


3 Stereoscopic 3D video compression

Your goals for this “Introduction” chapter are to learn about:

- 2D video encoding.
- Scalable video encoding.
- Stereoscopic video encoding.
- Performance evaluation of different encoding approaches for 3D video.

This chapter describes the state of the art video coding approaches for 3D video. An introduction to the 2D video coding algorithms is provided. Moreover, scalable video coding approaches which can be utilized in scaling 2D video applications into immersive video are discussed. Then the potential coding approaches for 3D video in general and more specifically for stereoscopic video are discussed.

.....Alcatel-Lucent 

www.alcatel-lucent.com/careers

What if
you could
build your
future and
create the
future?

One generation's transformation is the next's status quo. In the near future, people may soon think it's strange that devices ever had to be "plugged in." To obtain that status, there needs to be "The Shift".



3.1 2D Video Coding

The primary aim of video coding is the removal of spatial and temporal redundancies present in raw images captured from a video camera. Video coding allows video to be used in communication applications with reduced storage and bitrate requirements. The block-based transform coding and subband-based decomposition of images are commonly utilized as the basic coding principles. Video coding has been standardized (H.261 in 1990, MPEG-1 Video in 1993, MPEG-2 Video in 1994, H.263 in 1997, MPEG-4 Visual or part 2 in 1998, H.264/AVC in 2004, HEVC in 2013), in order to facilitate the interoperability among different products and applications. The technology advances result in higher compression efficiency, different application support (video telephony-H.261, consumer video on CD-MPEG-1, broadcast of Standard Definition: SD/High Definition: HD TV- MPEG-2 and 4K/8K TV: HEVC) and network compliance (switched networks such as PSTN- H.263/MPEG-4 or ISDN- H.261 and Internet or mobile networks H.263/H.264/MPEG-4). Most of the video coding standards are based on hybrid video coding which employs block matching (i.e. Block Matching Algorithm: BMA) motion compensation and the Discrete Cosine Transform (DCT). The reasons for adopting hybrid video coding approach are that:

- A significant proportion of the motion trajectories found in natural video can be approximately described with a rigid translational motion model.
- Fewer bits are required to describe simple translational motion.
- Implementation is relatively straightforward and amenable to hardware solutions.

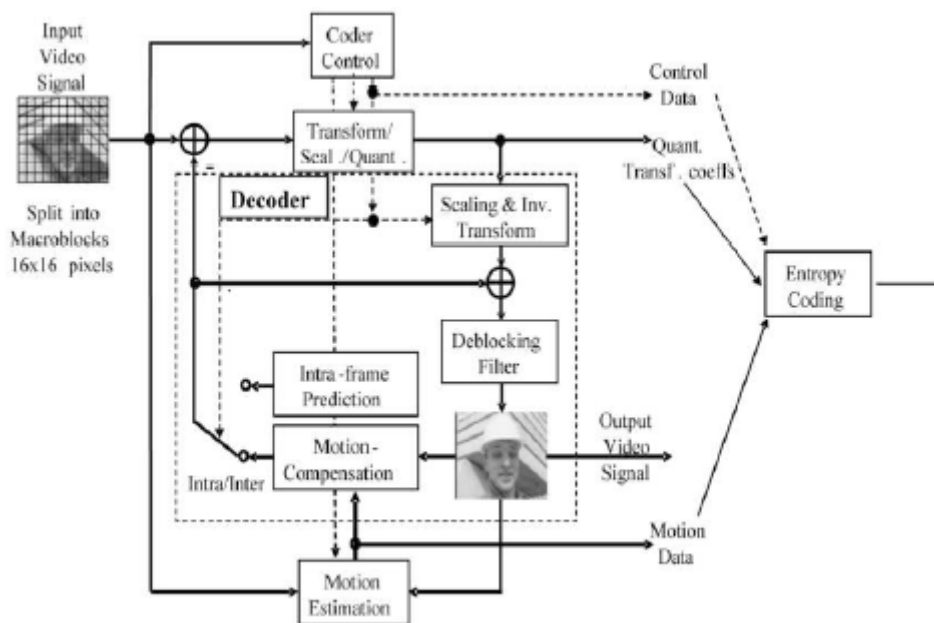


Figure 3.1: Basic structure of a hybrid coder [33]

H.265/High Efficiency Video Coding (HEVC) is the latest video coding technology standardized by the ISO/IEC Moving Picture Experts Group and the ITU-T Video Coding Experts Group [114]. This is primarily based on principles of previous H.264/AVC video coding standard [32]. However, HEVC provides much improved compression efficiency compared to H.264/AVC with added functionality [32]. Figure 3.1 shows the basic structure of a H.264/AVC coder. Similar to the most of the hybrid-video coders, this structure eliminates temporal and spatial redundancies through motion compensation and DCT based transform coding approaches respectively. The high compression efficiency and the network friendliness for interactive and non-interactive video applications are the main achievements in this latest standard [33][34]. The some of the coding features which assist H.264/AVC to gain superior video quality are variable block-size motion compensation with small block sizes, quarter-sample-accurate motion compensation, multiple reference picture motion compensation and in-the-loop de-blocking filter. H.264/AVC consists of two conceptual layers called Video Coding Layer (VCL) and Network Adaptation Layer (NAL). NAL renders a network adaptive bit-stream using the coded bit-stream available at the VCL interface (see Figure 3.2). This close integration of two layers allows H.264/AVC to be used in low bitrate video communication applications across heterogeneous networks.

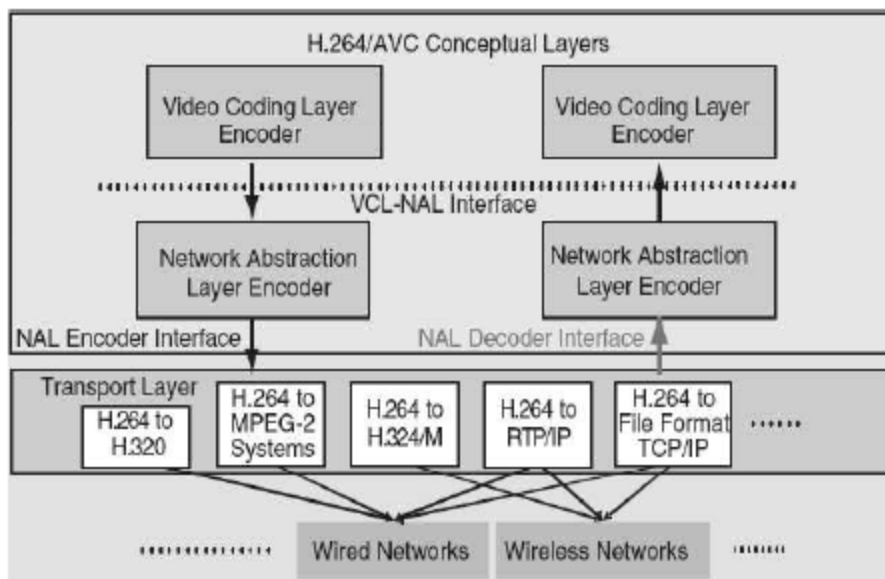


Figure 3.2: H.264/AVC in a transport environment [35]

In addition to the new features used for high compression gain, it consists of several error resilience and concealment features in order to provide more robust and error free video over communication channels. For example it supports slice structure (flexible slice sizes, redundant slices, Flexible Macroblock Ordering-FMO and Arbitrary Slice Ordering ASO), data partitioning, Parameter set structure, NAL unit syntax structure and SP/SI synchronization pictures [33] to be used in error prone environments. The potential tools can be employed in wireless video communication applications and H.264/AVC coded video over best-effort IP networks as described in [35] and [36] respectively.

3.2 Scalable Video Coding

Modern video transmission and storage systems are typically characterised by a wide range of access network technologies and end-user terminals. Varying numbers of users, each with their own time varying data throughput requirements, adaptively share network resources resulting in dynamic connection qualities. Users possess a variety of devices with different capabilities, ranging from cell phones with small screens and restricted processing power, to high-end PCs with high-definition displays. Examples of applications include virtual collaboration system scenarios, as shown in Figure 3.3, where a large, high powered terminal acts as the main control/commanding point and serve a group of co-located users. The large terminal may be the headquarters of the organization and consists of communication terminals, shared desk spaces, displays and various user interaction devices to collaborate with remotely located partners. The remotely located users with a small, fixed terminal will act as the local contact and provide the local information. Mobile units (distribution, surveying, marketing, patrolling, etc.) of the organization may use mobile terminals, such as mobile phones and PDAs, to collaborate with the headquarters.

Maastricht University *Leading in Learning!*

Join the best at the Maastricht University School of Business and Economics!

Top master's programmes

- 33rd place Financial Times worldwide ranking: MSc International Business
- 1st place: MSc International Business
- 1st place: MSc Financial Economics
- 2nd place: MSc Management of Learning
- 2nd place: MSc Economics
- 2nd place: MSc Econometrics and Operations Research
- 2nd place: MSc Global Supply Chain Management and Change

Sources: Keuzegids Master ranking 2013; Elsevier 'Beste Studies' ranking 2012; Financial Times Global Masters in Management ranking 2012

Visit us and find out why we are the best!
Master's Open Day: 22 February 2014

Maastricht University is the best specialist university in the Netherlands (Elsevier)

www.mastersopenday.nl



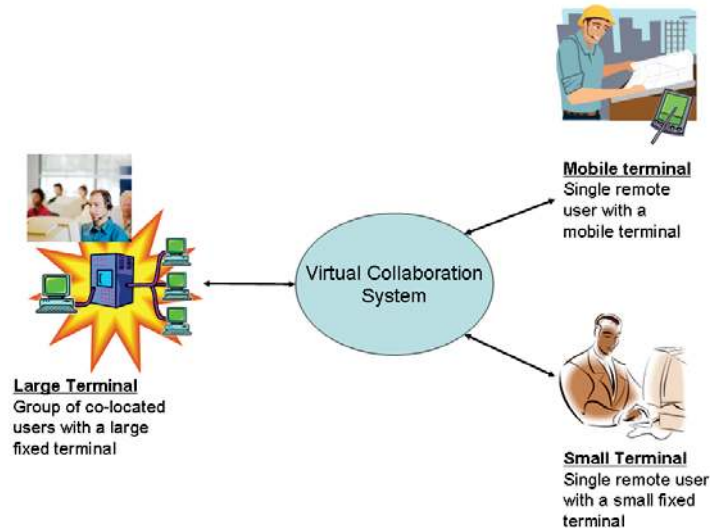


Figure 3.3: Virtual collaboration system diagram

In order to cope with the heterogeneity of networks/terminals and diverse user preferences, the current video applications need to consider not only compression efficiency and quality but also the available bandwidth, memory, computational power and display resolutions for different terminals. The transcoding methods and the use of several encoders to generate different resolution (i.e. spatial, temporal or quality) video streams can be used to address the heterogeneity problem. But above mentioned methods impose additional constraints such as unacceptable delays and increase bandwidth requirements due to redundant data streams. Scalable video coding is an attractive solution for the issues posed by the heterogeneity of today's video communications. Scalable coding produces a number of hierarchical descriptions that provide flexibility in terms of adaptation to user requirements and network/device characteristics. The characteristics of the scalable video coding concept can be utilized to scale the existing 2D video applications into stereoscopic video. For example, colour and depth video can be coded into two scalable descriptors and depending on the receiver terminal capabilities, the users could either render stereoscopic video or shift back to conventional 2D video [37]. This book investigates the adaptability of the scalable video coding concept into backward compatible stereoscopic video applications. Therefore, the background related to scalable video coding is provided.

3.2.1 Scalable coding techniques

At present video production and streaming is ubiquitous as more and more devices are able to produce and distribute video sequences. This brings the increasingly compelling requirement of sending an encoded representation of a sequence that is adapted to the user, device and network characteristics in such a way that coding is performed only once while decoding may take place several times at a different resolution, frame rate and quality. Scalable video coding allows decoding of appropriate subsets of bitstream to generate complete pictures of size and quality dependent on the proportion of the total bitstream decoded. A number of existing video compression standards support scalable coding, such as MPEG-2 Video and MPEG-4 Visual. Due to reduced compression efficiency, increased decoder complexity and the characteristics of traditional transmission systems the above scalable profiles are rarely used in practical implementations. Recent approaches for scalable video coding are based on motion compensated 3D wavelet transform and motion-compensated temporal differential pulse code modulation (DPCM) together with spatial de-correlating transformations [38-41].

The wavelet transform proved to be a successful tool in the area of scalable video coding since it enables to decompose a video sequence into several spatio-temporal subbands. Usually the wavelet analysis is applied both in the temporal and spatial dimensions, hence the term 3D wavelet. The decoder might receive a subset of these subbands and reconstruct the sequence at a reduced spatio-temporal resolution at any quality. The open-loop structure of this scheme solves the drift problems typical of the DPCM-based schemes whenever there is a mismatch between the encoder and the decoder. The scalable video coding based on 3D wavelet transform is addressed in recent research activities [38] [39]. The scalable video coding profiles of existing video coding standards are based on DCT methods. Unfortunately, due to the closed loop, these coding schemes have to address the problem of drift that arises whenever encoder and decoder work on different versions of the reconstructed sequence. This typically leads to the loss of coding efficiency when compared with non-scalable single layer encoding.

In 2007, the Joint Video Team (JVT) of the ITU-T VCEG and the ISO/IEC MPEG standardized a Scalable Video Coding (SVC) extension of the H.264/AVC standard [40]. This SVC standard is capable of providing temporal, spatial, and quality scalability with base layer compatibility with H.264/AVC. Furthermore, this contains an improved DPCM prediction structure which allows greater control over the drift effect associated with closed loop video coding approaches [41].

Bit-streams with temporal scalability can be provided by using hierarchical prediction structures. In these structures, key pictures are coded at regular intervals by using only previous key pictures as references. The pictures between the key pictures are the hierarchical B pictures which are bi-directionally predicted from the key pictures. The base layer contains a sequence of the key pictures at the coarsest supported temporal resolution; while the enhancement layers consist of the hierarchically coded B pictures (see Figure 3.4). A low-delay coding structure is also possible by restricting the prediction of the enhancement layer pictures from only previous frame.

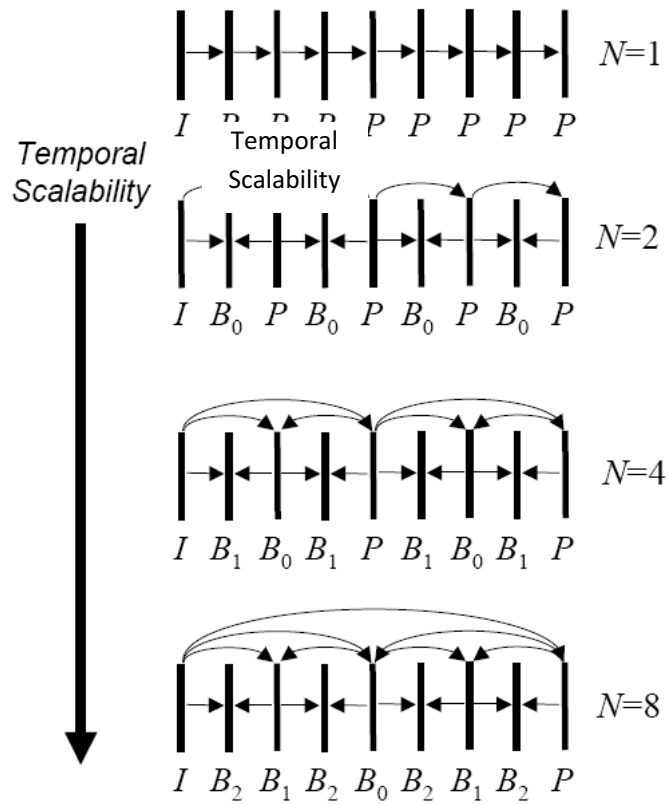


Figure 3.4: Prediction structure for temporal scalability.

> **Apply now**

REDEFINE YOUR FUTURE
**AXA GLOBAL GRADUATE
PROGRAM 2015**

redefining / standards **AXA**

agence.cdg. © Photonistop

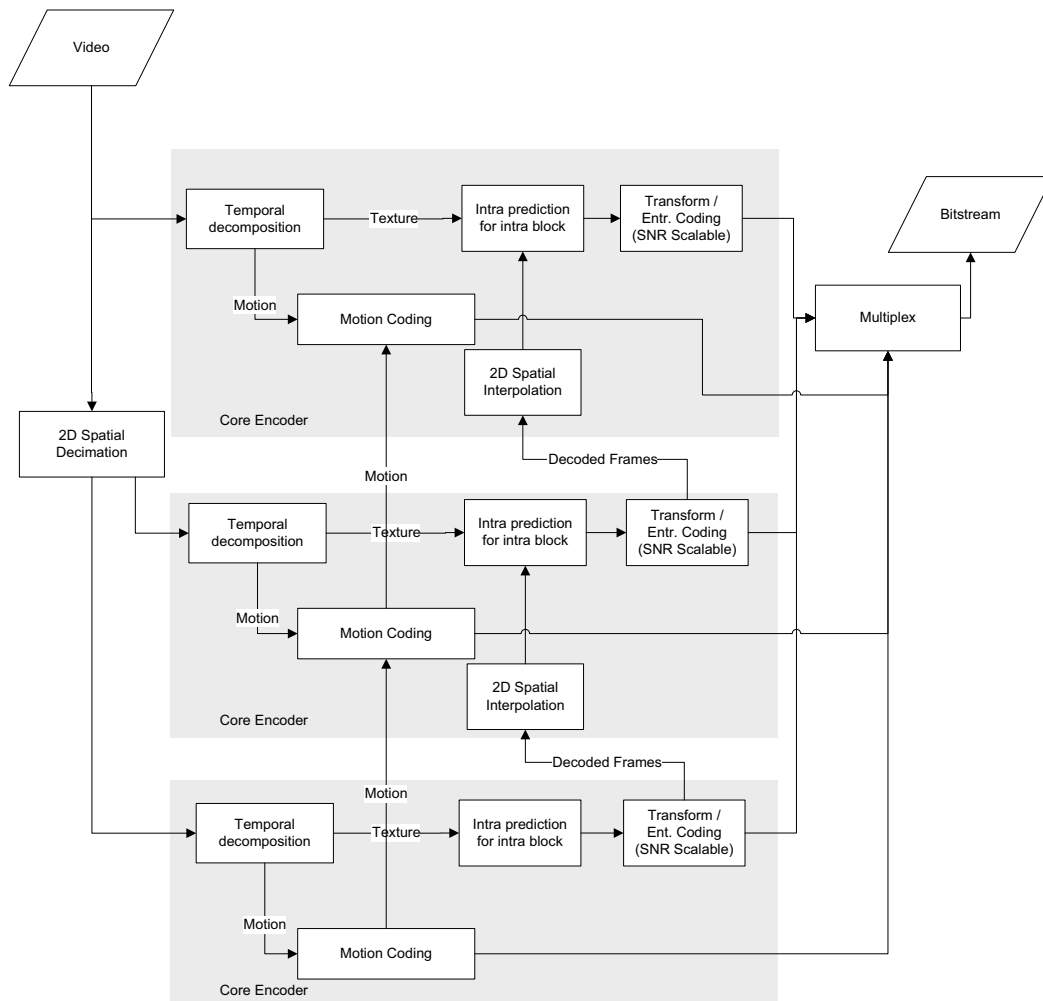


Figure 3.5: Scalable encoder using a multi-scale pyramid with 3 levels of spatial scalability [40]

Spatial scalability is achieved using a multi-layer coding approach in prior coding standards, including MPEG-2 and H.263. Figure 3.5 shows a block diagram of a spatially scalable encoder. In the scalable extension of H.264/AVC, the spatial scalability is achieved with an over-sampled pyramid approach. Each spatial layer of a picture is independently coded using motion-compensated prediction. Inter-layer motion, residual or intra prediction mechanisms can be used to improve the coding efficiency of the enhancement layers. In inter-layer motion prediction, for example, the up-scaled base layer motion data is employed for the spatial enhancement layer coding.

Quality scalability can be considered as a subset of spatial scalability where two or more layers are having similar spatial resolutions but different quality levels. The scalable extension of H.264/AVC also supports quality scalability using coarse-grain scalability (CGS) and medium-grain scalability (MGS). CGS is achieved using spatial scalability concepts with the exclusion of the corresponding up-sampling operations in the inter-layer prediction mechanisms. MGS is introduced to improve the flexibility of bit-stream adaptation and error robustness.

The SVC technology is fast moving towards the deployment of practical, real-time scalable applications [42]. The associated technologies which support scalable applications over communication channels are emerging. For example, Wenger *et al* draft a RTP (Real-time Transport Protocol) payload format for SVC video [43]. This book investigates the possibility of utilizing SVC concept to deliver backward compatible stereoscopic video applications over the networks. Moreover, the scalability features will be employed to encode colour plus depth map stereoscopic video more efficiently with minimal effect for the perceived quality of 3D video.

3.3 3D Video Coding

Scenery captured from different angles/viewpoints simultaneously results a large amount of raw video data. For instance, stereoscopic video, which is one of the simplest forms of 3D video, requires twice the size of storage capacity and double the bandwidth compared to the requirements of 2D video. Thus, 3D video coding is crucial if immersive video applications are to be available for the mass consumer market in the near future. The MPEG Ad-hoc group which worked on exploration of 3DAV (3D Audio-Visual) application scenarios and technologies, identified several MPEG coding tools which are directly applicable for different representations of 3D video and also identified the areas where further standardization is required [5]. The coding approaches for 3D video may be diverse depending on the representation of 3D video. For example, available principles of classical video coding can be utilized to compress pixel-type data including stereo video, multi-view video, and associated depth or disparity maps. However, efficient coding approaches and standardization for some 3D representations such as multi-view video are yet to be discovered. The compression algorithms for 3D mesh models have also reached a high level of maturity. Even though there are efficient compression algorithms for static geometry models, compression of dynamic 3D geometry is still an active field of research. A survey of coding algorithms for different 3D application scenarios is presented in [44]. The coding of image-based 3D representations (e.g. stereoscopic video) is addressed in this book. Henceforth, the encoding of image based 3D representations is discussed in this chapter.

The 3D image/video encoding approaches aim at exploiting inter-view statistical dependencies in addition to the conventional encoding approach for 2D video, which removes the redundancies in the temporal and spatial domains. The prediction of views utilizing the neighbouring views and the images from the same image sequence are shown in Figure 3.6. In general temporal prediction tends to be more efficient compared to inter-view prediction and combined prediction [45]. However, the efficiencies of these prediction methods are varied depending on the frame rate, camera distances and the complexity of content (e.g. motion and spatial details). The JVT (Joint Video Team) standardized the multi-view extension of H.264/AVC [46]. The adopted approach is based on the hierarchical B-frames syntax [see Figure 3.7]. T_0, T_1, \dots, T_{100} in Figure 3.7 represent consecutive time instances of image capture whereas S_0, S_1, \dots, S_7 represent consecutive camera positions. Each image of the multi-view image sequence is encoded using spatial, temporal or inter-view predictions. This multi-view coding strategy has shown improved coding gain through exploiting inter-view statistical redundancies compared to encoding each view separately. However, the gain is significant only for dense camera settings but not for dome-type (i.e. cameras are set far apart) arrangements. Moreover, the proposed improvements for existing multi-view coding algorithms can be found in the research literature, which enhance the application requirements including random access, low delay and memory optimization [47]. Specific coding tools for MVC such as illumination compensation, view interpolation prediction, inter-view direct mode are also under JVT investigation at present.



BI

Business
Strategic Marketing Management
International Business
Leadership & Organisational Psychology
Shipping Management
Financial Economics

BI NORWEGIAN BUSINESS SCHOOL

EFMD
EQUIS
ACCREDITED

Empowering People. Improving Business.

BI Norwegian Business School is one of Europe's largest business schools welcoming more than 20,000 students. Our programmes provide a stimulating and multi-cultural learning environment with an international outlook ultimately providing students with professional skills to meet the increasing needs of businesses.

BI offers four different two-year, full-time Master of Science (MSc) programmes that are taught entirely in English and have been designed to provide professional skills to meet the increasing need of businesses. The MSc programmes provide a stimulating and multi-cultural learning environment to give you the best platform to launch into your career.

- MSc in Business
- MSc in Financial Economics
- MSc in Strategic Marketing Management
- MSc in Leadership and Organisational Psychology

www.bi.edu/master



1st order neighbours

- T (temporal)
- S (inter-view)
- L/R (combined)

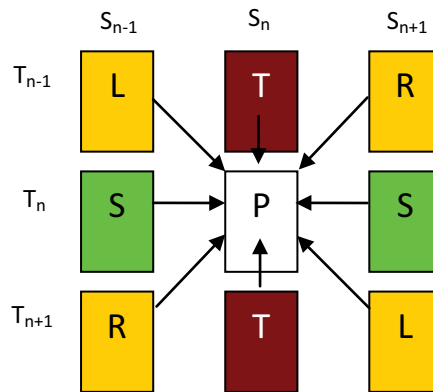


Figure 3.6: Coding of multi-view video

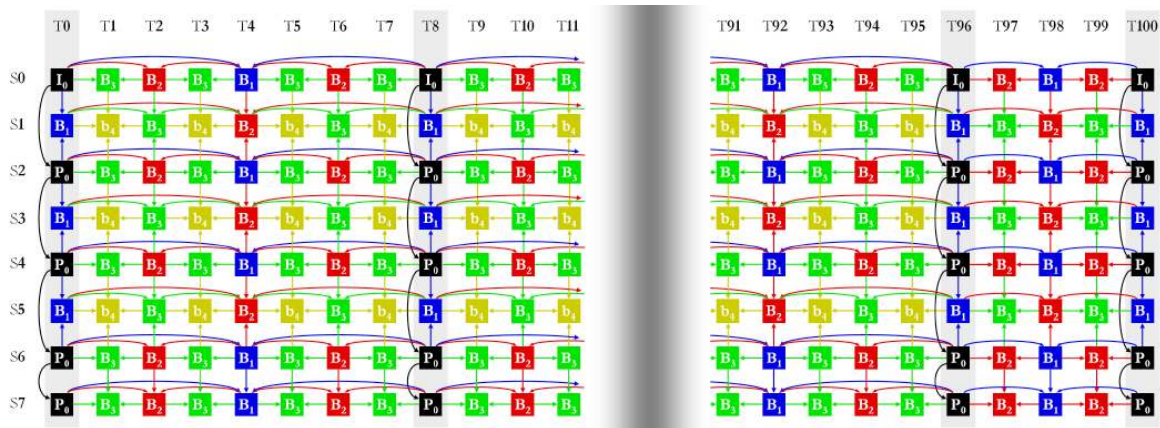


Figure 3.7: Multi-view coding [46]

3.4 Stereoscopic Video Coding

Conventionally, stereoscopic video is captured by two cameras a small distance apart producing two views which differ somewhat of the same object. Therefore, stereoscopic video coding aims to exploit the redundancies present in two adjacent views while removing the temporal and spatial redundancies. For example, an extended H.264 video codec for stereoscopic video, exploiting spatial, temporal, disparity and world-line correlation is described in [48]. Instead of coding the left and right views separately, the coding of left view and the disparity compensated right image has been an efficient method of coding. Disparity estimation algorithms are utilized to identify geometric correspondence of objects in stereo pair [49]. A disparity compensated residual image coding scheme for 3D multimedia applications is described in [50]. The proposed stereoscopic video coding algorithm in [51] employs an object-based approach in order to overcome the artefacts caused by block-based disparity estimation approaches. However, the disparity-compensated prediction approach can not be applicable to colour plus depth video as colour image and associated depth image have different texture information and number of objects. In addition, mixed-resolution video coding for the left and right views is based on the binocular suppression theorem which is based on the response of the human visual system for depth perception. This theory suggests that 3D perception is not affected if one of the view can be in high quality compared to the other view [134]. The high quality view drives 3D perception by compensating the slight loss in the other view. H.264/AVC based stereoscopic video codec described in [52] utilizes asymmetric coding of the left and right images in order to improve the compression efficiency. The results in [52] show that stereoscopic video can be coded at a rate about 1.2 times the monoscopic video using spatial and temporal filtering of one image sequence. Moreover, a 3D video system based on H.264/AVC view coding described in [53] examines the bounds of asymmetric stereo video compression. The mixed-resolution coding concept can be applied to colour plus depth video in different perspectives, i.e. to code the colour and depth video based on their influence towards better perceptual quality. For example, the depth image can be coded differently compared to the corresponding colour image [130].

The multi-view profile (MVP) of the MPEG-2 video coding standard can be utilized in coding stereoscopic video with Disparity-Compensated Prediction (DCP). This profile is defined based on the Temporal Scalability (TS) mode of MPEG-2 [54]. However, due to the coarser disparity vectors which are defined on a block-by-block basis of 16x16 pixels, the inter-view prediction error is larger compared to the motion compensated error [55]. Therefore, the gain obtained with this coding approach is not significant compared to the coding of left and right views separately. However, the proposed improvements (e.g. global displacement compensated prediction [56]) for MPEG-2 multi-view profile have shown improved picture quality at low bitrates. Due to the disparity estimated prediction structure of the MPEG-2 multi-view profile this approach is not suitable for encoding colour plus depth video.

The MPEG-4 Multiple Auxiliary Component (MAC) is a generalization of the grayscale shape coding [57]. This allows image sequences to be coded with a Video Object Plane (VOP) on a pixel-by-pixel basis which contains data related to video objects such as disparity, depth and additional texture. However, this approach needs to send the grayscale/alpha shape with any other auxiliary component (e.g. disparity) and as a result the coding efficiency is affected. Even though MPEG-4 MAC is suitable to encode colour plus depth video due to the presence of grayscale reduces the coding gain. Therefore, the performance of a modified MPEG-4 MAC coder with no alpha plane is analyzed in this book in the last part of this chapter.

The multi-view codec developed by JVT can also be utilized in stereoscopic video coding with inter-view disparity compensation [46]. However, this method is not applicable for colour plus depth stereoscopic video due to the availability of different texture information in each image sequence. Consequently, the amount of redundancies which can be exploited through inter-layer prediction is minimal. The recent interest on multi-view plus depth representation may be suitable to encode monoscopic video plus depth information [29]. However, these approaches are developed for interactive multi-view applications such as free-viewpoint video and may not be so effective towards more simple representations like stereoscopic video.

Need help with your dissertation?

Get in-depth feedback & advice from experts in your topic area. Find out what you can do to improve the quality of your dissertation!

Get Help Now



Go to www.helpmyassignment.co.uk for more info



The MPEG-4 Animation Framework eXtension (AFX) standard ISO/IEC 14496-16 defines two image-based depth representations that can be used to encode colour plus depth video [58]. The two categories are the simple Depth Image (DI) and Layered Depth Image (LDI) representation which is utilized to store occluded areas of the original image. However, this approach seems very much computer graphics oriented and suitable for synthetic image synthesis [58]. Therefore, this approach is not suitable for encoding natural scenes using monoscopic video plus depth representations of 3D video.

The quality of the virtual left and right views generated through DIBR is dependent on the quality of the colour video, associated depth map and the camera geometry. The relationship can be attributed to the Quantization Parameters (QPs) of each image and the virtual camera position (see Equation 3.1).

$$\text{Quality} = f(QP_{\text{colour}}, QP_{\text{depth}}, \text{Camera position}) \quad \text{Equation 3.1}$$

However, the effect of colour image quality is more influential with this approach than the depth image quality as texture information is directly viewed by the users. Therefore, effects of colour and depth coding on the perceptual quality needs to be studied thoroughly. This chapter investigates the perceptual bounds of depth map coding for stereoscopic video applications. For example, spatially down-sampled depth maps would be accurate enough to generate good quality left and right views. Moreover, the efficient coding architectures for colour plus depth map coding and their suitability for communication applications are reported in this chapter.

Due to the characteristics of the depth image sequence (e.g. smoothness of real world objects), it can be efficiently compressed with the existing coding algorithms than the coding of corresponding monoscopic video sequence [59]. According to [59], H.264/AVC outperforms MPEG-2 and MPEG-4 video coding standards in depth map coding. For example, depth maps can be encoded with H.264/AVC at a bitrate less than 20% of the MPEG-2 coded colour image sequence for Standard Definition (SD) resolution image sequences. The following subsection describes potential encoding approaches for stereoscopic 3D video.

3.4.1 Exploration of Efficient Stereoscopic Video Coding Configurations

Simulcast configurations:

- This approach generates two bit-streams
- Needs multiplexing and de-multiplexing modules after and before the video codec
- Inter-view redundancies are not removed, and hence less efficient

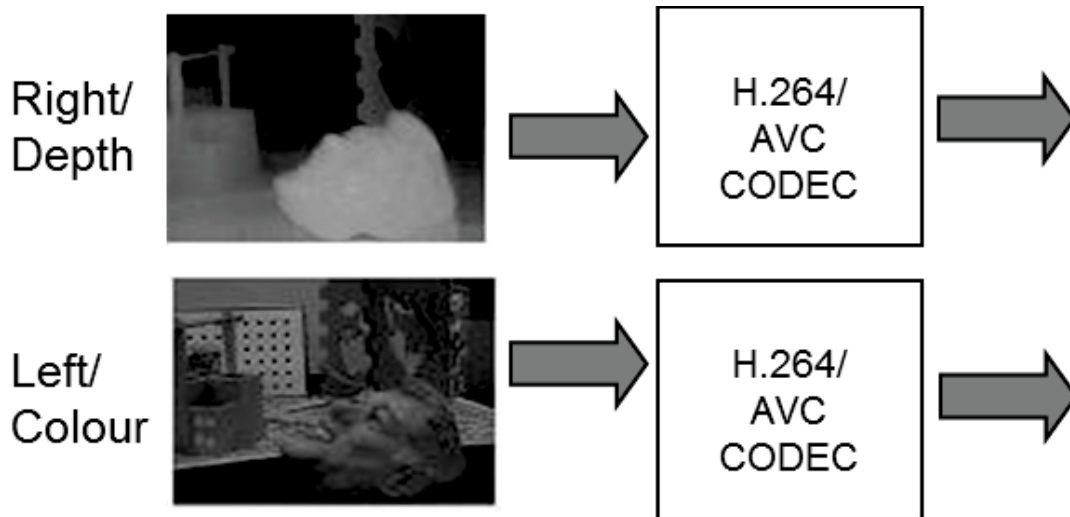


Figure 3.8: Parallel/Simulcast encoding approach for 3D video

Pre-processing configurations: Also known as frame packing 3D format. This refers to the combination of two frames, one for the left eye and the other for the right eye, into a single “packed” frame that consists of these two individual sub-frames. The key difference of a Frame Packing signal is that each sub-frame for each eye is still at full resolution, i.e., 1920×1080 for a 1080p Frame Packing signal, and 1280×720 for 720p Frame Packing 3D content.

Side-by-Side approach:

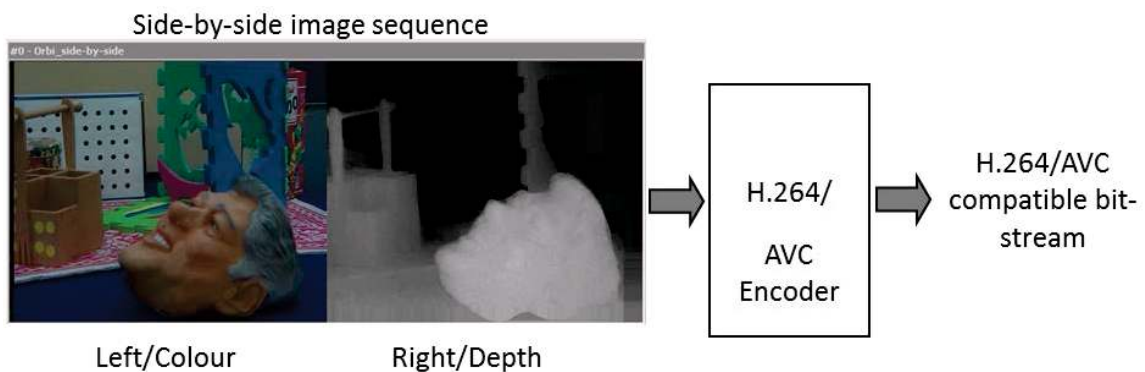


Figure 3.9: Frame packed configuration- side-by-side format

Top-Bottom approach:

Interlaced image sequence

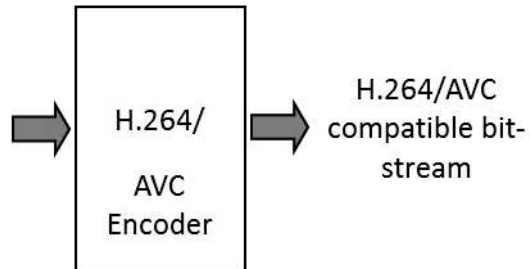


Figure 3.10: Frame packed configuration- top-bottom format Type 1 (interlaced format)

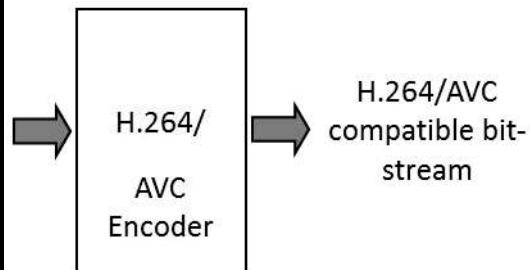
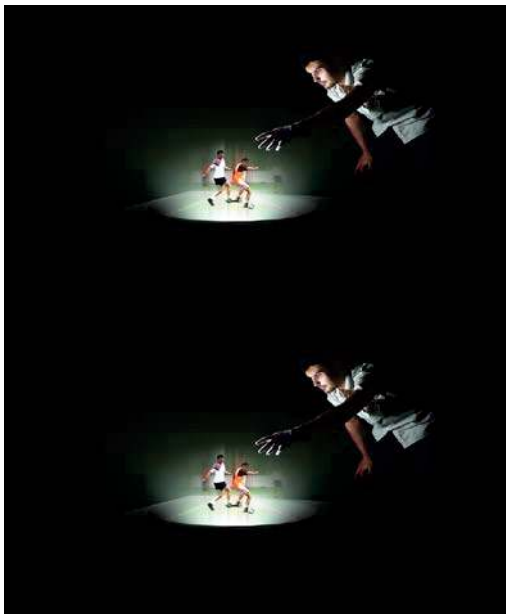


Figure 3.11: Frame packed configuration- top-bottom format Type 2

- This approach generates a single output bit-stream
- A single coded can be utilized for encoding
- Inter-view redundancies are not removed, and hence less efficient

Frame compatible 3D: Both Frame Compatible 3D and Frame Packing 3D formats involve forming a single frame that contains “sub-frames” for the left and right eye. In both cases, the Sub-frames can be packaged together into a single frame via the Side-by-Side 3D format or the Top-and-Bottom 3D format. The key difference of a Frame Compatible signal is that each sub-frame for each eye is down sampled along one axis to lower the resolution of each sub-frame along one axis. As a result, the total dimension of a Frame Compatible Frame is the same as a regular 2D HD frame (since each sub-frame has half the resolution along either the horizontal or vertical dimension). This is the reason this format is called Frame Compatible.



Brain power

By 2020, wind could provide one-tenth of our planet's electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generating costs relate to maintenance. These can be reduced dramatically thanks to our systems for on-line condition monitoring and automatic lubrication. We help make it more economical to create cleaner, cheaper energy out of thin air.

By sharing our experience, expertise, and creativity, industries can boost performance beyond expectations. Therefore we need the best employees who can meet this challenge!

The Power of Knowledge Engineering

Plug into The Power of Knowledge Engineering.
Visit us at www.skf.com/knowledge

SKF

Download free eBooks at bookboon.com



Click on the ad to read more

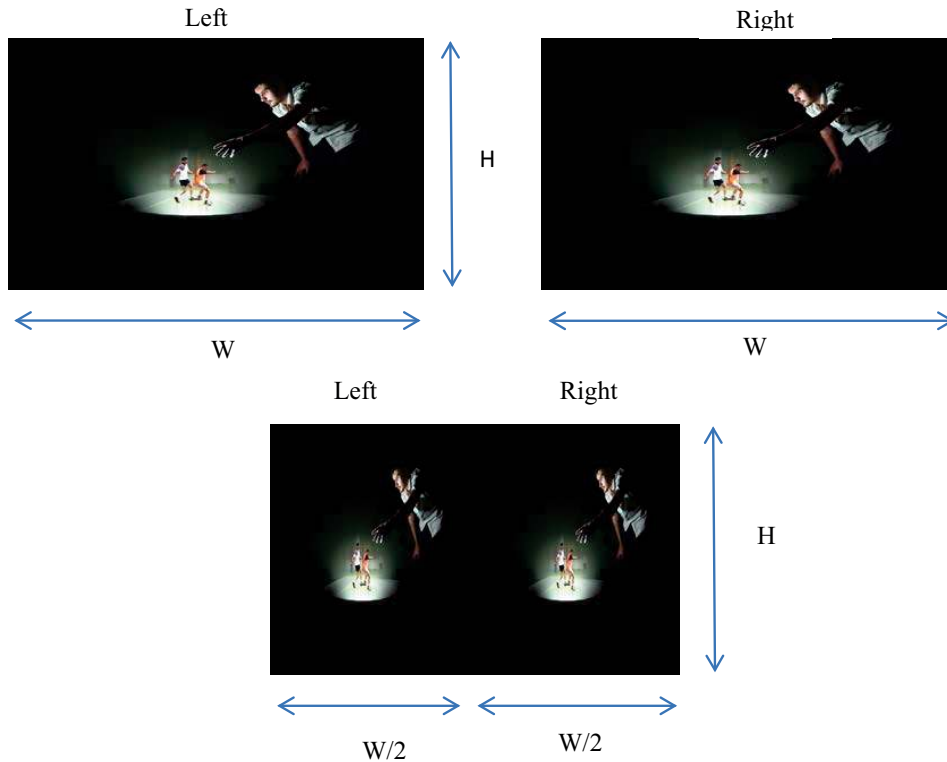


Figure 3.12: Frame compatible format for 3D video

Layered encoding configurations (e.g., MPEG-2 TS, Scalabel H.264/AVC, Multi-View Coding):

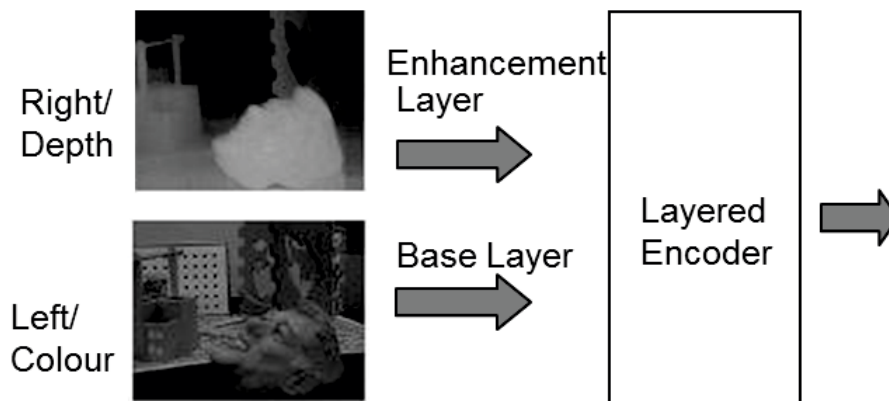


Figure 3.13: Layered encoding approach for 3D video

- Generates a single bit-stream
- Removes inter layer redundancies
- Backward compatibility (i.e., the base layer is compatible for 2D video decoding)
- Asymmetric coding support using temporal, spatial and quality scalability options available with H.264/SVC

Two layered encoding approaches for stereoscopic 3D video are describe in detail below.

Download free eBooks at bookboon.com

Stereoscopic Video Coding Configuration Based on Scalable Video Coding (SVC): The Scalable Video Coding (SVC) encodes video data in different resolutions (spatial/temporal) and quality (refer to the SVC section above). Depending on the receiver capabilities the users can decode the full-resolution image or sub-version of the image. Most of the SVC methods utilize the layered coding approaches to encode scalable image sequences. This study proposes a stereoscopic video coding configuration based on the layered coding architecture of SVC. The proposed method is implemented on the scalable extension of H.264/AVC. Scalable H.264/AVC is developed and standardized by JVT and supports spatial, temporal and quality scalability for video coding [40]. In this coding configuration, the colour and depth image sequences are coded at the base and enhancement layers respectively as shown in Figure 3.14. In addition, this layered coding architecture can also be utilized to encode left and right view based stereoscopic video. The coding of left and right views with this coding configuration is considered in this chapter during the performance comparison of colour plus depth coding vs. left and right view coding.

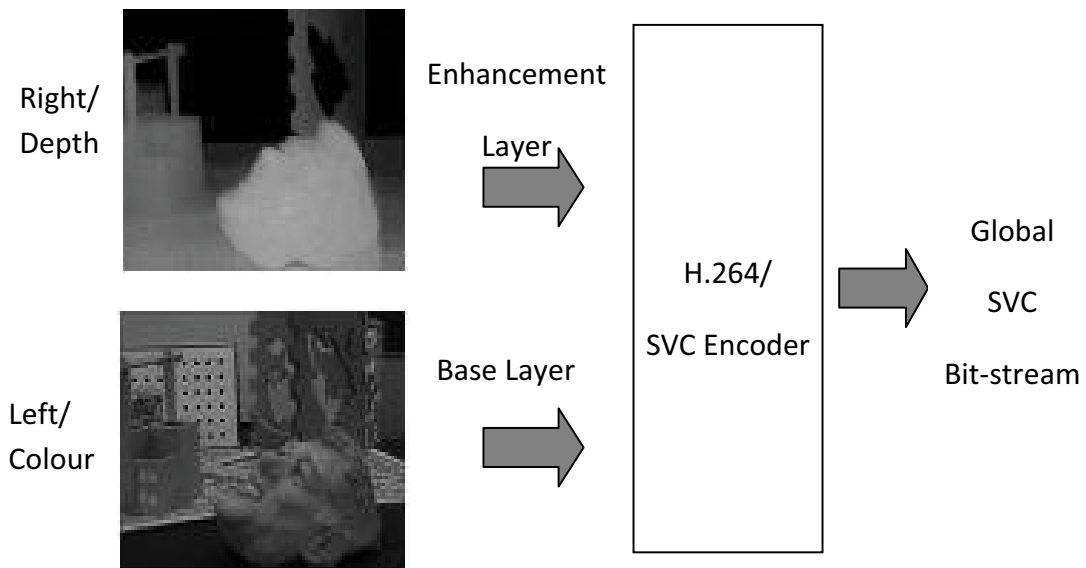


Figure 3.14: Stereo video coding configuration based on scalable H.264/AVC.

As the base layer of this configuration is compatible with H.264/AVC decoders, users with a H.264/AVC decoder will be able to decode the colour image sequence, whereas users with a SVC decoder and 3D rendering hardware will be able to decode the depth map sequence and experience the benefits of stereoscopic video. Therefore, the backward compatibility nature of this scalable coding configuration can be employed to enhance or scale existing 2D video applications into stereoscopic video applications. Furthermore, this configuration can be utilized to exploit asymmetric coding of the colour and depth map image sequence since scalable H.264/AVC supports a range of temporal, spatial and quality scalable layers. For instance, during encoding, the resolution of colour and depth image sequences can be independently changed based on their inherent characteristics, without affecting the perceptual qualities of 3D video. The proposed asymmetric coding methods with this scalable coding configuration are presented in [131]. Furthermore, inter-layer redundancies can be exploited based on the correlation between the colour and depth images. The results presented here make use of the adaptive inter-layer prediction option in the JSVM (Joint Scalable Video Model) reference software codec, which selects the coding mode for each MB (Macro-Block) using Rate-Distortion (R-D) optimization. Moreover, this coding configuration complies with the ISO/IEC 23002-3 (MPEG-C part 3) standard, which specifies the inter-operability among applications based on colour plus depth/disparity sequences [28]. The output of this coding configuration is a single SVC bit-stream which can be partially decoded. Therefore, the transmission and synchronization of this content over communication channel is not difficult compared to sending colour and depth streams separately. Furthermore, the supportive technologies (e.g. RTP format for SVC) are emerging to support the delivery of SVC data streams over communication channels [43]. Subsection 3.5.1 discusses the coding efficiency of the proposed SVC configuration compared to MPEG-4 MAC and H.264/AVC based stereoscopic video coding approaches.

Stereoscopic Video Coding with MPEG-4 Multiple Auxiliary Components (MAC): The MAC (Multiple Auxiliary Components) is added to Version 2 of the MPEG-4 Visual part [57] in order to describe the transparency of video objects. The MAC is defined for a video object plane (VOP) on a pixel-by-pixel basis and contains data related to video objects such as disparity, depth, and additional texture. As MPEG-4 MAC allows the encoding of auxiliary components (e.g. depth, disparity, shape) in addition to the Y, U and V components present in 2D video, this coding approach can be utilized to compress colour plus depth stereoscopic video [108] or disparity compensated stereoscopic video [109]. The coding of monoscopic video plus depth map with MPEG-4 MAC is illustrated in Figure 3.15. Even though the encoding of shape information is compulsory with any other auxiliary data stream (i.e. depth map), the MAC coder utilized in this book is modified to remove the shape coding requirement of MPEG-4 MAC. Consequently, this coding configuration is expected to present improved coding efficiency compared to the original MPEG-4 MAC coder [55].

MPEG-4 MAC is a good mechanism for generating one-stream stereoscopic video coding output. The one-stream approach facilitates end-to-end video communication applications without a system level modification (avoid multiplexing and de-multiplexing stages for different streams) and can be in compliant to the ISO/IEC 23002-3 (MPEG-C part 3) standard. In this work, the R-D performance of MPEG-4 MAC coded colour plus depth stereoscopic video is compared against results obtained with the H.264/AVC and scalable H.264/AVC video coding standards. However, it should be noted that the MPEG-4 MAC utilized is based on the ISO/IEC 14496-2 standard (MPEG-4 Visual), rather than the AVC technology, which is ISO/IEC 14496-10.



"I studied English for 16 years but...
...I finally learned to speak it in just six lessons"
Jane, Chinese architect

ENGLISH OUT THERE

Click to hear me talking before and after my unique course download

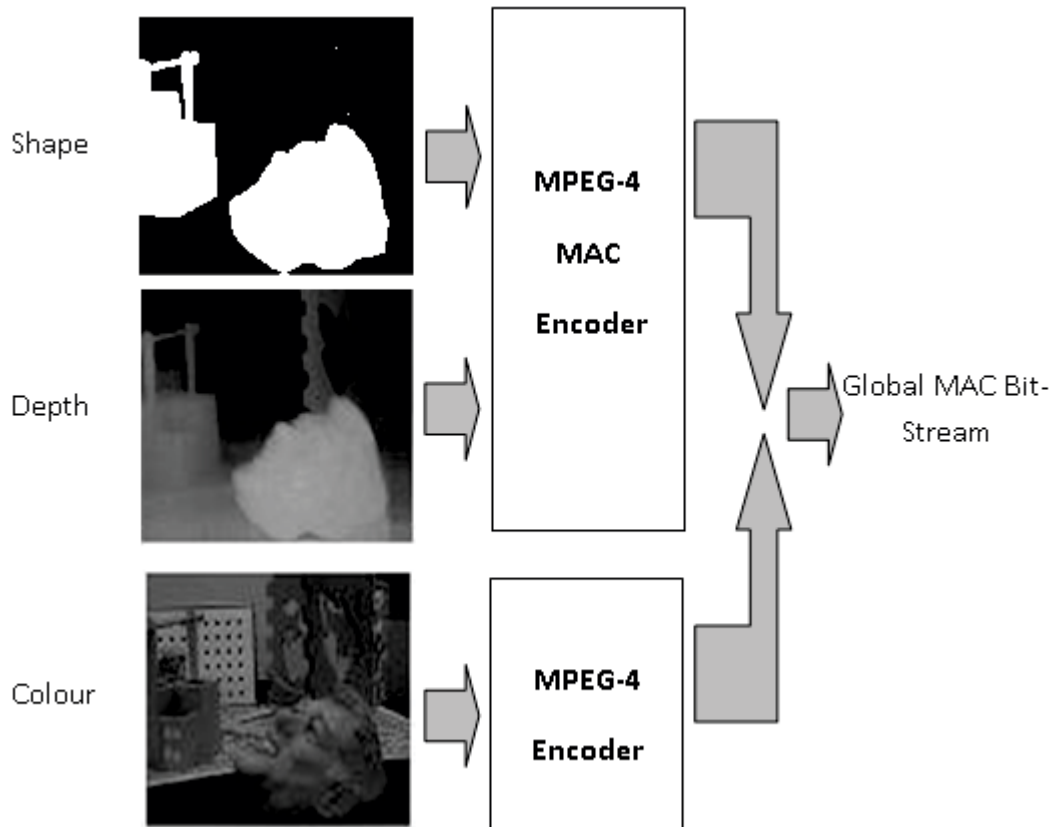


Figure 3.15: Stereo video coding configuration based on MPEG-4 MAC.

3.5 Performance analysis of different encoding approaches for colour plus depth based 3D video and comparison of left and right view encoding vs. colour plus depth map video encoding

The following encoding configurations are used in this evaluation;

- Layered encoding with H.264/AVC Scalable Video Coding
- MPEG-4 MAC Encoding
- H.264/AVC Encoding with side-by-side stereoscopic 3D video (not frame compatible format)

The Orbi and Interview test sequences are utilised to obtain the Rate-Distortion (R-D) results. The experiments are carried out using CIF (Common Intermediate Format, 352×288) format image sequences in order to evaluate the performance of low bitrate 3D video applications. As the original resolution of these sequences is 720×576, the image sequences are initially cropped (i.e. 704×576) and then down-sampled into CIF (352×288) format video. During the down-sampling process, four nearest pixel values of the original format are averaged to represent the pixel value of the down-sampled image.

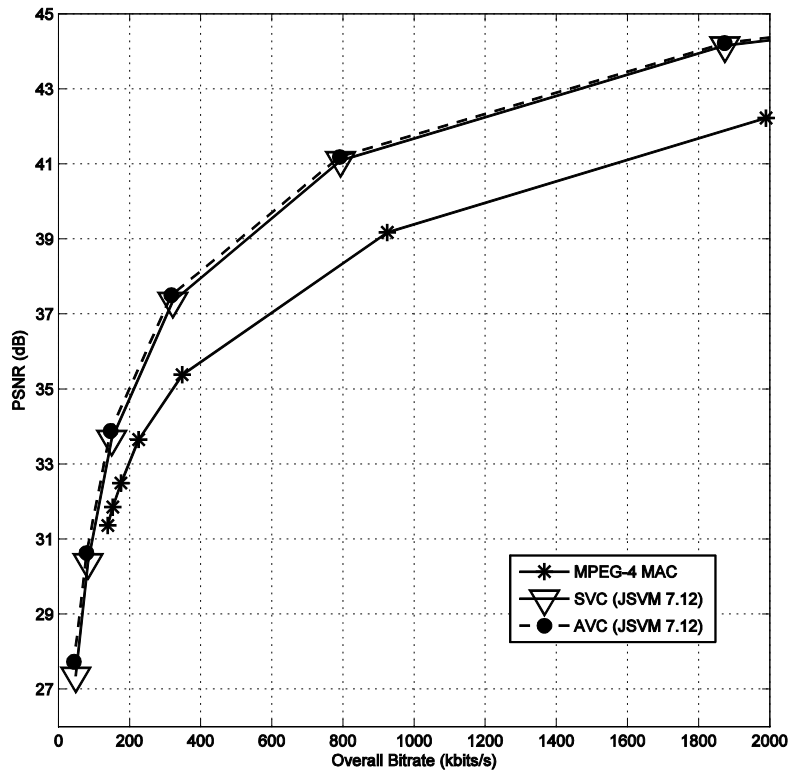
The image sequences are encoded with the three encoding configurations based on the existing video coding standards, MPEG-4 MAC, H.264/AVC (using the scalable H.264/AVC single layer coding) and the scalable extension of H.264/AVC. The base layer encoder of the scalable H.264/AVC (i.e. encode using the single layer configuration) is utilized to obtain H.264/AVC coding results as the base layer bit-stream is backward compatible for H.264/AVC, and to avoid differences arising from the different software implementations of H.264/AVC and the scalable extension of H.264/AVC. The basic encoding parameters used are shown in Table 3.1. The Quantisation Parameter (QP) in the configuration file is varied to obtain the bitrate range shown in the R-D curves. The same QP is used for encoding both base and enhancement layers of scalable H.264/AVC. The R-D curves show the image quality measured in PSNR against the resulting average bitrate. Experiment 1 compares the coding performance of the stereoscopic video coding architectures. The coding performance of colour and depth video vs. left and right video with the proposed SVC coding configuration is discussed in Experiment 2.

Encoding Parameter	Value
Test sequences	Orbi, Interview
No. of frames	300
Sequence format	IPPP...
Reference frames	1
Search range	16 pixels
Entropy coding	VLC (Variable Length Coding)

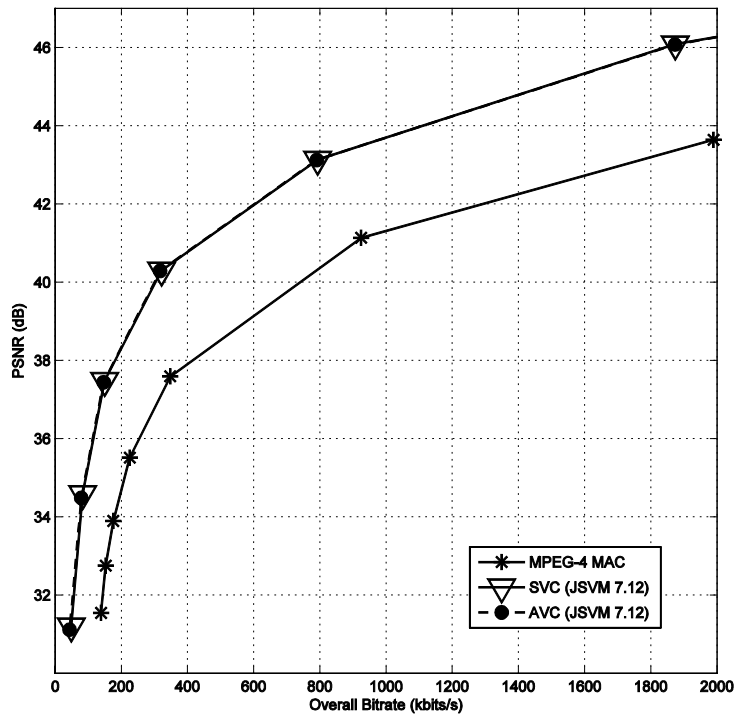
Table 3.2: Encoding parameters

3.5.1 Experiment 1: Comparison of Stereoscopic Video Encoding Configurations

The R-D performance of Orbi and Interview colour and depth sequences using MPEG-4-MAC (layered encoding), H.264/AVC (frame packing, side-by-side format) and scalable H.264/AVC coding (layered encoding) configurations are shown in Figures 3.16 and 3.17 respectively. All of the results are plotted against the overall bitrate (output bitrate of the SVC codec), which includes all of the overhead, texture, colour, motion vector bits of both colour and depth map video. In order to highlight the R-D performance at a range of bitrates, the final bitrate is shown from 0 Kbits/s to 2Mbits/s. The H.264/AVC coded stereoscopic video sequences (i.e. side-by-side images) are separated into colour and depth map video in order to calculate the PSNR with respect to their original colour and depth image sequences.

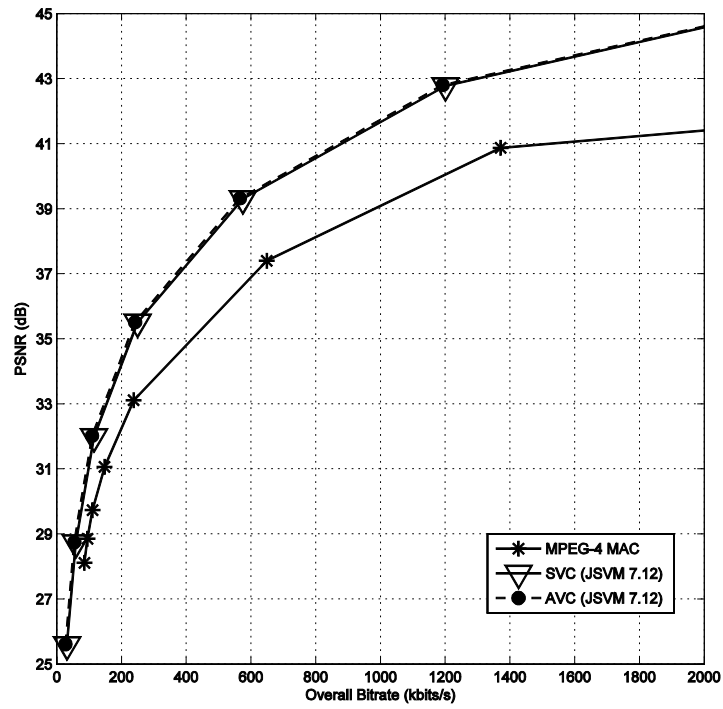


(a)

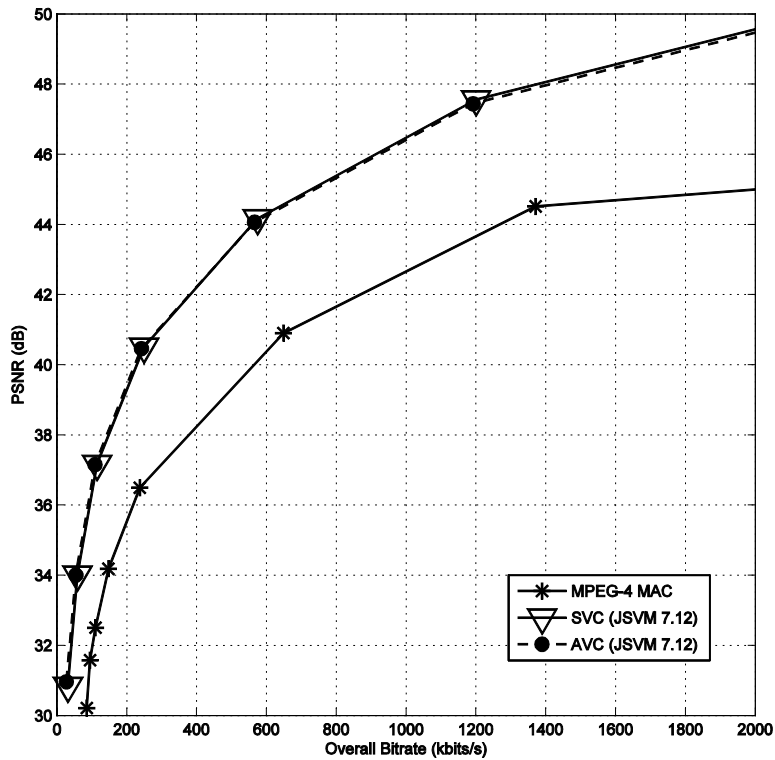


(b)

Figure 3.16: R-D curves for Orbi sequence (a) Colour image sequence (b) Depth image sequence.



(a)



(b)

Figure 3.17: R-D curves for Interview sequence (a) Colour image sequence (b) Depth image sequence.

The R-D curves for both Orbi and Interview sequences show that the proposed configuration with scalable H.264/AVC performs similar to the H.264/AVC configuration, and outperforms the MPEG-4 MAC based configuration at all bitrates. The SVC configuration has not outperformed the H.264/AVC configuration due to the negligible usage of inter-layer prediction between the base and enhancement layers and extended header data associated with SVC. Even though, common object boundaries are present in colour and depth map sequences, the depth image sequence has different texture information, and less number of objects and high frequency components compared to the colour video. Consequently, the number of inter-layer coding modes used is very small during the encoding process. In addition, the R-D performance of the scalable H.264/AVC configuration is slightly affected due to the overhead bits added to the SVC bit-stream to recognize enhancement layer data. However, the flexibility of the SVC configuration in stereoscopic video coding, such as asymmetric coding support (temporal, spatial and quality scalability for depth image sequences), single bit-stream output, and backward compatibility, facilitates end-to-end stereoscopic video communication chain to a greater extent. The flexible macro-block (MB) sizes and skipped MB features available in H.264/AVC standard have helped to achieve better coding performance with the H.264/AVC and scalable H.264/AVC based configurations compared to the MPEG-4 MAC configuration at all bitrates. Furthermore, it can be observed that the configurations based on H.264/AVC provide reasonable image quality at very low overall bitrates compared to the MPEG-4 MAC configuration.

What do you want to do?

No matter what you want out of your future career, an employer with a broad range of operations in a load of countries will always be the ticket. Working within the Volvo Group means more than 100,000 friends and colleagues in more than 185 countries all over the world. We offer graduates great career opportunities – check out the Career section at our web site www.volvogroup.com. We look forward to getting to know you!

VOLVO
AB Volvo (publ)
www.volvogroup.com

VOLVO TRUCKS | RENAULT TRUCKS | MACK TRUCKS | VOLVO BUSES | VOLVO CONSTRUCTION EQUIPMENT | VOLVO PENTA | VOLVO AERO | VOLVO IT
VOLVO FINANCIAL SERVICES | VOLVO 3P | VOLVO POWERTRAIN | VOLVO PARTS | VOLVO TECHNOLOGY | VOLVO LOGISTICS | BUSINESS AREA ASIA

Download free eBooks at bookboon.com



The H.264/AVC and scalable H.264/AVC configurations outperform the MPEG-4 MAC configuration by a considerable margin for depth image quality at all overall bitrates (see Figure 3.16 (b) and 3.17 (b)). The fewer number of objects and unavailability of high frequency components help to achieve superior depth map quality for H.264/AVC based configurations. These smooth depth images can be highly compressed using flexible MB sizes and skipped MB modes available in H.264/AVC [110] compared to other MPEG video coding standards. The gain is more visible in the Interview sequence, which has less motion and stationary background.

The subjective image qualities of the Orbi colour and depth map sequences are illustrated in Figures 3.18 and 3.19 respectively. The image sequences are obtained at an overall bitrate of 150 Kbits/s, utilizing the coding configurations based on scalable H.264/AVC and MPEG-4 MAC. According to Figure 3.18, subjective quality of the SVC coded colour image is better compared to that of the MPEG-4 MAC coded colour image. This is more visible in the depth image sequences as shown in Figure 3.19. The scalable H.264/AVC coded depth image demonstrates a sharp and better image quality compared to that of the MPEG-4 MAC coded depth image sequence at the given low bitrate of 150 Kbits/s

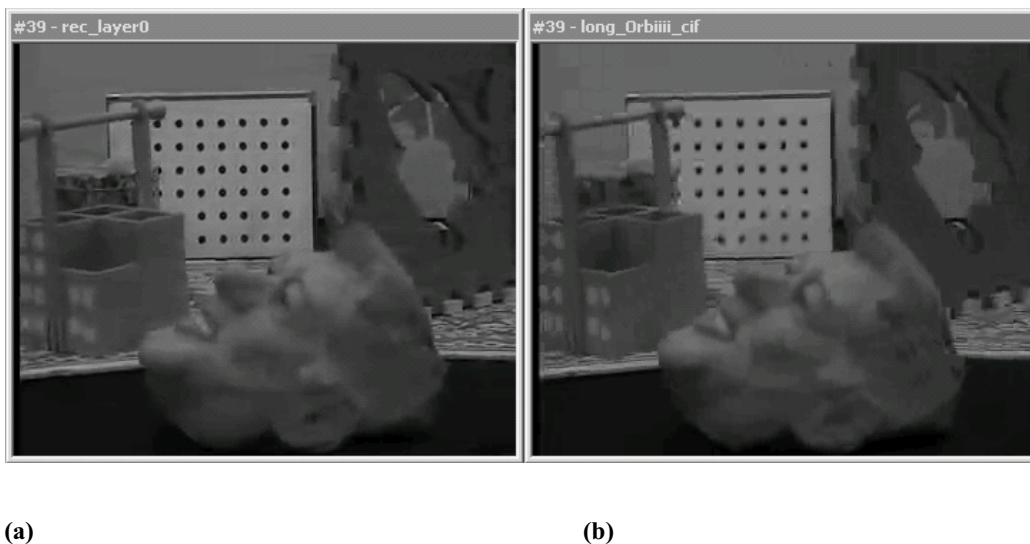


Figure 3.18: Subjective image quality of the Orbi colour sequence at an overall bitrate of 150 Kbits/s (a) Scalable H.264/AVC configuration (b) MPEG-4 MAC configuration.



(a)

(b)

Figure 3.19: Subjective image quality of the Orbi depth sequence at an overall bitrate of 150 Kbits/s (a) Scalable H.264/AVC configuration (b) MPEG-4 MAC configuration.

qaieteye[®]
Challenge the way we run

**EXPERIENCE THE POWER OF
FULL ENGAGEMENT...**

**RUN FASTER.
RUN LONGER..
RUN EASIER...**

**READ MORE & PRE-ORDER TODAY
WWW.GAITEYE.COM**



Table 3.3 shows the image quality of both sequences at an overall bitrate of 200 Kbits/s. This shows that the proposed stereoscopic video coding configuration based on scalable H.264/AVC provides superior quality compared to the MPEG-4 MAC configuration at overall bitrates as low as 200Kbits/s. The high performance and flexible features (backward compatibility and temporal, spatial and quality scalability) associated with the SVC architecture can be employed to scale low bitrate conventional video applications into stereoscopic video applications. Furthermore, at an overall bitrate of 200Kbits/s the Orbi depth image sequence can be coded at 49% of the Orbi colour image bitrate using the proposed configuration based on scalable H.264/AVC. The depth image bitrate requirements could be further reduced by using a high QP value or reduced temporal or spatial scalability at the enhancement layer (depth image) encoder without affecting the perceptual quality of the stereoscopic video application. Moreover, the occlusion problems associated with the DIBR method can also be resolved using this proposed SVC configuration with the use of several pairs of colour plus depth images, which is known as Layered Depth Images (LDI).

Encoding configuration	Orbi Y-PSNR (dB)		Interview Y-PSNR (dB)	
	Colour	Depth	Colour	Depth
Proposed (Scalable H.264/AVC)	34.74	38.31	34.22	39.29
MPEG-4 MAC	33.05	34.68	32.25	35.52
H.264/AVC	35.01	38.33	34.41	39.41

Table 3.3: Image quality at an overall bitrate of 200 Kbits/s

3.5.2 Experiment 2: Colour plus Depth Coding vs. Left and Right View Coding

This experiment compares the R-D performance of colour and depth image sequences vs. left and right image sequences encoding using the efficient scalable H.264/AVC configuration. In order to produce left and right image sequences, the Orbi and Interview sequences are projected into virtual left and right image sequences using the DIBR as described in Equation 3.2 and coded as the base and enhancement layers with the SVC coding configuration. The rendered left and right image sequences can be considered as stereoscopic video captured using a stereo camera pair. Then, the coded colour and depth image sequences at the base and enhancement layers (i.e. using the scalable H.264/AVC) are also converted to virtual left and right video to be compared with the coded left and right video.

$$P_{pix} = -x_B \frac{N_{pix}}{D} \left[\frac{m}{255} (k_{near} + k_{far}) - k_{far} \right] \quad \text{Equation 3.2}$$

Where, N_{pix} and x_B are the number of horizontal pixels of the display and eye separation respectively. The depth value of the image is represented by the N -bit value m . k_{near} and k_{far} specify the range of the depth information behind and in front of the picture, relative to the screen width N_{pix} respectively. The viewing distance is represented by the parameter value D .

Figures 3.20 and 3.21 show the R-D performance of Orbi and Interview sequences respectively at low bitrates up to an overall bitrate of 500 Kbits/s. Both Orbi and Interview sequences demonstrate better performance for colour and depth image coding than coding projected left and right view video with scalable H.264/AVC. Due to the characteristics of depth image sequences, they can be highly compressed using H.264/AVC coding tools [110]. As a result the coded depth content requires less average bitrate compared to its corresponding colour image sequence. The coded colour and depth content can be squeezed into a given target bitrate with less degradation in colour and depth image quality, which is vital to render two high quality left and right sequences using DIBR method. In case of left and right view coding with scalable H.264/AVC, both image sequences require considerable amount of bitrate due to the availability of similar texture information in both images. Furthermore, the amount of disparity between the projected left and right images has hindered the use of adaptive inter-layer prediction more effectively between the base and enhancement layers of the proposed SVC configuration. Consequently, left and right view coding has failed to achieve improved quality at low bitrates. However, at higher bitrates left and right video coding may achieve better performance due to the possibility of squeezing good quality left and right sequences into a given bitrate.

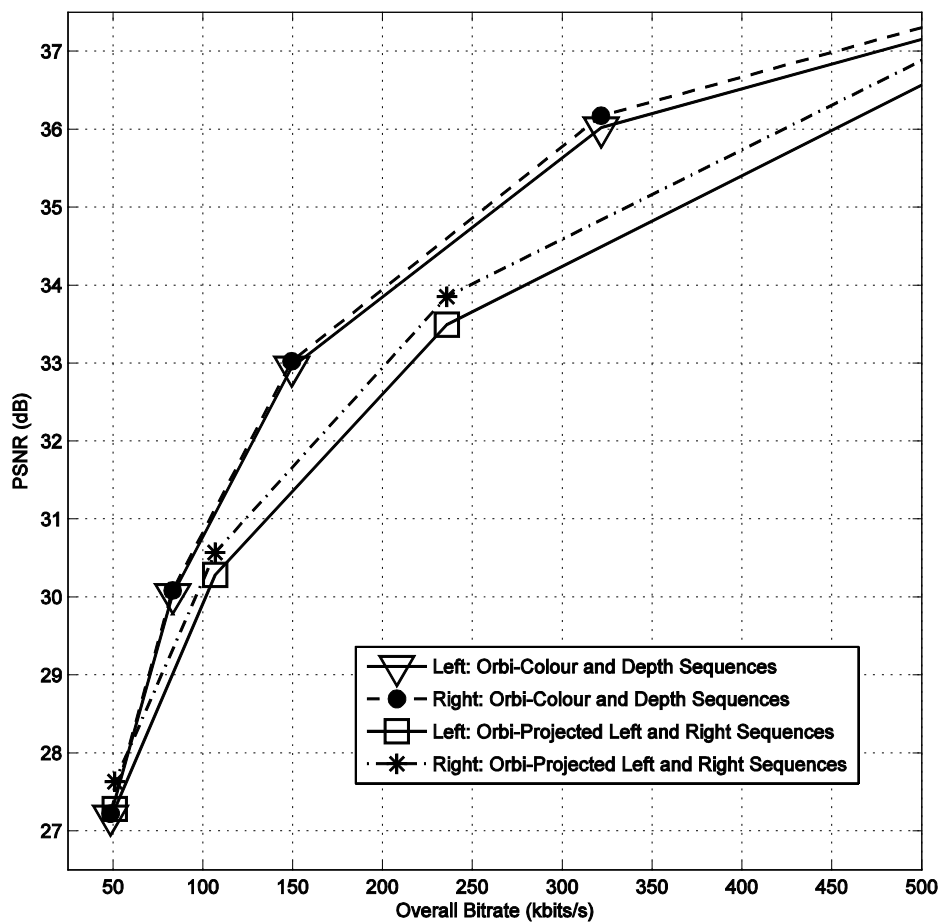


Figure 3.20: R-D curves for Orbi (using colour and depth sequences and projected left and right sequences).

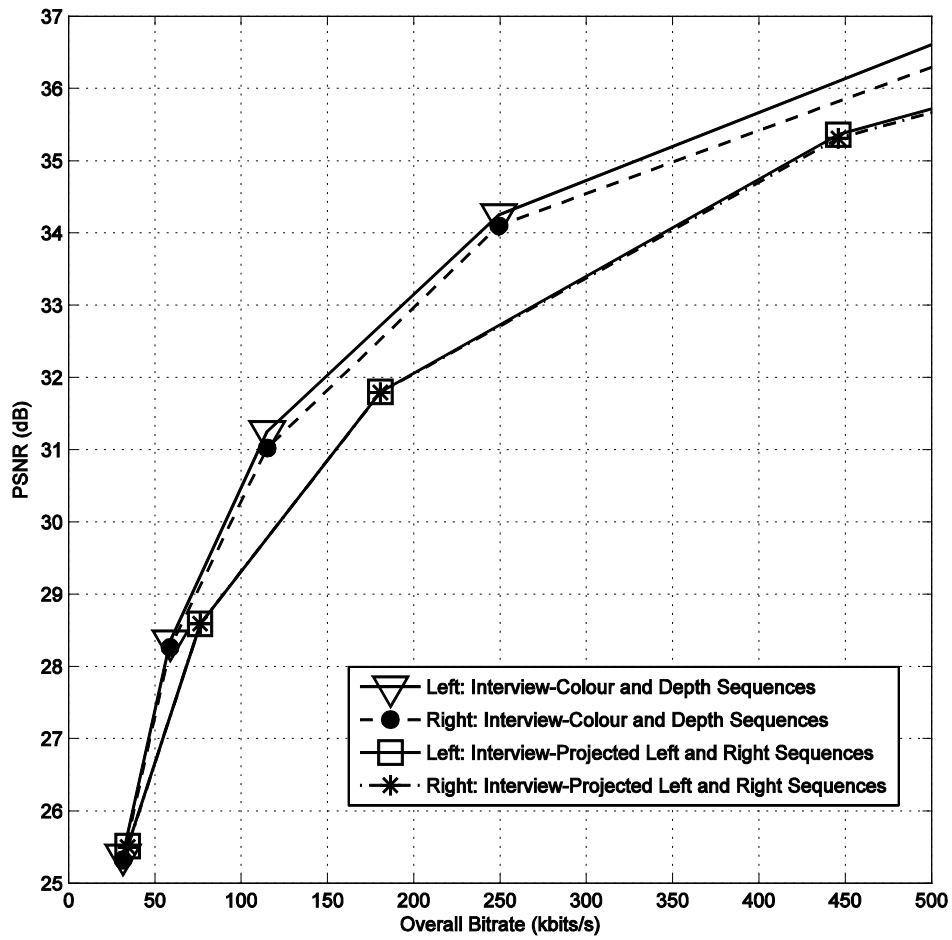


Figure 3.21: R-D curves for Interview (using colour and depth sequences and projected left and right sequences).