UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

BRUNO BOESSIO VIZZOTTO

Efficient Algorithms for Process and Communication of Multiview Videos: Contributions to Rate Control and Thread Management for 3D-Videos

Tese apresentada como requisito parcial para a obtenção do grau de Doutor em Ciência da Computação.

Prof. Dr. Sergio Bampi
Orientador

Porto Alegre
2017

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Vizzotto, Bruno Boessio

Efficient Algorithms for Process and Communication of Multiview Videos: Contributions in Rate Control and Thread Management for 3D-Videos [Tese] / Bruno Boessio Vizzotto. – 2017.

95 f.:il.

Orientador: Sergio Bampi.

Tese (Doutorado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação. Porto Alegre, BR – RS, 2017.

1.Introduction. 2.Background and Related Work 3.Multiview Encoding Analysis 4.Rate Control Algorithms for Multiview Videos 5. Power Eficient Thread Management for Multiview video encoding. 6.Conclusions I. Bampi, Sergio. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Reitor: Prof. Rui Vicente Oppermann
Vice-Reitor: Profa. Jane Fraga Tutikian
Pró-Reitor de Pós-Graduação: Prof. Celso Giannetti Loureiro Chaves
Diretor do Instituto de Informática: Profa. Carla Maria Dal Sasso Freitas
Coordenador do PPGC: Prof. João Luiz Dihl Comba
Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

A todos os colegas e amigos que contribuíram ativamente para esta tese. Ao Prof. Sergio Bampi por todo tempo dedicado a minha orientação, ao Grupo de Microeletrônica da Universidade Federal do Rio Grande do Sul pelo suporte, ao Programa de Pós-Graduação em Ciência da Computação da UFRGS pela disposição de suas excelentes estruturas e de seus distintos profissionais. Agradeço a CAPES e ao CNPq que apoiaram financeiramente a realização deste trabalho.

Aos colegas do laboratório 67/215 que contribuíram das mais variadas formas para que este trabalho fosse possível. Vagner Rosa, Bruno Zatt, André Luís Del Mestre Martins, Leonardo Bandeira Soares, Eduarda Monteiro, Daniel Palomino, Felipe Sampaio, Thaisa Leal, Dieison Silveira, Débora Matos, Cláudio Diniz, Mateus Grellert, Guilherme Paim.

Agradeço ao prof. Jörg Henkel que ofereceu seus recursos do Chair for Embedded Systems do Karlsruhe Institute of Technology. Aos colegas que me acolheram e me ajudaram das mais diversas formas, Muhammad Shafique, Muhammad Usman Karim Khan, Sajjad Hussain, Mohammad Salehi, Fazal Hameed, Jorge Castro-Godinez, Marvin Damschen, Volker Wenzel, Florian Kriebel, Martin Buchty e Gull-Nida Amjad.

Agradeço especialmente a Carmen Sax, Doris Arnitz e sua família pela excelente recepção toda gentileza e atenção em minha estadia em Etlingen.

Agradeço à Universidade Federal do Pampa, e em especial aos professores Alessandro Girardi e Marcia Cristina Cera (*in Memoriam*) pelo esforço de possibilitar meu intercâmbio. Aos meus colegas da Engenharia de Telecomunicações pelo suporte durante meu período de ausência.

Dedico este trabalho a minha família e agradeço a todo apoio recebido durante este período. Ao amor da minha vida, Daniela, e aos meus filhos Alice e João Pedro deixo este trabalho como prova do meu maior esforço. Aos meus pais Assis e Mariza, meus irmãos Cassiano e Mariana (e família - Luciano e Betina), minhas avós Célia e Carmen. A minha família em Rio Grande, Seu Carlos e Dona Sandra, Augusto, Daniel, Saulete e Valentina.

SUMMARY

ABBREVIATIONS

3DTV	*Three-dimensional Television*
AVC	*Advanced Video Coding*
BU	*Basic Unit*
BIPS	*Billion Instruction per Second*
CODEC	*Coder/Decoder*
CTU	*Coding Tree Unit*
CU	*Coding Unit*
DE	*Disparity Estimation*
DPB	*Decoded Picture Buffer*
DV	*Disparity Vector*
fps	*Frames per Second*
FRExt	*Fidelity Range Extensions*
FTV	*Free Viewpoint Television*
GOP	*Group of Pictures*
GGOP	*Group of GOP*
HBP	*Hierarchical Bi-prediction*
HEVC	*High Efficiency Video Coding*
HRC	*Hierarchical Rate Control*
HVS	*Human Visual System*
IEEE	*Institute of Electric and Electronics Engineers*
INTRA	*Intra Prediction*
ISO	*International Organization for Standardization*
ITU-T	*International Telecommunication Union – Telecommunication*
IT	*Inverse Transform*
JVT	*Joint Video Team*
KIT	*Karlsruher Institut für Technologie*
MAD	*Mean Absolute Differences*
MB	*Macroblock*
MBEE	*Mean Bit Estimation Error*
MD	*Mode Decision*

MDP	*Markov Decision Process*
ME	*Motion Estimation*
MPC	*Model Predictive Control*
MPEG	*Moving Picture Experts Group*
MV	*Motion Vector*
MVC	*Multiview Video Coding*
PID	*Proportional Integral Derivative*
PMV	*Predicted Motion Vectors*
POMDP	*Partially Observed MDP*
PSNR	*Perceptible Signal-to-Noise Ratio*
Q	*Quantization*
IQ	*Inverse Quantization*
QP	*Quantization Parameter*
RC	*Rate Control*
RD	*Rate-Distortion*
RDO	*Rate-Distortion Optimization*
RGB	*Red, Green, Blue*
RoI	*Regions of Interests*
SP	*Switching P*
SI	*Switching I*
SVR	*Support Vector Regression*
T	*Transform*
IT	*Inverse Transform*
TOPS	*Trillion Operations per Second*
UFRGS	Universidade Federal do Rio Grande do Sul
UHD	*Ultra High Definition*
VCEG	*Video Coding Experts Group*
YCbCr	*Luminance, Chrominance Blue, Chrominance Red*
YCgCo	*Luminance, Chrominance Green, Chrominance Orange*
WP	*Weighted Prediction*

LIST OF FIGURES

LIST OF TABLES

RESUMO

Esta tese propõe contribuições para o processo de codificação de vídeos de múltiplas vistas. Uma análise dos padrões atuais de codificação de vídeos de múltiplas vistas é apresentado destacando os principais desafios destes codificadores considerando comunicação e processamento. Esta tese apresenta duas contribuições. Primeiramente, técnicas de controle de taxa e ajuste de fluxo de dados são propostos nos níveis de quadros e unidades básicas, objetivando melhor precisão na saíta do bitstream do codificado enquanto entregando uma determinada qualidade visual, ao considerar as restrições impostas pelo sistema de transmissão. Técnicas preditivas no nível de quadros associadas com um algoritmo de região de interesses no nível de unidades básicas gerando aprendizagem por reforço no modelo de controle geral apresentam significativa redução na variação da taxa de bits. O modelo proposto não excede 1% de variação nos dados de saída. Ademais, a qualidade visual sofre uma perda máxima de 1,5%. Segundo, um gerenciador de threads associado a um balanceador de carga de trabalho e controle de potência para processamento de vídeos de múltiplas vistas em plataformas de múltiplos núcleos. Esta técnica aplicada a um sistema de 32 núcleos atinge até 51% de economia no consumo de energia com uma degradação visual na qualidade do vídeo de até 2% se comparada ao software de referência.

Palavras-chaves: Codificação de Vídeo Digital, Vídeos de Múltiplas Vistas, Controle de Taxa, Balanceamento de Workload, Gerenciamento de Threads.

ABSTRACT

This thesis proposes contributions for the encoding process of multiview videos. Analysis of current multi-view video coding standards is presented, aiming to understand the key challenges of these encoders considering communication and processing. This thesis presents two contributions. Firstly, techniques of rate control and data flow adjustment are proposed in the frame and basic unit levels, targeting best accuracy in the output bitstream of the encoder while delivering the desired video quality, considering the restrictions imposed by the transmission system. The predictive techniques at frame level associated with the regions of interest algorithm at the basic unit level to generate a reinforcement of learning in the overall control model present a significant reduction in the bitrate variations. The proposed model does not exceed 1% of the variation in the output data. Also, the visual quality suffered a maximum loss of 1.5%. Second, a thread management associated with workload balancing and power control for multi-view video processing on multi-core platforms. The results obtained by the proposed techniques show that the thread management jointly with coding adjustments allows a significant reduction in complexity. This technique applied to a 32-core system reached up to 51% saving in energy consumption with up to 2% degradation in the visual quality of the video compared to the reference software.

Keywords: Digital Video Coding, Multiview Videos, Rate Control, Workload Balance, Threads Management.

1 INTRODUCTION

Multiview video processing has become the key point to the development of 3D-video technology. The new 3D-video technology became widely applied in cinema, television, games, smartphone, etc. To obtain the best coding rates on the new processing technologies and in turn to get a satisfactory final result the consumer expectations, multi-view coding standards such as Multiview Video Coding (MVC) (JVT, 2008) and 3D High Efficiency Video Coding (3D-HEVC) (Müller, et al., 2013) have emerged. However, these new standards and their extensions have brought a high computational cost associated with a high energy demand. Adapting the new standards to new processing technologies is a key point to power constrained devices. The key challenges that arises are: How to provide high and smooth visual quality while considering restricted bandwidth? How to adapt new multiview video standards to multiple core platform and appropriately take advantage of its potential? To address these questions are the main objective of this thesis.

This thesis proposes the development of techniques in algorithms to reach high and smooth visual quality using the multiview encoder that transmits its bitstream under a constrained bandwidth. Moreover, this thesis presents a workload balance scheme and thread management for multiview videos encoded over multiple core platform. This work targets the overall multiview process over multiple core system while the focus in the rate control algorithm-level since it operates to adjust the bit distribution over the levels in the multiview encoder. There is no work in the literature using thread management considering multiview video content by adapt to the multi-core platform. Considering the Rate Control perspective, there are different works for specific scenarios as well for specific levels. Considering multiview videos, the disparity estimation (DE) prediction tool complete changed the situation. Although the motion and disparity are conceptually similar, the disparity presents an entirely distinct behavior regarding average vector length, ideal search pattern shape, local minima and computational effort. Additionally, by considering both motion and disparity data, the amount of information and correlation in the different domains are much vast.

The goal of this thesis is to address the above-discussed challenges through research and development of power-efficient algorithms and system level techniques for low-

power multiview video processing, while at the same time avoiding a significant loss in the video quality. This thesis shows the previous results of investigating distributed, scalable, and adaptive resource and power management techniques at both algorithm and system level in multi-core systems while accounting for the knowledge of emerging multiview video processing algorithms and multiview video content to enable power-/energy-efficient management of processor computational and memory resources.

The Chapter 2 of this thesis presents the basic concepts of digital video coding, the novel standards for video coding and its Multiview component tools. It also describes the module of primary interest to this thesis, the rate control, followed by a revision of the related work. Moreover, Chapter 3 describes the detailed studies performed to characterize and quantify the possible correlation of the multiview video content. This evaluation is key for the development of an efficient algorithm for Rate Control and the thread management for multiview video processing in multiple core processors. Chapter 4 describes the proposed algorithms and schemes for the Rate Control and presents two main topics: rate control techniques at the frame level and bit allocation at the basic unit level. Then it is shown the results and comparisons of multi-granularity Rate Control methods based on Model Predictive Control and Markov Decision Process. Chapter 5 proposes thread management with workload adaptation for multiview video encoding over multiple core processing platform and its results. Finally, Chapter 6 presents the conclusions and plans for future works related to these topics. In the end, the list of bibliographic references.

2 BACKGROUND AND RELATED WORK

In this chapter, the basic notions about digital video coding, multiview video and the evolution of Multiview Video standards are presented. Moreover, it is presented the basic of rate control module. The workload balance and thread management aspects for multiview videos are detailed since they are the main focus of this thesis. Finally, a brief introduction to the state-of-the-art of each topic is also presented and discussed.

2.1 Basics of Video Coding

A sequence of pictures (or frames) of a scene captured at a given time produces a basic single view video, denoting a frame rate providing to the viewer the sensation of motion. The frame rate can vary in frames per second (fps) depending on the predefined requirements. The picture is a compact by a given number of dots known as picture elements, i.e. pixels (Richardson, 2010). The video resolution, is the number of pixels in each frame, i.e. the number of horizontal and vertical pixel lines. The resolutions depend typically on the target application. Some examples of applications include mobile devices that handle with low resolution and low frame rate sequences (as 480p at 24 fps) while home cinema targets higher resolution and higher frame rates (as HD1080p at 60 fps) (Pourazad, et al., 2009).

A different representation of color spaces is used to represent raw or encoded videos, while the most usual are the RGB (Red, Green, Blue) and YUV standard (Sullivan, et al., 2005). The reason is that most of the regular computer monitors operate at the RGB space while most of video coding standards work over the YUV space. As the RGB, three channels composes the YUV space: one channel dedicated to luminance (Y) and two chrominance channels (U and V). The reason for adopting the YUV color space for video coding is its smaller correlation between color channels making it easier the coding of these channels independently. Considering that the Human Visual System (HVS) is more sensitive to luminance when compared to chrominance, so, it is possible to reduce the amount of information in the chroma channel with reduced effect on the overall perception. The reduction of chroma information is performed by using a technique of sub-sampling (also known as pixel decimation) (Sullivan, et al., 2005). The most popular

color sub-sampling in video encoding is the pattern YUV 4:2:0 that stores one U and one V sample for each four luminance samples, reducing in half the total amount of raw video data (Richardson, 2010).

Block coding is the base of all of the current well-known video encoding standards. In other words, they just divide each frame into pixel blocks to encode the video in minor pieces (Zatt, et al., 2007). These blocks are named accordingly to the standard. In the H.264, macroblock (MB), while in the state-of-the-art HEVC it is called coding unit (CU). In this thesis, basic unit (BU). The H.264 standard uses MBs with blocks of 16x16 luma pixels and the associated chroma samples (see Figure 2.1). An encoded group of this block is called slice (Wiegand, et al., 2003). One or more MBs, contiguous or not, forms the slice. In the same way, one or more slice of the same type forms a frame. In turn, each slice can be classified mostly in one of three different types: Intra (I), Predictive (P) and Bi-predictive (B) slices. Figure 2.1 is composed of three slices being one contiguous (Slice 0) and two noncontiguous slices (Slices 1 and 2). Note that the terminology used here is based on the H.264 standard and is directly applicable to the Multiview Video Coding standard either (Richardson, 2010) (JVT, 2009).

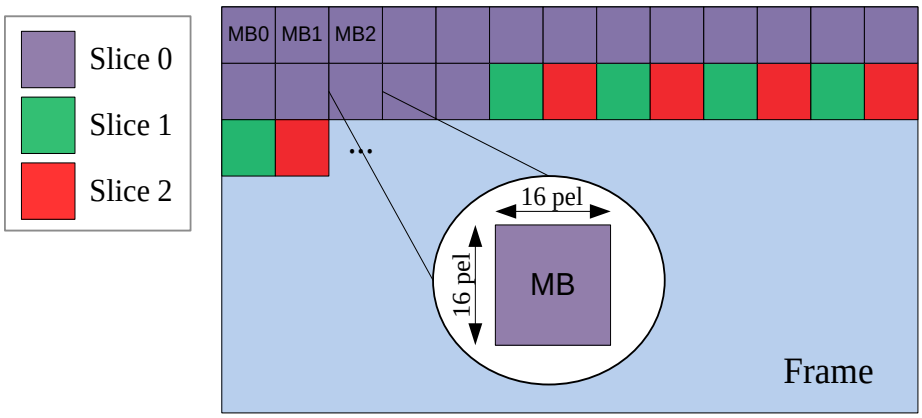

Figure 2.1: Macroblocks and slices possible distribution in a frame.

Source: (Zatt, et al., 2010)

To understand the difference in the slice types it is necessary to consider the two primary prediction modes that were firstly defined in the H.264 standard and adopted by the state-of-the-art video encoder (HEVC). The intra-frame prediction only exploits the spatial redundancy by using neighboring pixels to predict the current MB (Wiegand, et al., 2003). The inter-frame prediction uses the similarity between different frames by using areas from other frames, called reference frames, to better predict the current MB. Intra (I) macroblocks use the intra-frames prediction information while predictive (P),

and bi-predictive (B) macroblocks use the inter-frame prediction. Particularly, while P macroblocks only use past frames as a reference (in the time of the video sequence), the B macroblocks can use frames from the past, future (due to IBP structure) or both as references. The intra slices are formed only by I MBs. Predictive slices support I and P macroblocks, and Bi-predictive slices support I and B macroblocks as observed by (Richardson, 2010) and (JVT, 2009).

2.2 Multiview Video Coding Properties

The multiview video sequence is composed of a finite number of single view video sequences captured from independent cameras in the same 3D scene (Merkle, et al., 2007), even they are arranged in a particular set up, and this scheme information is required to provide the best encoding results. In this way, usually, these cameras are carefully calibrated, synchronized and positioned. Typically they are in a parallel array of one or two dimensions. However, there are systems where the cameras are placed in arch or cross shapes (Kauff, et al., 2007). The spacing between camera is typically 5cm, 10cm or 20cm for most of the available test sequences (Su, et al., 2006). In Figure 2.2 a multiview video with four views and the captured frames along the time axis are presented.

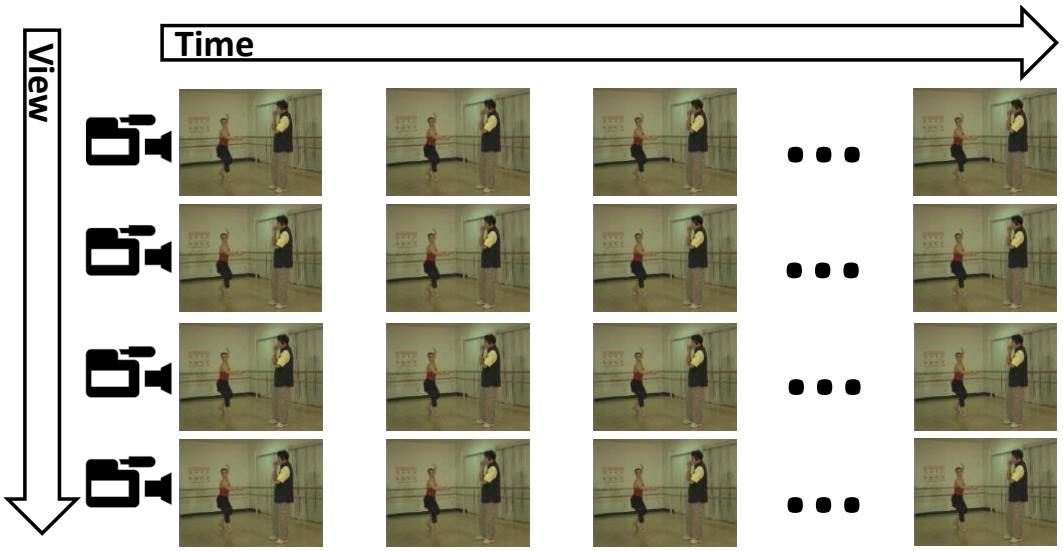

Figure 2.2: Multiview video sequence
Source: Modified from (Zatt, et al., 2010)

Figure 2.3 presents a complete multiview system required to capture, encode, transmit, decode and display the multiview videos (Chen, et al., 2009). The sequence is first captured and encoded by an MVC encoder firstly encodes the obtained sequence to decrease the amount of data to be stored or transmitted. Then, the generated bitstream it

shall be forwarded by using broadcast, internet or can be stored in a media servers or local storage. At the decoder side, the bitstream, or part of it, is decoded and displayed according to the displaying technology available at the receiver end. If the target is a simple single view display, the decoder will consider only the base view that is decodable with a regular Advanced Video Coding (AVC) video decoder. If the application is a Free Viewpoint Television (FTV) system, the user selects the desired viewpoint within the 3D scene, and the video decoder selects which view to decoding. For multiview displays, all the views displayed are decoded added by the views used to reconstruct them.

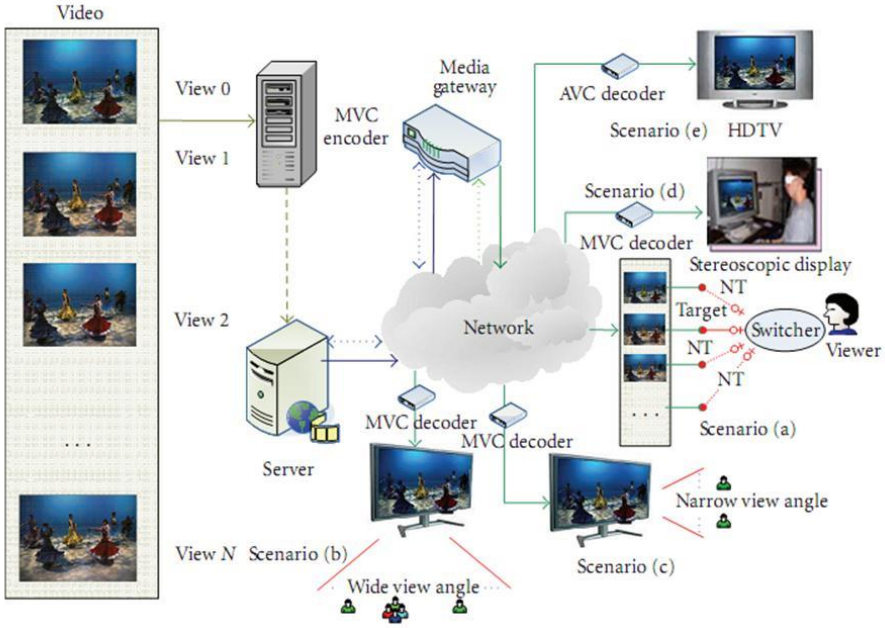

Figure 2.3: Multiview video capture, (de)coding, transmission and display system
Source: (Chen, et al., 2009)

The most important scenarios to display a multiview stream are the Three Dimensional Television (3DTV) and the FTV. Three Dimensional Television is the extension to the traditional 2D with the viewer depth sensation made possible (Smolic, et al., 2007). In this kind of application, multiple views are decoded and displayed simultaneously. The simplest 3D displays are stereoscopic that show two simultaneous views but usually require the use of a special glass to provide 3D sensation. In another hand, the evolution is the auto-stereoscopic displays that eliminate the need for glasses. Real multiview displays can decode and show a higher number of views at the same time increasing the observer freedom allowing head parallax (i.e. the viewpoint changes when the observer changes its position) (Pourazad, et al., 2009). This system provides more realism and

interactivity to the user. The display technology may vary from 2D televisions to multiview displays.

The encoding process can use different techniques. The most primitive are the so-called simulcast where a single view video coding standard is used to encode each view independently. Figure 2.4 shows the simulcast approach that considers the intra-frame prediction and inter-frame prediction (motion estimation - ME) exploiting the spatial and temporal redundancy, in the meantime, not considering the disparity redundancy (the redundancy between frames of different views). Multiview encoders as MVC uses the inter-view prediction (disparity estimation) to obtain an advantage of the similarities between these views from the same scene. The inter-view prediction represented by the red arrows in Figure 2.4 are responsible for a bitstream reduction of 20-50% for the same video quality (Müller, et al., 2013). More details on the multiview tools of the former state-of-the-art MVC, coding efficiency and complexity are discussed in the next section.

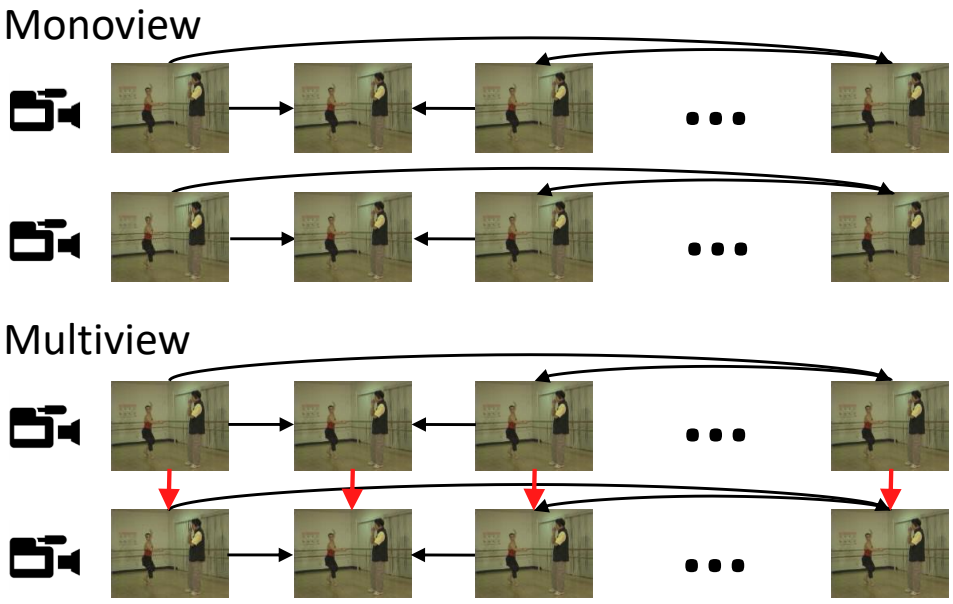

Figure 2.4: Prediction comparison between monoview (Simulcast) and multiview
Source: Modified from (Zatt, et al., 2010)

2.3 Multiview Video Standardization

Considering the first multiview standard, in a strict definition the Multiview Video Coding is not a coding standard but an extension to the H.264/AVC or MPEG-4 Part 10 (JVT, 2003). The MVC was defined by the Joint Video Team (JVT) in March 2009 (JVT, 2009). The JVT is the group of experts formed by the Motion Picture Experts Group (MPEG) from ISO/IEC and the Video Coding Experts Group (VCEG) from ITU-T.

The H.264 and the HEVC standard works over the YCbCr color space. The color spaces are also used in different subsampling patterns. Including 4:2:0 (four luminance samples for one sample of each chrominance channel), 4:2:2 (two luminance samples for one sample per chrominance channel) and 4:4:4 (one luminance channel for one sample in each chrominance channel). The supported combinations and a set of tools supported depend on the profile of video coding operation (JVT, 2009).

The first published version of H.264 defines three profiles: Baseline, Main and Extended. The first one, the Baseline profile focus on video calls and video conferencing supporting only I and P slice and the CAVLC entropy coding method aiming simpler coding for power constrained devices. The most popular, the Main profile was developed to provide tools for the high definition displaying and video broadcasting. Besides the tools defined by the Baseline profile, it also includes the support of B slices, interlaced videos, and the most complex CABAC entropy coding. The Extended profile targets video that streaming on channels with high package loss and defines the SI (Switching I) and SP (Switching P) slices (Richardson, 2010). Finally, the Fidelity Range Extension (FRExt) set the High profiles: High, High 10 (in which, uses 10 bits per Y, Cb or Cr sample), High 4:2:2 and High 4:4:4 targeting high fidelity videos (JVT, 2009). The state-of-the-art standard HEVC adopts these set of definitions with reviews.

The extension of multiview video introduced to the H.264 standard a new set of SEI (Supplemental Enhancement Information) messages to simplify parallel decoding and the transmission of sequence parameters (JVT, 2009). Additionally, proposing the disparity estimation or inter-view prediction (Merkle, et al., 2007). The disparity estimation is the most important innovation in the MVC extension that allows the exploration of similarities between different views. The focus of this extension is to find the best matching for the current macroblock in a reference frame within the reference view. The search criteria, search patterns, and objective are similar to the motion estimation. However, the behavior of the disparity estimation differs significantly (depicted in Chapter 4).

The bit rate required for coding multiview video with the MVC extension of H.264/AVC increases linearly with the number of encoded views. Therefore, MVC is not appropriate for delivering 3D content for autostereoscopic displays, due to a lack of depth

map processing. Hence, the HEVC standard applies depth maps to de new 3D-HEVC standard.

2.3.1 Multiview Encoding Process in the HEVC

The HEVC standard targets to process Ultra High Definition (UHD) 3840 × 2160 pixels videos. 3D video encoding of ultrahigh resolution videos requires an enormous amount of computational power and memory. The raw video data of these 3D videos ranges from 1Gbps to 15Tbps. Besides this, higher pixel data representations from 8 to 32 bits pixel to delivery high-dynamic range videos. Moreover, larger image or video frame rates are required for such different application areas like 30 to 60 fps for automotive and security while 60 to 120 fps for medical imaging. Finally, a higher number of views like 2 to 6 in automotive, 8 to 16 in medical teleoperation theaters, more than a hundred in football/Olympics stadium lead to a massive data processing requirement.

Depending on the application scenario, video with many views and resolution, 3D video encoding requires a significant computational power of several thousand of Billion Instruction per Second (BIPS) up to a hundred of Tera Operations per Second (TOPS) from the underlying platform. This power/processing requirement was increased more than five times compared to the previous MVC standard (JVT, 2009).

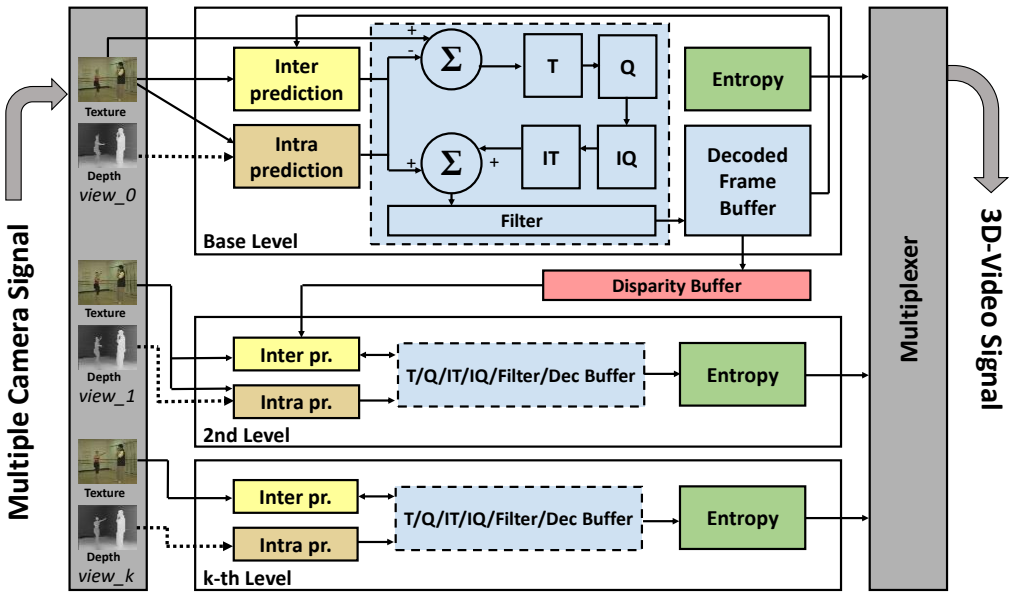

Figure 2.5: 3D-HEVC encoder block diagram

Figure 2.5 shows the high-level block diagram of the 3D-HEVC encoding process. Like all hybrid standards, three phases composes the coding process: prediction,

transforms, and entropy coding. The transform and entropy phases are similar to H.264/AVC, except for the new coding units (CU) to be encoded by the entropy encoder. The main difference from simulcast HEVC is the legacy from MVC where the prediction phase that incorporates the inter-view prediction.

The base view, the first one to be encoded, is encoded in compliance to the HEVC standard. So, the prediction has two options, the intra-frame or the inter-frame prediction. The complete encoding process is described in this section considering the Main profile tools in YCbCr color space with 4:2:0 sub-sampling, while further extensions available in the High profiles omitted for simplicity.

2.4 Rate Control Process

Usually, multiview video sequences are captured using a high sample rate, over 30 fps, to improve the motion flow and give the observer a smoother motion sensation. The high frame rate applied implies in a high redundancy or similarity between neighbor frames in the time (and also disparity) axis. As noticed in Figure 2.6, frames S0T0 and S0T1 are very similar. Therefore, only the differences between them have to be transmitted.

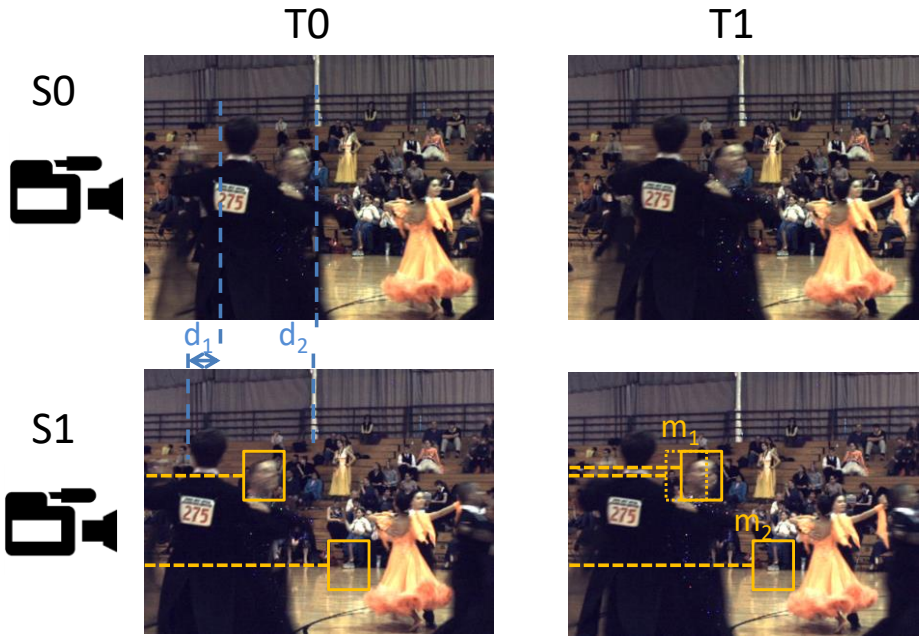

Figure 2.6: Temporal and disparity similarities
Source: Modified from (Zatt, et al., 2010)

The algorithm that exploits this inter-frame similarity is the motion estimation (ME). It searches in the temporal neighbor frames, known as a reference, the region that represents the best match for the current block or macroblock (Zatt, et al., 2010). Once the best matching block is found, a vector pointing to that position, the motion vector

(MV) is generated. Consider, for example, a background region (one of the yellow boxes in Figure 2.6); there is no motion between T0 and T1, so the motion vector m2 is zero. The dancers moving (woman's face in the yellow box) present a displacement along the time, this movement is represented by m1. The set of motion vectors of a given frame is called motion field and represent valuable information to understand the motion of an object as time progresses.

Despite the high coding efficiency provided by MVC, the transmission and storage of 3D-videos remain a big challenge, especially for services operating over bandwidth/buffer-constrained infrastructures. It becomes even more challenging due to changing input video properties, run-time variations on video encoder state, battery level and user preferences. Thus, to provide high video quality while meeting channel bandwidth/buffering constraints, it is necessary to further optimize the bandwidth usage by intelligently regulating the bits allocation. Therefore, a Rate Control (RC) block is implemented to dynamically find a good compromise between the coding efficiency and video quality by adapting the Quantization Parameter (QP).

The Rate Control regulates the output coded bitstream to produce high video quality at a given target bitrate (Li, et al., 2003). An efficient RC scheme must be able to provide increased video quality for a given target bitrate with smooth visual quality variation along the time, for different views and within the frames. Also, the RC should keep the bitrate as close as possible to the target bitrate (optimizing the bandwidth usage) while avoiding sudden bitrate variations.

2.4.1 Rate Control Related Work

There are several Rate Control schemes found in the current literature. Most of them developed targeting single-view encoders such as H.264 or HEVC. Recently, a few works peculiar to the multiview video have been proposed focusing on frame and BU level RC. In this section, it is present an overview of the state-of-the-art on Rate Control.

2.4.1.1 Single-View

Considering the single-view domain, the majority of proposals are extensions to the RC implemented in the H.264 reference software that employs a quadratic model for MAD (Mean Absolute Differences) distortion prediction (Li, et al., 2003). However, the quadratic model leads to limited control performance, as discussed in (Tian, et al., 2010). Aware of this limitation, the authors in (Jiang, et al., 2004) and (Merrit, et al., 2007)

propose improved MAD prediction techniques. The scheme presented in (Kwon, et al., 2007) implements both distortion and rate prediction models while in (Ma, et al., 2005) the RC exploits rate-distortion optimization models. A RC based on a PID (proportional–integral–derivative) feedback controller is presented in (Zhou, et al., 2011). A RC scheme for encoding H.264 traffic surveillance videos is proposed in (Wu, et al., 2009), which uses RoI to highlight regions containing significant information. In (Agrafiotis, et al., 2006), the technique is used to highlight preset regions of interests (RoI) using priority levels. However, single-view approaches do not fully consider the correlation available in the spatial, temporal and view domains and, consequently, cannot effectively predict the bit allocation or distortion resulting in inefficient RC performance. Finally, in (Zhang, et al., 2017) a new R-D model is proposed by classifying blocks into different depth and ROI groups. Moreover, a machine learning approach is applied to enhance the accuracy of the distortion estimation.

2.4.1.2 Multiview

The majority of RC proposals targeting multiview video are based on a simple extension of single-view approaches (Li, et al., 2003) and are still unable to exploit multiview properties entirely. Novel solutions, however, have been proposed and most of them are limited to frame level. In (Yan, et al., 2009) and (Yan I, et al., 2009) the information from previously encoded frames is used to predict the current frame QP accurately. But these are clear solutions unable to adequately handle the complex Hierarchical Bi-Prediction (HBP) structure of the multiview video, limiting the number of input samples and the Rate Control learning. The scheme in (Xu, et al., 2011) considers a single fixed HBP structure and does not consider the inter-Group of Picture (GOP) correlation. These limitations at frame level are partially addressed in (Vizzotto, et al., 2012) and extended to deal with any possible HBP structure in (Vizzotto, et al., 2013).

In (Su, et al., 2014) it is proposed a dynamic adaptive rate control system and its associated rate-distortion model for the High-Efficiency Video Coding (HEVC) multiview video. The work presents evaluation in Rate of Quality of Experience model over a subjective test. The (Lie, et al., 2014) paper presents a new 3D video encoding system featuring 3D quality optimization and joint rate control between color and depth components. The Support Vector Regression (SVR) model presented in the paper show relevance with results considering specific "IPPP..." structures (Song, et al., 2016) proposes an improved Largest Coding Unit (LCU) level rate control algorithm for the 3D

video. The proposed algorithm includes bit allocation for extended views in combination with temporal and inter-view MAD predictions. Finally, (Tan, et al., 2017) presents an inter-view dependency-based rate control (RC) algorithm for 3D-HEVC. The authors present a complete scheme of rate control restricted for frame level, including bit allocation, quantization parameter, and depth level.

As discussed, to deal with distinct image regions within a frame there is a need for a BU/CU-level Rate Control. Moreover, to find an optimal global solution a common frame- and BU-level Rate Control scheme must be designed. Recent works have proposed solutions for the BU-level RC in MVC and CU in 3D-HVC. In (Park, et al., 2009) another extension to the quadratic model (Li, et al., 2003) is proposed using the classic MAD prediction to set the QP value and, consequently, falling in the problem of not considering MVC/3D-HEVC view domain. The authors of (Lee, et al., 2011) use the concept of RoI, based on the Just-noticeable difference (Liu, et al., 2010), to determine relevant regions and allocate more bits to them. However, this solution does not employ feedback-based control and just considers the coding information from the reference frame.

To cover the gap between frame-level and BU-level Rate Control and address the limitations inherent to the state-of-the-art solutions it is required a dynamically adaptive Rate Control Scheme able to jointly consider all Rate Control actuation levels to provide optimized bandwidth usage (bitrate allocation) and similar video quality in spatial, temporal and view domains.

In this thesis, it is inherited the proposed Hierarchical Rate Control (HRC) for Multiview Video Coding (Vizzotto, et al., 2013) that employs a joint solution for the multiple levels of Rate Control. The proposed HRC uses a Model Predictive Control (MPC)-based Rate Control that jointly considers GOP-level and frame-level stimuli to accurately predict the bit allocation and define an optimal control action at coarse-grain. To further optimize the bit allocation within the frames the HRC implements a Markov Decision Process (MDP) to refine the control action at BU-level taking into consideration image properties to define and prioritize Regions of Interest (RoI). Finally, novel Reinforcement Learning techniques are used to feedback MPC and to update the MDP states transitions probabilities.

- **MPC-based frame-level Rate Control**: It is responsible for predicting the bitrate allocation and defining an optimal QP value for the current frame while minimizing

a performance cost function. The proposed MPC-based RC deals with multiple stimuli superposition building the input horizon using previously encoded frames from temporal and view neighborhood. The proposed scheme also incorporates the GOP-phase for accurate bitrate prediction.

- **MDP-based Basic Unit-level Rate Control**: The BU-level RC receives the QP defined at frame level and adjusts the QP for each BU. The proposed Markov Decision Process-based RC takes the decisions over a map of states based on a set of possible actions (QP adaptations) and the associated rewards. The map of states is linked to the texture-based map of Regions of Interest and provides the structure to make decisions.
- **Coupled Reinforcement Learning**: It is responsible for adapting MPC and MDP models to the dynamic system behavior. After an action is taken at BU-level, the RL reads the system response and, updates the transition probabilities and the associated rewards in the MDP model. Once the frame is fully encoded, the resulting map of states is used to update the fame-level MPC. This strategy integrates frame-level and BU-level guaranteeing consistency and avoiding modeling mismatches.

Summarizing, the available Rate Control techniques do not fully exploit the correlation potential available in the spatial, temporal and view domains of MVC and 3D-HEVC. Also, they are unable to adapt to multiple HBP structure and cannot employ the inter-GOP periodic behavior for RC optimization.

2.5 Workload Balance and Thread Management in HEVC

The reference software for 3D-HEVC video encoder (3D-HEVC-Software) adopted the *wavefront* based solution (Zhang, et al., 2014) developed to improve the parallelism in HEVC. However, this approach does not consider the disparity and the scalability of 3D video content.

2.5.1 Related Work for Workload Balance

Recently, state-of-the-art works looked into the workload balancing and complexity management problems for single view video coding. In (Correa, et al., 2011), the authors propose a complexity control addressing power-constrained devices. In (Shafique, et al., 2010) an adaptive complexity reduction scheme using mode exclusion is presented for earlier H264/AVC. Therefore, in (Khan, et al., 2013) the authors present a collaborative complexity reduction targeting the Intra-mode of HEVC. In (Sanchez, et al., 2014) the

authors present complexity reduction considering depth maps of Intra mode 3D-HEVC. Finally, in (Kang, et al., 2014) the authors present a low complexity scheme considering the disparity modes of 3D content by exploiting neighbor blocks. However, this work does not present any concern with thread allocation nor power efficiency. So, to fully utilize the underlying resources (for power-efficiency), the challenge is to balance the workload of 3D-HEVC among every core by intelligently regulating the allocation of processing jobs to the associated threads. Therefore, a thread manager is required, which should dynamically find a good tradeoff between the coding/power efficiency and the video quality by adapting the resource allocation. This problem has not yet been addressed by state-of-the-art works towards parallel 3D-HEVC systems.

This work aims to reach high throughput using parallelized multiview video encoder on a multi-core framework while upgrading the power utilization of the framework. It is introduced a thread management plan to adaptively disseminate the workload of 3D-HEVC as specific occupations among parallel threads. The objective is to adjust the workload and likewise tune the voltage-frequency of the cores, considering that the end goal of the power utilization of the multi-core framework is limited. Furthermore, it is employed application-aware with content-aware complexity management scheme that adaptively tunes the application's parameters at runtime. In summary, it is proposed a Workload Balanced Thread Management that is a technique to workload balancing and dispatch encoding jobs to respective threads, such that the application's throughput requirements are met. Moreover, a Run-time Power Manager works to optimize the voltage-frequency levels of each core individually in the multi-core system, this action it is enough to provide the workload allocated to an individual core.

Moreover, an efficient resource allocation scheme for 3D-HEVC must be power-efficient while delivering smooth visual quality. Since the workload of 3D-HEVC is considerable, this system must exploit the application-specific properties (e.g., tuning application's configurations like quantization levels) to achieve maximum power efficiency while meeting given constraints (i.e. target bitrate). However, the challenge remains to find the application's knobs which impact the power (or performance) of the 3D-HEVC video encoder the most, appropriate configuration setting and addressing the workload variations associated with these knobs.

3 MULTIVIEW ENCODING ANALYSIS

This chapter presents an overall analysis of the multiview encoding process while observe and evaluate the high related correlation available in the 3D-neighborhood of a multiview video sequence. The 3D-neighborhood is defined by the three following domains: spatial, temporal and disparity (or view). For each of the domain, it is evaluated the correlation of a set of macroblocks in relation to the current one.

3.1 Multiview Video Characteristics and Observations

In observations presented by (Zatt, et al., 2010) it was noticed that the same objects present in a 3D scene are typically spotted in different views (except for occlusions). Figure 3.1 presents different views in S0 to S7 of a multiview video sequence while T0 to T8 are the temporal frames for each view and I, P and B are the types of each frame: intra, predictive and bi-predictive. Moreover, the motion sense spotted in one view is directly related to the motion perceived in the neighboring views (Deng, et al., 2009).

Considering the pattern of 3D videos with parallel cameras, the motion field is similar in these views (Kim, et al., 2007). In the same way, the disparity of some object perceived in more than one camera remains the same for different time instances where motion occurs. Furthermore, for other kinds of motion the disparity is highly correlated. The same observation has been carried out by (Kim, et al., 2007) (Deng, et al., 2009) (Shen, et al., 2010). The next sections discuss and detail the three correlation domains available in a multiview video sequence.

3.1.1 Spatial Domain

The spatial domain associated to Intra is a correlation that is related to the similarity within a frame/picture. The previous picture and video coding standards such as JPEG2000 and H.263 were able to exploit the pixel similarity between the neighbor blocks in the image. This kind of behavior is a consequence of the fact that neighbor MBs tends to belong to the same image object or regions and, consequently, present the same video properties. Some exception occurs, one of them happen in object borders were the image properties may change abruptly. Using the example in the previous Figure 3.1, all the MBs in the white background share the same video properties. The same happens for

the MBs within one of the objects. The discontinuity happens when an object border is found.

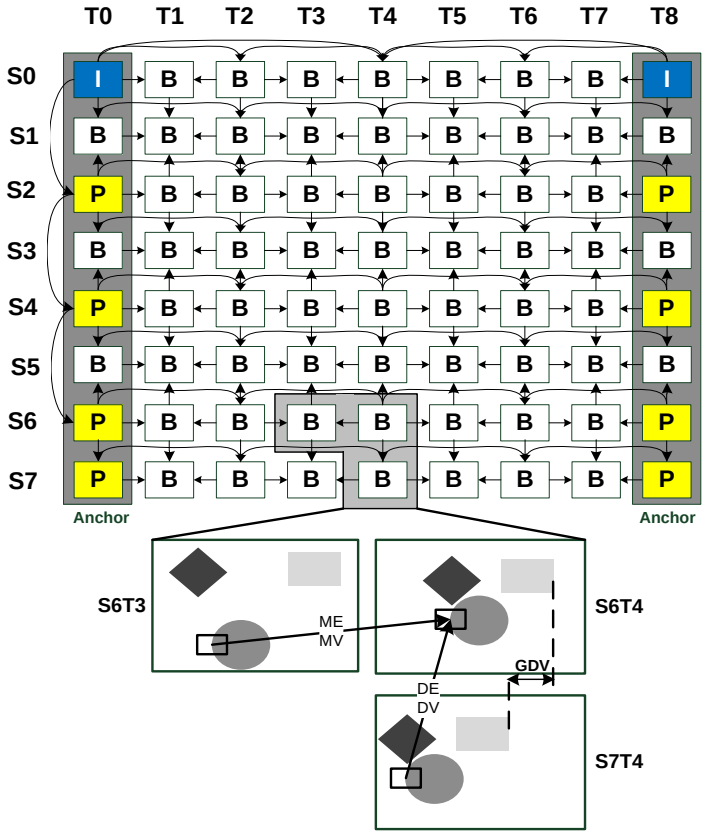

Figure 3.1: Multiview coding structure with detailed Motion and Disparity Vectors representation.
Source: (Zatt, et al., 2010)

In this evaluation, the interest is not to exploit the pixel-level correlation but the correlation of coding information as a coding mode, motion vectors, and disparity vectors. In the H.264 standard, some techniques try to exploit this kind of correlation, the differential coding of intra prediction modes inside a macroblock is a good example. These techniques corroborate to the observations that there is a high correlation between MBs in the neighborhood. It is important to note that for the focus of this analysis it is referred to the spatial correlation as one dimension, but it is composed of two dimensions, the width, and height of a picture.

3.1.2 Temporal Domain

The temporal correlation represents the similarities between different frames in the same view of a video sequence. Since the usual sample rate for standard and high definition videos is equal or higher than 30 frames per second, the objects of a given

frame are usually present in neighbor temporal frames with a displacement that depends on its motion. Consider the frames S6T3 (view 6, time 3) and S6T4 (view 6, time 4) in Figure 3.1; the same objects seem in both frames with a small displacement. Each object has the same image properties along the time. Thus the coding information should be similar. In other words, for the same object at the time, the same set of coding methods and motion intensity tend to be used. The correlation is lost when there is an occlusion, or the object moves out of the captured scene.

In the same way to the spatial correlation, there are tools enabled to exploit the temporal correlation at the pixel level, i.e. the motion estimation. At the coding information level, an attempt to use this correlation was proposed in the H.264 standard by using the direct prediction for motion vectors. This prediction used the collocated MB motion vector to predict the current one.

3.1.3 Disparity Domain

The multiview video introduced the whole new domain: disparity. This domain refers to the similarities between frames in different views. The redundancies at the pixel level, are exploited by the disparity estimation tool. However, no tool can exploit this correlation at the coding information level.

As shown in Figure 3.1 in the frames S6T4 (view 6, time 4) and S7T4 (view 7, time 4), the same objects are present in the neighbor views displaced by the disparity vector. Since they are the same objects, the same image properties are shared, and the same coding information tends to be used in different views. Moreover, the disparity correlation in a neighborhood is lost when a given object is out of the area captured by a given camera, or there is an object occlusion for a given camera point of view.

Aiming to obtain an accurate evaluation of the possible correlation for the primary purpose of this work, an extensive analysis of multiview videos were performed. For this analysis, it is used different multiview video sequences following the MVC test recommendation by JVT (Su, et al., 2006). These sequences have the coding structures similar to the one presented in Figure 3.1. In the next section, the multiview video encoding process analysis is performed for both Rate Control and Workload Distribution focus.

3.2 Multiview Video Encoding Process Analysis

To analyze the complexity of multiview encoding, this work focus on the processing analysis, using the reference software for the well know standards. For efficient compression, the software offers multiple ways of encoding a macroblock. These include a choice of different macroblock partition sizes, prediction directions, reference frames and search window sizes. In the standard implementation, all the possible options are exhaustively checked, and the ones resulting in the lowest rate-distortion cost are finally selected.

The large correlation space in the multiview to be evaluated and potentially used to improve the efficiency encoding process over multi-core platform and the smooth visual quality by using schemes in Rate Control, it is presented a detailed analysis of this correlation.

A visual evaluation is performed to detect the typical coding behavior for the different block of coding units characteristics considering video properties, coding variables and neighborhood properties in the three correlation domains: spatial, temporal and disparity. This visual evaluation will be the input to the statistical analysis that quantifies the correlation within the neighborhood using video properties as additional information to smartly predict the coding properties of the current block of coding. The statistical analysis must consider different video sequences, quantization parameters (QPs) and encoder configuration scenarios to provide data for accurate decision-making at the algorithm design phase.

3.2.1 Rate Control Analysis

In this section, it is presented a detailed bitrate distribution analysis to provide a better understanding towards the bitrate distribution during the MVC encoding process and its correlation with spatial, temporal and view neighborhood. The analysis is presented in a top-down approach starting with the view-level related discussion, following to frame-level and concluding with BU-level considerations. In this way, it is used eight views of the “flamenco2” VGA video sequence encoded at a fixed QP, that is, without Rate Control, for an IBP view coding order (0-2-1-4-3-6-5-7) and Hierarchical Bi-prediction (HBP) at the temporal domain, as depicted in Figure 3.1. One Basic Unit is equivalent to one Macroblock (MB).

Figure 3.2 shows the uneven bitrate distribution along different views. This distribution is highly related to the prediction hierarchy inside a Group of GOP (GGOP). The View 0 or Base View is encoded independently with no inter-view prediction. It leads to reduced possibilities of prediction and, consequently, worse prediction, more residues, and higher bitrate. B-Views (View 1, 3 and 5) fully exploit the inter-view correlation by performing disparity estimation (in addition to spatial and temporal predictions) to upper and bottom neighboring views. This increased prediction decision space results in improved prediction quality and tends to lead to reduced bitrates. P-Views (View 2, 4, 6, and 7) represent the intermediate case performing disparity estimation about a single neighboring view. P-Views typically present bitrate in the range between Base View and B-Views bitrates. Note, in Figure 3 the View 7 is a P-View, but its reference view is closer if compared to other P-Views. While View 2 is two views distant to its reference view (View 0), View 7 is just one view distant to View 6. Typically a reduced bitrate is required for View 7 due to a relatively better disparity estimation prediction.

The bitrate relations associated to prediction hierarchy, however, are not always correct and vary with the video/image properties of each view. For instance, in the example provided in Figure 3.2, View 6 (P-View) present reduced bitrate to View 1 and View 3 (both B-Views). Thus, we can conclude that even employing Bi-prediction at disparity domain the View 1 and 3 are harder to predict to View 6 and produce higher bitrate. A similar observation is the increased bitrate generated by View 7 if compared to other P-Views. Reduced bitrate is expected for View 7, but increased bitrate is measured. These observations show that in addition to the dependence on the prediction structure (as discussed above), the bitrate distribution has a high dependence on the video content of each view. Hard-to-predict views typically present high texture and/or high motion/disparity objects and require more bits to reach similar video quality.

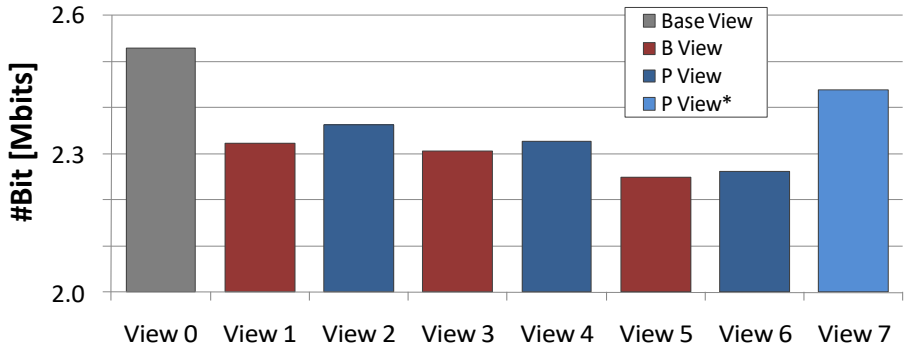

Figure 3.2. View-level bitrate distribution (Flamenco2, QP=34)

The bitrate distribution at frame level presented in Figure 3.3 shows that inside each GOP the frames that give higher bitrate are located at lower hierarchical prediction levels. This is related to error propagation and the distance of temporal references, the farther the reference, the harder to find a good prediction. Therefore, more error is inserted resulting in higher bitrates. In B-Views this effect is attenuated once these views are less dependent on the temporal references due to the higher availability of disparity references. Figure 3.3 illustrates that for neighboring GGOPs the frames at same relative position exhibit similar and periodic rate distribution pattern, the GOP-Phase.

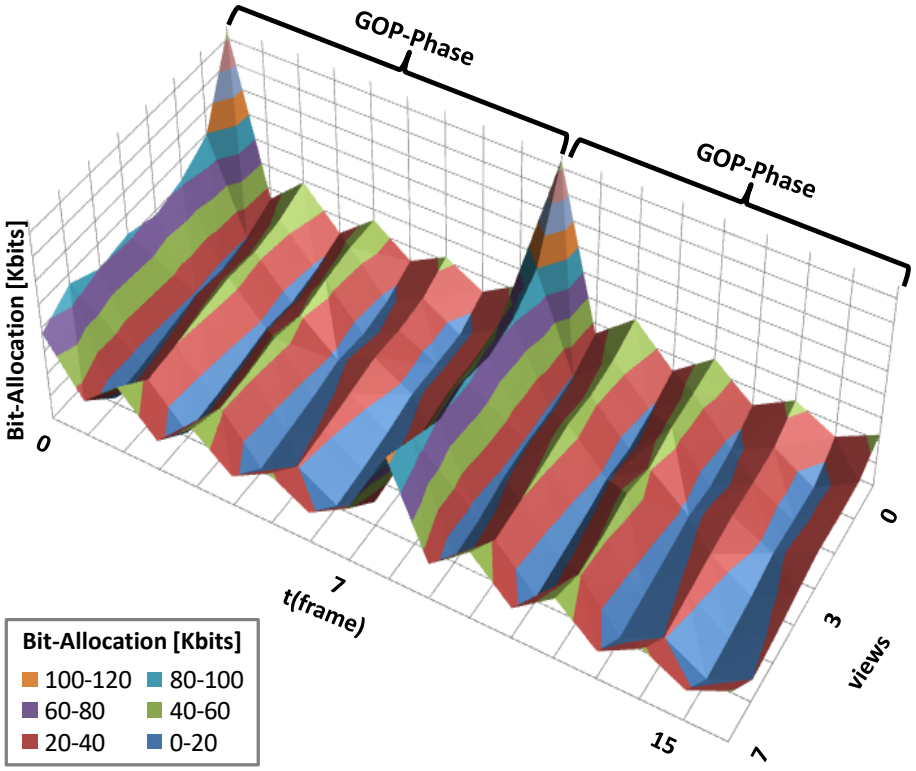

Figure 3.3. Frame-level bitrate distribution for two GGOPs (Flamenco2, QP=34)

Inside each frame, the number of bits generated for each BU is also related to the video content. Figure 3.4 shows that the similar and low motion/disparity background requires lower bitrate if compared to the dancer's region and the textured floor for similar quality. However, the Human Visual System (HVS) requires a higher level of details for texture and border regions to perceive good quality and, consequently, these areas deserve higher objective quality. Therefore, textured regions must be detected and receive further increased the number of bits during the encoding process through QP reduction.

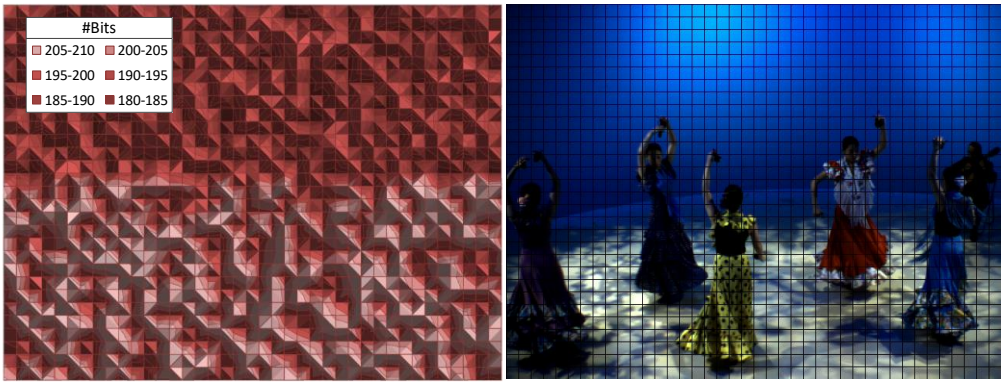

Figure 3.4. Basic Unit-Level bitrate distribution (Flamenco2, QP=34)

Summarizing the Rate Control analysis conclusions:

- The frame-level bitrate distribution depends on the prediction hierarchy and the video content of each frame. An effective Rate Control must consider the neighboring frames at temporal, view and GOP-phase domains.
- The video properties have to be considered at BU-level in order to locate and prioritize regions that require higher quality.

3.2.2 Workload Distribution Analysis

The latest High-Efficiency Video Coding (HEVC) standard (Sullivan, et al., 2012), introduced in early 2013, aims at doubling the video compression ratio compared to its predecessor (i.e. H.264/AVC) (Bossen, et al., 2012). Several new coding tools like the recursive partitioning of the Coded Tree Blocks (CTB), additional filters and Intra prediction modes lead to improved compression efficiency. However, simulations in (Khan, et al., 2013) using reference software (HEVC-Software) show that these improvements come at the cost of an increased computational complexity of ~1.7× compared to H.264. The following analysis illustrates that this corresponds to an increase of 47% in energy consumption. These issues are amplified linearly for simultaneously encoding multiple videos/views like in 3D video encoding (see Figure 3.5). Not only the data-rate increases sharply, but now, the workload of encoding each video/view must be balanced. However, the workload of each independent view may differ by a large amount (see Figure 3.5), thus, making workload balancing particularly harder for encoding multiple videos/views simultaneously.

The 3D version of HEVC standard (3D-HEVC) (Müller, et al., 2013) aims at reducing the data rate by 50% compared to HEVC simulcast. It exploits correlations among images of the same scene from different views, and compression modes of these views, to partially address these issues. However, the increased compression comes at the cost of greater complexity (for instance, the complexity increases by 172% for a 2-view encoding sequence compared to a single view sequence encoding).

- **Target Research Problem and Preliminary Analysis**

Despite the high coding efficiency delivered by 3D-HEVC, the encoding of 3D videos while meeting timing deadlines imposes a huge challenge regarding complexity and power efficiency. Additionally, varying video properties, run-time encoder state (e.g., resources allocated to the encoder by OS), and user preferences, along with the throughput constraints of the encoder accumulate towards additional design issues. Besides keeping the high video quality needs to be high, the complexity of the encoder should not increase beyond reasonable limits, such that 3D-HEVC can be realized on real-world, multi-camera systems.

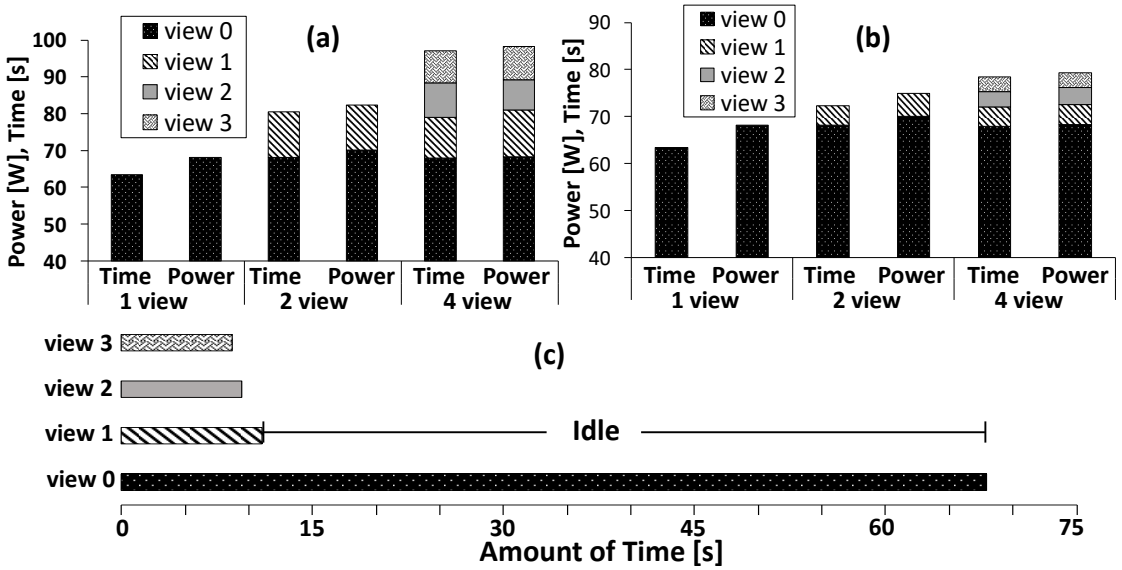

Figure 3.5: Power and Time-Complexity for "Poznan Hall" sequence performed in 1, 2 and 4 views for (a) 1 core and (b) 1 view per core. (c) Core usage Encoding time for 4 views encoding for 1 view per core.

Parallel processing of 3D-HEVC on a multi-core system can meet these design goals. The multi-threaded 3D-HEVC application can divide the workload and assign jobs to each thread, to execute in parallel.

Figure 3.5 (a) and (b) shows the power and time consumption to encode the 3D video sequence “Poznan Hall” in 1, 2 and 4 views using the reference software (3D-HEVC-Software). Figure 3.5 (a) shows the case when only a single core is allocated to process all views, while Figure 3.5 (b) portrays the scenario whereby each view is assigned an independent core to realize parallel processing. Figure 3.5 (c) presents the time consumed to encode each view on independent cores. As noticed, the core utilization for views other than view 0 is less than 10%, denoting that the workload is extremely unbalanced among the cores, resulting in low throughput and high (leakage) power consumption. Thus, a naïve allocation of cores to each video/view will not solve workload balancing problem, and a sophisticated scheme is required.

4 RATE CONTROL ALGORITHMS FOR MULTIVIEW VIDEOS

This Section presents the approach of this work targeting rate control for Multiview Video in the Multiview Video Coding standard and the novel control techniques. The content of background and MPC controller presented in this chapter were presented in (Vizzotto, et al., 2012), the content of the master thesis. Besides the techniques that exploit the frame-level rate control distribution along different video sequences is fundamental to define an efficient and fast QP prediction, able to provide high precision distribution of bitrate at a negligible cost concerning efficiency (rate-distortion tradeoff). Additionally, at fine grain level rate control, a novel prediction algorithm is proposed to offer adaptivity to the changing scenarios..

4.1 Background Knowledge

The next subsections presents the background concepts employed in the proposed scheme for rate control in multiview videos. At first, it is introduced an overview and basic of the Model Predictive Control, which provides the foundation for developing the frame-level RC. In the following, it is shown the statistical supporting to the Markov Decision Process that is implemented at the fine grain level of RC. Finally, the concepts related to Reinforcement learning are introduced.

4.1.1 Model Predictive Control

The Model Predictive Control (MPC) present by (Garcia, et al., 1989) and (Morari, et al., 1997) has demonstrated to accurately predict the response of multiple stimuli dynamic systems such as multiview video encoders by employing the control-theory superposition principle (Tatjewski, 2010). It outperforms traditional feedback controllers by efficiently integrating input stimuli to state space constrains while providing dynamic flexibility by employing phase concept (periodic behavior) through rolling input and output horizons.

The main goal of the MPC is to define the optimal sequence of actions to lead the system to a desired and safe state by considering the system's feedback to previous states and previously taken actions (presented by the conceptual MPC behavior in Figure 4.1).

The actions are taken based on the prediction of the future system behavior associated to a set of predicted future actions, i.e., accurate prediction leads to better actions. Figure 4.1 shows how MPC improves the prediction (predicted output) performance with respect to the actual system output (reference trajectory) using past knowledge (measured output and past control inputs) and predicted future control actions (predicted control inputs). To define this sequence of actions the MPC minimizes the performance function presented in Eq. 4.1. It minimizes the cost by defining a set of outputs y based on a set of inputs u, where $u[k + i - 1|k]$, $i = \{1,\dots, m\}$ denotes the set of process inputs with respect to which the optimization is performed; u is known as the control horizon or input horizon in the MPC theory. The input variation between two contiguous time instants is represented by $\Delta u[k + i - 1|k]$, *i.e.* $\Delta u[k + i - 1|k] = u[k + i - 1|k] - u[k + i - 2|k]$. Similarly, $y[k + i|k]$, $i = \{1,\dots, p\}$ is the set of outputs, named prediction horizon or output horizon (see Figure 4.1). The control horizon determines the number of actions to find. The prediction horizon determines how far the behavior of the system is predicted. *m* and *p* are the size of control/input and prediction/output horizons, respectively. *m* is the index of input horizon while *p* defines how many outputs are predicted, i.e. how many future actions are considered in the optimization processes. w_i is a weighting coefficient that denotes the relative importance of a given output with respect to the future control interaction. r_i is the reference variable, in other words, the value of the reference trajectory (ideal trajectory to be tracked) in instant *i*. *k* is the horizons index and represents the *k-th* input/output horizon. y^{SP} defines the output set point that limits the prediction horizon.

$$\min_{u[k|k]\dots u[k+p-1|k]} \sum_{i=1}^{p} w_i \left(y[k+i|k] - y^{sp}\right)^2 + \sum_{i=1}^{m} \Delta u\left(r_i[k+i-1|k]^2\right) \quad (4.1)$$

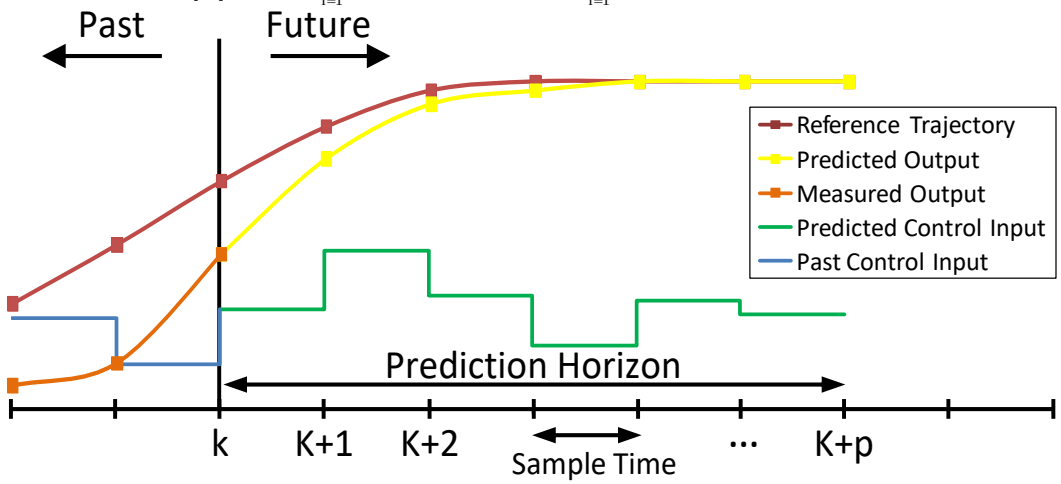

Figure 4.1: Conceptual behavior of the Model Predictive Control (Behrendt, 2009)

4.1.2 Markov Decision Process

The Markov Decision Process (MDP) is a mathematically-based optimization model of discrete state, sequential decision making in a stochastic environment that depends only on the current state and not on the previous states (Bellman, 1957). However, if a controlled MDP is considered the transition probabilities are affected by previous actions. In this scenario, applying Reinforcement Learning (RL) can solve MDP with no explicit probabilities definition.

MDP is formally defined by 4-tuples *(S,A,P(.,.),R(.,.))* composed by a finite set of states *S={s0, s1,...}*, actions *A={a0, a1,...}*, rewards *R={r0, r1,...}* and transition probabilities *P={p0, p1,...}*. The *S* includes all possible states assumed by the controlled system, actions *A* are the possible acts to be taken by the decision-maker in face of a given system state. *P(S)* is the probability distribution of transitions between system states and, finally, *R(S)* is the reward related to a given action for a given state. At each discrete time step *t* the process lays in a state $s \in S$ and the decision maker may choose any action $a \in A$ that will lead the process to a new state $s' \in S$ providing a shared reward $R_{at}(s,s')$. The rewards are used by the decision maker in order to find an action that maximizes, for a given policy, the total accumulated reward, as shown in Eq. 4.2 (where $0 \leq \gamma \leq 1$ denotes the discount factor). Eq. 4.3 defines the probability P_a that an action *a* in the state *s* at time *t* will lead to state *s'* at time *t+1*. In a controlled Markov process, the probabilities are obtained through the state change rewards defining a Markov cyclic chain.

$$\sum_{t=0}^{\infty} \gamma^t R_{at}(s_t, s_{t+1}) \tag{4.2}$$

$$P_a = R_{at}(s_{t+1} = s' | s = s_t, a_t = a) \tag{4.3}$$

4.1.3 Reinforcement Learning (RL)

The reinforcement learning model is an agent to improve autonomous systems performance through trial and error by learning from previous experiences instead from specialists (Barto, 1994), that is, the agent learns from the consequences of actions. In reinforcement learning model the agent is linked to the system to observe its behavior and take actions. RL theory is based on the Law of Effect, that is, if an action leads to a satisfactory state the tendency to produce this action increases. For each discrete time step *t* the RL agent receives the system state $s \in S$ and rewards *R(S)* to take an action $a \in A$ that maximizes the reward $R_{at}(s,s')$. This action may lead the system to a new state $s' \in S$ and produce a system output, in terms of a scalar reinforcement value, used to define the new

reward $R_{a(t+1)}(s,s')$ according to Eq.4.4. The general representation of reinforcement learning value is given by *RL* in Eq.4.5, where *U* denotes the function that changes the system state from *s* to *s'* and h_R denotes the learning history.

$$RL_{a(t+1)}(s,s') = RL_{at}(s,s') + RL \tag{4.4}$$

$$RL = U(s,s') + h_R \tag{4.5}$$

4.2 Hierarchical Rate Control

The Hierarchical Rate Control model is presented originally in (Vizzotto, et al., 2013), adapted to the current version of the encoder. The overall flow of the Hierarchical Rate Control (HRC) for multiview videos is presented in Figure 4.2. The HRC is responsible for controlling the encoder output bitrate, in accordance to the user preferences and/or channel limitations, by monitoring the multiview video encoder and actuating through Quantization Parameter (QP) adaptation. It can be conceptually divided in two actuation levels: frame-level (that encapsulates GOP and frame levels) at coarse grain and; at fine grain level (FG-Level). The contributions of this thesis it is majority in the FG-Level.

The multiview encoder receives the video sequences as input along with all user preferences and configurations to start the encoding process. The Model Predictive Control-based frame-level RC models the system behavior considering the encoding hierarchy and predicts the bitrate allocation at frame-level considering temporal, view and GOP-phase (inter-GOP) correlation. It defines the optimal QP for the predicted frames, the base QP, and forward it to the Markov Decision Process-based Fine Grain level RC. At FG-level, a fine grained-decision is taken to define the QP variation considering the image properties in terms of Regions of Interest. The decision maker considers the previous knowledge, by implementing the Reinforcement Learning method, to increase or decrease the QP in relation to the base QP. To couple the frame- and FG-level in HRC, the Reinforcement Learning unit feedbacks both the MPC and the MDP to keep system consistency and avoid mismatches. The HRC employs an observer unit able to read, store and manage the multiview encoder feedback (generated bitrate) and variables that define the encoder system state (target bitrate, QP, input constraints, etc) in order to support the bitrate prediction and actions/decision taking. Also, an image properties extractor is employed to build the Regions of Interest map used for FG-level RC.

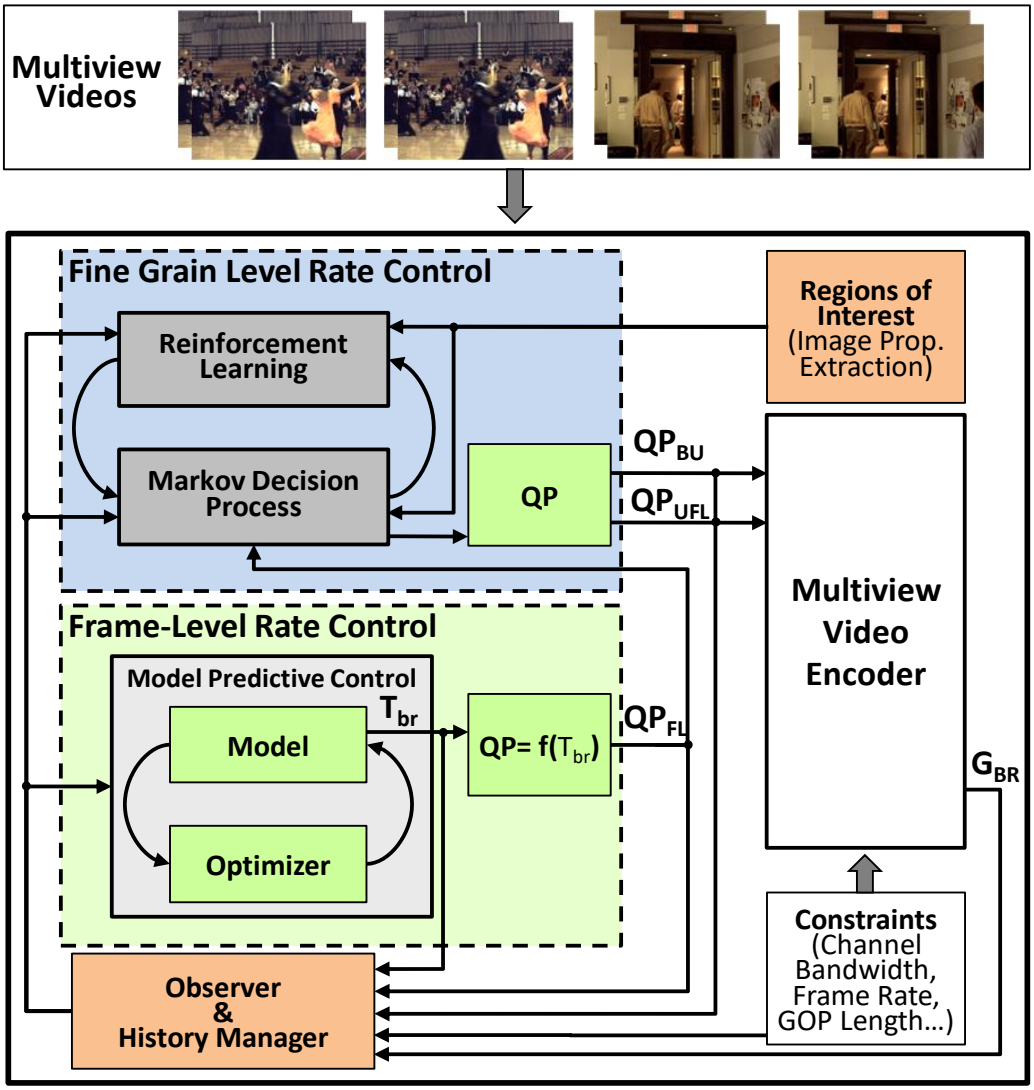

Figure 4.2: Hierarchical Rate Control system diagram
Source: The Author

The different components of the Hierarchical Rate Control scheme is discussed in the following subsections along with the model equations that describe the whole controller behavior. For simplicity it is provided in Table 4.1.

4.2.1 Frame-Level Rate Control

As discussed, the main goal of a Model Predictive Controller (MPC) is to predict the future behavior of a system state and/or output over a finite time horizon as well as compute the future input signals at each step. These actions occur by minimizing a cost function under inequality constraints on the manipulated control or the controlled variables (Zheng, 2010). In this work, the MPC operates at frame level predicting the bitrate and providing the quantization parameter (QP) for each frame to be encoded. The rate controller tries to define a sequence of actions and then induces the system to a desired state while the negative effects of this action are reduced respecting restrictions and taking

constraints into account. In other words, the RC defines a QP that optimizes the bandwidth or bit allocation while maximizing the visual quality and reducing bitrate/quality sudden variations.

Table 4.1: Variables Definitions

Variable	Description
Frame-Level Rate Control	
T_{BR}	Target bitrate for one frame (bits per frame)
BW	Channel bandwidth (bits per second)
FR	Frame rate (frames per second)
BA	Bit allocation (absolute)
w_I, w_P, w_B	I, P and B weight respectively (absolute)
$\bar{w}_{GOP}$	Average w for the current GOP (absolute)
L_{GOP}	GOP Length (# of frames)
ω	Frame weight (absolute)
N_A	Number of anchor frames (# of frames)
BR	Bitrate (#bits)
H_{QP}	QP History (absolute)
QP_{FL}	Quantization Parameter at Frame-level RC (discrete)
QP_{CLP}	Quantization Parameter in last process (discrete)
QP_{st}	Initial Quantization Parameter (discrete)
Q	Quantization Parameter in the optimization loop (discrete)
N_{FR}	Number of frames encoded in the GOP
Fine Grain-Level Rate Control	
M_S	RoI- Normalized Variance Matrix (absolute 0 – 1)
$M(\delta)$	MDP Reward Matrix (matrix of absolute RD)
σ^2	Variance of a given BU
μ	Average of BUi
N_{BU}	Number of BUs
QP_{UFL}	Updated Quantization Parameter at Frame-level RC (discrete)
QP_{BU}	Quantization Parameter at BU-level RC (discrete)
T_{BR}	Target bitrate for one frame (bits per frame)
R_S	BU Reward “Shared” (absolute)
R_{Learn}	Reinforcement Learning Value (vector of HR)
$f(s,\delta)$	Probability of state transition
P_R	Probability results from RL vector of “phase” actions. Actions of RL in a range of at least 2 horizons.
$\Delta\ \delta$	Variation between actual BU δ and the δ of anchor frame

M_f	Variation of variance matrix values
H_R	History of RL
G_{BR}	Generated bitrate (bits per frame)
$U(s,s')$	Function to update the matrix from s to s'

The bitrate prediction is performed considering the neighborhood correlation at temporal, view and inter-GOP domains. As discussed before, there is a high correlation in the temporal and view neighboring frames inside the same GOP. Moreover, there is also a periodic pattern that repeats at GOP level, the GOP-Phase. With the MPC-based rate control enabled, the scheme is able to exploit these correlations in order to accurately predict the future bitrate. Figure 4.3 represents the previously encoded frames used for prediction (control horizon) and the current frame to be predicted (prediction horizon) for a given multiview video encoder prediction structure. As depicted in Figure 4.3, the input horizon is composed of disparity and temporal neighbor frames in the same GOP plus frames belonging to the same temporal instant from the previous GOP. The output horizon is composed by the current frame to be encoded. This method extends the work proposed in (Vizzotto, et al., 2012) by employing a variable weighting factor for frames considering their positions in relation to the current frame. In (Vizzotto, et al., 2012) the weight of the feedback of each frame in the control horizon was defined based on its frame type (I, P or B) and its relative position inside a fixed GGOP structure.

The variable weighting factor is calculated considering the number of references and their distance to the current frame. With this extension the fame-level RC may be directly implemented in any hierarchical bi-prediction structure (HBP) while still catching the GOP-phase correlation.

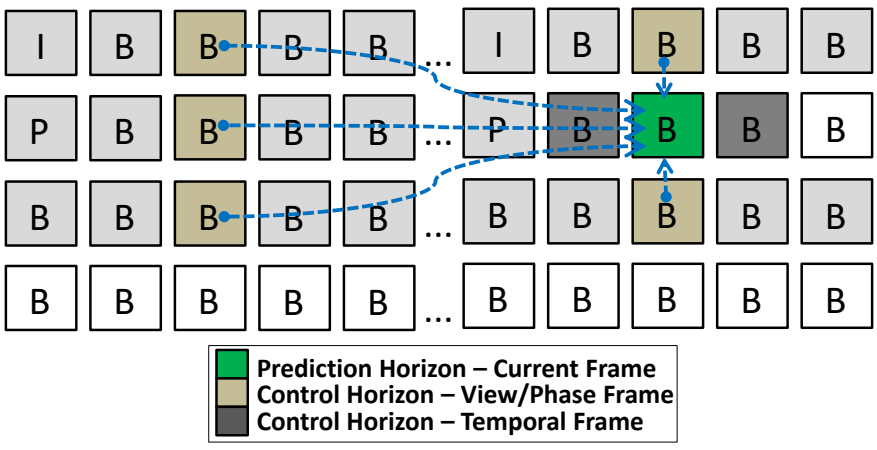

Figure 4.3: MPC-based RC Horizons.
Source: The Author

Figure 4.4 shows the MPC optimization process and how the component functions interact with each other. The Rate Model generates, based on the neighborhood correlation, a bitrate prediction for the current frame, the target bitrate. Based on the prediction an optimal QP is defined and the internal model is updated. The system feedback and the actually used QP defined in the fine grain level RC are received through the observer. Figure 4.4 illustrates the connections of different components of the proposed control scheme, such that important model equations are mentioned in their corresponding boxes. This provides the inter-linkage of different equations.

Figure 4.4 maps the real functions of the proposed model to the MPC conceptual model. The input horizon is the past TBR given by Eq. 4.6. The output horizon is represented by the matrix of bitrate variation (ΔTBR) in Eq. 4.12.

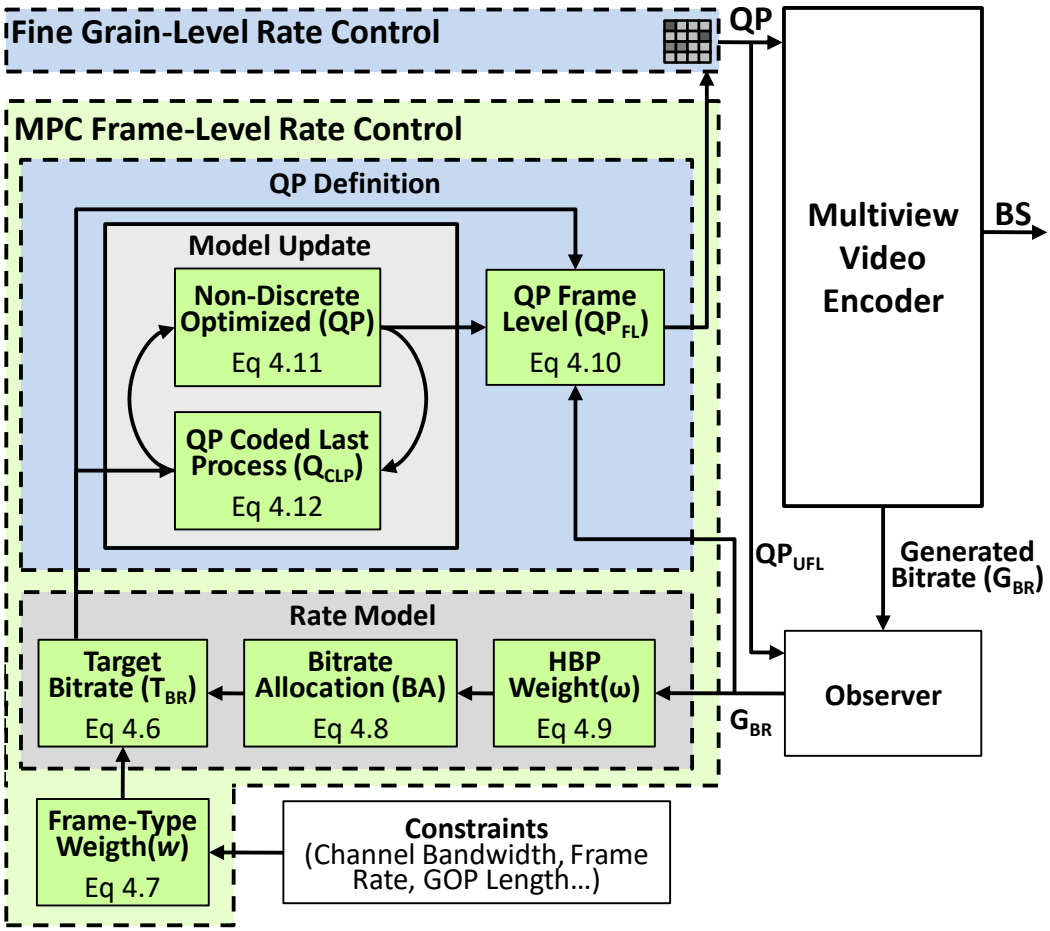

Figure 4.4: Frame-Level Rate Control Diagram.
Source: The Author

4.2.1.1 Rate Model

The MPC-based Rate Control defines the target bitrate ($T_{BR(f)}$) considering the bandwidth (*BW*) and frame rate (*FR*) constrains along with the neighboring frames weights (*w*) and their frames bit-allocation (*BA*), as shown in Eq. 4.6. The term w*BA denotes the weighted bit allocation.

$$T_{BR(f)} = \frac{BW}{FR} + w * BA \qquad (4.6)$$

The feedback and the correlation between frames vary with the type of each frame. The bitrate range of distinct frame types (I, P and B) lie in different ranges. Thus, the weighting factors for each frame type must be different. A weight (w_I) is statically predefined for I frames (Li, et al., 2003) while P and B-frame weights (w_P and w_B) are calculated dynamically considering the weights of temporal neighboring frames. Note w_P corresponds to the weight of the nearest P-frame in the GOP. Eq. 4.7 originally presented by (Li, et al., 2003) shows how the weights are calculated considering the HBP in order to respect the local linearity inside the current GOP; where $\bar{w}_{GOP}$ is the average of *w* computed over all previously-encoded frames in the current GOP, *f* represents the f-th frame of a given type (I, P or B) in the processing order, L_{GOP} denotes the GOP length, $u=1/(L_{GOP-1})$ is defined to provide a clearer representation for Eq. 4.7, and w_{f-1} denotes the weight of the previous frame of the same type. For a smooth weighting propagation, *w* is limited according to a statistically-defined range.

$$\begin{aligned} w_I &= 0.75 \\ w_P &= max\left\{w_{f-1} - 2u, min\left\{\bar{w}_{GOP} - .25, w_I - 2u\right\}\right\} \\ w_B &= max\left\{w_{f-1} - 4u, min\left\{\bar{w}_{GOP} - .25, w_P - 2u\right\}\right\} \end{aligned} \qquad (4.7)$$

The target bit allocation (*BA*) is given by a history-based weighted model to optimize MPC for best target bit allocation, as shown in Eq.4.8. $BA_{(f-1)}$ represents the bit allocation for the last frame of the same type. The original MPC Rate Control proposed in (Vizzotto, et al., 2012) was designed to differentiate between anchor and non-anchor frames through implementing three weights for the *BA* definition: anchor ω_A, non-anchor ω_{NA} and P/B anchor frames ω_{PBA}. However, this thesis extends the model to capture the differences between all frames according to their hierarchical level in HBP and their number of references (0..2 temporal + 0..2 disparity reference frames). The goal is to calculate this weight considering the position of the current frame in the GOP and the number of reference frames used by this frame. Note, frames with different number of references

have a distinct interaction with the GOP bitrate allocation, i.e., the more references used by a frame, the lower the bitrate tend to be (because of more efficient prediction) and this behavior must be considered during bitrate prediction step.

This extension allows the HRC to better respond to variations inside the GGOP and to become more flexible by adapting, without further extensions, to any HBP structure. The weights $\omega_{i,j}^{m,n}$ (where *i* and *j* are the frame time instant and view; *m* and *n* denotes the number of references in the temporal and view domains, respectively) calculation is presented in Eq. 4.9. The term $\omega_{i,j}^{m,n} \Big/ \sum_{m=0}^{X}\sum_{n=0}^{Y}\omega^{m,n}$ represents the normalized weight of a specific frame with respect to the total accumulated weight. The term $BR_{i,j}^{m,n}$ represents the bit number of the corresponding reference frame. The term $\mathrm{QP}_{i,j\ (\mathrm{f}-1)}^{m,n}$ represents the QP of the previously encoded frame within the GOP (*i*,*j*) with specific number of references (*m*,*n*) in temporal and disparity domains.

$$\mathrm{BA}_{(\mathrm{f})} = \left(\mathrm{BA}_{(\mathrm{f}-1)} - \frac{\mathrm{BA}_{(\mathrm{f}-1)}}{\mathrm{N_A} - 1} + \frac{\omega_{i,j}^{m,n}}{\sum_{m=0}^{X}\sum_{n=0}^{Y}\omega^{m,n}} - 1 \right) \times \frac{\mathrm{BW}}{FR} \times L_{\mathrm{GOP}} \qquad (4.8)$$

$$\omega_{i,j}^{m,n} = \frac{(BR_{i,j}^{m,n} \times \mathrm{QP}_{i,j\ (\mathrm{f}-1)}^{m,n}) + (L_{\mathrm{GOP}} - 2)\omega_{i,j\ (\mathrm{f}-1)}^{m,n}}{L_{\mathrm{GOP}} - 1} \qquad (4.9)$$

4.2.1.2 Quantization Parameter Definition

Once the prediction is performed, the RC must define a proper action in terms of QP. The QP is determined by summation of all target bitrate ($T_{BR(f)}$) in the prediction horizon, the summation of all generated bitstream in the control horizon (*BR*), and the history of QPs (H_{QP}, harmonic mean of previously-used frame-level QPs within the GOP), as shown in Eq.4.10. *p* and *m* come directly from the MPC definition in Eq. 4.1 where *p* is the index associated to the output horizon (target BR for future frames) and *m* is the index associated to the input horizon (past frames BR output). Note, the QP defined in the frame-level (QP_{FL}) RC is not directly used by the multiview encoder but forwarded to the BU-level RC to refine the QP selection.

$$QP_{FL} = H_{\mathrm{QP}} \times \left(\sum_{\mathrm{i}=1}^{\mathrm{p}} T_{\mathrm{BR(i)}} \Big/ \sum_{\mathrm{i}=1}^{\mathrm{m}} BR_{\mathrm{(i)}} \right) \qquad (4.10)$$

To maintain the performance of the proposed MPC controller there is a need to update the QP model. For that, the HRC implements an optimization loop with non-discrete steps

(k) where Q_{CLP} denotes the quantization parameter for the frame coded in the last process, i.e., the previous encoded frame. Eq. 4.11 and Eq. 4.12 describe the update process where the QP value is constrained to a variation range of ±2 QP points for smooth update. Note that unlike Q_{FL}, Q_{CLP} does not represent the actually used QP in the previous frame. Rather it is the QP representation inside the model optimization process of the MPC. QP_{max} and QP_{min} are the upper and lower bounds for QP values. These limits are defined by the application user in order to limit quality boundaries or assumed to be ±4 (for this specific implementation) with respect to the initial QP (Q_{st}). In Eq. 4.12, M is the transposed matrix of ω multiplied by target bitrate variation ($\Delta T_{BR(f)}=T_{BR(f)}-T_{BR(f-1)}$) for the frames belonging to the control horizon. The $\sum$ operator represents the sum for all previously encoded frames from zero to the length (L) of the GOP. ΔQ^{j}_{k} denotes the frame-level QP variation between subsequent frames ($\Delta Q^{j}_{k}=Q^{j}_{k}-Q^{j-1}_{k}$) while N_{FR} is the number of frames already encoded belonging to the GOP. Q_{st} is the initial QP. Note, the number of iterations to update is constrained between 1 and GOP/2+1.

$$\begin{aligned} \mathrm{Q_k} &= min\left\{\mathrm{Q_{(k-1)}}+2, max\left\{\mathrm{Q_{(k-1)}}-2, \mathrm{Q}\right\}\right\} \\ \mathrm{Q_{(k-1)}} &= min\left\{\mathrm{QP}_{max}, max\left\{\mathrm{QP}_{min}, \mathrm{Q}_{CLP}\right\}\right\} \end{aligned} \tag{4.11}$$

$$\mathrm{Q_{CLP}} = \sum_{i=0}^{L} \mathrm{Q}_i \times \det\left(\mathrm{M}\left(\omega \times \Delta \mathrm{T_{br}}\right)^{\mathrm{T}}\right) \times \mathrm{Q_{st}} \times \frac{\sum_{j=0}^{L} \widetilde{\Delta \mathrm{Q_k}}^{\mathrm{j}}}{\mathrm{N_{Fr}}} \tag{4.12}$$

4.2.2 Fine Grain-Level Rate Control

As part of the HRC is proposed a FG-level Rate Control employing Markov Decision Process along with Reinforcement Learning able to consider the image properties through a texture-based Region of Interest (RoI) map, as detailed along this section.

Figure 4.5 depicts the diagram of the proposed FG-level RC that works as an amendment or refinement of the frame-level RC. In order to refine the accuracy of bit allocation and provide smooth visual quality, FG-level RC incorporates the concept of Region of Interest (RoI) into a Markov Decision Process (MDP). In this case, MDP additionally employs Reinforcement Learning in order to adapt to dynamic encoder and input variations. At each decision step, the RC monitors the state of the system and determines the next action to take based on constraints observations and the control policy. Firstly, the HRC defines the RoIs for anchor frames generating a map of weights M_S that will determine the importance of each BU inside the picture. Secondly, the

weights map is linked to a map of states *M(δ)* in the MDP that corresponds to the QP for each BU. The MDP fits to the MVC encoder behavior by providing the structure to make decisions partly randomly and partly under a control. Finally, to dynamically adjust the matrix of states for next decision, the RL is responsible to feedback the system response to the current state for both BU-level and frame-level control.

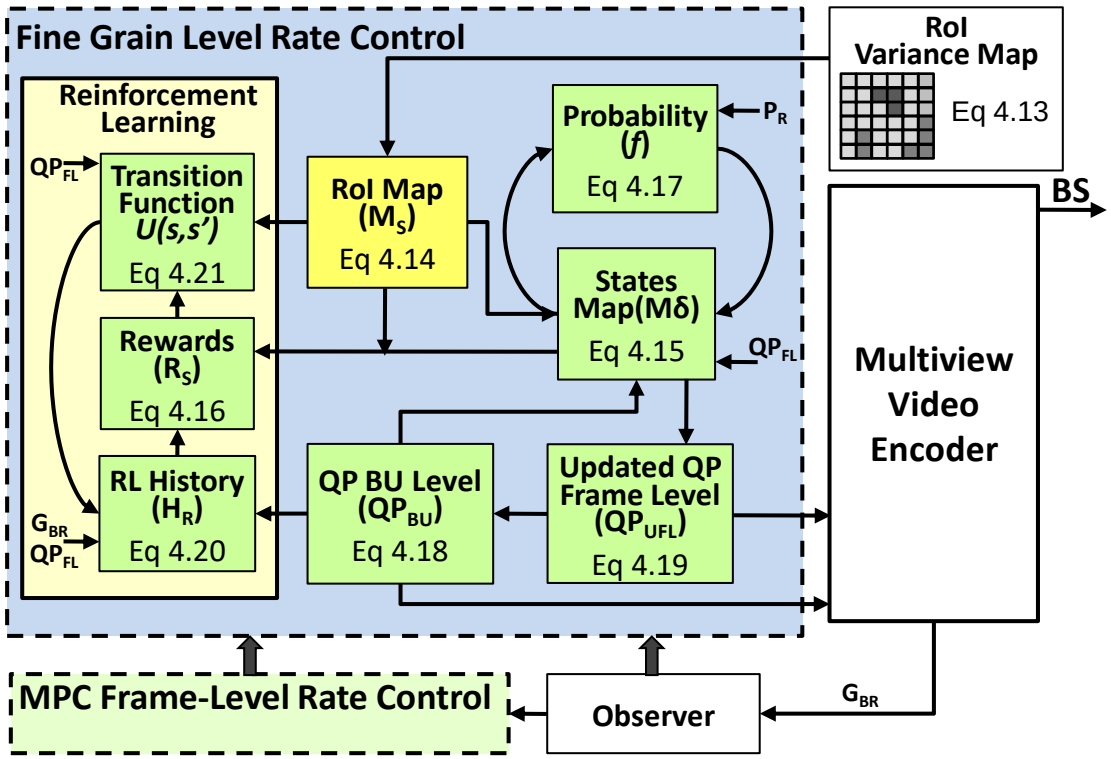

Figure 4.5: Fine Grain Level Rate Control Diagram.
Source: The Author

4.2.2.1 Regions of Interest Concepts

As previously discussed in Chapter 2, frames are composed of regions with distinct image properties requiring a variable number of bits to be encoded. Regular video encoders use the same QP to encode all Basic Units within a frame leading to inefficient bitrate distribution and undesirable quality variations inside the frame. However, it is possible to define regions to receive special treatment, the Regions of Interest (RoI), as defined in (Lee, et al., 2011), (Agrafiotis, et al., 2006) and (Wu, et al., 2009). The BUs belonging to RoIs may be prioritized by the Rate Control unit to protect the quality of those regions. In this work, the whole frame is considered to have the same semantic relevance (this leave space for further application specific extensions (Wu, et al., 2009)) but regions that present a hard-to-predict content must be allowed to use more bits through QP reduction. According to the presented analysis, textured regions tend to generate more residue and, consequently, require higher bitrate.

In this solution, the RoI is determined through a normalized variance map for all anchor frames. Eq. 4.13 defines the variance σ^2 where ρ_i denotes the luminance of pixel i and μ represents the average luminance of the pixels block. The normalized variance map is given by M_S in Eq. 4.14, where $\sigma^2(BU_i)$ represents the variance of the i^{th} basic unit. N is the number of basic units in a frame. Figure 4.6 presents one example of variance map. Additionally, HRC also keeps a second matrix of states where each value represents a bitrate of a frame inside a GGOP encoding history to incorporate temporal and view neighborhood information to the MDP process. The matrixes data are used by the MDP and RL to define the rewards associated to each state and action taken by the control. For non-anchor frames, the HRC uses statistics of anchor-frames with reinforcement learning R_{Learn}.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\rho_i - \mu)^2; \quad \mu = \frac{1}{n}\sum_{i=1}^{n}\rho_i \tag{4.13}$$

$$M_{S(j,k)} = \frac{(\sigma^2(BU_i) - \mu)^2}{N - 1} \tag{4.14}$$

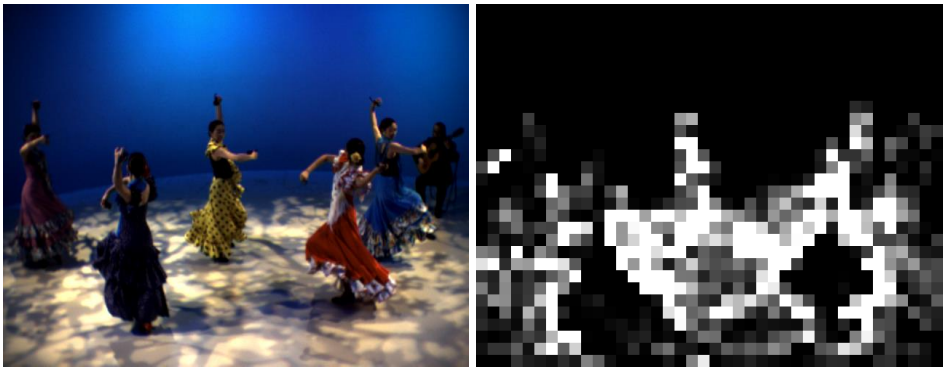

Figure 4.6: Variance-based Region of Interest map (Flamenco2).
Source: The Author

4.2.2.2 Markov Decision Process

The original HRC presented by (Vizzotto, et al., 2013) implements the FG-level RC by employing the Markov Decision Process. Considering that other techniques besides dynamic programming can be used to handle the control problem and to simplify the RL step, MDP stage was restricted to a Partially Observable MDP approach (POMDP-a). This method includes restriction to avoid data-intensive computing without side effects observed in the final definition of the QP. Traditionally, the MDP works over a matrix of independent states $M_f(s)$ representing the QPs of each BU within a frame; this is the same action taken in the POMDP-a. The states and transitions modeled are depicted in Figure 4.7. Each BU has a set of possible actions A with associated rewards R_S and transition

probabilities $f(s,\delta)$. In this model the possible actions are: (i) increment QP (if $f(s,\delta)\geq 1$), (ii) decrement QP (if $f(s,\delta)\leq -1$), and (iii) maintain the QP value defined at frame-level (if $1<f(s,\delta)< -1$); see Eq.4.18. A matrix of coefficients $M(\delta)$ is used to define the reward for each action according to Eq. 4.15. In Eq.4.15, Max_{QP} represents the value of the maximum frame QP within the interval of one GOP length to the past. Note that this is different from QP_{max}, which is a user-provided parameter to restrict the maximum allowable QP for quality reasons. The parameter BS denotes the number of generated bits and is related to the basic unit in that specific matrix position. The rewards R_S are calculated based on the RoI map M_S, the matrix of coefficients $M(\delta)$ and the Reinforcement Learning R_{Learn}, as shown in Eq. 4.16. Note that the parameter R_{Learn} is different from RL of Eq. 4.5. For each action, there is a probability of transition $f(s,\delta)$ defined by Eq. 4.17. Note, $f(s,\delta)$ is same as Pa. However, Pa denotes the probability in the theoretical model while $f(s,\delta)$ denotes the probability in this algorithm. Therefore, $f(s, \delta)$ is the generic representation for any transition according to the MDP theoretical definition while $f(s_i, s_j)$ is this mathematical representation for a given transition from state i to state j. The parameter $\Delta\delta$ denotes the exceeded limit of variation used when a BU receives more than two on QP variation reward. The BU-level QP (QP_{BU}) is derived using a process similar to the one given in JVT-G012, where the quantization step is incremented or decremented by one or left unchanged; see Eq. 4.18.

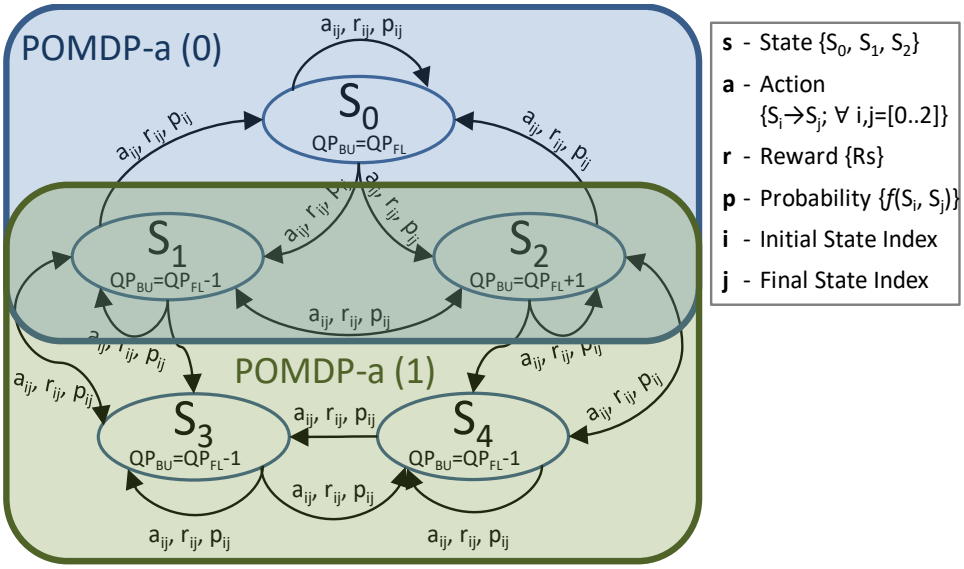

Figure 4.7: Partially Observable Markov Decision Process approach (POMDP-a).

$$M(\delta_{i,j}) = \sum \frac{QP_{BU}{}^{i}_{j} \times BS^{i}_{j}}{Max_{QP} \times (T_{BR}{}^{i}_{j} / N_{BU})} \qquad (4.15)$$

$$R_S = R_{Learn\,j}^{\,i} \times |M(\delta) - M_S| \tag{4.16}$$

$$f(s,\delta) = P_R + \Delta\delta \tag{4.17}$$

$$QP_{BU} = \begin{cases} QP_{FL} + 1 & if \quad f(s,\delta) \geq +1 \\ QP_{FL} - 1 & if \quad f(s,\delta) \leq -1 \\ QP_{FL} & if \quad -1 < f(s,\delta) < +1 \end{cases} \tag{4.18}$$

4.2.2.3 Reinforcement Learning

The Reinforcement Learning incorporates the knowledge of previous events in the decision-making process through monitoring the multiview video encoder system response and updating state transitions probabilities and rewards at both frame- and FG-level. In this thesis, it was used a linear quadratic model approach based on (Sivan, et al., 1972) and (Hager, et al., 1998), where the FG-level feedback occurs by an update at the history of reinforcement learning H_R. So, in Eq. 4.20 it is computed H_R[1] considering the variation of target bit rate (ΔT_{BR}), sum of all BU level QP values ($\sum QP_{BUL}$) of current encoding view (k), sum of generated bitrate at the BU level ($\sum G_{BRL}$) of current encoding view (k), and variation of QP at the frame level (ΔQP_{FL}). ΔT_{BR} and ΔQP_{FL} are defined as the difference between the two last encoded frames for target bitrate and QP at frame level, respectively. QP_{FL} is the basis QP for the whole frame and must be considered when taking local (FG level) decision within that specific frame. In this way, this scheme accounts for the FG-level history knowledge to enable adaptation of BU-level QP, which leads to a reduced Mean Bit Estimation Error (MBEE).

The main improvement of this thesis considering the technique adopted in (Vizzotto, et al., 2013) it is the temporal difference learning given by Bellman equation to solve the dynamic decision problem presented with the coupled solution in the Markov Decision Process. Eq. 4.21 denotes the state transition policy by giving the final MDP state matrix that is used as obtained knowledge for the upcoming frames. The term "$-1 > f(s,\delta) > +1$" occurs when we have a scenario without I-P or I-B transition and the result of Eq. 4.17 is near to zero. Note that $\Delta\delta$ will be between 0 and 1 and the value of P_R will lie between ±2. It is important to note that although P_R denotes a probability, it is represented in the -2 and 2 range (for implementation reasons) and can be remapped to the (0..1) interval by a simple linear scaling where the interval (0..0.25) is mapped to the interval (-2..1),

[1] Note, h_R corresponds to the history of state changes between S and S' considering the value of the states and state change probabilities. H_R also incorporates the history of contexts of target bitrate, QPs at FG level and QPs at frame level.

interval (0.25..0.5) is mapped to the interval (-1..0) and so on. This is similar to Eq. 4.18 and Eq. 4.21 where $f(s,\delta)$ represents a probability remapped to the interval (-10..10). The QP of the frame is updated using Eq. 4.21 and calculated according to Eq. 4.19. In Eq. 4.19, *trunc* represents the truncation operator, $\sum M_{f(s)}$ is the sum of all positions in the matrix of state transition probabilities and N_{BU} is the number of basic units in the GOP. M_f (s, s') denotes the probability of transition from state s to state s'. QP_{UFL} provides feedback to the MPC at frame-level.

$$QP_{UFL} = \text{trunc}\left(QP_{FL} * \sum M_f(s) \Big/ N_{BU} \right) \tag{4.19}$$

$$H_R = \frac{\Delta T_{BR} \times \sum QP_{BUL}^{k}}{\sum G_{BRL}^{k} \times \Delta QP_{FL}} \tag{4.20}$$

$$U(s,s') = QP_{FL} \begin{cases} M_f(s,s') \; if \; -1 > f(s,\delta) or \; f(s,\delta) > +1 \\ M_f(s,s) \; if \qquad -1 \le f(s,\delta) \le +1 \end{cases} \tag{4.21}$$

To find the policy of pi (π) that maximizes the FG-level Reinforcement Learning (given by Eq. 4.22) it is used backward solution technique – where *E* denotes the expected value – considering a GGOP as a finite horizon of action. The algorithm used is given by function U in Eq. 4.23, where t denotes one frame inside a GGOP, which is composed of a set of frames T ($t \in T$), *a* is the action in the set of actions *A*, *s* denotes the states and *r* is the reward while *p* denotes the probability of transition.

$$U_\pi = E_{\pi,s}\left\{ \sum_{t=1}^{T} r(s_t, a_t) \right\} \tag{4.22}$$

$$U_t(s) = \max_{a \in A}\left\{ r(s,a) + \sum_{s' \in S} p(s' \mid s,a) U_{t+1}(s') \right\} \tag{4.23}$$

4.2.3 Results and Evaluation

This section presents the experimental results for this work. In the following lines, it is described the simulation environment and the experimental setup. Finally, the detailed results considering control accuracy, coding efficiency, and video quality are presented with in-depth discussion for the view, frame and BU-level perspectives.

4.2.3.1 Experimental Setup

The simulation environment is based on the MVC reference software, the JMVC 8.5 (JMVC-Software, 2012), with the required extensions to implement the HRC and state-of-the-art solutions. As test sequences eight sequences with 8 views each are considered:

four VGA (640 x 480 pixels) sequences - “ballroom”, “exit”, “vassar” and “flamenco2” - , two XGA (1024 x 768 pixels) - “Breakdancers” and “Uli”, and two HD1080p (1920 x 1080 pixels) - “GT Fly” and “Poznan Hall2”. The view coding order follows the IBP [2] pattern, 0-2-1-4-3-6-5-7, to consider all possible view types (I, P and B-views). The GOP size was defined as 8 with temporal Hierarchical Bi-prediction Prediction (HBP) structure as recommended by VCEG-AA10 report (Tan, et al., 2005). The Basic Unit is defined as one macroblock. All sequences were encoded using four target bitrates along 13 GGOPs (105 frames per view); 256, 392, 512, 1024 kbps for VGA; 512, 768, 1024, 2048 kbps for XGA; and 1024, 1536, 2048, 4096 kbps for HD1080p. CABAC and FRExt were enabled.

To provide representative results for a wide range of video content it is selected sequences with high (“flamenco2”) and low motion (“vassar”) and with high (“Uli”) and low disparity (“vassar”). Also, to guarantee a fair comparison to the state-of-the-art, the solutions proposed in the literature were implemented in the infrastructure and evaluated under the same settings applied for the HRC. For comparison, well known metrics were used to measure the RC accuracy (MBEE), coding efficiency, and objective video quality (Bjøntegaard Delta Bitrate and Bjøntegaard Delta PSNR). Additionally, encoded pictures are presented to attest the subjective video quality.

Note that the JMVC has no RC defined or implemented. Therefore, it is not possible to set the JMVC encoder to output a given bitrate. However, it is possible (by using experimentation) to select the QP value that delivers a bitrate that best matches a given target. For this reason the MBEE provided by JMVC is the highest one. Regarding the comparison with state-of-the-art, comparisons with both frame-level (Yan I, et al., 2009), (Lee, et al., 2011), (Vizzotto, et al., 2012) and FG-level (Li, et al., 2003) rate control schemes are provided. For simulations and comparison with state-of-the-art, it is used the same configuration file provided by JMVC reference software. For this, it is extended the configuration file with an additional subsection Rate Control in the Section “Encoding”.

4.2.3.2 Control Accuracy and Bitrate Precision

As previously discussed, the RC is supposed to sustain the bitrate as close as possible to the target bitrate (optimizing the bandwidth usage) while avoiding sudden bitrate variations. To measure the RC accuracy, that is, how close the actual generated bitrate (Ra) is in relation to the target bitrate it is used the Mean Bit Estimation Error (MBEE), defined in Eq. 4.24. The mean is calculated over all Basic Units (NBU) along 8 views and 13

GGOPs for each video sequence. Table 4.2 presents the accuracy in terms of MBEE (less is better) for the HRC compared to the state-of-the-art solutions (Yan I, et al., 2009), (Lee, et al., 2011), (Vizzotto, et al., 2012) and (Li, et al., 2003) respectively [a], [b], [c] and [d]. On average, the Hierarchical Rate Control provides 0.95% MBEE while ranging from 0.7%-1.37%. The competitors (Yan I, et al., 2009), (Lee, et al., 2011), (Vizzotto, et al., 2012) and (Li, et al., 2003) present, on average, 2.55%, 1.78%, 2.03% and 1.18%, respectively. This illustrates that HRC scheme performs better than the MPC-only frame-level RC (Vizzotto, et al., 2012). The superior accuracy is a result of the ability to adapt the QP jointly at frame and BU-levels considering the neighborhood correlation and the video content properties.

$$\mathrm{MBEE} = \left\{ \sum_{i=0}^{N_{BU}} \frac{|R_t - R_a|}{R_t} \times 100 \right\} \Big/ N_{BU} \quad (4.24)$$

Table 4.2: Control Accuracy comparison

Sequence		Bit-Rate [kbps]							MBEE [%]					
		Target	JMVC	[a]	[b]	[c]	[d]	HRC	JMVC	[a]	[b]	[c]	[d]	HRC
VGA	*Ballroom*	256	268	263	260	262	259	258	4,64	2,63	1,48	2,43	1,17	0,75
		392	408	402	397	401	396	395	4,06	2,61	1,32	2,21	1,07	0,78
		512	529	523	520	521	518	516	3,33	2,16	1,59	1,83	1,13	0,78
		1024	1058	1048	1041	1045	1032	1032	3,30	2,35	1,63	2,04	0,81	0,78
	Exit	256	267	261	259	258	258	258	4,29	2,10	1,18	0,86	0,88	0,94
		392	408	402	397	402	397	396	3,99	2,55	1,36	2,46	1,29	0,92
		512	528	523	521	520	519	516	3,21	2,25	1,83	1,65	1,36	0,83
		1024	1056	1048	1043	1043	1038	1031	3,14	2,34	1,85	1,90	1,38	0,72
	Flamenco2	256	268	263	259	263	258	258	4,79	2,91	1,30	2,89	0,84	0,71
		392	409	402	400	398	397	395	4,34	2,56	2,01	1,60	1,25	0,71
		512	530	526	522	524	519	516	3,56	2,68	1,96	2,36	1,36	0,84
		1024	1059	1049	1044	1043	1040	1031	3,41	2,44	1,98	1,87	1,58	0,70
	Vassar	256	267	262	261	262	258	258	4,27	2,30	1,91	2,27	0,81	0,75
		392	407	402	399	401	396	395	3,73	2,50	1,71	2,18	1,00	0,72
		512	528	525	519	520	517	516	3,13	2,54	1,39	1,56	1,07	0,86
		1024	1056	1049	1037	1038	1035	1033	3,15	2,46	1,28	1,38	1,10	0,86
XGA	Break dancers	512	525	524	521	522	519	518	2,47	2,41	1,82	1,98	1,37	1,23
		768	801	788	782	784	778	776	4,33	2,54	1,88	2,09	1,26	1,08
		1024	1052	1050	1044	1044	1036	1034	2,72	2,56	1,91	1,99	1,21	1,00
		2048	2101	2109	2093	2094	2072	2070	2,58	2,99	2,19	2,24	1,15	1,06
	Uli	512	525	525	521	522	519	519	2,46	2,54	1,84	2,03	1,29	1,37
		768	801	789	783	784	777	776	4,28	2,72	1,90	2,14	1,20	1,08
		1024	1052	1052	1043	1044	1036	1034	2,74	2,72	1,87	1,97	1,16	0,95
		2048	2101	2101	2092	2095	2071	2069	2,59	2,60	2,17	2,28	1,13	1,05
HD	GT Fly	1024	1050	1049	1043	1045	1038	1037	2,54	2,44	1,86	2,05	1,37	1,27
		1536	1581	1575	1565	1568	1556	1553	2,93	2,54	1,89	2,08	1,30	1,11
		2048	2104	2101	2087	2089	2073	2069	2,73	2,59	1,90	2,00	1,22	1,03
		4096	4202	4219	4186	4188	4143	4140	2,59	3,00	2,20	2,25	1,15	1,07
	Poznan Hall2	1024	1049	1050	1043	1045	1037	1038	2,44	2,54	1,86	2,05	1,27	1,37
		1536	1582	1578	1565	1569	1555	1553	2,99	2,73	1,89	2,15	1,24	1,11
		2048	2104	2104	2086	2089	2072	2068	2,73	2,73	1,86	2,00	1,17	0,98
		4096	4202	4203	4185	4190	4143	4139	2,59	2,61	2,17	2,29	1,15	1,05
Total Average									3,31	2,55	1,78	2,03	1,18	0,95

Source: The Author.

In Figure 4.8 the long term behavior of distinct Rate Control schemes in terms of accumulated bitrate. A more accurate RC maximizes the use of available bandwidth and, consequently, the accumulated bitrate are presented. After a few initial GGOPs (4-5) required for control stabilization, the HRC curve fits better to the target bitrate followed by (Vizzotto, et al., 2012), as shown in Figure 4.8. JMVC with no RC presents the worst bandwidth usage, as expected.

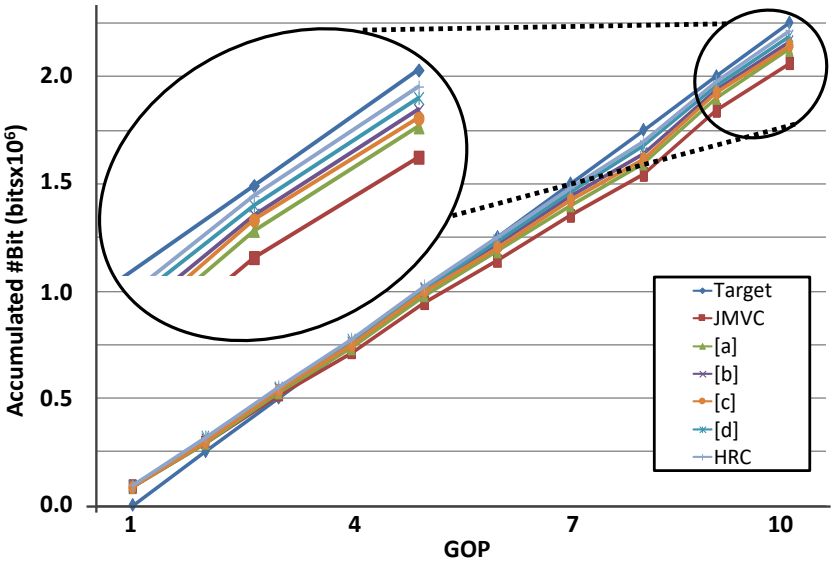

Figure 4.8: Accumulated number of bits for test video “flamenco2”.
Source: The Author

4.2.3.3 Video Quality

Once the accuracy of the HRC is proven it is presented the rate-distortion (RD) results to show that overall video quality and quality smoothness are not compromised. Figure 4.9 and Figure 4.10 summarize the objective rate-distortion in terms of BD-PSNR (Bjøntegaard Delta PSNR) and BD-BR (Bjøntegaard Delta Bitrate) (Tan, et al., 2005) in relation to JMVC with no RC. The HRC provides 1.86dB BD-PSNR increase or BD-BR reduction of 40.05%, on average. When compared to the scheme of (Lee, et al., 2011) that provides the best RD performance among all related works, the HRC provides 0.06dB increased BD-PSNR and 3.18% reduced BD-BR. Note, in addition to the superior RD performance, HRC also outperforms (Lee, et al., 2011) in terms of bit estimation accuracy (1.08% MBEE reduction).

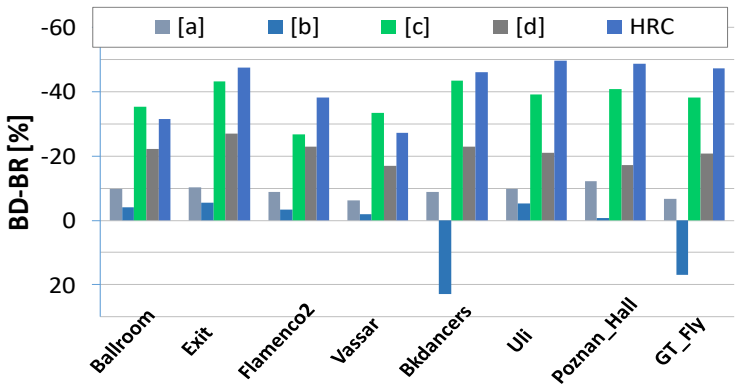

Figure 4.9: BD-BR reduction compared to JMVC.
Source: The Author

There are few cases in Figure 4.9 and Figure 4.10 (like comparison to (Li, et al., 2003) for ballroom sequence), where this proposed scheme does not achieve the best RD-performance. Note that the first goal of an RC solution is to provide an accurate bitrate allocation respecting the available bandwidth. This is crucial to meet the bandwidth requirements and buffer design in order to avoid underflow and overflow case. For this, MBEE is typically employed for evaluation of rate controllers.

Providing high rate-distortion efficiency is the second goal. It is noteworthy that for a few situations where the RD performance is inferior to the state-of-the-art (e.g., compared to (Li, et al., 2003) for ballroom sequence) the proposed scheme outperforms the scheme of (Li, et al., 2003) in terms of control accuracy, as demonstrated in Table 4.2. In summary, for a proper comparison it is important to keep in mind that RD and accuracy must be considered together where accuracy is the main goal of a RC.

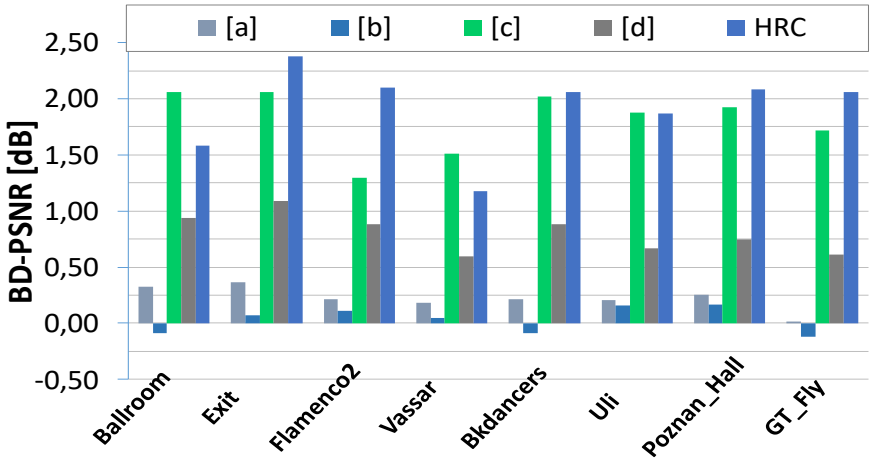

Figure 4.10: BD-BR increase compared to JMVC.
Source: The Author

4.2.3.4 Detailed Results

In this section it is presented the HRC detailed results for “flamenco2” sequence encoded at 1024kbps. For simplicity, it is analyzed only the first 4 views. Figure 4.11 shows the target bitrate, the total accumulated bitrate and the accumulated bitrate for each view. The presented bitrate distribution is smooth also at view level without abrupt oscillations. As expected from the previous discussion, the base view (View 0, I-view) is more bitrate hungry followed by P-views (View 2) and B-views (View 1, 3).

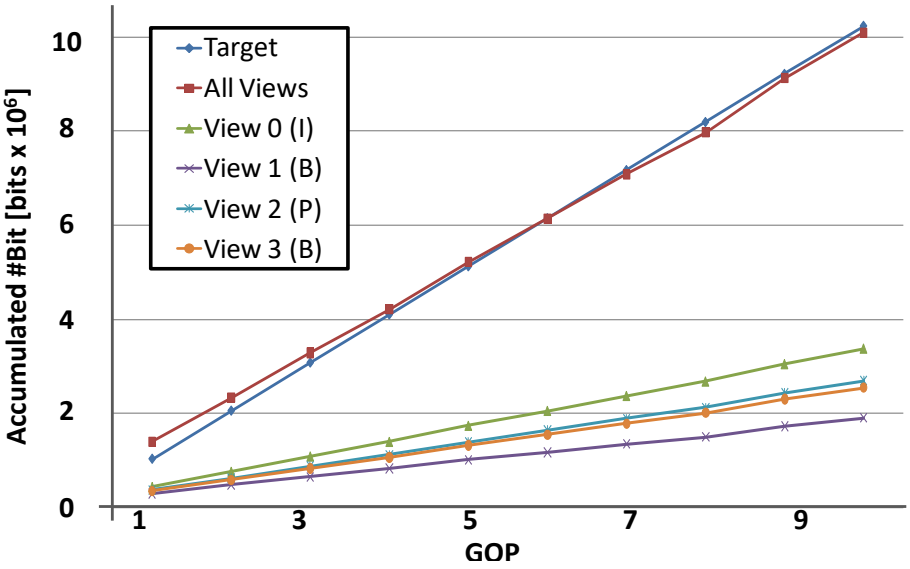

Figure 4.11: View-level bitrate distribution (Flamenco2).

Source: The Author

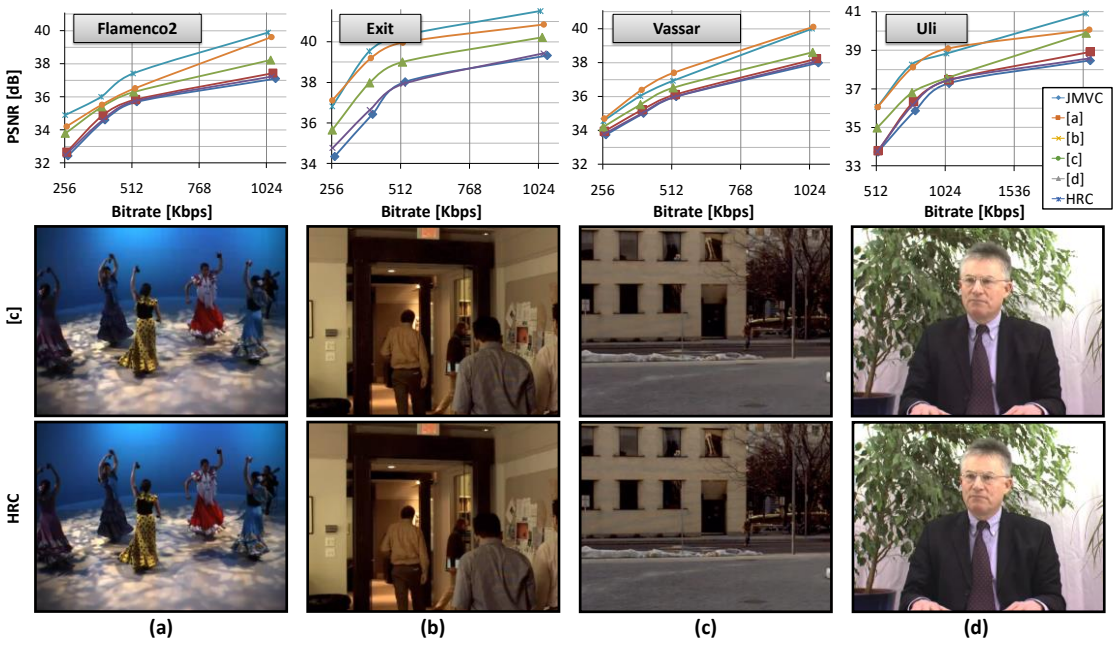

Figure 4.12: Rate-Distortion Results.
Source: The Author

Figure 4.12 shows the RD curves for four different video sequences considering high (Figure 4.12a) and low motion (Figure 4.12b), low (Figure 4.12c) and high disparity (Figure 4.12d) sequences and an intermediate case with moderate motion/disparity activity (Figure 4.12b). The HRC shows its superiority in relation to the state-of-the-art for most of the RD curves (except for the low motion/disparity sequence “vassar”). Decoded frames encoded using the HRC and (Lee, et al., 2011) are also presented in Figure 4.12 for

subjective quality considerations. Note that HRC does not insert visual artifacts such as blurring and blocking noise. Moreover, it does not compromise the borders sharpness typically lost in case of bad QP selection.

In Figure 4.13 it is detailed the controller behavior along the time at frame level. Each point represents the average bitrate or PSNR for all frames in a given time instant. It is possible to note that the bitrate (Figure 4.13a) and PSNR (Figure 4.13b) oscillations tend to reduce along the time because of the RC stabilization. The HRC stabilization is clearly noticed by comparing the first GGOP (dotted box) with, for instance, the GGOP #8 (dashed box). GGOP #8 delivers improved bandwidth utilization (actual bitrate closer to target), better video quality and reduced quality oscillation.

To quantify the video quality smoothness is measured, for the experiment presented in Figure 4.13a, the PSNR variance for HRC and (Lee, et al., 2011). HRC provides 0.47dB PSNR variance while for (Lee, et al., 2011) the variance is 0.60dB, that is, the HRC delivers a video quality with reduced PSNR oscillation in comparison to the state-of-the-art. Figure 4.13b highlights the superior HRC performance in terms of bandwidth usage (better accuracy, as discussed in Section VII.B) while delivering superior and smoother video quality (Figure 4.13a) compared to related work solutions.

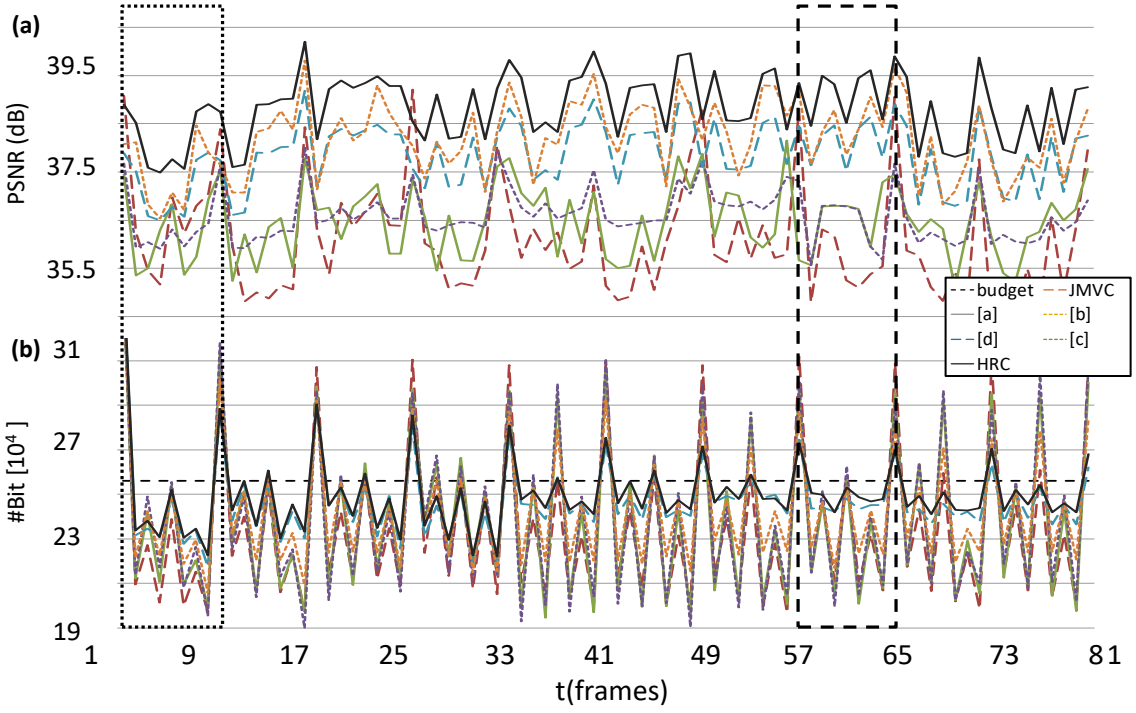

Figure 4.13: Controller behavior results considering (a) quality and (b) accuracy.
Source: The Author

The frame-level bitrate distribution is further detailed for the GGOP #8 (highlighted in Figure 4.13) in Figure 4.14. It shows, graphically, the smooth bitrate and PSNR variations delivered by the proposed solution considering frame-level. Note, the HRC surface presents no sudden variations for both bitrate and PSNR. Compared to the other solutions, it is clear that the bitrate and quality provided by HRC are significantly smoother even when compared to previous solution presented by (Vizzotto, et al., 2012) (this solution provides the lowest MBEE among all competitors).

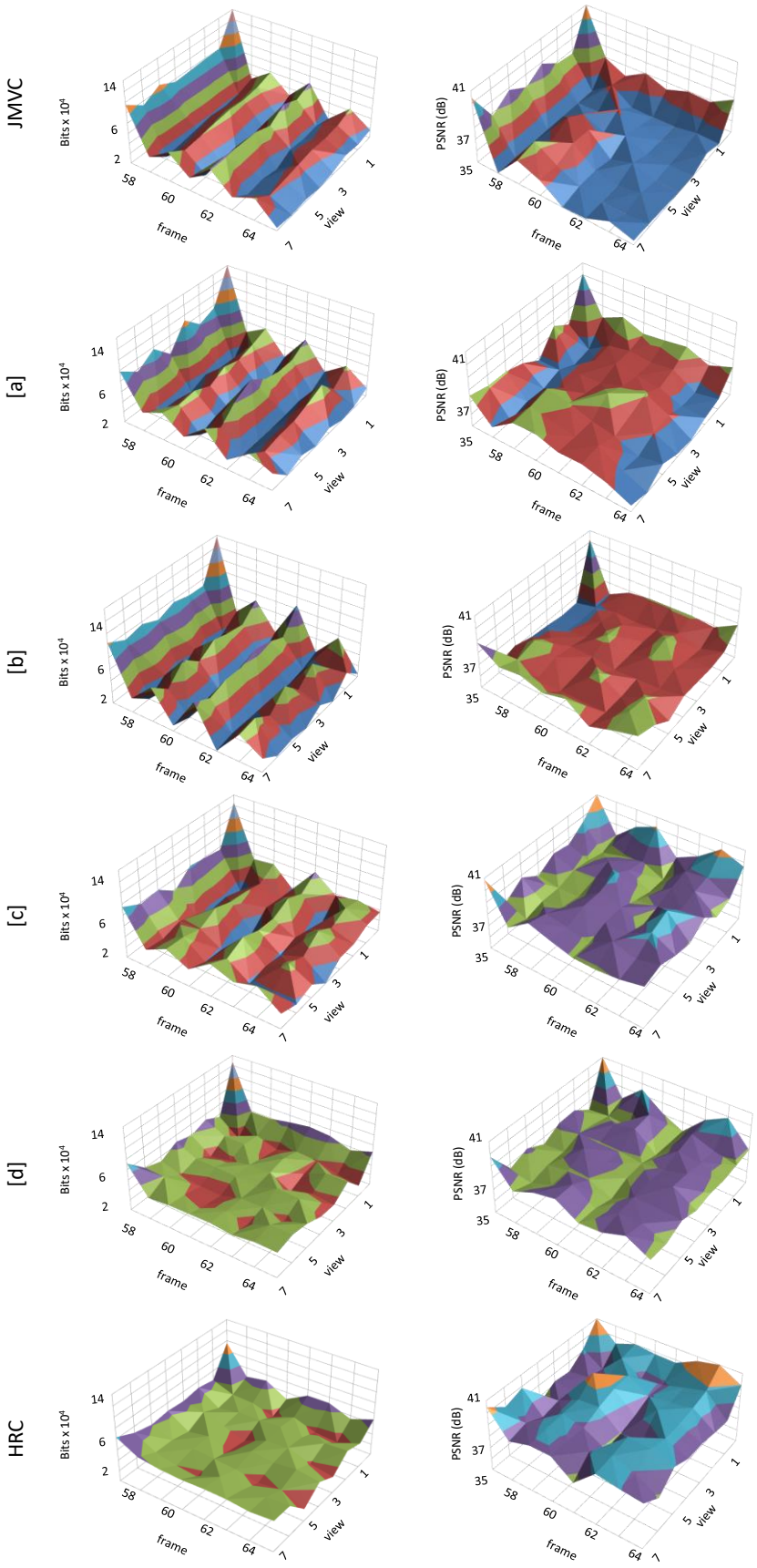

Figure 4.14: Bitrate and PSNR distribution at frame level (GOP #8).
Source: The Author

Analogous analysis was performed to demonstrate the behavior of the RC at BU level. Figure 4.15 shows the bitrate distribution for a frame region (zoomed image) in sequence “flamenco2”. Observe that for HRC the bitrate varies with the texture complexity due to RoI-aware (see texture map in Figure 4.15) MDP implementation. In case of the homogeneous background fewer bits are spent while in case of the textured objects and borders (dancer) more bits are allocated. Note that, in Figure 4.15, the HRC bitrate distribution surface plot fits the object shapes. This behavior prioritizes the regions where the HVS requires a higher level of details leading to superior overall quality. State-of-the-art techniques are unable to accurately react to the image content. Among related works, (Yan, et al., 2009) adapts better to the image. In addition, the HRC also results in smoother variations within the same region (dancer’s body or background) if compared to the state-of-the-art, as shown in Figure 4.15. It avoids sudden quality variations and coding artifacts inside those regions.

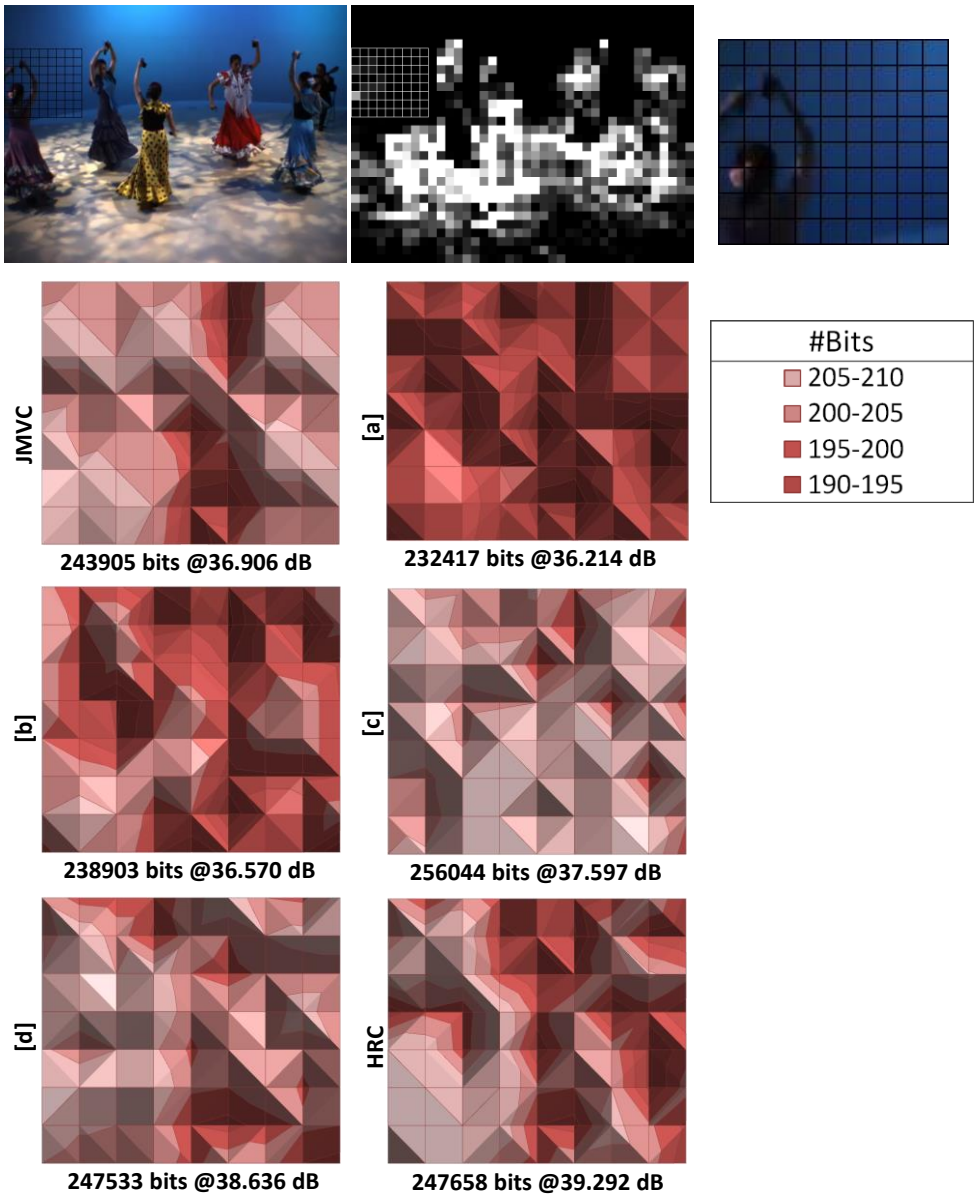

Figure 4.15: Bitrate distribution at BU level (GOP #8).
Source: The Author

4.2.3.5 Detailed Results for Showing the Effects of Different Features of the proposed Rate Control Scheme

Table 4.4 illustrates the bit rate, MBEE, and PSNR results for different features of the rate control scheme. The following three cases are discussed:

1) MPC is the frame-level model predictive rate controller. This rate controller adapts the bit budgets at the frame level and can only handle the variations at a coarse-granularity. However, it cannot handle the bit variations at a fine-

granularity, like Basic Unit (BU)-level due to different texture and motion properties of different objects present in the same video frame.

2) MPC+MDP is the frame-level MPC with BU-level Markov Decision Process based rate controller. BU-level rate control handles the variations at the block level due to different objects with diverse motion and texture properties. However, a pure BU-level rate control may lead to long durations for controller stability. Therefore, typically a BU-level rate control is coupled with a frame-level rate control, which reaches to a stable state faster compared to the BU-level rate controller. The MPC rate controller thereby determines the control range for the MDP-based rate control, which can then refine the bit allocation decisions to reduce the MBEE.

3) MPC+MDP+RL is similar to case 2 but additionally employs reinforcement learning for the MDP. In this case, the reinforcement learning enhances the decisions of the MDP based BU-level rate control to adapt to the run-time changing scenarios of different objects and image variations. The primary limitation of incorporating reinforcement learning is additional computational

In the following, a discussion about the results of different features of the proposed hierarchical rate control scheme.

MPC brings significant quality improvements both in terms of bit rate reduction and PSNR improvement compared to the JMVC at a given target bit rate. Though the Rate-Distortion (RD) performance of MPC and MPC+MDP are similar, MDP provides 2% improved MBEE accuracy, which improves the buffer behavior. However, when employing the reinforcement learning of MDP, the joint solution may achieve an accuracy of approximately 23%. The employment of reinforcement learning also provides improved RD-performance compared to the MPC-only case. When jointly considering the Table 4.3 and Table 4.4, a breakdown of benefit can be observed, where MPC provides 67% improvement in terms of MBEE reduction when compared to JMVC, MDP provides 2%, and finally reinforcement learning provides 21% MBEE improvements. In absolute terms, the MPC-only solution incurs 1.18% MBEE. By employing MDP and reinforcement learning, the BU-level RC reduces the MBEE to 0.95%.

The PSNR results in Table 4.4 show that using coupled MPC, MDP and RL techniques, the proposed RC reaches ~ 1.7dB gain compared to reference software. MPC technique represents 55% of total visual quality gain, MPC + MDP provide 14% of this gain and finally MPC + MDP + RL provide 31%. Overall, it is noteworthy that reinforcement learning and MDP jointly provide improved RD-performance compared to MPC-only solution. Overall, the integrated solution provides significantly improved RD-performance compared to state-of-the-art (see Table 4.3).

Table 4.3: Bit-Rate, MBEE, and PSNR results for different features of Rate Control Scheme

Sequence		Bit-Rate [kbps]				MBEE [%]			PSNR [dB]			
		Target	MPC	MPC+MDP	MPC+MDP+RL	MPC	MPC+MDP	MPC+MDP+RL	JMVC	MPC	MPC+MDP	MPC+MDP+RL
VGA	*Ballroom*	256	259	259	258	1,17	1,03	0,75	32,37	33,825	33,87	34,03
		392	396	396	395	1,07	0,97	0,78	34,81	35,444	35,527	35,933
		512	518	517	516	1,13	1,06	0,78	35,73	36,373	36,59	37,042
		1024	1032	1032	1032	0,81	0,80	0,78	37,20	38,362	37,892	39,228
	Exit	256	258	258	258	0,88	0,90	0,94	34,35	35,663	35,925	36,824
		392	397	397	396	1,29	1,25	0,92	36,43	37,98	38,131	39,53
		512	519	519	516	1,36	1,33	0,83	38,02	39,003	39,283	40,297
		1024	1038	1038	1031	1,38	1,35	0,72	39,32	40,217	40,217	41,509
	Flamenco2	256	258	258	258	0,84	0,77	0,71	32,44	33,794	33,915	34,901
		392	397	397	395	1,25	1,16	0,71	34,61	35,392	35,396	36,004
		512	519	519	516	1,36	1,31	0,84	35,70	36,292	36,604	37,412
		1024	1040	1040	1031	1,58	1,55	0,70	37,10	38,229	38,888	39,883
	Vassar	256	258	258	258	0,81	0,75	0,75	33,74	34,206	34,343	34,555
		392	396	396	395	1,00	0,97	0,72	35,01	35,522	35,739	36,027
		512	517	517	516	1,07	1,02	0,86	36,00	36,513	36,61	36,93
		1024	1035	1035	1033	1,10	1,07	0,86	37,99	38,609	39,254	40,008
XGA	Break dancers	512	519	519	518	1,37	1,33	1,23	33,38	34,925	35,643	36,032
		768	778	777	776	1,26	1,23	1,08	35,74	36,523	37,205	37,735
		1024	1036	1036	1034	1,21	1,17	1,00	36,83	37,423	38,142	38,543
		2048	2072	2071	2070	1,15	1,13	1,06	38,23	39,36	40,136	40,814
	Uli	512	519	518	519	1,29	1,23	1,37	33,72	34,97	35,49	36,055
		768	777	777	776	1,20	1,15	1,08	35,86	36,797	37,522	38,262
		1024	1036	1035	1034	1,16	1,12	0,95	37,28	37,553	37,904	38,825
		2048	2071	2071	2069	1,13	1,11	1,05	38,47	39,884	40,116	40,904
HD	GT Fly	1024	1038	1038	1037	1,37	1,33	1,27	33,52	34,794	34,854	34,891
		1536	1556	1556	1553	1,30	1,28	1,11	34,02	35,129	35,305	35,324
		2048	2073	2073	2069	1,22	1,20	1,03	35,11	36,138	36,311	36,334
		4096	4143	4142	4140	1,15	1,13	1,07	37,84	38,914	38,973	38,987
	Poznan Hall2	1024	1037	1037	1038	1,27	1,26	1,37	33,67	35,038	34,934	34,941
		1536	1555	1554	1553	1,24	1,18	1,11	34,23	35,332	35,405	35,462
		2048	2072	2071	2068	1,17	1,13	0,98	35,39	36,365	36,522	36,56
		4096	4143	4143	4139	1,15	1,14	1,05	38,00	39,054	39,146	39,151
Total Average						1,18	1,14	0,95	35,69	36,676	36,931	37,467

Source: The Author.

4.2.3.6 Complexity Results

For the MPC (at frame level) the number of calls and the number of samples is reduced as it is proportional to the size of the GOP. Similarly the processing effort for MDP (at BU level) is proportional to the frame size. Compared to the MVC encoding case without a rate controller, the proposed hierarchical rate control scheme incurs an average encoding time increase of 2.25% (worst case: 3.11%). The detailed encoding time overhead is presented in the Table 4.4. It is worthy to note that, the overhead of the rate control scheme is still smaller than that of the reference (Lee, et al., 2011). Considering the quality improvement of the scheme, the overhead of 2.25% may be acceptable.

Table 4.4: Complexity results for the proposed Scheme

Encoding Time Overhead						
Sequence		[a]	[b]	[c]	[d]	Our
VGA	Ballroom	0,39%	0,70%	4,61%	1,00%	1,81%
		0,20%	0,35%	4,11%	0,48%	1,95%
		0,24%	0,31%	4,27%	0,53%	1,88%
		0,22%	0,53%	4,21%	0,66%	1,82%
	Exit	0,33%	0,59%	4,65%	1,11%	1,89%
		0,20%	0,44%	4,40%	0,92%	1,64%
		0,24%	0,42%	4,41%	0,92%	1,94%
		0,42%	0,77%	4,44%	0,85%	1,88%
	Flamenco	0,30%	0,30%	4,23%	0,74%	1,94%
		0,28%	0,52%	4,37%	0,66%	2,12%
		0,18%	0,55%	4,33%	0,53%	2,04%
		0,15%	0,51%	4,27%	0,63%	2,04%
	Vassar	0,09%	0,31%	4,23%	0,90%	1,64%
		0,04%	0,30%	4,44%	0,89%	1,60%
		-0,06%	0,04%	4,19%	0,63%	1,34%
		0,22%	0,33%	4,26%	0,61%	1,95%
XGA	Break dancers	0,37%	0,67%	2,68%	1,65%	2,68%
		0,26%	0,49%	2,47%	1,41%	3,11%
		0,23%	0,65%	2,44%	1,48%	3,05%
		0,23%	0,55%	2,33%	1,49%	2,87%
	Uli	0,14%	0,35%	2,33%	1,36%	2,37%
		0,00%	0,26%	2,44%	1,24%	2,22%
		0,05%	0,31%	2,35%	1,25%	2,46%
		0,32%	0,31%	2,39%	1,22%	2,47%
HD	GT_Fly	0,37%	0,49%	2,56%	1,59%	3,01%
		0,11%	0,16%	2,16%	1,18%	2,81%
		0,18%	0,25%	2,24%	1,05%	2,78%
		0,26%	0,18%	2,25%	1,17%	2,81%
	Poznan Hall	0,33%	0,46%	2,37%	1,49%	2,45%
		0,53%	0,39%	2,31%	1,48%	2,42%
		0,25%	0,31%	2,09%	1,22%	2,15%
		0,36%	0,28%	2,29%	1,20%	2,76%
Average		0,23%	0,41%	3,35%	1,05%	2,25%

Source: The Author.

5 POWER EFFICIENT THREAD MANAGEMENT FOR MULTIVIEW VIDEO ENCODING

To address the increased bandwidth due to multiple views, the Multiview Video Coding can provide up to 50% bit-rate reduction compared to simulcast by exploiting the inherent correlation between different views of a multiview video. However, this improved bit-rate reduction comes at the cost of significantly increased computational complexity due to inter-view prediction.

To meet the throughput requirements of a parallelized 3D-HEVC video encoder on a multi-core system, while optimizing the power consumption of the system. In the next lines it is presented a thread management scheme to adaptively distribute the workload of 3D-HEVC as individual jobs among parallel threads. The goal is to balance the workload and accordingly tune the voltage-frequency of the underlying cores, such that the power consumption of the multi-core system is minimized. Further, it is employed application- and content-aware complexity management scheme which adaptively tunes the application's parameters at runtime. A reduced complexity results in a smaller operating frequency to meet the deadline, and thus, complexity management scheme results in higher power-efficiency. Summarizing:

- **Workload Balanced Thread Management**, which employs workload balancing technique to pack encoding jobs and dispatch them to respective threads, such that the application's throughput requirements are met.
- **Run-time Power Manager**, which optimizes the voltage-frequency levels of each individual core in the multi-core system, that would be enough to sustain the workload allocated to a particular core.

To the best of our knowledge, this is the first work in the direction of power-efficient parallelized 3D-video coding. Moreover, it is proposed an open-sourced 3D-HEVC video encoder with multi-threading capabilities as a service to the research community

5.1 Complexity Analysis and Estimation

The impact of disparity estimation on MV/3D-HEVC encoding is presented in Figure 5.1 showing the relation between inter-frame prediction within a view defined by the

motion estimation and the inter-view modes in the disparity estimation for encoding process of “Poznan Hall” sequence. It can be noted that DE/ME modes relation grows with the increasing number of views. Moreover, can be demonstrated the percentage of correlated prediction mode for temporal and disparity in spatial neighbors CUs. Where correlated prediction refers to neighbors block that use the same mode in its prediction, like the same motion vector. Figure 5.1 (b) presents the time complexity distribution for 6-views encoding considering I-B-P coding structure order. This distribution is highly correlated to the prediction hierarchy structure. The base view (*View 0*) is encoded only with intra- and inter-frame modes with no inter-view prediction leading to reduced possibilities of prediction. On the other hand, bi-predicted views fully exploit the inter-view correlation by performing DE - in addition to spatial and temporal predictions - to upper and bottom neighboring views. Moreover, Figure 5.1 (a) shows that the most disparity modes used, the least the complexity.

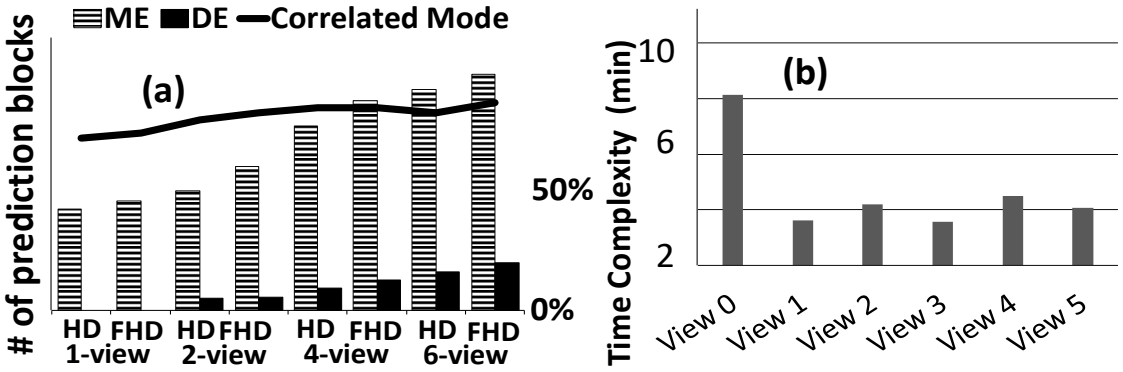

Figure 5.1: (a) Comparison between number of Disparity/Motion modes for 1, 2, 4 and 6 views in HD and FHD resolution. (b) Time for each of 6 view encoded “Poznan Hall” sequence (0-5-2-4-1-3 order)

In order to achieve fast encoding at minimal video quality loss, it is proposed the use of complexity management at Coding Tree Unit (CTU) level. To evaluate this scenario, Figure 5.2 shows the histogram of percentage difference in complexity (number of cycles) and generated bitstream (number of output encoded Bytes) between spatial neighbor CTUs within a tile. These neighbors are considered in relative base views and consecutive frames in the encoding process. Moreover, in Figure 5.2 (a), the horizontal axis presents the percentage of generated byte difference, θ between neighbors CTU of consecutive video frames. θ is given by Eq. 5.1:

$$\theta = \frac{B_0(c,i) - B_0(c,i-1)}{B_0(c,i)} \times 100 \quad (5.1)$$

Here, $B_0(c,i)$ represents the bytes generated to encode the first CTU of frame i. As seen, the generated bitstream for neighbor CTU is highly correlated. The same equation applies for Figure 5.2 where encoding time replaces the total of encoded bytes. Similarly, the generated bitstream per CTU, the time complexity by neighbor CTUs are correlated. Moreover, these curves can be estimated via a Gaussian distribution. Thus, the correlation between neighboring CTUs can be exploited to estimate the bitrate and time complexity of the current CTU, which can be translated to determine the workload.

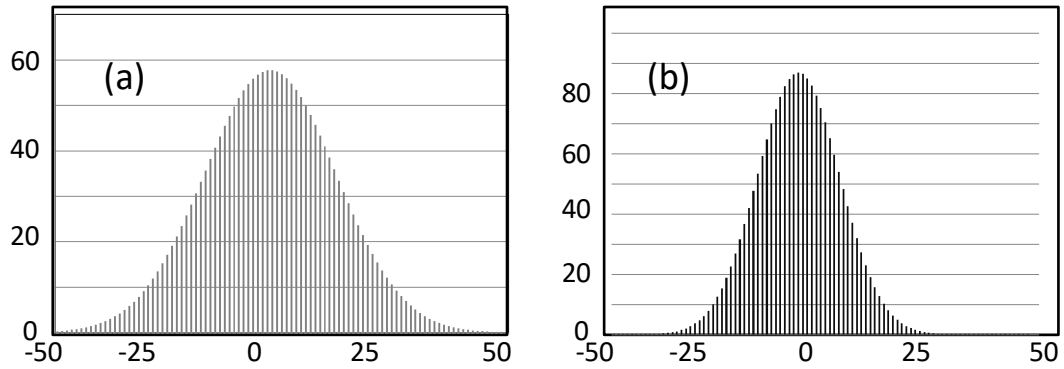

Figure 5.2: Difference histogram of (a) Bytes and (b) Encoding time for neighbor CTU of "Poznan Hall" sequence (FHD) for 80 frames

Summary of Observations:

- Base view is the most complex view to encode due to the lack of disparity data.
- The increase number of views decrease the average complexity.
- Neighboring CTUs within a tile present more than 80% of correlated prediction modes.

5.2 Initial Configuration

Before starting the MV/3D-HEVC, the proposed scheme sets up the hardware, depending upon the throughput requirements (e.g., resolution, number of views and target bitrate) and hardware characteristics, a tile structure is generated by the offline setup. This setup predicts the number of cycles that a CTU will consume given the size of the frame and the throughput requirements. With this information the function can predict the number of CTUs in a tile for a nominal clock frequency of the cores. Thereafter, in the second interval, with the complexity prediction, it maps the CTU-assignment policy

considering its collocated base view CTUs complexity. There are several complexity adjustment knobs in MV/3D-HEVC which can be tuned to reduce the total workload at the cost of increased size of the compressed output (i.e., resulting Bits).

5.2.1 Complexity Prediction

The Complexity Prediction adjusts the number of performed predictions/tests per CTU (Intra, Inter-Frame, Inter-view and Skip, given by γ that denotes the average of neighbors complexity) to adjust the computational complexity. For example, HEVC tests 35 modes to determine the best Intra mode. However, the number of tests can be reduced at the expense of reduced video quality (i.e., more compressed bytes will be generated). For Intra, it is selected the γ most popular occurrences in the past encoded frame. For inter-frames, the complexity is predicted by normalizing (1-35) the average complexity of collocated neighbors given statistically considering both spatial and disparity predictions.

In this scheme, the selective reduction of candidate modes is performed by using statistics provided by the neighboring CUs that have highly related encoding process, sharing residual information at CTU level. Figure 5.3 presents the algorithm of the proposed scheme for the encoding process. First, the number of available CTU per core (*h*) is calculated by the spatial resolution and the number of views of a given 3D sequence.

Initial Configuration and Complexity Estimation Algorithm

```
1. function CTUcomp (Hres[], Wres[], Nview[], Ncore[], PDmode[], CTUwidth[], CUsize[], B[])
2. Nctu ← (Hres x Wres x Nview)/(CTUwidth²);
3. h = Nctu/Ncore;
4. t ← mod(h);
5. if (freq < maxfreq) {t←1; γ←PDmode[Intra]}
6. else{
7. foreach (tile)
8.  foreach (CTU)
9.      switch (mode_prediction)
10.  case (skip): CTUc ← 1; break;
11.  case (inter frame):
12.          ACCc (m) ← ∑ CUc (m) x 2^(CTUwidth/CUw);
13.          CTUc (m) ←ACCc(m) x NF; break;
14.  case (inter view):
15.          ACCc (m) ← SMO(m) x 2^(CTUwidth/CUw);
16.          CTUc (m) ←ACCc(m) x NF; break;
17.  case (intra): CTUc ← SMO(m) x NF; break;
18.  endloop
19. w ← B x Nctu x CTUc;    //Equation (2)
20. γ ← PDmode [ CTUc];
21. endloop
```

Figure 5.3 Initial Configuration, Complexity Estimation for Workload Adaptation and Thread Management with selective approach.

Here, *h* represents the number of CTUs within a frame per views and hence in the number of tiles (line 4). Note that due to the MV/3D-HEVC standard targeting ultra-high-resolution videos, it is not expected to have a higher number of cores than the number of CTUs being assigned. The total number of threads and γ for the CTU within tile T (line 3). γ is given according to normalized complexity (line 13-15) and it is adjusted to proper compute the workload for the current CTU (line 19-20).

5.3 Workload Adapter and Thread Management

The proposed scheme for CTU-based workload balancing of MV/3D-HEVC is presented in Figure 5.4 that aims minimizing the power consumption of the system. The Workload Balancing and Thread Manager controls the computational complexity of processing CTUs, using online statistics and then adapting the encoding parameters. Moreover, it also adapts the number of threads and the number of CTUs assigned to a thread. Furthermore, the Power Manager is responsible for determining the number of cores used and to dynamic scale their voltage and frequencies.

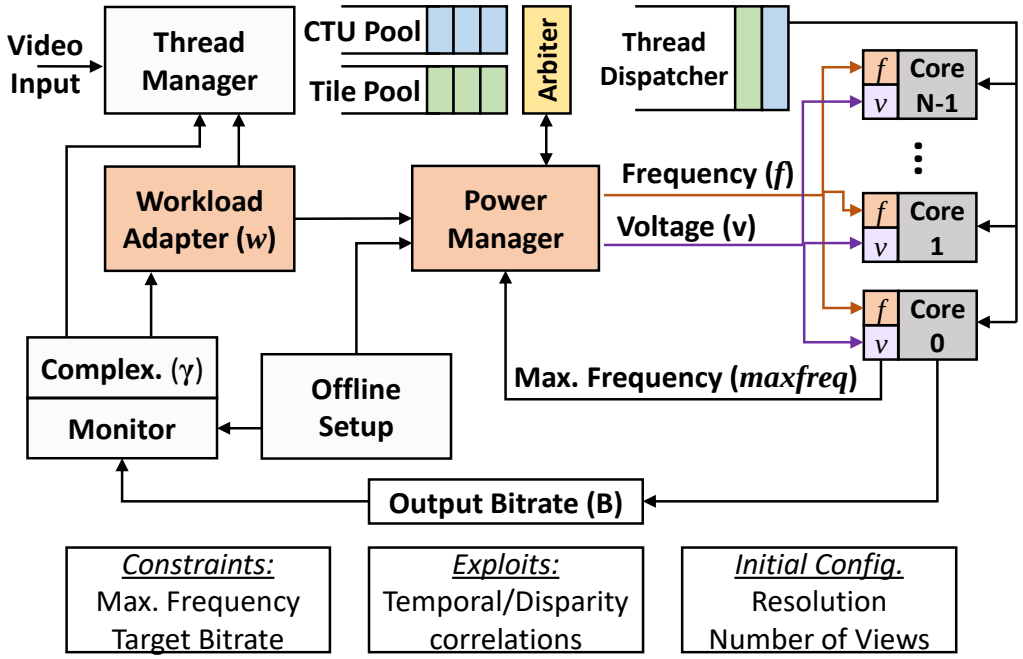

Figure 5.4 Workload balancing scheme for 3D-HEVC on Multi-core.

To reach high compression efficiency while minimizing the video quality degradation (qualified as rate-distortion, which is a well-adopted metric in the video community), it is dispatched the CTU and Tile threads to use all the available cores with maximum frequency. The algorithm ends when γ for all CTUs in a Tile are defined as well the thread

pool of tile level is complete. The workload w of CTU c in tile t, with the target bitrate constraints of B is given by:

$$w_c(\gamma, QP) = B_\gamma \times \sum_{t=1} CTU \times C_{(\gamma,QP,T)} \quad (5.2))$$

Where, $C_{(\gamma,QP,T)}$ is the number of cycles consumed by a CTU of a frame with T tiles, with the given γ and QP values. The summation pertains to the total number of CTU of tile T. Note that w denotes the total number of cycles consumed per second for the given tile.

For workload balancing, the hierarchical prediction structure of 3D-videos is used as adaptation interval IN (see Figure 5.5 (a)), whereby the number of Group of Pictures are adapted according to the statistics. The starting frame (Anchor) of this interval is always intra-frame and achieve the best rate-distortion compression. The anchor frame in base view is used as reference to the Non-Anchor Predicted (P) and Bi-Predicted (B) frames.

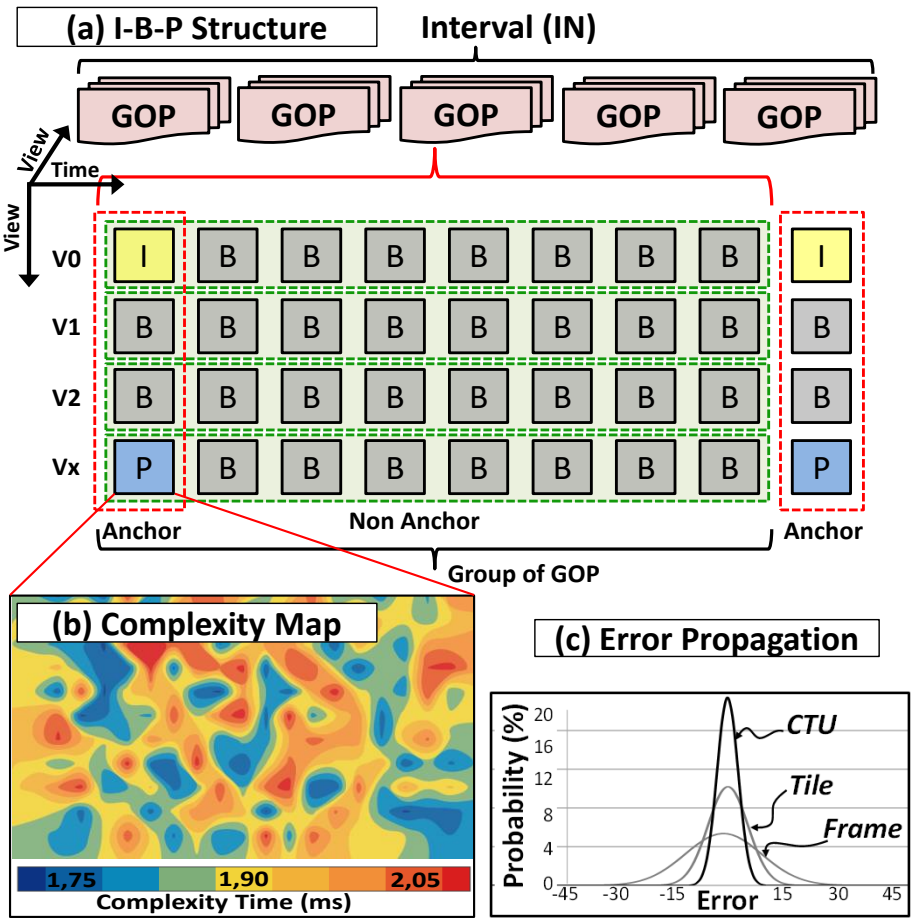

Figure 5.5: Interval in (a) I-B-P structure with (b) related complexity map for Inter frame and (c) complexity prediction error propagation.

As show in Chapter 2, the inter-frame/view modes of spatial neighboring CTUs are highly correlated. Figure 5.5 (b) presents a map where γ is gradually adjusted to each

CTU by the complexity prediction and the workload balance for. Figure 5.5 (c) presents the error propagation in predicted complexity for each structure in an interval. It can be noted that CTU has the lowest probability of error compared to Tile and Frame.

The thread manager uses the balanced workload as policy to selective map threads in two pools. One pool is CTU level and the other one is reserved for tiles. The thread manager uses the workload of each CTU within a tile to compare to complete tile workload and by using a Metropolis-Hasting distribution of CTUs in the last interval to choose witch what thread will be allocated as shown in Eq. 5.3.

$$if\left(w_t > \sum_{i}^{N} w_{CTU}\right) then\ \{Tile\ Pool \leftarrow_{Thread}(w_t)$$

$$else\ \{\ CTU\ Pool \leftarrow_{Thread}(w_{CTU})\} \quad (5.3)$$

Finally, in Eq 5.4 the arbiter selects between Tile and CTU pool to dispatch the thread by considering power manager information (available cores, frequency and voltage).

$$if\left((available\ core) and\left(freq < maxfreq * \frac{norm[\gamma]}{100}\right)\right)$$

$$then\ \{dispatch\ (_{CTU\ Pool}) else\ \{\ dispatch\ (_{Tile\ Pool})\} \quad (5.4)$$

5.4 Run-time Adaptive Power Control

The Run-time Power Control scales the operating frequency and voltage of each core, depending upon the workload of the thread assigned. The frequency of each core is adjusted by an offset based on the predicted complexity of the CTU in the same tile of the hierarchical structure (disparity neighbor). The number of CTU in a tile determines the core operating frequency and voltage $k(f,v)$, and the $offset$ to control the number of CTU in the neighbor views, $offset\ (k)$, are defined by Eq. 5.5 and Eq. 5.6.

$$k(f,v) = maxfreq \times \left(\frac{CTUtotal}{N} + offset(k)\right) \quad (5.5)$$

$$offset(k) = offset(k-1) + \left(\underbrace{CTUc}_{SMO} - \frac{1}{\gamma}\right) \quad (5.6)$$

In these equations, $k(f,v)$ is the operating frequency and voltage of k^{th} core, k is the core index. Also, N is the number of tiles in a frame (and neighbors view), and CTUtotal is the number of CTUs in the *IN* (interval). In Eq. 5.6, the frequency and voltage offset

for each core are set according to the time complexity of CTU. The proposed algorithm adopts an adaptive operating frequency and voltage method, considering the difference between the ideal complexity for each CTU or tile thread, and the predicted complexity, which leads to good tradeoff between power efficiency and RD.

5.5 Results and Analysis

5.5.2 Simulation Setup

The 3D-HEVC reference software (latest version: HTM-14.1) is provided by Fraunhofer institute (3D-HEVC-Software). However, it does not have threading capabilities available. Moreover, it is very difficult to map the source code to include tile threading. Therefore, the time consumed by setup and irrelevant test conditions are too costly and needlessly intrude into the coding complexity. In this way, it was developed an in-house, functionally compliant 3D-HEVC encoder in C++ with multi-thread capability in our lab (in collaboration with the Chair of Embedded System of the Karlsruhe Institute of Technology).

Hardware platform simulation is performed in the Sniper x86 multi/many-core simulator (Carlson, et al., 2011) with support to dynamic voltage and frequency scaling. The measurements of power efficiency are generated with McPAT (Li, et al., 2013).

The experimental results were generated using the video sequences present in the Common Test Conditions by the Joint Collaborative Team on 3D Video Coding Extension Development JCT3V (Rusanovskyy, et al., 2013). It is used five video sequences with different spatial/disparity behavior in three different resolutions: *Poznan Hall*, *Poznan Street*, and *GT_Fly* in Full HD (1920x1080) and HD (1280x720); *Kendo* and *Balloons* in XGA (1024x768). The experiments were performed using 2, 4 and 6-views sequences, QP={22,27,32,37}, GOP = 8 frames and TZ Search.

5.5.3 McPAT Simulation Framework

Figure 5.6 presents a block diagram of the McPAT framework (Li, et al., 2013). The McPAT software uses an XML-based interface with the performance simulator. The use of this interface allows both the passing of dynamic activity statistics and the specification of the microarchitecture configuration parameters generated by the performance simulator. Moreover, the software sends runtime power dissipation back to the

performance simulator through the interface, then the performance simulator can respond to power or temperature data.

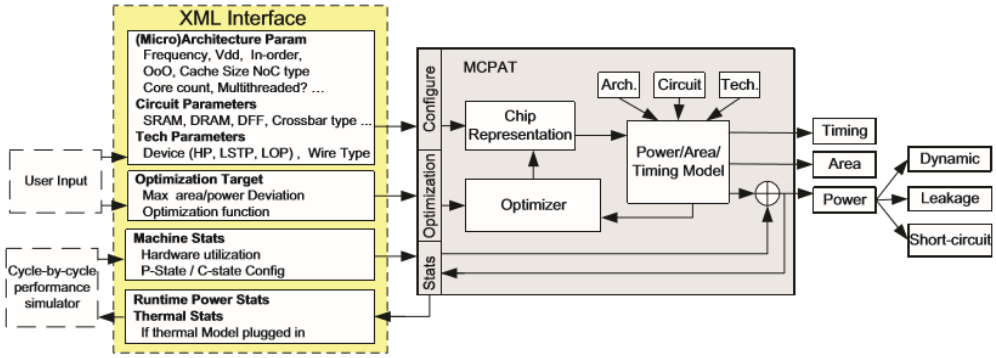

Figure 5.6: Block diagram of the McPAT framework (Li, et al., 2013).

The components of McPAT are described as follows. Firstly, the hierarchical modeling of power, area, and timing. At second, the circuit-level implementations for optimization. Finally, the internal chip representation to present the analysis of power, area, and timing. The input parameters directly set the majority of the parameters in the internal chip representation, such as core issue width and cache capacity.

The hierarchical structure of McPAT allows to model structures at a low level, as well as allows the developer to focus on the high-level configuration. The optimizer determines unspecified parameters in the internal chip representation, focusing on two major regular structures: interconnects and arrays. The developer can set the frequency and bisection bandwidth of the network-on-chip, the capacity and the characteristics of caches, or the number banks. In the same way, letting the tool to determine the implementation details such as the choice of metal planes, the effective signal wiring pitch for the interconnect. As stated by (Li, et al., 2013), these optimizations lessen the burden on the architect to figure out every detail, and significantly lowers the learning curve to use the tool. Users always have the flexibility to turn off these features and set the circuit-level implementation parameters by themselves.

The focus of McPAT is to provide accurate power and area modeling, and a target clock rate is used as a design constraint. The software applies the optimization function to report the final power and area values to the developer configurations considering the power and area deviation. The module power and timing models together with the final chip representation generated by the optimizer are used to compute the final peak power.

The peak power of individual units and the machine utilization statistics are used to calculate the final runtime power dissipation.

5.5.4 Power Efficiency Results

TABLE 5.1 shows the total power consumption of each sample for 1, 2, 4 and 6 views 3D video in respectively 4, 8, 16 and 32 cores set up. For comparison, it is also shown power consumption on a baseline implementation without workload adaptation (without workload balance nor thread management). It is noticed that the workload adapter and power manager contribute significantly to power reduction. The power savings is given by a simple percentage difference between the power consumption by considering adaptive (dynamic voltage and frequency scaling) and non-adaptive workload balance.

Table 5.1: Power Consumption and Rate-Distortion Comparisons

Sequence	Cores	Views	Power [W]		Power Savings	PSNR [dB]	
			Adaptive	Non Adaptive		Adaptive	Non Adaptive
Poznan Hall	4	1	340.047	463.320	26.61%	37.93	38.12
	8	2	323.120	470.151	31.27%	38.01	38.47
	16	4	317.850	483.839	34.31%	37.54	38.09
	32	6	312.874	482.928	35.21%	37.53	37.92
Poznan Street	4	1	347.422	580.395	40.14%	36.93	37.78
	8	2	338.901	583.881	41.96%	37.14	37.80
	16	4	327.191	590.144	44.56%	37.16	38.01
	32	6	324.865	595.593	45.46%	37.50	38.19
GT_fly	4	1	350.567	578.847	39.44%	37.09	37.33
	8	2	349.488	580.993	39.85%	37.12	37.97
	16	4	345.451	596.666	42.10%	37.09	37.74
	32	6	340.629	593.454	42.60%	37.16	37.68
Kendo	4	1	348.005	678.330	48.70%	37.14	37.96
	8	2	347.594	682.035	49.04%	37.20	38.02
	16	4	342.221	690.920	50.47%	37.23	38.12
	24	6	340.109	697.726	51.25%	37.30	38.11
Ballons	4	1	351.186	595.131	40.99%	37.67	38.26
	8	2	349.931	613.943	43.00%	37.68	38.21
	16	4	346.627	625.007	44.54%	38.01	38.67
	32	6	340.440	640.066	46.81%	38.11	38.72

5.5.5 Time Complexity and Rate-Distortion

THE VIDEO QUALITY AND TIME COMPLEXITY FOR SELECTED SEQUENCES IS PRESENT IN

Table 5.2. Moreover, the proposed scheme reaches lower bitrate than the non-adaptive solution.

Table 5.2: PSNR and Time Complexity comparison.

Sequence	Adaptive		Non Adaptive	
	PSNR [dB]	Time [msec]	PSNR [dB]	Time (msec)
Kendo	38.46	80	38.62	102
Ballons	37.71	104	37.80	123
Poznan Hall	38.03	101	38.14	128
Poznan Street	38.16	106	38.25	122
GT_fly	38.11	110	38.22	135

AS SHOWN IN

Table 5.2, the proposed scheme in CTU level minimizes the error propagation along the prediction structure, since the future ME/DE have the closest quality references from neighborhood. Regarding the frame level case, the proposed scheme causes negligible losses in rate-distortion efficiency.

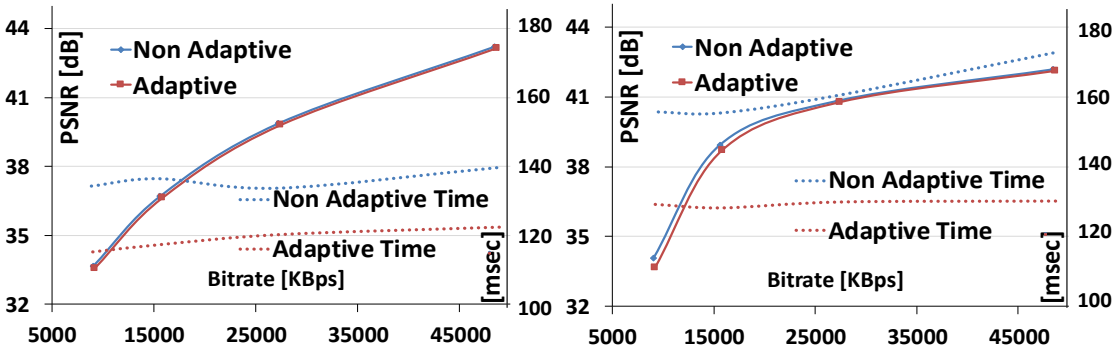

Figure 5.7 Time Complexity and Rate-distortion comparison for 4 different target bitrates of “Poznan Hall” sequence.

Figure 5.7 presents the comparison of time complexity and Rate-distortion for “Poznan Hall Sequence” encoded with 4 views in 16 cores by using adaptive workload balanced thread management and without using for four target bitrate.

Figure 5.8 presents a detailed analysis of core frequencies and γ for the "Poznan Hall" sequence in for four different views associated with 32 cores. The thread manager balances the workload, while the power manager accurately regulates the frequency.

Moreover, Figure 5.8 (e) present the time per core to encode 80 frames of 4 views FullHD "Poznan Hall" sequence.

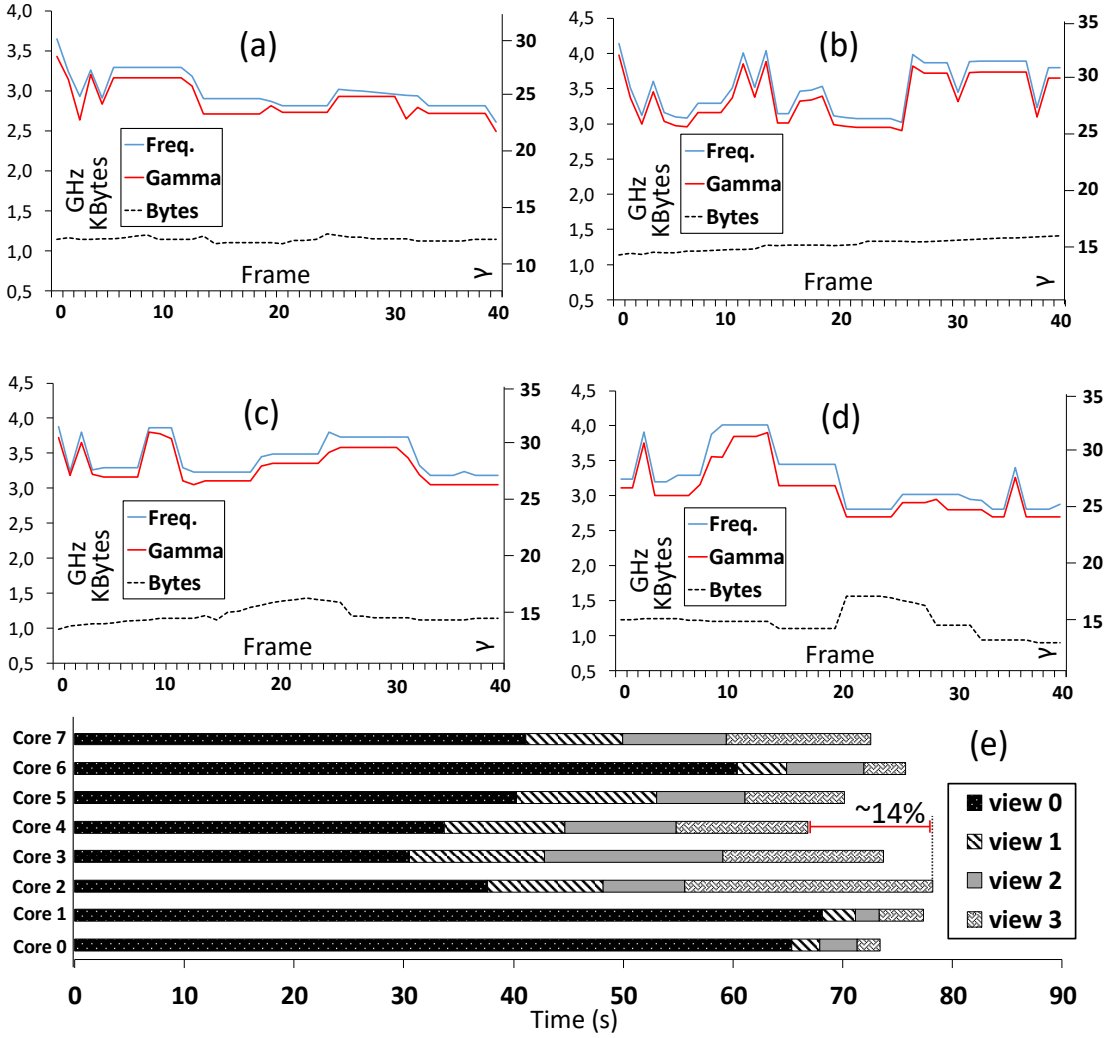

Figure 5.8 Bitrate, frequency and γ adaptation of CTUs in (a) base view (b) view 1 (c) view 2 and (d) view 3 of "Poznan Hall" Sequence encoded in 16 cores. (e) Time occupancy of 8 core encoding "Poznan Hall" with 4 views.

6 CONCLUSIONS

This thesis presents two main contributions focusing on three-dimensional videos. This work presents the challenges for process and transmits multiview videos over a restricted bandwidth. A comprehensive proposal shows Rate Control solution at multiple levels of processing. Furthermore, a strategy for workload balancing generated by a multi-view encoder processed on a multi-core platform, this solution includes a thread manager and a run-time power control. Both proposals based on an analysis of the execution flow of the multiview video coding standards.

The related work discussion pointed the drawbacks of current literature in the lack of 3D-oriented solutions, the incomplete exploration of 3D-neighborhood and the lack of joint consideration of algorithmic and system solution. Based on these points this work defines a clear statement about the need for a Rate Control scheme designed for the specific needs of multiview videos acting in fine and coarse grain level. Moreover, this proposal included an algorithm that takes into account the knowledge about the specific 3D application. This approach seems to be an efficient way to make feasible the accurate bitstream transmission of high definition multiview video coding. The results demonstrate the accuracy of prediction and the smooth visual quality delivery by the proposed scheme.

Along this thesis is presented a workload balancing mechanism with thread manager while using dynamic voltage and frequency scaling to deal with the complexity and power challenges of 3D video processing on multi-core platforms. This scheme exploits the spatial and disparity redundancy characteristics by using techniques to perform thread management that adapts to the video content. To reduce the power consumption while delivering high visual quality with minimizing rate-distortion, a strategy to dynamically scale core frequency and voltage considering the 3D video hierarchy is adopted. This scheme provides up to 51% power consumption savings along with less than 2% degradation in rate-distortion efficiency. The power consumption reduction along with minimal drops in the 3D encoding efficiency demonstrated the potential of the proposed scheme for the validation of dedicated hardware for 3D-video processing devices.

As future works, it is planned to include results of the Rate Control Scheme over the 3D- and MV-HEVC and comparisons with the state-of-the-art. Additionally, it is intended to include results of the workload manager over actual multi-core architectures used in mobile devices. This approach can give support to the overall process of coding multiview video in dedicated hardware for multi-core platforms.

This text summarizes the main results achieved along the Ph.D. in the *Programa de Pós-Graduação em Ciência da Computação* (PPGC) of the Federal University of Rio Grande do Sul (UFRGS), begun in 2012 (from 2010 Master extension), including the research developed during the internship at the Karlsruhe Institute of Technology (KIT - Germany), in 2015.

6.1 Publications

- **Bruno Boessio Vizzotto,** Bruno Zatt, Muhammad Shafique, Sergio Bampi, Jörg Henkel: Model Predictive Hierarchical Rate Control With Markov Decision Process for Multiview Video Coding. **IEEE Trans. Circuits System Video Technology**. 23(12): 2090-2104 (2013)

- **Bruno Boessio Vizzotto**, Volnei Mazui, Sergio Bampi: Area efficient and high throughput CABAC encoder architecture for HEVC **IEEE International Conference on Electronics, Circuits, and Systems** (ICECS), Cairo, 2015, pp. 572-575. doi: 10.1109/ICECS.2015.7440381

- Muhammad Shafique; Semin Rehman; Florian Kriebel; Muhammad U. K. Khan; Bruno Zatt; Arun Subramaniyan; **Bruno Boessio Vizzotto**; Jörg Henkel; Application-Guided Power-Efficient Fault Tolerance for H.264 Context Adaptive Variable Length Coding. **IEEE Trans. on Computers** ISSN: 0018-9340 (2016)

REFERENCES

3D-HEVC. 3D-HEVC Reference Software. **3D HEVC Extension**, 2014. Disponivel em: <https://www.hhi.fraunhofer.de/en/departments/vca/research-groups/image-video-coding/research-topics/3d-hevc-extension.html>. Acesso em: 11 set. 2014.

AGOSTINI, L. **Desenvolvimento de Arquiteturas de Alto Desempenho Dedicadas à Compressão de Vídeo Segundo o Padrão H.264/AVC**. [S.l.]: [s.n.], 2007.

AGRAFIOTIS, D. et al. **Multiple Priority Region of Interest Coding with H.264**. IEEE International Conference on Image Processing (ICIP). Atlanta, GA, USA: IEEE. 2006. p. 4.

BARTO, A. G. Reinforcement learning control. **Current Opinion in Neurobiology**, v. 4, p. 6, 1994.

BEHRENDT, M. **A basic working principle of Model Predictive Control**, 2009. Disponivel em: <http://en.wikipedia.org/wiki/File:MPC_scheme_basic.svg>. Acesso em: 22 out. 2012.

BELLMAN, R. A Markovian Decision Process. **OTS The Rand**, Santa Monica, CA, USA, p. 15, April 1957.

BOSSEN, F. et al. HEVC Complexity and Implementation Analysis. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 22, p. 12, December 2012.

BUGDAVI, M.; SZE, V.; MINHUA, Z. **HEVC ALF decode complexity analisys and reduction**. 18th International Conference on Image Processing (ICIP). Brussels: IEEE. 2011. p. 4.

CARLSON, T. E.; HEIRMAN, W.; EECKHOUT, L. **Sniper:** exploring the level of abstraction for scalable and accurate parallel multi-core simulation. International Conference for High Performance Computing, Networking, Storage and Analysis. New York: ACM. 2011. p. 12.

CHEN, Y. et al. The Emerging MVC Standard for 3D Video Services. **EURASIP Journal on Advances in Signal Processing**, v. 2009, n. 13, p. 13, 2009.

CORREA, G. et al. Complexity control of high efficiency video encoders for power-constrained devices. **IEEE Transactions on Consumer Electronics**, v. 57, p. 8, November 2011.

DENG, Z.-P. et al. **A Fast View-Temporal Prediction Algorithm for Stereoscopic Video Coding**. International Congress on Image and Signal Processing (CISP). Tiajin, China: IEEE. 2009. p. 5.

GARCIA, C.; PRETT, D.; MORARI, M. Model predictive control: theory and practice - a survey, v. 25, 1989.

HAGER, W.; PARDALOS, P. **Optimal Control**. 1. ed. [S.l.]: Springer, v. 15, 1998.

HEVC-SOFTWARE. 3D-HEVC reference software. Disponivel em: <https://hevc.hhi.fraunhofer.de/3DHEVC>. Acesso em: 03 ago. 2015.

JIANG, M.; YI, X.; LING, N. **Improved frame-layer rate control for H.264 using MAD ratio**. International Symposium on Circuits and Systems (ISCAS). Vancouver, BC, Canada: IEEE. 2004. p. 4.

JMVC-SOFTWARE. **Multiview Video Coding References**, 2012. Disponivel em: <http://h264.hhi.fraunhofer.de/mvc>. Acesso em: 20 ago. 2012.

JVT. **JVT-G050 - Draft ITU-T Rec. and final draft international standard of joint video specification**. Joint Video Team. [S.l.], p. 12. 2003.

JVT. **JVT-AB204 - Joint Draft 8.0 on Multiview video coding**. Joint Video Team. [S.l.], p. 14. 2009.

KANG, J.-W. et al. **Low complexity Neighboring Block based Disparity Vector Derivation in 3D-HEVC**. IEEE International Symposium on Circuits and Systems (ISCAS). Melbourne, VIC, Australia: IEEE. 2014. p. 4.

KAUFF, P. et al. Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability. **Signal Processing: Image Communication**, v. 22, p. 18, February 2007.

KHAN, M. U. K. et al. **"Hardware-Software Collaborative Complexity Reduction Scheme for the Emerging HEVC Intra Encoder**. Design, Automation & Test in Europe Conference & Exhibition (DATE). Grenoble, France: IEEE. 2013. p. 4.

KIM, Y.; KIM, J.; SOHN, K. Fast Disparity and Motion Estimation for Multi-view Video Coding. **IEEE Transactions on Consumer Electronics**, v. 53, n. 2, p. 8, July 2007.

KWON, D.-K.; SHEN, M.-Y.; JAY KUO, C.-C. Rate Control for H.264 Video With Enhanced Rate and Distortion Models. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 17, p. 13, April 2007.

LEE, P.-J.; LAI, Y.-C. **Vision perceptual based rate control algorithm for multi-view video coding**. International Conference on System Science and Engineering. Macao, China: IEEE. 2011. p. 4.

LI, S. et al. **McPAT:** An Integrated Power, Area, and Timing Modeling Framework for Multicore and Manycore Architectures. 2009. MICRO-42. 42nd Annual IEEE/ACM International Symposium on Microarchitecture. New York, NY, USA: ACM Transactions on Architecture and Code Optimization. 2009. p. 12.

LI, Z.; PAN, K.; LIM, P. **Adaptive basic unit layer rate control for JVT - JVT-G012**. Thailand, p. 16. 2003.

LIE, W.-N.; LIAO, Y.-P. **Rate control technique based on 3D quality optimization for 3D video encoding**. IEEE International Conference on Image Processing (ICIP). Paris, France: IEEE. 2014. p. 4.

LIU, A. et al. Just Noticeable Difference for Images With Decomposition Model for Separating Edge and Textured Regions. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 20, p. 5, October 2010.

MA, S.; GAO, W.; LU, Y. Rate-distortion analysis for H.264/AVC video coding and its application to rate control. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 15, p. 12, December 2005.

MERKLE, P. et al. Efficient Prediction Structures for Multiview Video Coding. **IEEE Transactions on Circuits and Systems for Video technology**, v. 17, n. 11, p. 1461-1473, Novembro 2007.

MERRIT, L.; VANAM, R. **Improved Rate Control and Motion Estimation for H.264 Encoder**. IEEE International Conference on Image Processing (ICIP). San Antonio, TX, USA: IEEE. 2007. p. 4.

MIANO, J. **Compressed Image File Formats:** Jpeg, Png, Gif, Xbm, Bmp. 1st. ed. Boston: ACM Press, v. I, 1999.

MORARI, M.; LEE, H. Model Predictive Control: Past, Present and Future. **Computers & Chemical Engineering**, v. 23, n. Elsevier, p. 16, May 1997.

MÜLLER, K. et al. 3D High-Efficiency Video Coding for Multi-View Video and Depth Data. **IEEE Transactions on Image Processing**, v. 22, p. 3366-3378, September 2013.

PARK, S.; SIM, D. **An efficienct rate-control algorithm for multi-view video coding**. IEEE 13th International Symposium on Consumer Electronics, 2009. ISCE '09. Kyoto, Japan: IEEE. 2009. p. 115-118.

POURAZAD, M.; NASIOPOULOS, P.; WARD, R. **An Efficient Low Random-Access Delay Panorama-Based Multiview Video Coding Scheme**. IEEE Conference on Image Processing. Cairo: IEEE. 2009. p. 2945-2948.

POURAZAD, M.; NASIOPOULOS, P.; WARD, R. **An Efficient Low Random-Access Delay Panorama-Based Multiview Video Coding Scheme**. IEEE Conference on Image Processing. Cairo: IEEE. 2009. p. 2945-2948.

RICHARDSON, I. **The H. 264 advanced video compression standard**. 2nd. ed. [S.l.]: John Wiley and Sons, 2010.

RUSANOVSKYY, D.; MÜLLER, K.; VETRO, A. **Common test conditions for 3DV Core Experiments - JCT3V-G1100**. Geneva, Switzerland, p. 7. 2013.

SANCHEZ, G. et al. **Complexity reduction for 3D-HEVC depth maps intra-frame prediction using simplified edge detector algorithm**. International Conference on Image Processing (ICIP). Paris, France: IEEE. 2014. p. 3209-3213.

SHAFIQUE, M.; MOLKENTHIN, B.; HENKEL, J. **An HVS-based Adaptive Computational Complexity Reduction Scheme for H.264/AVC video encoder using Prognostic Early Mode Exclusion**. Design, Automation &

Test in Europe Conference & Exhibition (DATE). Leuven, Belgium: ACM. 2010. p. 1713-1718.

SHEN, L. et al. View-Adaptive Motion Estimation and Disparity Estimation for Low Complexity Multiview Video Coding. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 20, p. 925-930, March 2010.

SIVAN, R.; KWAKERNAAK, H. **Linear Optimal Control Systems**. 1st. ed. [S.l.]: John Wiley & Sons, Inc, v. I, 1972.

SMOLIC, A. et al. Coding Algorithms for 3DTV - A Survey. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 17, n. 11, p. 1606-1621, Novembro 2007.

SONG, Y.; JIA, K.; WEI, Z. **Improved LCU Level Rate Control for 3D-HEVC**. Visual Communications and Image Processing (VCIP). Changdu, China: IEEE. 2016.

STOYKOVA, E. et al. 3-D Time-Varying Scene Capture Technologies: A Survey. **IEEE Transactions on Circuits and Systems for Video Technology** , v. 17, n. 11, p. 1568-1586, October 2007.

SU, T. et al. **A DASH-based 3D multi-view video rate control system**. 8th International Conference on Signal Processing and Communication Systems (ICSPCS). Gold Coast, QLD, Australia: IEEE. 2014.

SU, Y.; VETRO, A.; SMOLIC, A. **Common Test Conditions for Multiview Video Coding - JVT-T207**. Klagenfurt, Austria. 2006.

SULLIVAN, G. J. et al. Overview of the High Efficiency Video Coding (HEVC) Standard. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 22, p. 1649-1668, December 2012.

SULLIVAN, G. J.; WIEGAND, T. Video Compression - From Concepts to the H.264/AVC Standard. **Proceedings of the IEEE**, v. 93, n. 1, p. 18-31, January 2005.

SULLIVAN, G.; WIEGAND, T. Rate-Distortion Optimizatoin for Video Compression. **IEEE Signal Processing Magazine**, v. 15, p. 74-90, November 1998.

TAN, K. T.; SULLIVAN, G.; WEDI, T. **Recommended Simulation Conditions for Coding Efficiency Experiments - VCEG-AE010**. Marrakech, Morocco. 2005.

TAN, S. et al. Inter-View Dependency-Based Rate Control for 3D-HEVC. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 27, n. 2, p. 337-351, February 2017.

TATJEWSKI, P. Supervisory predictive control and on-line set-point optimization. **International Journal of Applied Mathematics and Computer Science**, v. 20, p. 483-495, March 2010.

TIAN, L.; SUN, Y.; ZHOU, Y. **Analysis of quadratic R-D model in H.264/AVC video coding**. 17th IEEE International Conference on Image Processing (ICIP). Hong Kong, China: IEEE. 2010. p. 2853-2856.

UGUR, K. et al. High Performance, Low Complexity Video Coding and the Emerging HEVC Standard. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 20, p. 1688-1697, December 2010.

VIZZOTTO, B. et al. **A Model Predictive Controller for Frame-Level Rate Control in Multiview Video Coding**. IEEE International Conference on Multimedia and Expo (ICME). Melbourne, Australia: IEEE. 2012. p. 485-490.

VIZZOTTO, B. et al. Model Predictive Hierarchical Rate Control With Markov Decision Process for Multiview Video Coding. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 23, p. 2090-2104, December 2013.

WIEGAND, T. et al. Overview of the H.264/AVC Video Coding Standard. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 13, n. 7, p. 560-576, Julho 2003.

WU, C.-Y.; SU, P.-C. **A Region of Interest Rate-Control Scheme for Encoding Traffic Surveillance Videos**. International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP). Kyoto, Japan: IEEE. 2009.

XU, L. et al. **Priority pyramid based bit allocation for multiview video coding**. IEEE Visual Communications and Image Processing (VCIP). Tainan, Taiwan: IEEE. 2011.

YAN I, T. et al. **Frame-layer rate control algorithm for multi-view video coding**. ACM/SIGEVO Summit on Genetic and Evolutionary Computation (GEC). New York, NY, USA: ACM. 2009. p. 1025-1028.

YAN, T. et al. **Rate Control Algorithm for Multi-View Video Coding Based on Correlation Analysis**. Symposium on Photonics and Optoelectronics. Wuhan, China: IEEE. 2009.

ZATT, B. et al. **Memory Hierarchy Targeting Bi-Predictive Motion Compensation for H.264/AVC Decoder**. IEEE Computer Society Annual Symposium on VLSI (ISVLSI). Porto Alegre, Brazil: IEEE. 2007. p. 445 - 446.

ZATT, B. et al. **A Multi-Level Dynamic Complexity Reduction Scheme for Multiview Video Coding using 3D-Neighborhood Correlation**. Design, Automation and Test in Europe (DATE). Brussels, Belgium: IEEE. 2010.

ZATT, B. et al. **An Adaptive Early Skip Mode Decision Scheme for Multiview Video Coding**. Picture Coding Symposium (PCS). Nagoya, Japan: IEEE. 2010. p. 42-45.

ZHANG, S.; ZHANG, X.; GAO, Z. **Implementation and improvement of Wavefront Parallel Processing for HEVC encoding on many-core platform**. International Conference Multimedia and Expo Workshops (ICMEW). Chengdu, China: IEEE. 2014.

ZHANG, Z. et al. A New Rate Control Scheme For Video Coding Based On Region Of Interest. **IEEE Access**, v. PP, n. 99, March 2017.

ZHENG, T. **Multi-objective nonlinear model predictive control:** Lexicographic method. Shanghai, China: SCIYO ISBN 978-953-307-102-2, 2010.

ZHOU, Y. et al. PID-Based Bit Allocation Strategy for H.264/AVC Rate Control. **IEEE Transactions on Circuits and Systems II: Express Briefs**, v. 58, p. 184-188, March 2011.