ABSTRACT
This paper is concerned with the subjective assessment of compressed 3D video. The generated test 3D video sequences are full HD 3D video sequences consisting of the original uncompressed 3D video sequences and their compressed versions which were generated at different bit-rates using recent standard video codecs including H.264/AVC and its Scalable Video Coding (SVC) and Multi-View Video Coding (MVC) extensions. Subjective tests are conducted, focusing on the spatio-temporal visual quality, depth perception and naturalness. Experimental results show that MVC exhibits the highest performance on average for 3D video both in terms of R-D curves and subjective tests.

Index Terms—3D, Video, Quality, Depth, Naturalness, Subjective

1. INTRODUCTION
3D video is being increasingly used in entertainment, gaming, interactive interfaces, and consumer electronics. Existing 2D video coding methods [1, 2] could be used to encode 3D video in a compatible format, and more efficient 3D video compression methods have been developed [3].

The 3D video quality can be affected by the used encoding format and employed compression method and bit-rate. Therefore, there is a need to assess the perceived quality of compressed 3D video sequences in order to understand the effect of various compression methods on the perceived quality of experience (QoE).

Currently, there is a lack of 3D video databases that include 3D video test content and associated subjective scores. In this work, subjective tests are conducted to assess the QoE of full HD 3D video sequences that are compressed at different bit-rates using state-of-the-art standard codecs, namely the ITU-T H.264/AVC and its Scalable Video Coding (SVC) and Multi-View Video Coding (MVC) extensions. The 3D video QoE is evaluated in terms of the spatio-temporal visual quality, which we refer to hereinafter as visual quality for short, depth perception, and naturalness. The obtained subjective scores are then used to analyze the perceived visual performance of the state-of-the-art standard video codecs.

The remainder of this paper is organized as follows. Section 2 presents the employed 3D video compression methods. Section 3 describes the conducted 3D video subjective tests, along with the perceived performance of the employed video codecs. A conclusion is presented in Section 4.

2. 3D VIDEO COMPRESSION METHODS
The stereo video is composed of a left view and a right view. The simplest way to encode the stereo video is to encode the left view and right view separately, which is referred to as simulcast. This simulcast approach could achieve good quality and could use a conventional 2D HD encoder. However, it is not quite efficient for transmission as double equipment and bandwidth are needed. Furthermore, with simulcast, it is difficult to maintain L/R sync.

Broadcasting companies are currently considering different alternatives in terms of compression methods and encoding formats for achieving a full HD 3D video service, including using the Frame Sequential (FS) or Side-by-Side (SBS) encoding format and current standard video codecs, namely H.264/AVC and its SVC and MVC extensions.

Table 1 shows the International Telecommunication Union (ITU) categorization on Plano-stereo 1st generation 3D TV [1]. Level 1, also known as conventional high-definition display compatible, uses optimized color anaglyph. Despite having a huge cost benefit, this kind of 3D imaging does not allow high quality reproduction of stereo color video.

Level 2, also known as conventional HD frame compatible (HDFC), is obtained using a spatial multiplexing that combines the left and right video sequences into one video sequence in which the left and right views are combined in a single frame and which can be encoded using conventional 2D video codecs. This allows handling the 3D video as a 2D HD video using typical channels and equipment. Two popular ways to perform spatial multiplexing include Side-by-Side (SBS) and Top-and-Bottom (TB). However, this conventional HD frame compatible format reduces the horizontal or vertical resolution by half. This led to the development of the HD frame compatible format (Level 3), which combines the frame compatible format with the H.264 SVC spatial scalability. The H.264 SVC spatial scalability is used to encode additional information in order to achieve full HD resolution for both views. This provides backward compatibility and also enables the functionality to access full HD for both views. However, this method does not consider the inter-view correlation.
Table 1: ITU 3D TV Categorization on Plano-stereo 1st generation 3D TV ([1]).

<table>
<thead>
<tr>
<th>Compatibility level</th>
<th>Matrix of signal format for 3D TV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Conventional High Definition display Compatible (CDC) Optimized colour anaglyph</td>
</tr>
<tr>
<td>Level 2</td>
<td>Conventional HD Frame Compatible (CFC) Frame compatible Left and Right in same High Definition frame</td>
</tr>
<tr>
<td>Level 3</td>
<td>High Definition Frame Compatible Frame compatible plus H.264/MPEG4-VC resolution (for example, H.264 SVC)</td>
</tr>
<tr>
<td>Level 4</td>
<td>Conventional High Definition Service Compatible (CSC) 2D HD + MVC (ie, H.264 MVC) Left and Right formed by matrixing</td>
</tr>
</tbody>
</table>

Since there is a relatively strong correlation between the left and right views, MVC was developed to explore the inter-view dependencies. One view is predicted from the other.

Based on the above, three different compression methods are utilized in this work for comparison. The first one uses JM MVC to encode the left and right views both in full HD FS format. The second one uses JM H.264/AVC to encode the 3D video in a HDFC format (left view and right view are placed SBS in a 2HD frame). The third method uses SVC with a JSVM base layer (BL) used for encoding the 3D video base layer in a Conventional Frame Compatible (CFC) format (down-sampled HD SBS), followed by a JSVM spatial enhancement layer (EL) for encoding the 3D video in a HDFC format (2* HD SBS).

The rate control algorithm in JM and JSVM is not advanced and sophisticated, while JMVC does not have an implemented rate control mechanism. So, QP is used to control the bit-rate of the compressed stream. Firstly, we use the same QP setting for each combination of test sequence and compression codec (QP=24, 28, 32, 36, 40). Then a fitting procedure between QP and bit-rate is applied to each combination of test sequence and compression codec as follows:

\[ QP = p_1 \cdot \left( \frac{(\text{Bit-Rate}/p_2) - p_3}{p_4} \right) + p_5 \]

(1)

In order to cover the visual quality and transmission bandwidth range, the anchor bit-rates are chosen as 1.3, 3, 5, 7, 9, 11, and 15 Mbps. Different QP are set to reach the considered bit-rate for each combination of test sequence and compression codec. The proper QP value is calculated through target bit-rate and fitted parameters using (1).

For illustration, Fig. 1 shows the obtained R-D curves for 300 frames of the 1920 * 1080p Nict4K3D digest sequence, at 30 frames per second (fps), from the NICD 3D video database, and compressed using JM AVC, JSVM and JM MVC at the bit-rates of 1.3, 5, 7, 9, 11, and 15 Mbps, respectively. For every codec, cascaded QP is enabled, GOP size = 8, GOP structure is IBBBBBBBP, search range = 128. The following codec versions and parameters are used for encoding the video sequences. For JM AVC, JM version 17.2 is used, ProfileIDC is set to 100 as high profile, LevelIDC is set to 51. The input YUV video frames are each of size 3840*1080. The employed JSVM version is version 9.19.11, SearchFuncFullPel and SearchFuncSubPel are set to SAD. ProfileIDC for BL is set to 100 as high profile, ProfileIDC for EL is set to 86 as scalable high profile. The BL frame size is 2*960*1080, and the EL frame size is 3840*1080. The employed JM MVC version is 17.2, ProfileIDC is set to 100 as high profile, LevelIDC is set to 51. The input files are two 1920*1080 sequences corresponding to the left view and right view. The NumberOfViews parameter is set to 2, the QP of view1 is larger than QP of view2 by 3. Both MVCEnableInterViewFlag and MVCInterViewReorder are enabled.

From Fig. 1, it can be seen that, for the relatively high-motion sequence, JM MVC with the FS format (MVC+FS) achieves the best R-D performance followed by JM AVC with the SBS format (AVC+SBS), and then by JSVM with the SBS format (SVM+SBS).
3. SUBJECTIVE QUALITY ASSESSMENT

The reliable assessment of video quality plays an important role in improving the end user's quality of experience (QoE), especially in 3D video services. It is thus important to explore how 3D video compression affects the user's experience in terms of visual quality, depth perception and naturalness. The 3D video quality can be used in evaluating the performance of video transport systems, including compression method, compression parameters, etc. Controlling and monitoring the QoS parameters of the individual system components by appropriately selecting the compression methods and compression parameters is important for efficiently achieving high overall system performance and QoE.

Subjective testing is an important component for evaluating the user’s QoE. In this testing, a group of human subjects are asked to judge the quality of the video sequence under predefined system conditions. The scores given by observers are averaged to produce the Mean Opinion Score (MOS).

Recent efforts on 3D subjective testing include the work of the ITU-R Working Party 6C on the requirements for subjective testing of 3DTV [4, 8]. More work is needed towards finding a standardized way to measure the perceived 3D video quality.

Some recommended guidelines for 3D video subjective testing can be summarized as follows. The test sequences should cover different content such as indoor and outdoor, low motion and high motion, low texture and high texture, and varied edge content. The spatio-temporal characteristics of the video sequences should be representative of the service under study. Some 2D video characteristics (e.g., motion, blockiness, sharpness and blur) are also applicable to 3D video subjective testing. In addition, stereoscopic 3D video exploits the human binocular visual system to recreate the depth of objects based on the offset in relative positions in corresponding stereo views.

Several perceptual characteristics affect the 3D video QoE including the visual quality, depth quality, naturalness, and visual comfort. The visual quality refers here to the perceived spatio-temporal visual quality of the video, which is a main component for both 2D and 3D video.

Depth quality refers to the sensation of depth. While features in monocular cues, such as linear perspective and blur can provide some sensation of depth, stereoscopic 3D images/videos contain extra depth information, which may lead to an enhanced sensation of depth.

Naturalness refers to the degree of the truthful representation of reality for perceived 3D video. In [5], it is found that the judgments of naturalness can be split into 75% based on the perceived 2D video quality and 25% based on the perceived depth.

Visual (dis)comfort refers to the degree of comfort when viewing 3D video. Improperly captured, artifacts due to compression, transmission errors, improperly displayed stereoscopic images could be a source of discomfort [6].

---

### Table 2 Sequences used for the subjective test.

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Codecs</th>
<th>Bit Rates (Mbps)</th>
<th>No. of Seq</th>
<th>Total No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>MVC+SVC+H</td>
<td>1, 3, 5, 9, 15</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>MVC+SVC+H</td>
<td>1, 3, 5, 9, 11, 15</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>D</td>
<td>MVC+SVC+H</td>
<td>1, 5, 15</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>MVC+SVC+H</td>
<td>1, 5, 15</td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>

The recommended viewing conditions generally match those used for 2D [7]. During the subjective testing, it is recommended [8] to use a viewing distance of 3H for a 1920×1080 video, where H is the frame height. The suggested viewing distance is 18 to 24 inches for the best NVIDIA 3D Vision experience on a 120Hz, 22-30 inches, LCD [9].

Observers should be screened for visual acuity, color, and stereoscopic vision. The latter could be assessed using clinical tests, such as Randot, Titmus, or Frisby stereo tests. The session duration might be as long as 20-40 minutes with breaks. Sequence duration is usually chosen as 8-10 seconds.

For the subjective testing methodologies, many of the standard methods in Rec. BT. 500 [7] (e.g., Single-Stimulus (SS), Double Stimulus Continuous Quality Scale (DSCQS), Stimulus-Comparison (SC), Single Stimulus Continuous Quality Evaluation (SSCQE)) could be used. For example, DSCQS was adopted in [10] and [11], while the Subjective Assessment Methodology for Video Quality (SAMVIQ) [12] was used in [13]. SAMVIQ can be defined as a multi-stimulus continuous quality scale method using explicit and hidden references. Each sequence can be played and assessed as many times as the observer wants.

In this paper, the subjective test content consisted of 55 10-second 1920×1080p 3D video sequences (Table 2), including compressed sequences using the aforementioned codecs and bit-rates, and their corresponding uncompressed versions. In addition, 2D versions (uncompressed, and compressed using JM AVC) were included in the test content.

Ten subjects participated in the subjective testing. All subjects were screened for visual acuity (20/20), depth perception (Randot), and color-blindness. The subjects viewed each video sequence using active shutter glasses and a 23-inch LCD monitor (DELL Alienware2310) with a 120Hz refresh rate, at a distance of 24 inches. The room illumination was 500 lux.

---

![Fig.2 Presentation timeline of designed subjective test.](image-url)
SS is used as the subjective test methodology as shown in Fig. 2. Subjects were asked to score the visual quality, depth perception, and naturalness using an ACR five-grade scale (Excellent, Good, Fair, Poor and Bad), as illustrated in Fig. 3.

Fig. 4 shows the relationship between bit-rate and the mean opinion score for visual quality (MOS-VQ) for JM MVC+FS, JM AVC+SBS and JSVM+SBS for a high motion sequence and a low motion sequence. It can be seen that JSVM+SBS achieves the lowest performance at the lower bit-rates (corresponding to IP video).

5. CONCLUSION
This paper is concerned with the assessment of the QoE for compressed full HD 3D video. First, R-D curves are obtained for full HD 3D video sequences with various spatial and temporal content using state-of-the-art video codecs (JM AVC, JSVM and JM MVC). Subjective tests are conducted in order to determine perceptual 3D characteristics in terms of spatio-temporal visual quality, depth perception and naturalness. Both R-D performance results and subjective tests show that MVC has the best overall performance. The subjective scores of SVC-coded 3D videos are close to those of MVC/AVC-coded 3D videos for high bit-rates, while SVC has the advantage of backward-compatibility and scalability. However, SVC achieves a significantly lower performance at lower bit-rates (below 10Mbps). Future research includes the development of objective 3D video quality metrics that highly correlate with subjective scores and, thus, reflect the perceived 3D video quality.
4. REFERENCES