4 The transmission aspects of 3D video

Your goals for this "The transmission aspects of 3D video" chapter are to learn about:

- 3D video broadcasting.
- 3D video transmission over IP networks and 3D video streaming.
- 3D video error resilient and concealment techniques.
- Challenges for 3D video over unreliable networks.

With the availability of well defined scene representations, efficient coding approaches and affordable 3D displays, a need for efficient 3D video transmission technology has become acute. Unlike 3D movies, which can be traced back to 1903, 3D broadcast and streaming/communication applications require rate-adaptive coding approaches, transport streams for delivery and signalling (e.g. MPEG-2 Transport Stream (TS)), control protocols and error resilient tools to render good quality 3D video content to the end-users [60]. The practical demonstrations (e.g. 3D video streaming) and prototype 3D video transmission systems have being studied in the past [61-63]. However, the technology is not matured up to the level of 2D video transmission technologies as yet. Moreover, most of the 3D delivery techniques will be built on top of the existing infrastructure and tools for conventional video applications. This section presents the background related to potential 3D transmission technologies and supportive technologies (e.g. error recovery) necessary for efficient transmission of 3D video content over broadcast and communication links.

3D-TV broadcasting has been an interesting application scenario from the early ages of analogue TV [64] [65]. However, these services were not continued for a long time due to the low quality pictures visible with different viewing aids (e.g. polarized glasses) and requirements for specific display systems [65]. With the introduction of digital transmission technologies, the possibilities of broadcasting 3D-TV pictures are further studied and necessary technologies (e.g. MPEG-2 Multi View Profile) are developed. For example, the integration of stereoscopic TV with HDTV is studied in [66] and [67] using separate transmission of left and right views and side-by-side arrangement (frame packing format) of left and right image sequences respectively. These approaches utilize existing infrastructures and technologies for HDTV transmission system and thus not so efficient due to the limitations (e.g. resolution, bandwidth) imposed for 3D video. For instance, frame compatible formats yield low resolution for left and right images (both left and right images are horizontally down-sampled). Due to the flexibility of rendering interactive stereoscopic video, colour plus depth representation has been tested over broadcasting channels [68] [24]. The approach in [68] utilizes the "private" or "user data" of the MPEG-2 Transport Stream to send the depth information [69], whereas the multiplexing of colour plus depth within the MPEG-2 Transport Stream is studied in [24]. Moreover, the latter approach is now standardised as an amendment ISO/IEC 13818-1 (MPEG-2 Transport Stream) [70].

The availability of voice over IP (e.g. VoIP), TV pictures over IP (e.g. IPTV) has influenced 3D video services to select the IP packet network as one of the main transport mediums. The wide spread usage of IP to deliver video services over wired/wireless channels will enable 3D video over large application space. Moreover, the flexible integration of IP transmission aspects with the other layers of the protocol stack allows more space for adapting different 3D scene representations and optimum coding methodologies. The 3D streaming services are classified into four main categories in [71] namely;

- Server unicasting to a single client
- Server multicasting to several clients
- Peer-to-peer (P2P) unicasting
- P2P multicasting.

The emerging 3D video streaming applications would prefer RTP/DCCP/IP protocol [72] over conventional RTP/UDP/IP protocol due to effective video rate adaptation for streaming applications by DCCP (Datagram Congestion Control Protocol) to match the TCP-Friendly Rate Control (TFRC) [73]. The rate adaptation strategies for stereo and multi-view video offer more flexibility than the conventional video due to the availability of more than one view during transmission. For example, the rate for the stereoscopic video application can be adapted by choosing different resolutions for the right image sequence together with full-resolution left image sequence according to the binocular suppression theorem[52]. Several studies can be found for open loop and closed loop rate adaptation schemes for 3D video over UDP and DCCP protocols [74-76]. For example, the client-driven multi-view video streaming described in [75] reduces the bandwidth need through selecting only a small number of views for transmission based on the user's head position.

Interactive 3D video streaming will enable seamless, more involving and adaptable delivery of 3D content to end users. However, 3D video streaming over band-limited and unreliable communication channels can introduce artifacts on the transmitted 3D content. The effect could be much more significant compared to conventional 2D video streaming. For instance, the nature of 3D video source format (e.g. colour plus depth images vs. left and right views) and the way our Human Visual System (HVS) perceives channel introduced artifacts in 3D video is different from 2D video; as an example, colour plus depth map 3D video presentation may have to utilize impaired depth map information at the receiver-side to render novel views.

Video streaming over the Internet has become one of the most popular applications and Internet 3D video streaming is expected to become more popular in the future, also thanks to the recently standardized wireless systems, including WIMAX, 3GPP LTE / LTE advanced, the latest 802.11 standards, and advanced short range wireless communication systems, enabling the transmission of high bandwidth multimedia data. For such applications the target of the system design should be the maximization of the final quality perceived by the user, or Quality of Experience (QoE), rather than only of the performance of the network in terms of "classical" quality of service (QoS) parameters such as throughput and delay. 3D video services, and in particular those delivered through wireless and mobile channels, face a number of challenges due to the need to handle a large amount of data and to the possible limitations due to the characteristics of the transmission channel and of the device. This can result in perceivable impairments originated in the different steps of the communication system, from content production to display techniques.

The overall enjoyment or annoyance of 3D video streaming applications or services is influenced by several factors such as human factors (e.g., demographic and socio economic background), system factors (e.g., content and network related influences) and contextual factors (e.g., duration, time of the day and frequency of use). The overall experience can be analysed and measured by QoE related parameters which quantify the user's overall satisfaction about a service [110][111]. Quality of Service (QoS) related measurements only measure performance aspects of a physical system, with main focus on telecommunications services. Measuring QoS parameters is straightforward since objective, explicit technological methods can be used, whereas measuring and understanding QoE requires a multi-disciplinary and multi-technological approach. The added dimension of depth in 3D viewing influences several perceptual attributes such as overall image quality, depth perception, naturalness, presence, visual comfort, etc. For instance, an increased binocular disparity enhances the depth perception of viewers, although in extreme cases this can lead to eye fatigue as well. Therefore, the overall enjoyment of the 3D application could be hindered by the eye strain experienced by the end user. The influence of these attributes on the overall experience of 3D video streaming users is yet to be investigated.

The effect of transmission over band-limited and unreliable communication channels (such as wireless channels) can be much worse for 3D video than for 2D video, due to the presence in the first case of two channels (i.e., stereoscopic 3D video) that can be impaired in a different way; as a consequence the 3D reconstruction in the human visual system may be affected. Some networks introduce factors directly related to temporal domain de-synchronization issues. For instance delay in one view could lead to temporal de-synchronization and this can lead to reduced comfort in 3D viewing. The methods employed to mitigate these artifacts (e.g., error concealment) need to be carefully designed to suit 3D video applications. The simple application of 2D image/video methods would not work effectively in this case, as discussed in [112] for different error concealment algorithms for 3D video transmission errors. In [112] it is observed that is some cases switching back to the 2D video mode is preferred to applying 2D error concealment methods separately for left and right views to recover missing image information during transmission. There could be added implications introduced by these artifacts into our HVS. Therefore artifacts caused as a result of 3D video streaming can be clearly appreciated only by understanding how our HVS perceives different 3D video artifacts. Frame freezing mechanisms employed to tackle missing frames caused by transmission errors or delay could lead to temporal de-synchronization where one eye sees delayed content compared to the other eye. There are two implications associated to the case where one view is affected by transmission impairments:

- Binocular suppression
- Binocular rivalry

Our HVS is still capable to align and fuse stereoscopic content if one view is affected by artifacts due to compression, transmission, and rendering. Binocular suppression theory suggests that in these situations the overall perception is usually driven by the quality of the best view (i.e., left or right view), at least if the quality of the worst view is above a threshold value. However this capability is limited and studies show that additional cognitive load is necessary to fuse these views [113]. Increased cognitive load leads to visual fatigue and eye strain and prevents users from watching 3D content for a long time. This directly affects user perception and QoE. If one of the views is extremely altered by the transmission system, the HVS will not be able to fuse the affected views, and this causes binocular rivalry. This has detrimental effects on the final QoE perceived by the end user. Recent studies on 3D video transmission [112] have found that binocular rivalry is causing the overall perception to be affected and this effect prevails over the effect of binocular suppression. To avoid the detrimental effect of binocular rivalry, the transmission system could be designed appropriately taking this issue into account. For instance, the transmission system parameters can be updated "on the fly" to obtain 3D views with minimum distortions, according to the feedback on the measure of 3D video quality at the receiver-side. In case of low quality due to different errors in the two views, if the received quality of one of the views is significantly low, the transmission system could be informed to allocate more resources to the worse view or to increase the error protection level for that 3D video channel to mitigate the quality loss in the future. This increases the opportunity to fuse the 3D video content more effectively and improve the final QoE of users.

Video over IP often suffers from packet losses. The congestion and noise/interference/fading are the main causes for packet losses in wired and wireless links respectively. Therefore, effective error correction approaches (e.g. Automatic Repeat-reQuest (ARQ), Forward Error Correction (FEC)), joint source channel coding techniques and error concealment techniques are necessary to send 3D services over IP networks more effectively. The performance of 3D video transmission over wireless channels which is considered as bandwidth limited and error prone, can be further improved with Joint Source Channel Coding (JSCC) approach which is an effective method to overcome such challenges [77]. The study carried out in [78] proposes a JSCC scheme for colour plus depth stereoscopic video. The advanced channel coding approaches for 3D video are addressed in several studies [79]. For example, in [80], stereoscopic video streaming using FEC techniques are investigated. The left and right image frames are classified into layers based on their contribution towards final reconstructed quality. Then the layers are transmitted based on the principal of Unequal Error Protection (UEP) using different error correction codes. Furthermore, Multiple Description Coding (MDC) approaches are proposed for stereoscopic video which allows sending video with acceptable quality bounds at bad channel conditions [81]. However, the use of error correction codes with 3D video is somewhat unjustifiable due to the high demand for bandwidth by 3D video content itself. Therefore, this research explores different perspectives to protect 3D video over networks without sending additional data. For example, transmission power can be allocated unequally for 3D video components depending on their contribution towards perceptual quality. In a way the adapted approaches exploit cross-layer design using the perceptual aspects of 3D video.

Error concealment is necessary to perform at the decoder in order to reduce the temporal error propagation caused by unpreventable packet losses. The conventional error concealment tools for 2D video can be adapted in recovering the errors of corrupted 3D video [82][83]. Error concealment algorithms have been proposed for left and right view based stereoscopic video utilizing the additional data from the corresponding image sequence [84] [85]. Bilen et al propose two methods for full frame loss concealment in left and right based stereoscopic video using overlapped block motion and disparity compensation [86]. Moreover, in [87] an approach to recover the entire right frame in stereoscopic video transmission is described based on the relativity of prediction modes for right frames. However, the above mentioned studies are aimed at providing error resilience for left and right view based stereoscopic video. Therefore, error concealment of colour plus depth stereoscopic video is considered in this book. Three error concealment methods are proposed based on the correlation of scene characteristics (e.g. motion correlation) and shape concealment approaches.