Resilience of Video Streaming Services to Network Impairments

Maria Torres Vega, *Member, IEEE,* Cristian Perra, *Senior Member, IEEE,*4b and Antonio Liotta, *Senior Member, IEEE*

Abstract—When dealing with networks, performance management through conventional Quality of Service (QoS) based methods becomes difficult and is often ineffective. In fact, quality emerges as an end-to-end factor, for it is particularly sensitive to the end-user perception of the overall service, i.e. the user's Quality of Experience (QoE). However, the two are not independent from each other and their relationship has to be studied through metrics that go beyond the typical network parameters. To better explore the value of assessing QoE alongside QoS in high-speed, lossy networks, this paper presents an experimental methodology to understand the relation between network QoS onto service QoE, with the aim to perform a combined network-service assessment. Using video streaming services as the testcase (given their extended usage nowadays), in this work, we provide studies on three network-impaired video-sets with the aim to provide a comprehensive evaluation of the effects of networks on video quality. First, the ReTRIeVED video set provides the means to understand the most impairing effects on networks. Furthermore, it triggered the idea to create our own sets, specialized in the most impairing conditions for 2D and 3D: the LIMP Video Quality Database and the 3D-HEVC-Net Video Quality Database. Our study and methodology are meant to provide service providers with the means to pinpoint the working boundaries of their video-sets in face of different network conditions. At the same time, network operators may use our findings to predict how network control policies affect the user's perception of the service.

Index Terms—2D Video streaming services, 3D Video streaming services, quality of experience, network impairments

I. INTRODUCTION

Satisfying users, devices and services' requirements is fundamental to successfully manage an ever-growing world-wide wireless network [1]. To achieve this, it is essential to perform a real-time monitoring of the applications and networks, and act upon them when improvements are required. This monitoring task has traditionally been studied in the context of network Quality of Service (QoS) management. However, due to the variability in channel conditions, services over wireless networks incurs quality degradations even when there is sufficient nominal capacity. QoS factors reflect the status of individual networks but do not comprehensively capture the end-to-end features that affect the overall quality delivered to

M. Torres Vega is with the IDLab–imec, Ghent University, Ghent, Belgium.
A. Liotta is with the Data Science Centre, University of Derby, UK.
C. Perra is with the Department of Electrical and Electronic Engineering, University of Cagliari, Italy.

the user. To address these elements, Quality of Experience (QoE) management represents a much better proposition [2].

When it comes to Internet services, humans become the quality meters [3], and their expectations, perceptions and needs carry a great value. The overall goal of QoE management is therefore to optimize the end-user QoE (end-user perspective), while making efficient use of network resources (current and future) and maintaining a satisfied customer base (provider perspective). Successfully managing QoE for a specific application requires understanding and identifying the multiple factors that affect it (subjective and objective), from the point of view of various actors in the service provisioning chain. Among all factors, QoS impairments have demonstrated to have a great influence on QoE [4]. Understanding the relationship between network-based QoS parameters and user-perceived QoE provides important input for the QoE management process, in particular to network providers with control over network resource planning and provisioning mechanisms. Extensive research has been done to find the relation between QoS and QoE [5]. However, the effect of QoS is highly non-linear; thus, developing a generic model that could work over a broad range of conditions becomes a challenging task.

Particularly crucial is the case of video streaming services, which are the most widespread and used service types nowadays [6]. New streaming protocols increase bandwidth requirements and transmission complexity, which are critical elements for service and network providers [7]. In this type of services the relation between QoS parameters and their effect on the experienced quality becomes more challenging. The effect of the QoS impairments on the user's perceived quality will heavily depend on the transmission protocol used for streaming. This also means that QoE has to be modeled in a protocol-dependent way. For example, a streaming service based on HTTP/TCP (such as MPEG-DASH [8]) is more susceptible to temporal impairments (retransmissions alleviate the influence of the lost packets). As such, the methods used to assess quality have to focus on temporal impairments. Examples of these are the temporal approach of Duanmu et al. [9], the DASH quality assessments of Zegarra et al. [10], Rodríguez et al. [11] and Zhao et al. [12] or even the Video Multimethod Assessment Fusion (VMAF) of Netflix. In contrast, when the streaming service is RTP-based (teleconferencing systems or IPTV), due to the fact that RTP does not provide retransmission mechanisms, it is easier to study the direct effect of the QoS impairments (packet

loss, jitters, throttles). These can be measured by traditional objective metrics such as the Peak to Noise Ratio (PSNR) or Structural Similarities (SSIM). However, generalizing the effect of each one of the impairments isolated under a broad range of videos and conditions is still under investigation.

In this work, we present a general experimental methodology to assess the effects of QoS on QoE for RTP-based video streaming services. We have applied our methodology to three network-impaired video-sets with the aim to provide a comprehensive evaluation of the effects of networks on video quality. First, the ReTRIeVED video set [13] provided the means to understand the most impairing effects on networks. Furthermore, it triggered the idea to create our own sets, specializing in the most impairing conditions for 2D and 3D: the LIMP Video Quality Database [14] and the 3D-HEVC-Net Video Quality Database. Our study and methodology are meant to provide service providers with the means to pinpoint the working boundaries of their video-sets in face of different network conditions. At the same time, network operators may use our findings to predict how network control policies affect the user's perception of the service.

The remainder of this paper is organized as follows. Section II provides a short introduction to QoE and how it is measured for RTP video streaming services. Looking for a general, systematic and comprehensive methodology for understanding the effects of QoS parameters onto QoE and pinpointing the breaking conditions, we set up the controlled evaluation environment presented in Section III. Section IV, presents the quantitative results of the deployment of this experimental method considering a broad range of network impaired video contents (the three video sets). Section V, provides further insights on the non-linear relations between QoS and QoE, as well as setting a group of boundary rules for the benefit of network and service providers. Finally, Section VI draws key conclusions.

II. BACKGROUND: VIDEO QUALITY ASSESSMENT OF RTP VIDEO STREAMING SERVICES

Measuring QoE of RTP video streaming services in a fast and accurate manner is still an open area of research. Due to its subjective essence, the legitimate judges of visual quality are the humans, whose opinion can be obtained through subjective analyses [15]. In practice, presented stimuli (for example impaired video sequences) are rated by subjects under controlled conditions. These ratings express the subjective QoE (sQoE) described typically by the Mean Opinion Score (MOS). However, due to the time-consuming nature and bias of subjective experiments, in the last years great effort has been placed onto developing objective quality metrics which could provide with a valid alternative, i.e. objective QoE (oQoE) [15]. In this work, we have explored the latter avenue, which allows not only to automate the QoE measurement and control loop, but also doing so in real-time.

The ultimate goal of the oQoE metrics is to provide the best possible correlation to subjective studies and the human vision system (HVS) by means of only the reference (original) and the received (distorted) material. Depending on the amount of reference information necessary to perform the assessment, the oQoE approaches are classified in three categories: Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR). FR metrics require the original material to perform their assessment. Thus, they are, in general, considered more accurate than the other two categories [16]. One simple example of these metrics is the Peak-Signal to Noise Ratio (PSNR) algorithm, which relies on the root mean square error (RMSE) for its assessment. Although very simple and computationally efficient, PSNR fails to provide an accurate enough assessment, particularly when videos have been distorted by network impairments [17].

The Structural Similarities (SSIM) algorithm has shown better correlation to the HVS [18]. SSIM was originally developed to assess still images, but was later adapted to video. It is based on the observation that a natural image or frame in a video is highly structured [19]. Structural information is defined as the attributes that represent the structure of objects in the scene, independent of the average luminance and contrast. Hence SSIM combines comparisons in terms of luminance, contrast and structure. Since the HVS is highly adapted to structural information, the performance of SSIM is better correlated with sQoE than other simpler oQoE metrics such as PSNR.

Specifically, because of its good correlation with subjective values, the Video Quality Metric (VQM) [20] is broadly chosen as a better option when subjective studies are not available. The VQM calculation involves extracting perception based features (spatial alignment, valid region estimation, gain and level offset calculations and temporal alignment), computing video quality measurements and combining the parameters to provide the overall quality measurement. This process is computationally more expensive than SSIM and PSNR. Thus, VQM is commonly used as the benchmark FR metric [21] or when real-time measurements are not required.

Finally, another very accurate FR model is the MOVIE (MOtion-based Video Integrity Evaluation) algorithm [22]. MOVIE integrates both spatial and temporal aspects of distortion in its assessment. It uses explicitly motion information from the reference video and evaluates the quality of the test video along the motion trajectories of the reference video. Despite its accuracy, MOVIE is so computationally demanding that it can hardly be used on large video datasets, which is why we have not used it in our assessment.

III. QOS-QOE ASSESSMENT METHODOLOGY

The driving motivation for our work is to provide a systematic and general analysis of the effect of QoS parameters on the end-user experience of the service (QoE). For this reason, our first step was to deploy streaming services in real networks and assess the end-user quality while the network was monitored. Examples of these studies are presented in [23], [24]. These studies, while providing solutions shaped to the networks under scrutiny, lacked on general behavioral conclusions. In addition, due to the changing conditions of real-network systems, it is very difficult to obtain a benchmark solution. For this reason, our next research question was to understand the
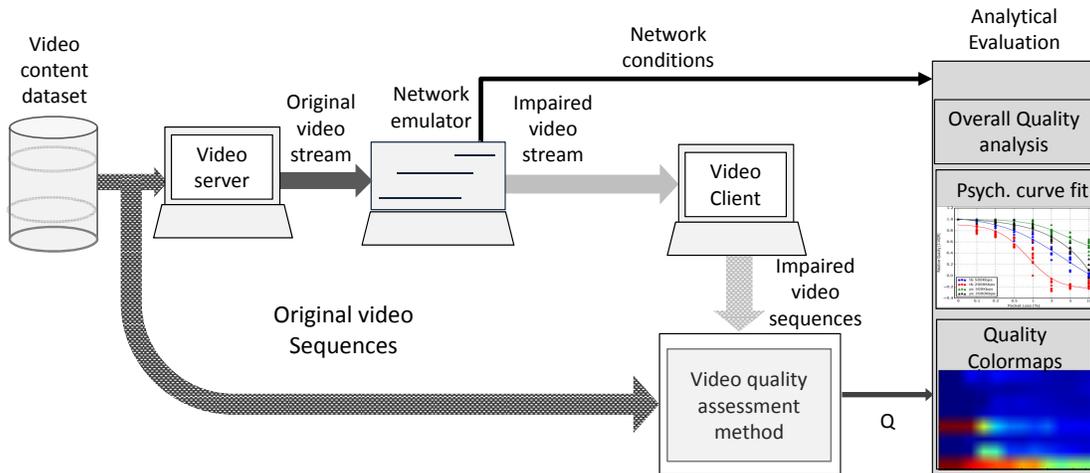
Fig. 1: Experimental Evaluation Methodology for mapping the network QoS to the end-user QoE.

influence of the different QoS parameters on video streaming services on a broad, general and systematic manner. Such analysis is meant to provide video and network managers with the means to perform actions, both on the videos streams and on the network control plane. To achieve this, it was necessary to have a controlled environment in which network and video conditions could be assessed in an isolated manner. Herein, we present our controlled evaluation set-up.

We engineered the indoor network video evaluation system shown in Figure 1. In it, an RTP-video server streams to a connected RTP-video client. Between client and server a network emulator is located. This emulator (Hurricane II from PacketStorm is able to emulate network conditions in real-time as well as full ITU network impairment models [25]. On video reception, the client performs an objective video quality assessment by means of a full-reference comparison between the original and the impaired materials. At the same time, the network emulator provides the real-time assessment of the conditions employed. Finally, both QoS and QoE assessments are used to understand how the first affects the second one. This analysis consists of an overall quality degradation assessment, quality colormaps in which conditions and videos are unfolded. In addition, as introduced in Section II, the relation between QoE and QoS can be modeled by means of psychometric curves [5]. In this QoS-QoE evaluation we include a final sigmoid psychometric fit.

Wireless networks suffer from instantaneous changes on conditions derived from congestion in densely populated areas, low reception (due to long distances to the access point), etc. These changes translate in QoS impairments which can be classified in four basic types: delay, jitter, packet loss (PLR) and throughput constraints (in terms of throttle) [26]. For our systematic analysis, instead of using the full network models installed on the network emulator, we decided to do an isolated analysis of each of these four QoS impairments.

As we explained in the background section, PSNR has been deemed inaccurate. On the other hand, the excessive computational times of MOVIE made it inadequate for large

scale assessment. VQM and SSIM were best suited metrics to act as benchmark and could be used for our general and systematic methodology. The decision of which benchmark to use was made on the accuracy of each of the candidates to the HVS, as it will be shown next.

IV. QUANTITATIVE EVALUATION OF THE NETWORK IMPAIRED VIDEO SETS

In this section we present the quantitative results of applying our experimental QoS-QoE method to three video datasets. Our first step in the assessment was to pick a quality benchmark in line with subjective assessment. We performed this validation on the ReTRiEVED dataset, which comes with MOS (Section IV-A). Once the objective FR benchmark was selected, our purpose was to find and pinpoint the most impairing conditions. For it, we again evaluated the ReTRiEVED dataset (Section IV-B). Based on these results, Section IV-C and Section IV-D present the analysis of our own network impaired videosets, which provide a deeper and broader analysis of the effect of the most impairing conditions (PLR and bandwidth) for 2D and 3D videos, respectively.

A. Objectivizing the Subjective Perception of Network Impaired Videos

QoE is inherently subjective [27] and the most extended procedure to measure perceived quality is by means of MOS indexes. However, as we introduced in the previous section, subjective studies, while providing accurate measurements, are unsuitable to perform quality analyses at large scale. Furthermore, their biased essence makes them difficult to map to a general viewer condition. For these reasons, objective metrics, which focus on the objective degradation of videos, are better suited to measure quality degradations. The first investigation we performed was to look for an objective metric that would fit the user's subjective experience accurately.

In the previous Section we picked SSIM and VQM as benchmark quality candidates. Thus, the subjective benchmark fit was performed using SSIM and VQM.

http://packetstorm.com/packetstorm-products/hurricane-ii-software/

TABLE I: ReTRIEVED Video Dataset parameters in terms of video types (acronym, name, description) and network impairments. Videos are subjected to each of the 23 different conditions individually. This makes a total of 184 impaired videos to evaluate.

		Video type					Network condition			
Acronym	Name	Size [Pixel]	FR [fps]	Length [s]	Description	Delay [s]	Jitter [ms]	Throughput [Mbps]	PLR [%]
cr	Crowdrun	704×576	25	9	Still; forest; people running	0	0	No	0
do	Duckstakeoff	704×576	25	9	Still; water; ducks take off	0.1	1	5	0.1
hb	Harbour	704×576	30	9	Still; water; Sailing boat race	0.3	2	3	0.4
ic	Ice	704×576	30	7	Still; outdoor ice rink; young people ice skating	0.5	3	2	1
pj	Parkjoy	704×576	25	8	Horizontal tor train; river edge; people running	0.8	4	1	3
sc	Soccer	704×576	30	7	Horizontal tor train; close up football match	1	5	0.5	5
re	Restaurant	720×576	25	9	Horizontal tor train; restaurant; man drinking				8
ru	Running	720×576	25	8	Horizontal tor train; forest; two persons running				10

TABLE II: PCC indexes of the FR metrics VQM and SSIM to the subjective MOS for all videos of the ReTRiEVED data set averaged per video type and network condition. Delay is given in seconds, jitter in milliseconds, bandwidth throttle in Mbps and packet loss ratio in percentage. Cell colors give qualitative correlation levels: green (best); and red (worst).

Video Type	VQM					SSIM				
	Network Condition					Network Condition				
	Delay[s]	Jitter[ms]	Through.[Mbps]	PLR[%]	ALL	Delay[s]	Jitter[ms]	Through.[Mbps]	PLR[%]	ALL
cr	0.308	0.994	0.954	0.842	0.775	-0.162	-0.425	0.776	0.394	0.146
do	-0.32	0.747	0.754	0.78	0.49	-0.341	0.71	0.649	0.335	0.337
hb	0.112	0.996	0.99	0.923	0.755	-0.05	0.749	0.726	0.14	0.392
ic	0.7553	0.971	0.974	-0.456	0.871	0.756	0.836	0.535	0.186	0.578
pj	0.7323	0.895	0.928	0.879	0.859	0.412	0.83	0.894	0.238	0.594
sc	-0.773	0.985	0.934	0.887	0.508	-0.73	0.777	0.529	0.283	0.217
re	-0.42	0.98	0.98	0.828	0.59	-0.486	0.823	0.795	0.623	0.439
ru	0.507	0.95	0.99	0.86	0.83	0.695	0.887	0.86	0.77	0.803
All	0.113	0.9398	0.939	0.84	0.709	0.0134	0.648	0.72	0.371	0.438
	±0.57	±0.08	±0.08	±0.05	±0.4	±0.55	±0.44	±0.14	±0.22	±0.19

In order to assess the accuracy of these two FR metrics, we picked the ReTRIEVED Video Quality database [26], [13]. This set is composed by 184 test videos, obtained from 8 different original sources. These videos (encoded to MPEG2) are characterized by a broad range of spatial and temporal information, which allows drawing general conclusions out of the assessment. The 8 original videos are then subjected to practical transmission impairment scenarios, generated by a network emulator (NETEM) and Video LAN [28]. Packet loss rate, jitter, delay, and throughput have been considered as possible distortions resulting from video transmission, whereas their values are chosen based on ITU and ETSI recommendations [28], [26]. Table I shows the characteristics of this dataset in terms of video types (acronym, name, size, frame rate, length and description) and the four condition types and levels of impairment. Each of the 184 impaired videos of this dataset was evaluated by a set of 40 people using the well-known standardized MOS scoring system.

VQM and SSIM indexes were obtained for all the videos of the video set. Their values were normalized to provide quality levels, i.e. one for "full quality" and zero for "full degradation". Subsequently, the accuracy of the full reference assessments was evaluated by means of the Pearson Correlation Index [29] to the average MOS.

Table II presents the overall correlation values of VQM and SSIM to MOS, considering all the videos of the dataset and network conditions. While each of the columns corresponds to the results of the videos subjected to a particular impairment, the rows show the average Pearson Correlation Coefficient (PCC) [29] per video type. Overall correlations across all network conditions and video types are shown in the last column and the last row, respectively. The PCC correlations are performed along the impairment level. This means, for example, that to assess the PCC of VQM to MOS for video 'cr' at 1% packet loss, the correlation is done for all quality indexes from 0 (no impairment) up to 1%.

The first conclusion from this analysis is that VQM clearly outperforms SSIM in most of the videos and conditions. Only three out of the 8 videos of the delay assessment provide better correlation with SSIM than with VQM. Furthermore, VQM has an overall correlation of over 70% in the whole dataset. This gives already an idea of the good performance of VQM. However, the standard deviation is close to 40%. The reason for this behavior can be found looking at the impairments one by one. While the results on jitter (column 2), throughput (column 3), and packet loss (PLR) (column 4) are very good (reaching values between 85% and 95%), the overall correlation in the case of the delay barely reaches 20%. This comes from the fact that the perception of delay on videos varies greatly among subjects, while the objective metric gives a more general (more objective) measurement.

In order to explore these discrepancies in greater detail, Figure 2 shows the correlation results unfolded for the level of the impairment both for VQM (Figure 2a) and SSIM (Figure 2b). Each of the four colormaps shows one impairment type (Delay, Jitter, Throughput and PLR) in the two figures. The x-axes indicate the impairment level for each of the four conditions, while the y-axis labels each of the video types with a different number (1 to 8). Dark blue is presented for very high or close to 1 PCC correlation between the metrics. When

(a) Correlation of VQM to the subjective studies on the ReTRIeVED Video Quality Database.

(b) Correlation of SSIM to the subjective studies on the ReTRIeVED Video Quality Database.
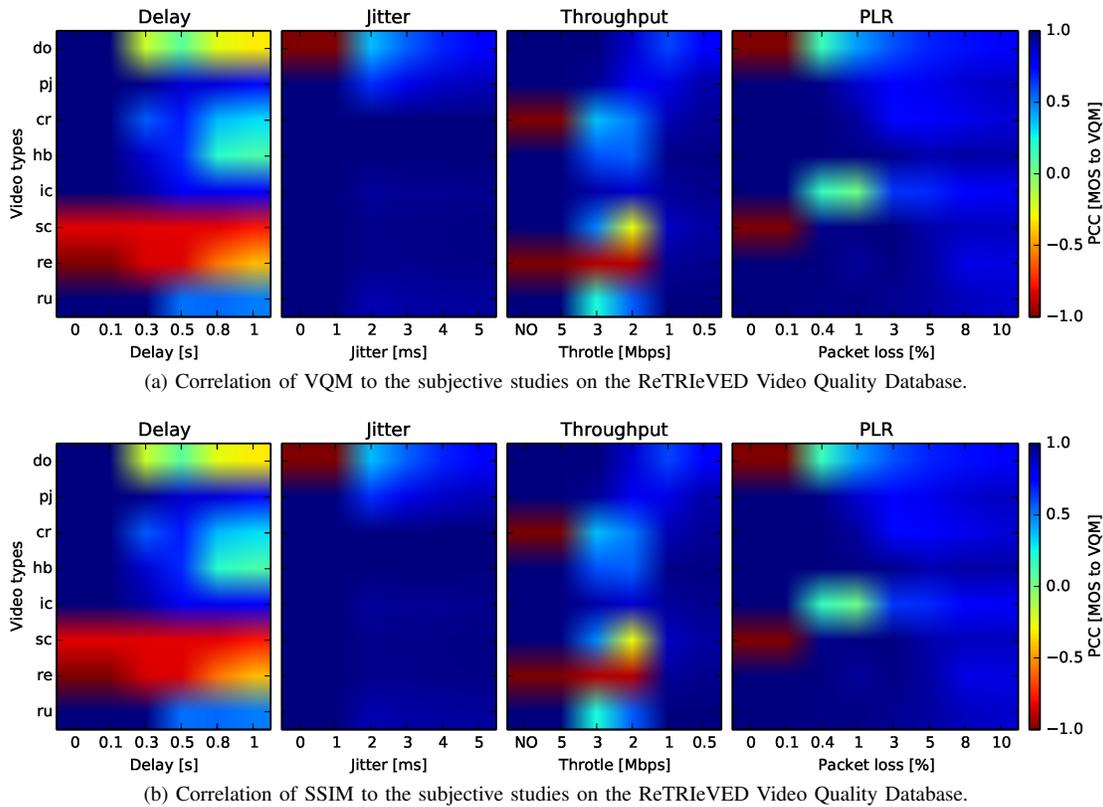
Fig. 2: PCC colormaps of the two FR candidates to MOS on the ReTRiEVED Video Quality database.

the correlation starts degrading, the blue starts fading to yellow (0 or no correlation) to a final red (-1 or anticorrelation).

Looking at the colormaps provides a more evident reasoning behind the low correlations in the delays. Basically the videos 'sc' (6th line) and 're' (7th line) show full anticorrelation to the MOS in the case of delays, while the correlation of the video 'do' degrades dramatically after 0.3 ms. Apart from that, colormaps are dominated by blue, indicating an overall good (though not perfect) correlation to the subjective assessment results (MOS). This imperfect correlation between objective (VQM) and subjective (MOS) metrics is well-known and certainly within the acceptability boundary. These results made us select VQM as the benchmark quality of preference for those cases where subjective assessment available. The evaluations shown next are performed using VQM.

B. Discovering the Most Impairing Network QoS condition: Assessing the ReTRIeVED Video Quality Database

The next step was to pinpoint the most affecting conditions on video streaming services. To achieve this, we again turned to the ReTRiEVED video quality database [26], [13] (Table I), which focuses on the effect of the most known impairments on Standard Definition (SD) videos. We followed the experimental QoS-QoE method presented in Section III.

We first calculated the VQM indexes for all the videos on the dataset. VQM assesses the degradation of videos, returning values between 'zero' (no degradation) and 'one' (full degradation) [20]. In order to provide a more intuitive assessment in terms of quality, the VQM values are inverted.

In this way, the benchmark quality index returns 'one' if the video when the highest achievable quality and 'zero' when the degradation is complete.

Table III provides the quality results of all the videos, on the four different network conditions, with three selected sensing points between the lowest to the highest level of the impairment. Furthermore, the videos have been reorganized from the least to the most resilient one. In addition to the table, Figure 3 presents the colormaps that show the evolution of the quality from unimpaired to fully-impaired network conditions.

Different video types experiment diverse degradation. This can be seen in the behavioral pattern of the most affected videos (do and pj), which for most of the network conditions suffer a higher degradation than the others. This is a known concept broadly discussed in literature. However, if we take a comparative look at the videos, it is possible to group their behavioral patterns in three different classes, according to the effect that networks have on the degradation of the videos. As such, while the videos 'do', 'pj' and 'cr' are heavily affected by the four impairments (corresponding to the first class), the videos 'hb' and 'ic' and 'sc' are less affected and follow a very similar patter (second class). Finally 're' and 'ru' are the most resilient videos and can be grouped into the third class. Looking at the composition of the videos inside each defined category, it is possible to understand the videos similar behavior. For example, the videos 'do', 'pj' and 'cr' present very complex scenes and a substantial degree of motion. The videos 're' and 'ru' present nearly static scenes and very little motion. Based on this early classification, it is possible to

TABLE III: Quality degradation by means of the benchmark quality (1-VQM) for all the videos and conditions of the ReTRiEVED Video quality Database. Videos are organized from the least to the most resilient. For each of the network impairments, three sensing points are given. Cell colors provide a qualitative degree of quality (where 0 means maximum degradation and 1 full quality): red (0-0.1), orange (0.1-0.25), yellow (0.25-0.5), green (0.5-0.75), turquoise (0.75-0.95) and dark blue (1-0.95).

Video Type	Network Condition											
	Delay[s]			Jitter[ms]			Through.[Mbps]			PLR[%]		
	0.1	0.5	1	1	3	5	5	2	0.5	0.1	3	10
do	0.55	0.549	0.52	0.49	0.27	0.14	0.67	0.65	0	0.53	0.35	0.12
pj	0.65	0.54	0.54	0.52	0.24	0	0.68	0.68	0	0.59	0.43	0.206
cr	0.67	0.6	0.68	1	0.12	0	1	1	0.1	0.6	0.49	0.21
hb	0.91	0.96	0.96	0.88	0	0	0.41	0.46	0.1	0.91	0.62	0.31
ic	0.96	0.961	0.96	0.95	0.15	0.09	0.84	0.78	0	0.96	0.59	0.345
sc	1	0.993	.9	0.86	0.07	0	0.56	0.52	0	0.97	0.48	0.17
re	0.9	0.9	0.9	0.85	0.02	0	0.64	0.64	0	0.88	0.62	0.41
ru	1	0.87	0.98	0.93	0.1	0	0.9	0.7	0	0.96	0.46	0.03
All	0.83 ±0.2	0.78 ±0.2	0.8 ±0.2	0.81 ±0.15	0.098 ±0.1	0.01 ±0.1	0.71 ±0.2	0.7 ±0.18	0 ±0.1	0.8 ±0.2	0.5 ±0.1	0.22 ±0.1
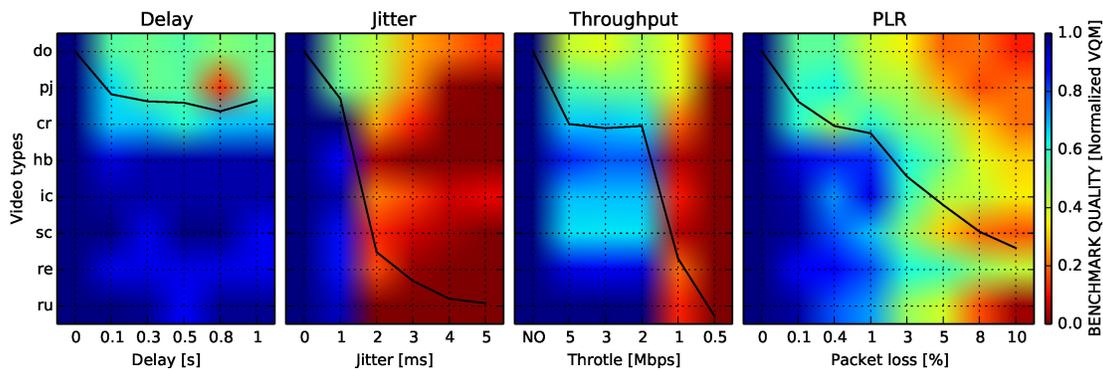
Fig. 3: Benchmark quality in all the videos of the ReTRIeVED Video Quality Database.

derive some general conclusions of the resilience about these videos to networks.

In general, the least impairing artifact is the delay. Only three of the videos (do, pj and cr, the first class) are affected by it, while the other 5 provide good resilience (average of 0.8 quality with maximum averages delays of 1 second). Moreover, the effect of delays is flat, meaning that for the videos more affected and the ones more resilient there is no significant difference between low levels of delay (such as 0.1s) or the highest ones (1s).

The jitter shows a generalized behavioral pattern for all the videos. Nearly full quality or 50% for the two most affected videos, do and pj) up to 1 ms to full degradation for any higher level of impairment. The reason for this comes from the buffering effect of the encoders on the receiver side. Jitter affects the most when the buffer is full but data is still missing. In that case, quality is completely lost. More recent encoding techniques than MPEG2 (such as H.264 and H.265) already show buffering mechanisms to counter the effects of the jitter.

Finally, the throughput and packet loss provide the broadest range of impairing effects. As a general rule, in terms of transmission/compression shrinking (throughput), as the band reduces down to 1Mbps, the quality is lost for most of the videos. Regarding the packet loss, in general videos still have reasonably good quality up to 3% drops. However, three of the videos (again 'do', 'pj' and 'cr') start loosing quality already with 0.4% (quality decreases down to roughly 0.6).

To sum up, video types and compositions get affected by network impairments in a different manner, but some general patterns can be extrapolated from a comparative analysis. Delay infers a constant and low degradation. Jitter, although very much affecting the video streaming services, is providing the same behavior independently from the video type under scrutiny. This allows for countering actions by the network provider, such as maintaining it below 1ms to stream. Finally, the effect of both network drops (packet loss) and bandwidth constraints is strong on the videos. However, it is heavily dependent on video type and condition. Thus, it is not easy to reach general rules with the assessment of this first video set.

Based on this reasoning, we decided to explore in more detail the effect of both packet loss and bandwidth. Thus, we generated and assessed the two video-sets presented and evaluated in the next two subsections.

C. Broad Analysis of 2D-SD Videos: Generating and Assessing the LIMP Video Quality Dataset

Given the lessons learned from the assessment of the ReTRiEVED dataset [26], [13], the purpose of this second analysis was to develop and assess a video quality database focusing on the effect of packet loss and bandwidth compression. For it, we made use of the test-bed presented in Section III and streamed the LIMP (Loss Impaired) Video Quality Database [14].

TABLE IV: LIMP Video Quality dataset parameters range in terms of video types (acronym, name and description), compression and network packet loss ratio. Compression bitrate is provided in kbps and the packet loss ratio in percentages. Each of the 10 original video sources is subjected to all of the 8 bandwidth constraints and the 12 packet loss levels. This means a total of 960 videos to evaluate.

Video type						Condition		
Acronym	Name	Size [Pixels]	FR [fps]	Length [s]	Description	Bitrate [kbps]	PLR	[%]
bs1	Blue Sky	768 × 432	25	10	Circular motion; Blue sky and trees	64	0	0.5
mc1	Mobile Calendar	768 × 432	25	10	Pan, horizontal tor train; Calendar vertical	640	1	1.5
pa1	Pedestrian Area	768 × 432	25	10	Still; People on intersection	768	2	2.5
pr1	Park Run	768 × 432	25	10	Pan; Person across a park	1024	3	3.5
rb1	River Bed	768 × 432	25	10	Still; River bed, pebbles in the water	2048	4	4.5
rh1	Rush Hour	768 × 432	25	10	Still; Rush hour traffic on the street	3072	5	10
sf1	Sunflower	768 × 432	25	10	Still; Bee over sunflower	4096		
sh1	Shields	768 × 432	25	10	Pan, still, zoom; Person across display	5120		
st1	Station	768 × 432	25	10	Still; Railway track, train and people			
tr1	Tractor	768 × 432	25	10	Pan; Tractor across the fields			

TABLE V: Quality degradation by means of the benchmark quality for all the videos and conditions of the LIMP Video Quality Database. Videos are organized from the least to the most resilient to packet loss and compression. 5 sensing points are provided between the minimum and maximum level of network losses. Bitrate compression of each of the videos is given in kbps and the packet loss level in percentage. Cell colors provide a qualitative degree of quality (where 0 means maximum degradation and 1 full quality): red (0-0.1), orange (0.1-0.25), yellow (0.25-0.5), green (0.5-0.75), turquoise (0.75-0.95) and dark blue (1-0.95).

Video Type	Bitrate [kbps]	PLR [%]					Video Type	Bitrate [kbps]	PLR [%]				
		0	1	3	5	10			0	1	3	5	10
sh1	5120	0.61	0.53	0.44	0.32	0.1	pr1	5120	0.78	0.65	0.5	0.46	0.24
	4096	0.6	0.54	0.35	0.29	0.01		4096	0.75	0.64	0.46	0.36	0.21
	3072	0.58	0.47	0.43	0.36	0.06		3072	0.64	0.55	0.44	0.4	0.24
	2048	0.55	0.49	0.5	0.38	0.24		2048	0.56	0.47	0.45	0.39	0.29
	1024	0.51	0.55	0.4	0.31	0.23		1024	0.4	0.38	0.34	0.32	0.22
	768	0.47	0.41	0.33	0.33	0.2		768	0.36	0.33	0.3	0.27	0.27
	640	0.45	0.37	0.37	0.36	0.17		640	0.33	0.3	0.3	0.25	0.24
	64	0.024	0.011	0	0	0		64	0.04	0.04	0.04	0.04	0.03
rb1	5120	0.79	0.64	0.54	0.45	0.42	tr1	5120	0.89	0.63	0.51	0.34	0.15
	4096	0.76	0.61	0.52	0.44	0.4		4096	0.86	0.7	0.45	0.4	0.09
	3072	0.67	0.58	0.49	0.47	0.39		3072	0.81	0.6	0.47	0.33	0.18
	2048	0.6	0.51	0.46	0.44	0.36		2048	0.74	0.61	0.51	0.36	0.22
	1024	0.46	0.44	0.42	0.4	0.37		1024	0.59	0.48	0.41	0.4	0.29
	768	0.42	0.39	0.36	0.36	0.32		768	0.53	0.41	0.32	0.37	0.31
	640	0.39	0.36	0.36	0.33	0.32		640	0.49	0.36	0.32	0.31	0.27
	64	0.22	0.23	0.24	0.21	0.21		64	0.02	0	0	0.01	0.01
pa1	5120	0.94	0.58	0.42	0.33	0.18	mc1	5120	0.88	0.69	0.41	0.34	0.12
	4096	0.91	0.64	0.4	0.34	0.15		4096	0.85	0.62	0.35	0.29	0.18
	3072	0.86	0.61	0.44	0.34	0.19		3072	0.81	0.66	0.43	0.32	0.09
	2048	0.78	0.575	0.44	0.39	0.29		2048	0.76	0.67	0.46	0.4	0.17
	1024	0.59	0.51	0.4	0.36	0.29		1024	0.67	0.56	0.34	0.35	0.21
	768	0.51	0.45	0.35	0.33	0.25		768	0.62	0.49	0.47	0.32	0.27
	640	0.47	0.41	0.35	0.31	0.28		640	0.58	0.43	0.41	0.37	0.25
	64	0.13	0.12	0.12	0.12	0.08		64	0.18	0.16	0.15	0.15	0.14
rh1	5120	0.96	0.76	0.52	0.36	0.17	st1	5120	0.96	0.81	0.61	0.47	0.3
	4096	0.94	0.64	0.53	0.3	0.16		4096	0.91	0.78	0.53	0.5	0.32
	3072	0.9	0.69	0.5	0.31	0.14		3072	0.88	0.81	0.62	0.55	0.36
	2048	0.83	0.74	0.51	0.41	0.17		2048	0.83	0.75	0.65	0.45	0.43
	1024	0.67	0.58	0.52	0.36	0.24		1024	0.71	0.65	0.57	0.56	0.44
	768	0.59	0.47	0.38	0.39	0.28		768	0.67	0.6	0.58	0.51	0.39
	640	0.54	0.44	0.38	0.35	0.19		640	0.64	0.56	0.54	0.47	0.29
	64	0.16	0.14	0.13	0.14	0.1		64	0.23	0.23	0.23	0.19	0.21
sf1	5120	0.96	0.84	0.59	0.52	0.31	bs1	5120	0.95	0.85	0.61	0.51	0.29
	4096	0.96	0.77	0.7	0.54	0.23		4096	0.94	0.81	0.6	0.52	0.45
	3072	0.93	0.84	0.61	0.48	0.29		3072	0.91	0.8	0.58	0.49	0.4
	2048	0.88	0.76	0.67	0.53	0.36		2048	0.87	0.8	0.64	0.55	0.41
	1024	0.78	0.67	0.62	0.61	0.39		1024	0.78	0.69	0.64	0.52	0.44
	768	0.72	0.65	0.57	0.51	0.36		768	0.74	0.67	0.6	0.59	0.46
	640	0.69	0.56	0.59	0.49	0.37		640	0.71	0.62	0.57	0.51	0.5
	64	0.12	0.12	0.11	0.1	0.07		64	0.3	0.29	0.28	0.28	0.29

Ten original raw high quality videos were obtained from the Live Quality Video Database [30], [31] (10 seconds and 25fps). Each of these videos is of a different dynamic composition and type (Table VI). These videos were compressed

to H.264/MPEG4 with a resolution of 768x432 at 8 different bitrates (64kbps, 640kbps, 768kbps, 1024kbps, 2048 kbps, 3072kbps, 4096kbps and 5120 kbps). We selected H.264 given its higher robustness and efficiency than MPEG2. The selection of the compression bitrates was done to obtain the most diverse variety of video qualities. For example, one could think that with the new ultra-high speed networks (such as the one envisioned by 5G or optical wireless), very low quality transmissions (64kbps) would not be necessary anymore. However, these rates are still visible in congested mobile networks or in adaptive streaming applications that bootstrap at the lowest bitrate and then slowly adapt the quality to the actual network conditions. With this variety of bitrates, our dataset covers a broad range of video types, which allows for a comprehensive QoS-QoE analysis.

In addition to bandwidth constraints, packet loss has been demonstrated to be the main cause of degradation in RTP video transmissions [17], [32], as we also showed in the previous Section. Thus, for the generation of the full dataset, 80 original videos (10 types at 8 compression levels each) were transmitted from server to client through a lossy network (using the experimental test set-up presented in Section III). Each video type and bitrate was streamed through the network 11 times. Each iteration the video was subjected to a different levels of packet loss (0.5%, 1%, 1.5%, 2%, 2.5%, 3%, 3.5%, 4%, 4.5%, 5% and 10%). The final output was a video set consisting of 960 videos, obtained from 10 original videos, each encoded at 8 different bitrates and 12 different conditions (1 compression degradation + 11 compression degradation and packet loss).

As in the previous case, we performed a VQM assessment of all the videos. The original reference video was obtained from the original 10 videos encoded at maximum quality (given by the encoder ffmpeg [33]). Once the VQM indexes were obtained, they were inverted to measure quality instead of degradation.

Table V presents the quantitative results of each of the videos for the different compression rates. In this case five sensing points are provided (0%, 1%, 3%, 5% and 10%). The videos are again organized from the least to the most resilient. Two sets of colormaps present the views of the data focused on the video type (Figure 4a) and the compression rate (Figure 4b). Finally, averages per video types and bitrate are provided in Figure 4c and Figure 4d. These five elements provide a comprehensive analysis of the effect of packet loss and compression on 2D-SD videos.

First, we can have a look at the effect of networks losses and compression on each of the video types. As it is to be expected, some of them are more resilient than others. While sh1 is the most impaired video (Figure 4a, first row first column) provides a very impaired pattern during both sweeps of compression and packet loss, bs1 (Figure 4a, second row, fifth column) maintains good levels of quality (over 70%) up to 3% packet loss. This gives an idea of the variety of the videos of the set.

However, looking at the video averages (Figure 4c), it is possible, as in the previous case, to classify the videos according to their resilience to networks. On the least resilient

end, the videos sh1 and pr1 show already from the start, very bad performance (meaning that they are heavily influenced by the compression impairment). Furthermore, their average quality drops below 0.4 from 2% packet loss on. On the other end of the spectrum, bs1, sf1 and st1 maintain average quality values higher than 0.5% up to 5%. In between those two patterns, rb1 has the flattest pattern of all, during the full packet loss sweep the quality degrades barely 20%. This gives the impression that the complexity of this video is more affected by video compression than by network losses. The final four videos have a similar trend, keeping average values over 0.5 up to 2%. It is worth noting that the low average results come from the influence of the 64kbps version of the videos, which brings the average qualities down. As in the case of the ReTRiEVED video set, there is a high correlation between the composition of the videos and their resilience to networks. While sh1 and pr1 present very rich scene and fast motion, sf1, bs1 and st1 present very low motion and scene complexity.

If we look at the behavior from the compression point of view, as a general note, as the bitrate decreases, the videos' robustness to network losses increases. This counter-intuitive effect can be seen on Table V. Bitrates from 768kbps to 64kbps suffer a maximum of 30% degradation over the full packet loss sweep, while higher bitrates, such as 5, 4 or 3 Mbps, suffer 70% degradation in the least. This effect can also be seen in the figures (the three last colormaps on Figure 4b and the grey, brown and cyan lines of Figure 4d). In addition, all of the averaged qualities except the 64kbps, are kept over 0.4 up to 5% packet loss.

The bitrates can be easily classified according to their strength against packet loss. 64kbps, as it was expected, provides the lowest quality at all times. 640kbps and 768kbps provide very low quality but similar quality. In the same way, 1 and 2 Mbps have their own performance group and the highest bitrates (5, 4 and 3 Mbps) as well. Putting all these notions to service and network providers could provide many hints as to how to use certain videos and networks, as we show in Section V.

D. Broad Analysis of Stereoscopic 3D Videos: Generating and Assessing the 3D-HEVC-Net Video Quality Dataset

3D video streaming application are booming [34], to fulfill the demand of high-quality media [35]. However, with their high quality and bandwidth requirements come more stringent constraints on networks [36]. In this situation, it is fundamental to understand how the network limitations impact quality [23]. Early studies, such as the one presented by Hewage et al. [37] or Politis et al. [36] have shown the effects of QoS artifacts onto user QoE in small sets of video-samples. Our purpose was to study the non-linear distortion effects that lossy networks have onto 3D videos in a general, systematic manner with a broad range of video types and conditions.

A freely available dataset of high quality 3D Stereoscopic video sequences was chosen for the proposed experimental analysis [35]. This was composed of 10 high-definition stereoscopic sequences. The resolution of the sequences is
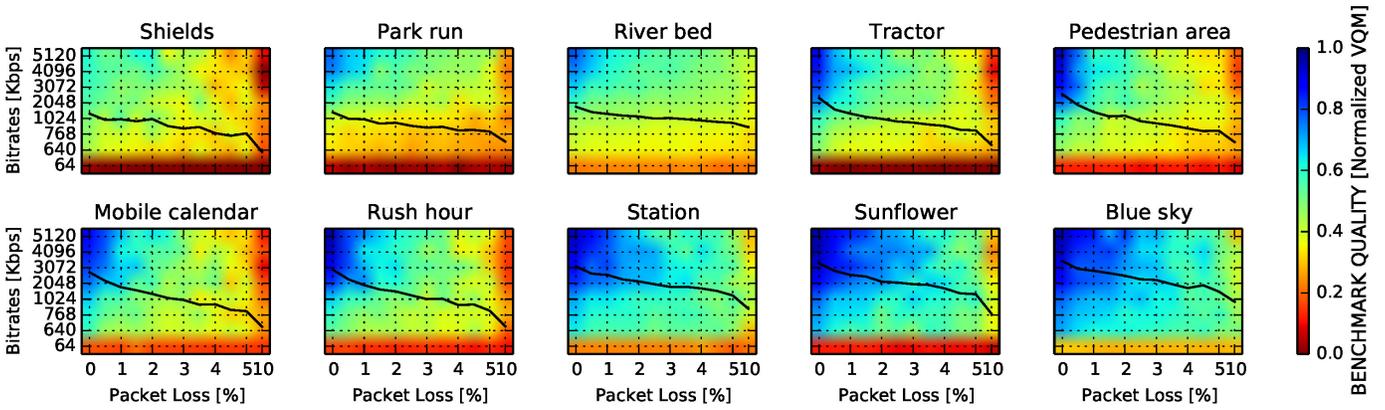
(a) Quality colormaps of the LIMP Quality Database organized per video type.
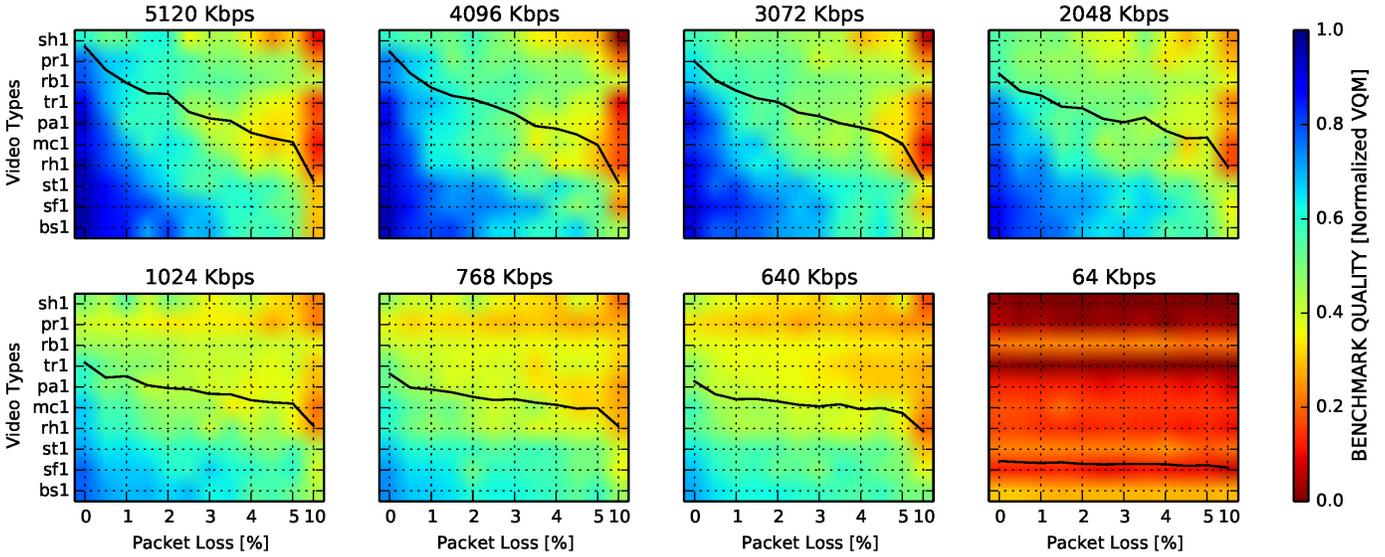
(b) Quality Colormaps of the LIMP Video Quality Database organized per bitrate compression.
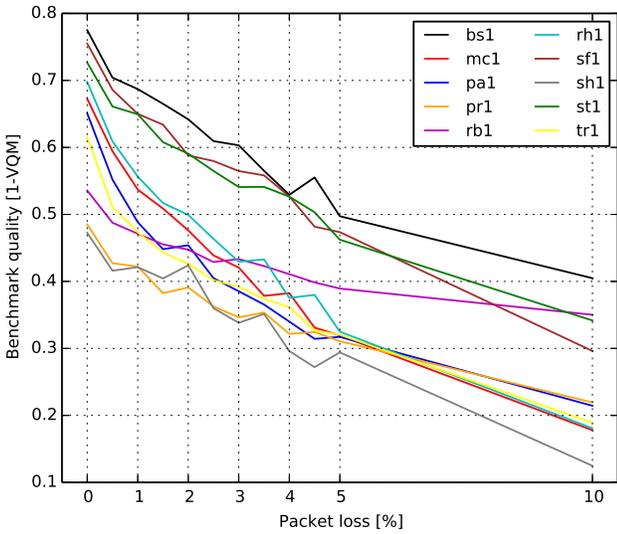
(c) Averages degradation per video type.
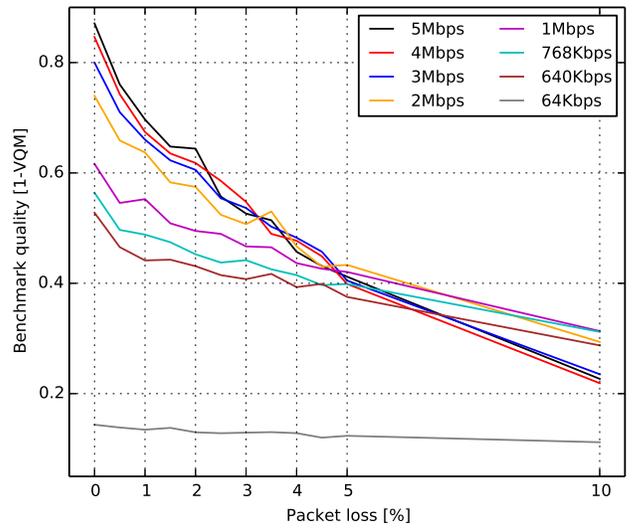
(d) Averages degradation per bitrate.

Fig. 4: Quality results for all the videos, bitrates and packet loss levels of the LIMP Video Quality database. The videos are organized from the least to the most resilient to impairments. The bitrates go from lowest to highest compression.

1920×1080 pixels, the color sampling $YUV420$, the frame rate $25fps$, and the duration $16s$ for all the sequences except

for the last one ($13s$). The stereoscopic video sequences were first processed in order to obtain a side-by-side dataset of Full-

HD videos.

Common formats for 3D video are conventional stereo video (CSV), mixed resolution stereo (MRS), and video plus depth (V+D) [38]. For the generation of the dataset, we chose the CVS format with frame-compatibility in side-by-side arrangement [39]. The frame-compatibility approach combines the left and the right views into a standard 2D video format, keeping in this way the compatibility towards 2D video coding tools. The stereoscopic frames can be arranged in a 2D frame in several ways. Common approaches are side-by-side, top-and-bottom, row interleaving, column interleaving. A side-by-side frame is obtained by subsampling the left and the right frame of a stereoscopic sequence along the horizontal axis by a factor of two. Hence the new left and right views embedded in a side-by-side had resolution 960×1080.

Table VI provides a description of the video types, descriptions, bitrates and network conditions. We compressed each of the original stereoscopic, 3D, two-views, 10 sequences to H.265/HEVC [40] at six compression levels (0.5Mbps, 1Mbps, 1.5Mbps, 2Mbps, 3Mbps, and 4Mbps) with a GOP of 15 frames (1 I, 3 P, 11 B). Subsequently, each of the compressed sequences was merged into a single stream for each bitrate. Each stream (total length $157s$) was sent over the emulated network configured at 7 levels of packet loss (0.1%, 0.2%, 0.5%, 1.0%, 3.0%, 5.0%, 10.0%), for a total of 42 testing condition. Each streaming experiment was repeated 10 times for each bitrate and for each drop rate. On reception, the video sequences were split and assessed in a full-reference manner. This made a total of 4200 videos for analysis (10 iterations of 10 videos at 6 compressions and 7 packet loss levels) [41].

As in the previous two cases, we performed a VQM assessment of all the videos. For it, the original reference video was picked on the highest bitrate HEVC versions of the videos (i.e. 4Mbps). In this way, 1/6 of the data set (videos encoded at 4Mbps) evaluated only the effect of the network, while the remaining 5/6 provided compression and network loss assessment. Furthermore, we used VQM due to its proven performance in 3D Stereoscopic [37]. Subsequently, values were inverted, as in the previous datasets.

As for the 2D videos, we show a two folded (videos and compression) analysis by means of five elements. Table VII presents the quantitative results of each of the videos at the different compression rates. In this case, as there are more packet loss levels, six points of the evolution are given (0%, 0.5%, 1%, 3% , 5% and 10%). The videos of the table are organized from the least to the most resilient to drops. Two sets of colormaps present the views of the data focused on the video type (Figure 5a) and the compression rate (Figure 5b). Finally, we presents the averages over the video types and bitrates across all videos (Figure 5c) and bitrates (Figure 5d).

We can first have a look at the effect of networks losses and compression per video type. Similarly to 2D videos, some of the 3D videos present more resilient patterns than others. While 'tb' (Figure 5a, first row first column) degrades dramatically already for low packet loss (with a 0.5% of losses, its quality is already down to 0.3) for all its compression variants, 'pc' (Figure 5a, second row, fifth column) maintains acceptable levels of quality (higher than 0.5) up to 3% packet loss.

In the same way as it occurred with the 2D videos, it is still possible to classify the behavioral pattern of each of the videos within different categories, taking into account their performance against the network impairments. On one side of the spectrum, the video 'tb' conforms its own category having the worst performance of all from the start. It keeps average values below 0.15, from 1% packet loss. On the other end, 'nr' and 'pc' provide values close to 0.7 for 1% and only drop below 0.5 for losses higher than 3%. In between those two patterns, 'sc', 'hl' and 'ul' outperform 'tb'. Finally 'lb', 'bg', 'bk', and 'bx' are in average degraded to their 50% with more than 1% packet loss.

Studying the behavior from the point of view of the compressions provides the first striking result. Unlike, 2D videos, in which the quality was gradually dropping as the losses increased, 3D videos show dramatic drops, going from qualities of 0.9-0.8 at 0.1%-0.2% to roughly 0.5 for 1% packet loss and down to 0.25 for 3%. The reason for this comes from the higher resolutions of the videos and bitrates, apart from the fact that 3D videos transport more complex information within the two views. In addition, as in the 2D case, as the bitrate decreases, the resilience of the videos to packet loss increases (Figure 5b).

Finally, the bitrates can be easily classified according to their strength against packet loss. The highest compression (500Kbps) provides the lowest quality of the group, up to 3%. From that moment on, it over-passes all the other compressions. This proves again the statement that lower compressed videos are more resilient to network losses. The second highest compression (1Mbps) has a similar behavioral pattern but with a higher slope, providing lower quality than all the lower compressions (1.5 Mbps to 4Mbps) up to 1% losses. After that, it offers better performance than all of them together. The behavioral pattern of each of the other compressions can be classified in two groups. Medium-low compression (1.5, 2 Mbps) starts lower than the highest bitrates, but provides a better average performance than the highest rates up to 1%. Finally, 3Mbps and 4Mbps conform their own category with a very similar pattern. This is also interesting, given the fact that 4Mbps only carries network degradation, while 3Mbps has degradations from both compression and drops.

These notions and the ones extracted from the 2D videos can be used in a real network or video server to obtain a grasp of the state of their network and videos and adjust in case it is necessary. The next section provides some hints for a real case scenario.

V. Discussion: Hints for Service and Network Providers

The purpose of this section is to provide insights derived from the study of the three datasets which could, in turn, be instrumental to video services and network providers. Suppose a video service provider has stored video content both on 2D-SD and 3D-HD formats. For illustration purposes we have focused on the LIMP and the 3D-HEVC-Net datasets. In order to adapt the streamed videos to the network and client

TABLE VI: 3D-HEVC-Net Video Quality dataset parameters range in terms of video types (acronym, name and description), compression and network packet loss ratio. The videos bitrates are given in Mbps. The packet loss ratios are given in percentages. Each of the videos was streamed at one of the 6 bitrates and one of the 8 PLR levels. Each of the experiments was repeated 10 times. This makes a total of 4000 videos to evaluate.

Video type							Condition	
Acronym	Name	Size [Pixels]	FR [fps]	Length [s]	Description		Bitrate [Mbps]	PLR [%]
bg	Barrier gate	1920×1080	25	16	Still on gate. Barrier opens, car passes.		0.5	0
bk	Basket	1920×1080	25	16	Still on basketball game; Pan, action from one side to another.		1	0.1
bx	Boxers	1920×1080	25	16	Still on boxer training;still on two boxers fighting.		1.5	0.2
hl	Hall	1920×1080	25	16	Still on hotel hall from ceiling;two people meet in one corner.		2	0.5
lb	Lab	1920×1080	25	16	Still on two scientists working on a lab.		3	1
nr	News Report	1920×1080	25	16	Still on two men sitting behind table, reading the news.		4	3
pc	Phone Call	1920×1080	25	16	Still on man behind a table picking up the phone.			5
sc	Soccer	1920×1080	25	16	Still from behind the soccer goal; men scoring.			10
tb	Tree Branches	1920×1080	25	16	Still on tree branches; leaves moved by the wind.			
ul	Umbrella	1920×1080	25	13	Still; man opens umbrella and starts turning it.			

TABLE VII: Quality degradation by means of the benchmark quality for all the videos and conditions of the 3D-HEVC-Net Video Quality Database. Videos are organized from the least to the most resilient to packet loss and compression. Compression bitrate is provided in Mbps; Packet loss ratio in percentages. Cell colors provide a qualitative degree of quality (where 0 means maximum degradation and 1 full quality): red (0-0.1), orange (0.1-0.25), yellow (0.25-0.5), green (0.5-0.75), turquoise (0.75-0.95) and dark blue (1-0.95).

Video Type	Bitrate [Mbps]	PLR [%]						Video Type	Bitrate [Mbps]	PLR [%]					
		0	0.5	1	3	5	10			0	0.5	1	3	5	10
tb	4	1	0.3	0	0	0	0	hl	4	0.86	0.47	0.35	0	0	0
	3	0.78	0.34	0	0	0	0		3	0.86	0.46	0.35	0	0	0
	2	0.7	0.4	0.27	0	0	0		2	0.8	0.47	0.36	0	0	0
	1.5	0.61	0.43	0.34	0	0	0		1.5	0.74	0.47	0.08	0	0	0
	1	0.48	0.3	0.29	0.06	0	0		1	0.63	0.42	0.37	0.21	0.02	0
	0.5	0.3	0.24	0.19	0.15	0.13	0		0.5	0.42	0.34	0.29	0.15	0.12	0
sc	4	1	0.48	0.16	0	0	0	ul	4	1	0.62	0.24	0	0	0
	3	0.9	0.46	0.21	0	0	0		3	0.78	0.5	0.33	0	0	0
	2	0.85	0.45	0.23	0.04	0	0		2	0.7	0.45	0.4	0.001	0	0
	1.5	0.78	0.47	0.34	0.02	0	0		1.5	0.63	0.44	0.42	0.17	0	0
	1	0.67	0.5	0.34	0.1	0.02	0		1	0.52	0.42	0.37	0.26	0.07	0
	0.5	0.42	0.34	0.3	0.17	0.14	0		0.5	0.36	0.32	0.29	0.23	0.22	0.06
bk	4	1	0.66	0.53	0	0	0	lb	4	1	0.62	0.43	0	0	0
	3	0.86	0.6	0.45	0	0	0		3	0.83	0.51	0.38	0.02	0	0
	2	0.79	0.54	0.45	0.15	0	0		2	0.75	0.43	0.43	0.14	0	0
	1.5	0.73	0.53	0.46	0.24	0.03	0		1.5	0.67	0.49	0.42	0.24	0	0
	1	0.63	0.47	0.46	0.32	0.18	0		1	0.56	0.45	0.35	0.24	0.16	0
	0.5	0.43	0.38	0.33	0.24	0.19	0.09		0.5	0.35	0.32	0.27	0.16	0.14	0.05
bg	4	1	0.64	0.43	0	0	0	bx	4	1	0.73	0.58	0.17	0	0
	3	0.84	0.54	0.39	0	0	0		3	0.85	0.62	0.5	0.19	0	0
	2	0.77	0.55	0.43	0.13	0	0		2	0.78	0.56	0.52	0.3	0.11	0
	1.5	0.71	0.52	0.4	0.26	0	0		1.5	0.72	0.62	0.51	0.37	0.16	0.002
	1	0.61	0.5	0.4	0.26	0.16	0		1	0.61	0.56	0.42	0.32	0.25	0.03
	0.5	0.41	0.34	0.32	0.22	0.17	0.12		0.5	0.42	0.38	0.37	0.28	0.22	0.12
nr	4	1	0.82	0.7	0.39	0.02	0	pc	4	1	0.83	0.63	0.32	0	0
	3	0.93	0.74	0.67	0.37	0.2	0		3	0.92	0.78	0.66	0.41	0.2	0
	2	0.88	0.75	0.67	0.39	0.28	0.02		2	0.87	0.77	0.73	0.51	0.41	0.06
	1.5	0.83	0.73	0.63	0.43	0.33	0.06		1.5	0.83	0.73	0.7	0.52	0.43	0.18
	1	0.75	0.69	0.62	0.43	0.31	0.18		1	0.74	0.67	0.64	0.47	0.44	0.29
	0.5	0.56	0.53	0.52	0.39	0.35	0.2		0.5	0.56	0.54	0.52	0.43	0.39	0.29

requirements, the content and compressions available could be classified. In the previous section we introduced a possible classification of the content according to the video resilience to networks, by direct analysis and grouping of video behavior. (Table VIII). In addition, the resilience is heavily dependent on the content type and characteristics (for example in terms of scene composition and video motion). This notion could be used to cluster and pre-classify content types according to their composition (for example using a classification similar to the previous one). As new videos are added into the system, these could be classified according to their characteristics (such as

complexity, motion, etc.) within one of the known classes. This means that the administrators would only be required to have a small set of analysed video classes available (such as the one we present in Table VIII) in order to understand the behavior of a wide range of video services, making the solution scalable.

The next step would be to understand the sensitivity of the videos and compression classes to network losses. This can be done by a psychometric curve fitting. With this notion, the quantitative quality indexes of the videos, and the classifications (Table VIII), it is possible to fit a sigmoid psychometric
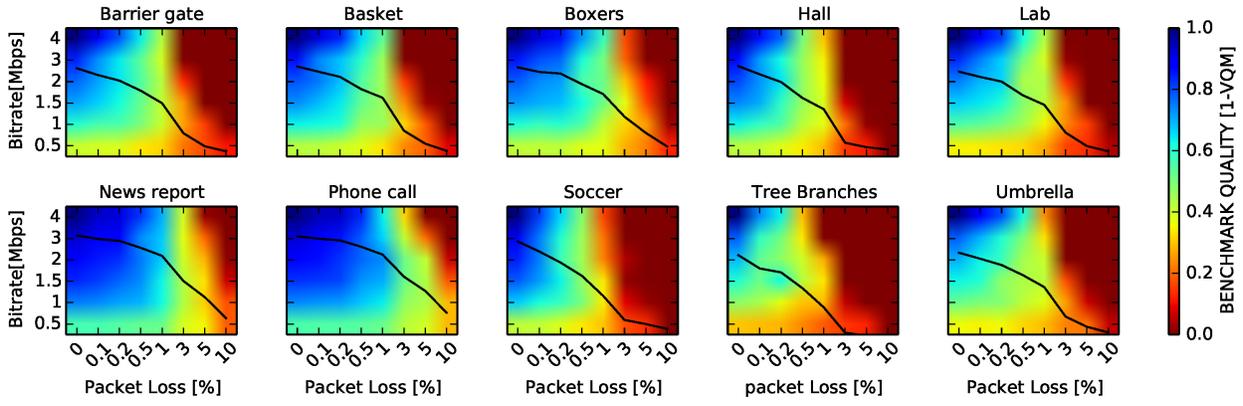
(a) Quality colormaps for the 3D-HEVC-Net Video Quality Database organized per video type.
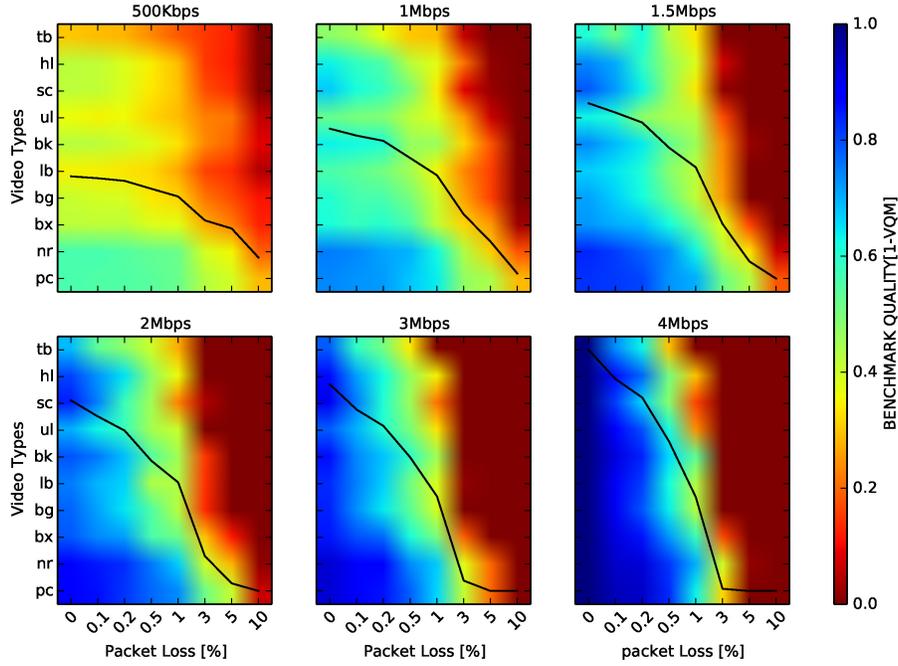
(b) Quality colormaps for the 3D-HEVC-Net Video Quality Database organized per bitrate.
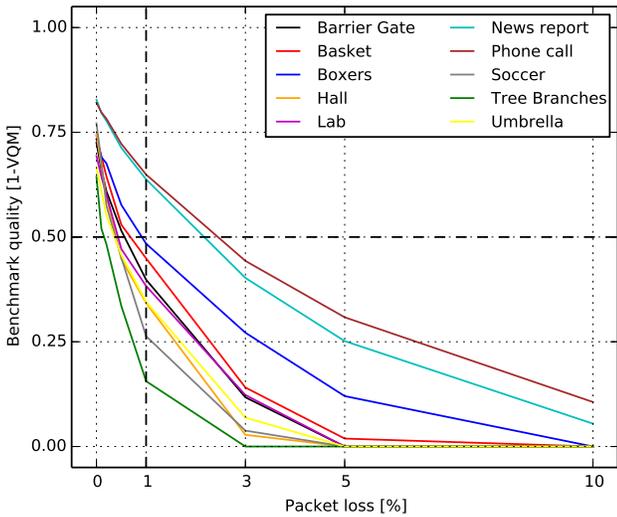
(c) Averages degradation per video type.
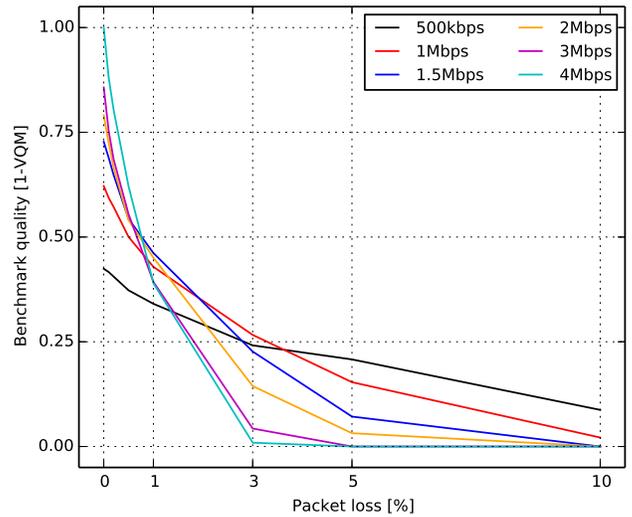
(d) Averages degradation per bitrate.

Fig. 5: Quality plots for all the videos, bitrates and packet loss levels of the 3D-HEVC-Net Video Quality database. The videos are organized from the least to the most resilient to impairments. The bitrates go from lowest to highest compression.

TABLE VIII: Classification of the videos and bitrates of the LIMP and the 3D-HEVC-Net video quality databases. Video classes are organized according to their resilience to network losses in four classes per video-set, ranging from 1 (least resilience to network drops) to 4 (most resilient to packet loss). Bitrates classes are organized in four categories: VHC (Very High Compression, very low bitrates), MHC (Medium High Compression, medium low bitrates), MLC (Medium Low Compression, medium high bitrates) and VLC (Very Low Compression, very high bitrates).

2D-SD Videos (LIMP)				3D-HD Videos (3D-HEVC-Net)			
Video Class				Video Class			
Class 1	Class 2	Class 3	Class 4	Class 1	Class 2	Class 3	Class 4
sh1	rb1	tr1	st1	tb	sc	lb	nr
pr1		pa1	sf1		hl	bg	pc
		mc1	bs1		ul	bk	
		rh1			bx		
Comp. MP4/H.264 Part 10				HEVC/H.265			
VHC	MHC	MLC	VLC	VHC	MHC	MLC	VLC
64Kbps	640Kbps	1Mbps	3Mbps	500Kbps	1000Kbps	1500Kbps	3000Kbps
	768Kbps	2Mbps	4Mbps			2000Kbps	4000Kbps
			5Mbps				
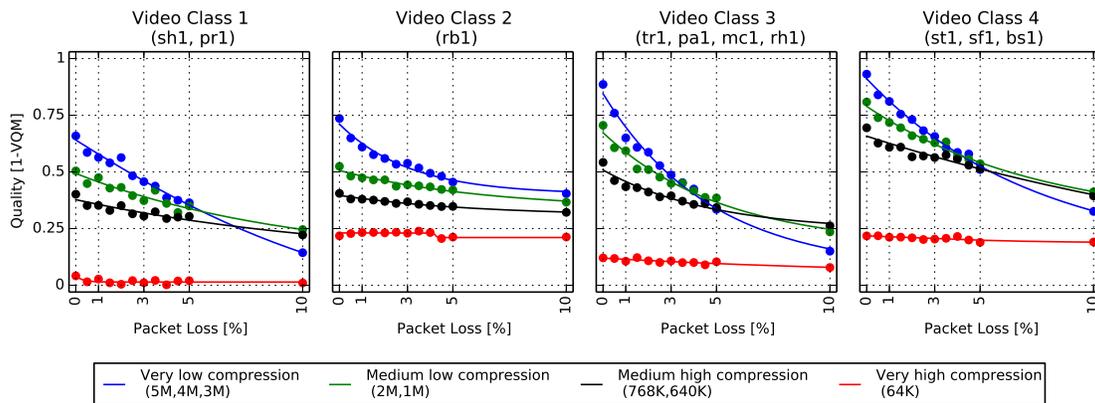
(a) Psychometric fits per class video types and compression of the LIMP Video Quality Database.
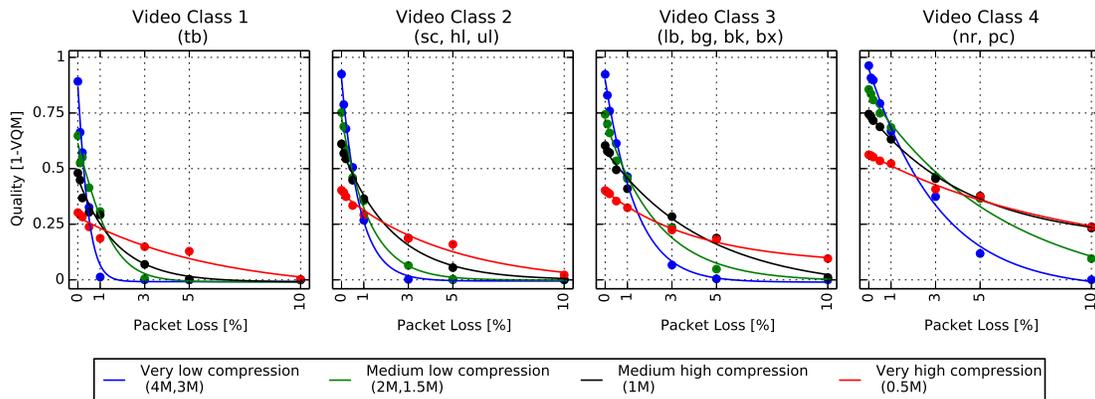
(b) Psychometric fits per class video types and compression of the Stereoscopic 3D videos of the 3D-HEVC-Net Video Quality Database.

Fig. 6: Psychometric fits for the videos of the two video datasets. a) LIMP Video Quality Database; b) 3D-HEVC-Net Video Quality Database.

curve to each of the video and bandwidth classes (Figure 6). From these two sets of psychometric curves it is evident to see the difference on resilience to packet loss between 2D and 3D. While for the 2D videos transmitting with highly lossy networks (3-5%) can still be recommendable for transmission for most video types, transmitting any of the 3D videos with network losses higher than 1% inflict degradations higher than 50%.

The information contained in this type of plots brings with

it the understanding of the trade-off between video classes and compressions to the next level. Such an analysis can be used from the point of view of the network provider to adjust/improve the quality of the network, when a given video class and compression type are required by a client. One example of this type of assessment and management can be found on Table IX. It shows an analysis of the best suited video-bandwidth to the various conditions. In it, the lossy network conditions are split in five intervals, ranging from

TABLE IX: Video and compression class combinations depending on the sensed network conditions interval. Videos from the two selected sets 2D-SD (LIMP) and 3D-HD (3D-HEVC) are shown classified according to the video and bandwidth classes presented in Table VIII. Packet loss ratios are divided in five intervals: no packet loss (0%); low packet loss (0.1-1%); medium packet loss (1-3%); high packet loss (3-5%); very high packet loss (5-10%).

Videos		Packet loss Interval [%]				
Set	Class	0	0.1-1	1-3	3-5	5-10
2D SD	1	VLC	VLC	VLC	VLC	MLC
	2	VLC	VLC	VLC	VLC	VLC
	3	VLC	VLC	VLC	MLC	MHC
	4	VLC	VLC	VLC	MLC	MHC
3D HD	1	VLC	MLC	VHC	VHC	VHC
	2	VLC	MHC	VHC	VHC	VHC
	3	VLC	VLC	MHC	VHC	VHC
	4	VLC	VLC	MLC	MHC	VHC

no losses (0% packet loss) to a fully impaired scenario (5%-10% packet loss). Based on the psychometric curve fits shown in Figure 6, and the loss intervals defined, Table IX presents the best suited bandwidth, for each of the video categories both in 2D and in 3D. Tables such as this one will provide valuable to network and service providers, to pursue resource efficiency and user satirsfaction (trades-offs). On the one hand, from the point of view of the video service provider, if a client were to request a 3D-HD video of class 2, this would only be transferred at the highest quality if the network is clean (0% loss). However, where any disturbances were sensed on the network, the video provider could decide to transmit the video on a higher compression rate, to prevent loss of quality on the client side. On the other hand, if we take the network manager perspective, when the requested 3D-HD video class 2 were to be set to the maximum transmission rate (the client requests the best quality for its 3D video transmission), the network manager would be required to ensure for the network to remain clean (i.e., having virtually no losses) during the whole streaming session. When this could not be guaranteed, other improving actions would be needed. Such actions can range from prioritizing the client's traffic (in the case of end-to-end networks) or increasing transmission power (for indoor wireless communications).

VI. CONCLUSION

When dealing with lossy networks, a mere QoS-based analysis comes insufficient. We find, that in this case quality emerges as an end-to-end factor, for it is particularly sensitive to the end-user perception of the overall service, i.e. the user's experienced quality (QoE).

To better explore the value of assessing QoE alongside QoS in high-speed, lossy networks, in this work we have presented an experimental, systematic, general evaluation methodology to map QoS onto QoE. To demonstrate the generality of our approach and understand the behavioural differences among video compressions and types, we have generated three different video sets (one of 2D videos and one of stereoscopic 3D videos).

Derived from our analysis, we have shown the high resilience of RTP-video streaming services against temporal impairments such as network delay and jitter. In addition, we have pinpointed the resilient differences among different type of 2D video streaming services compared to 3D videos.

Our study and methodology are meant to provide service providers with the means to pinpoint the working boundaries of their video sets in face of different network conditions. Based on the results shown on the video sets, we have used psychometric functions to provide with a comprehensive analysis of the implications this type of evaluation can have for service and network providers.

REFERENCES

[1] L. Atzori, C. W. Chen, T. Dagiuklas, and H. R. Wu, "QoE management in emerging multimedia services," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 18–19, 2012.

[2] J. Joskowicz and R. Sotelo, "A model for video quality assessment considering packet loss for broadcast digital television coded in H.264," *International Journal of Digital Multimedia Broadcasting*, vol. 2014, 2014, cited By 8.

[3] S. Barakovic and L. Skorin-Kapov, "Survey and Challenges of QoE Management Issues in Wireless Networks," *Journal Comp. Netw. and Communic.*, vol. 2013, pp. 165 146:1–165 146:28, 2013.

[4] R. Stankiewicz and A. Jajszczyk, "A Survey of QoE Assurance in Converged Networks," *Comput. Netw.*, vol. 55, no. 7, pp. 1459–1473, May 2011.

[5] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A Generic Quantitative Relationship Between Quality of Experience and Quality of Service," *Netwrk. Mag. of Global Internetwkg.*, vol. 24, no. 2, pp. 36–41, Mar. 2010.

[6] Cisco, "VNI forecast," http://ciscovni.com/vniforecast/index.htm.

[7] Y. Han, Z. Yuan, and G. M. Muntean, "An Innovative No-Reference Metric for Real-Time 3D Stereoscopic Video Quality Assessment," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 654–663, Sept 2016.

[8] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet," *IEEE Multimedia*, vol. 18, no. 4, pp. 62–67, 2011.

[9] Z. Duanmu, K. Zeng, K. Ma, A. Rehman, and Z. Wang, "A quality-of-experience index for streaming video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 154–166, Feb 2017.

[10] D. Z. Rodríguez, R. L. Rosa, E. C. Alfaia, J. I. Abraho, and G. Bressan, "Video Quality Metric for Streaming Service Using DASH Standard," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 628–639, Sept 2016.

[11] S. Zhao, Z. Li, D. Medhi, P. Lai, and S. Liu, "Study of user qoe improvement for dynamic adaptive streaming over http (mpeg-dash)," in *2017 International Conference on Computing, Networking and Communications (ICNC), Sillicon Valley, USA*, Jan 2017, pp. 566–570.

[12] M. Zhao, X. Gong, J. Liang, W. Wang, X. Que, Y. Guo, and S. Cheng, "Qoe-driven optimization for cloud-assisted dash-based scalable interactive multiview video streaming over wireless network," *Signal Processing: Image Communication*, vol. 57, pp. 157 – 172, 2017, medical Image Communication, Computing and Security.

[13] F. Battisti, M. Carli, and P. Paudyal, "ReTRiEVED Video Quality Database," http://www.comlab.uniroma3.it/retrieved.htm.

[14] M. Torres Vega and A. Liotta, "LIMP-Video Quality Database," https://www.tue.nl/index.php?id=53688.

[15] M. Shahid, A. Rossholm, B. Lövström, and H. Zepernick, "No-reference image and video quality assessment: a classification and review of recent approaches," *EURASIP J. Image and Video Processing*, vol. 2014, p. 40, 2014.

[16] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, Sep. 2008.

[17] 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, Krakow, Poland*, May 2014, pp. 1–5.

[18] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[19] A. K. Moorthy and A. C. Bovik, "Visual quality assessment algorithms: What does the future hold?" *Multimedia Tools Appl.*, vol. 51, no. 2, pp. 675–696, Jan. 2011.

[20] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[21] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.

[22] K. Seshadrinathan and A. C. Bovik, "Motion-based perceptual quality assessment of video," in *Human Vision and Electronic Imaging, San Jose, USA*, ser. SPIE Proceedings, B. E. Rogowitz and T. N. Pappas, Eds. SPIE, p. 72400.

[23] M. Torres Vega, S. Zou, D. C. Mocanu, E. Tangdiongga, A. M. J. Koonen, and A. Liotta, "End-to-end performance evaluation in high-speed wireless networks," in *The Int. Conf. on Network and Service Management (CNSM), Rio de Janeiro, Brazil*, 2014.

[24] D. C. Mocanu, G. Santandrea, W. Cerroni, F. Callegati, and A. Liotta, "Network performance assessment with quality of experience benchmarks," in *Int. Conf. on Network and Service Management (CNSM), Rio de Janeiro, Brazil*, 2014.

[25] ITU, "Itu-t g.1050 : Network model for evaluating multimedia transmission performance over internet protocol," 2011.

[26] P. Paudyal, F. Battisti, and M. Carli, "Impact of video content and transmission impairments on quality of experience," *Multimedia Tools and Applications*, vol. 2016.

[27] P. Le Callet, S. Moeller, A. Perkis, and eds., "Qualinet white paper on definitions of Quality of Experience," in *COST Action IC 1003*, Laussanne, Switzerland, 2012.

[28] P. Paudyal, F. Battisti, and M. Carli, "A Study On The Effects Of Quality Of Service Parameters On Perceived Video Quality," in *Procs. of 5th European Workshop on Visual Information Processing, EUVIP 2014, Paris, France*, December 2014.

[29] M. G. Kendall, A. Stuart, and J. K. Ord, Eds., *Kendall's Advanced Theory of Statistics*. New York, NY, USA: Oxford University Press, Inc., 1987.

[30] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *Trans. Img. Proc.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[31] ——, "A subjective study to evaluate video quality assessment algorithms," in *Human Vision and Electronic Imaging, San Jose, USA*, ser. SPIE Proceedings, vol. 7527. SPIE, 2010, p. 75270.

[32] F. Suárez, A. García, J. Granda, D. García, and P. Nuño, "Assessing the QoE in Video Services Over Lossy Networks," *J. Netw. Syst. Manage.*, vol. 24, no. 1, pp. 116–139, Jan. 2016.

[33] ffmpeg, "A complete, cross-platform solution to record, convert and stream audio and video," Available at https://ffmpeg.org/.

[34] I. STANDARD, "3333.1.1-2015 - ieee standard for quality of experience (qoe) and visual-comfort assessments of three-dimensional (3d) contents based on psychophysical studies," 2015.

[35] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. Le Callet, J. Gutierrez, and N. Garcia, "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," 2012, pp. 109–114, cited By 34.

[36] I. Politis, L. Dounis, C. Tselios, A. Kordelas, T. Dagiuklas, and A. Papadakis, "A model of network related QoE for 3D video," in *2012 IEEE Globecom Workshops, Anaheim, USA*, Dec 2012, pp. 1335–1340.

[37] C. T. E. R. Hewage and M. G. Martini, "Quality of experience for 3D video streaming," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 101–107, 2013.

[38] A. Gotchev, D. Strohmeier, K. Mueller, G. B. Akar, and V. Petrov, "Source and channel coding recipes for mobile 3D television," in *Digital Signal Processing (DSP), 2011 17th International Conference on, Corfu, Greece*. IEEE, 2011, pp. 1–8.

[39] G. Piro, C. Ceglie, D. Striccoli, and P. Camarda, "3D Video transmissions over LTE: A performance evaluation," in *Eurocon 2013, Zagreb, Croatia*, July 2013, pp. 177–185.

[40] x265, "HEVC/H.265," Available at http://x265.org/hevc-h265/.

[41] M. Torres Vega, C. Perra, D. Mocanu, and A. Liotta, "Effect of Lossy Networks on Stereoscopic 3D-Video Streams," in *12th IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting (BMSB2017), Cagliari, Italy*, june 2017.