

Color-plus-Depth Level-of-Detail in 3D Tele-immersive Video: A Psychophysical Approach*

Wanmin Wu, Ahsan Arefin, Gregorij Kurillo[†], Pooja Agarwal, Klara Nahrstedt, Ruzena Bajcsy[†]

Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

[†]Department of EECS, University of California at Berkeley, Berkeley, CA, USA.

{wwu23, marefin2, pagarwl, klara}@illinois.edu, †{gregorij, bajcsy}@eecs.berkeley.edu

ABSTRACT

This paper presents a psychophysical study that measures the perceptual thresholds of a new factor called Color-plus-Depth Level-of-Detail peculiar to polygon-based 3D tele-immersive video. The results demonstrate the existence of Just Noticeable Degradation and Just Unacceptable Degradation thresholds on the factor. In light of the results, we describe the design and implementation of a real-time perception-based quality adaptor for 3D tele-immersive video. Our experimental results show that the adaptation scheme can reduce resource usage while considerably enhancing the overall perceived visual quality.

Categories and Subject Descriptors

H.1.2 [Information Systems]: Models and Principles—*human factors*; H.4.3 [Information Systems Applications]: Communications Applications—*Computer conferencing, teleconferencing, and videoconferencing*

General Terms

Design, Experiment, Human Factors, Measurement

Keywords

Tele-immersive Video, Perception, Psychophysics, Color-plus-Depth, Level-of-Detail, Adaptation

1. INTRODUCTION

The past few years have witnessed a surge of interest in telepresence video collaboration technologies. Several industrial systems have arisen [1][2], yet with a limited application domain as most are still concentrated solely on

*Area chair: Pål Halvorsen. This material is based upon work supported by the National Science Foundation under Grant No. CNS 0834480. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.
Copyright 2011 ACM 978-1-4503-0616-4/11/11 ...\$5.00.

desktop video-conferencing scenarios. Emerging is the 3D tele-immersion technology that expands the horizon by supporting full-body interaction of physical activities in virtual reality environments. Applications have been found in cyber-archeology, rehabilitation, collaborative dancing, and gaming (e.g., [28][29]).

Despite great potential, today's tele-immersive systems still face significant challenges due to the high interactivity requirement and resource demand. Great research efforts have been devoted to making the systems more efficient [5][21][28], with notable improvement, but the focus has been primarily on system-centric, algorithmic optimizations. The systems still struggle with heavy temporal and spatial overheads and complexities, limiting their broader deployment.

Since the ultimate goal of tele-immersion is to deliver a satisfying experience to end users, we believe that taking a more human-centric perspective would be beneficial. It is known that the Human Visual System (HVS) has perceptual limitations, so the research question is whether it is possible to exploit these limitations and reduce data load and/or rate without impairing much perceived quality. Similar questions have been studied in traditional video-conferencing systems for factors such as jitter [7], audio-visual sync [27], latency [11], and frame rate [3][16]. However, 3D tele-immersive video possesses unique characteristics whose perceptual impact is little understood.

Perhaps the most important trait that distinguishes 3D tele-immersion from the traditional video-conferencing is its color-plus-depth video format as the visual representations of users. Therefore, the density and accuracy of the texture and depth maps is a new and critical factor in tele-immersive video, which we combine and refer to as the *Color-Plus-Depth Level-of-Detail (CZLoD)*. In this work, we make the first attempt to psychophysically study this factor in *polygon-based* 3D tele-immersive video. We employ the method of limits from psychophysics [8] to examine two perceptual thresholds - Just Noticeable Degradation (JNDG) and Just Unacceptable Degradation (JUADG). We evaluate forty stimuli in four varied conditions with different contents and pixel resolution settings. The results indicate that the threshold levels are actually fairly generous (i.e., a fair amount of degradation can be suffered) and are related to both activity type and resolution. In general, fine motor activity exhibits lower threshold levels than gross motor activity, and lower resolution video exhibits lower threshold levels than higher resolution levels.

In light of the results, we design and implement a perception-

based real-time adaptation scheme for CZLoD in 3D tele-immersion. Implemented as a closed feedback loop, the adaptor monitors various interdependent Quality-of-Service (QoS) parameters to determine the appropriate degradation ratio for CZLoD. The actual degradation, nevertheless, is achieved by controlling a *detailing parameter*, whose mapping to the degradation ratio is unpredictable as it varies with environments and activities. Thus a learning algorithm is used to learn the quantitative model of the relationship between the detailing parameter and the CZLoD degradation ratios. We evaluate the adaptation scheme in a real-world 3D tele-immersive testbed, and the experimental results demonstrate that the proposed scheme can achieve considerable improvement in frame rate without impairing perceived detailing quality. We also record the generated tele-immersive video with and without adaptation respectively and conduct a crowdsourcing subjective study to compare their overall quality. The collected responses show that over 90% of users thought the video with adaptation was better than the unpaired video.

Our work is partly inspired by the level-of-detail concept in 3D graphics [18]. Nevertheless, we believe there are four main factors that distinguish our work: (1) we are the first to psychophysically measure perception thresholds of level-of-detail in 3D video. (2) Tele-immersion is a distributed, highly interactive pipeline, so almost all of our design choices are made carefully for the interactive timing bounds (Section 5). (3) Since tele-immersion is about real-time video where scene complexity is unpredictable, the relationship between the detailing parameter and the resulting CZLoD is unpredictable too; therefore, we learn a predictive model about their relations at run-time (Section 5.2.3). We also show that the predictor can produce very high accuracy in real time (Section 5.3). (4) We demonstrate how the perceptual thresholds obtained psychophysically can be applied to practice for real-time resource management.

In this work, we develop an intra-stream data adaptation scheme that reduces level-of-detail within each stream at the sending side without users being aware of it. This human-centric approach effectively alleviates the data load for computation-intensive operations, thus improving the temporal efficiency of the systems. Yet even with intra-stream data reduced, spatial (bandwidth) efficiency is still a problem due to the multi-stream/multi-site nature of tele-immersive collaboration. In our other work [33], we have actually addressed the spatial challenge with an inter-stream data adaptation scheme at the networking phase that reduces the number of streams with minimal disruption to visual quality.

We begin by reviewing related work (Section 2), and then discuss tele-immersive video in greater depth (Section 3). A psychophysical study of perceptual limitation and acceptability on the detailing factor is then described (Section 4). We present the construction of our perception-based quality adaptation scheme (Section 5). The paper concludes by discussing the limitations of the study and implications of our research (Section 6).

2. RELATED WORK

The ultimate goal of 3D tele-immersion is to enable people to interact across distance just as if they were co-located physically. This is achieved by fusing the 3D representations of geographically distributed users into a virtual reality en-

vironment in real time. The history of 3D tele-immersion can be traced back to about a decade ago, when researchers demonstrated the first networked tele-immersive application that could run at 2 to 3 frames per second (fps) [28]. Various efforts have been made to improve the efficiency of the systems in diverse components such as depth reconstruction [15][34], coordinated data transport protocol [21], rendering [28], as well as real-time 3D video compression [35]. Despite the notable improvement, tele-immersive systems are still far from becoming a commodity due to the high interactivity demand and heavy computational complexities. In this paper, we tackle the challenge from a different perspective by examining data redundancy in terms of psychophysical principles. We believe our approaches are orthogonal to the system-centric algorithmic improvements, and thus can be combined to provide greater performance benefits.

Psychophysics is not new to the multimedia community. The JPEG codecs [12], for example, compress images by eliminating high frequency details that are invisible to human eyes. Audio compression algorithms, such as MP3, exploit psychoacoustic principles to reduce information that is less audible to human ears [9]. Recently, psychophysics is also being applied to haptic feedback where the samples with imperceptible changes are removed from network transmission [26]. Perhaps the most relevant to our work is the recent psychophysical study conducted by De Silva *et al.* that considered the Just Noticeable Difference in Depth (JNDD) in 3D video [24]. However, the context is very different from this work in that the video content therein is for offline-generated 3D-TV. The real-time requirement of polygon-based tele-immersive video leads to the emergence of a new definition of CZLoD that is not applicable in 3D-TV video (this will become more apparent in Section 3). We also develop (to our knowledge) the first perception-based adaptation scheme for polygon-based 3D tele-immersion.

3. TELE-IMMERSIVE VIDEO

To avoid any confusion, we have to first point out that tele-immersive video is different from the commonly known stereoscopic video (as in 3D movies) which creates depth illusion with two offset imaging sequences for the two eyes of viewers respectively. Unlike such stereoscopic video, tele-immersive video refers to the *color-plus-depth* video, created and visualized *in real time*.

3.1 Real-time Acquisition

The acquisition pipeline in 3D tele-immersion mainly consists of three stages: capture, transmission, and visualization. Tele-immersive video is often captured by an array of synchronized cameras surrounding the physical environment. Unlike conventional multi-view video conferencing/lecturing systems, each camera here is a stereo unit, typically equipped with binocular or trinocular lenses, and connected to a host computer via IEEE 1394 (FireWire) interface. At interactive rates, the host computer grabs image frames synchronously from all lenses and produces *color-plus-depth* frames. A more detailed description of this 3D reconstruction process is deferred to Section 3.2. After the 3D reconstruction is completed with some post-processing (filtering, smoothing, regularity enforcement), the 3D frame is sent to the collaborating site(s) for visualization in a virtual-reality environment where all participants can seamlessly interact.

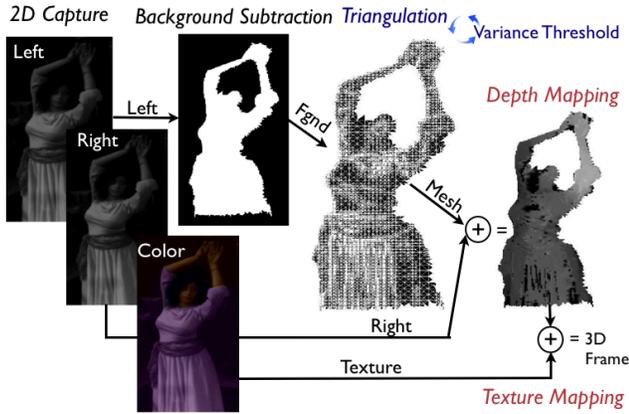


Figure 1: Hybrid depth and texture mapping process: expensive depth cross-correlation is only performed on vertices after meshing, and depths for the other pixels are linearly interpolated (adapted from [29]).

3.2 Color-Plus-Depth Level-of-Detail

In this section, we describe the generation of color-plus-depth frames in greater detail. As Figure 1 illustrates, after the raw frames are fetched from the stereo camera, they are first preprocessed (e.g., resizing, rectification). Then one of the images is used as the “reference frame” (e.g., the image from the left eye of the camera as shown in Figure 1), where background subtraction is performed. The next major step normally would be to reconstruct the 3D information of the frame for each foreground pixel. However, obtaining an accurate, dense (per pixel) depth map on commodity hardware turns out to be very time-consuming. In addition, transmitting the full-sized texture information would be quite costly in network bandwidth as well as in visualization latency, given the multi-site, multi-stream nature of tele-immersion. For these reasons, several tele-immersive systems use the polygonal modeling approach [15][20][30], to alleviate both temporal and spatial overheads.

In this approach, after background subtraction the reference frame is decomposed into polygons (e.g., triangles) based on texture homogeneity. This is done by recursively refining bisection until the variance within every polygon is less than a threshold TH_{var} . Importantly, afterwards the expensive depth-correlation operation is only performed on mesh vertices [29]. The depth calculation for the other pixels can thus be largely simplified by linear interpolation. Similar subdivision can be applied for textures as well. Since now only the coordinates and textures of vertices need to be transmitted (and those of the remaining pixels to be approximated at the receiving/rendering side), such region-based representations (and the accompanying hybrid depth and texture mapping algorithms) lead to a reduction of frame size as well as data manipulation time in all stages of the tele-immersion pipeline, making them favorable for resource-intensive tele-immersion applications. Due to its clear benefits compared to the conventional point cloud approach (with substantially less data to process and transmit), this polygon modeling approach is very likely to become more widely adopted in the future.

We can observe that the number of foreground vertices after meshing regulates the color-plus-depth granularity of

Notations	Descriptions
$JNDG$	Just Noticeable Degradation Ratio
$JUADG$	Just Unacceptable Degradation Ratio
p	Probability for determining psychophysical thresholds
f_i	2D reference frame with frame number i
F_i	Color-plus-depth frame reconstructed from f_i
$CZLoD(F_i)$	Number of foreground vertices in F_i
W	Size of running window for computing frame rate
TH_{var}	Variance threshold (detailing parameter)
TH_{fr}^h, TH_{fr}^l	Upper/lower limit of normal frame rate range
$DR^a(F_i)$	Actual degradation ratio of CZLoD in F_i
$DR^t(F_i)$	Target degradation ratio of CZLoD in F_i
$N_0(F_i)$	CZLoD of F_i when $TH_{var} = 0$
$N_v(F_i)$	CZLoD of F_i when $TH_{var} = v$
\mathcal{F}	Mapping function from DR^t to TH_{var}
$err(F_i)$	$= DR^a(F_i) - DR^t(F_i) $, error of degradation ratio
TH_{err}	Threshold of err to trigger variance learning
Δ_u, Δ_d	Decrease and increase sizes for DR^t adjustment

Table 1: Notations.

the mesh. It also determines the density/accuracy of tele-immersive video due to the disparate treatment of vertices and non-vertices in 3D reconstruction and texture mapping. Hence, we refer to this metric as the *Color-plus-Depth Level-of-Detail (CZLoD)* metric, which characterizes the spatial (including z-axial) and textural richness and accuracy of polygon-based tele-immersive video. Clearly, it is largely impacted by the setting of the *variance threshold* $TH_{var} (\in \mathbb{Z})$. The smaller the variance threshold is, the finer the meshing is, and the more dense/accurate the depth and texture maps will be. Therefore, the variance threshold TH_{var} is a *detailing parameter* for the CZLoD of tele-immersive video.

We are concerned about whether there are perceptual limits on the *degradation* of CZLoD for the purpose of data reduction. We thus mathematically formulate the metric of *degradation ratio* (DR). Suppose we denote the 2D reference frame as f_i (i is frame number), and the 3D frame generated from it as F_i . Assume $N_0(F_i) \in \mathbb{N}^0$ is the number of foreground vertices computed on f_i if TH_{var} were set to 0, and $N_v(F_i) \in \mathbb{N}^0$ is the number of foreground vertices computed on f_i if TH_{var} were set to v ($v \geq 0$), the degradation ratio of CZLoD on the frame F_i can then be expressed as

$$DR(F_i) = 1 - \frac{N_v(F_i)}{N_0(F_i)} \quad (1)$$

where $0 \leq DR(F_i) < 1$. For convenience, Table 7 lists all the notations used throughout the paper.

Obviously, the lower the degradation ratio, the more vertices the frame F_i will contain, resulting in higher definition and accuracy in both texture and depth, and leading to sharper and more stereoscopic images (depending on perceptual mechanisms).

3.3 Stimulus Generation Engine

While 2D video quality studies can utilize a pool of 2D video sequences offered by the Video Quality Expert Group (VQEG) [32], there are no standard test data for tele-immersive

video. Since 3D tele-immersion is essentially a live pipeline from cameras to renderers, a naïve way of obtaining test sequences would be to record different test sequences multiple times with different configurations of *treatment factors* (which refer to the sources of variation that are of particular interest to the experimenter [19]). However, this approach not only requires a large amount of effort but also suffers from uncontrollable varying conditions such as captured content and illuminance. Therefore, we propose a stimulus generation engine suitable for general tele-immersive video studies. To ensure the treatment factors are only varied within homogeneous blocks [19], we decouple the capturing from the 3D reconstruction so that different configurations can be applied during each phase, even on the exact image samples if desired.

Figure 2 depicts the three distinct stages of the engine. In Stage 1, a number of frames of the experiment’s activities are synchronously captured, preprocessed, and stored. Numerous parameters can be configured at this stage, including the desired pixel resolution, whether to use rectification or background subtraction, the number of frames to take, etc. To generate lower pixel resolution images, the raw frames can be downsampled. In Stage 2, the engine retrieves the specified 2D frames and repeatedly performs 3D reconstruction with varying parameters such as the variance threshold TH_{var} , whether to use trinocular or binocular stereo matching, etc. The host computers then send their reconstructed frames to a renderer that aggregates the frames and writes them to disk storage. In the final stage, the 3D frames are replayed as stimuli with possibly varying parameters such as frame rate. In short, the systematic decomposition allows automatic generation of stimuli with the flexibility of controlling desired treatment factors, while keeping blocking and nuisance factors (e.g., content) fixated [19].

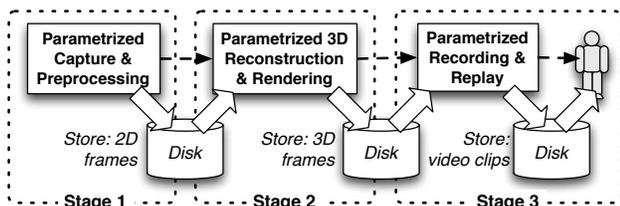


Figure 2: Stimulus generation engine: in Stage 1, 2D frames are taken, preprocessed, and stored; in Stage 2, 3D reconstruction is repeatedly applied with varied parameters; in Stage 3, the stored 3D frames are replayed with parameterization to generate the final stimuli.

4. PSYCHOPHYSICAL EXPERIMENT

The purpose of the psychophysical experiment is to measure two perceptual thresholds of CZLoD degradation: (a) Just Noticeable Degradation (JNDG), and (b) Just Unacceptable Degradation (JUADG). Identification of these thresholds can guide us to develop perception-based CZLoD adaptation mechanism for resource saving without impairing the perceived visual quality. We employed the Ascending Method of Limits [8] as the experimental methodology. It is one of the oldest and most widely used approaches in psychophysics for determining thresholds of sensations. The methodology, originally designed to measure singular intensity such

	Gross Motor	Fine Motor
Low Res. 320x240	Exercise-L	Lego-L
High Res. 640x480	Exercise-H	Lego-H

Table 2: Stimulus block codes

as light luminance and sound frequency, was slightly modified in order to measure degradation level by means of comparison. In our study, CZLoD conditions were presented in sequential pairs, one being an unimpaired reference, and one being the same video impaired. The magnitudes of impairment were presented in an ascending order.

4.1 Stimuli

With the experimental methodology in mind, we generated forty stimuli. Below we discuss their conditions, properties, and resource characteristics.

There are two factors that may impact the perception of CZLoD impairment: sequence content and pixel resolution of raw frames (called “blocking factors”). To explore their relationship with the treatment factor CZLoD, we created 2 (contents) \times 2 (pixel resolutions) groups (called “blocks”) of stimuli, each having a different configuration of blocking factors. For the first factor (content), we categorized the most frequent tele-immersive activities into two types and recorded a representative video for each type: (a) gross-motor activities such as Tai-Chi training, dancing, and physical rehabilitation that involve large body movement; and (b) fine motor activities such as telemedicine, cyberarcheology, object/tool instructions that involve finer body movement (e.g., on hands), and manipulation of objects. For the former type, we recorded a person (performer) doing an elbow exercise (commonly used in telepresence physical therapies), while for the latter type, we recorded the performer showing a small Lego house where details were not only more demanding for the object but also for the finger movement of the performer. For the second blocking factor (pixel resolution), we chose two levels that had been mostly used in tele-immersive systems [28][31]: (a) high - 640×480 , and (b) low - 320×240 . The four stimulus blocks were coded as in Table 2.

The stimulus generation engine (Figure 2) was then employed to generate 10 levels of stimuli in each block. In the first stage, the Dragonfly2 stereo camera (Point Grey Inc.) was employed for recording the 2D images for each of the four block conditions. After 2D frames were acquired, 3D frames were generated for each block by repeatedly running binocular stereo matching on the same stored images with varied TH_{var} for obtaining different CZLoD degradation settings. Finally, the 3D color-plus-depth frames were rendered with a fixed frame rate 10 fps (chosen according to [16]) for the subjects. Following the ITU-R BT.500 standard [13], each test sequence was 10 seconds long, so about 100 frames were included.

The 10 stimuli for each block were coded as S_0, S_1, \dots, S_9 with increasing levels of degradation ratio in CZLoD (Figure 3(a)), with an approximate step size of 10% (degradation ratio). Figure 4 shows the snapshots of the lowest and highest stimulus conditions for each block. For each stimulus, the degradation ratio was calculated by averaging across all frames (relative standard deviation measured to be 2.08% - 2.83% for all stimuli). Therein, S_0 was the unimpaired reference stimulus ($TH_{var} = 0$). The TH_{var} values for other

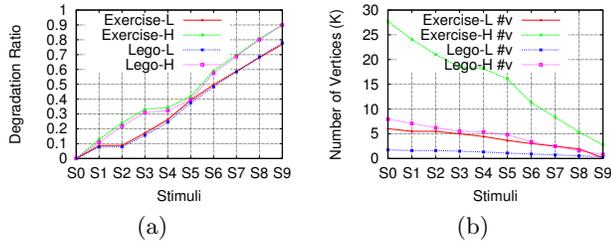


Figure 3: Stimulus statistics: (a) the degradation ratios gradually increase with the stimulus levels, (b) the actual numbers of vertices ($\#v$).

stimuli were manually chosen to approximately achieve the expected degradation ratio (it was impossible to be exact). Two sets of TH_{var} values were used, one for the lower pixel resolution blocks (Exercise-L/Lego-L), and the other for the higher resolution blocks (Exercise-H/Lego-H). Figure 3(b) presents the actual number of vertices after the meshing process.

4.2 Participants and Procedures

We followed the ITU standard in conducting the experiment [13]. Sixteen adult participants were recruited from University of Illinois at Urbana-Champaign, primarily graduate students and staff in the Department of Computer Science¹. All had normal or corrected vision. Four participants were Indian, three were American, two were Chinese, two were German, three were Bangladeshi, one was Mexican, and one was South African. The sample consisted of 6 women (37.5%) and 10 men (62.5%). Regarding the level of experience with tele-immersive video, the sample consisted of 5 experts (31.25%) and 11 novices (68.75%).

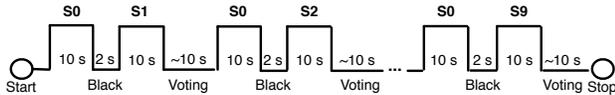


Figure 5: Experimental procedure: sequential {unimpaired, impaired} pairs of stimuli were shown, with ascending degradation ratios. Each stimulus was 10-sec long, the interval showing a black screen was 2-sec long within pair, and the voting period between pairs was about 10-sec long [13].

The sequence of blocks presented was: Exercise-L, Exercise-H, Lego-L, and Lego-H. Figure 5 shows the experimental process (adapted from [13]) within each block. Pairs of stimuli were presented automatically using a script with the ascending levels of degradation. For each pair, the first video was the unimpaired reference video, shown to mitigate memory effect [22], and the second video was the impaired one. In between the pair, there was a 2-second interval with black screen [13]. The voting period after each pair was about 10 seconds long, when the observer was asked if he/she could tell any difference between the two clips, and whether he/she thought any video had unacceptable quality. The subjects

¹The exact age distribution is unknown because some subjects expressed unwillingness to disclose age.

were told that they could take a break at any time during the experiment.

4.3 Apparatus

The experiment was conducted in the MONET (Multi-media Operating and Networking) laboratory at the University of Illinois at Urbana-Champaign. Participants were asked to be seated in front of a LCD monitor during the experiment with a standard viewing distance [13]. The detailed specification of the monitor used is listed in Table 3. 3D displays were available but not used mainly for usability concerns. Despite their rapid growth, today’s state-of-the-art 3D displays are not yet ready to be deployed for tele-immersive activities. For example, typical stereoscopic displays require observers to wear goggles to perceive the depth effect, which is intrusive and thus unsuitable for physical activities often conducted in tele-immersive environments. The new autostereoscopic displays eliminate the need for wearing glasses; however our previous experience with them indicates that the technology was far from mature as they caused considerable discomfort for viewers. Lambooij et al. gave a general review of the visual discomfort caused by stereoscopic and autostereoscopic displays [17]. Therefore, in this experiment we resorted to using conventional displays for visualization. However, it is still worth noting that 3D displays are only designed to hypothetically *improve* depth perception [14], not to *enable* it. In fact, depth perception is achieved by a variety of visual cues (such as shading, texture gradient, linear perspective, motion parallax, occlusion, etc.) that are still relevant in tele-immersive video regardless of the type of display used [10]. We chose to trade the possible increase of depth perception for the visual comfort of users, which was believed to be more important.

Table 3: Detailed specification of the monitor used.

LCD Monitor Model	Acer X222W
Dimensions (WxDxH)	51.4 cm x 20.4 cm x 41.8 cm
Resolution	1680 x 1050 / 60 Hz
Dot Pitch	0.282 mm
Response Time	5 ms
Brightness	300 cd/m ²

4.4 Evaluation Results

In psychophysics, perceptual thresholds are defined to be the stimulus intensities (in our case, CZLoD degradation ratios) that can be detected/accepted some p portion of the time, with $p = 50\%$ often used [8]. Figure 6(a) shows the measured JNDG and JUADG thresholds in four blocking conditions using probability 50% (equivalent to taking medians in the data). There are several observations:

Existence of perceptual thresholds: There do exist perceptual thresholds on the magnitude of CZLoD degradation that viewers can detect and accept. The average JNDG across all conditions is 61.5%, suggesting that degradation below 61.5% is not noticeable to average users. This implies that we can *transparently* reduce a considerable amount of resource usage by degrading CZLoD without actually impairing the perceived quality. The existence of JUADG (average 81.25%) indicates the degradation should be bounded by this upper limit otherwise it might make the overall quality unacceptable.

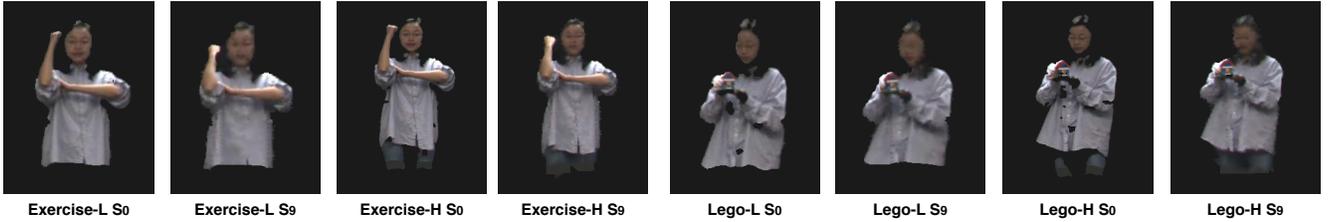


Figure 4: Stimulus snapshots: please refer to Figure 3(a) for the stimulus codes and their degradation ratios.

Impact of pixel resolution: JNDGs in both content blocks (Exercise and Lego) are lower for the two 320x240 conditions (Exercise-L/Lego-L) than for the corresponding 640x480 conditions (Exercise-H/Lego-H), indicating that it might be easier for subjects to notice degradation with lower pixel resolution than with higher resolution. This may be possibly because lower resolution already loses much more detail than the higher resolution (in our case, four times); thus any further degradation would become more noticeable. Likewise, JUADGs of Exercise-L/Lego-L are lower than those of Exercise-H/Lego-H.

Impact of content: Within the same pixel resolution condition, the thresholds for the 320x240 resolution (Exercise-L/Lego-L) vary with content, while those for the 640x480 resolution (Exercise-H/Lego-H) do not. Exercise-L has a higher JNDG than Lego-L, meaning that it is harder to notice the degradation in the exercise scene that contains only elbow movement than in the Lego video involving finer granularity object and finger movement. This is partly due to the fact that the arm in elbow exercises requires much less details than the Lego object. Since viewer attention normally focuses on the arms in motion, any changes in other parts of the video tend to go unnoticed (a phenomenon often referred to as “change blindness” in visual perception [25]). Similarly, the tolerance for degradation (JUADG) is higher in Exercise-L than in the Lego-L.

Figure 6(b) presents the cumulative probability distribution of the response data. We have the following observations:

Relationship between noticeability and acceptability thresholds: For every condition, the A curve (for JUADG) is always on the right side of the N curve (for JNDG). This indicates that as degradation increased, at some point subjects started to notice the distortion yet felt it was still acceptable, but after some further degradation, they started to feel the video quality was unacceptable.

Noticeability and acceptability offset: The offsets between the N and A curves in each condition are similar - mostly close to 10%-20%. Considering the step size in our stimuli was about 10%, this means about 0%-20% more degradation than the noticeable region would cause the quality to be perceived as unacceptable. The reason why we consider step size is as follows: Assume for the first six stimulus levels presented (S_0 versus S_1, S_2, \dots, S_6 , Figure 5), the subjects did not notice a difference between the unimpaired and impaired video. Then assume at S_7 they suddenly started to notice the degradation (say, of ratio 70%, meaning JNDG = 70%). It is unclear any ratio between that of S_6 and S_7 is detectable or not due to the discrete stimulus levels we used. Hence it is safer to take the step size into consideration here.

Impact of pixel resolution and content: The observations we

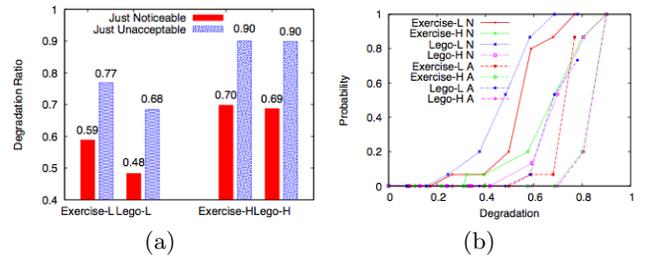


Figure 6: (a) perceptual thresholds on four conditions taken by medians, and (b) psychometric curves of JNDG and JUADG (in the legends, “N” stands for JNDG, and “A” stands for JUADG).

have from Figure 6(a) about the impact of the two blocking factors (when p is 50%) are also generally true for other p values (y -axis). For example, Lego-L has lower JNDG and JUADG than Exercise-L. Lego-H and Exercise-H mostly have the same responses. The lower resolution blocks (Exercise-L and Lego-L) generally have lower thresholds than the higher resolution blocks (Exercise-H and Lego-H).

5. ADAPTATION SCHEME

5.1 Overview

QoS parameters are characterized by spatial (intra-frame) and temporal (inter-frame) requirements. For tele-immersive video, the spatial parameter refers to the spatial (including z -axial) resolution, and the temporal parameter corresponds to the frame rate. Naïve tele-immersive applications are often implemented without considering the temporal-spatial balance. The results from our psychophysical study suggest that CZLoD provides tele-immersion developers a powerful tool to control the detail complexity of the video, and in turn control the frame processing time or the frame rate. A major implication of our findings is that transparent degradation on spatial resolution (CZLoD) is possible to achieve “free” saving on resources without users being aware of it, i.e., by degrading CZLoD to a level where the distortion is just unnoticeable. In addition, when frame rate drops to a level that it hurts overall usability, the degradation on CZLoD can further increase (yet within acceptable ranges) to reduce data load and thereby elevate frame rate. Furthermore, past research has implied a curvilinear relationship between spatial quality and frame rate, e.g., improvements in frame rate become less noticeable above approximately 10 frames per second [16]. Therefore, when frame rates are found to be higher than necessary, the CZLoD degradation

ratio can be lessened (if possible) to recover or promote the detailing quality thereby reducing frame rate. In short, we can manipulate the CZLoD degradation ratio to achieve a balance between the temporal quality (frame rate) and the spatial quality (CZLoD).

Based on these principles, we propose a novel, human-centric, real-time adaptation scheme (at the sender side) for tele-immersive video. We design the adaptor as a closed feedback loop [4] for the control of detailing quality in tele-immersive video. Figure 7 illustrates the schematic process of the adaptation. It has three major components: QoS Monitor, Decision Engine, and Variance Calculator. QoS Monitor is responsible for collecting and analyzing time series of QoS parameters (e.g., frame processing time, frame size, reconstruction time), and extracting meaningful information online to notify Decision Engine for triggering adaptation. Decision Engine computes an appropriate target CZLoD degradation ratio for the 3D reconstruction process. Since the degradation ratio is actually controlled by manipulating the variance threshold (Section 3.2), a Variance Calculator component is used to compute the correct variance threshold given a target degradation ratio from Decision Engine. Yet a challenge is that the mapping from a desired CZLoD degradation ratio to a variance threshold is unpredictable due to its dependency on scenes (e.g., clothing texture, skin colors, presence of objects, lighting illuminance). Therefore, Variance Calculator dynamically learns a quantitative model between the CZLoD degradation ratio and the appropriate variance threshold. Based on the model, it computes the proper variance threshold given a target degradation ratio, and feeds it into the 3D reconstruction pipeline for video quality adaptation.

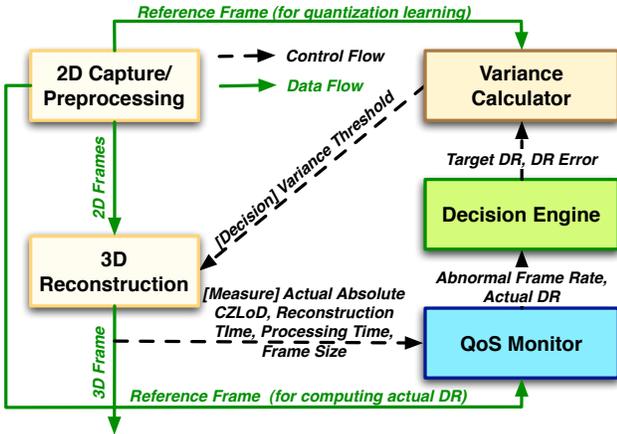


Figure 7: Adaptation scheme overview: refer to Section 5.1. DR stands for Degradation Ratio.

5.2 Design and Implementation

5.2.1 QoS Monitor

Various CZLoD-related QoS parameters are inter-dependent in tele-immersion. Figure 8 depicts the most relevant parameters and their dependencies identified using the Granger-causality graphs [23] over profile data. QoS Monitor continuously collects time-series meta-data of these parameters

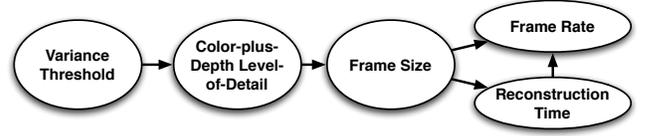


Figure 8: Dependency graph of parameters.

for each frame, and performs online analysis and profiling. It then provides “feedback” to Decision Engine. The feedback includes two types: (a) *frame rate events* (excessively high or low) for triggering increase or decrease of the degradation ratio, and (b) *actual degradation ratio* $DR^a(F_i)$ of every frame F_i .

Since providing real-time feedback in the control loop is key, a simple yet efficient range checking approach is used for evaluating frame rate. Essentially, if the frame rate drops below a lower-limit threshold (TH_{fr}^l), Decision Engine is notified to increase degradation ratio (for lower CZLoD quality); if the frame rate increases beyond an upper-limit threshold (TH_{fr}^h), Decision Engine is notified to decrease degradation ratio (for higher CZLoD quality). The thresholds should be set according to the perceptual characteristics of the frame rate [16]. Compared to the single threshold method where $TH_{fr}^l = TH_{fr}^h$, range thresholding is important for avoiding the flickering effect that can occur when a parameter constantly switches between low and high levels as it hovers near the threshold. For the same reason, the frame rate is not computed on a per frame basis, but averaged over a running window of size W (frames).

Apart from the frame rate reports, QoS Monitor also evaluates the actual degradation ratio of each frame F_i , $DR^a(F_i)$, and reports it to Decision Engine for taking corrective measure. $DR^a(F_i)$ needs to be measured because the sequence complexity and resource condition are constantly changing, meaning it is possible that a target degradation ratio would not be achieved exactly as desired. It is worth pointing out that the precise computation of $DR^a(F_i)$ requires the original 2D frame f_i be reconstructed by setting $TH_{var} = 0$ (refer to Equation 1). To facilitate the computation, the 2D capture/preprocessing component of the live 3D tele-immersion pipeline periodically sends a reference frame f_r to QoS Monitor, on which it applies 3D reconstruction with $TH_{var} = 0$ and computes the reference CZLoD expressed as $N_0(F_r)$ (Section 3). Since this is a relatively expensive operation, it is only periodically performed. We believe this is reasonable considering that performer motion cannot change dramatically within a short period of time; i.e., $N_0(F_r)$ would be very close to $N_0(F_i)$ due to their temporal proximity². Using the latest available $N_0(F_r)$ to approximate $N_0(F_i)$, QoS Monitor can then compute $DR^a(F_i)$ (Equation 1) and report it to Decision Engine.

5.2.2 Decision Engine

The foundation of the adaptation logic in Decision Engine is based on the perceptual thresholds (JNDG and JUADG) on the color-plus-depth spatial resolution of the video. The thresholds decompose the CZLoD quality of tele-immersive

²In this approximated computation, an out-of-range result (i.e., $DR^a(F_i) > 1$) is theoretically possible, but very unlikely. If it does occur, $DR^a(F_i)$ is corrected to 1.

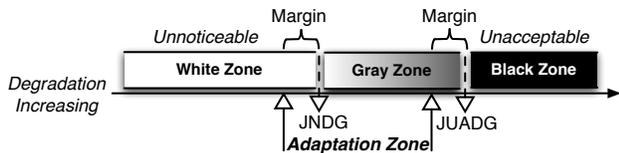


Figure 9: The Just Noticeable (JNDG) and Just Unacceptable (JUADG) thresholds decompose the range of CZLoD degradation to three zones.

video into three zones: *white zone*, where distortion is minimally noticeable; *gray zone*, where the distortion gradually becomes noticeable yet still acceptable; and *black zone*, where the degradation is unacceptable. The basic idea of Decision Engine is to dynamically adjust the target degradation ratio primarily in the gray zone, with some margins (Figure 9). The margins are introduced to account for the step size in our psychophysical experiment (as discussed in Section 4.4) as well as environmental and user dynamics. Hence, if we denote the margin size as B_n and B_a ($0 \leq B_n, B_a \leq 1$) for noticeability and acceptability thresholds respectively, the adaptation zone can be defined as $[JNDG - B_n, JUADG - B_a]$ in terms of degradation ratio.

As mentioned above, Decision Engine receives two types of information from QoS Monitor: (a) abnormal frame rate, and (b) DR^a of every frame. Upon receiving alarms of abnormal frame rate, Decision Engine computes an appropriate target degradation ratio. For this purpose, a linear control mechanism is used. Basically, an abnormally low frame rate ($FR_i < TH_{fr}^l$) means the need for lower CZLoD quality (or higher degradation ratio). Thus the engine computes the target degradation ratio as $DR^t(F_i) = DR^a(F_{i-1}) + \Delta_d$ where $DR^t(F_i)$ denotes the target degradation ratio (greater DR means more degradation), $DR^a(F_{i-1})$ denotes the actual degradation ratio of the last frame F_{i-1} (reported by QoS Monitor), and Δ_d denotes the adjustment size for increasing DR . This ratio is then used for all frames until the next time adaptation is triggered. Similarly, an unnecessarily high frame rate ($FR_i > TH_{fr}^h$) triggers the engine to produce a desired degradation ratio as $DR^t(F_i) = DR^a(F_{i-1}) - \Delta_u$ where Δ_u is the adjustment size for decreasing DR .

The settings of Δ_d and Δ_u can follow various protocols, e.g., AIMD (Addictive Increase Multiplicative Decrease), or proportional to frame rate deviation from normal mean. Although much more complicated or aggressive changes may result in faster reaction time, they may also incur more overhead or abrupt changes in the detailing resolution of the video. We find that simple constant small sizes are sufficiently effective in responding to frame rate anomalies while maintaining a gradual and graceful change that is less noticeable. Further, simple models provide superior performance benefits that are critical in such real-time environments. We utilize the measured thresholds from Section 4 to guide the adaptation zone (Figure 9). When the frame rate is excessively low (i.e., to increase DR), if $DR^t(F_i)$ reaches the upper limit of the adaptation zone (close to the black zone in Figure 9), Δ_d is set to 0; i.e., no further degradation is allowed otherwise the quality would become unacceptable. Likewise, when the frame rate is excessively high (i.e., to de-

crease DR), if $DR^t(F_i)$ reaches the lower limit of the adaptation zone (close to white zone), Δ_u is set to 0; i.e., further improvement on the detailing quality would not be noticeable anyway thus is unnecessary. Besides the calculation of the target degradation ratio, Decision Engine also computes the adaptation error between the actual and target degradation which will be used for Variance Calculator (as explained below).

5.2.3 Variance Calculator

Given the target CZLoD degradation ratio DR^t , Variance Calculator is responsible for determining the proper value for the detailing parameter TH_{var} in the 3D reconstruction. However, the mapping \mathcal{F} from DR^t to TH_{var} is nontrivial because it highly depends on external conditions such as scene complexities. Therefore, we dynamically learn a quantitative model in order to predict the correct TH_{var} value for a desired DR^t . The learning process is performed only when Decision Engine finds that the adaptation error $err = |DR^a(F_i) - DR^t(F_i)|$ is larger than some threshold $err > TH_{err}$, meaning that significant changes in scenes might have happened that make the previous model less applicable. To learn a new model, Variance Calculator repeatedly applies 3D reconstruction on the frame f_i with exponentially increasing TH_{var} values, and the resultant $DR^a(F_i)$ values are logged. This process runs in parallel with the actual 3D tele-immersion pipeline and is thus unobtrusive. The TH_{var} values and their resultant $DR^a(F_i)$ values are then fed into a least-square regression module to develop an exponential model as follows [6]:

$$\mathcal{F} : TH_{var} = e^{a \cdot DR^t + b} \quad (2)$$

where e is the Euler number, and a and b are constants. The choice of the above exponential model comes from our extensive curve-fitting experiments with TI video, where their relationships are all a form of Equation 2. With this simple model we are able to achieve a high accuracy (median residual of 0.022%) with as few as 10 training points (refer to Section 5.3). With the model available, Variance Calculator is then able to set a proper variance threshold after 2D preprocessing and before 3D reconstruction for a desired degradation ratio.

5.3 Performance Evaluation

We evaluated the adaptation scheme in a real-world 3D tele-immersive system. The Bumblebee2 stereo camera (Point Grey Inc.) was used. It was connected to a host computer (Intel Xeon quad-core CPU 2.8GHz and 2GB RAM) via an IEEE 1394b card. The pixel resolution used was 320x240. Two professional lamps (Brightline SeriesONE) were employed to produce soft and diffused lighting condition. The rendering machine had an Intel Xeon CPU 2.3GHz, 2GB RAM, and an NVIDIA GeForce 9800 graphics card. The scene was an experimenter performing arm exercises in front of the camera. We compared two conditions — with and without adaptation, using the same experimental setup. Both objective and subjective measurements were collected. The technical metrics such as frame rate, frame size, target and actual degradation ratios were logged, and the rendered video was recorded on the renderer for subjective evaluation. Settings for the algorithmic parameters were: $W = 5, TH_{fr}^l = 8, TH_{fr}^h = 12, B_n =$

-3	-2	-1	0	+1	+2	+3
Much Worse	Worse	Slightly Worse	The Same	Slightly Better	Better	Much Better

Table 4: Rating scale used to compare videos with and without adaptation.

$B_a = 10\%$, $JNDG = 59\%$, $JUADG = 77\%$, $\Delta_u = \Delta_d = 5\%$ (Table 1).

Figure 10(a) shows the frame rate comparison. In this case, the adaptation scheme achieved an average of 27% improvement on frame rate from about 8 fps to 10 fps. According to [16], any improvement below a frame rate of 10 fps is considerably noticeable to users. It is also worth noting that even with the adaptation mechanisms added, the frame rate improves and real-time performance is guaranteed, meaning that the overhead of our algorithms is minimal. We also compared the frame rate with some CPU stress. For this, a process was run together with 3D tele-immersion that took at peak 16% of the CPU load. This simulates conditions where the CPU is less powerful or higher pixel resolution is configured. As Figure 10(b) shows, the frame rates without adaptation dropped to about 6-7 fps with several sudden dips to 3 fps. On the other hand, the frame rates achieved with adaptation (with the same conditions) remained relatively stable around 9 fps (with average improvement being 39.6%). Figure 10(c) shows the actual degradation ratios used with the projected target ratios. The prediction accuracy was high, with a median residual of 0.022%. The average degraded ratio was 22.7%, with a standard deviation of 0.063%. Considering that the JNDG is around 60% (Figure 6), there was still room for much more reduction if the frame rates were below desired thresholds.

We also conducted a user study to compare the visual quality of the recorded video. The crowdsourcing methodology was used due to the simplicity of the experiment. Following the ITU standard for double stimuli video comparison study [13], we made a short video with the following structure (in sequential order): (a) five seconds of text illustrating the purpose of the study, (b) two seconds of text indicating “Video 1” to be shown, (c) ten seconds of Video 1, (d) two seconds of text indicating “Video 2” to be shown, and (e) ten seconds of Video 2, and (f) ten seconds of text asking the rating question: “Compared to Video 1’s quality, Video 2’s quality is: [the scale shown in Table 4]?” [13]. The video was uploaded to Youtube and was advertised to a mailing list (containing graduate students, staff, and professors in Department of Computer Science). The ranking data were collected anonymously through an online Google Doc Form. A total of 81 responses were collected. Three of them were discarded because respondents notified the experimenter that they submitted by mistake. Figure 10(d) shows the collected ratings. Among the 78 responses, 96.2% of the users thought the video with adaptation turned on was better quality than the video with adaptation turned off, and 3.8% thought they were the same. 12.8% (of total) gave a “(+1) Slightly Better” ranking, 51.3% gave a “(+2) Better” ranking, and 32.1% gave a “(+3) Much Better” ranking. Clearly, our adaptation scheme not only saves system resources (i.e., CPU load), but also improves subjective video quality.

6. CONCLUDING REMARKS

This paper identifies a new critical quality factor called Color-plus-Depth Level-of-Detail (CZLoD) in 3D tele-immersive video. A psychophysical study of the perceptual thresholds of CZLoD is performed and presence of two perceptual thresholds - Just Noticeable Degradation (JNDG) and Just Unacceptable Degradation (JUADG) is demonstrated. Taking CZLoD as a guiding parameter, we design an online human-centric QoS adaptation scheme to dynamically adapt the video quality. Our experiments show that the adaptation scheme considerably reduces the resource demands while enhancing the perceived visual quality.

Future Work. Currently, N_0 is only periodically evaluated in our adaptation scheme (Section 5.2.1); we will investigate real-time statistical approaches for predicting N_0 . Also, we employ a simple constant adaptation model for adjusting degradation ratio due to its efficiency (Section 5.2.2), but complex models are worth investigating. As hardware advances, it is possible that the increasing number of cores could run more complex models with demanded time bounds. Further, we measure the perception thresholds on 2D display, and we will perform the study on 3D displays as they mature. Finally, as we have stated previously, we only consider the temporal challenge in this work and address the spatial challenge in our previous work [33]. In future work, we will study their seamless integration.

Discussion. Subjective video quality research suffers from the limitation that the results might vary with the sequence content, and this study is no exception. While we attempt to be representative in choosing the tele-immersive activities for the psychophysical study, we do not intend to draw any general conclusion about the specific values of JNDG and JUADG in all tele-immersive applications. Rather, the main contribution of our study is the identification of the existence of perceptual thresholds on a unique factor that has (to our best knowledge) never been explored in real-time color-plus-depth video. The measurements of the thresholds provide practical guidelines on their estimation in the field. We also demonstrate that by applying these thresholds to practice, we can adapt the detailing quality and achieve considerable resource saving as well as enhancement on the perceived video quality.

7. REFERENCES

- [1] Cisco TelePresence, <http://www.cisco.com/en/us/products/ps7060/index.html>.
- [2] Polycom Telepresence, <http://www.polycom.com/telepresence>.
- [3] R. Apteker, J. Fisher, V. Kisimov, and H. Neishlos. Video acceptability and frame rate. *IEEE Multimedia*, 2(3):32–40, 1995.
- [4] K. J. Åström and R. M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2008.
- [5] H. H. Baker, N. Bhatti, D. Tanguay, I. Sobel, D. Gelb, M. E. Goss, W. B. Culbertson, and T. Malzbender. Understanding performance in Coliseum, an immersive videoconferencing system. *ACM TOMCCAP*, 1(2):190–210, 2005.
- [6] A. Björnc. *Numerical Methods for Least Squares Problems*. SIAM, 1996.
- [7] Y.-C. Chang, T. Carney, S. A. Klein, D. G.

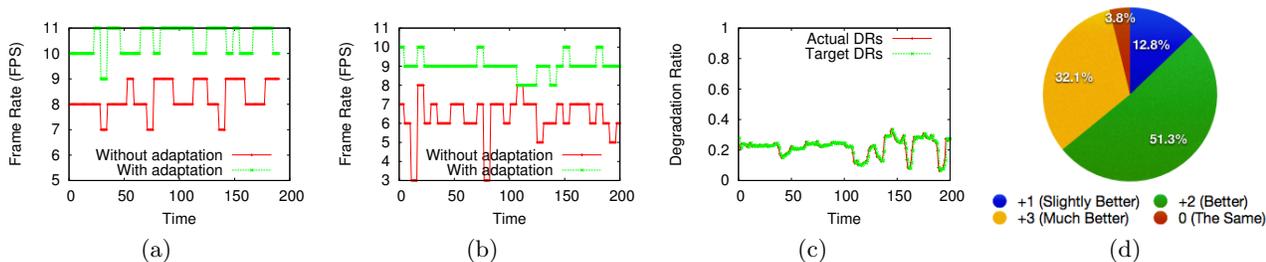


Figure 10: Adaptation performance: (a) the frame rates with and without adaptation where the CPU has no additional load other than tele-immersion, (b) the same comparison with a 16% CPU stress, (c) the degradation ratios (DRs) are well below the just noticeable threshold, with high prediction rate regarding variance threshold, and (d) the subjective evaluation results comparing the quality of the adapted video against the unimpaired video.

Messerschmitt, and A. Zakhor. Effects of temporal jitter on video quality: Assessment using psychophysical methods. In *Proc. of SPIE HVIP'98*, 1998.

[8] G. Gescheider. *Psychophysics: The Fundamentals*. Psychology Press, 3rd edition, 1997.

[9] S. Hacker. *MP3: The Definitive Guide*. O'Reilly, 1st edition, 2000.

[10] I. Howard. *Seeing in Depth. Volume 2: Depth Perception*. I Porteous, 2002.

[11] S. Iai, T. Kurita, and N. Kitawaki. Quality requirements for multimedia communication services and terminals - interaction of speech and video delays. In *Proc. of GLOBECOM '93*, 1993.

[12] ITU. Information technology - digital compression and coding of continuous-tone still images - requirements and guidelines. In *Rec. T.81*, 1992.

[13] ITU. Methodology for the subjective assessment of the quality of television pictures. In *Rec. BT.500*, 2009.

[14] Z. X. Jin, Y. J. Zhang, X. Wang, and T. Plocher. Evaluating the usability of an auto-stereoscopic display. *Proc. of HCI'07*, pages 605–614, 2007.

[15] T. Kanade, P. Rander, and P. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE MultiMedia*, 4(1):43–54, 1997.

[16] J. K. Kies. *Empirical methods for evaluating video-mediated collaborative work*. PhD thesis, Virginia Polytechnic Institute and State University, 1997.

[17] M. Lambooi and W. IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: A review. In *JIST*, volume 53, 2009.

[18] D. Luebke, M. Reddy, J. Cohen, A. Varshney, B. Watson, and R. Huebner. *Level of Detail for 3D Graphics*. Morgan Kaufmann, 2002.

[19] D. C. Montgomery. *Design and Analysis of Experiments*. Wiley, 7 edition, 2008.

[20] P. J. Narayanan, P. W. Rander, and T. Kanade. Constructing virtual worlds using dense stereo. In *Proc. of ICCV'98*, pages 3–10, 1998.

[21] D. E. Ott and K. Mayer-Patel. A mechanism for TCP-friendly transport-level protocol coordination. In *Proc. of USENIX'02*, pages 147–159, 2002.

[22] M. Pinson and S. Wolf. Comparing subjective video quality testing methodologies. In *Proc. of SPIE VCIP'03*, pages 573–582, 2003.

[23] M. E. R. Dahlhaus. *Causality and graphical models for time series*. University Press, Oxford, 2003.

[24] D. D. Silva, W. Fernando, G. Nur, E. Ekmekcioglu, and S. Worrall. 3D video assessment with just noticeable difference in depth evaluation. In *Proc. of ICIP'10*, pages 4013–4016, 2010.

[25] D. J. Simons and R. A. Rensink. Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9(1):16–20, 2005.

[26] E. Steinbach, S. Hirche, J. Kammerl, I. Vittorias, and R. Chaudhari. Haptic data compression and communication. *IEEE SPM*, 2011.

[27] R. Steinmetz. Human perception of jitter and media synchronization. *IEEE JSAC*, 14(1):61–72, 1996.

[28] H. Towles, W. chao Chen, R. Yang, S. uok Kum, H. F. N. Kelshikar, J. Mulligan, K. Daniilidis, H. Fuchs, C. C. Hill, N. K. J. Mulligan, L. Holden, B. Zeleznik, A. Sadagic, and J. Lanier. 3D tele-collaboration over Internet2. In *ITP*, 2002.

[29] R. Vasudevan, G. Kurillo, E. Lobaton, T. Bernardin, O. Kreylos, R. Bajcsy, and K. Nahrstedt. High quality visualization for geographically distributed 3d teleimmersive applications. In *IEEE Transactions on Multimedia*, volume 13, pages 573–584, 2011.

[30] R. Vasudevan, G. Kurillo, E. Lobaton, T. Bernardin, O. Kreylos, R. Bajcsy, and K. Nahrstedt. High quality visualization for geographically distributed 3d teleimmersive applications. *IEEE ToM*, 2012.

[31] R. Vasudevan, E. Lobaton, G. Kurillo, R. Bajcsy, T. Bernardin, B. Hamann, and K. Nahrstedt. A methodology for remote virtual interaction in teleimmersive environments. In *Proc. of ACM MMSys'10*, pages 281–292, 2010.

[32] VQEG. <http://www.its.bldrdoc.gov/vqeg/>.

[33] W. Wu, Z. Yang, I. Gupta, and K. Nahrstedt. Towards multi-site collaboration in 3d tele-immersive environments. In *IEEE ICDCS'08*, 2008.

[34] S. Würmlin, E. Lamboray, and M. Gross. 3D video fragments: Dynamic point samples for real-time free-viewpoint video. Technical report, Institute of Scientific Computing, 2003.

[35] Z. Yang, Y. Cui, Z. Anwar, R. Bocchino, N. Kiyancilar, K. Nahrstedt, R. H. Campbell, and W. Yurcik. Real-time 3D video compression for tele-immersive environments. In *Proc. of MMCN*, 2006.