Deduction of Video Quality Degradation in Transport Networks

Georg Wenzel (1), Gerhard Gehrke (1), Martin Belzner (2), Dieter Stoll (1), Wolfgang Thomas (1)

1 : Alcatel-Lucent Deutschland AG, Thurn & Taxisstr. 10, D-90411 Nuremberg
e-mail: { gwenzel | gerhardgehrke | dieterstoll | wolfgangthomas } @ alcatel-lucent.com

2 : Chair of Information Transmission, Friedrich-Alexander University Erlangen-Nuremberg, Cauerstr. 7, D-91058 Erlangen, e-mail: belzner @ LNT.de

Abstract

It is already known that significance marking of IP video packets in conjunction with a corresponding evaluation in the transport network considerably protects the quality of videos against degradation from network congestion. The focus of this paper therefore is the investigation of possibilities to take benefit from such significance marks. In particular the interest is to estimate the video quality in the network without having to spend considerable effort for decoding or even decrypting of the video stream.
While the merit of significance markings has already been proven by simulation, now the benefits have been demonstrated by experiments on actual transport network hardware.
Thereby, network element typical Ethernet performance counters registering the amount of received and lost bytes have been modified for differentiation by the significance marks. Those counters, retrieved per second, were compared with the quality of the video.
It is discussed how to construct a metric for video quality, and the role of a suitable definition of an objective reference for the calibration of this metric: The metric has to regard specifics of the video compression, video encoding and transport network characteristics. Results from these experiments are provided and challenges for further refinement and implementation are pointed out.

1. Introduction

During recent years, transport networks have become the convergence platform for all type of communication services. While there still exist separate infrastructures for some services like ISDN or cable TV, there is a clear tendency to concentrate all services on the same infrastructure. As consequence of this convergence, services have to compete against each other for network resources and the quality of a service is influenced not only by rare events in an infrastructure which is designed for high availability (so called 5 9's) but also by other services during normal working conditions.

Therefore it is more important – but also more difficult – to supervise service quality inside the network, given the constraint that the transport provider must not care for contents of the transported streams. As such metrics are currently not available, transport providers tend to prioritize packet video and telephone transmission over standard internet traffic either through separated transport resource allocation or via priority markings in the packet headers. These priority markings are evaluated during transport by the traversed network elements, which can therefore influence the queuing and packet drop decision and/or the prioritized scheduling of dedicated services [1]. Besides the assurance of a reliable packet video and telephone transmission, the monitoring of the transmission quality of such streams becomes important. Constantly determining the actual, customer perceived quality of a transported video stream is not only important for customer or inter transport provider billing and reclamation issues but also for reacting to sudden und unforeseen congestion events. An explicit monitoring of the video quality, e.g. through video decoding within the network elements or direct customer feedback is however not possible for the transport provider for reasons of effort and probably missing rights of access or keys to decrypt the transported content.

The goal of our work is to identify potentials for in network measurements of video quality degradation which do not require video stream decoding. We expect such means to better support the implementation of "emerging" services: Operators e.g. like the "Deutsche Telekom" expand their traditional data transport business by offering "Triple Play" beside data and voice transport. Video transportation of IP-TV or video on demand services shall play an important role. In the following paragraphs and sections we present the results gained from video quality measurement in different error scenarios.

2. Video Transmission in Packet Networks

2.1 Basic Video Structure

In order to understand the impact of packet loss in a packet video transport we take first a closer look at the basic video structure. A video stream is composed of subsequent picture frames, which are often presented with a rate of ~25 frames per second (fps). In order to reduce the required transport data rate the video pictures are compressed during the video encoding process by reduction of irrelevant and minor information. Independently of the picture encoding method one can distinguish between two different basic frame types, Reference and Non-Reference picture frames. Reference pictures serve as decoding basis for other picture frames while information from Non-Reference pictures does not propagate to consecutive or preceding pictures. This is shown in Figure 1. The independently decodable intra coded I-Frame reference pictures serve as resynchronization points in case of errors and thus limit the error propagation spread. A group of inter-dependent pictures (between two subsequent I-Frames), with respect to the decoding, is organized in a so-called "Group of Pictures" (GOP). Arrows in Figure 1 show the information propagation from reference frames (intra coded or predictively encoded) to dependent frames (predictively encoded).

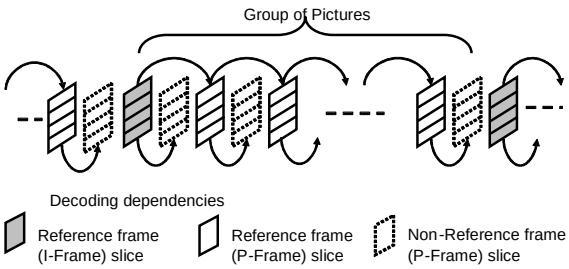

Figure 1: Video Picture Frame Structure

For a further increase in the error resilience a picture frame can be subdivided into independent decodable picture slices. As consequence errors from one slice will not propagate over slice boundaries. The encoded video data slices are mapped depending on the used transport technology into packets. In the packetization process, it is important to keep the separation between reference and non reference information thus reducing the error propagation in a packet loss event. Additionally, the information of two different slices will not be packed into the same transport packet. By using a separation into independent slices a video picture frame stream can be regarded as multiple parallel independent video slice streams. In Figure 2 an example video packet stream structure is shown. On a packet level a repeated sequence of reference and non-reference packets can be seen.

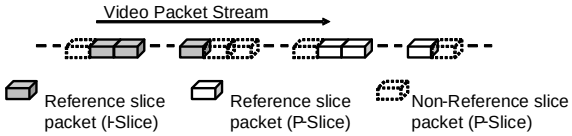

Figure 2: Video Packet Stream Structure

2.2 Expected Video impairments in packet loss scenarios

As described in the previous paragraph there exist interdependencies between different picture frames within a GOP. By mapping this onto the packet transport layer these result in certain interdependencies between packets within the transport stream. The loss of a packet might not only impact the quality of the picture frame it belongs to, but can also have an impact on future picture frames depending on the picture frame type. In case of packet loss in reference pictures more severe video quality degradations can be expected compared to packet loss in non-reference pictures [2]. So basically we have to distinguish between two packet loss scenarios. A measured packet loss in non-reference pictures will provide a certain amount of video quality degradation which is limited to a short amount of time, depending on the picture frame rate. So it can be expected that the observable video impairment for a standard 25 fps video stream will be temporally limited to a maximum of 40 ms. Furthermore through the video slice structure the quality degradation will also be spatially limited. Through this error propagation limitation for non-reference packet loss a correlation between the currently measured packet loss event and the expected video quality can be expected. However, to correlate any observed packet loss in reference picture frames with the actual video quality becomes difficult. At first, it is not known to the monitoring network element how many other pictures are impaired. Packet loss in the first reference picture of a GOP will have significantly more impact on the video quality than packet loss in the last reference picture of a GOP. As a consequence of this, packet loss in reference pictures leads to a decorrelation between the observed packet loss and the actual video quality. E.g. while monitoring zero packet loss at a certain moment, the video quality could nevertheless be degraded due to packet loss in the past. The expected observable video impairments will be temporarily longer persistent, while still being spatially limited due to the slice structure as in the non reference packet loss scenario. In the worst case

scenario, this results in impairments for the total duration of a group of pictures.

3. Significance marking and network significance Awareness

As described in the previous section the packets in a video data stream are of different significance in the video decoding process and thus packet loss have a different impact on the perceived video quality. In order to be able to distinguish the packets within the network elements, based on their significance, a corresponding significance marking on the reference and non-reference packets is required [3]. This significance marking has to be added within the video packetization process by the video encoder. Furthermore significance aware network elements are required so that in case of indicated or sudden congestion events the network elements along the transmission path can decide which packets to enqueue and which packets to drop (Figure 3).

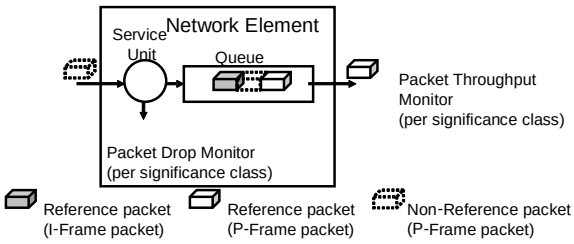

Figure 3: Basic Network Element Structure

In the queuing mechanism two basic properties have to be considered. For the quality of real time applications (e.g. IP-TV) which does not allow for end-to-end retransmissions, it is preferable to use the same queue in the network element for non-reference and reference packets, instead of using separate queues with different priorities. This prevents that packets to overtake each other during the scheduling process of the queues which could cause quality degradations at the receiver side [4]. However using the same queue for high and low priority packets can lead to scenarios where low priority packets fill up the queue thus preventing high priority packets from entering the queue. So despite of using the simple Tail Drop queuing mechanism where arriving packets are either enqueued or dropped depending on only whether the queue is fully filled, it is in general preferred to drop the non-significant frames 'earlier' than the significant ones. By configuring the used network element's queue with different thresholds per priority class it can be assured that non reference packets can not congest the queue and thus block reference packets with a higher priority. This can be even combined with a random element for the low priority class, which in addition sometimes drops arriving non-reference frames if the queue is almost filled up to the specific 'early' threshold. This - of course - creates artificial packet drops although the queue would practically be able to temporarily store the packets. However this is far less critical to the video quality than the loss of a reference packet.

In the packet monitoring process, it is also important to differentiate the packet drop and throughput counters for the different packet significance classes (Figure 3).

Dependent on the used transport medium, different transport stream techniques are available for further packetization and encapsulation. Recently, video and other real time applications are often transported via the Real Time Transmission Protocol (RTP). The RTP protocol contains important information for the video presentation, e.g. timestamps or the markers for the significance [5]. In the scenario considered in this paper, the packet video RTP stream is transported via IP-Multicast over an IP/Ethernet capable transport network. That means the encoded RTP packets are directly encapsulated into IP packets and then mapped into Ethernet packets. Both the IP encapsulation as well as the Ethernet encapsulation provides the possibility for a packet significance/priority marking. The IP packet priority can be signaled in the IP DS field of the IP header [6] - also known as Type of Service (TOS) field, while the Ethernet packets can indicate the priority and drop eligibility in the PCP fields in the Ethernet VLAN tags [7]. By means of an adequate network element configuration by the transport provider a mapping can be defined between the network element packet dropping rules and the IP DS respectively the VLAN tag PCP fields. Because currently no standard for video packet significance markings does exist neither for the IP nor the ETH layer, we used self-selected arbitrary values for both the significance aware network element as well as the video encoder.

4. Network Element Monitoring Impairments

Monitoring the performance quality of a packet stream within a network element is currently based on evaluating dropped and forwarded packets. These network performance counters are usually polled by the transport provider in time intervals of the magnitude of several minutes. Therefore network elements are often designed, while internally monitoring every single packet, to provide only aggregated information. This might be fast enough for a posteriori performance evaluation (e.g. for billing), but it is by far to slow to allow a detailed real-time quality monitoring, which would be required to capture and maybe to react to unforeseen traffic changes. Furthermore the

monitoring often does not capture the actual amount of lost bytes, assuming packets of different lengths. Therefore we used in our work a modified network element with congestion drop and throughput performance counters on a byte basis instead of traditional packet performance counters.

The actual network element which we used could only be modified to allow the polling of the performance counters in one second intervals which corresponds to ~25 picture frames. Unfortunately this did not allow detecting the GOP structure of the video streams even if the latter could be isolated from each other.

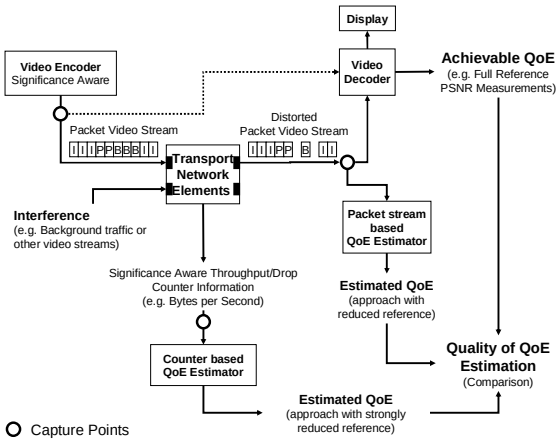

Figure 4: Test set-up structure

5. Measurement Test Setup

5.1 Network Environment

For this work a network environment as well as network equipment of Alcatel-Lucent was used. As a central element for the quality measurements a modified metro switching and transport network element was used. The network element allowed a distinction of the packets by their markings on IP and ETH level, which in our tests corresponded to the different video packet significances. While mainly acting as IP/Ethernet switch with the QoS queuing functionality it additionally performed the mapping of IP/Ethernet packets into SDH/SONET TDM signal payloads. The Virtual Concatenation (VCAT) technique in SDH/SONET allows dimensioning stepwise the allocated transport resources. They have been set thus that the video traffic stream under test could pass the network element without being influenced. Additionally, background traffic stream was sent into the network element. The video stream as well as this background traffic stream were internally mapped onto the same egress port thus that they had to compete for the shared link. This way, artificial congestion was induced to evaluate the impact of packet loss on the quality of the video stream under test. The background traffic was structured so that it emulated concurring aggregated video traffic. The data rate of the background traffic was following a predefined temporal profile in order to generate differently severe congestion events. The video packet stream was captured via wireshark software [8]. This allowed an offline evaluation of video loss and throughput even on a packet level granularity, which equals a quality evaluation at the fastest possible sampling rate. Figure 4 shows an overview schematic of the test setup.

5.2 Video Encoder/Decoder and Video Content

A H.264/AVC video encoder provided by the Fraunhofer Institute for Integrated Circuits IIS was used to generate the video stream. This system encodes an analogous received video signal, which in our case is generated by a DVD player, and provides the resulting packet stream with significance markings for reference and non-reference video packets. Note that using as input an analogous video signal, which is impacted by not fully removable interlacing effects and other hardware induced noise, the video encoder would encode a difference even for original identical images, if it was not fine-tuned to compensate such artifacts.

The video encoder generates an intra coded picture at the start of each GOP. All other pictures within the GOP are predictively encoded. Reference and non-reference pictures are strictly alternating within the GOP.

As video samples we used standardized SD-TV video sequences with a PAL resolution provided by the Video Quality Expert Group [9]. In the scenario presented in the following we used the “F1-Car” sequence representing a video with fast changing content, the “Tree” sequence representing a video with no or only slow changing video content as well as the "Moving graphic" sequence for a video with medium fast changing content.

The video decoding of the distorted video stream after the network element as well as of the original input video stream was done via an open the source software called FFmpeg [10].

6. Video Quality Estimation and Measurements

The goal of this investigation is to derive a video quality estimation metric solely based on byte throughput and on byte drop performance counters measured by network elements in the video transport path. In order to evaluate the accuracy of the video quality estimation derived from a network performance counter, a comparison to the actual video quality is required. We used a typical measurement for the video encoding/transmission quality as quality reference, the so called Peak Signal to Noise (PSNR) measurement. The PSNR measurement is normally a full reference method meaning the received video signal after decoding is compared to the originally transmitted one after decoding [11]. The PSNR is measured in db and is defined as the ratio of the peak signal value (here the 8 Bit luminance value) to the Mean Squared Error (MSE) (1), which is derived from comparing the original/reference signal value (Ref((i,j)) with the received/distorted signal value (Dist(i,j)) of each pixel of each picture frame. S_h and S_v represent the image size in the horizontal and vertical dimension. (2)

$$PSNR = 10 * \log_{10}\left(\frac{255^2}{MSE}\right) \qquad (1)$$

$$MSE = \frac{\sum_{i=1}^{S_h}\sum_{j=1}^{S_v}\left(\mathrm{Ref}(i,j) - \mathrm{Dist}(i,j)\right)^2}{S_h * S_v} \qquad (2)$$

As it can be seen from the definition, the PSNR metric has an inherent major drawback for quality evaluation: In order to measure the actual impairments arising from the loss of packets during the video transport, we compare the quality of the decoded video stream after the network element with the quality of the decoded video stream before the network element. The PSNR metric is, however, not well defined for comparing identical images. In a case, where no packet loss occurs, the PSNR value becomes infinite. So in order to prevent an arbitrary clipping of infinite values we instead used the Video Quality Metric (VQM) (3) proposed by the International Telecommunication Union (ITU) [11] as video quality reference metric.

$$VQM_{PSNR} = \frac{1}{1 + e^{0.15*(PSNR - 19.7818)}} \qquad (3)$$

The VQM metric is based on a mapping of PSNR measurements to subjective video quality perceptions. The values used for the scaling were experimentally determined by the ITU. So instead of an infinite value for identical images the VQM metric will return the value zero.

7. Measurement and Monitoring Results

The performance throughput and loss counters provided by the network element were latched every second. Thus each sample value represents for an average of about 60 packets the stream quality per second. This makes a correlation between packet loss and the resulting video quality difficult. Therefore we gathered instead the raw monitoring information from a stream analysis on packet granularity. Internal to a future network element this could be easily done as well. The following results are based on this identification of IP packets.

In Figure 5 it can be seen that a measured byte loss ratio on a per picture frame basis of ~40ms can lead to a wide spread of different video qualities. The video quality is measured in 1-VQM. A large value on the vertical axis corresponds to a high PSNR value and thus good image quality. A direct correlation between the packet loss event and a related picture quality seems not to be possible. Quality measurements on a per picture frame basis are thus not recommended, especially not in non significance aware networks. As mentioned earlier, the observed variance is an immediate consequence of the correlation of picture frames amongst each other so that a picture may be affected not only by a lost packet directly belonging to it, but also by the loss of a packet, which belongs to a reference frame. In a significance aware network, the quality spread is less extreme as less reference packets are lost, however there is still a considerable difference between the measured different quality values for a certain byte loss ratio.

In order to manage these fluctuations we increased the measurement and averaging interval to the videos' GOP size. This makes sense as it is known from the general video structure that errors can only propagate within the GOP borders.

In Figure 6 the measured video quality on a group of picture (50 pictures in this scenario) granularity is presented. One can see that in a significance aware network, where only non reference picture frame packets are dropped, the achievable quality correlates well with the byte loss ratio. The narrow shape of the

curve allows a mapping between the measured byte loss in the network element and a certain video quality. In a non significance network the quality estimation is significantly more difficult.

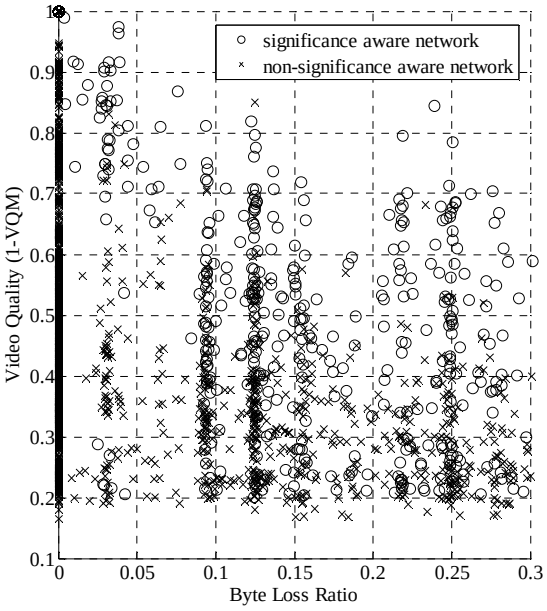

Figure 5: Quality vs. byte loss ratio (per picture frame)

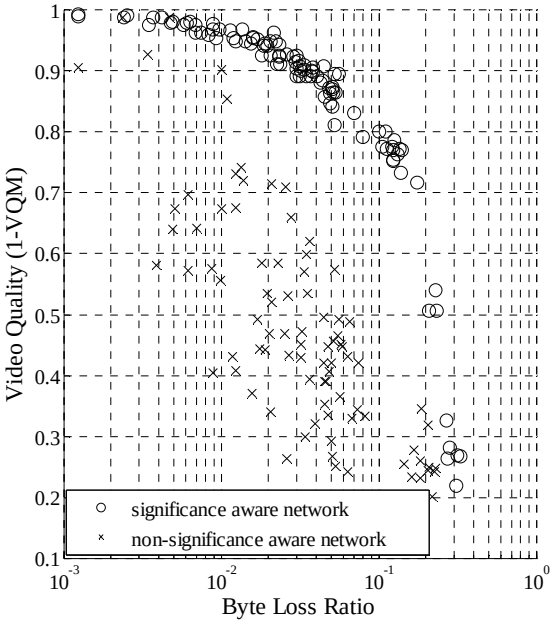

Figure 6: Quality vs. Byte loss ratio (per GOP) - fast

An upper bound of the video quality in non significance aware networks is the curve which is obtained in the significance aware case which is present if only non-reference frames are dropped by the network element. Again it can be seen that the measured video quality in a non significance aware network can greatly vary for identical byte loss ratios thus that it seems unlike that in such networks the correct video quality can by estimated through byte loss counts.

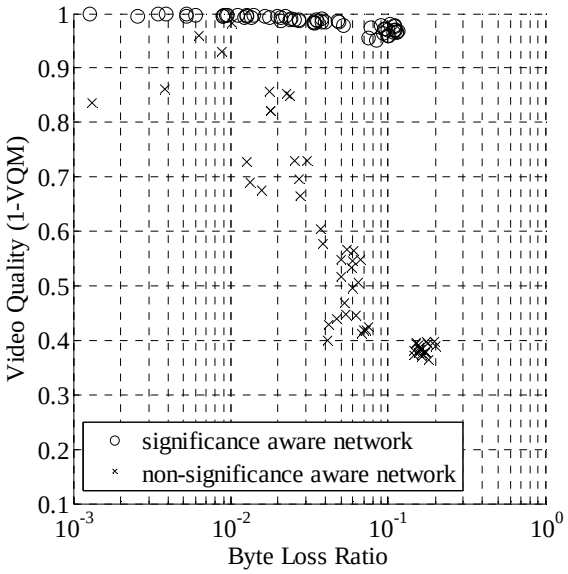

Figure 7: Quality vs. byte loss ratio (per GOP) - slow

Comparing the quality measurement results from Figure 6, which represents a video with fast changing content, to Figure 7 which shows the achievable video quality for a slow changing content, it can be seen that the mapping between byte loss and video quality is furthermore depending on the internals of the transmitted video itself. Since this information cannot be deduced without decoding the video, it is unknown to the transport service provider, and thus renders the definition of a suitable quality estimation difficult, even in a significance aware network.

8. Conclusion

It has been shown that in service aware networks even very simplistic metrics can provide video quality degradation information without the need for decoding the video stream. In spite of this positive observation it turns out that on absolute scales the details of videos, which are unknown to the metric, would need to be known for calibrating the metrics correctly. Therefore it is left to future work to further improve the presented concepts and derive a metric which is de-

ployable in networks. From our investigations we see two major aspects to be looked at:

1. In video distribution networks the goal is certainly not to supervise quality of individual streams. Rather it is to asses the delivery of bundles of video streams. Thus it should be investigated to what extend statistical effects help in the analysis of video quality. It is likely that either bundles have on average the same internal behavior or that at least bundle specific calibration could be derived.
2. The VQM metric which was used in this work is not very correlated to human perception. It should be investigated to what extend other metrics, like MPQM (moving picture quality metric, [12]) or MOScor (PSNR-based corrected Mean Opinion Score [13]), provide better quality metrics.

9. Acknowledgment

This work has been partially funded by the German Ministry for Research and Education (BMBF Grant MUNAS 01BP554).

Special thanks go to the Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33 in 91058 Erlangen, Germany, which supported us by technical consulting and with video encoding hardware. http://www.iis.fraunhofer.de/bf/amm/

10. References

[1] G. Gehrke, R. Hocke, G. Wenzel, “QoE estimation of compound services in significance aware packet networks”,
Bell Labs Technical Journal Vol. 13, No. 3

[2] Cheng-Han Lin et al., „The Packet Loss Effect on MPEG Video Transmission in Wireless Networks”, Proceedings of the 20th International Conference on Advanced Information Networking and Applications - Volume 1, 2006

[3] Jirka Klaue et al., “Semantic-aware Link Layer Scheduling of MPEG-4 Video Streams in Wireless Systems”, Proc. of Applications and Services in Wireless Networks (AWSN), Bern, Switzerland, 2003

[4] IETF RFC 2309, “Recommendations on Queue Management and Congestion Avoidance in the Internet”, April 1998

[5] IETF RFC 3984, “RTP Payload Format for H.264 Video,” Feb. 2005

[6] IETF RFC 2474, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", Dec. 1998

[7] IEEE 802.1ad, “Virtual Bridged Local Area Networks – Amendment 4: Provider Bridges,” August 2005

[8] Wireshark, http://www.wireshark.org/

[9] Video Quality Expert Group
http://www.its.bldrdoc.gov/vqeg/

[10] FFmpeg, http://ffmpeg.mplayerhq.hu/

[11] Working Group on Multimedia Communications Coding and Performance: Objective Video Quality Measurement Using a Peak-Signal-to-Noise-Ratio (PSNR). Committee T1 - Telecommunications. T1A1.1/2001-026R7, 2001

[12] C. van den Branden Lambrecht, O. Verscheure, “Perceptual quality measure using a spatio-temporal model of the human visual system”, Proceedings of the SPIE, Vol. 2668 (1996) 450-461

[13] Olivia Nemethova, Michal Ries, Matej Zavodsky and Markus Rupp - "PSNR-Based Estimation of Subjective Time-Variant Video Quality for Mobiles" Proc. of MESAQIN 2006, Prague, Czech Republic, June 2006