each of the different metrics, including the eight classical NR metrics, SSIM, and our three regression metrics. A sample of test cases is captured in the colormaps of Figures 3 and 2. Each colormap shows correlation values to VQM, considering different video types and a range of bitrates (y-axis) and packet-loss values (x-axis). Full correlation to VQM (100% accuracy) is denoted in dark blue, whereas full anti-correlation to VQM is indicated in red. Looking at the color patterns, it is clearly visible how our metrics consistently outperform all other NR metrics and, in most cases, even SSIM.

5. CONCLUSION

The work presented herein, was triggered by our preliminary assessment of the drawbacks of classic NR metrics, which we published in MoMM15 [37] and later further extended in [38]. Having concluded therein that no existing metric would be accurate enough to evaluate mobile streaming services, we set off to develop the regression-based method presented in this paper.

However, our earlier work provided the key to developing an efficient hybrid metric based on the existing ones. It was observed that existing metrics did operate correctly but only under confined conditions, and that the whole spectrum of video conditions was collectively well-covered by the metrics set. The intuition was then that an appropriate combination of existing metrics would lead to a breakthrough in NR assessment. We looked at machine learning, and, regression methods offered a computationally-viable avenue. The obtained results confirm that regression is also an accurate approach - in fact, comparably as accurate as FR metric VQM, and considerably better than FR metric SSIM.

Our method, as depicted in Figure 1, is meant to be a general framework to employ supervised learning in real-time streaming services. Given the positive findings, we intend to try other machine learning algorithms, with the possibility of further improving NR video quality assessment based on cognitive methods.

6. ACKNOWLEDGMENTS

This work has been carried out in the context of the European Research Council project BROWSE (Beam-steered Reconfigurable Optical-Wireless System for Energy-efficient communication - Grant 291632) and the ICT COST Action 3D-ConTourNet (IC1105).

7. REFERENCES

- [1] F. Agboma and A. Liotta. Addressing User Expectations in Mobile Content Delivery. *Journal of Mobile Information Systems, special issue on Improving Quality of Service in Mobile Information Systems*, 3(3), 2007.
- [2] F. Agboma and A. Liotta. Quality of Experience management in mobile content delivery systems. *Telecommunication Systems*, 49(1), 2012.
- [3] S. Borer. A model of jerkiness for temporal impairments in video transmission. In *in proc. of the Second International Workshop on Quality of Media Experience (QoMEX)*, 2010.
- [4] T. Brandão and M. Queluz. No-reference quality assessment of h.264/avc encoded video. *IEEE Trans. on Circuits and Systems for Video Tech.*, 20(11):1437–1447, November 2010.
- [5] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam. Objective video quality assessment methods: A classification, review, and performance comparison. *TBC*, 57(2):165–182, 2011.
- [6] M. G. Choi, J. H. Jung, and J. W. Jeon. No-reference image quality assessment using blur and noise. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, 3(2):184 – 188, 2009.
- [7] A. G. Ciancio, A. L. N. T. da Costa, E. A. B. da Silva, A. Said, R. Samadani, and P. Obrador. No-reference blur assessment of digital pictures based on multifeature classifiers. *IEEE Transactions on Image Processing*, 20(1):64–75, 2011.
- [8] R. Ferzli and L. J. Karam. Human visual system based no-reference objective image sharpness metric. In *ICIP*, pages 2949–2952. IEEE, 2006.
- [9] D. Freedman. *Statistical Models : Theory and Practice*. Cambridge University Press, August 2005.
- [10] P. Gastaldo, S. Rovetta, and R. Zunino. Objective quality assessment of mpeg-2 video streams by using cbp neural networks. *Neural Networks, IEEE Transactions on*, 13(4):939–947, Jul 2002.
- [11] S. S. Hemami and A. R. Reibman. No-reference image and video quality estimation: Applications and human-motivated design. *Signal Processing: Image Communication*, 25(7):469–481, 2010.
- [12] J. Hu and H. Wildfeuer. Use of content complexity factors in video over ip quality monitoring. In *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, pages 216–221, 2009.
- [13] C. Keimel, J. Habigt, and K. Diepold. Hybrid no-reference video quality metric based on multiway pls-r. In *EUSIPCO 2012: 20th European Signal Processing Conference*, pages 1244–1248, Sep 2012.
- [14] P. Le Callet, S. Moeller, A. Perkis, and eds. Qualinet white paper on definitions of quality of experience. In *COST Action IC 1003*, Lausanne, Switzerland, 2012.
- [15] P. Le Callet, C. Viard-Gaudin, and D. Barba. A convolutional neural network approach for objective video quality assessment. *Neural Networks, IEEE Transactions on*, 17(5):1316–1327, Sept 2006.
- [16] A. Liotta, D. C. Mocanu, V. Menkovski, L. Cagnetta, and G. Exarchakos. Instantaneous video quality assessment for lightweight devices. In *Proc. of Int. Conference on Advances in Mobile Computing, MoMM '13*, pages 525:525–525:531, New York, NY, USA, 2013.
- [17] H. Liu and I. Heynderickx. A no-reference perceptual blockiness metric. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008.
- [18] G. Matheron. *Principles of geostatistics*, volume 58. Society of Economic Geologists, 1963.
- [19] Matlab. Statistics and machine learning toolbox. Available at <http://nl.mathworks.com/products/statistics/>.
- [20] V. Menkovski, G. Exarchakos, and A. Liotta. Machine learning approach for quality of experience aware

- networks. In *in proc. of Computational Intelligence in Networks and Systems*, November 2010.
- [21] V. Menkovski, G. Exarchakos, and A. Liotta. Online learning for quality of experience management. In *Proceedings of the 19th Machine Learning conference of Belgium and The Netherlands*, 2010.
 - [22] V. Menkovski, G. Exarchakos, and A. Liotta. The value of relative quality in video delivery. *J. Mob. Multimed.*, 7(3):151–162, Sept. 2011.
 - [23] D. Mocanu, A. Liotta, A. Ricci, M. Vega, and G. Exarchakos. When does lower bitrate give higher quality in modern video services? In *Network Operations and Management Symposium (NOMS), 2014 IEEE*, pages 1–5, May 2014.
 - [24] K. Pandremmenou, M. Shahid, L. P. Kondi, and B. Löfstrom. A no-reference bitstream-based perceptual model for video quality estimation of videos affected by coding artifacts and packet losses. In *Human Vision and Electronic Imaging*, Feb. 2015.
 - [25] C. Perra. A low Computational Complexity Blockiness Estimation Based on Spatial Analysis. In *in IEEE 22nd Telecommunications Forum*, 2014.
 - [26] M. H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on broadcasting*, 50(3):312–322, Sept. 2004.
 - [27] J. R. Quinlan. Learning efficient classification procedures and their application to chess end games. In *Machine Learning. An Artificial Intelligence Approach*, pages 463–482, 1983.
 - [28] K. Rank, M. Lendl, and R. Unbehauen. Estimation of image noise variance. *IEEE Proceedings Visual, Image, Signal Processing*, 146(2):80–84, 1999.
 - [29] A. Reibman, S. Sen, and J. der Merwe. Analyzing the spatial quality of internet streaming video. In *in Proceedings of International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, January 2005.
 - [30] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack. Study of subjective and objective quality assessment of video. *Trans. Img. Proc.*, 19(6):1427–1441, June 2010.
 - [31] M. Shahid, J. Panasiuk, G. V. Wallendael, M. Barkowsky, and B. Löfstrom. Predicting full-reference video quality measures using HEVC bitstream-based no-reference features. In *Seventh International Workshop on Quality of Multimedia Experience, QoMEX 2015, Pilos, Messina, Greece, May 26-29, 2015*, pages 1–2, 2015.
 - [32] M. Shahid, A. Rossholm, B. Löfstrom, and H. Zepernick. No-reference image and video quality assessment: a classification and review of recent approaches. *Journal on Im. and Vid. Proc.*, 40(1), 2014.
 - [33] T. Shanableh. Prediction of structural similarity index of compressed video at a macroblock level. *IEEE Sign. Proc. Lett.*, 18(5), 2011.
 - [34] B. Staehle, A. Binzenhöfer, D. Schlosser, , and B. Boder. Quantifying the influence of network conditions on the service quality experienced by a thin client user. In *14. GI/ITG Konferenz Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen (MMB 2008)*, Dortmund, Germany, Apr. 2008.
 - [35] N. Staelens, D. Deschrijver, E. Vladislavleva, B. Vermeulen, T. Dhaene, and P. Demeester. Constructing a no-reference h.264/avc bitstream-based video quality metric using genetic programming-based symbolic regression. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(8):1322–1333, Aug 2013.
 - [36] M. Torres Vega, E. Giordano, D. C. Mocanu, and A. Liotta. Cognitive no-reference video quality assessment for mobile streaming services. In *in proc. of the 7th International Workshop on Quality of Multimedia Experience (QoMEX)*, May 2015.
 - [37] M. Torres Vega, V. Sguazzo, D. C. Mocanu, and A. Liotta. Accuracy of no-reference quality metrics in network-impaired video streams. In *13th International Conference on Advances in Mobile Computing and Multimedia (MoMM2015)*, Dec. 2015.
 - [38] M. Torres Vega, V. Sguazzo, D. C. Mocanu, and A. Liotta. An experimental survey of no-reference video quality assessment methods. *Int. Journal of Pervasive Computing and Communications*, 12(1), 2016.
 - [39] S. Winkler. *Digital Video Quality: Vision Models and Metrics*. Wiley, 1 edition, Mar. 2005.
 - [40] S. Winkler and P. Mohandas. The evolution of video quality measurement: From PSNR to hybrid metrics. *IEEE Transactions on Broadcasting*, 54(3):660–668, Sept. 2008.
 - [41] H. R. Wu and M. Yuen. A generalized block-edge impairment metric for video coding. *Signal Processing Letters, IEEE*, 4(11):317–320, Nov. 1997.