# Multiview coding and compression for 3D video

Seif Allah El Mesloul Nasri

BADJI MOKHTAR- ANNABA UNIVERSITY, ALGERIA

Faculty of Engineering Sciences
Department of Electronics

This thesis is submitted for the degree of
*Doctor of Philosophy*

Title

# Multiview coding and compression for 3D video

Option: Multimedia and Digital Communications

By

Seif Allah El Mesloul Nasri

Supervisor: Prof. Khaled Khelil — University of Souk Ahras

Co-supervisor: Prof. Noureddine Doghmane — University of Annaba

**In Front of the JURY**

| | | |
|---|---|---|
| President: | Prof. Mohamed Fezari | University of Annaba |
| EXAMINERS: | Prof. Abdul Hamid Sadka | Brunel University London, UK |
| | Dr. Saliha Harize | Université de Annaba |
| | Dr. Boukari Karima | Université de Annaba |
| Invited: | Dr. Amara Bekhouch | University of Souk Ahras |

1st July 2018

وَقُلْ رَبِّي ارْحَمْهُمَا كَمَا رَبَّيَانِي صَغِيرًا

To my mother, my father
and my brothers.

# Abstract

Multiview video (MVV) is an advanced representation of 3D video technology. The MVV content depends on several parameters such as the deployed cameras number, recording angles and the captured scene. Each used camera inherently generates an extra amount of data compared to the 2D conventional video that only needs one camera. MVV needs specific coding techniques that take into consideration the visual similarities between the viewpoints set. Both temporal and interview correlations could be exploited to improve the compression ratios. Multiview video coding (MVC) is the extended profile of H.264/AVC video codec. MVC offers compression efficiency improvements achieving 50 % over simulcast video coding. However, better compression performance leads to higher random access complexity which might hamper MVC usage in applications such as Free viewpoint Television and 3D TV broadcasting. It also degrades the viewers quality of experience due to the lack of interactivity.

This thesis aims to study MVC encoding and to solve the aforementioned problem of Random Access (RA) ability by proposing faster interview prediction approaches. The thesis first presents a brief description of the 3D video concepts from capturing to displaying. Furthermore, it focuses on video coding fundamentals and its multiview extension form. The thesis then proposes two novel techniques to enhance the random access ability of the MVC encoder. The first proposed approach (PBI) aims to lower the cost of randomly accessing any picture at any position and instant, with respect to the multiview reference model JMVM and other state-of-the-art methods.

The proposed (PBI) scheme is mainly based on the use of two base views (I-views) in the interview structure with selected positions instead of only a single reference view as in the standard structure. This provides a direct interview prediction for the remaining views and ensures faster random access ability while maintaining a competitive compression performance. PBI achieves a random access gain of 20 % relative to the reference model MVC. The second proposed approach (PIP) surpasses PBI structure by achieving a random access gain of 53.33% compared to the benchmark standard MVC.

A novel random access ability evaluation method ($G_R$) has been suggested and adopted throughout the thesis experimental parts. It allows more accurate assessment by considering all pictures types of the tested multiview schemes. A comparative investigation of proposed and reported multiview schemes is also presented in this thesis. Results of the conducted tests allow classifying the examined multiview schemes and GOP sizes preferences according to their effects in terms of random access ability and compression efficiency. Finally, more experiments have been conducted comparing MVC and MV-HEVC standards in terms of compression efficiency using different multiview video sequences that have different textures and resolutions.

**Key Words** : Multiview video , 3D video, video coding, MVC, MV-HEVC, compression efficiency, random access.

# ملخص

الفيديو متعدد المناظر (**ف.م.م**، **MVV**) يمثل صيغة متقدمة لتكنولوجيا الفيديو الثلاثي الأبعاد. الفيديو متعدد المناظر يوفر تجربة ثلاثية أبعاد غنية مقارنة بالفيديو ثلاثي الابعاد ذو المنظرين فقط. يتعلق محتوى **ف.م.م** بعدة متغيرات من بينها عدد الكاميرات المستخدمة، زويا التصوير إضافة إلى المشهد المُصوّر. ينتج عن كل كاميرا مستخدمة محتوى معلومات إضافي بالمقارنة مع الفيديو الثنائي الأبعاد التقليدي. ال **ف.م.م** يحتاج الى تقنيات ضغط وترميز خاصة تأخذ بعين الاعتبار التشابه الموجود بين مجموعة المناظر أو الكاميرات المستخدمة. يمكن استغلال كلٌ من التشابه الزمني والجهوي لتحسن كفاءة الضغط. يمثل ضاغط الفيديو متعدد المناظر **MVC** الامتداد النموذجي لضاغط الفيديو **H.264**، حيث يوفر كفاءة ضغط أفضل تصل الى نسبة 50 بالمئة بالمقارنة مع **H.264** المنفرد. رغم ذلك، كفاءة ضغط أفضل يمكن أن تقودنا الى تعقيد أكثر في خاصية الوصول العشوائي مما قد يعيق استخدام **MVC** في تطبيقات مثل **FVT** والارسال المباشر عبر **3DTV** . كما قد يخفض ذلك من جودة التجربة للمشاهدين بما أن خاصية التفاعل أبطأ. تهدف هذه الرسالة الى دراسة ال **MVC** بطريقة معمقة وحل الإشكالية السابق ذكرها بخصوص قدرة الوصول العشوائي عبر اقتراح مقاربات هياكل تنبؤ أسرع. تهدف المقاربة الأولى المقترحة (**PBI**) إلى خفض تكلفة الوصول العشوائي إلى أي صورة في أي موقع ولحظة مقارنة مع النموذج المرجعي **JMVM** ونماذج حديثة أخرى. يستند المخطط **PBI** بالأساس إلى استخدام جهتين مرجعيتين (I-views) محددتي التموقع ضمن هيكلة ما بين المناظر، هذا مقابل استخدام جهة مرجعية واحدة في النموذج الأصلي. يتيح هذا تنبؤا مباشرا للجهات المتبقية ويضمن قدرة وصول عشوائي بأقل تكلفة مع الحفاظ على معدل ضغط منافس. مقارنة مع هيكلة **MVC** الأصلية، تحقق **PBI** ربحا في الوصول العشوائي بمقدار 20 بالمئة بالنسبة. تتفوق المقاربة الثانية المقترحة **PIP** على هيكلة **PBI** وتحرز ربحا بمقدار 53.33 بالمئة بالمقارنة مع هيكلة **MVC** الأصلي. تم اقتراح واعتماد طريقة جديدة $G_R$ لتقييم قدرة الوصول العشوائي عبر الأجزاء التجريبية للرسالة. تسمح هذه الطريقة بتقييم أكثر دقة حيث تأخذ بعين الاعتبار جميع أنواع الصور الموجودة ضمن الهياكل المتعددة المناظر المدروسة. كما تعرض الرسالة مقارنة استقصائية بين هياكل تنبؤ متعددة المناظر مقترحة ومنقولة. نتائج هذه التجارب تسمح بترتيب الهياكل المدروسة إضافة الى خيارات حجوم GOP بالتناسب مع تأثيراتها على قدرة الوصول العشوائي وكفاءة الضغط. أخيرا، أجريت المزيد من التجارب للمقارنة بين MVC و MV-HEVC فيما يتعلق بكفاءة الضغط، باستخدام فيديوهات متعددة مناظر لها أشكال ومقاييس متباينة.

**الكلمات المفتاحية:** الفيديو المتعدد المناظر، الفيديو ثلاثي الأبعاد، ضغط الفيديو، فعالية الضغط، الوصول العشوائي

# Résumé

La vidéo multi-vues (MVV) est une représentation avancée de la technologie vidéo 3D. Le contenu MVV dépend de plusieurs paramètres tels que le nombre de caméras déployées, les angles d'enregistrement et la scène capturée. Chaque caméra utilisée génère intrinsèquement une quantité de données supplémentaire par rapport à la vidéo classique en 2D qui ne nécessite qu'une seule caméra. Le MVV nécessite des techniques de codage spécifiques qui prennent en compte les similitudes visuelles entre les points de vue définis. Les corrélations temporelles et d'interview pourraient être exploitées pour améliorer les taux de compression. Le codage vidéo multi-vues (MVC) est le profil étendu du codec vidéo H.264/AVC. Le MVC offre des améliorations de l'efficacité de la compression atteignant 50 % par rapport au codage vidéo en diffusion simultanée. Toutefois, de meilleures performances de compression entraînent une plus grande complexité d'accès aléatoire qui pourrait entraver l'utilisation du MVC dans des applications telles que la télévision à point de vue libre et la diffusion de télévision en 3D. Elle dégrade également la qualité de l'expérience des téléspectateurs en raison du manque d'interactivité. Cette thèse vise à étudier l'encodage MVC et à résoudre le problème susmentionné de la capacité d'accès aléatoire (RA) en proposant des approches plus rapides de prédiction des interviews. La thèse présente d'abord une brève description des concepts de la vidéo 3D, de la capture à l'affichage. De plus, elle se concentre sur les principes fondamentaux du codage vidéo et sa forme d'extension multi-vues. La thèse propose ensuite deux nouvelles techniques pour améliorer la capacité d'accès aléatoire de l'encodeur MVC. La première approche proposée (PBI) vise à réduire le coût de l'accès aléatoire à n'importe quelle image, à n'importe quelle position et à n'importe quel instant, par rapport au modèle de référence multi-vues JMVM et à d'autres méthodes de pointe. Le schéma proposé (PBI) est principalement basé sur l'utilisation de deux vues de base (I-vues) dans la structure d'interview avec des positions sélectionnées au lieu d'une seule vue de référence comme dans la structure standard. Cela permet de prévoir directement les entretiens pour les autres vues et d'assurer un accès aléatoire plus rapide tout en maintenant une performance de compression compétitive.

La PBI permet d'obtenir un gain d'accès aléatoire de 20 % par rapport au modèle de référence MVC. La deuxième approche proposée (PIP) surpasse la structure PBI en réalisant un gain d'accès aléatoire de 53,33 % par rapport au modèle MVC standard de référence. Une nouvelle méthode d'évaluation de la capacité d'accès aléatoire (GR) a été suggérée et adoptée tout au long des parties expérimentales de la thèse. Elle permet une évaluation plus précise en considérant tous les types d'images des schémas multi-vues testés. Une étude comparative des schémas multi-vues proposés et rapportés est également présentée dans cette thèse. Les résultats des tests effectués permettent de classer les schémas multi-vues examinés et les préférences de taille des GOP en fonction de leurs effets en termes de capacité d'accès aléatoire et d'efficacité de la compression. Enfin, d'autres expériences ont été menées pour comparer les normes MVC et MV-HEVC en termes d'efficacité de compression en utilisant différentes séquences vidéo multi-vues qui ont des textures et des résolutions différentes.

**Mots clés** : Vidéo multi-vues, vidéo 3D, codage vidéo, MVC, MV-HEVC, efficacité de la compression, accès aléatoire.

# Acknowledgement

I would like to express my sincere thanks and ultimate gratitude to my mother Lynda and my father Nasser, who dedicated great efforts to enable me to attain this level of education, no nothing would have been achievable without their love and support. Many thanks also to my two brothers, Amine and Zoheir, for their support and encouragement.

I would like to acknowledge and thank my academic supervisor Prof. Khaled Khelil for providing me the finest supervision, guidance and advices during my PhD research period.

My great gratitude goes to my co-supervisor Prof. Noureddine Doghmane who introduced me the research world and taught me throughout ten academic years, since my bachelor's degree. His supervision, knowledge and visionary advices are sincerely appreciated.

My Special thanks and respect go to Prof. Abdul Hamid Sadka for all kinds of support he provided me during my academic stay in his centre for Media Communications Research (CMCR)at Brunel University. He was an outstanding academic mentor and a unique wise advisor for research and beyond.

I take this opportunity also to acknowledge my appreciation to members of CMCR and CEPROQHA project for the exceptional research experience that I had with them.

Many thanks for all my colleagues and friends of the LASA laboratory at Annaba University, for the fruitful discussions and the prestigious memories that we had.

I am also thankful to Prof. Mohamed Fezari, Prof. Abdul Hamid Sadka, Dr. Saliha Harize and Dr Boukari Karima for kindly accepting to examine and review this thesis.

Last but not least, I would like to thank my grandparents, my big family members, all my teachers, all Akhlak Madina project members and all my friends. Many thanks for your encouragements and prayers.

# List of tables

# List of Figures

# List of abbreviations

| | |
|---|---|
| **2D** | Two-dimensional |
| **3D** | Three-dimensional |
| **3DTV** | Three-dimensional Television |
| **AVC** | Advanced Video Coding |
| **AU** | Access Unit |
| **CABAC** | Context-based Adaptive Binary Arithmetic Coding |
| **CAVLC** | Context-based Adaptive Variable Length Coding |
| **CIF** | Common Intermediate Format |
| **DC** | Disparity Compensation |
| **DCT** | Discrete Cosine Transform |
| **FPS** | Frame Per Second |
| **FVV** | Free Viewpoint Video |
| **GGOP** | Group of Group Of Pictures |
| **GOP** | Group Of Pictures |
| **HBP** | Hierarchical B Pictures |
| **HD** | High Definition |
| **HEVC** | High Efficiency Video Coding |
| **HVS** | Human Visual System |
| **IDR** | Instantaneous Decoder Refresh |
| **ISO** | International Organisation for Standardisation |
| **ITU-T** | International Telecommunication Union – Telecommunications Standardization Sector |

| | |
|---|---|
| **JMVC** | Joint Multiview Video Coding |
| **JMVM** | Joint Multiview Video Model |
| **JTC** | Joint Technical Committee |
| **JVT** | Joint Video Team |
| **MB** | Macroblock |
| **MC** | Motion Compensation |
| **ME** | Motion Estimation |
| **MERL** | Mitsubishi Electric Research Laboratories |
| **MOS** | Mean Opinion Score |
| **MPEG** | Moving Pictures Experts Group |
| **MSE** | Mean Squared Error |
| **MV** | Motion Vector |
| **MVC** | Multiview Video Coding |
| **MVV** | Multiview Video |
| **NAL** | Network Abstraction Layer |
| **PSNR** | Peak Signal to Noise Ratio |
| **QP** | Quantisation Parameter |
| **RD** | Rate-Distortion |
| **RDO-MD** | Rate Distortion Optimized Mode Decision |
| **SD** | Standard Definition |
| **SR** | Search Range |
| **SVC** | Scalable Video Coding |
| **VCL** | Video Coding Layer |

# Contents

# Chapter 1
# Introduction

## 1.1 Background and motivation

Recording and transmitting visual information was always a major issue for humanity. Since the prehistoric time, before 40,000 years, human ancestors used cave painting with primitive drawing kits for recording important events, communicating with one another and jumping imagination borders. Painting and sculpting were the only means to record and preserve visual information for an extended historical period.

Development in mathematics, geometrics and optics sciences engendered the actual different types of digital cameras, notably, works of Ibn Al-Haytham or Alhazen in *Latin* (965-1040 AD), one of the optics founders, who first identified the basic principles underlying the modern camera in his manuscript "book of optics" [1].

Nowadays, we are witnessing an exponential advancement in communication and information technologies including digital video communication. It is estimated that the video traffic on the internet will occupy 82 percent of all transmitted data by 2021 [2].

Rising demand for more advanced and interactive video technology has also known a fast growth. 3D video technology is one important type of the advanced video technologies.

3D video end-users benefit from an immersive visual experience including advanced approximation to the real perception, more low-level features, depth sensation and motion parallax.

The broad adoption of 3D video technology can be measured through the rising sales of the 3D display and the increasing number of the 3D cinema screens. Though, 3D video technology is not only dedicated to entertainment and leisure. 3D video technology is also applied in critical domains such as education, surveillance, healthcare issues and cultural heritage preservation. In fact, 3D imaging [3][4] has multiple systems; each one is characterised by its specific capturing, coding and visualisation techniques. 3D imaging systems [5] can be divided into three main categories, briefly described in Figure 1.1.

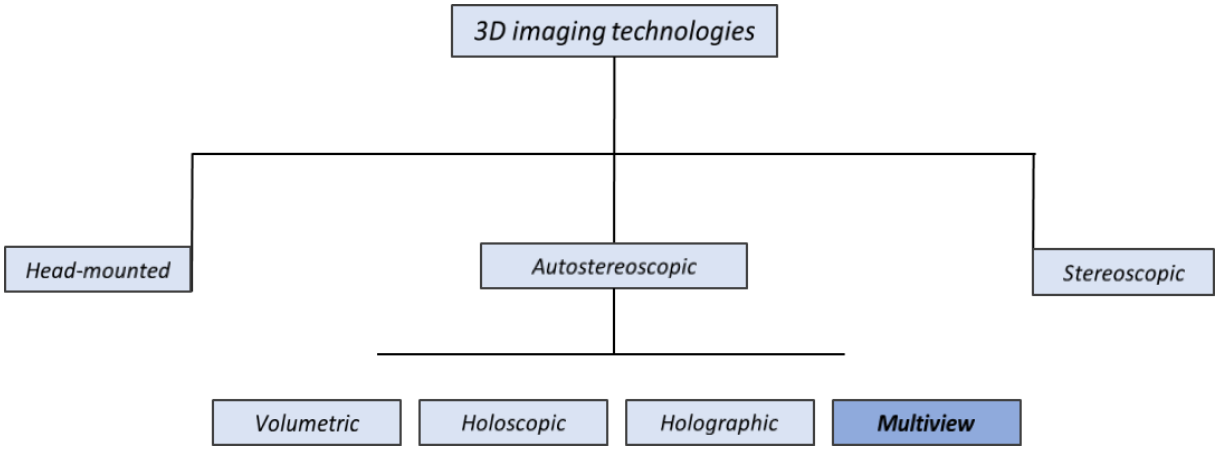



Figure 1.1 Diagram of 3D imaging systems

A. Stereoscopic 3D (S3D) imaging [6]: S3D is the simplest and normalised system that only needs two images to produce a 3D perception. It mimics human eyes perception by using left and right cameras. Depth effect is generated from the slightly different angle of the two cameras. The viewer has to wear special glasses to separate the mixed stream and perceive the scene' depth. There are three types of glasses corresponding to the used technique to mix the right and left images: Anaglyph, polarised and shutter glasses [7]. Apart from that viewers are obliged to wear glasses, S3D technology has other drawbacks such as causing stress and eye fatigue if watching for an extended period [8], in addition to its fixed and narrow view angle.

B. Head-mounted system: Head-mounted displays (HMDs) or Head-worn displays (HWDs) [9] are designed as on-body devices and coupled with the human eyes to support mobile users. Binocular HMDs offer a 3D immersive experience that is generally used for civilian and military training in addition to medicine, sport and gaming. HMDs highly support virtual reality applications. Other types of HMDs provide a see-through feature for augmented or mixed reality applications (Figure 1.2). However, wearing a head helmet could not be suitable for everyone and may cause eye and brain fatigue if used for a long period.

Figure 1.2 Mixed reality (MR) application in medicine of the HMD [10]

C. Autostereoscopic imaging systems: offer viewers a glasses-free 3D experience and ensure depth sensation as well as motion parallax. There are several types of 3D autostereoscopic systems from which we briefly mention four leading technologies:

- Volumetric 3D technology: Permits to spatially reconstructing recoded images or videos in a transparent display volume which generally takes a spherical form [11]. Volumetric 3D systems, mostly use a rotating projector and series of liquid crystal or gas panels to generate three-dimensional images spread around 360° [12]. Volumetric 3D systems are still expensive and complicated to manufacture which limits their adoption to only certain applications such as military and health domains.
- Holographic 3D technology: offers true 3D perceptions of the reconstructed scene in an open space. Basically, holography technology enables recording the scattered light field of objects and reproduces it later in a 3D space [13].

Holography has several techniques [14] depending on the applications domain which includes arts, medicine and military. However, holography is still unaffordable in the market due to its complex techniques and high cost.

- Holoscopic 3D imaging (H3D): also knowing as integral imaging and/or light field imaging, H3D uses a specific single aperture camera to capture the 3D information. Microlens array (MLA) is added to the optical components of a 2D camera to acquire the scene light field [15]. MLA is composed of tiny microlenses where each one allows to record a micro image from a slightly different perspective. The recorded micro-images are processed to extract different types of formats including 2D images, stereoscopic images and multiview images [16]. Although H3D capturing process is as comfortable as shooting with a typical 2D camera, extraction is highly complex and takes a long time during the post-processing stage. In addition, the quality of the resulted images still requires further enhancements.

- Multiview Video (MVV) technology: Multiview video is generated when a set of synchronised and carefully calibrated cameras capture the same scene from different perspectives. The number of the adjacent cameras and their geometric arrangement depends on the targeted application. Figure 1.3 shows an example of a multi-camera setup used in the production of multiview videos.

Figure 1.3 Multiview video acquisition example [17]

Multiview video can be displayed on: 1) 3D stereoscopic (3DS) displays where eye-glasses are required to feel the depth sensation, or 2) Autostereoscopic multiview displays (AMDs) where the 3D effect is perceived without using any eye-glasses. AMDs offer broader view angles and horizontal motion parallax compared to 3DS. Quality of autostereoscopic 3D experience depends on the deployed display technologies and their capabilities [18][19].

Multiview video offers an enriched 3D experience that qualifies it to successfully compete with the aforementioned 3D technologies in many application fields such as gaming, 3D cinema, education, surveillance, cultural heritage and medicine. However, representation of raw multiview videos needs an enormous amount of data. If no compression techniques are applied, storage or transmission of MVVs could be difficult or even impossible with the conventional storage devices and bandwidth capabilities. For example, if an MVV sequence of eight cameras with a frame size of 1024×768 is transmitted over a network at a rate of 30 frames/second, then about 4.5 Gbit/s (=1024×768×3×8×8×30) of bandwidth is required. In fact, the highest average connection speed in the world, as marked in the 2017 state of the internet report [20], is equal to 26.8 Mbps (recorded in South Korea), which is still too low to meet the increasing demands of 3D/HD videos. Innovative and advanced compression techniques must be deployed to overcome the current networks and storage devices limitations.

Video coding and compression techniques [21]considerably reduce the amount of data while preserving an excellent visual quality. Video compression exploits the redundancies that exist within and between the video frames.

Since multiple cameras are used to capture the same scene from different locations to produce MMVs, there will be indeed similarities between the adjacent views. Exploiting these interview similarities reduces the massive amount of data which MVV generates. Thereby, multiview video codecs are proposed as fundamental techniques to efficiently compressing and adapting MVV for storage and transmission over different networks. Recent video coding standards such as H.264 [22] and H.265 [23], provide extended profiles which take advantage of the interview redundant information in MVVs.

Apart from compression efficiency, multiview video coding must meet a requirements list to ensure a decent 3D service for MVV users. Random access ability is one of the vital requirement that minimises the coding complexity and improves users' interactivity with the 3D content.

## 1.2 Thesis scope

As presented in the previous section, there are ample varieties of 3D videos technologies that are currently facing challenges and attracting researchers' interests. Multiview video technology is a broad research field composed of different parts similar to other video communication systems. However, this thesis deals exclusively with multiview video technology as a study case.

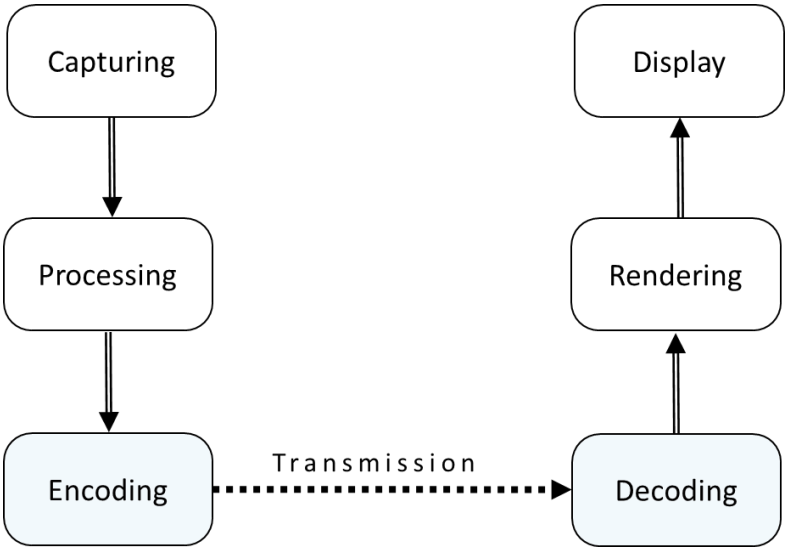



Figure 1.4 Typical blocks of an MVV communication system

From producing MVV content to delivering and displaying it for end-users, all blocks that appear in Figure 1.4 are in fact active research fields in MVV systems.

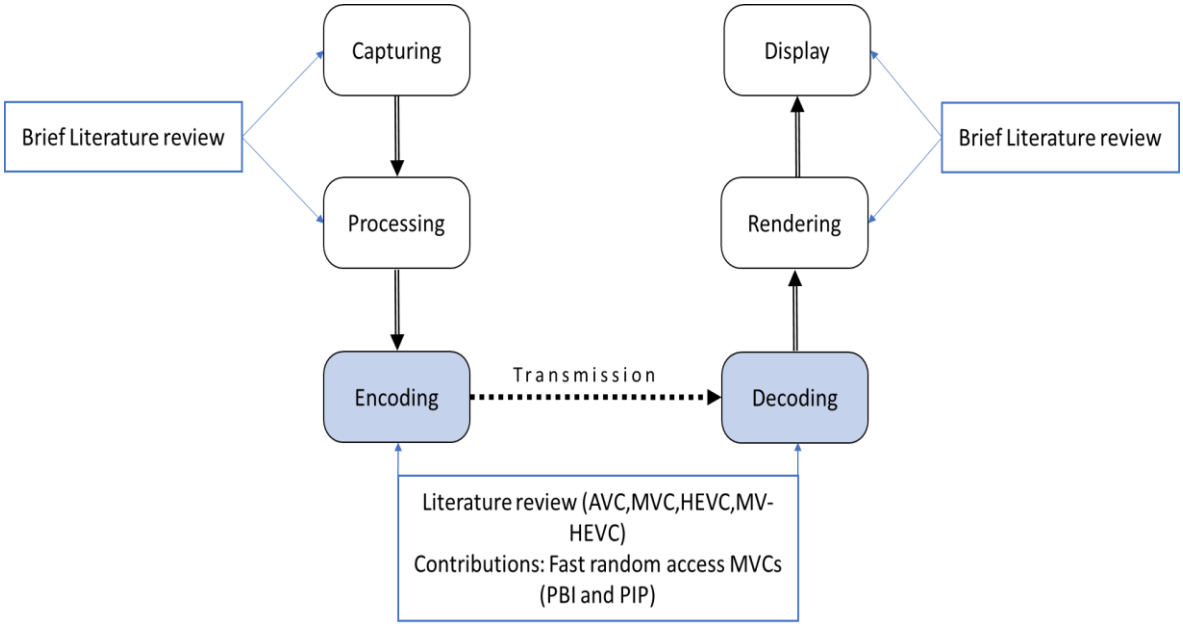



Figure 1.5 Considered subjects in our thesis

Thus, as shown in Figure 1.5, our thesis mainly focuses on the Multiview Video Coding part of the MVV system.

Therefore, the thesis presents a literature review of the fundamental concepts of 2D and multiview video coding by describing codecs such as AVC/H.264[22], MVC/H.264 [24], HEVC [23] and MV-HEVC [25]. Furthermore, ameliorated multi-view coding schemes are proposed in this thesis to enable fast random access and ensure proper compression ratios.

Exploiting temporal and interview dependencies during the encoding process is essential to provide efficient compression. On the other hand, this exploitation creates more complex multiview schemes which deteriorate the random access ability. In fact, the low-delay random access feature is essential to offer users the ability to quickly access an arbitrary selected view or frame, which means improving users' interactivity with the 3D scene. Rendering a randomly selected view of the multiview video with acceptable response time is called by computer graphics experts the interactive frame-rate.

Consequently, we aim in this thesis to design balanced multiview coding schemes that facilitate random access whereas providing good compression performance.

## 1.3 Thesis contributions

The main contributions of the thesis are as follows:

- A novel prediction structure PBI [26] that exploits both temporal and interview correlations while providing a faster random access ability exceeding 30% compared to the standard JMVM reference software. The proposed approach could be applied in multiview video coding regardless of the deployed cameras number. This approach is presented, tested and discussed in Chapter 3.

- A novel global random access evaluation metric [26] which considers all frames types of the evaluated structure. Global random access $G_R$ evaluation is composed of three sub-metrics: $G_R$ for anchor picture, $G_R$ for non-anchor pictures and $G_R$ which includes both anchor and non-anchor pictures.

- A second proposed prediction structure PIP [27] which renders faster random access ability compared to the IBP scheme and the proposed PBI structure. PIP achieves 53.33% of random access ability improvement with respect to MVC standard. In addition, different GOP sizes have been used to study their direct effects on multiview video coding in terms of compression efficiency and random access ability.

We have finally presented a theoretical and experimental comparison between MVC standard and the recent MV-HEVC codec, paving the way for future research work on MV-HEVC.

## 1.4 Outline of the thesis

The thesis is structured as follows:

Chapter 2 . addresses fundamentals of the multiview video system where much focus is given to the coding part. It begins with an introduction about the 3D history and concepts. Furthermore, it briefly presents the whole multiview video system chain covering acquisition and display techniques as well as basic multiview coding notions. Chapter 2 also highlights the video coding notions by providing an overview of the H.264 video coding standard, its architecture and its main features. Multiview video coding (MVC) extension of H.264, its requirements list, features and applications are all described in this chapter.

Chapter 3 firstly describes the default MVC interview prediction and other related approaches. The Proposed PBI interview prediction structure is presented in the second part of chapter 3. The proposed approach PBI is highlighted throughout schemes and specific equation models. PBI random access ability and compression efficiency are evaluated and compared against the considered structures.

Chapter 4 proposes an advanced interview prediction structure PIP which provides a faster random access ability.

Moreover, Chapter 4 presents an exhaustive investigation of the examined multiview video coding schemes based on different interview prediction structures and various group of pictures (GOP) sizes. PIP approach achieves better random access results against the first proposed approach PBI along the four used GOPs.

Chapter 5 presents an overview of the HEVC coding concepts while focusing on the multiview extended profiles MV-HEVC. Additionally, experiments are carried out to comparing MVC and MV-HEVC in terms of compression performance. Multiview video sequences with different resolutions and textures were subject to these tests. The reported results show the outperformance of the MV-HEVC over the MVC in terms of bitrate saving yielding a gain of over 70 % for a selected video sequence.

Chapter 6 finalises the thesis with a summary of its major findings, general conclusion and recommended suggestions for future research works.

# Chapter 2
# Fundamentals

## 2.1 Introduction

This chapter provides some essential concepts of the multiview video systems with a particular focus on their acquisition, display and coding techniques. It starts with a brief history of the 3D progression in Section 2.2 . The multiview system is then described including its two major applications: 3DTV and FTV in Section 2.3 .Different multiview video acquisition arrangements are explained and illustrated in Section 2.4 this is coupled with the cameras arrangement conditions that need to be considered to produce a satisfactory quality MVV. Section 2.5 introduces the multiview autostereoscopic display with its parallax barrier and lenticular techniques. An overview of the multiview video coding is presented in the last section. It first includes the multiview video coding history and the actual MVC requirements list to be respected throughout the development processes. The H.264 standard is then described, and its tools and key features are presented.

Finally, the multiview extension of H.264, its typical interview prediction and the MVC block diagram are given at the end of this chapter.

## 2.2 3D video history

3D video technology is a success story which might be traced to more than 180 years ago when Sir Charles Wheatstone constructed the first stereoscope in 1832. The device provides viewers depth illusion where two mirrors reflect two slightly different views of the same picture at 90-degree angle (Figure 2.1). The result is a merged 3D perception of the original picture. Sir Charles' research and observations on the "Phenomena of Binocular Vision" were presented to the Royal Society of London on two occasions [28][29].

Figure 2.1 Stereoscope of Charles Wheatstone [30]

Meanwhile, in 1838, Sir David Brewster designed his stereoscope which produced 3D vision from images. Later, in 1840, photography was invented. Therefore, drawings and painting were replaced by photographs in the existing stereoscopic devices. In 1844, Brewster improved further his stereoscope by adding prismatic lenses to fuse and enlarge stereo images and improving its portability [31].

The brothers Lumière were the first to make 3D stereoscopic films publicly available by using stereoscopic projector in 1903.

In addition to demonstrating the world' first colour television transmission in 1928, the Scottish engineer John Logie Baird also demonstrated the first stereoscopic television during the same year.

The 1950s witnessed the first golden age of 3D film industry. Mainly because Hollywood tried to attract more audience after the box office incomes dropping due to the keen competition with the television. 3D technology has since constantly upgraded which allowed the emergence of several 3D techniques for different application domains.

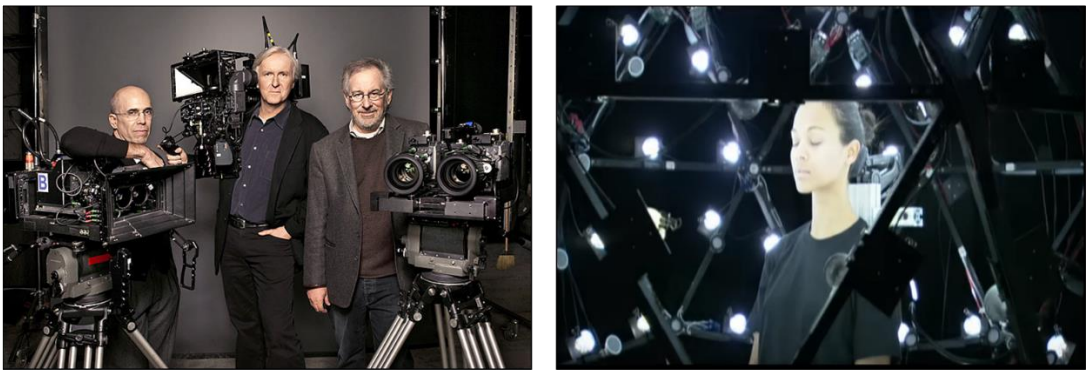
Progress of the 3D technologies over 170 years in addition to a definitive taxonomy covering the field up to the year 2000 were presented in Benton's book [32].

Figure 2.2 Multiview system used in The Matrix film [33]

The famous film The Matrix [33] released in 1999, was a successful application of the multiview system cameras. A set of 120 precisely triggered cameras was deployed to produce time freezing and slow motion virtual travelling effects, known nowadays as bullet time effect. This effect creates the illusion of a freely moving camera in a freeze time. Bullet time effect is actually one of the FTV applications.

Devices' digitisation was the new wave of the 21st century start which opened the doors for more advanced 3D content generated from all-digital media production chain. However, at that time, the ultimate challenge was to upgrade video services from the standard analogue definition to the digital high definition. This fact prevented the 3D video from being widely adopted.

The 3D cinema golden age has marked its second renaissance just after the success of Avatar [34] which created a new 3D genre at the beginning of the 2010s. Avatar combined computer-generated characters with real actors and used different 3D technologies (See Figure 2.3). To understand the exponential growth of 3D video adoption during the last decade, we only need to look at the rising global revenues of 3D enabled consumer devices and the rising number of digital 3D screens in cinemas. (See Figure 2.4)



(a) (b)

Figure 2.3 3D systems used to film Avatar [34]:

(a) 3D stereoscopic digital cameras, (b) omnidirectional multiview system.

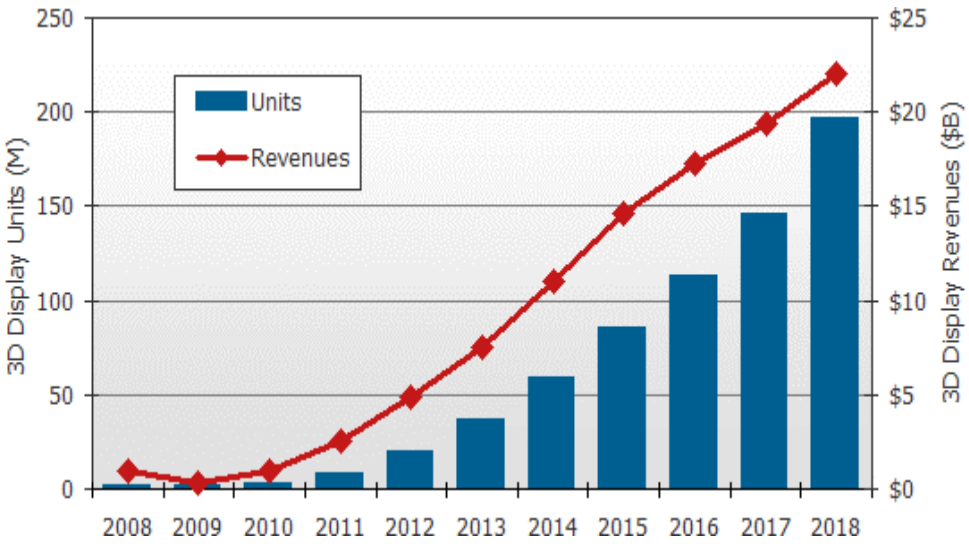



Figure 2.4 3D display technology progress [35]

## 2.3 Multiview Video System

Multiview video (MVV) is an extension of the conventional stereo video with a higher number of views. As shown in Figure 2.5, MVV provides more visual content of the captured scene compared to the stereoscopic video. MVV is based on recording multiple texture views of the same scene from closely located angles. This type of 3D video is mainly applied in two visual media scenarios: Free Viewpoint Video (FVV) and 3DTV.


(a) Stereoscopic video


(b) Multiview video

Figure 2.5 Example of 3D texture video

FTV or Free viewpoint video (FVV): This 3D system offers an immersive visual experience that allows users to freely move around the captured scene. An FVV system is comprised of several cameras set up around the scene. These cameras are connected to a network and controlled to capture the same scene from multiple positions simultaneously. Since the number of cameras is limited, the selected viewpoint does not always correspond to a real camera but could be synthesised from the captured visual data of the limited cameras. One of the notable examples of FVV application was previously mentioned in Section 2.2 the bullet time effect in the film "The Matrix".

FVV is also widely adopted in sports events, a notable example of that is the Iview joint project by BBC and Surrey University which focused on covering football and rugby sports scenarios where a minimum number of four cameras was used to generate a synthesised free viewpoint video. However, good visual quality requires a higher number of cameras [36][37].

3D TV: provides the viewers with an impression of 3D depth. The concept is inspired by the Human Visual System (HVS), where each eye perceives the same view with a slightly different position. Stereoscopic 3D is the basic application of the multiview system where only two pictures are transmitted to the viewers. Otherwise, N views (where N>2) are deployed depending on capturing and displaying capabilities. Figure 2.6 illustrates an end-to-end multiview system, where multiple cameras simultaneously record a scene which is later exposed through an autostereoscopic glasses-free screen that employs a lenticular sheet to separate the stereo pairs and offers a motion parallax effect.

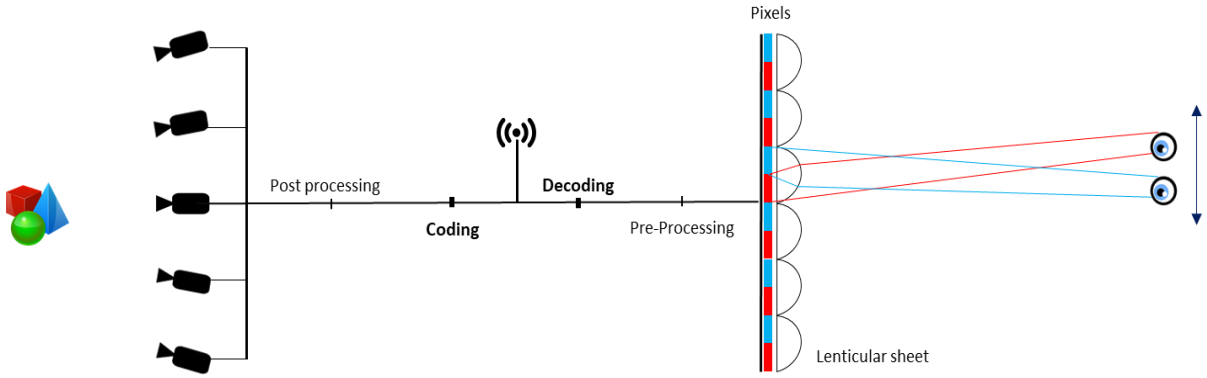



Figure 2.6 Typical Multiview video system

Further details about multiview video acquisition, display and coding are presented in the following sections.

## 2.4 Multiview Video Acquisition

Multiple cameras arrays layout, number and setting, can greatly vary depending on the targeted application [38]. The most common multiview cameras arrangements are illustrated in Figure 2.7 and described as follows:

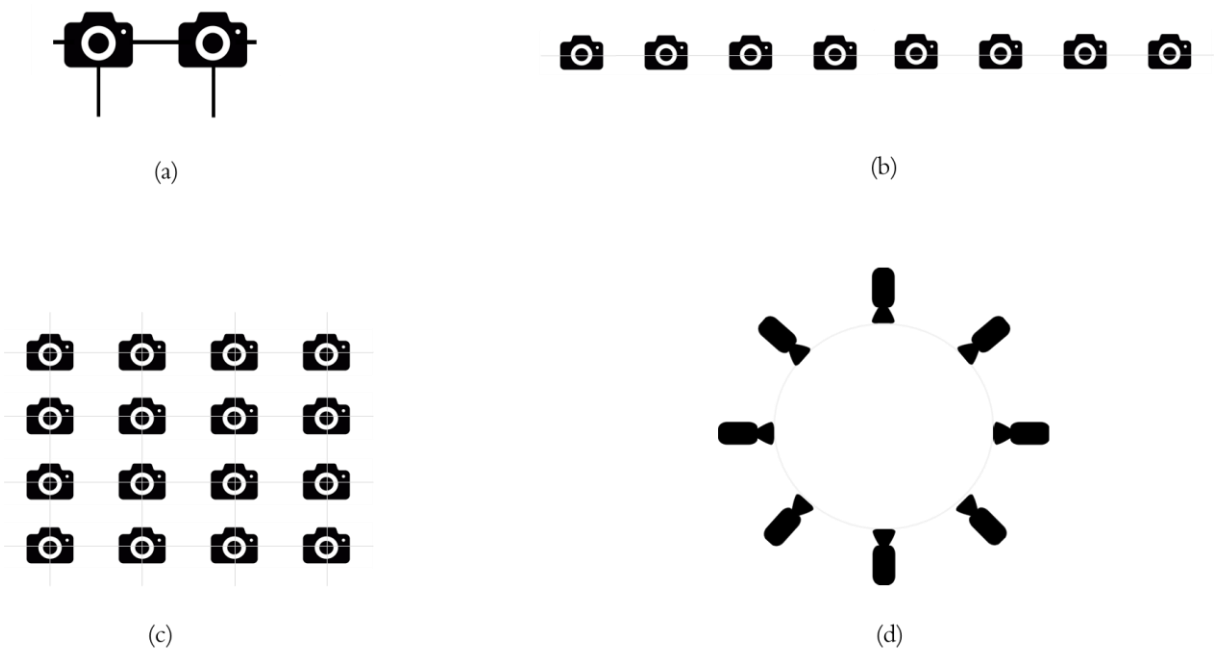



Figure 2.7 Multiview video acquisition arrangements: (a) binocular system, (b) linear system, (c) bidimensional arrays system and (d) omnidirectional system

a) Binocular or stereoscopic system is the basic multiview capture system which includes two close cameras that stimulate the HVS. This system is applied in 3D stereo visualisation and requires specific glasses for depth perception.
b) Linear system: cameras are regularly spaced and placed in one horizontal array. Although this configuration provides a single plan for viewpoint navigation, it highly facilitates the depth of scene estimation. This system produces 3D content for autostereoscopic displays.
c) Bidimensional system: cameras are placed in vertical and horizontal arrays to form a 2D linear planar. This system supports both horizontal and vertical motion parallax.

d) Omnidirectional or global system deploys multiple cameras in convergent setup toward around the scene centre. This system is mainly designed for free viewpoint video navigation, bullet time effect and motion capture (MoCap).

All multiview acquisition systems must consider intrinsic parameters such as ISO, shutter speed and aperture for every single camera to produce decent videos. Moreover, multicamera requirements need to be respected to ensure coherent and synchronised multiple videos. These requirements include:

1. Accurately synchronise the multiple cameras and using similar frame rate in order to facilitate the temporal integration of the multiview video data.

2. Multiple cameras should be accurately placed, and viewings fields should be defined to integrate the recorded multiview video geometrically.

3. The different surfaces of the scene should be perceived at least by two cameras to allow 3D reconstruction and depth estimation.

## 2.5 Multiview Video Display

Multiview video data can be displayed over a broad range of 3D display technologies. In this section, we will only consider autostereoscopic displays which support MVV, offer immediate three-dimensional effect and do not rely on any specific eyewear. Rather than exposing only one right and one left image, multiview displays include n>2 views forming n-1 successive stereo pairs. The autostereoscopic displays distribute several stereoscopic pairs to set of viewing zones. Thus, observers move horizontally and perceive multiple view-windows which creates the motion parallax effect. Two main optical methods are used to manufacturing autostereoscopic multiview displays, namely parallax barrier and lenticular sheet (see Figure 2.8).

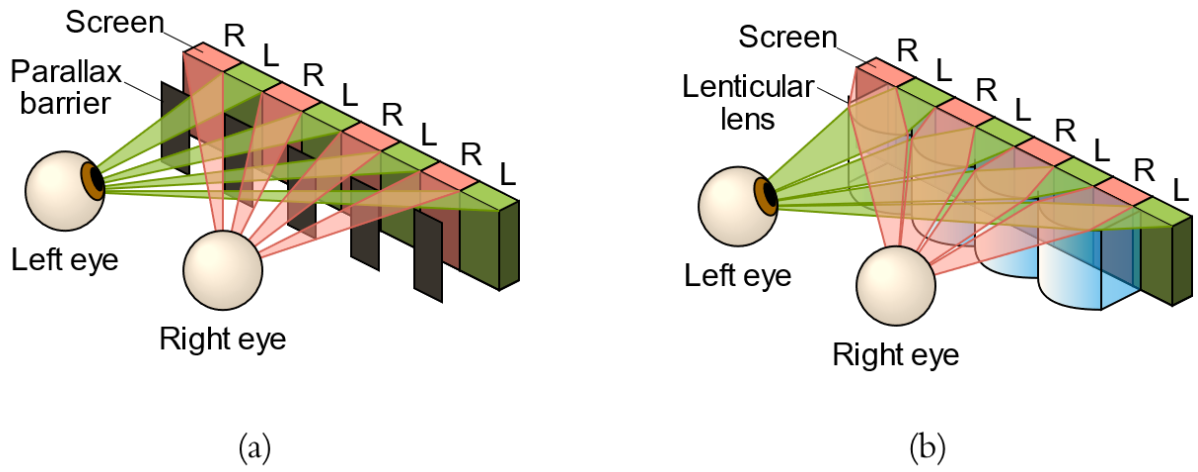



Figure 2.8 Autostereoscopic multiview display based on: (a) parallax barrier technique and (b) Lenticular technique [5]

a) Parallax barrier technique: A parallax barrier is placed in front of an LCD or other pixelated emissive displays. The barrier's vertical apertures separate alternately left and right eye image columns; hence light can only pass to the desired viewing zone. Users have to be within a defined distance range to benefit from 3D effects. Parallax barrier technology suffers from lighting issues due to the light occlusion that occurs a dim brightness.

b) Lenticular technique: This technology is based on the same principle of the viewing-windows. However, cylindrical lenses are installed in front of the pixel raster instead of the spaced barriers to avoid lighting loss [39]. The cylindrical lenslets act like tiny magnifying glasses, allowing each eye to only perceive one point of view among the multiple mixed views in each image. In fact, the most commercially available multiview displays use the lenticular technology. Vertical lenticular lenses may occur interference phenomenon which is known as Moiré effect. This issue has been tackled by slanting the cylindrical lenses with respect to the flat pixel grid [40]. In fact, nowadays, the commercially available multiview displays mostly deploy lenticular filter technology.

## 2.6 Multiview Video Coding

The coding process is a fundamental part of the video communication chain. It primarily includes compression of the video data and adapts it for transmission and storage. The Multiview video inherently introduces a significant amount of data compared to the conventional video. If, for example, a 2D video has (x) data, a multiview video will have N multiplied by (x) data where N is the views' number. It is soon apparent that compression is even more compulsory stage to deliver MVV content through the existing transmission channels. To do so, specific coding techniques that consider the amplified data volumes have to be used. Fortunately, a significant correlation exists between the adjacent views of the multiview video. This correlation can be exploited to further improve the compression ratio. Figure 2.9 illustrates the 3D dimensional correlation that exists within the multiview video content. Besides the interview correlation noted in Figure 2.9 as angular similarity, MVV content already yields spatial and temporal correlations within every 2D video of the views set. Therefore, the synchronised sequences of the multiview video have to be coded jointly and simultaneously by only one video codec (see Figure 2.10).

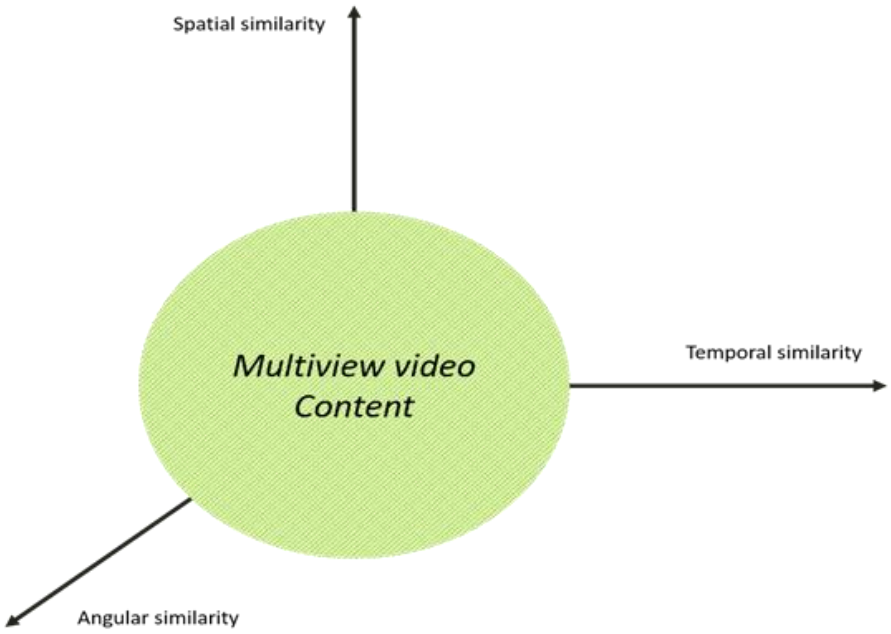



Figure 2.9 Multiview video correlation types

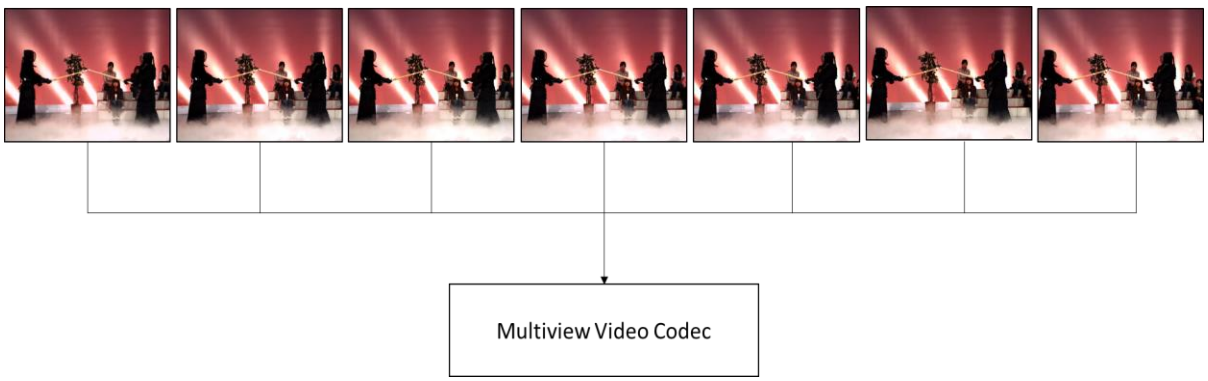



Figure 2.10 Jointly coded multiview video

Figure 2.11 portrays a simplified example of how spatial, temporal and interview correlations can be exploited. It shows a multiview video sequence with two adjacent views each has only two frames. The sequence represents a small blue square moving horizontally within an empty grey space. Instead of transmitting four frames, only information of frame 1 of view 1 in addition to its motion and disparity vectors can be included in the MVV bitstream to significantly reduce the data volume. Hence, effective motion-compensated and disparity prediction algorithms are used to eliminate the redundant information between the successive frames and adjacent views, respectively.

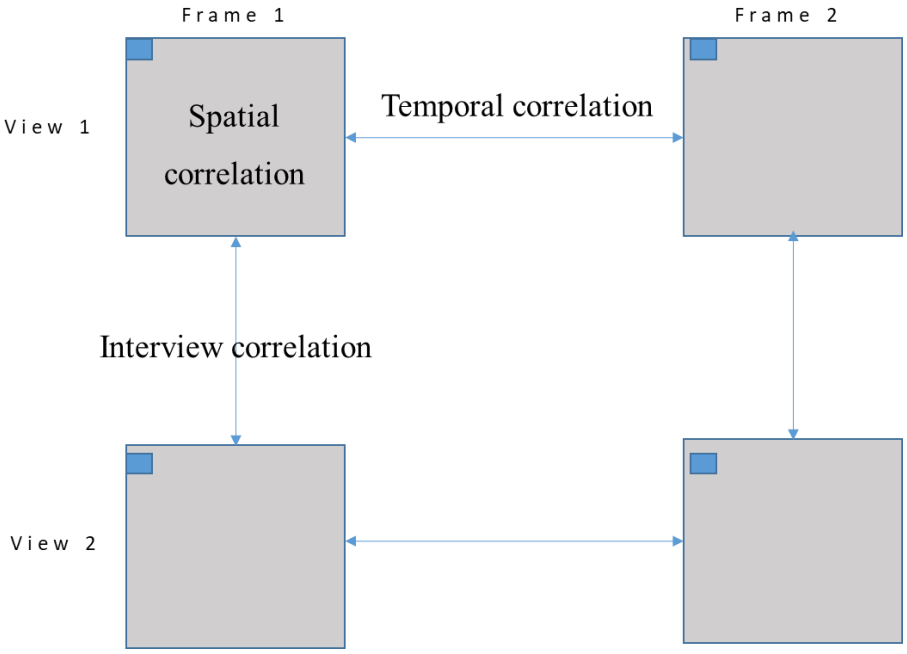



Figure 2.11 Simplified example of the redundant information in MVV

### 2.6.1 Multiview Video Coding History

It is hard to pinpoint an exact date in which the multiview coding concept was first introduced. However, one of the earliest noted propositions was done by Lukacs [41] in 1986. This research introduced the concept of interview prediction. Other approaches and experiments, then followed up, notably works of Dinstein et al. [42] in 1989, and research of Perkins [43] who described a mixed resolution coding structure as well as a transform-domain technique for disparity-compensated prediction.

The first international standard that supports multiview video coding was presented in 1996 [44] and consisted of extending H.262/MPEG-2 [45] to only support encoding of two views. In this first multiview standard, the left view was chosen as the base view which offers compatibility with the conventional H.262/MPEG-2 decoder. The right view was referred to an enhancement view which used pictures of the base view as references. The used coding tool features of this extended scheme had originally been developed for supporting temporal scalability [46][47].

At that time, the ultimate challenge was to upgrade video services from the standard analogue definition to the digital high definition. This fact prevented the multiview extension of H.262/MPEG-2 from being applied and developed.

Following the progress in video compression technologies and multimedia services, the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) joint forces and form a collaborative team namely Joint Video Team (JVT) in 2001. The JVT later released the first version of the H.264/AVC standard in May 2003. Meanwhile, the subgroup 3DAV of MPEG triggered the standardisation process of MVC in 2005 after receiving evidential outputs of some proposed multiview video coding schemes [48]. The Call for Proposal (CfP) document [49] launched in July 2005 was followed by responses [50][51][52]of different participants that proposed designs and tools with respect to the MVC requirements list [53]. After experimental evaluations of different proposed technologies, the MVC scheme [24] based on H.264 codec with interview structure and hierarchical temporal prediction, was selected as the default MVC design. The first MVC standard was then published including the multiview high profile in a reference model named the Joint Multiview Video Model (JMVM) [54].

### 2.6.2 MVC Requirements

A list of requirements has to be respected when developing a video coding schemes. In fact, MVC requirements [55] vary depending on the targeted multimedia service. A non-exhaustive list of MVC main requirements is detailed as follows:

- Compression efficiency:

High compression efficiency is considered as a central requirement for any video coding model. Compression efficiency is expressed through trade-offs between the bitrate gain and the video quality. It can be evaluated in terms of PSNR (dB) versus bitrate of the compressed video. An MVC with good compression efficiency must reach a significant gain against a simulcast scheme where no dependency between views is employed, as well as against benchmark schemes.

The compression efficiency of an MVC is mainly affected by the applied interview prediction scheme and the group of pictures (GOP) architecture. Additionally, proper cameras calibration and configuration ensure adequate similarities between views which allows better interview exploitation during the compression process.

- Random access:

Low delay random access ability comes at the top of any video coding requirements list. Random access ability ensures that any picture within the multiview video structure can be reached, coded, decoded and displayed with a relatively minimum delay. 2D video coding schemes only consider temporal random access. However, since MVC introduces interview dependency, both temporal and view random access are required. Fast random access improves the users' interactivity and navigation within the multiview video content. For applications where views switching is indispensable such as FVV, MVC schemes should be carefully designed in a way to minimise the number of the decoded frames between different views.

- Scalability:

Scalability is also a required functionality for video coding models. It allows decoders to access a portion of a bitstream while still being able to generate a decent video and display it on the terminal device. It reflects that any part of the video bitstream can be accessed by the decoder to produce an adjusted video quality. Scalability enhances interoperability of the same video bitstream over different networks and terminals. It offers multiple resolution levels and various frame rates of the same video. Scalability enables MVV content to be displayed on screens with limited views number capability.

- Backward compatibility:

MVC should be compliant with conventional decoders such as H.264/AVC and allowing single view extraction. Therefore, the base view of the multiview structure should be independently coded.

- Low-delay coding:

Low delay coding must be ensured by the MVC. It is much required for real-time applications such as live streaming and video conferences.

Parallel processing strategy [56] is employed to reduce delays. Its implementation enables encoding multiple views simultaneously.

- Camera parameters (intrinsic and extrinsic):

It is primarily required to send camera parameters within the bitstream to support view interpolation, depth perception and feature detection at the decoder side.

- Resource consumption:

MVC should be efficient in terms of resources consumption, such as processing power, used memory and bandwidth occupation. The MVC should be able to exploit interview similarity without heavily increasing the coding complexity because it might hamper the smooth 3D displaying. This requirement also includes energy consumption, especially when the MVC codec is integrated into embedded systems where energy saving become primordial.

## 2.6.3 Basics of H.264/AVC

MVC standard is an extended profile of H. 264 video codec. In this section, we provide a brief description of the H.264 key features which are also used with slight differences in MVC core.

### 2.6.3.1 Colour Space

H.264/AVC standard utilises Y Cb Cr colour space where Y represents the luminance component, Cb and Cr represent the chrominance component. Y value component represents the brightness level, whereas Cb and Cr values measure grey deviation towards blue and red respectively. Since the HVS is more sensitive to luminance than colour information, Y samples are represented with a higher resolution compared to Cb and Cr samples.

H.264/AVC main profile uses a sampling format where the luminance has a double size of the chroma components; This sampling structure is known as 4:2:0 format in which each sample is represented in 8 bits (See Figure 2.12).

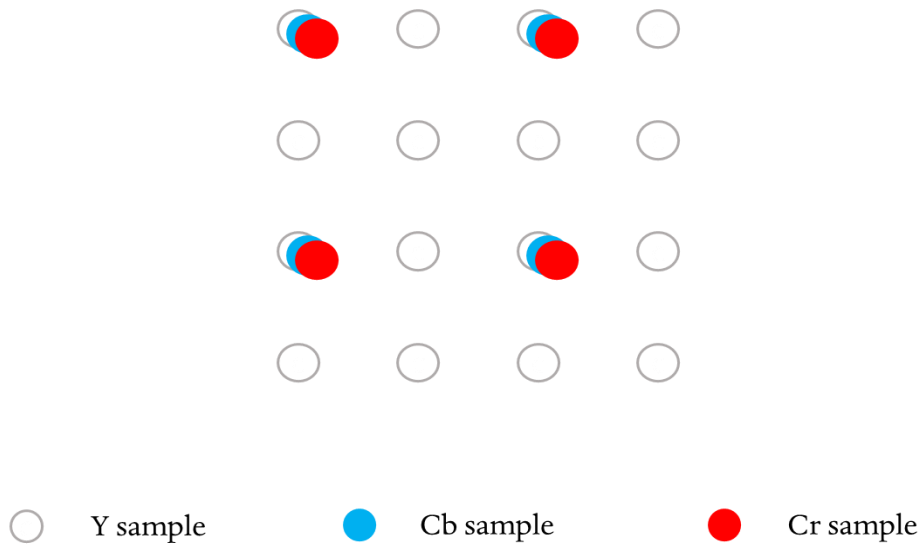



Figure 2.12 4:2:0 Format pattern

### 2.6.3.2 Marco Blocks

A Macroblock (MB) is the basic frame unit adopted in H.264 coding. Each frame of the video sequence is split into MBs of size 16 x 16 luminance samples (Y) and 8x8 samples for each chrominance component (Cb, Cr), all in 4:2:0 format. Figure 2.13 illustrates the MB structure within a 4:2:0 format.

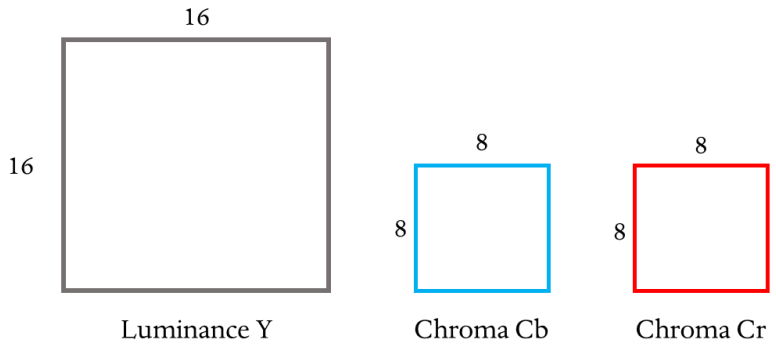



Figure 2.13 Macroblock structure in H.264

### 2.6.3.3 Slices

A video coded frame is composed of slices collection or only one slice. A set of contiguous macroblocks makes up a slice in which the number of macroblocks is not constant; it can be just one or all frame' macroblocks. The slice structure enables coding flexibility at different rates. There are three types of slices reflecting the prediction type that can be used for coding their macroblocks:

I slice: uses only intra prediction

P slice: uses both intra prediction and inter prediction for coding its macroblocks. However, only one direction is allowed for inter prediction.

B slice: uses intra prediction and inter prediction from two directions to coding its macroblocks.

### 2.6.3.4 Intra Frame Prediction

In this type of prediction, macroblock samples are predicted from the same slice. H.264 supports two intra prediction modes: Intra_4×4 and Intra_16×16.

The intra_4×4 mode is more suitable for regions with significant details. The macroblock is split into 4×4 partitions; each one can use eight specific directional modes and a DC mode. Intra_16×16 is more suitable for smooth regions of a frame. In this type, the macroblock is predicted as a whole and can use four modes: DC, planar mode, vertical mode and horizontal mode.

### 2.6.3.5 Inter Frame Prediction

Inter frame prediction exploits the correlations between successive frames to reduce data volume. The previously encoded frames are regarded as references that can be used to predict other frames. Motion compensated prediction is used to find the best match of a current frame' macroblock in a reference frame. The motion estimation is carried out on macroblocks of 16×16 or smaller block sizes of 16×8, 8×16 and 8×8. If the 8×8 mode is selected, it can be further divided into sub-macroblocks of sizes 8x4, 4x8 or 4x4 (Figure 2.14). Both P and B slices use similar macroblock portioning techniques. However, Macroblocks of B slices are predicted from two reference frames (Figure 2.15).

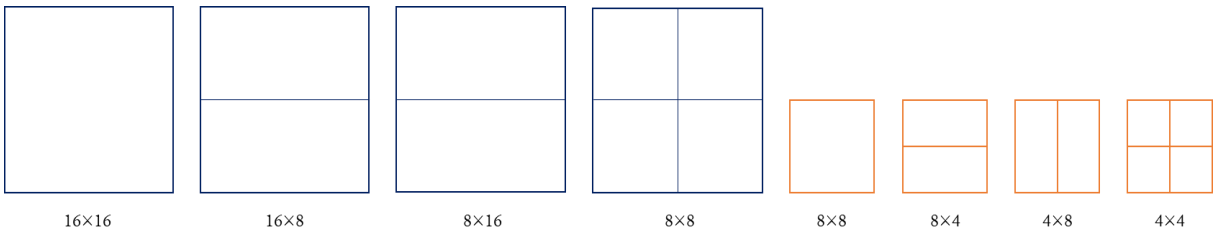



Figure 2.14 Macroblocks portioning modes

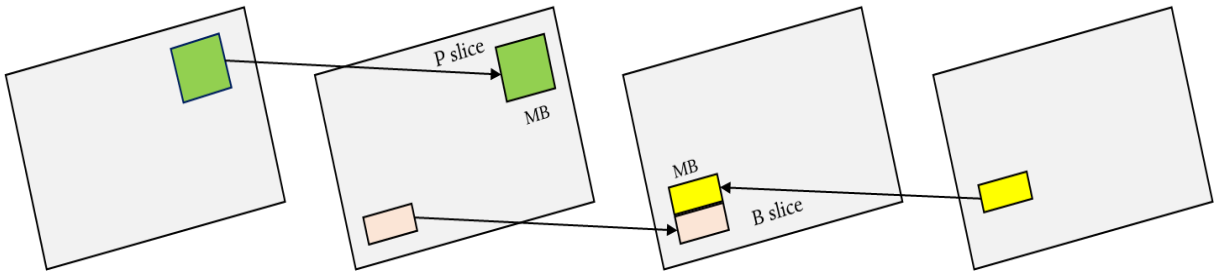



Figure 2.15 Inter frame prediction for macroblocks

#### 2.6.3.6 H.264/AVC Encoder

Figure 2.16 illustrates a simplified H.264 blocks diagram. A macroblock has two options: 1. it can be coded using intra prediction from spatially adjacent samples of the current slice which have been already coded and reconstructed. 2. The macroblock can be processed using inter prediction from one or two reference frames. The generated predicted signal of both cases is then subtracted from the original macroblock resulting in residual macroblock. Discrete Cosine Transform (DCT) is applied on the residual macroblock. The quantisation is then used to eliminate the less significant coefficients. Afterwards, the resulted signal goes through entropy coding, where two techniques are defined in the H.264 standard: Context Adaptive Variable Length Code (CAVLC) [58] and Context Adaptive Binary Arithmetic Code (CABAC) [59]. Entropy coding is also applied to other signals such as the reference frames list, motion vectors and macroblock portioning modes. The reconstructed frames are used as references frames for spatial and temporal prediction. In fact, an opposite path (green lines in Figure 2.16) is functionalised to reconstruct macroblocks. An inverse quantisation followed up by an inverse transformed residual are added to the predicted signal of the direct path. A deblocking filter is then used to decrease the undesirable blocking artefact effects. The reconstructed frames are temporarily stocked in the Reference Picture Buffer and are consequently ready to be used as references.

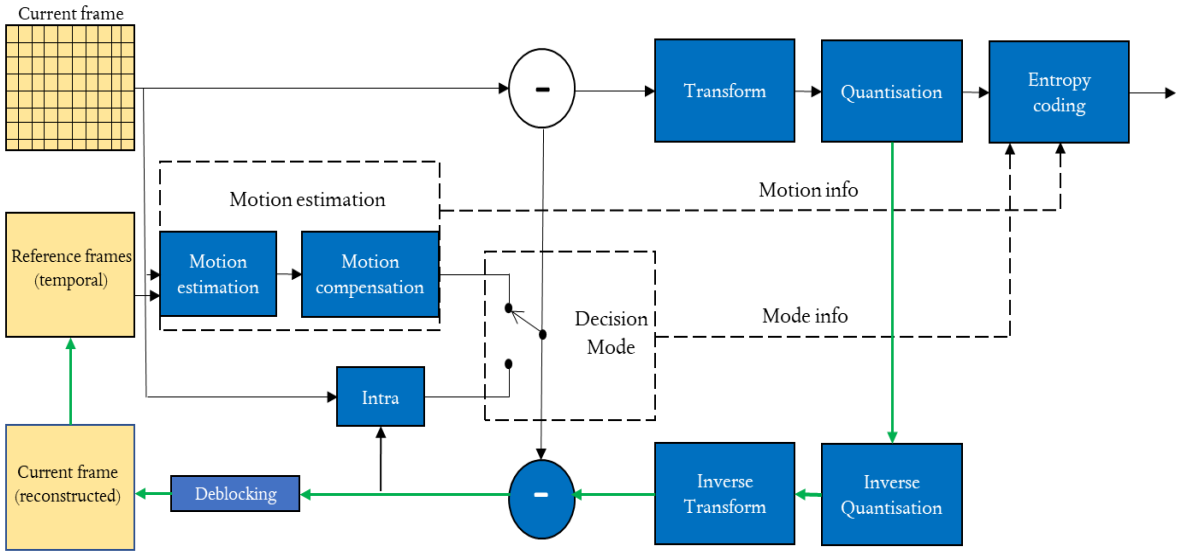



Figure 2.16 H.264/AVC encoder Diagram

## 2.6.4 Multiview Extension of H.264

Although MVC employs the H.264/AVC coding principles, it has some specific features that qualify it to efficiently encode the MVV media. The MVC extension [60] defines two profiles: The stereo high profile which is limited to two views and the multiview high profile for encoding multiple views.

While H.264/AVC only exploits two types of redundancy by making use of the intra and inter frames predictions, MVC enables interview prediction and exploits three redundancy types that usually exist within an MVV content:

(a) Spatial redundancy between the frame regions.

(b) Temporal redundancy between frames of the same view.

(c) Spatial redundancy between frames of the neighbouring views.

Hence, the reference frame lists capacity of H.264/AVC are extended to include indices of frames from adjacent views in addition to the frames of the view indices.

It is important to notice that interview prediction is possible between frames of adjacent views which only correspond to the same time instance. This set of frames is known as Access unit (AU). Figure 2.17 illustrates typical interview prediction process for MVC profiles.

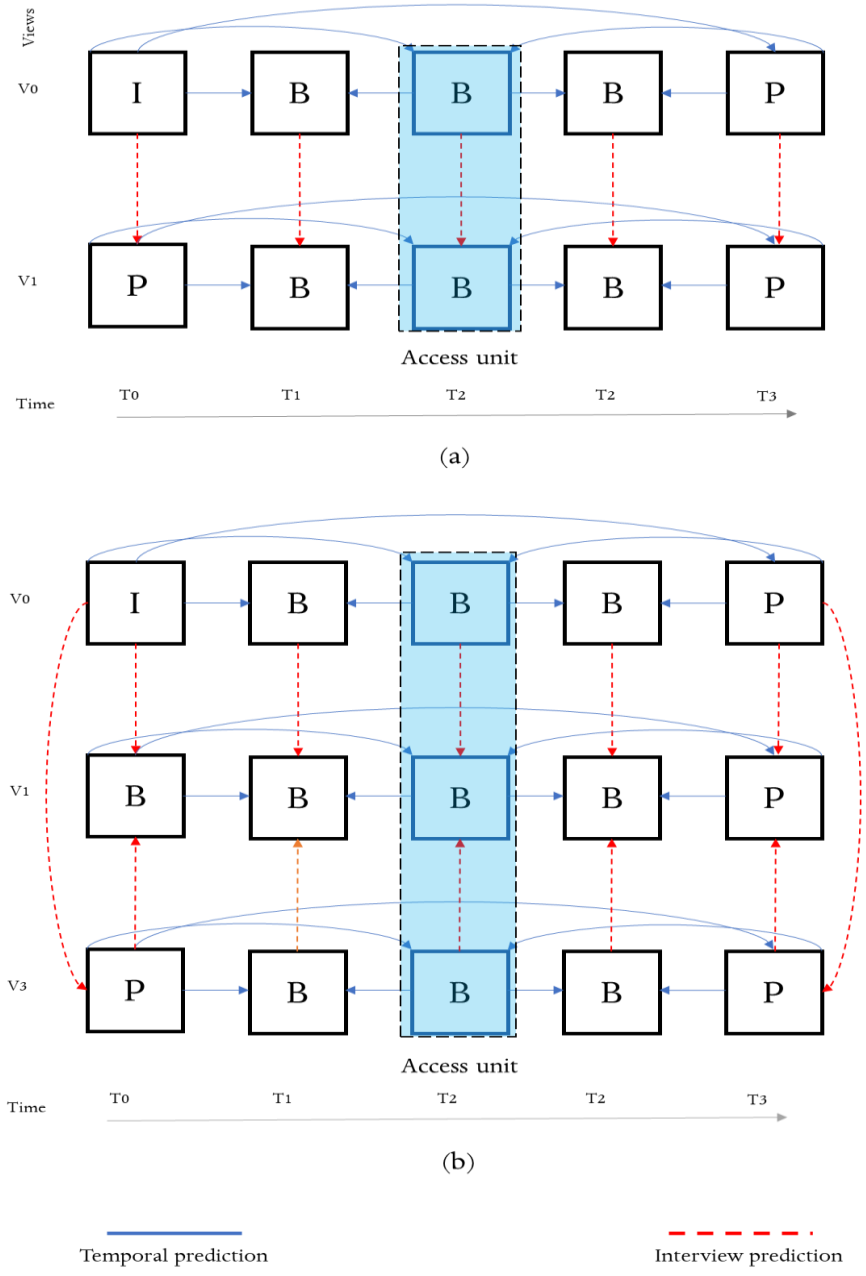



Figure 2.17 MVC prediction schemes for: (a) stereo high profile and (b) multiview high profile

Figure 2.18 depicts the typical processing blocks that are included in H.264 encoder to build up an MVC encoder. In fact, the MVC codec uses the same strategy of H.264 to encode the base view marked as V0 in both profiles. An improved mechanism is applied to the non-base views in which MVC employs Disparity Estimation (DE) and Disparity compensation besides motion estimation technique. Therefore, the MVC decision mode has three options: Intra frame, Inter frames and Interview predictions. The selected mode is normally determined by the rate-distortion (RD) optimisation. The standard equation of RD is given as follows:

$$J = D + \lambda R \tag{2.1}$$

J is the RD cost, D is the distortion costs, and R represents the bitrate. J is calculated for all possible combinations of block sizes by DE and ME through all available reference frames. The best combination is chosen to encode the active macroblock. The standard reference MVC software employs the Rate-Distortion Optimized Mode Decision (RDO-MD) for mode selection.

The multiview stream includes an independently coded base view bitstream to ensure the backward compatibility of the MVC standard with the single view profile of the standard [61]. Therefore, the video data related to the base view is encapsulated in Network Abstraction Layer[1] (NAL) units that is originally defined for 2D videos. However, the video data related to the non-base views are encapsulated in an extension NAL unit which is designed for both multiview video and scalable video coding (SVC) [62]. A specific flag is added to distinguish the NAL unit type, whether it is an MVC or SCV bitstream.

[1] A coded video stream is organised into NAL units. Video coding layer (VCL) NAL units contain video content data. Non VCL NAL units contain associated additional information.

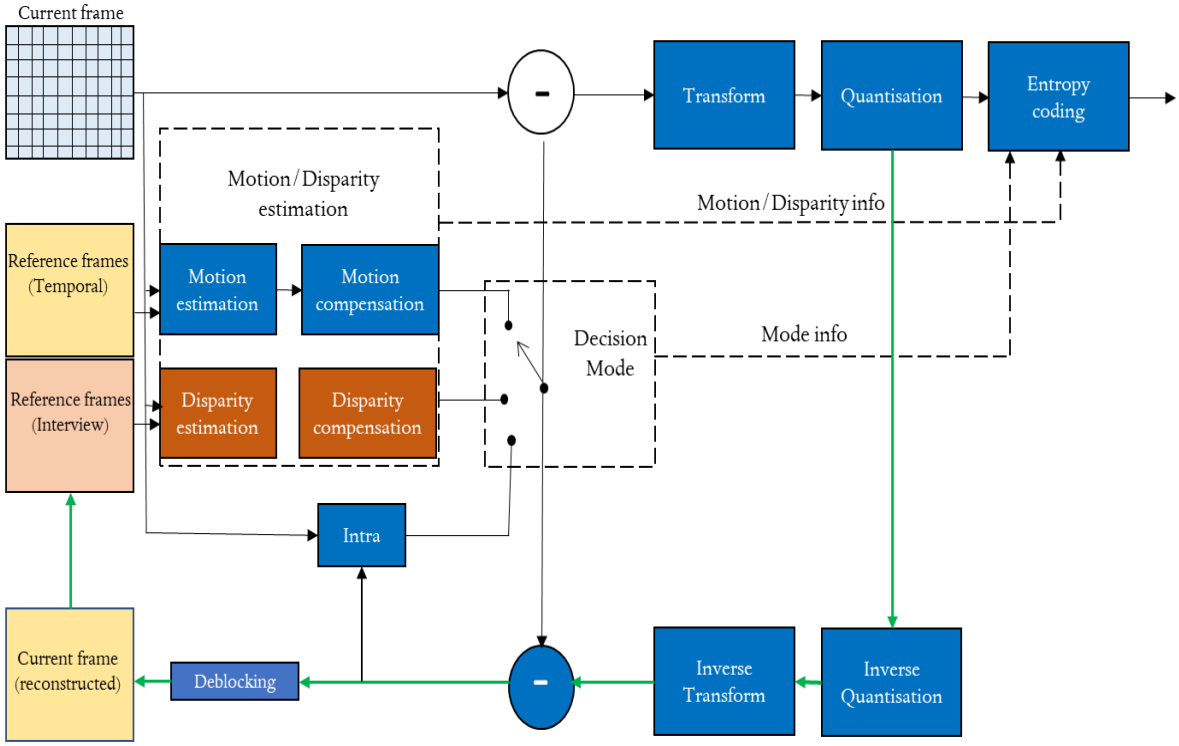



Figure 2.18 MVC/H.264 encoder Diagram

## 2.7 Conclusion

This chapter has reviewed the key concepts of the multiview video system . The recent history of 3D imaging, while tracing back its progression since 1832, was presented in the beginning of the chapter. Furthermore, multiview video acquisition and display techniques were briefly illustrated and described. More focus was dedicated to present the multiview video coding technology, its history, and its requirements. Finally, the multiview extension of H.264 standard was detailed to provide a strong background about the next chapter which will present a novel method to improve the random access ability as one important requirement of MVC.

# Chapter 3
# Random Access Enhancement for Multiview Video Coding

## 3.1 Introduction

Multiview video is a three-dimensional (3-D) scene captured by at least two cameras located at different viewpoints. Compared to the conventional 2-D video, the 3-D scene representation usually requires a much more substantial amount of data. Consequently, efficient compression for data storage or transmission represents a challenging task. The most straightforward solution for encoding this type of video is to encode each view independently with a standard video codec such as H.264/MPEG-4 AVC [22]. However, this method does not exploit the spatial correlations that exist between the adjacent views. An adequate compression of the MVV should combine both motion and disparity compensation to eliminate the redundant information and provide decent compression ratios. Merkle et al. [24] introduced an approach based on the exploitation of both temporal and interview prediction. This efficient approach is later adopted and implemented in a reference model named Joint Multiview Video model (JMVM) [54].

Besides the compression efficiency, reducing the complexity of the view random access is one of the most important requirements that should be considered in the multiview video coding. Many researchers proposed different MVC structures to meet the MVC requirements [53]. In [63], an MVC algorithm based on distributed source coding was proposed to tackle the free viewpoint switching problem of the coding efficiency. Even though it outperforms the solutions based on intra or closed-loop predictive coding, it provides less efficient compression compared to H.264/AVC standard. Similarly, in [64], three approaches were proposed, providing a reduced delay view random access; these methods include SP/SI frame coding, interleaved view coding and secondary representation coding. However, the compression performance was inferior to the multiview extension of AVC. Zhang et al. proposed, in [65], a method to adaptively select the best prediction mode among a set of predefined schemes. This approach is based on a spatiotemporal correlation analysis using Lagrange cost. It provides significant enhancement of the view random access but with an additional encoding delay and higher consumption of memory resource. In [66], Yang et al. suggest a prediction structure based on the enhancement of the encoding order of the B pictures and their reference frames as an extension of each independent view of the multiview video by applying a binary tree algorithm. This approach leads to a significant improvement in the bitrate performance. However, it slows down the view random access due to the increased coding complexity.

In this chapter, a novel interview prediction structure is proposed to improve the random access performance while maintaining high compression efficiency. It consists of using two base views (I-view) with selected positions in a scheme of eight views. An extended version of the proposed scheme, for structures of more than eight views, is then developed. Furthermore, a novel evaluation approach to fully assess the random access ability of the MVC coding schemes is introduced.

The rest of this chapter is themed as follows: Section 3.2 presents an overview of the multiview coding technologies through describing some related works such as simulcast scheme, sequential view prediction and distributed video coding for MVV. Relevant prediction structures such as IPP and IBP are presented in Section 3.4 .

Our proposed framework to improve the random access ability is detailed in Section 3.5 Evaluation methods and results of the random access ability are presented in Section 3.6 Section 3.7 presents the compression performance evaluation of our proposed approach against other relevant works. Summary of the evaluation results is given in Section 3.8 Finally, some concluding remarks about our framework achievements are provided in the conclusion.

## 3.2 Technologies for Coding Multiview Video

### 3.2.1 Simulcast

The straightforward solution to encode multiview video sequences is the simulcast scheme. Views in this structure are independently coded, and only temporal redundancy is exploited using standard video codecs such as AVC. The interview redundancy is neglected in the simulcast coding which makes it a low complexity solution and compatible with the 2D conventional decoders. By making use of H.264/AVC and the dyadic hierarchical prediction structure, video compression has been efficiently improved in comparison to the traditional simulcast coding structures [67]. Figure 3.1 depicts the dyadic hierarchical B prediction structure where the group of pictures (GOP) size is eight. A selected frame within the GOP can be predicted from previous and later frames of the closest higher hierarchical level. The first picture is independently coded as an instantaneous decoder refresh (IDR) picture, and the so-called anchor or key pictures are coded in regular intervals of seven frames. The B pictures, located between two I pictures and referred to as non-anchor frames, are temporally predicted using the concept of hierarchical B frames. This temporal prediction scheme was derived from the temporal scalability structure in Scalable Video Coding (SVC) [62].

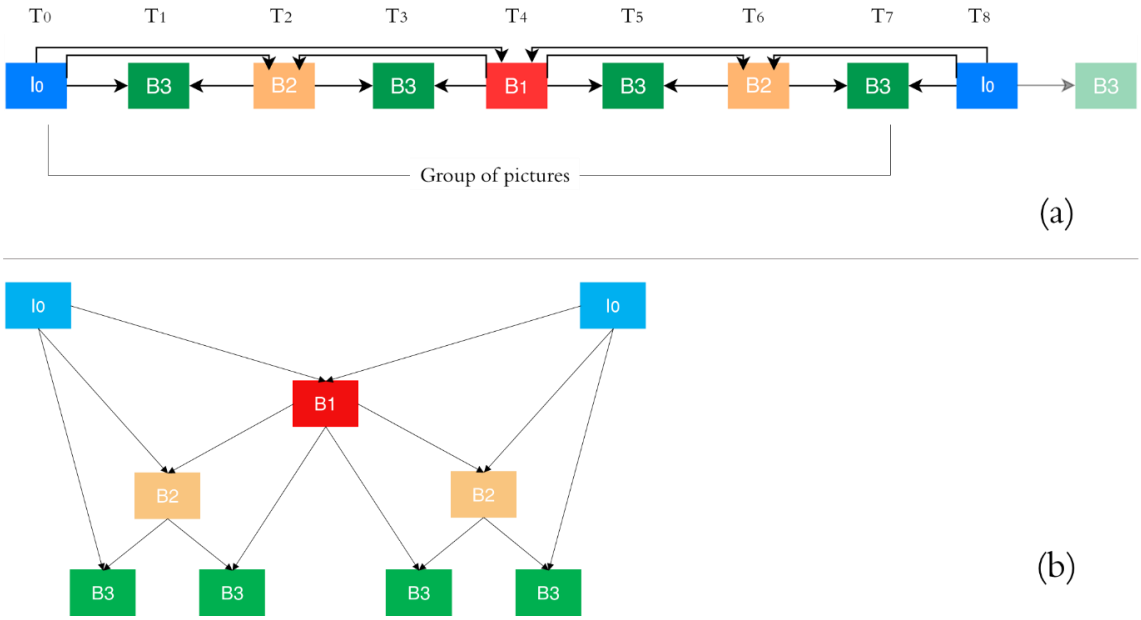



Figure 3.1 Hierarchical B pictures structure (a) group of pictures (b) levels decomposition

The simulcast scheme is illustrated in Figure 3.2. In the depicted case, each group of groups of pictures (GGOP) is composed of eight views and eight frames per GOP. Sn indicates the different views (cameras), while Tn represents the time location. Simulcast compression is characterised by its optimal interview random access as all key pictures used within the prediction structure are independently intracoded. However, its compression performance is not maximised as interview correlations are not utilised. Usually, simulcast coding is employed as a reference model for coding performance comparisons between different MVC schemes.

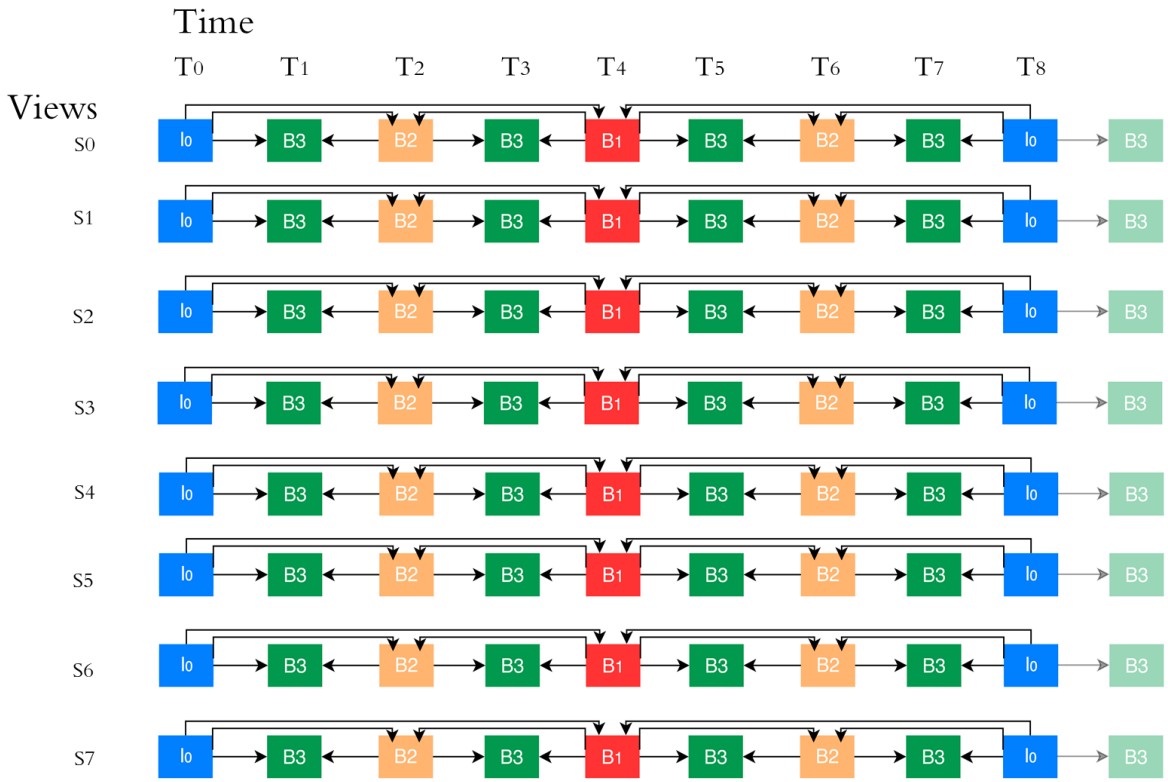



Figure 3.2 Simulcast structure for 8 cameras MVV

### 3.2.2 Hybrid video coding

Analytical results of multiview video sequences [68] indicate that significant correlation is present in the views level in addition to the temporal dependency. Therefore, prediction structures can take benefit from the combined temporal-interview dependencies. For instance, a frame can be predicted from reference neighbouring frames of the same view and reference frames of the adjacent views. Hence, to satisfy the main MVC requirements, many inter-view prediction structures were proposed based on using simultaneously temporal and interview dependencies. Some of the relevant propositions are presented as follows:

#### 3.2.2.1 Sequential view prediction

In this type of structures [69], each view can have a different role concerning the interview dependency. The first view is always encoded using only temporal prediction.

Frames of the remaining views are predicted using the corresponding frame of the previous view in addition to the temporal prediction. Bi-predicted frames may be used to improve the compression ratios. Figure 3.3 depicts a type of the sequential prediction structures.

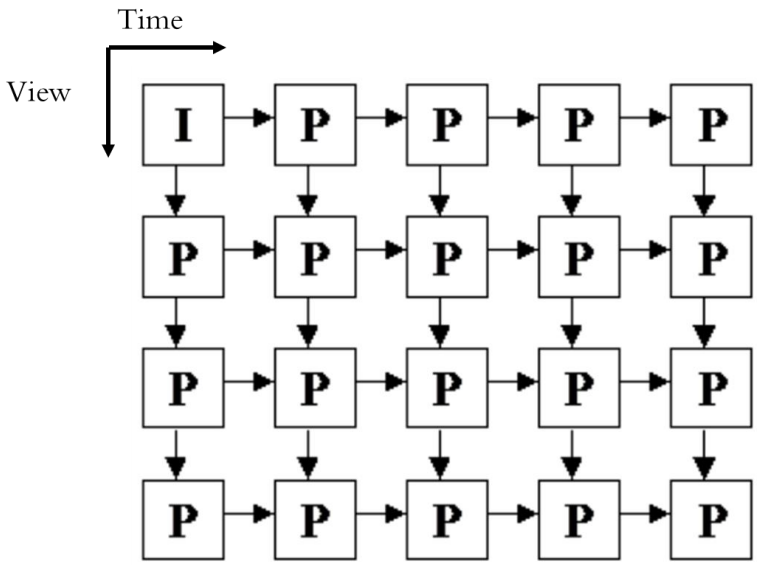
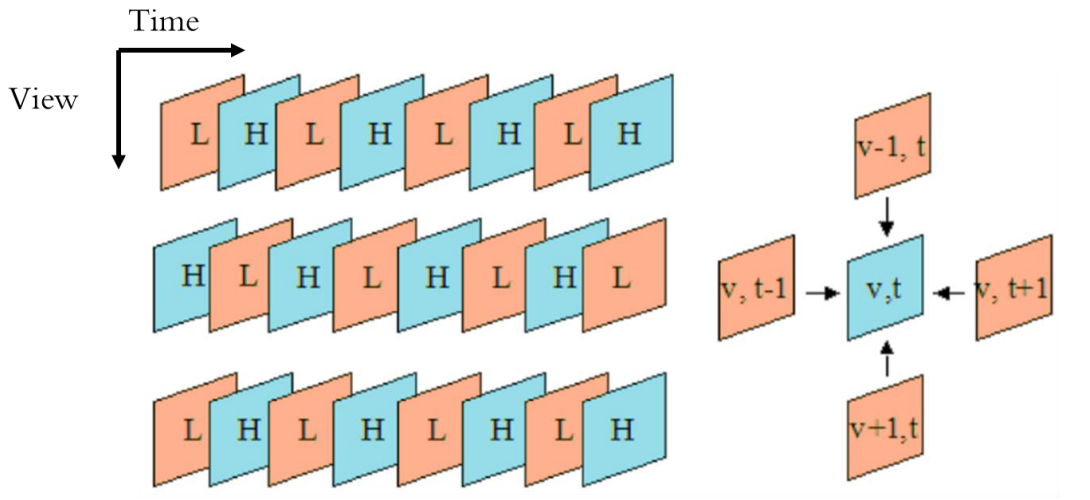



Figure 3.3 Sequential prediction structure [70]

### 3.2.2.2 Checkerboard decomposition

The multiview videos are decomposed in this prediction structure into low band frames, high band frames, and side information according to the selected prediction mode [71]. For each view, the sequence of low band frames is independently encoded, whereas, the sequence of high band frames is predicted using disparity and motion compensations from the neighbouring low band frames. Figure 3.4 shows an example of this prediction scheme.




(a) Checkerboard decomposition (b) High band frames prediction

Figure 3.4 Checkerboard decomposition prediction scheme [71]

### 3.2.2.3 Distributed video coding for MVV

Distributed video coding [72] is another approach that was core to develop some multiview video coding schemes [73]. It was essentially designed to support wireless cameras systems with low processing capacity and limited energy resources. Afterwards, the approach was further improved to reduce temporal and interview correlations within MVV [74][75]. Nevertheless, multiview distributed video coding schemes provide less compression performance compared to those based on hybrid coding [72].

### 3.2.2.4 Wavelet approach for multiview video coding

Video coding schemes based on wavelets have been extended to support interview prediction following the hybrid multiview video coding schemes [76].

4D wavelets based on coefficients decomposition in temporal and interview domain were also proposed [77] to support multiview video coding. Wavelet-based schemes provide scalability feature, and they are generally simple to implement.

### 3.2.3 Efficient prediction structures

Joint video team (JVT) was involved in the development of the MVC standard. Developing prediction structures using both temporal and interview predictions was a fundamental task of its works. The team evaluated different proposed approaches to finally adopt prediction structures presented in [24] based on their efficient RD behaviour. The proposed structures were later implemented on Joint Multiview Video Model (JMVM) [78] software. The reported study in [24] investigated the optimal configuration that combines temporal and interview predictions. The adopted solution in MVC standard was greatly based on the AVC coding tools. MVC brought changes to the High-level syntax of the AVC standard, enabling to encoding multiple views into a single encoded video stream. Views in MVC are encoded using the temporal prediction tools of AVC in addition to the interview prediction tools which extends the reference frames list of the encoded views to include frames from other views. Although MVC and AVC standards have limited changes mainly at the high level definition, other prediction tools such as MVC motion skip [79] and illumination compensation [80] were included in JMVM software. At the temporal level, the hierarchical B prediction structure, previously described in Section 3.2.1, was adopted due to its coding efficiency. Whereas, for the interview level, two interview prediction structures, illustrated in Figure 3.5, were adopted as non-normative multiview prediction schemes in JMVM.

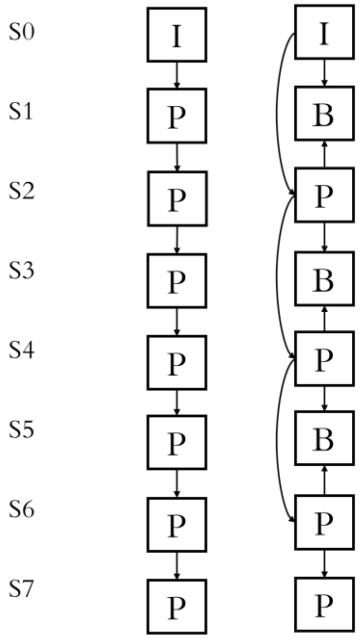



Figure 3.5 IPP and IBP interview prediction structures

Both IPP and IBP were tested in [24] for only anchor frames schemes and anchor and non-anchor frames structures. Obtained results showed that structures using interview prediction jointly for anchor and non-anchor frames achieve higher coding gain compared to those using only interview prediction for only anchor frames. Although IPP outperforms IBP in terms of compression efficiency, it increases the multiview encoding complexity and slows down the random access ability. More details about IPP and IBP structures will be provided in Section 3.4

## 3.3 Random Access Ability

Random access is an essential requirement for any video coding schemes that facilitates the view switching feature applied in free viewpoint video.

Temporal random access is provided by Instantaneous Decoder Refresh (IDR) frames which are independently coded from any other frames. IDR or I frames which represent natural random access point divide the video into multiple GOPs.

If a frame (x) within a GOP is selected to be accessed, the decoder searches for the closest I frame and starts decoding the depending frames until it arrives at frame (x).

Coding structures that minimise the number of the decoded frames to access a selected frame have better random access ability. Reducing the GOP size by inserting more intracoded frames improves the random access ability. However, this method costs increased data volume, as I frames provide more bit rates compared to P or B frames. Supporting temporal random access offers interactivity options such as fast backward and forward of the played video and selecting the desired playback position requested by the viewers in streaming applications. For MVC schemes, the process becomes much more complicated because the view dependencies are included in the video coding structure. Both temporal and view random access are involved in MVC. They ensure together that any frame can be accessed, decoded, and displayed with a minimum of intermediary decoded frames. Reducing coding dependencies between the encoded views and frames can be achieved by simply inserting more intracoded frames in both temporal and view levels. Therefore, the designed prediction structures have to consider the trade-off between random access ability and coding efficiency and provide balanced strategies. The next section examines some relevant prediction structures that consider the random access ability.

## 3.4 Relevant prediction structures

Many inter-view prediction structures have been proposed to satisfy the main MVC requirements such as coding efficiency and random access ability. The proposed schemes vary based on particular criteria such as the type of the deployed anchor pictures in the interview structure and the number of reference frames of the anchor and nonanchor pictures. In this section, for later comparison purposes, we present three relevant interview prediction structures, where two of them (IPP and IBP) have been proposed and adopted by MVC standard. The third interview prediction structure was based on JMVM and provided better random access ability compared to the former structures. Figure 3.6 depicts the IPP prediction structure of a multiview sequence employing eight cameras with GOP size of eight frames. This structure uses one IDR frame for each GGOP.

The first view S0 represents the base view that it is coded using only the temporal prediction. The remaining views from S1 to S7 are of type P. They start with an anchor P frame predicted from the previous anchor frame I/P. The nonanchor frames of the P views are predicted from both temporal and interview level, i.e., each nonanchor frame is coded from two frames of the temporal level and only one frame of the previous view level. E.g., the frame located in (S1, T4) is predicted from (S1, T0) and (S1, T8) at the temporal level, and from (S0, T4) for the interview level. IPP interview prediction scheme achieved better results in terms of compression efficiency compared to simulcast, IBP and all reported structures in [24]. Despite that, IPP is characterised by its increased complexity which deteriorates the random access ability.

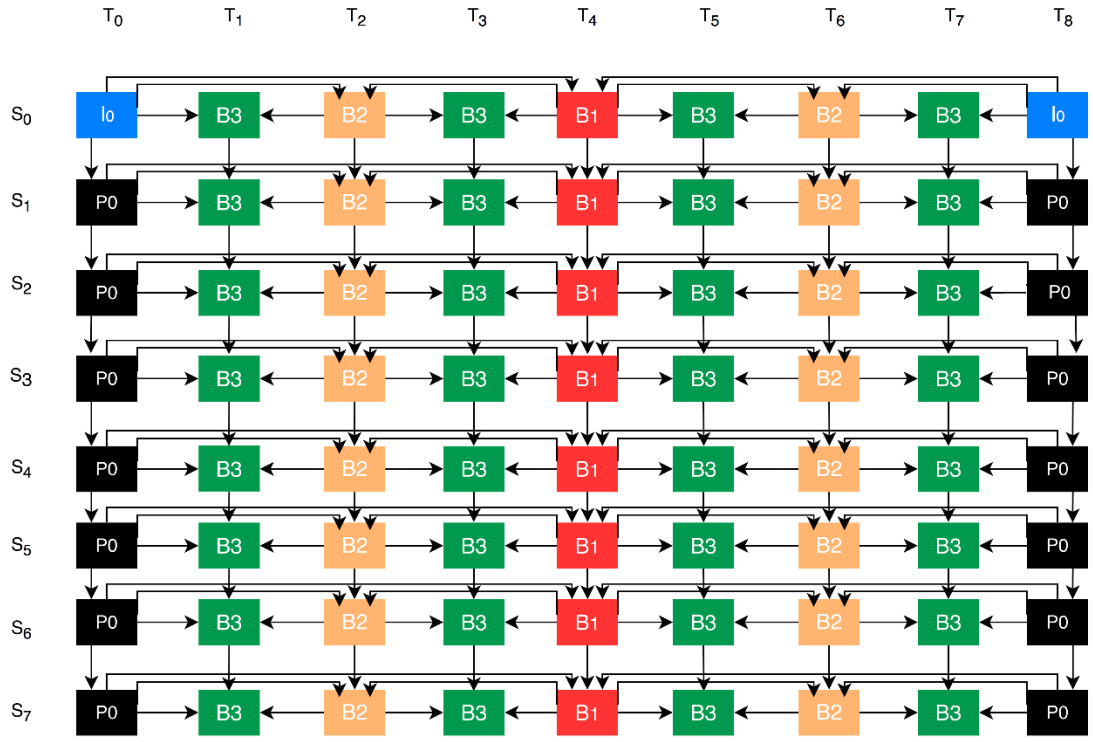



Figure 3.6 IPP interview prediction scheme

In [81] an evaluation method has been presented to assess the random access ability. It consists of calculating the maximum number ($N_{max}$) of reference images needed for coding or decoding the highest hierarchical level frame. For IPP scheme, $N_{max}$ is defined as follows:

$$N_{\max} = (H_{\max} + 1) + 5 \times [Nbr_{view} - 1] \quad (3.1)$$

Where $H_{max}$ defines the highest level of the B frames in the hierarchical B coding structure ($H_{max}$ is equal to 3 for IPP), and $Nbr_{view}$ represents the number of view in the structure.

The IBP interview prediction scheme, illustrated in Figure 3.7, uses the hierarchical B coding structure for both temporal and interview levels.

It employs three types of view: I view (S0, the base view of the structure), P views (S2, S4, S6 and S7) and B views (S1, S3 and S5). Each B view is set between two P views or between an I and P view. The B view starts with an anchor B frame which is bidirectionally coded from the I/P and P anchor pictures of the adjacent views. The nonanchor frames of the B views are predicted from four neighbouring frames: two from the temporal level and other two from the interview level. E.g., the frame (S1, T4) is predicted from (S1, T0) and (S1, T8) at the temporal level, and from (S0, T4) and (S2, T4) at the interview level. It has been shown in [24] that IBP provides a considerable gain in bitrate saving and video quality enhancement compared to the simulcast scheme. Furthermore, IBP scheme offers significant improvement in random access ability compared to IPP scheme. These two facts have qualified the IBP scheme to perform as a balanced interview prediction structure ensuring good compression efficiency and facilitating the random access ability. IBP was adopted as the default structure of the earlier JMVM and the latest Joint Multiview Video Coding (JMVC) [82] software for MVC standard.

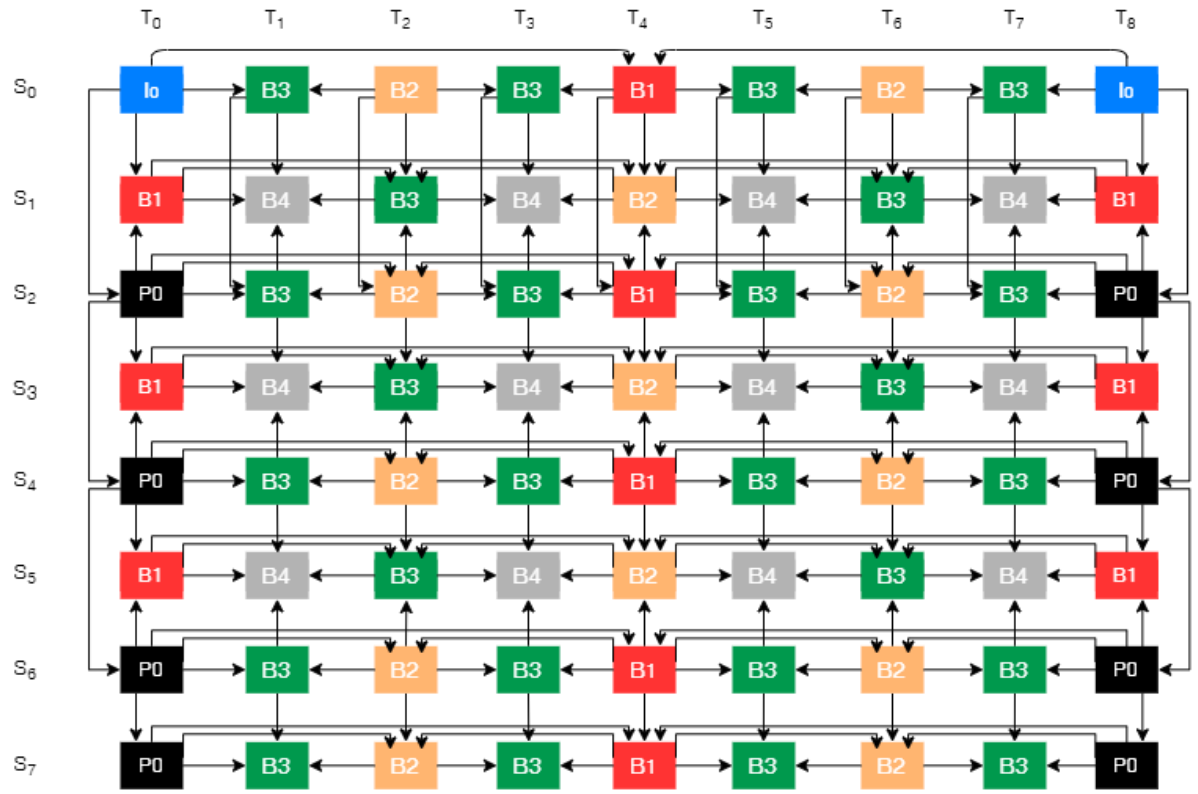



Figure 3.7 IBP interview prediction scheme

The following equation defines $N_{max}$ for IBP structure:

$$N_{\max} = 3 \times H_{\max} + 2 + 5 \times [Nbr_{view} - 1] \tag{3.2}$$

Note that $H_{max}$ is set to 4 for the IBP structure.

In [83], an approach based on using two successive B-views between two views of types I and P was proposed. It demonstrated that employing successive B-views can improve the bitrate saving gain and might enhance the random access if an appropriate interview prediction strategy was used. The hierarchical level of the used B frames was similar to that of the IBP structure. The obtained bitrate gain in this structure was due to the employment of more B-views which provide less bitrate when coded compared to I-view and P-views. The proposed scheme in [83] is referred to us as Amr structure.

The $N_{max}$ of Amr interview prediction structure [83] is given by:

$$N_{\max} = 3 + H_{\max} + 2 \times [(Nbr_{view} - 2)/3] \tag{3.3}$$

The obtained improvements of random access ability require further enhancement to ensure smoother view switching and better interactivity with the multiview video content. These improvements have to be done while paying attention to coding efficiency.

## 3.5 Proposed Framework for Random Access Enhancement

### 3.5.1 Proposed Approach

Random access ability can be improved if the chosen base view ensures a direct prediction to a maximum of non-base views within the multiview prediction scheme. Setting the I-view in a middle position [84] (instead of the first side position) and regularly involving B views could contribute to improving the random access ability.

The proposed approach [26] was designed to provide a direct interview prediction for all of the proposed structure' views. It is mainly based on employing two base views (I) per GGOP. The used base views are independently coded from any other views using temporal prediction based on B hierarchical algorithm.

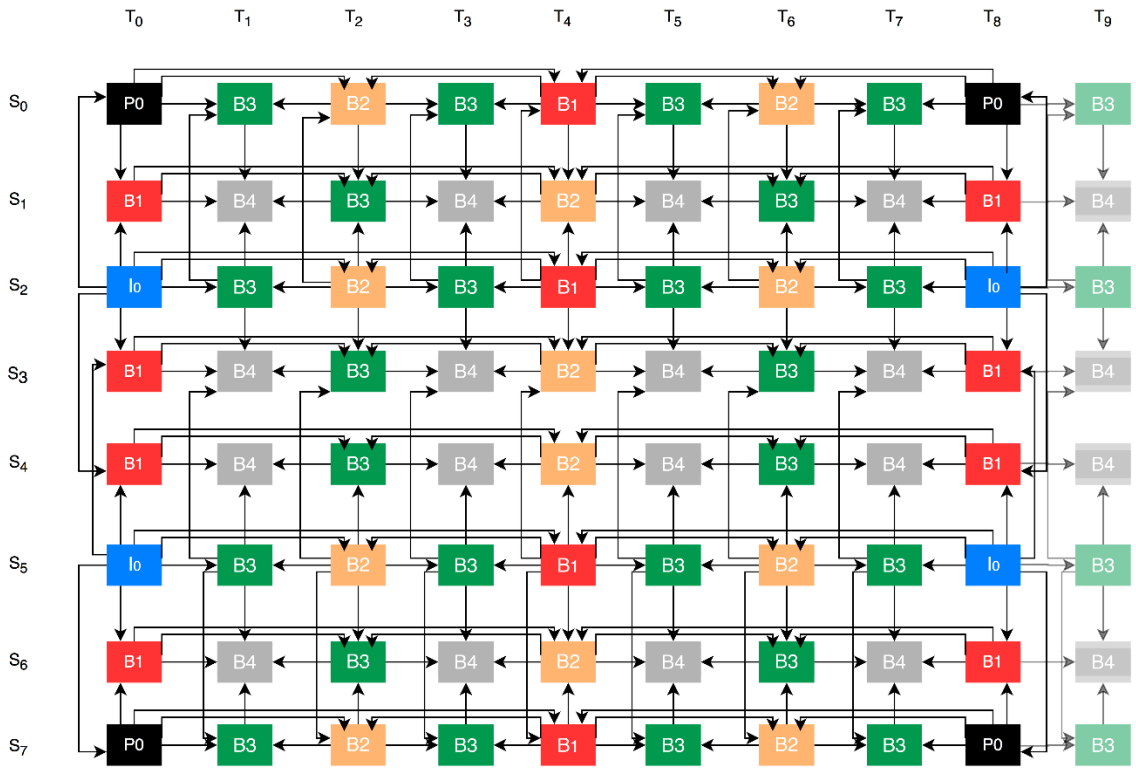



Figure 3.8 Proposed multiview prediction structure PBI

As shown in Figure 3.8, S2 and S5 are selected as optimal reference view positions allowing a direct interview prediction for all the remaining views (S0, S1, S3, S4, S6 and S7). The proposed scheme only contains two separated P-views (S0 and S7) leading to a less complex views dependencies. The P-views benefit from a direct dependency through S2 and S5, respectively. The P-view anchor frames are directly predicted from the adjacent I-view anchor frame. Whereas, the nonanchor frames of P-view have three reference frames, two from the same view in which they are located and one interview reference frame.

The proposed structure also includes four B-views (S1, S3, S4 and S6). The anchor frames of these views are bidirectionally predicted from two reference anchor pictures. The nonanchor frames of these views are predicted through four reference frames, equally divided between temporal and interview level. All these views benefit from a direct interview prediction. S3 and S4 were allocated in a way to form two sequential B-views which contributes to reducing the maximum number of the decoded frames for accessing a selected frame. B-views minimise the bitrate ratios compared to I-views and P-views.

Using B-views with these specific positions allowed our proposed design to achieve a competitive bitrate saving with good video quality. Furthermore, employing two I-views certainly reduces the encoding time and complexity, because no disparity compensation process will be used in all anchor and nonanchor frames of these views. Also, using two I-views results in doubling the number of the inserted intracoded frames (I frames) which means reducing the application of the motion compensation process. Hence, the proposed design further reduces the encoding time and complexity. I frames represent recovery point from errors in the video bitstream. Therefore, our design should provide an improved error robustness capability.

The proposed interview prediction scheme provides an additional backward compatibility due to the deployment of two I views that could be extracted by using the conventional H.264/AVC standard.

The proposed multiview prediction, referred to as PBI structure, provides two main view coding orders yielding the same coding results:

- $S_2$- $S_0$ - $S_1$ - $S_5$ – $S_3$ - $S_4$ - $S_7$ - $S_6$ I-P-B-I-B-B-P-B
- $S_2$ - $S_5$ - $S_0$ - $S_1$ - $S_3$ - $S_4$ - $S_7$ - $S_6$ I-I-P-B-B-B-P-B

Figure 3.9 illustrates the difference between the proposed structure PBI and the previously reported prediction structures by showing the anchor frames combinations of each scheme.

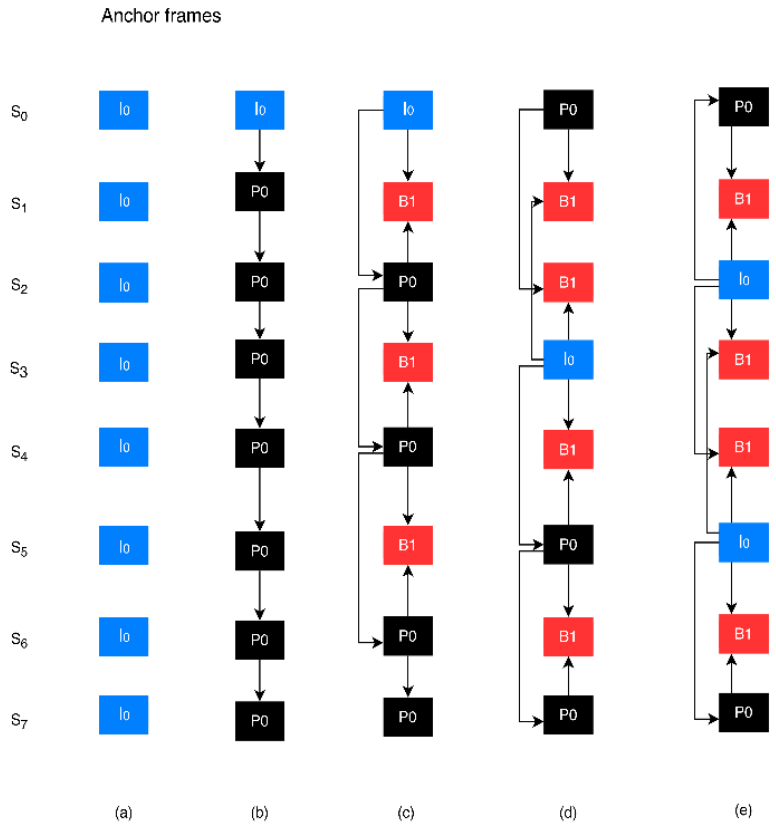



Figure 3.9 Interview schemes for anchor frames of: (a) Simulcast, (b) IPP, (c) IBP, (d) Amr, (e) proposed PBI [26]

The B4 frame ($S_6$, $T_1$) in Figure 3.10 has a hierarchical level of 4 which is the maximum level in our proposed structure PBI. Four reference frames in the temporal level are required to encode ($S_6$, $T_1$) frame. The locations of these four frames are ($S_6/T_0$, $S_6/T_2$, $S_6/T_4$ and $S_6/T_8$). The B4 frame ($S_6$, $T_1$) also needs five pictures from each adjacent view: I-view ($S_5$) and P-view ($S_7$). The locations of these interview frames are as follows: ($S_5/T_0$, $S_5/T_1$, $S_5/T_2$, $S_5/T_4$, $S_5/T_8$) and ($S_7/T_0$, $S_7/T_1$, $S_7/T_2$, $S_7/T_4$, $S_7/T_8$).

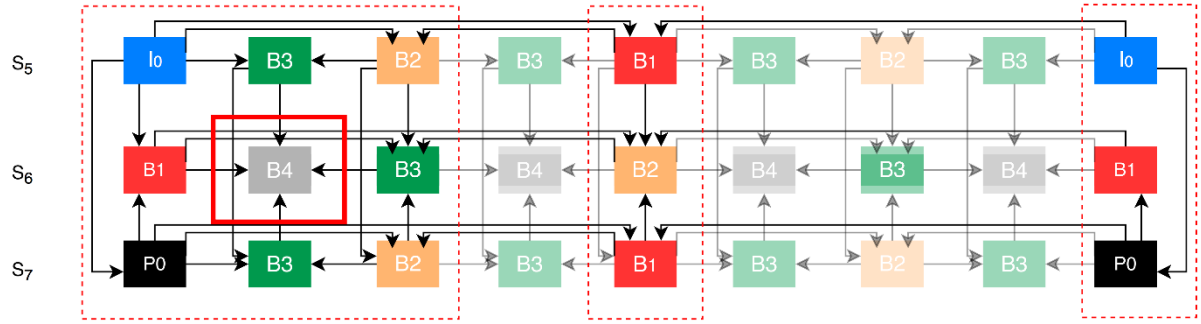



Figure 3.10 Random access scheme of the highest hierarchical B frame

The example of the B4 frame (S6, T1) is repeated four times in the same GOP of the same B-view (S6). This type of frame is also found in the rest of B-views (S1, S3 and S4). Consequently, the equation that describes the computation of the $N_{max}$ for the proposed inter-view prediction is deduced as follows:

$$N_{\max} = 3 \times H_{\max} + 2 \tag{3.4}$$

The maximum hierarchical level, $H_{max}$ in equation (3.4) is equal to four.

Table 3.1 compares $N_{max}$ equations of our proposed scheme PBI against IPP, IBP and Amr. It clearly shows the calculation simplicity that characterises PBI scheme. Only two arithmetic operations are used to calculate $N_{max}$ in PBI scheme, whereas the other schemes employ five arithmetic operations.

Table 3.1 $N_{max}$ Equations comparison

| | $N_{max}$ Equations |
|---|---|
| **IBP** | $N_{\max} = (H_{\max} + 1) + 5 \times [Nbr_{view} - 1]$ |
| **IPP** | $N_{\max} = 3 \times H_{\max} + 2 + 5 \times [Nbr_{view} - 1]$ |
| **Amr** | $N_{\max} = 3 + H_{\max} + 2 \times [(Nbr_{view} - 2)/3]$ |
| **PBI** | $N_{\max} = 3 \times H_{\max} + 2$ |

An additional random access evaluation metric was adopted in our research to provide an adequate comparison. In [83], $Nbr_{img}$ calculates the number of the decoded frames to access a given frame with regards to its type and position.

After analysing the different types of frame in our proposed scheme, $Nbr_{img}$ equations of PBI were deduced as follows:

A) For anchor frames of PBI:

- The number of encoded frames for accessing **I** picture is known to be equal to zero (**0**) in all structures.
- For **P** anchor frames (S0 and S7), **one** frame is needed for each of them.
- For the four **B** anchor frames (S1, S3, S4 and S6), each anchor B frame requires **two** frames to be encoded.

Equation 3.5 summarises the $Nbr_{img}$, for PBI anchor frames:

$$Nbr_{img} = \begin{cases} 0 & for\ I\ anchor\ frame \\ 1 & for\ P\ anchor\ frame \\ 2 & for\ B\ anchor\ frame \end{cases} \tag{3.5}$$

B) For non-anchor frames:

The number of the decoded pictures for accessing a given picture depends on its hierarchical level and its type of view (I, P or B). The different possibilities for calculating the $Nbr_{img}$ for the nonanchor frames of PBI scheme are given by:

$$Nbr_{img} = \alpha \times Hierarchy + \beta \tag{3.6}$$

$$Where \begin{cases} \alpha = 1,\ \beta = 0 & for\ I\ nonanchor\ frames \\ \alpha = 2,\ \beta = 1 & for\ P\ nonanchor\ frames \\ \alpha = 3,\ \beta = 2 & for\ B\ nonanchor\ frames \end{cases} \tag{3.7}$$

Table 3.2 regroups the various possible cases of calculating $Nbr_{img}$ in our PBI structure and the competitive schemes. It obviously shows the reduced complexity of our equations compared to the equations of the other schemes.

Table 3.2 Comparison between the equations computing the $Nbr_{img}$

| $Nbr_{img}$ | P anchor frames | B anchor frames | P non-anchor frames | B non-anchor frames |
|---|---|---|---|---|
| **PBI** | 1 | 2 | $2\times Hierarchy+1$ | $3\times Hierarchy+2$ |
| **IBP** | $\frac{Num_{view}}{2}$ | $1+\frac{Num_{view}}{2}$ | $(Hierarchy+1)+2\times[\frac{Num_{view}}{2}]$ | $3\times Hierarchy+2\times[\frac{Num_{view}}{2}]$ |
| **Amr** | $\frac{Num_{view}-2}{2}$ | $1+\frac{Num_{view}-3}{2}$ | $(Hierarchy+1)+2\times[\frac{Num_{view}-2}{3}]$ | $3\times Hierarchy+2\times[\frac{Num_{view}-1}{2}]$ |

$Num_{view}$ denotes the view number value, which can be 1, 2…8, and Hierarchy is the hierarchical level of the target frame.

## 3.5.2 Generalisation of the Proposed PBI

In this section, a general extension of the proposed PBI scheme, for cases when more than eight cameras are used, is developed. The PBI extended version maximises the use of B-views to ensuring a better compression efficiency and minimises the insertion of successive P-views to avoid slowing down the random access speed.

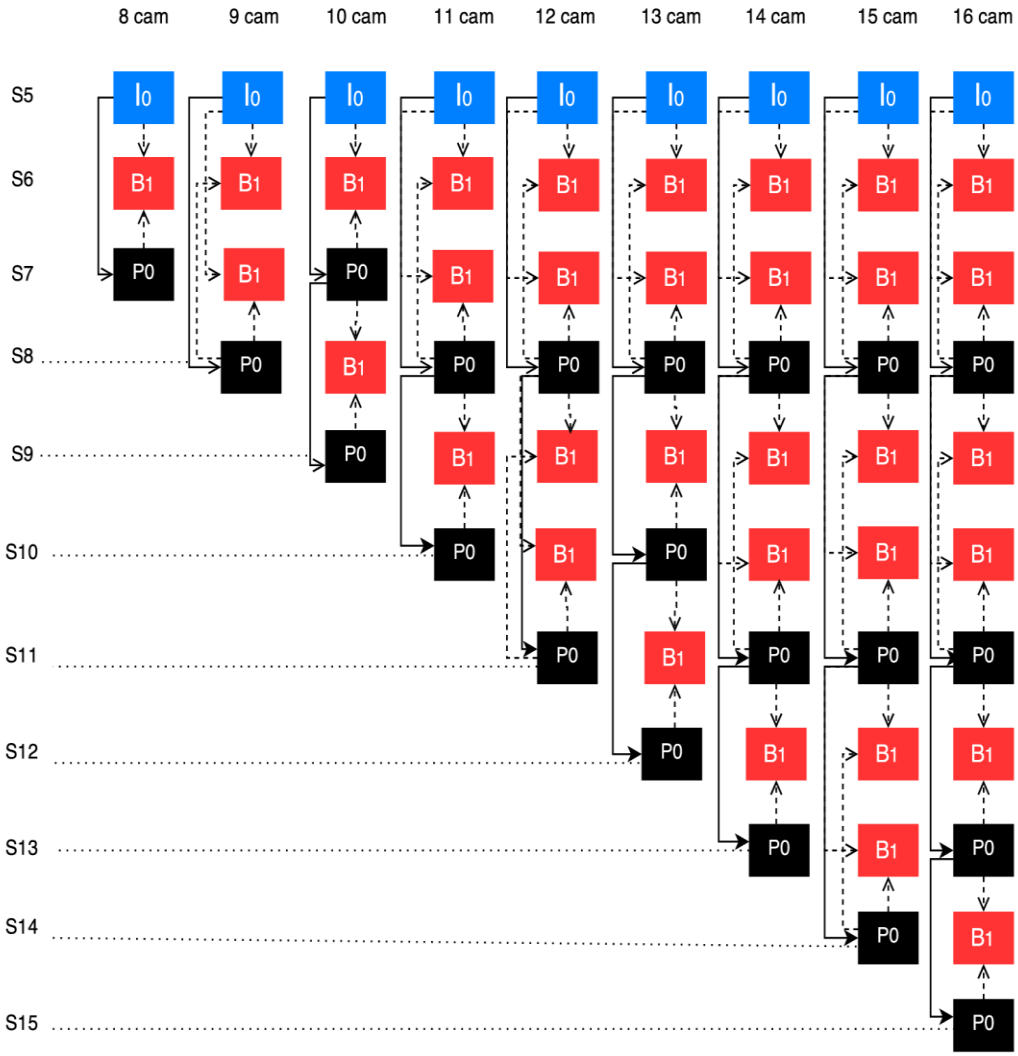



Figure 3.11 Anchor frames of the extended PBI scheme

Figure 3.11 illustrates the views order following the second base view S5 of the proposed scheme, where the ordering architecture differs based on the views number. A crucial condition is respected while designing the different scheme orders. It is about avoiding the use of successive P-views.

The PBI extended version defines three sequential orders according to the following formulas:

$$Order = \begin{cases} I/P, B, P & if\,(Nbr_{view}\ mod\ 3) = 2 \\ I/P, B, P, B, P & if\,(Nbr_{view}\ mod\ 3) = 1 \\ I/P, B, B, P & if\,(Nbr_{view}\ mod\ 3) = 0 \end{cases} \quad (3.8)$$

Each letter, I, P and B, corresponds to a view type, and the different orders represent the last views in every choice. The first order "$I/P, B, P$" is selected when the remainder of the division of the view number by 3 is equal to 2, such as 8, 11 and 14. Akin to the second choice "$I/P, B, P, B, P$", this order avoids the use of successive P-views. The third order "$I/P, B, B, P$" is the best possible choice as it allows the use of successive B-views and avoids the use of successive P-views. Additionally, it provides theoretically better results for both bitrate gain and random access ability. This order is selected when the remainder of the division of the view number by 3 is always equal to zero, e.g., the case of 9, 12 and 15 views. The three choices presented above are usually applied after the following order of views: "I, B, B" or "$P, B, B$"

The $N_{max}$ equation varies accordingly with the change of the views number in the interview structure. We have developed a general equation to calculate $N_{max}$, whatever is the number of views, as follows:

$$N_{\max} = 3 \times H_{\max} + 2 \times \left\lceil \frac{(Nbr_{view} - 5) + \alpha}{3} \right\rceil \quad (3.9)$$

$$Where \quad \alpha = \begin{cases} 0 & if\,(Nbr_{view} MOD\, 3) = 2 \\ -1 & if\,(Nbr_{view} MOD\, 3) = 0 \\ 1 & if\,(Nbr_{view} MOD\, 3) = 1 \end{cases} \quad (3.10)$$

$Nbr_{view}$ and $H_{max}$ denote respectively the views number and the highest hierarchical B frame level in the coding structure. The PBI extended structure allows a fast view random access, especially when several pairs of successive B-views are introduced in the design. It also provides an overall similar video quality, measured in PSNR, compared to IBP, IPP and Amr [83]. Details of the compression efficiency are provided in section 3.7.

Also, regardless the considered views number, equations which calculates the $Nbr_{img}$ for all PBI extended possible cases are developed as follows:

- For the anchor picture:

$$Nbr_{img} = \beta + \left\lfloor \frac{(V_p - 5) + \alpha}{3} \right\rfloor \tag{3.11}$$

$$Where \begin{cases} \beta = 0 \ for\ P\ anchor\ \ frames \\ \beta = 1 \ for\ B\ anchor\ \ frames \\ \qquad and \\ \alpha = \begin{cases} 1 & if\ (V_p\ MOD\ 3) = 1 \\ 0 & if\ (V_p\ MOD\ 3) = 2 \\ -1 & if\ (V_p\ MOD\ 3) = 1 \end{cases} \end{cases} \tag{3.12}$$

$V_p$ denotes the P-view number. It is also used to calculate the $Nbr_{img}$ needed for accessing the B-view' anchor frames. For example, for the case of 12 views, to calculate the $Nbr_{img}$ of the two anchors B frames S9 and S10, $V_p$ will be set to 12 and β to 1. Thus, the $Nbr_{img}$ is equal to three for both S9 and S10.

- For nonanchor picture:

$$Nbr_{img} = \beta \times Hierarchy + 2 \times \left\lfloor \frac{(V_p - 5) + \alpha}{3} \right\rfloor \tag{3.13}$$

$$Where \begin{cases} \beta = 1 \ \ for\ P\ nonanchor\ \ frames \\ \beta = 3 \ for\ B\ nonanchor\ \ frames \\ \qquad and \\ \alpha = \begin{cases} 1 & if\ (V_p\ MOD\ 3) = 1 \\ 0 & if\ (V_p\ MOD\ 3) = 2 \\ -1 & if\ (V_p\ MOD\ 3) = 0 \end{cases} \end{cases} \tag{3.14}$$

Hierarchy is the hierarchical level of the picture which can be set to 2, 3 or 4. For the nonanchor frames, $V_p$ is used in the same way as in equation (3.11).

## 3.6 Random Access Evaluation of the Proposed PBI Scheme

In this section, we provide the random access evaluation of the proposed framework PBI. Two methods are used during the evaluation process:

- The standard $N_{max}$ which is usually used to measure the random access performance of the multiview coding structures.
- A novel proposed metric, referred as the global random access ability $G_{RA}$, including three stages of assessment in which the multiview coding structure is holistically evaluated.

For both metrics the proposed framework compared against IBP and Amr prediction structures.

### 3.6.1 Random access assessment using $N_{max}$

The maximum number of reference frames $N_{max}$ value depends on multiple parameters, such as the views number of the system, the number of reference views (I-view) and their positions in the structure, the hierarchical level of the GOP design and the GOP size. The selected locations of the two reference views in the PBI structure have contributed to significantly reduce $N_{max}$ value.

The effectiveness in random access of any proposed scheme must be compared to the benchmark scheme IBP. Random access gain of any compared structure to IBP using $N_{max}$ is given as follows:

$$\Delta N_{\max} = \frac{N_{\max}(IBP) - N_{\max}(compared)}{N_{\max}(IBP)} \times 100\% \tag{3.15}$$

$N_{max}(compared)$ represents the $N_{max}$ value for either PBI or Amr [83] structure.

Although the proposed PBI structure provides good random access results regardless of the deployed GOP size, this latter is set to eight for all the considered schemes following the default design of the MVC standard.

Figure 3.12 and Figure 3.13 illustrate the superiority of the proposed PBI scheme through different views numbers, against IBP and Amr in terms of random access ability. The proposed scheme PBI maintains good random access ability despite the number of the deployed views. Noticeably, the best results are achieved when using more successive B-views. This was ensured by selecting the third choice mode presented in section 3.5.2 "$I/P, B, B, P$", where all B-views after the second base view are successive. For instance, for 15 views schemes, the $N_{max}$ of the proposed PBI structure is equal to 18 and that of the IBP structure is equal to 26. The $\Delta N_{max}$ gain, in this case, exceeded slightly 30% marking the largest obtained gain.

Figure 3.12 shows that the minimum $\Delta N_{max}$ gain of the proposed PBI over IBP was 20%, when both structures are composed of 10 views. This is due because of the absence of successive B-views and the increased use of I-views after the second base view.

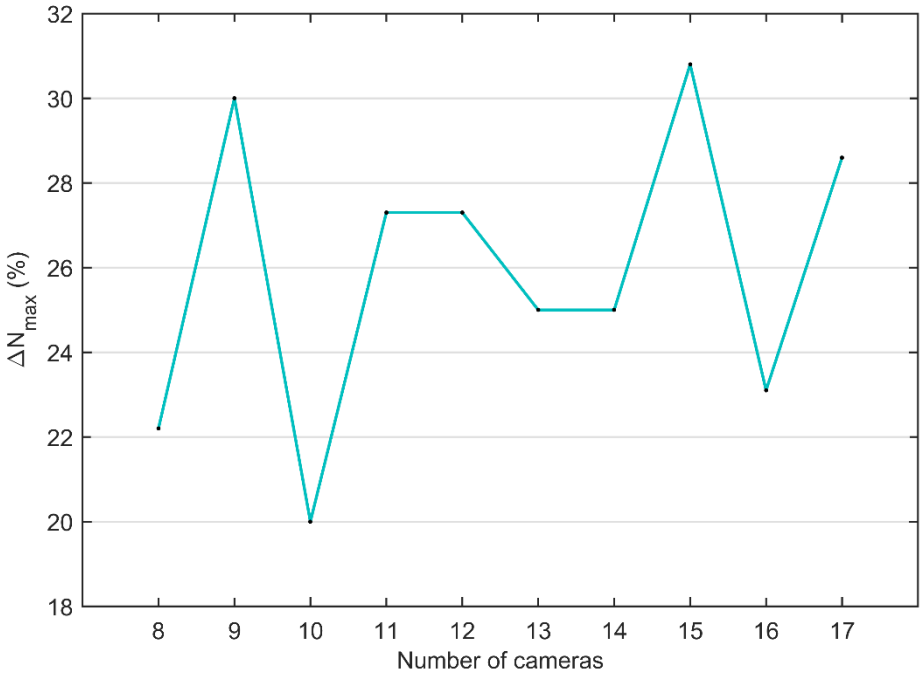



Figure 3.12 Random access gain of PBI scheme over IBP [26]

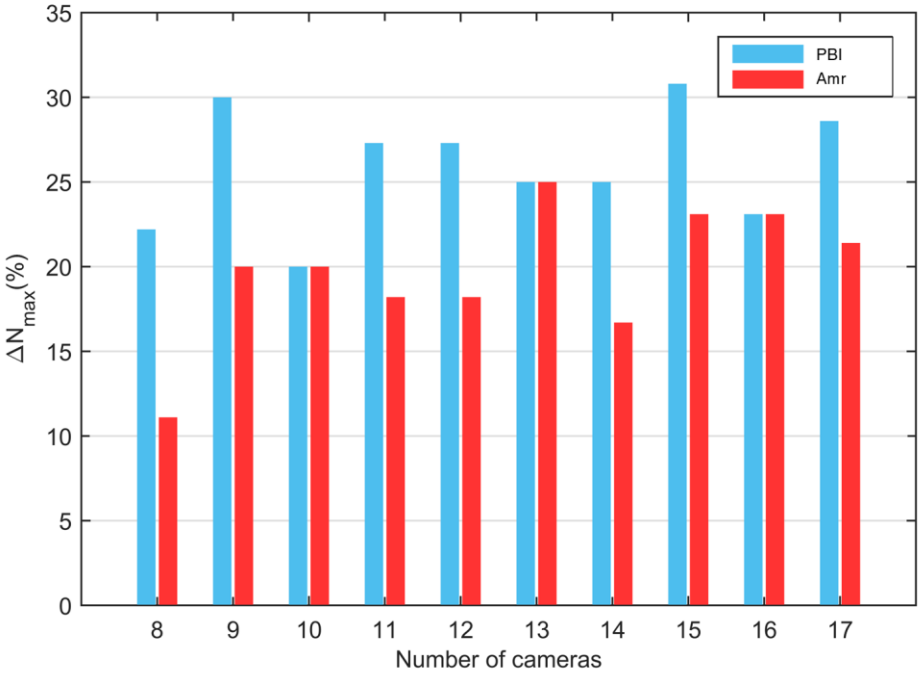



Figure 3.13 Random access gain with respect to IBP structure: comparison between PBI scheme and Amr structure [26]

Figure 3.13 illustrates a comparison, regarding the random access gain, between the proposed PBI [26] and Amr [83] schemes. It shows in overall that the proposed PBI scheme exceeds Amr structure with an average gain of about 11%. However, for 10, 13 and 16 views cases, both schemes exceptionally show similar behaviour. The reason for that is the use of the second choice mode, presented Section 3.5.2, equation (3.8) $"I/P,B,P,B,P"$. The obtained values of $N_{max}$ for the three reported structures as well as the relative gain $\Delta N_{max}$ of (PBI/IBP), (PBI/Amr), and (Amr/IBP) are reported in Table 3.3. The results show clearly that the proposed PBI structure significantly reduces the maximum number $N_{max}$ of the reference images needed for decoding a given frame, which in turn leads to a multiview video coding structure with enhanced random accessibility.

Table 3.3 Nmax and $\Delta$ Nmax gain of the proposed PBI structure in comparison to IBP and Amr structures. [26]

| View Number | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|
| $N_{max\ IBP}$ | 18 | 20 | 20 | 22 | 22 | 24 | 24 | 26 | 26 | 28 |
| $N_{max\ PBI}$ | 14 | 14 | 16 | 16 | 16 | 18 | 18 | 18 | 20 | 20 |
| $N_{max\ Amr}$ | 16 | 16 | 16 | 18 | 18 | 18 | 20 | 20 | 20 | 22 |
| $\Delta N_{max\ (PBI\ /IBP)\ (\%)}$ | 22.2 | 30.0 | 20.0 | 27.3 | 27.3 | 25.0 | 25.0 | 30.8 | 23.1 | 28.6 |
| $\Delta N_{max\ (PBI/Amr)\ (\%)}$ | 12.5 | 12.5 | 0.0 | 11.1 | 11.1 | 0.0 | 10.0 | 10.0 | 0.0 | 9.1 |
| $\Delta N_{max\ (Amr/IBP)\ (\%)}$ | 11.1 | 20.0 | 20.0 | 18.2 | 18.2 | 25.0 | 16.7 | 23.1 | 23.1 | 21.4 |

## 3.6.2 Random Access Evaluation using a Proposed Metric

The standard $N_{max}$ evaluates the random access performance through only one frame of the structure, i.e. the frame possessing the highest hierarchical level in the GGOP. All

the remaining frames are neglected during the evaluation process. $N_{max}$ could somehow lead to incorrect or not fair enough conclusions. This fact has been a strong motivation to propose and use a new approach to accurately evaluate the proposed PBI framework. The evaluation method allows a more in-depth look into the considered multiview scheme.

It is based on the calculation of the average cost for accessing each existing frame within the multiview prediction structure.

The novel metric is composed of three phases. Firstly, a global evaluation of the anchor pictures of each studied structure is carried out. Secondly, the nonanchor frames of different hierarchical level are evaluated. Finally, an evaluation covering the entire multiview prediction structure is performed.

The access speed to the anchor pictures ($G_{RA}$) is estimated through measuring the average cost of arbitrarily accessing all anchor pictures of the examined structure. We have developed $G_{RA}$ to be equal to the sum of the random access cost of the anchor frames divided by the number of views. $G_{RA}$ is given as follows:

$$G_{RA} = \frac{\sum_{i=1}^{V_n} Nbr_{img}(i)}{V_n} \tag{3.16}$$

$Nbr_{img}$: is the number of the encoded pictures to access an anchor frame.

$V_n$: is the number of the views in the structure.

Table 3.4 regroups $Nbr_{img}$(x)values of the Proposed structures PBI, IBP default structure and Amr scheme. The variable (x) refers to the view number.

Table 3.4 $Nbr_{img}$ to access anchor pictures following the view order

| | $Nbr_{img\ 0}$ | $Nbr_{img\ 1}$ | $Nbr_{img\ 2}$ | $Nbr_{img\ 3}$ | $Nbr_{img\ 4}$ | $Nbr_{img\ 5}$ | $Nbr_{img\ 6}$ | $Nbr_{img\ 7}$ |
|---|---|---|---|---|---|---|---|---|
| **IBP** | 0 | 2 | 1 | 3 | 2 | 4 | 3 | 4 |
| **Amr** | 1 | 2 | 2 | 0 | 2 | 1 | 3 | 2 |
| **PBI** | 1 | 2 | 0 | 2 | 2 | 0 | 2 | 1 |

By applying equation (3.16), we obtain the following $G_{RA}$ values for the three examined schemes: $G_{RA}IBP = 2.37$, $G_{RA}Amr = 1.62$ and $G_{RA}PBI = 1.25$.

To determine the $G_{RA}$ gain, we use the following (3.17) formula:

$$\Delta G_{RA} = \frac{G_{RA}(compared) - G_{RA}(PBI)}{G_{RA}(compared)} \times 100\% \tag{3.17}$$

$G_{RA}$ (compared) takes either the value of IBP or that of Amr. Figure 3.14 portrays the significant $G_{RA}$ gains achieved by our proposed structure against IBP and Amr. PBI structure speeds up the random access to the anchor frames by 47.25% and 22.83% compared to IBP and Amr structures respectively.

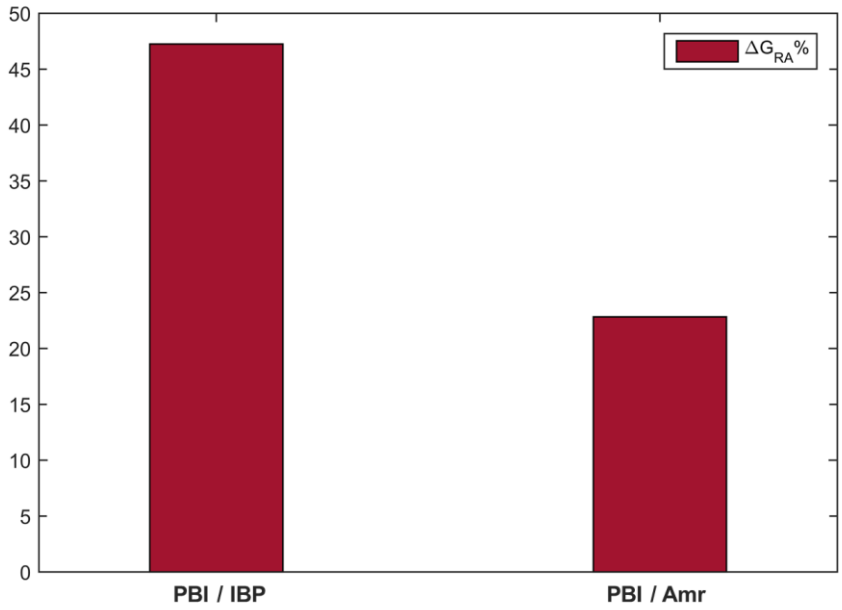



Figure 3.14 $G_{RA}$ gains of the proposed PBI structure compared to IBP and Amr [26]

The second evaluation stage consists of measuring the average cost of predicting nonanchor frames of the examined multiview structure. The global random access cost for nonanchor frames is developed as follows:

$$G_{RN} = \frac{\sum_{i=1}^{Vn} \sum_{t=1}^{GOP(size)-1} [Nbr_{img(i,t)}]}{GGOP(size) - V_n} \tag{3.18}$$

$G_{RN}$ is defined to be equal to the sum of $Nbr_{img}(i,t)$ divided by the number of the crossed frames during the calculation process which is equal to $GGOP(size) - V_n$. $Nbr_{img}(i,t)$ is the encoded picture number to access a non-anchor picture at view level position $i$ and instant position $t$. $V_n$ is the number of used views, and *GGOP(size)* is equal to $V_n$ multiplied by the GOP size.

After extracting 56 $Nbr_{img}(i,t)$ values of the nonanchor frames of the examined structures, formula (3.18) of $G_{RN}$ is applied separately to IBP, Amr and PBI. The obtained $G_{RN}$ results are $G_{RN}$ IBP = 10.41, $G_{RN}$ Amr = 9.85 and $G_{RN}$ PBI = 9.10.

By adapting the equation (3.17) for measuring $G_{RN}$ gains, and as shown in Figure 3.15, the PBI proposed structure enhances the random access ability by about 12.52% and 7.61% with respect to the IBP and Amr schemes, respectively.

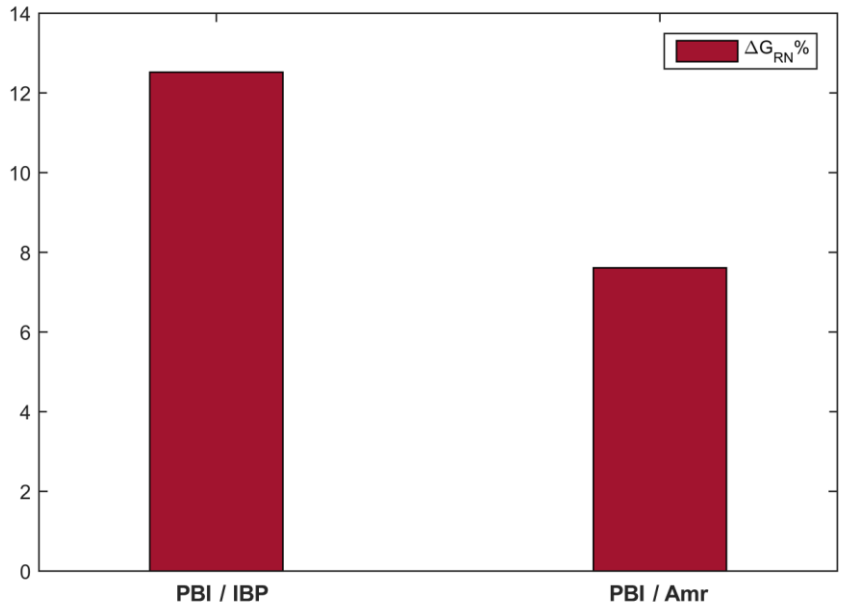



Figure 3.15 $G_{RN}$ gains of the proposed PBI scheme against IBP and Amr

The last phase of our proposed evaluation metric involves measuring the global random cost including anchor and nonanchor frames. $G_R$ is calculated through the following equation:

$$G_R = \frac{\sum_{i=1}^{V_n} \sum_{t=1}^{GOP(size)} [Nbr_{img}(i,t)]}{GGOP(size)} \tag{3.19}$$

The $G_R$ metric enables a holistic random access assessment of any examined multiview coding structure. It considers the cost of arbitrarily accessing each picture within the GGOP, regardless of its type and hierarchical level. Since $G_R$ reflects the average random access cost of the whole examined scheme.

$G_R$ might be considered as the fairest evaluation for any MVC structure. By making use of equations (3.19) and (3.17), the $G_R$ gains are reflected in Figure 3.16. It can be clearly inferred that PBI scheme offers better random access ability compared to IBP and Amr with increases of 13.5 % and 8 % respectively.

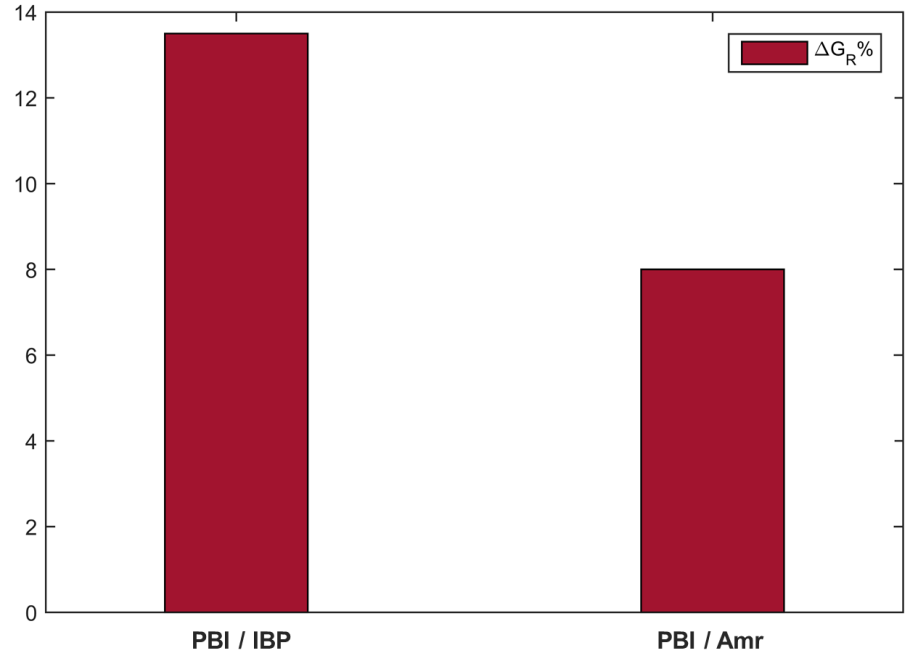



Figure 3.16 The global random access $G_R$ gain of the proposed PBI scheme[26]

## 3.7 Compression Efficiency evaluation of the PBI Scheme

This section presents and discusses the compression performance evaluation of our proposed multiview prediction structure PBI. Compression efficiency is reflected in video quality and data volume rate of the compressed video.

To evaluate compressed multiview video sequences, both subjective and objective methods can be applied.

The performance of the considered multiview video structure can be evaluated by simply comparing original and reconstructed multiview video sequences.

Certain conditions are required to be respected in order to conduct reliable subjective experiments based on HVS of video quality assessment. Subjective methods for the assessment of stereoscopic and multiview 3DTV systems are recommended by ITU-R (Radiocommunication Sector of ITU) in ITU-R BT.2021 [85].

Objective evaluation of multiview video coding relies on comparing the pixel values of the input and output video frames using some mathematical criteria such as signal-to-noise ratio (SNR), peak-to-peak SNR (PSNR) [86] and the mean-squared-error (MSE). Additionally, the bitrate of the compressed video must be expressed and included in the comparison for a fair evaluation of the different multiview video coding schemes.
In this research, the proposed PBI structure and the related structures were objectively evaluated using PSNR values expressed in decibel (dB) and bitrate variations expressed in bits per second (bit/s). The PSNR is given as follows:

$$PSNR = 10 \times \log_{10}\left(\frac{255^2}{MSE}\right) \tag{3.20}$$

The MSE represents the mean square error between the compressed and the original video signal. Typically, PSNR values for video compression lie within the range between 30 and 50 (dB).
When comparing multiview video coding algorithms, the average PSNR value of the different views is considered. Additionally, PSNR value of the luminance signal is practically sufficient for comparing the video quality.

### 3.7.1 Source Material and Test Conditions

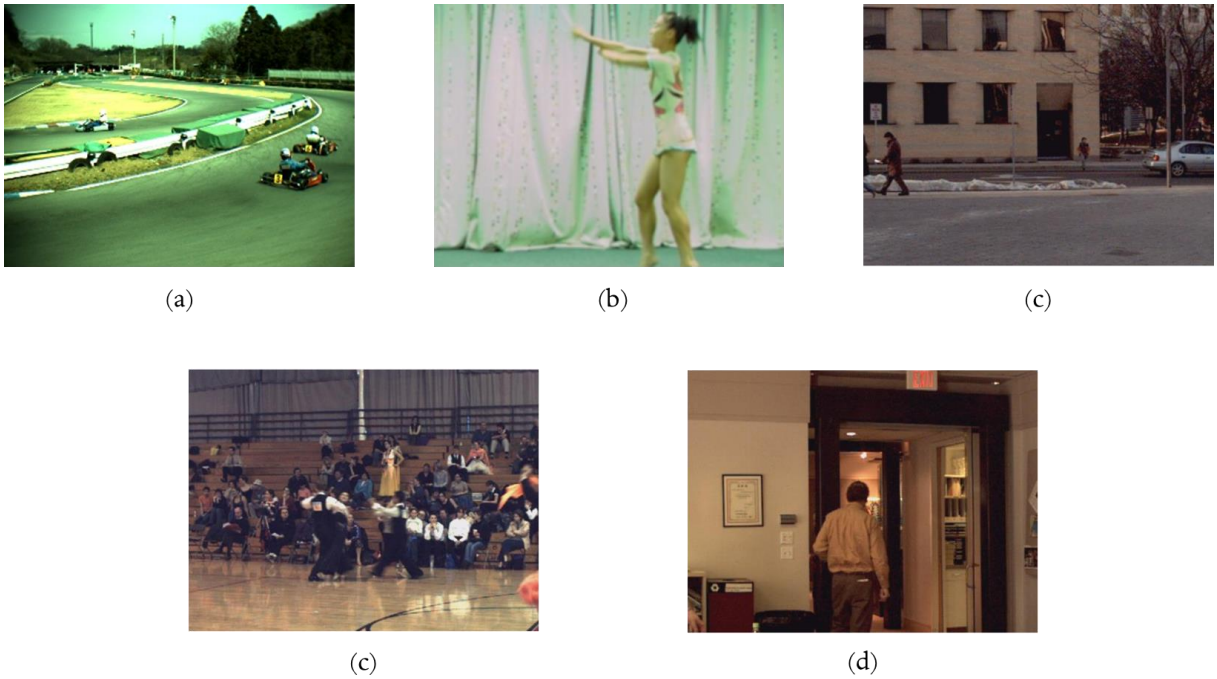
In order to evaluate compression efficiency of the proposed multiview scheme, various test sequences recommended by the JVT [87] have been used.

Table 3.5 reports the employed test sequences and their specific parameters including frame rate (fps), image resolutions, number of cameras, camera arrangements, and the distance between cameras. Exit and Vassar are examples of multiview sequences with low motion content whereas Race1, Ballroom and Rena are characterised with high motion content.

Figure 3.17 shows one view based samples of the utilised test sequences. Parameters of the tested sequences are included in the configuration file. This latter is an integral part of the supplemental enhancement information (SEI) messages of the bitstream that will be transmitted to the decoder side.

Table 3.5 Multiview video sequences for compression performance tests [26]

| Database | Sequences | Frame rate | Image resolution | Camera parameters |
|---|---|---|---|---|
| KDDI | Race1 | 30 fps | 640×480 | 8 cameras, 20 cm spacing, 1-D parallel |
| Tanimoto Lab | Rena | 30 fps | 640×480 | 100 cameras, 5cm spacing, 1-D parallel |
| MERL | Vassar | 25 fps | 640×480 | 8 cameras, 20 cm spacing, 1-D parallel |
| | Ballroom | 25 fps | 640×480 | 8 cameras, 20 cm spacing, 1-D parallel |
| | Exit | 25 fps | 640×480 | 8 cameras, 20 cm spacing, 1-D parallel |



(a) (b) (c)

(c) (d)

Figure 3.17 Multiview test sequences: (a) Race1, (b) Rena, (c) Vassar, (c) Ballroom and (d) Exit.

The same encoding configuration has been employed in order to come up with a fair judgement about the simulated multiview coding structures. The main encoding parameters are presented in Table 3.6.

Table 3.6 Encoding configuration

| Parameter | Setting |
|---|---|
| Symbol mode | CABAC |
| Quantisation parameter (QP) | 22, 27, 32, 37, 40 |
| GOP size | 8 |
| Search mode | Fast search |
| Search range | 64 |

The symbol mode specifies the used entropy coding mode which was the context-adaptive binary arithmetic coding (CABAC) for all tests. CABAC usually enhance the coding efficiency.

QP is the parameter that controls the compression ratio of the considered video codec. Results are presented for five QP. The GOP sizes is set to 8 for all simulations. The fast motion search algorithm is enabled for a maximum search range of 64.

### 3.7.2 Obtained Results and Discussions

Experimental tests are carried out to assess the compression efficiency of the proposed multiview video coding scheme PBI. The main results are expressed in graphs of PSNR (dB) versus bitrate (kbps).

All simulations were carried out using the five QP values that are mentioned in Table 3.6. The QP controls the compression performance of the coding structure. A negative relationship is distinguishable between QP and the compression efficiency. The bitrate and PSNR values increase as the QP decreases, and vice versa. Table 3.7 shows an example of the QP effects on the compression performance of the proposed coding structure PBI. The bitrate and PSNR values in Table 3.7 stand for the average value of eight views.

Table 3.7 QP effects on the compression efficiency of PBI

| QP | Q= 22 | Q= 27 | Q= 32 | Q= 37 | Q= 40 |
|---|---|---|---|---|---|
| Bit rate (kbps) | 1278.602 | 533.489 | 266.182 | 152.5808 | 109.564 |
| PSNR (dB) | 40.378 | 38.778 | 36.847 | 34.569 | 32.949 |

Figure 3.18 Compression performance views distribution of the PBI scheme shows the compression performance results of the PBI coding structure on ballroom MVV.

Eight video sequences are jointly encoded in this test. Figure 3.18 depicts the bitrate and quality distributions level of the eight encoded views.

It can be inferred that the two reference views S2 and S5 (I-views) are yielding larger bitrate values while providing a good video quality. It seems obvious that these results are due to the exploitation of the temporal prediction and neglection of the interview prediction.

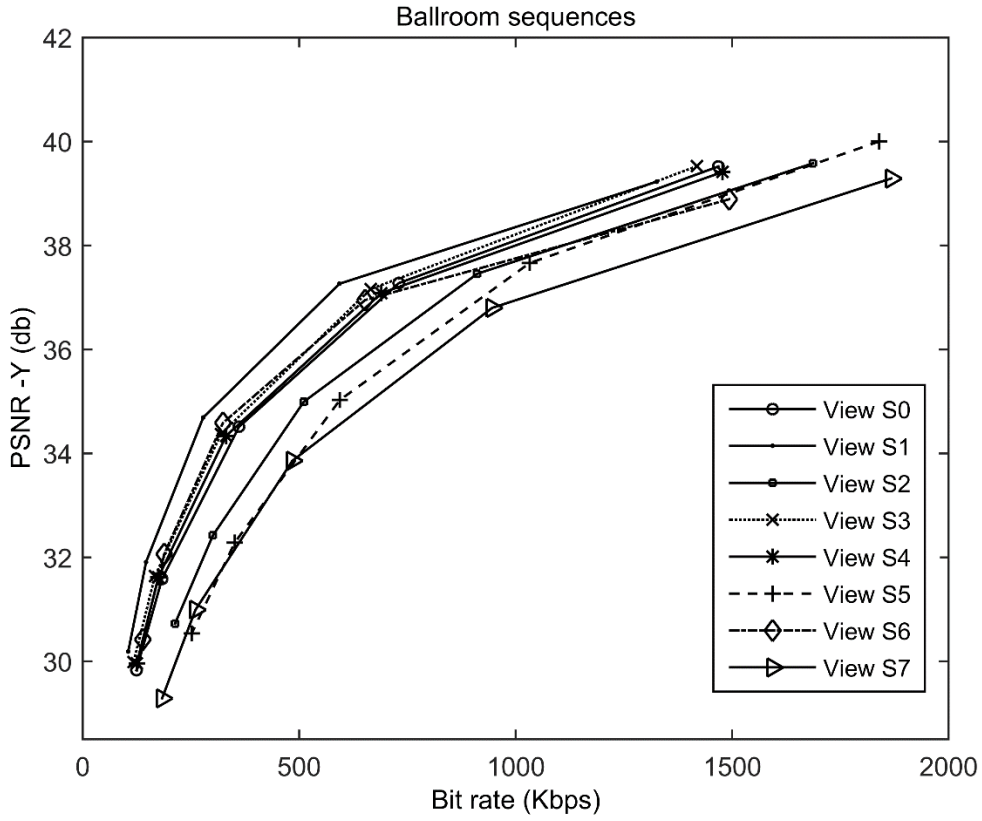
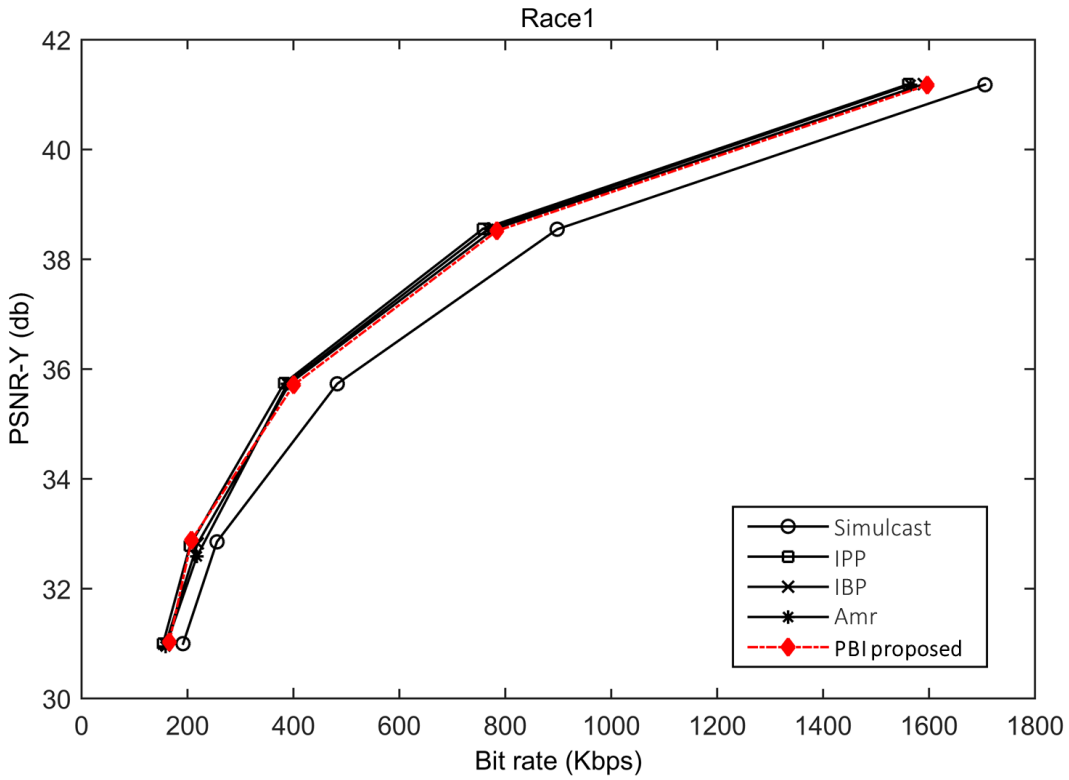
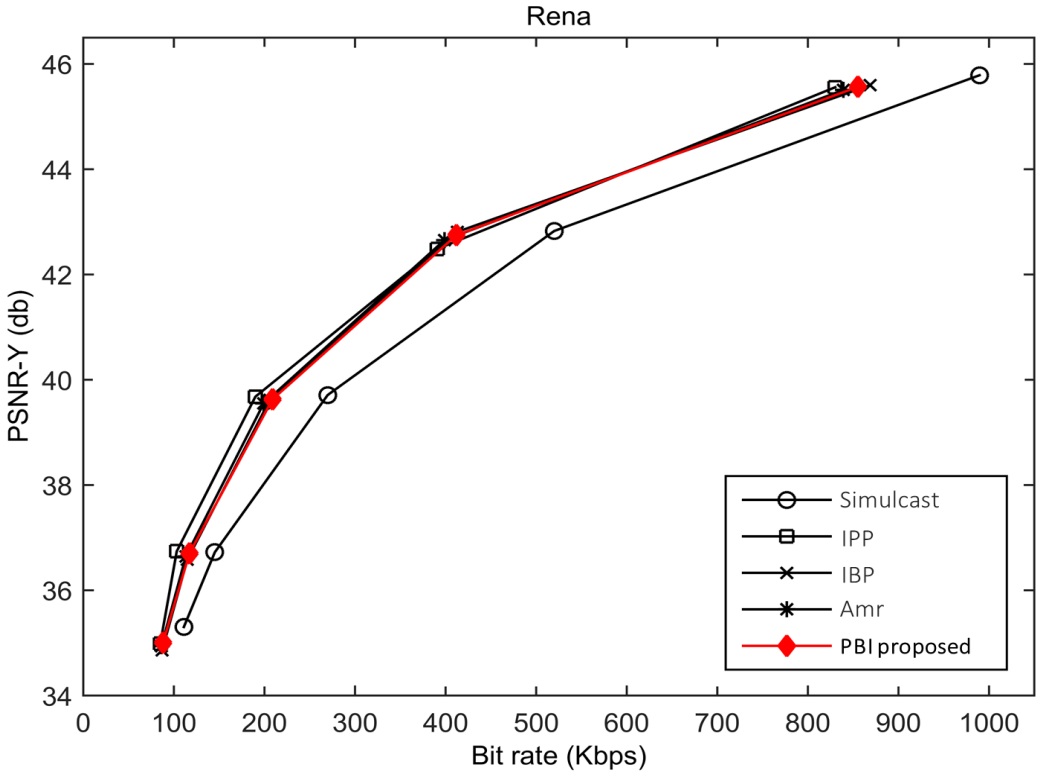



Figure 3.18 Compression performance views distribution of the PBI scheme

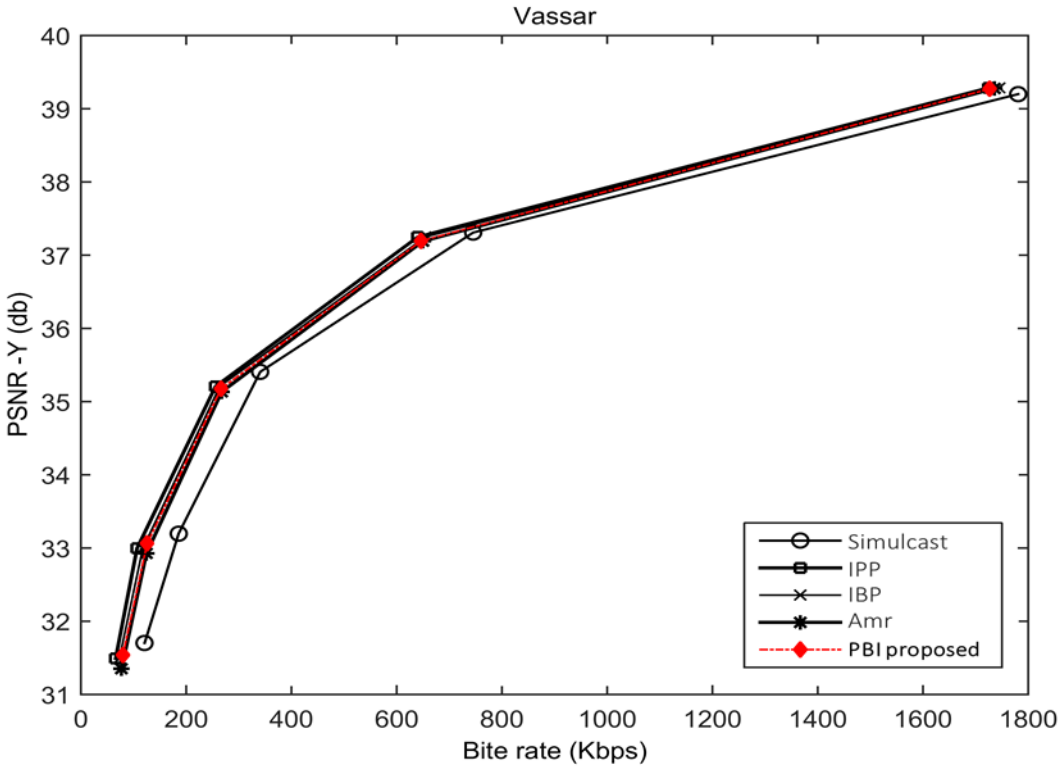
Figure 3.19 illustrates compression performance comparison between the proposed PBI scheme and four prediction structures (Simulcast, IBP, IPP and Amr) using the five MVV sequences as mentioned in Table 3.6.


Race1
PSNR-Y (db)
Bit rate (Kbps)
Simulcast
IPP
IBP
Amr
PBI proposed

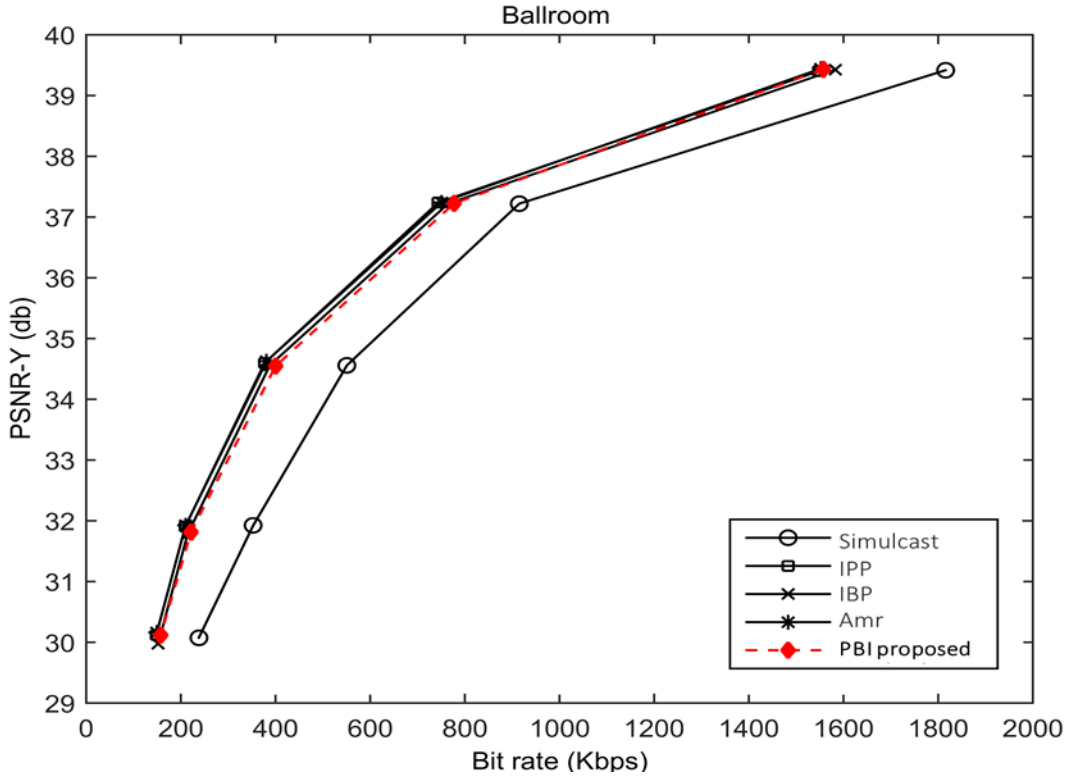
(a)


Rena
PSNR-Y (db)
Bit rate (Kbps)
Simulcast
IPP
IBP
Amr
PBI proposed

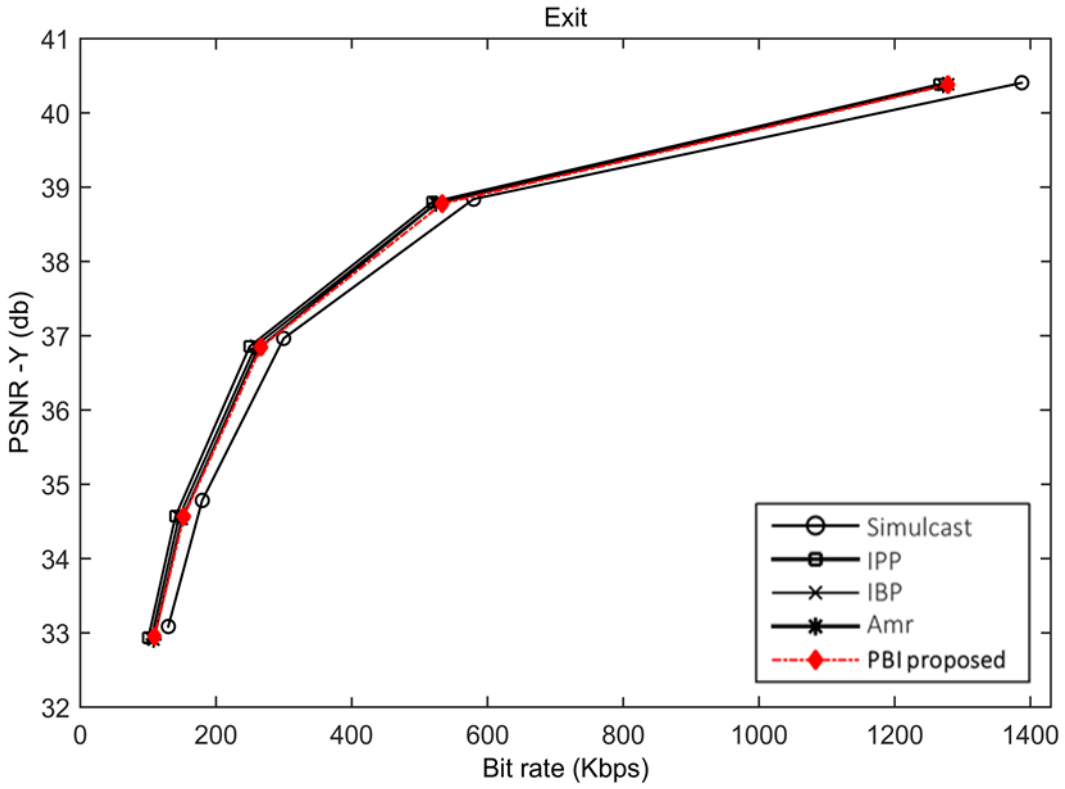
(b)


Vassar
PSNR -Y (db)
Bite rate (Kbps)
Simulcast
IPP
IBP
Amr
PBI proposed

(c)


Ballroom
PSNR-Y (db)
Bit rate (Kbps)
Simulcast
IPP
IBP
Amr
PBI proposed

(d)




(e)

Figure 3.19 Compression performance comparison using different MVV sequences: (a) Race1, (b) Rena, (c) Vassar, (d) Ballroom and (e) Exit [26]

It is easily observed from figure 3.19 that our proposed PBI structure outperforms simulcast structure and provide relatively similar compression performance compared to IPP, IBP and Amr. For the majority of the tested MVV sequences, IPP scheme produces a rather better bitrate saving despite its high complexity and slow random access ability against the other examined schemes. The video quality expressed in PSNR is practically similar for all the examined structures.

The compression performance of a given video encoding structure should be demonstrated at an appropriate quantisation parameter value that ensures a high video quality offering the viewer a clear and comfortable watching experience. Thus, the results obtained for QP = 22, which guarantee the best video quality among the five used QP values, are reported in Table 3.8.

Note that these results only cover the comparison of the proposed structure against IBP and Amr structures, because of the low bitrate saving and the poor random access performance that simulcast and IPP schemes provide, respectively.

Bitrate saving and PSNR gains are extracted from the following formulas:

$$\Delta_{BItrate} = \frac{bitrate_{compared} - bitrate_{proposed}}{bitrate_{compared}} \text{x}100\% \quad (3.21)$$

$$\Delta PSNR_Y = \frac{PSNR_Y(proposed) - PSNR_Y(compared)}{PSNR_Y(compared)} \times 100\% \quad (3.22)$$

$bitrate_{compared}$ and $PSNR_{compared}$ take the values of the considered structures to be compared against our proposed scheme.

Table 3.8 Compression efficiency evaluation of the proposed PBI structure

| | | IBP | | Amr | |
|---|---|---|---|---|---|
| | QP | Δ PSNR % | Δ bit-rate % | Δ PSNR % | Δ bit-rate % |
| Ballroom | 22 | 0.013 | 1.59 | -0.02 | -0.35 |
| Race 1 | 22 | - 0.043 | -0.41 | -0.016 | 0.0018 |
| Vassar | 22 | -0.0341 | 1.0426 | 0.0108 | 0.5156 |
| Exit | 22 | 0.13 | 0.17 | 0.007 | -0.53 |
| Rena 8 $_{views}$ | 22 | -0.07 | 1.558 | 0.131 | -1.9 |
| Rena 9 $_{views}$ | 22 | -0.015 | 2.94 | 0.116 | -1.72 |
| Rena 10 $_{views}$ | 22 | -0.061 | 0.557 | 0.087 | -0.189 |
| Rena 11 $_{views}$ | 22 | -0.029 | 0.676 | 0.17 | -3.177 |
| Rena 12 $_{views}$ | 22 | 0.32 | 2.86 | 0.131 | -3.63 |
| Rena 13 $_{views}$ | 22 | -0.002 | 1.60 | 0.17 | -3.74 |
| Rena 14 $_{views}$ | 22 | -0.021 | 3.18 | 0.131 | -3.15 |
| Rena 15 $_{views}$ | 22 | -0.012 | 2.11 | 0.109 | -1.61 |
| Rena 16 $_{views}$ | 22 | 0.0023 | -0.78 | 0.121 | - 2.56 |

Overall, the results in Table 3.8 indicate that the proposed PBI structure has almost similar PSNR behaviour compared to IBP and Amr. The proposed PBI structure provides better bitrate savings against IBP scheme, with an average of about 1.3%. Furthermore, PBI scheme seems to be slightly less efficient than Amr with an average bitrate loss of approximately 1.7%. However, this slight decline in bitrate saving is overlooked if we consider the significant random access improvement of PBI over Amr, which has achieved 22.83% using $G_{RA}$ metric.

## 3.8 Summary of PBI evaluation

Improving the random access is the main purpose of the proposed PBI scheme. This goal has been addressed and momentous results were obtained as previously detailed in sections 3.6.1 and 3.6.2. Moreover, the proposed PBI structure provides an interview scheme with less complexity for reducing the encoding time of any frame within the structure. Additionally, PBI scheme adequately satisfies the compression efficiency requirement. Figure 3.19 has illustrated the good compression performance of PBI scheme. Similar PSNR behaviour has been noticed across the reported structures in Table 3.8.

In general, the experimental evaluations have shown that the proposed PBI structure satisfies the compression efficiency requirement and enhances the random access ability. PBI would be the better choice for application where interactivity, navigation and view switching are highly required.

## 3.9 Conclusion

In this chapter, an overview of the MVC coding was presented. Different multiview coding technologies were reported such as checkerboard decomposition, distributed video coding and wavelet approach for multiview video coding. The benchmark multiview video coding standard has been studied as well as other related interview structures, namely simulcast, IPP and Amr.

The core of this chapter was dedicated to present and evaluate the proposed PBI multiview coding scheme. PBI scheme was designed to facilitate the random access ability of the MVC encoder. The proposed PBI structure is based on using two base views (S2, S5) per GGOP with positions allowing a direct interview prediction of the remaining views. Depending on the number of views and taking into consideration three possible choices following the second base view (S5), an extended PBI version was proposed, leading to improvement in the random access ability regardless of the views number. A new random access evaluation metric ($G_R$) was proposed to accurately assess any multiview coding structure. The PBI scheme showed distinguishable random access outperformance against the default structure IBP and Amr [83]. The outperformance of our scheme was measured using the standard metric $N_{max}$ and the proposed method $G_R$. Also, the rate-distortion evaluation demonstrated that our proposed scheme offers an efficient compression performance. The Proposed PBI is very suitable for free viewpoint video where smooth and instant navigation within 3D content is highly recommended.

The next chapter will also address the random access ability requirement aiming to achieve better improvements.

# Chapter 4
# Group of Pictures Effects on Proposed Interview Prediction structures

## 4.1 Introduction

Recent video coding standards such as H.264 [22] and H.265 [23], provide extension profiles allowing the exploitation of the inter-view resemblances for a better compression performance. Besides the compression efficiency, low-delay random access ability comes at the top of any video coding standard requirements list [55]. These two requirements are mainly affected by the applied inter-view prediction scheme and the group of pictures (GOP) architecture. An inter-view prediction scheme is built up by a combination of an interview dependency strategy and a temporal prediction structure. The GOP architecture defines the hierarchical level and the dependency nature between the temporal frames.

In this chapter, an investigation of multiview video coding schemes, based on a various group of pictures architectures, is presented.

The considered coding schemes in this chapter are: A new approach namely "PIP" [27], the proposed structure in Chapter 3 "PBI" and the MVC default structure "IBP".

Results of the conducted experiments allow to order the simulated schemes and GOP sizes preferences according to their impact on random access ability and compression performance. The proposed PIP approach achieves significant random accessibility improvements yielding a gain of 53.33 % and 36.36 % compared to MVC and PBI approach, respectively. "PBI" approach structure produces a substantial bit-rate saving compared to the aforementioned structures.

The rest of this chapter is organised as follows. Section 4.2 describes the different GOP architectures used for the experimental investigation. Section 4.3 introduces the new proposed prediction structure PIP. Random access ability and compression performance assessment is detailed.

## 4.2 Group of Pictures Arrangements

A GOP is a fixed pattern which is periodically repeated along the video sequence. The GOP architecture defines the temporal dependencies between the video frames. Three types of frames can be found within the GOP to reduce spatial and temporal redundancies:

- I frame: contains only intracoded macroblocks.
- P frame: has macroblocks which are temporally predicted from a single neighbouring frame.
- B frame: contains macroblocks that are temporally encoded from two adjacent frames.

The hierarchical B structure has been adopted as a default GOP pattern in both H.264 and H.265 due to its compression efficiency. Hierarchical B algorithm uses multiple levels of B frames along the GOP.

Although B frames have more complex encoding process using motion estimation and compensation, they significantly improve the compression efficiency compared to I and P frames. Multiview coding standards such as MVC and MV-HEVC also employ the hierarchical B structure for the temporal prediction.

In this section, we introduce four GOP patterns of the hierarchical B scheme. These patterns have different GOP lengths and multiple hierarchical B levels.

Figure 4.1 illustrates the hierarchical B scheme when GOP length is equal to four. The structure has two hierarchical levels with a total of three B frames per GOP. As illustrated in Figure 4.2, an additional hierarchy level is introduced for GOP size = 8. There are four B3 frames in the scheme which inherently slows down the random access ability compared to the GOPs=4 scheme.

A total of eleven B frames appears in Figure 4.3, where eight B frames of the third level are included. Therefore, the compression efficiency would be much better compared to schemes of Figure 4.1 and Figure 4.2.

Figure 4.4 shows the hierarchical B pattern for GOP length = 16. A fourth hierarchy level is introduced in this scheme with eight B4 frames.

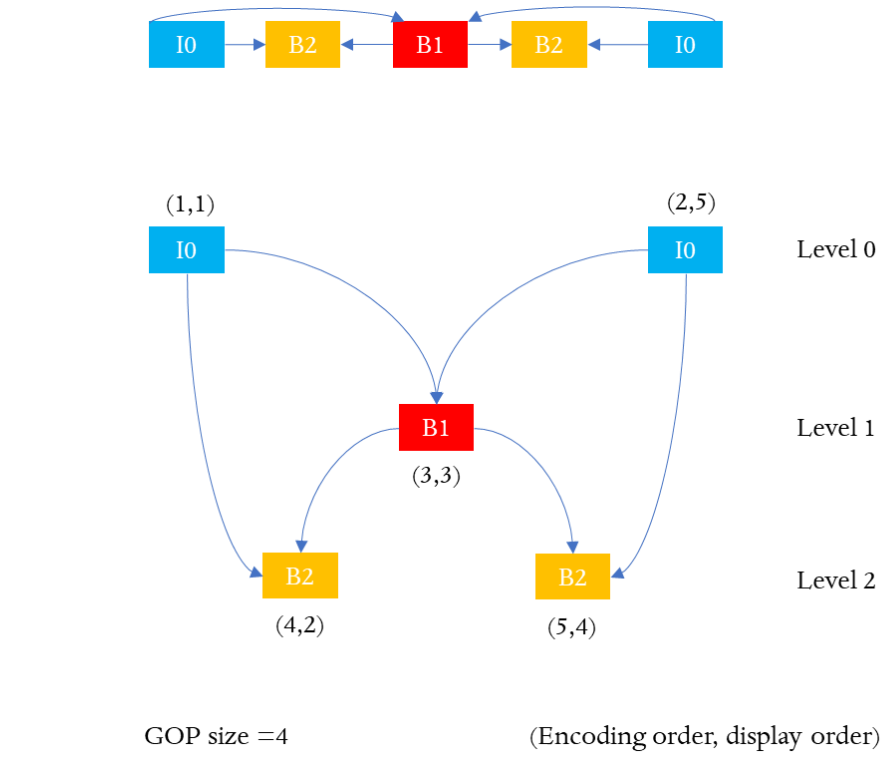



Figure 4.1 Hierarchical B pattern for GOP size = 4

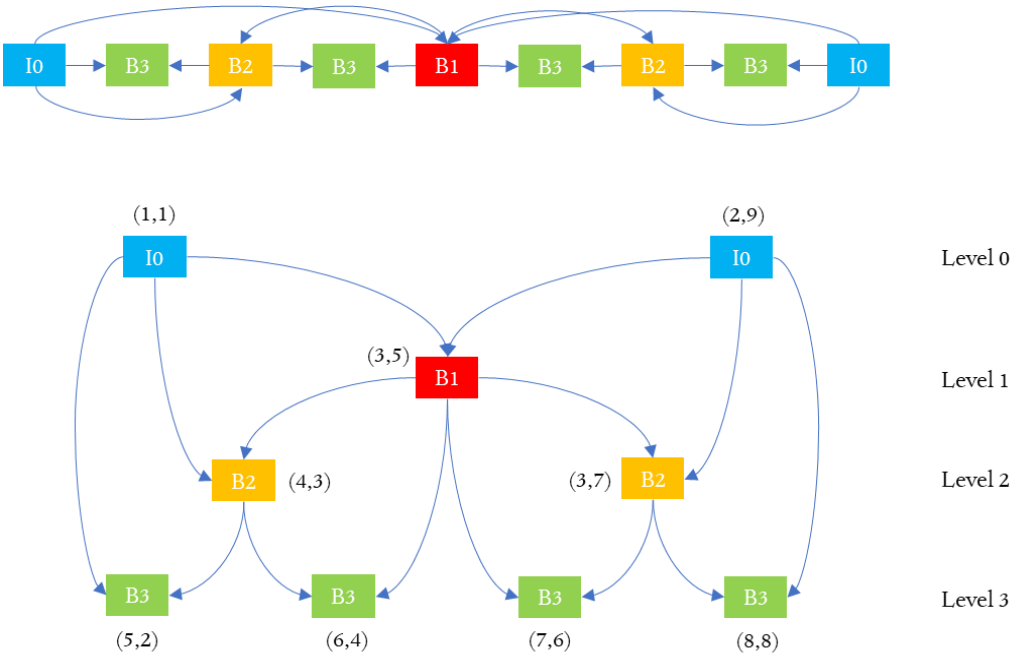



Figure 4.2 Hierarchical B scheme for GOP size = 8

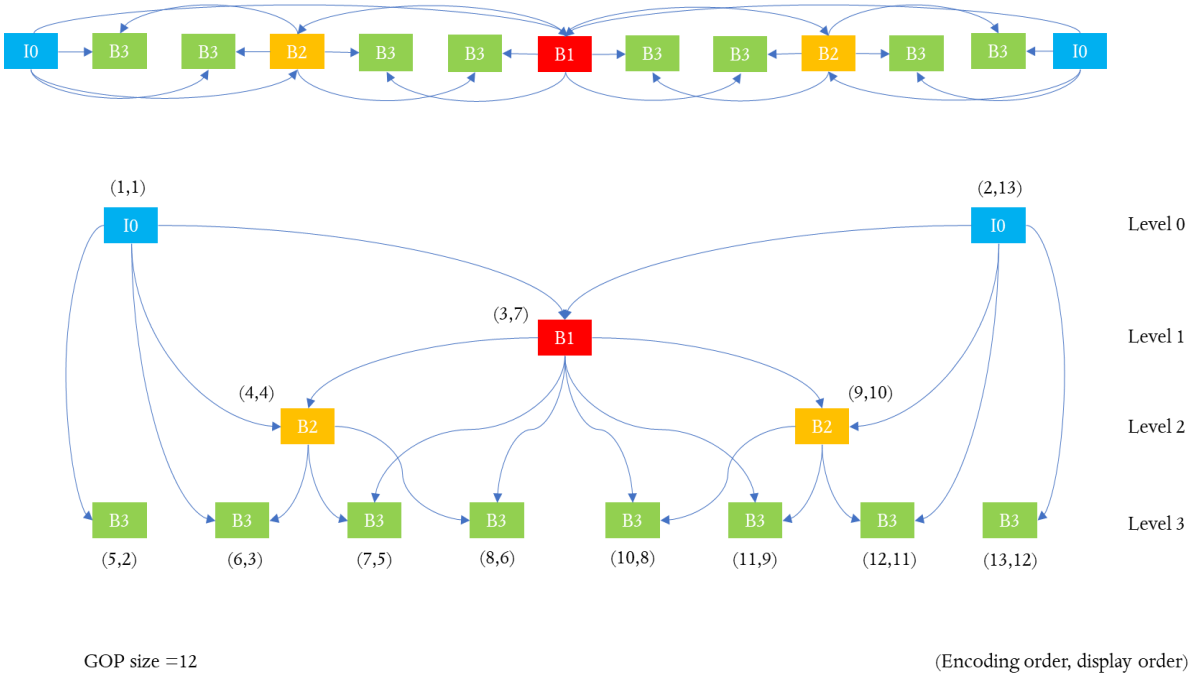



Figure 4.3 Hierarchical B scheme for GOP size = 12

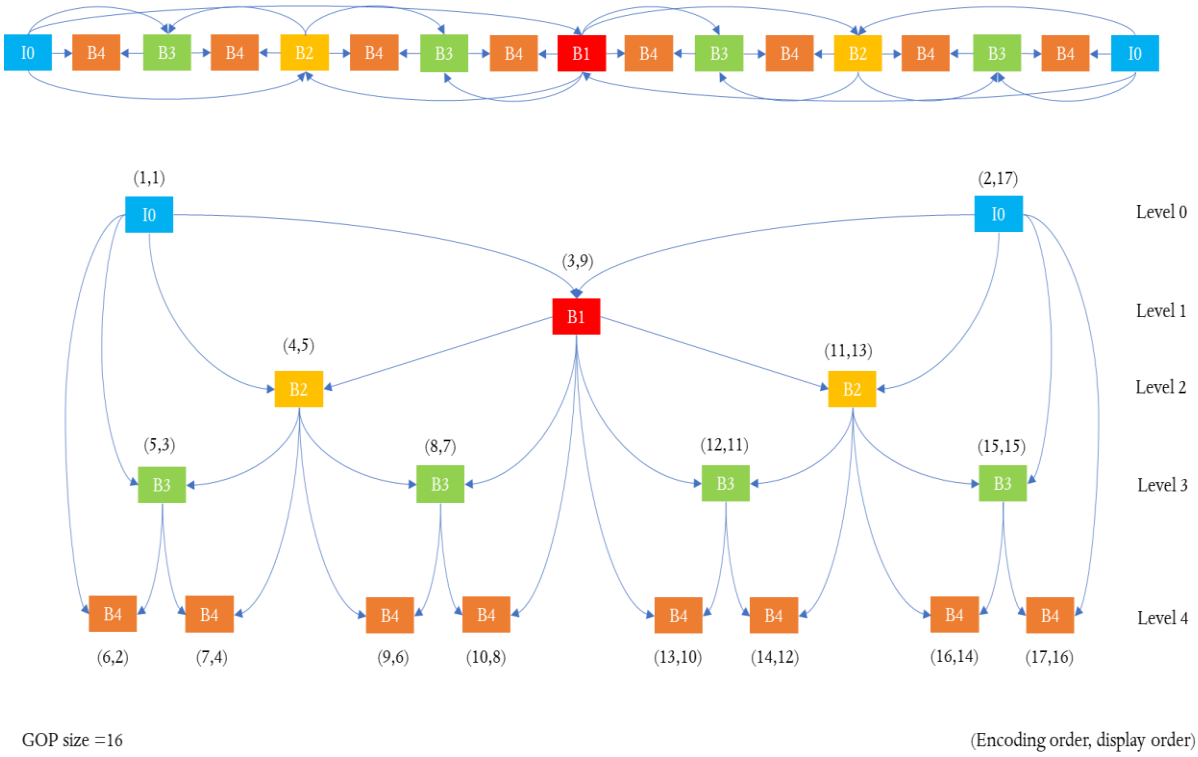



Figure 4.4 Hierarchical B scheme for GOP size = 16

To calculate the required frames number to encode a B frame within the temporal hierarchical structure regardless of its level, we use the following equation:

$$N_{img} = \text{Hierarchy level} + 1 \quad (4.1)$$

The four hierarchical patterns presented in this section will be employed within the three considered interview prediction structures to investigate their effects on the random access ability and compression performance.

## 4.3 The Proposed PIP Structure

The second proposed PIP multiview coding scheme is presented in this section. Figure 4.5 illustrates the PIP structure for a default design of eight views and GOP size of eight frames. The PIP proposed interview scheme is composed of two base views (I) and six predicted views (P) per group of groups of pictures (GGOP). The two reference views I-views (S2, S5) are independently coded from the other views. S2 and S5 are selected as optimal positions for the base views. These two positions allow a direct interview prediction for all the remaining P-views without any intermediate view.

The six P views are coded using jointly temporal and interview prediction techniques. The proposed design consists of two GGOPS which are completely independent of each other. Each sub-GGOP is constructed around an I-view. The first sub-GGOP is composed of S2 as a base view and S0, S1 and S3 as enhancement layers. The P-views' anchor frames are encoded through the I-view' anchor frame. Whereas the P-views' nonanchor frames are predicted from three reference frames, one from the interview level of the base view and two from their temporal level. The proposed PIP structure excludes using B-views that use at least two frames for encoding their anchor B-frames and requires a minimum of four frames to encode the nonanchor frames.

Using a structure of two base views and P-views, instead of B-views, significantly improves the random access ability. In other words, this combination reduces the number of the needed frames for coding or decoding any frame within the multiview video coding structure. This also leads to reducing the encoding time duration.

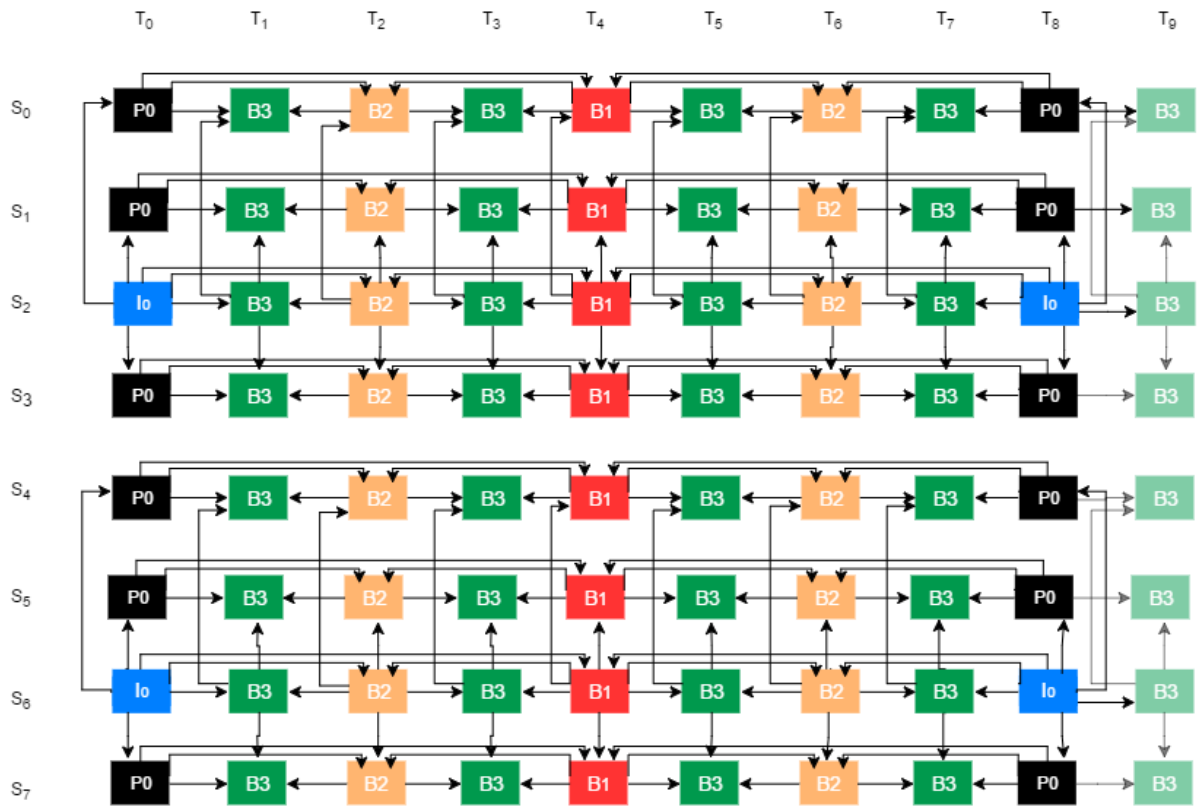



Figure 4.5 The PIP multiview prediction scheme [27]

The maximum hierarchical level of the PIP default design is equal to three, depicted in Figure 4.5 by the green colour (B3). This particular frame B3 can be found in both I and P-views. In order to calculate $N_{max}$ of the proposed PIP structure, we extract the example of B3 frame located in S0/T1 (Figure 4.5). For accessing this frame, four reference frames are required in the temporal level with these positions: S0/T0, S0/T2, S0/T4 and S0/T8, and five reference frames in the interview level with the following positions: S2/T0, S2/T1, S2/T2, S2/T4 and S2/T8. Consequently, $N_{max}$ equation of the proposed PIP scheme is deduced as follows:

$$N_{max} = 2 \times H_{max} + 1 \tag{4.2}$$

$H_{max}$ is equal to 4 for GOP size = 8.

Figure 4.6 illustrates a comparison between the anchor frames combinations of the IBP, PBI and PIP schemes.

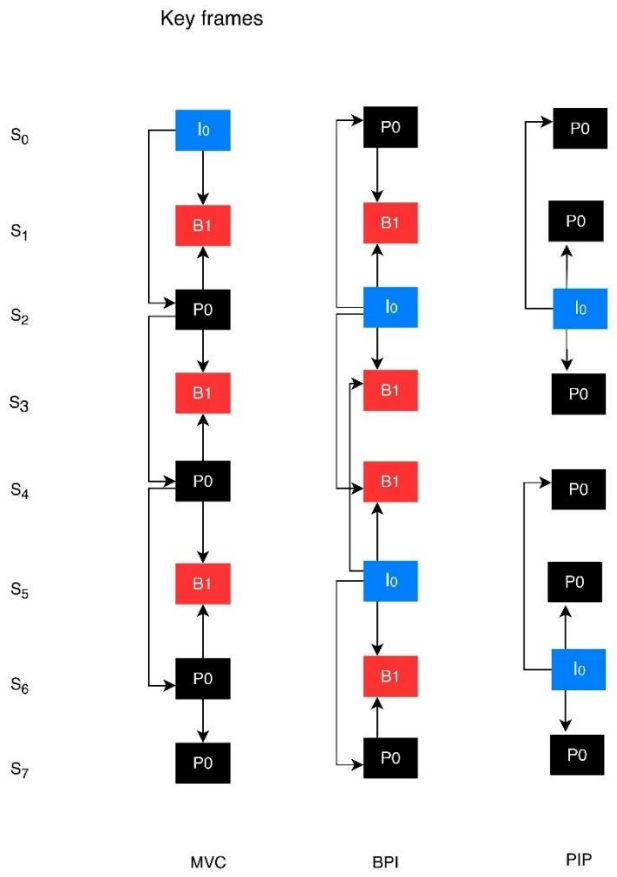



Figure 4.6 interview prediction schemes comparison between MVC, PBI and PIP structures [27]

Table 4.1 regroups the $N_{max}$ equations of MVC, PBI and PIP schemes. Table 4.1 clearly shows that the PIP structure provides the simplest equation compared to the reported structures.

Table 4.1 $N_{max}$ equations comparison [27]

| | $N_{max}$ equations |
|---|---|
| **MVC** | $N_{max} = 3 \times H_{max} + 2 + 5 \times [Nbr_{view} - 1]$ |
| **PBI** | $N_{max} = 3 \times H_{max} + 2$ |
| **PIP** | $N_{max} = 2 \times H_{max} + 1$ |

The following equations calculate the number of the crossed required frames for accessing any selected picture in the PIP scheme

- For anchor frames:

$$Nbr_{view} = \{0 \text{ for I frames, } 1 \text{ For P frames}\} \quad (4.3)$$

- For nonanchor frames:

$$Nbr_{view} = \alpha \times H_{level} + \beta \quad (4.4)$$

$$Where \begin{cases} \alpha = 1, \beta = 0 \; ForInonkeyframes \\ \alpha = 2, \beta = 1 \; ForInonkeyframes \end{cases}$$

$H_{level}$ takes the value of the hierarchical level of the frame. Equations (4.3) and (4.4) reveal the calculation simplicity for computing the random access of the proposed approach PIP.

## 4.4 Evaluation of The Proposed PIP Structure

### 4.4.1 Random access ability evaluation

In this section, we evaluate the random access ability of the proposed PIP scheme with respect to MVC and PBI schemes. Furthermore, we investigate the effects of using different GOP patterns over the random access performance.

Two metrics are employed for the assessment process. $N_{max}$ and the global random access cost which have been previously detailed in section 3.6.1 and 3.6.2, respectively.

The global random access evaluation allows a full assessment of the considered multiview scheme. It takes into account the random access cost of each existing frame in the GGOP design. $G_R$ calculates the average cost of predicting all frames in the GGOP of the multiview structure. Although the global random access assessment consists of three phases, we have only considered, in this section, the last phase $G_R$ which jointly evaluates the anchor and the nonanchor frames of the examined structure.

According to the presented GOP patterns in Section 4.2 , the GOP (size) takes four values in our tests: 4,8,12 and 16.

Table 4.2 regroups the $G_R$ obtained results of IBP, PBI and PIP structures following the GOP size values. From the collected data in this table, we can see that $G_R$ (PIP) resulted in the lowest values for all GOP sizes. These results indicate the random access outperformance of the proposed PIP structure compared to IBP and PBI.

Initially, when the GOP size is equal to 4, the $G_R$ attains the lowest values for all structures, where $G_R$ (IBP) = 7.84, $G_R$ (PBI) =5.57 and $G_R$ (PIP) = 4.45. However, the highest $G_R$ values are obtained when the GOP size is equal the 16. Moreover, the $G_R$ values are ascended generally with the increase of the GOP length from 4 to 16. This is due to the temporal reference scheme which varies depending on the GOP lengths.

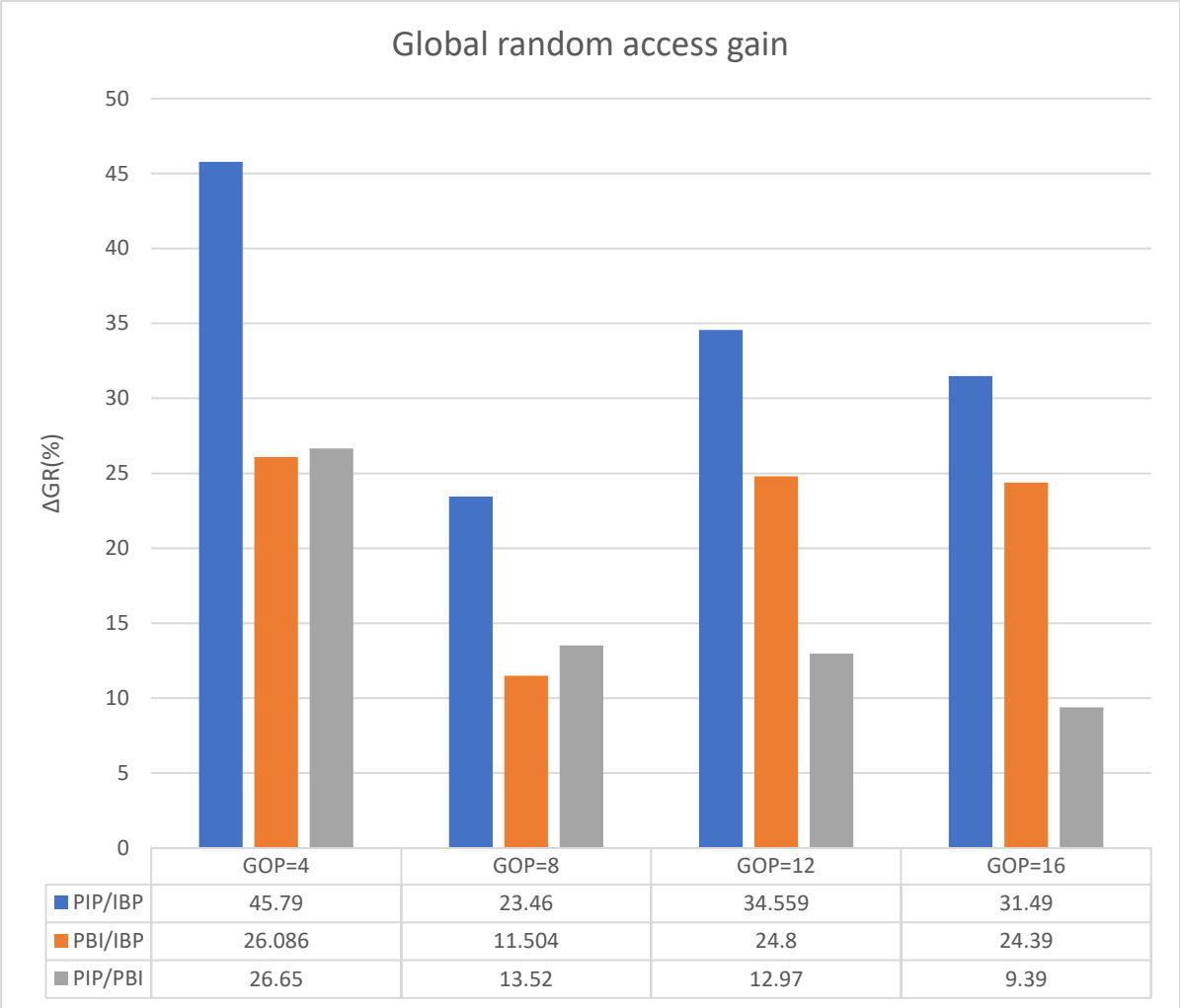
Table 4.2 $G_R$ Results of the considered prediction structures [27]

| GOP | $G_R$ (IBP) | $G_R$ (PBI) | $G_R$ (PIP) |
|---|---|---|---|
| 4 | 7.84 | 5.75 | 4.25 |
| 8 | 7.84 | 6.78 | 6 |
| 12 | 10.055 | 8.75 | 6.58 |
| 16 | 11.3125 | 10.25 | 7.75 |

The following formula is used to calculate the random access gain of our proposed structure:

$$\Delta G_R = \frac{G_R(\text{structure 1}) - G_R(\text{structure 2})}{G_R(\text{strcuture 1})} \times 100\% \qquad (4.15)$$

For instance, for calculating $\Delta G_R$ (PIP/IBP), $G_R$ (structure 1) takes the value of $G_R$ (IBP) and $G_R$ (structure 2) takes the value of $G_R$ (PIP).




| | GOP=4 | GOP=8 | GOP=12 | GOP=16 |
|---|---|---|---|---|
| PIP/IBP | 45.79 | 23.46 | 34.559 | 31.49 |
| PBI/IBP | 26.086 | 11.504 | 24.8 | 24.39 |
| PIP/PBI | 26.65 | 13.52 | 12.97 | 9.39 |

Figure 4.7 $\Delta G_R$ (%) comparison through different GOP sizes [27]

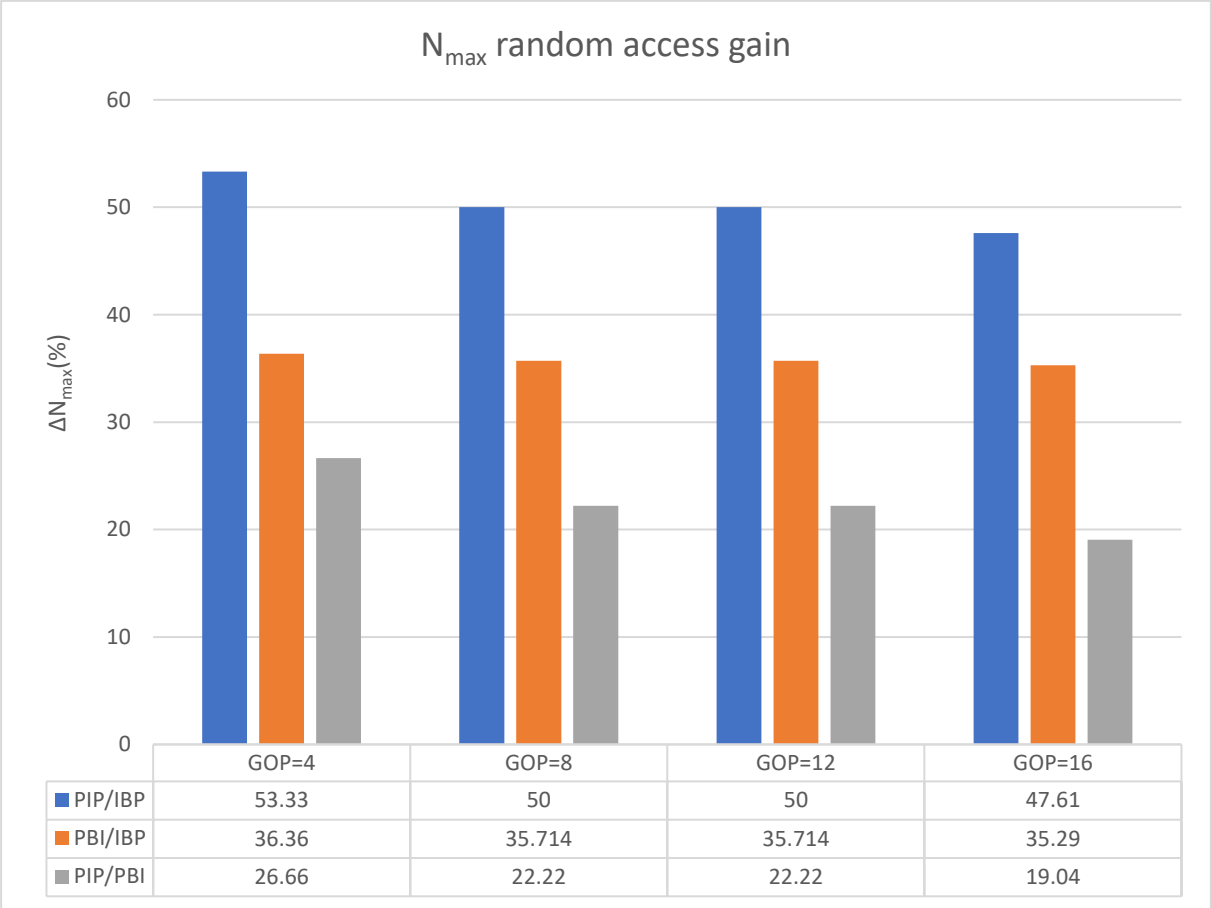
Figure 4.7 depicts the random access gain comparison between IBP, PBI and PIP structures, over the four selected GOP values. It can be clearly inferred that the "PIP" scheme is more effective, with an average gain of ~ 34 % and ~ 16 % relative to MVC and PBI, receptively. The largest gain of the proposed PIP structure is noted when GOP=4. It exceeds 45 % and 26 % compared to IBP and PBI, respectively.

Table 4.3 regroups the $N_{max}$ results of the examined schemes. The obtained values demonstrate that the PIP structure significantly reduces the maximum number of the reference frames required for accessing a given frame.

Improvements in random access ability achieved by the PIP structure, are also highlighted in Figure 4.8. The maximum Nmax gain of PIP exceeds 53 % and 26 % relative to IBP and PBI structures, respectively. Both $G_R$ and $N_{max}$ results demonstrate that the PIP multiview structure considerably reduces the complexity for accessing a given frame in the multiview video coding scheme, which leads in turn to an enhanced random access ability of the multiview video coding.

Table 4.3 $N_{max}$ Results of the considered prediction structures over different GOPs [27]

| GOP | $N_{max}$(IBP) | $N_{max}$ (PBI) | $N_{max}$(PIP) |
|---|---|---|---|
| 4 | 15 | 11 | 7 |
| 8 | 18 | 14 | 9 |
| 12 | 18 | 14 | 9 |
| 16 | 21 | 17 | 11 |




| | GOP=4 | GOP=8 | GOP=12 | GOP=16 |
|---|---|---|---|---|
| PIP/IBP | 53.33 | 50 | 50 | 47.61 |
| PBI/IBP | 36.36 | 35.714 | 35.714 | 35.29 |
| PIP/PBI | 26.66 | 22.22 | 22.22 | 19.04 |

Figure 4.8 $\Delta N_{max}$ (%) comparison through different GOP sizes [27]

### 4.4.2 Compression Efficiency Evaluation

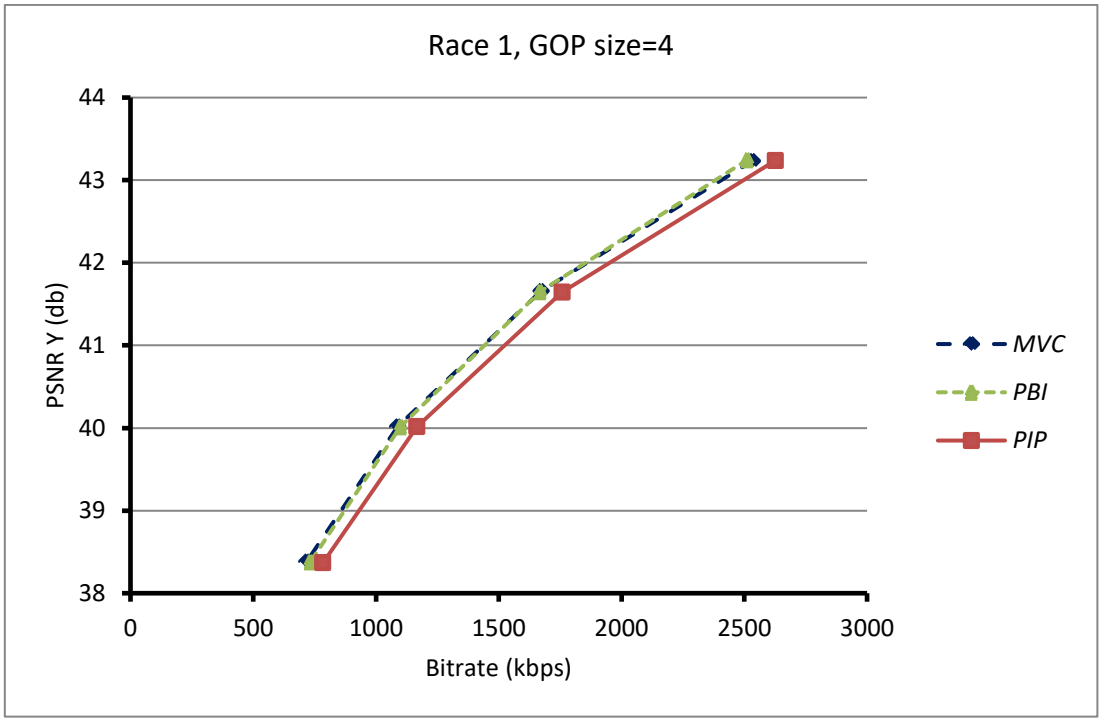
Compression performance results of the PIP prediction structure are presented and discussed in this section. The compression efficiency of the proposed PIP scheme is objectively evaluated against MVC(IBP) and PBI schemes, using graphs of PSNR(dB) versus bitrate (kbps). Common initial conditions and data resources are used to ensure fair comparisons. Table 4.4 regroups the used multiview video sequences and the common encoding configuration.

Table 4.4 Data materials and encoding configuration

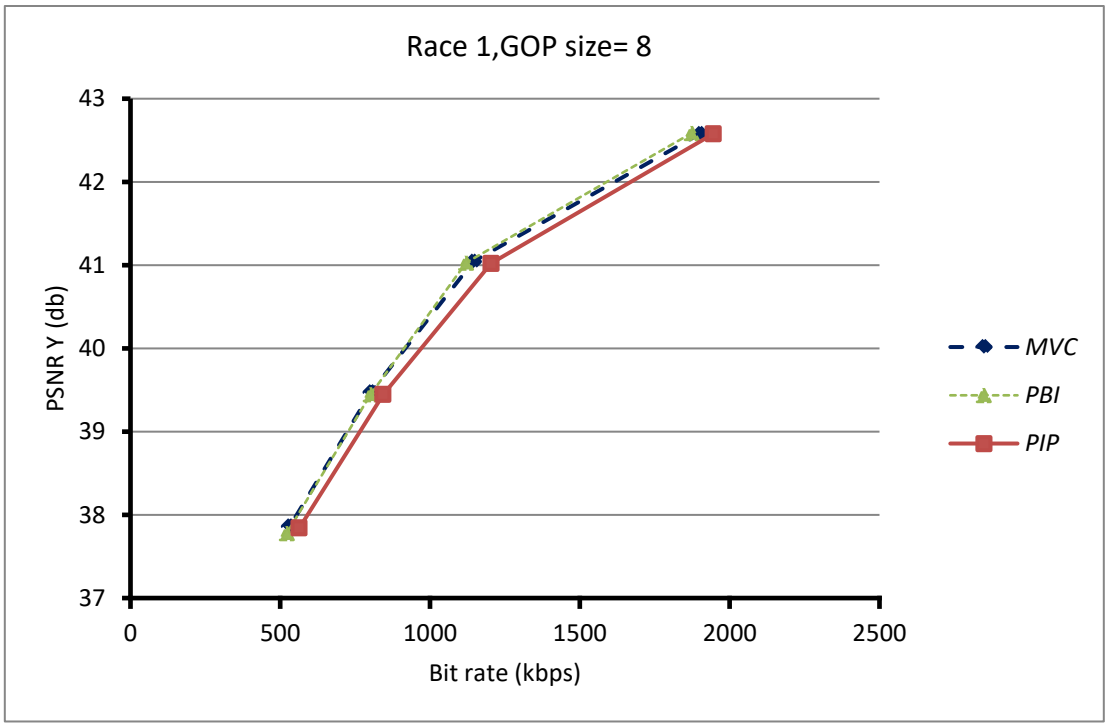
| Parameter | Setting |
|---|---|
| Video sequences | Race1, Exit and Ballroom |
| GOP size | 4, 8, 12 and 16 |
| Quantisation parameter | 20, 23, 26 and 29 |
| Symbol mode | CABAC |
| Search mode | Fast search |
| Search range | 64 |

Experimental investigations are conducted using three MVV sequences as reported in Table 4.4. Each MVV sequence is composed of eight parallel views. Similar to the random access assessment, the same GOP sizes are employed to evaluate the compression efficiency. Four QP values are used to control the rate-distortion variations; the lower the value of the QP, the higher the bitrate and video quality. The four QP values are used for the three considered multiview coding structures. In addition, four GOP lengths are employed for every QP value, which results in 48 different simulations for each MVV sequences. In other words, a total of 144 simulations were carried out over the investigation process.

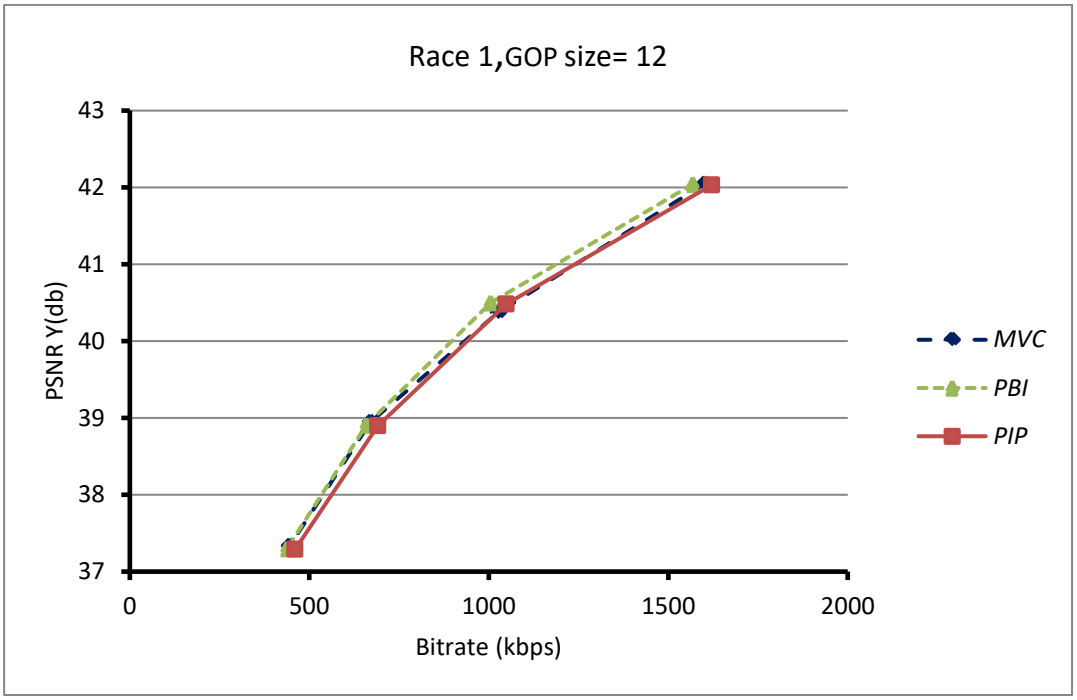
Obtained results are illustrated in Figure 4.9, Figure 4.10 and Figure 4.11. Each figure shows comparison between MVC (IBP), PBI and PIP structures using the same MVV sequence through different GOP sizes. Each figure is composed of graphs (a), (b), (c) and (d). The graph shows a comparison between MVC, PBI and PIP using the same MVV sequence and the same GOP size. For instance, the graph 4.9 (a) shows the rate-distortion curves of MVC, PBI and PIP when encoding Race1 sequence and employing a GOP size equal to four.


Race 1, GOP size=4
PSNR Y (db)
44
43
42
41
40
39
38
0
500
1000
1500
2000
2500
3000
Bitrate (kbps)
MVC
PBI
PIP

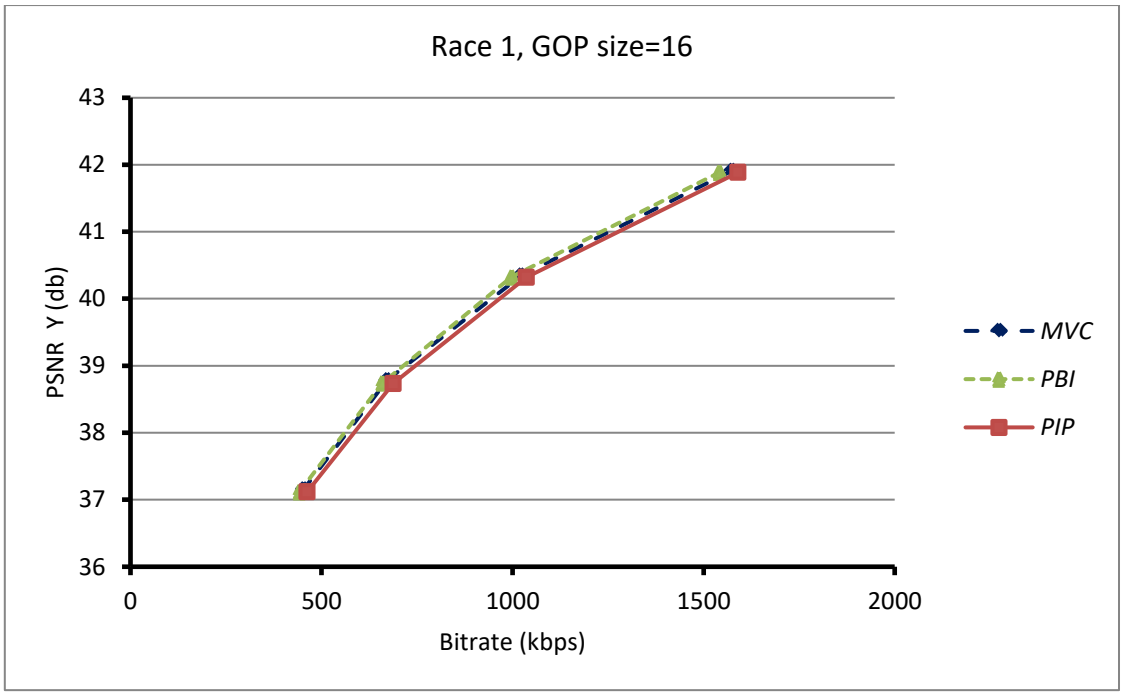
(a)


Race 1,GOP size= 8
PSNR Y (db)
43
42
41
40
39
38
37
0
500
1000
1500
2000
2500
Bit rate (kbps)
MVC
PBI
PIP

(b)

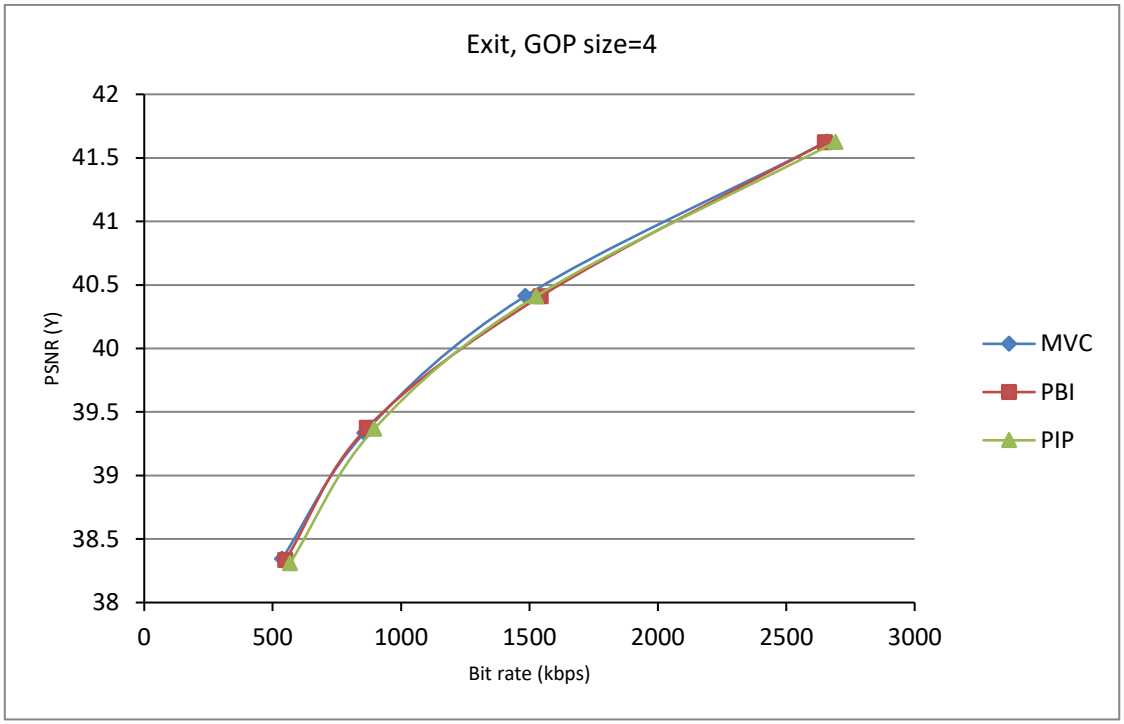



(c)

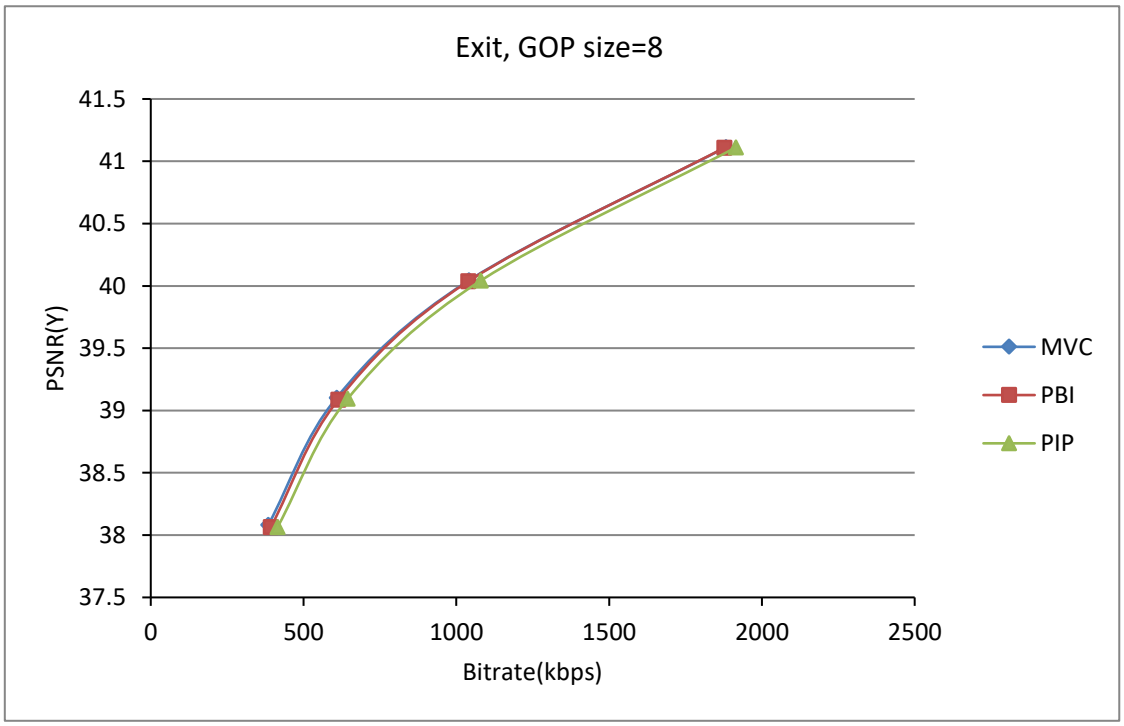



(d)

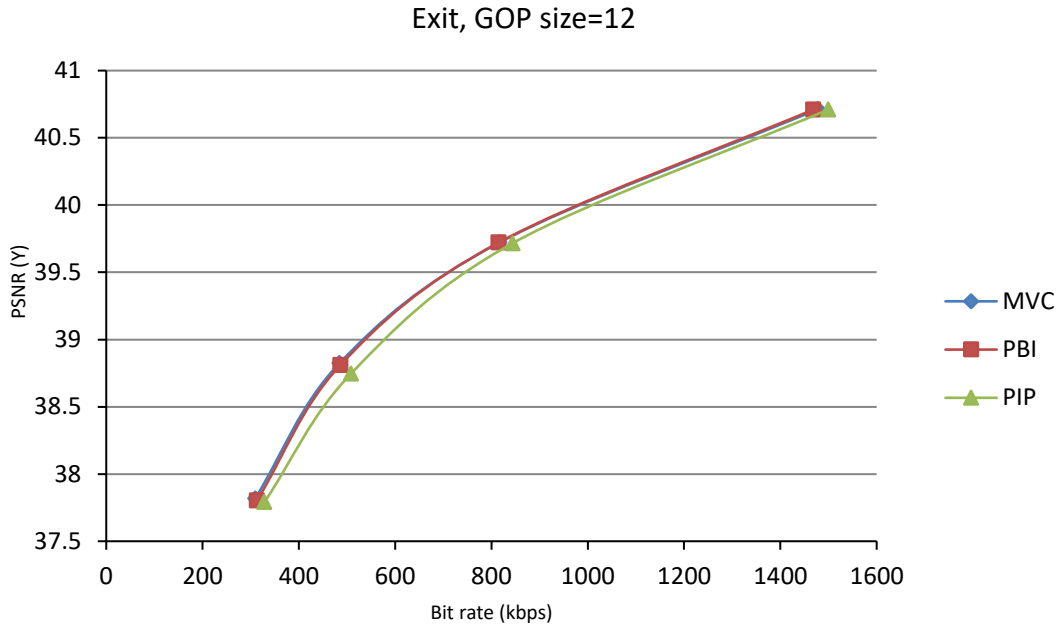
Figure 4.9 Rate-Distortion (RD) performance of MVC, PBI and PIP using Race1 sequence over four GOPs


Exit, GOP size=4
PSNR (Y)
42
41.5
41
40.5
40
39.5
39
38.5
38
0
500
1000
1500
2000
2500
3000
Bit rate (kbps)
MVC
PBI
PIP

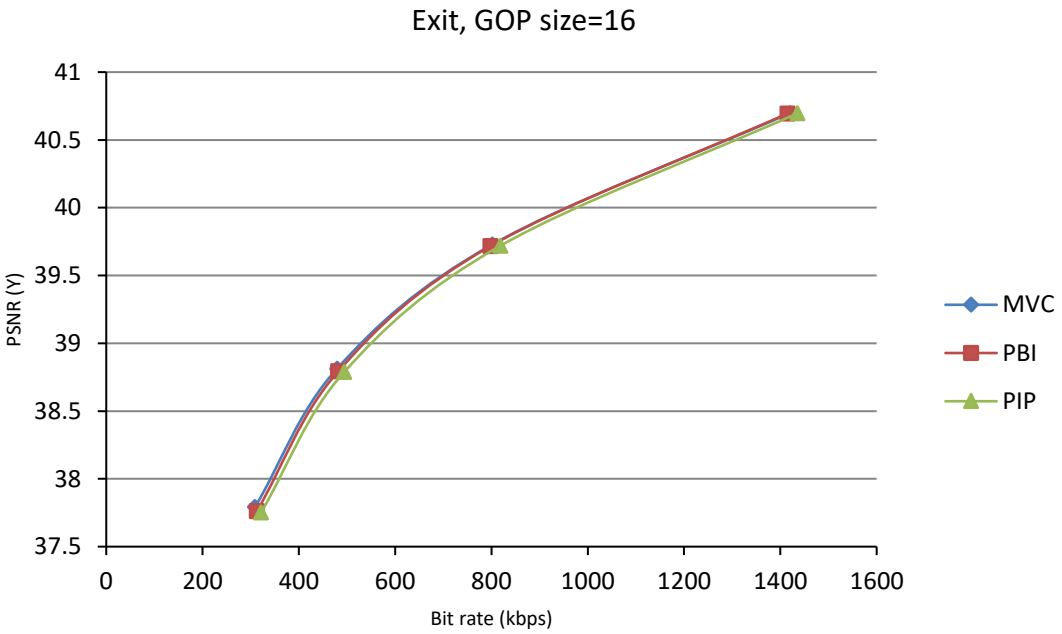
(a)


Exit, GOP size=8
PSNR(Y)
41.5
41
40.5
40
39.5
39
38.5
38
37.5
0
500
1000
1500
2000
2500
Bitrate(kbps)
MVC
PBI
PIP

(b)

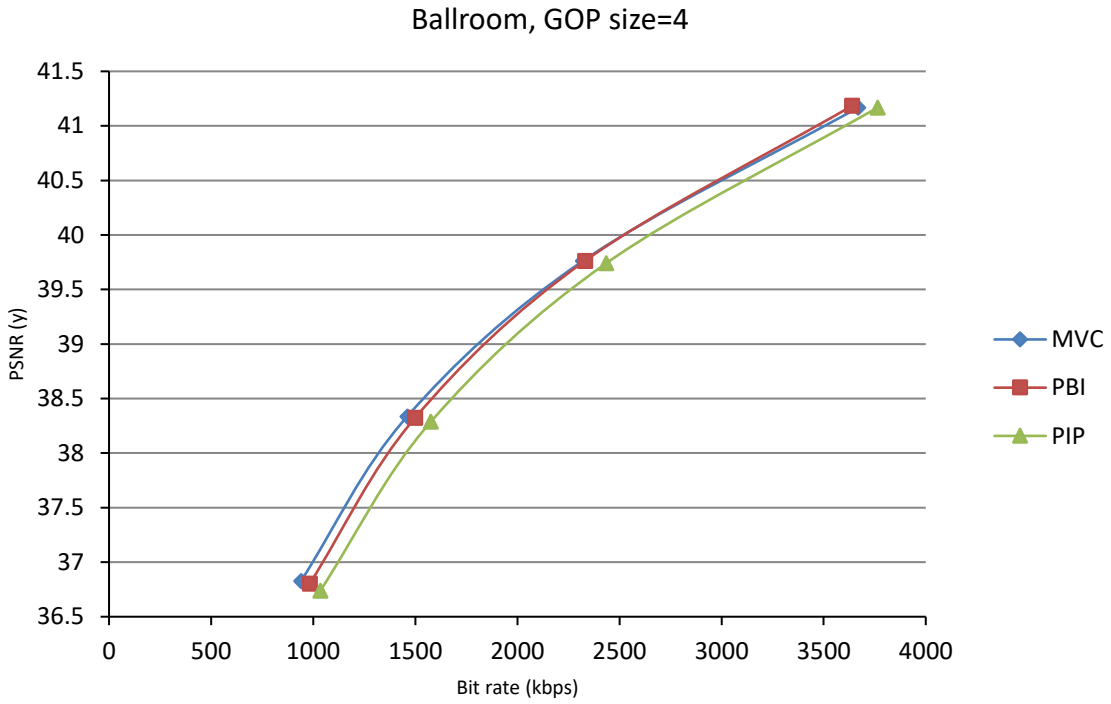



(c)

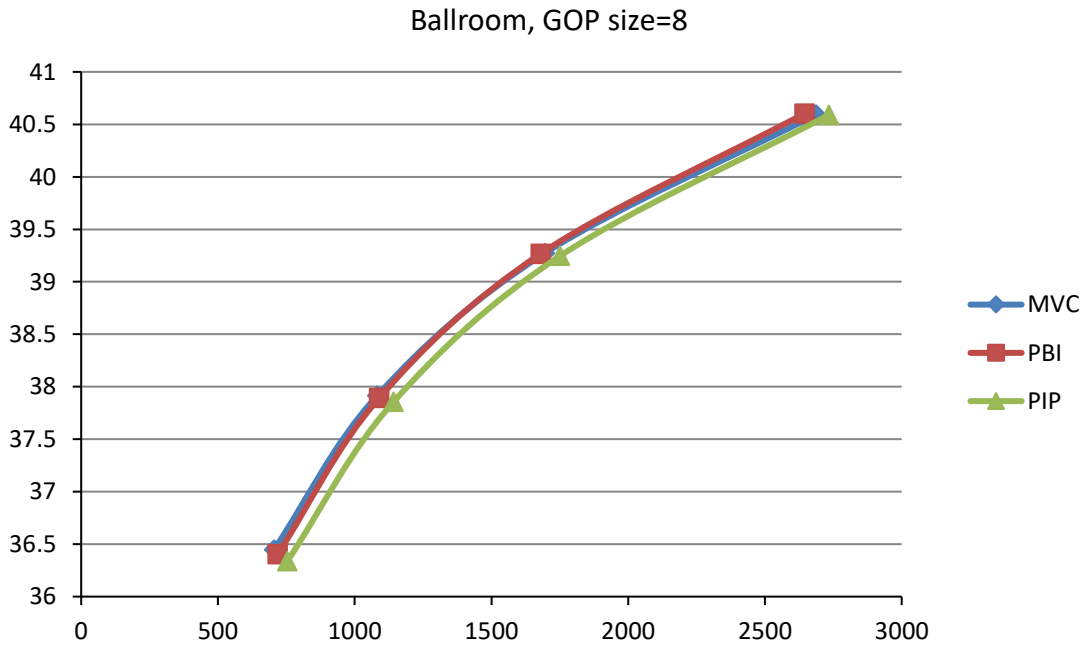



(d)

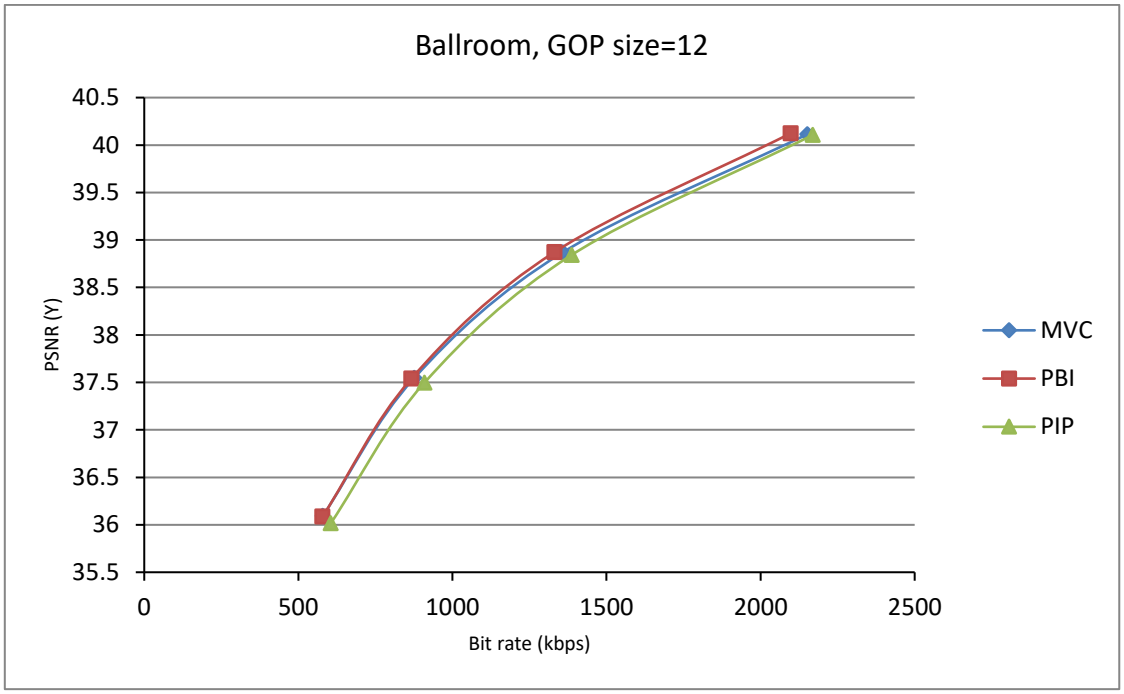
Figure 4.10 RD performance of MVC, PBI and PIP using Exit sequence over four GOPs


Ballroom, GOP size=4
PSNR (y)
41.5
41
40.5
40
39.5
39
38.5
38
37.5
37
36.5
0
500
1000
1500
2000
2500
3000
3500
4000
Bit rate (kbps)
MVC
PBI
PIP

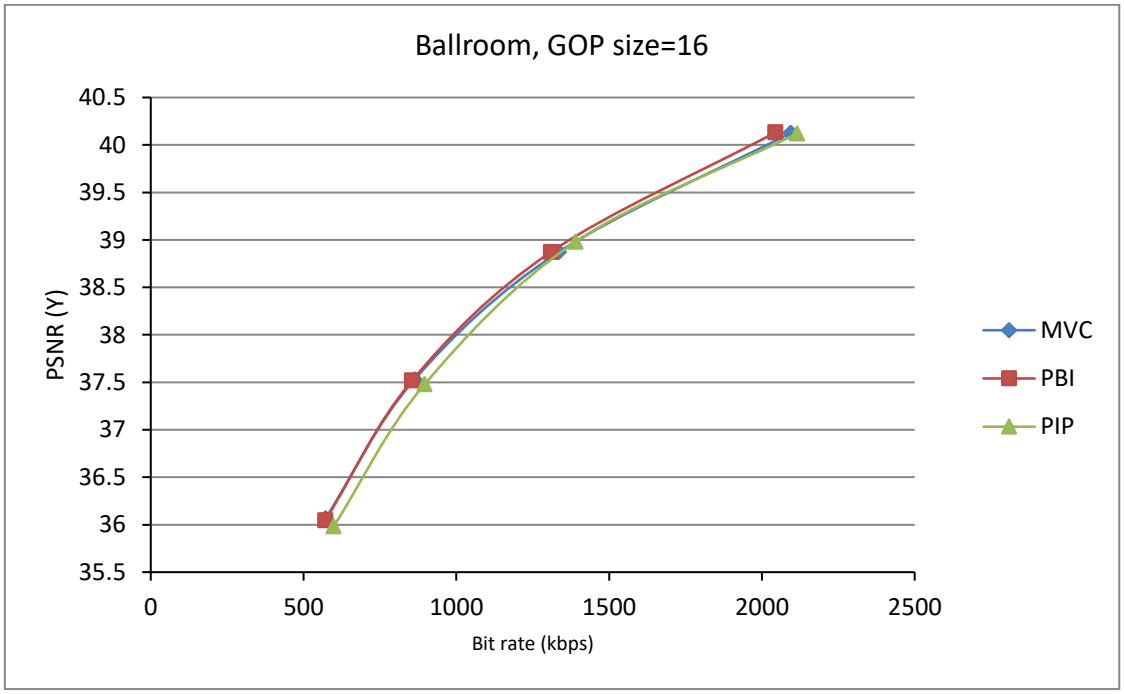
(a)


Ballroom, GOP size=8
41
40.5
40
39.5
39
38.5
38
37.5
37
36.5
36
0
500
1000
1500
2000
2500
3000
MVC
PBI
PIP

(b)




(c)




(d)

Figure 4.11 RD performance of MVC, PBI and PIP using Ballroom sequence over four GOPs

It is clearly inferred from these figures that the employment of a smaller GOP size, regardless of the considered multiview structure, always results in a better video quality compared to a longer GOP. The video quality measured in PSNR degrades each time the GOP size increases. However, a reduced GOP size generates larger bitrate values. A comparison between the four graphs of each figure ascertains this point. The video quality gradually decreases with the increase of the GOP size for the three considered structures regardless of the encoded multiview video. In addition, the resulting bitrate values are distinctly decreased for larger GOP sizes.

A shorter GOP size reduces the distance between the temporal reference frames, which results in higher levels of correlation between the frames of the same GOP. Consequently, both video quality and bitrate values are increased.

Figure 4.9, Figure 4.10 and Figure 4.11 reveal in overall that PBI scheme delivers better compression performance compared to MVC and PIP. The proposed PIP scheme provides a less efficient compression in terms of bitrate compared to MVC scheme. Nevertheless, this difference decreases by increasing the GOP size.

The structure composition is one of the main sources that generate the bitrate gain or loss. Including more I-views and P-views will obviously produce additional data during the compression process. This point has been clearly noticed through the PIP structure, which uses two I-views and six P-views. Additionally, the GOP length has a direct effect on the video compression, where the similarity between the successive frames is exploited by the temporal prediction, which is based on the reference frames that define the start and the end of a GOP. Hence, the similarity considered for exploitation increases as the GOP size is extended.

Therefore, further data is available to be removed, which leads to improvements in the compression efficiency. However, the resemblance between the frames will gradually decrease with the time progress. Therefore, it will not be useful to increase the GOP length without any specific limitation. This fact has been confirmed from the results when GOP= [12, 16] where practically similar graph lines are clearly distinguished. See graphs (c) and (d) of Figure 4.9, Figure 4.10 and Figure 4.11.

### 4.4.3 Prediction structures Trade-offs

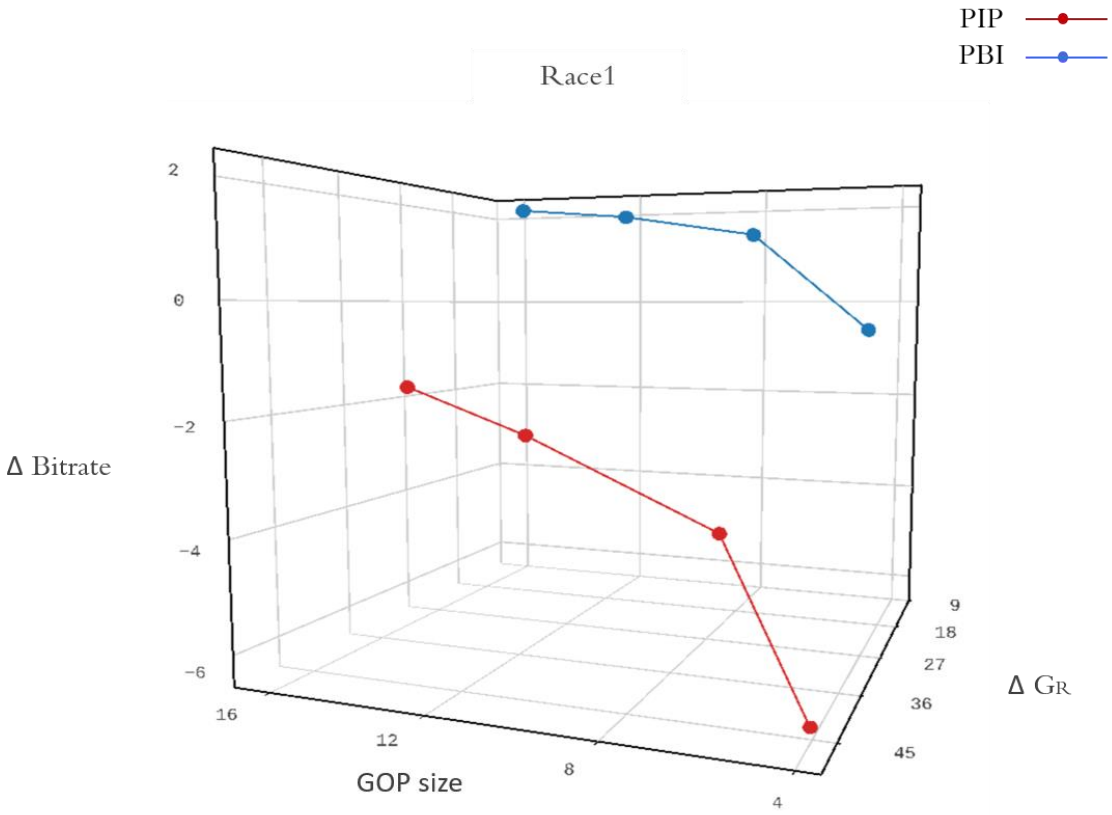
In this section, 3D graphs are presented to highlight the PBI and PIP structures trade-offs between the GOP size, the bitrate gain and the random access ability. Each 3D graph provides a holistic overview of the experimental outcomes, by illustrating the previous results in one graph composed of 3 axes. Where the (x) axis shows the four used GOP sizes; the (y) axis represents the average results of the random access ability relative to the MVC structures. The $\Delta$ bitrate results, which represent here the average value of the four used quantization parameters (QP=20, 23, 26 and 29), are projected along the vertical axis (z).

The $\Delta$ bitrate results are calculated by the following formula:

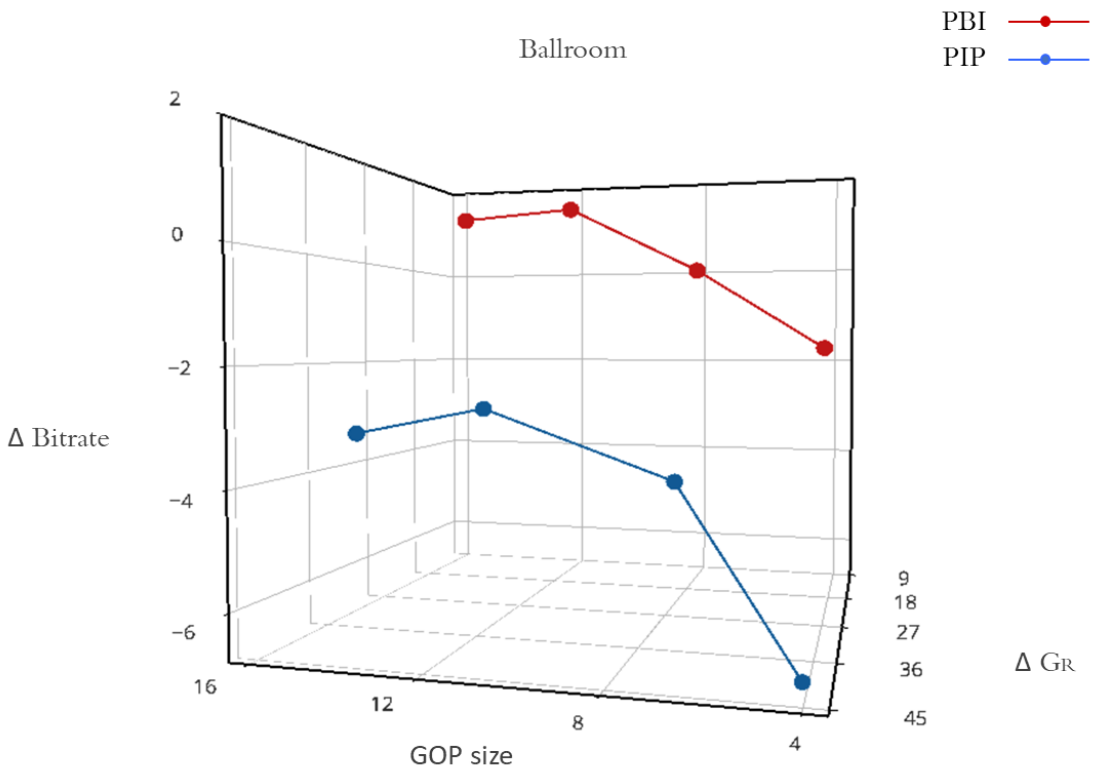
$$\Delta_{BItrate} = \frac{bitrate_{MVC} - bitrate_{proposed}}{bitrate_{MVC}} \text{x}100\% \qquad (4.16)$$

Bitrate (proposed) can take the value of PBI or PIP. The positive results of $\Delta$ bitrate will be taken as a gain, whereas the negative ones will be considered as a loss.

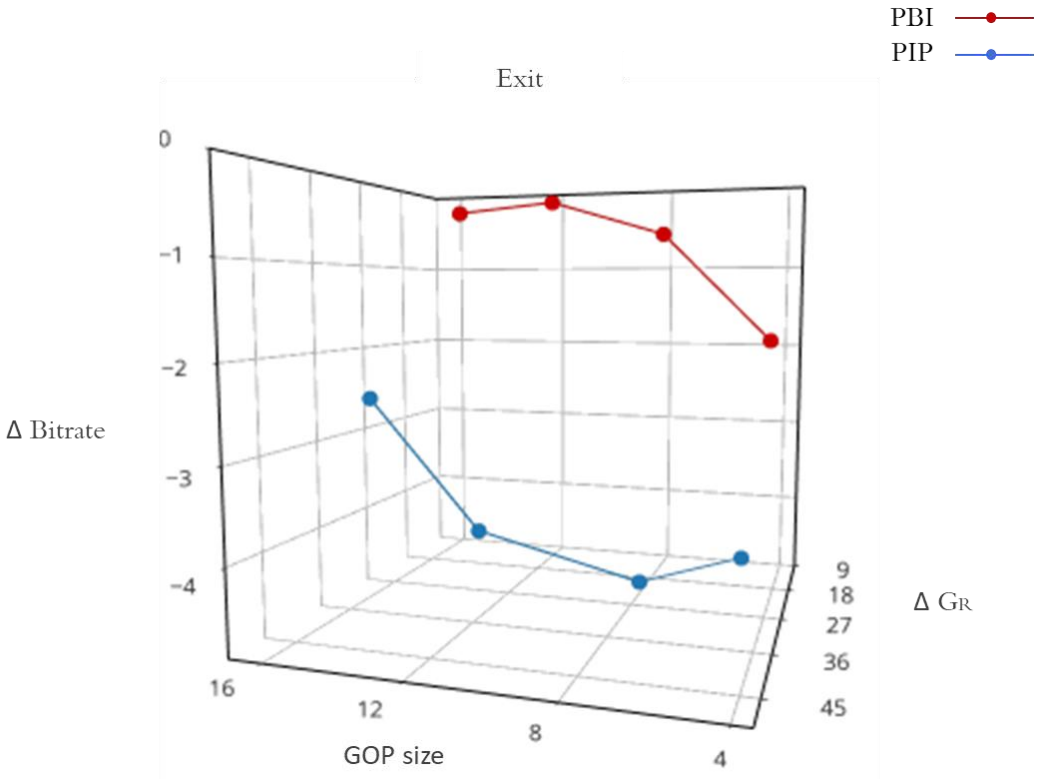
The 3D charts of Figure 4.12 show the evolutionary effect of changing the GOP size on both the random access ability and the compression efficiency in terms of bitrate. It can be easily deduced from the 3D charts that the PBI structure performs better than the PIP structure in term of bitrate saving. Conversely, the PIP outperforms PBI in its random access ability, where the optimum random access efficiency for PIP is achieved when GOP size is equal to 4. Hence, PIP is suitable for the multiview video and Free viewpoint video applications where the bitrate saving is less important and smoother interactivity is more required. The PBI can be considered as a balanced structure for standard applications as it provides improvements in both compression efficiency and random access ability relative to the benchmark coder MVC.


PIP
PBI
Race1
Δ Bitrate
2
0
−2
−4
−6
16
12
8
4
GOP size
9
18
27
36
45
Δ GR

(a)


PBI
PIP
Ballroom
Δ Bitrate
2
0
−2
−4
−6
16
12
8
4
GOP size
9
18
27
36
45
Δ GR

(b)




(c)

Figure 4.12 trade-offs of PBI and PIP structures

## 4.5 Conclusion

In this chapter, a new proposed PIP interview prediction structure was presented. The PIP scheme improved further the random access ability of the multiview video coder. PIP design is composed of two reference views and a remaining set of P-views to allow faster and direct interview prediction. Four different GOPs were applied to compare and investigate their effects on the random access performance and compression efficiency of the considered schemes. Evaluation results were divided into three parts. Firstly, the random access ability was evaluated using two metrics, namely $N_{max}$ and $G_R$. The PIP coding scheme achieved significant $G_R$ gains that exceed 45 % and 26 % compared to MVC and PBI structures, respectively.

Secondly, the compression efficiency was assessed in terms of bitrate saving and video quality using four GOPs, where for almost all cases the PBI approach achieved the best performance. The MVC structure comes in the second place, while the PIP approach showed a remarkable loss in bitrate saving. Finally, 3D trade-off charts were presented, showing the GOP size effects on the compression efficiency and random access ability for PBI and PIP coding structures relative to MVC. The results have ascertained the fact that using a reduced GOP length provides better random access ability and less bitrate saving. Conversely, larger GOP length leads to a slow random access and more bitrate saving.

# Chapter 5
# Multiview extension of HEVC/H.265

## 5.1 Introduction

Nowadays, the demand for both higher-resolution videos and 3D visual content is known an immense increase. It was already predicted that video traffic will occupy 82 percent of all the diffused data over the internet by 2021[2], and the 3D video with its different representations will be definitely a crucial part of this traffic. The multiview video, which has been the study case in our thesis, generates a considerable amount of data volume that needs progressive coding techniques to respond to the increased quality demands. Recent video coding standards such as H.264 [22] and H.265 [23], provide extended profiles that take advantage of the interview resemblances for better compression efficiency.

The first edition of the High Efficiency Video Coding standard (H.265) [88] was finalised and published in 2013 by the Joint Collaborative Team on Video Coding (JCT-VC). The H.265 standard can achieve 50% of bit-rate saving for equal perceptual video quality compared to H.264.

Back in July 2012, the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) was established by the ISO/IEC MPEG and ITU-T Video Coding Experts Group (VCEG) to develop the next generation of the 3D video coding standards. As a result, the HEVC second edition with scalability extension (SHVC) [89] and Multiview extension (MV-HEVC) [25] was completed in 2014 and published in early 2015.

In this chapter, MV-HEVC coding is introduced and evaluated in terms of compression performance relative to MVC. Before presenting MV-HEVC coding techniques in Section 5.3 we briefly describe HEVC and its main advanced coding tools in Section 5.2 Compression performance evaluation and comparison is performed in Section 5.4 The assessment is carried out using multiple MVV sequences of different quality and content.

## 5.2 HEVC Standard

The spread of high-definition video and beyond HD formats have pushed the need for more efficient video coding technologies to adapt the high-resolution visual content with the low bandwidth environments and the available storage devices. An emerging video compression standard namely HEVC has been released as a competent solution of the above requirement. HEVC was developed to cover almost all existing applications of the AVC standard with enhanced features, such as increasing compression performance, supporting videos with higher resolutions and improving parallel processing. Subjective evaluation results [90] show that HEVC/H.265 standard can reach the same quality levels as H.264/AVC whilst generating approximately 50% lower bitrate on average. To achieve this gain, HEVC standard adopted innovative tools such as accurate intra-/inter-predictions, in-loop sample adaptive offset filter and quadtree-based block partitioning. The reference software of HEVC which includes both encoder and decoder functionality is called (HM) [91]. Three main configuration structures of HEVC are defined in HM test model:

- All-intra structure: consists of encoding all frames using intracoded modes. This structure is dedicated for professional use where compression efficiency is not an important criterion.
- Random access structure: is the typical coding scheme for broadcast applications. It uses the hierarchical coding structure while intracoded (I) frames are regularly inserted every GOP.
- Low-Delay structure: recommended for applications where minimising end-to-end delay is highly required such as live video conference.

HEVC benefits from using variable pattern comparison and difference-coding areas starting from blocks of 16×16 to 64×64 pixels. The concept behind this is based on partitioning the frame into coding tree units (CTUs), which replace the macroblocks used in H.264. The coding structure is depicted in Figure 5.1.

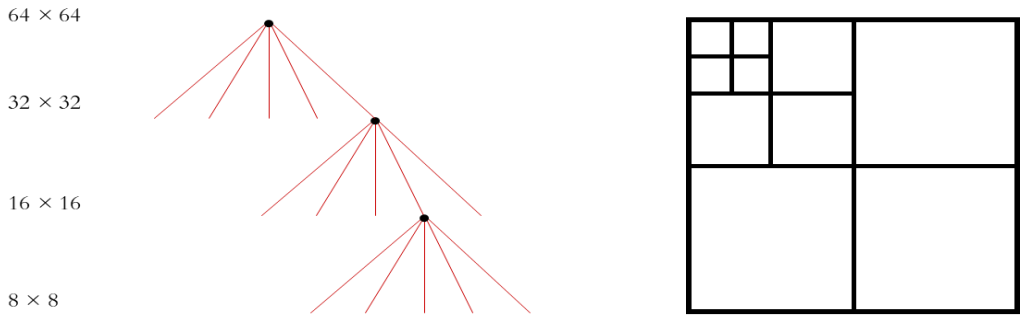



Figure 5.1 CTU partitioning and processing order in HEVC

Each CTU contains two chroma and one luma coding tree blocks CTBs. CTB size can be 16×16, 32×32 or 64×64, where larger pixel block size increases the compression efficiency. The CTBs are then divided into one or more coding units (CUs) as shown in Figure 5.2. The CU is split into prediction units (PUs), a basic entity for intra- and inter-predictions, variable in size from 64×64 to 4×4 pixels. Variable partition scenarios have been defined in the design of the HEVC encoder considering a certain attention to complexity.

For instance, to deal with critical case memory bandwidth in the decoding process, PUs coded using temporal inter-prediction are restricted to the minimum size of 8×8 if they are bi-predicted from two references, or 8×4 or 4×8 if they are predicted from a single reference.

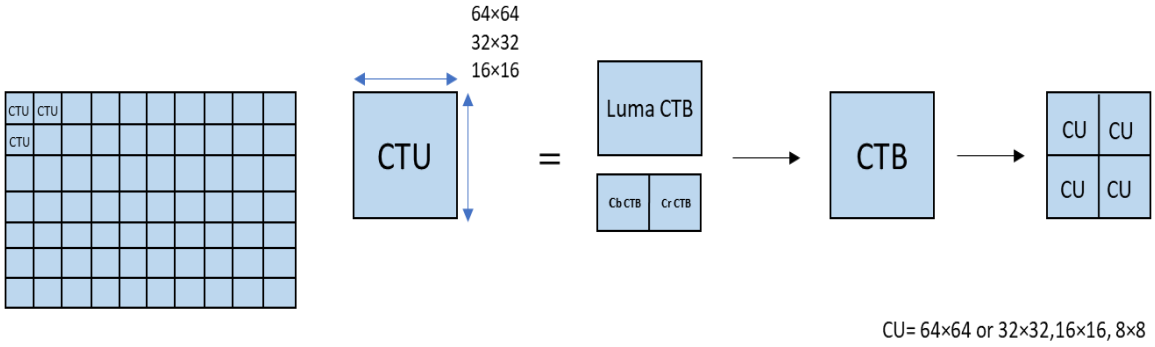



Figure 5.2 Luma and chroma CTB of HEVC

Both AVC and HEVC employ intra prediction mechanism. HEVC uses 33 intra prediction directions in addition to planar mode (0) and DC mode (1). However, only 8 directional modes are used in H.264. This significantly contributes to improve the compression efficiency of the intra frames of HEVC standard.

Furthermore, HEVC employs an improved motion prediction technique called Advanced Motion Vector Prediction (AMVP) compared to H.264 which uses Motion Vector Prediction (MVP) approach.

The HEVC bitstreams include an elementary unit called a network abstraction layer (NAL) unit, composed of playload and a header. The NAL header consists of a 5-bit NAL unit type, 6-bit layer identifier called nuh_layer_id, and a 3-bit temporal sub-layer identifier. A new video parameter set (VPS) structure has been included to the HEVC as metadata representation to allow the extension compatibility of the standard, and to include dependency between temporal sublayers. VPS also contains essential data that can be shared for the decoding process.

Tiles and wavefront parallel processing (WPP) are innovative features included in HEVC standard to enhance the parallel processing capability. Each of them may be useful in specific application contexts [23].

One of the distinguishable differences between HEVC and AVC standards is the resolution levels both codecs support. HEVC supports a resolution level up to 8K UHDTV (8192x4320) and frame rate up to 300 fps, whereas, H.264 is limited to 4K (4,096x2,304) resolution level and 59.94 fps.

## 5.3 MV-HEVC

A call for proposal [92] was issued in 2011 soliciting contributions from expert to develop further 3D video coding technology. Responses were good enough to facilitate establishing JCT-3V in July 2012. The JCT-3V main purpose was to develop a more advanced 3D video coding technology than the multiview video coding (MVC) extension of ITU-T H.264. The team concluded its works in June 2016, after analysing and defining 3D and multiview coding extensions for HEVC ITU-T H.265. 3D-HEVC was principally designed to compress video-plus-depth format whilst MV-HEVC addresses multiple textures video format.

MV-HEVC [93] was first integrated with the second edition of HEVC standard [94] and finalised later in February 2015.

MV-HEVC standard employs the same block coding tools of the HEVC main profile in addition to some specific features mainly related to the stereoscopic and multiview representations. MV-HEVC provides significant bitrate saving compared to the standard HEVC simulcast by enabling the exploitation of the interview references within the motion-compensated prediction. It is also noted that MV-HEVC utilises the same coding scheme (IBP) as the MVC standard. However, the concept of the interview has been replaced in MV-HEVC by the inter-layer prediction structure. The multi-layer approach is employed in all multi-layer extension [95], including MV-HEVC, 3D-HEVC as well as the scalable extension of HEVC (SHVC). A layer can represent a depth, texture or other auxiliary information related to a particular camera view. All layers of the same camera perspective are marked as a view; while layers representing the same type of information are denoted as components in 3D video. (See Figure 5.3)

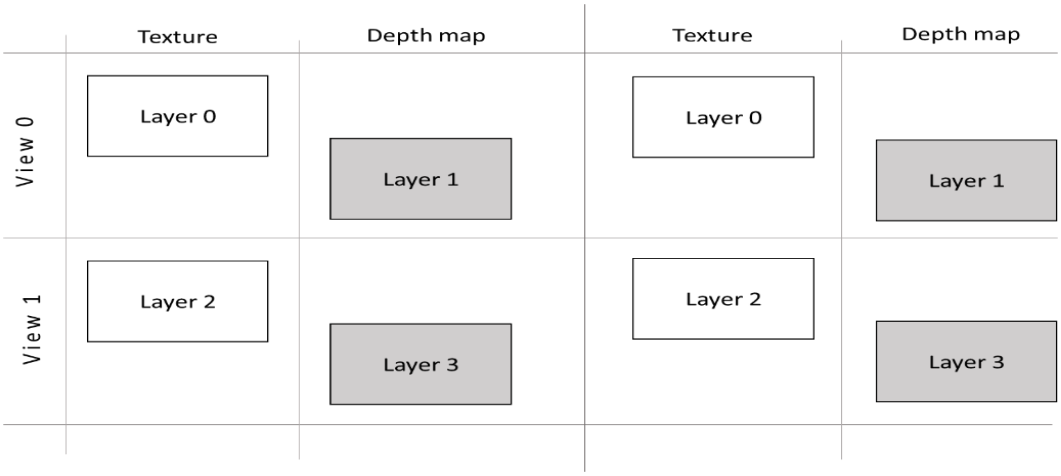



Figure 5.3 Layers division in MV-HEVC

MV-HEVC includes high-level syntax (HLS) additions [96] and can be implemented using existing 2D single-layer decoding cores. Moreover, MV-HEVC shares the same HLS with all HEVC multilayer extensions. The HLS enables the extraction of a single texture base view from MV-HEVC bitstream that is decodable by the main profile HEVC decoder.

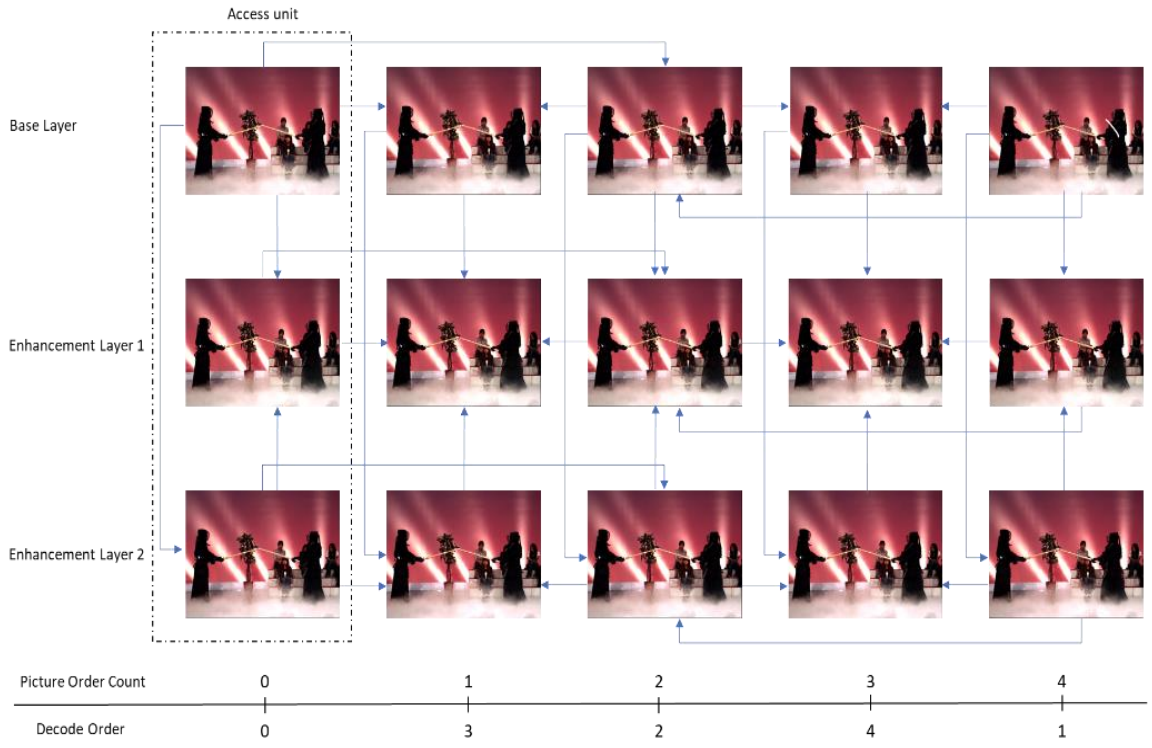



Figure 5.4 MV-HEVC bitstream with three texture views using IBP inter-view prediction

Figure 5.4 shows an example of an MV-HEVC bitstream with three texture views coded by the so-called IBP interview structure. The base layer (left view) is coded independently of other views using HEVC main profile. The MV-HEVC profile is enabled to code the two enhancement layers (ELs). EL2 (right view) utilises interview prediction from the base layer, and EL1 (centre view) is predicted from both left and right views.

## 5.4 Experimental evaluation

In this section, the performance of both MV-HEVC and MVC is compared and evaluated in terms of PSNR(dB) and bitrate (kbps) over several QP values. Four different video sequences have been used in the experiments. Table 5.1 describes the used multiview video sequences and their parameters. Also, samples of the tested sequences are shown in Figure 5.5.

Table 5.1 MVV sequences for compression efficiency evaluation

| Database | Video sequences | Frame rate | Image resolution | Camera parameters |
|---|---|---|---|---|
| MERL | Vassar | 25 | 640 × 480 | 8 cameras / 20 cm spacing |
| MERL | Ballroom | 25 | 640 × 480 | 8 cameras / 20 cm spacing |
| Fujii Lab | Kendo | 30 | 1024×768 | 7 cameras / 5cm spacing |
| Fujii Lab | Balloon | 30 | 1024×768 | 7 cameras / 5cm spacing |


Figure 5.5 First view frame of the used multiview video sequences

Table 5.2 regroups the common primary configuration that has been used to ensure a fair comparison. A total of 250 successive frames are encoded for each used sequence. The GOP size was equal to eight with guaranteed insertion of intra coded (I) frames at the end of each GOP. Four QP values have been chosen according to the standardisation tests defined in [97].

Table 5.2 Initial common encoding configuration

| | |
|---|---|
| Frame to be encoded | 250 |
| GOP size | 8 |
| Intra period | 8 |
| Quantisation Parameter | [25,30,35,40] |
| Search mode | Fast mode |
| Search range | 64 |

In addition, experiments were conducted using the latest available edition of each codec. Where HM 16.9 [98], which includes a multiview profile extension, was used for MV-HEVC. JMVC 8.5 codec was selected for MVC standard.

It should be mentioned that the used software models are developed using C++ programming language, and they are designed for research purposes and not for commercial applications. All simulations were carried out on a PC with Intel core i5 2.20 GHz CPU and 4 GB RAM.

As it was expected, results in Figure 5.6 and Figure 5.7 obviously show that MV-HEVC exceeds MVC in terms of bitrate saving and video quality. This outperformance ultimately covers all the carried out simulations through the different QP values and the various MVV sequences. The RD curves of the HD MVV sequences, illustrated in Figure 5.6, prove that MV-HEVC codec provides much better compression performance compared to MVC over the entire bitrate range. For example, when QP=20, the bitrate saving gain of MV-HEVC exceeds 25 % and 31 % for Balloon and Kendo sequences, respectively. Moreover, Figure 5.7 reveals that further bitrate saving gains were achieved by MV-HEVC for the standard definition MVV sequences, whereby a gain of 71% and 57% is marked for Vassar and Ballroom sequences, respectively.

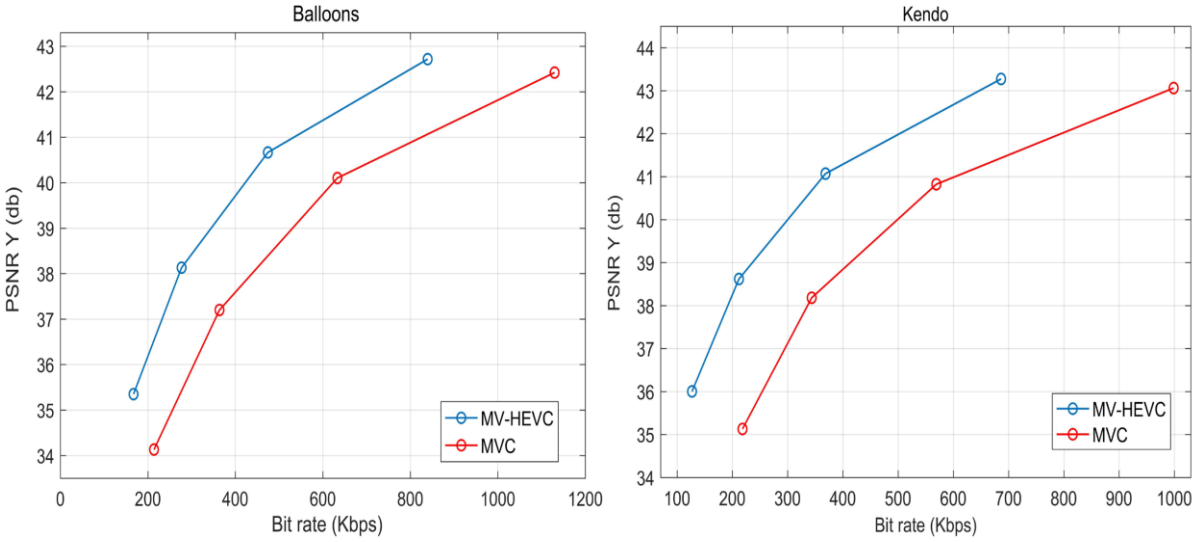



Figure 5.6 Compression performance comparison using HD Multiview video sequences

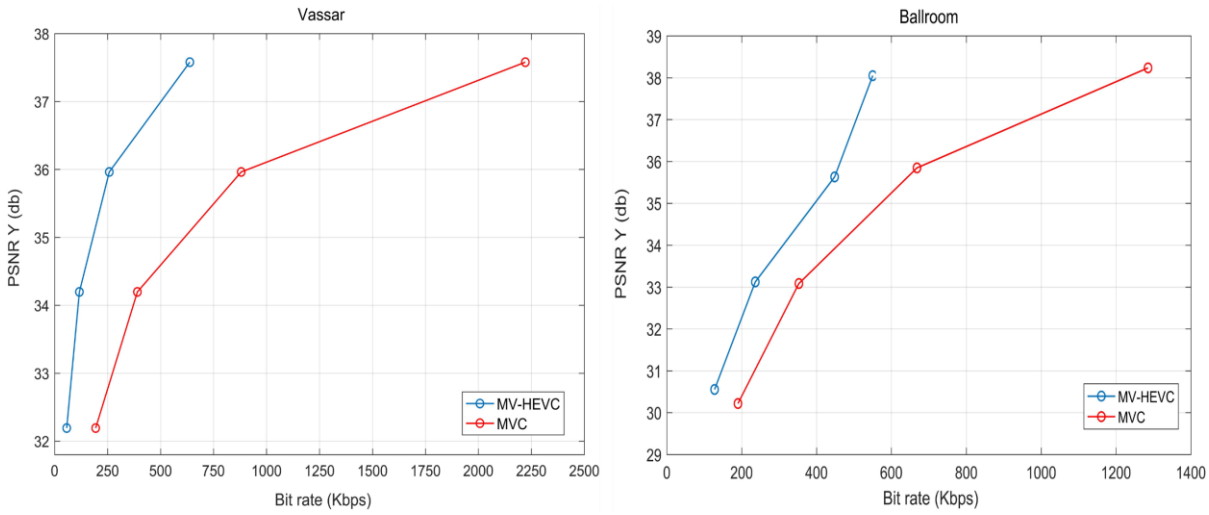



Figure 5.7 Compression performance comparison using SD Multiview video sequences

Figure 5.8 and Figure 5.9 illustrate a frame-based comparison between MV-HEVC and MVC codecs. Frame number 30 located in view 2 (camera 2) of the two chosen multiview video sequences is selected for this comparison. This frame which comes after three successive group of pictures is coded using both temporal and interview predictions. Also, the quantisation parameter QP=40 has been selected for this comparison to evaluate the performance of the reported codecs at the lowest level of perceptual image quality.

Figure 5.8 shows the comparison using a standard resolution video (Vassar), the degradation can be perceived in the compressed frame with MV-HEVC and MVC as well. However, the difference cannot be clearly seen between the two compressed frames. The MSE maps slightly highlight the difference between the two compressed frames, where extra red regions are observed in the frame compressed by MVC codec, which indicated a larger number of mismatching errors. However, the blue regions, which represent the matching between the original and the compressed frames, are distinctly perceived in the frame compressed by MV-HEVC. The PSNR values ascertain the MSE map results, where the PSNR (Y) gain of MV-HEVC is 0.7 dB, and the overall value is 0.69 dB.

Almost similar remarks can be deduced from Figure 5.9 where HD multiview video sequences have been used with the same quantisation parameter value for both codecs.

The results emphasise the same fact that MV-HEVC outperforms MVC in terms of image quality with a gain of 0.86 dB achieved for PNSR(Y) and 1.64 dB for the mean value which includes PSNR(Y), PSNR(U) and PSNR(V).

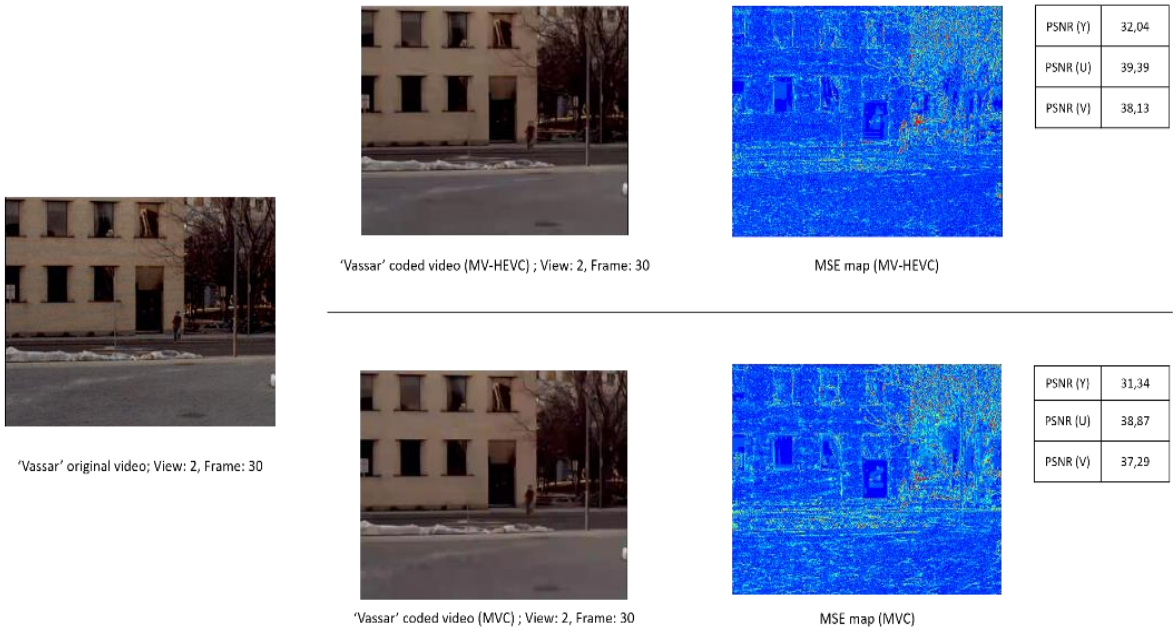



Figure 5.8 Image quality comparison between MV-HEVC and MVC using Vassar sequence

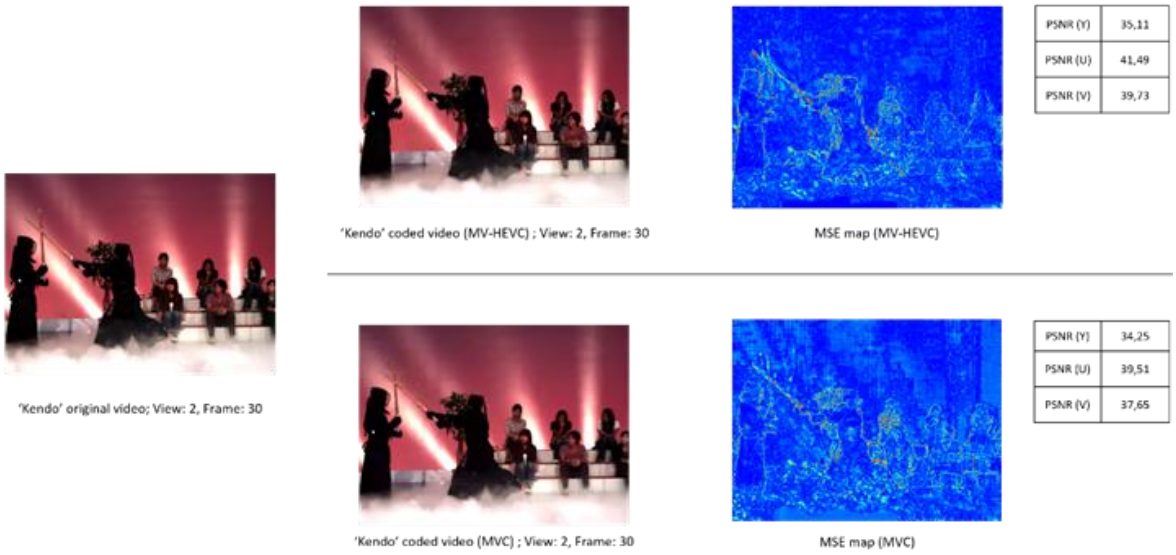



Figure 5.9 . Image quality comparison between MV-HEVC and MVC using Kendo sequence

## 5.5 Conclusion

This chapter briefly reviewed the theory and concepts behind the MV-HEVC coding standard. Some advanced tools of HEVC single layer codec were presented in Section 5.2 which are similarly used in the multilayer extension MV-HEVC. Both MVC and MV-HEVC employs the same IBP interview prediction schemes in addition to the hierarchical B structure for the temporal level. The principal differences between MV-HEVC and MVC consist of the base view single layer coding tools and features where MV-HEVC utilises the powerful tools of HEVC such as the innovative block partitioning to enhance the rate distortion performance. Both codecs were simulated and evaluated using common test conditions and different MVV sequences. Obtained results proved the outperformance expectations of MV-HEVC against MVC in terms of compression efficiency. The substantial bitrate saving gains started from 24% for Balloons sequences and achieved 70% for vassar sequences.

The main purpose of this chapter was to investigate the latest multiview video coding technology in order to identify the coding tools of our forthcoming research work. The next chapter will introduce our future workplan in the multiview video domain.

# Chapter 6
# Conclusion and Future Work

This chapter summarises the major findings of the PhD thesis and suggests research directions for potential future work. The thesis summary is provided in section 6.1 while the general conclusion and suggestions for future extension are presented in section 6.2 and 6.3 respectively.

## 6.1 Thesis Summary

Existing MVC codecs generally focus on optimising the compression performance while devoting less attention to other critical requirements such as the random access ability. This thesis addressed the low random access ability problem of the MVC to improve users experience with the 3D video content.

The thesis provided new solutions for this problem by proposing:

- An enhanced multiview prediction scheme (PBI) which provided a fast random access and decent compression performance compared to the benchmark model MVC.
- A second proposed framework (PIP) which surpassed the PBI random ability while maintaining a competitive bitrate/quality trade-off.

Chapter 2 briefly presented the 3D video progression history and its current market status. An overview study was then presented including the three principal stages of the 3D video production chain: 3D video acquisition, 3D display and 3D video coding. More focus was given to the 3D video coding part and more precisely to the multiview video coding, its concepts, requirements and the compression technology behind it.

Chapter 3 detailed the multiview coding technology through its relevant standard MVC/H.264 and other related multiview coding designs. After analysing the existing interview prediction solutions, the first proposition (PBI) was introduced to tackle the random access problem. PBI multiview prediction design was initially proposed within a multiview system of eighth parallel cameras. Moreover, PBI scheme has been extended to support multiview system of more than eight views depending on the implementer specifications and choices. PBI has been proved as a practical solution for the random access requirement under different conditions and multiple data inputs. Results of the proposed PBI were compared to default MVC structure and other related works.

Chapter 4 proposed a second approach (PIP) to further facilitate the random access capacity. An investigation of the GOP effects on the considered multiview video coding was also performed. Considerable gains were made by the PIP approach in terms of random accessibility compared to MVC default scheme and the PBI technique. 3D trade-off illustrations have been included, exposing the GOP size effects on the compression efficiency and random access ability of PBI and PIP coding structures relative to MVC.

Chapter 5 presented an overview of HEVC/H.265 and its multiview extension. Experimental tests have been carried out between the MVC codec and its successor MV-HEVC. Both codecs use the same IBP design and hierarchical B algorithm for the disparity compensation and the temporal level, respectively. Also, the MV-HEVC employs the powerful tools of HEVC such as the innovative block partitioning to improve the rate-distortion capability. Both codecs have been implemented and evaluated through different datasets and conventional test conditions. MVV texture sequences without considering depth map of SD and HD resolutions have been utilised as test elements. The obtained results proved the outperformance of MV-HEVC over MVC in terms of compression efficiency. Chapter 5 represents an introduction to trigger our future research on multiview video coding by making use of the latest testing model MV-HEVC.

## 6.2 General Conclusion

Many research efforts based on the MVC standards are being been made to improve the coding capability with regard to the multiview video coding requirements list. Switching and navigating within the 3D content is a primordial feature to enhance the 3D immersive experience. The technical aspect to boost this type of interactivity is enhancing random access ability functionality of the adopted multiview video codec. This thesis aimed to develop an improved multiview video coding frameworks that offer better content interactivity for the 3D video users. Multiview video coding schemes with improved random access ability have been proposed, tested and evaluated. Compression efficiency is also a principal requirement to optimise storage, bitrate streaming and video quality. This requirement has been also considered along with the proposed approaches. Two multiview prediction models have been proposed and tested within the scope of the MVC common test conditions defined by the standardisation community. The first approach PBI provides fast random access to the whole set of GGOP by enabling direct interview predictions from two base views to the remaining views of the scheme. Faster random access has been achieved compared to default MVC standard and other works.

A generalised adaptive model of the PBI scheme has been proposed depending on the views number in the MVV system. This latter has led to further enhancement in the view random access ability compared to IBP and Amr. An accurate evaluation method is primordial to ensure a fair comparison between the considered multiview schemes. To this end, new evaluation metrics were proposed allowing global random access assessment which takes into consideration each frame within the GGOP. $G_{RA}$, $G_{RN}$ and $G_R$ are the three proposed metrics to evaluate the random access performance. It was verified throughout the thesis that our proposed evaluation methods offer more accurate results compared to the standard $N_{max}$ metric. Future research on random access ability might utilise the $G_R$ evaluation approach for fair and accurate comparison between the examined multiview coding schemes. Additionally, a second approach PIP has been proposed to further enhance the random access ability of multiview video coding. "PIP" structure achieved significant results in terms of random accessibility yielding a gain of 53.33 % against the default MVC. Both proposed schemes "PBI" and "PIP" were evaluated and compared over different GOP sizes and quantisation parameters. Key remarks have been deduced from the conducted experimental tests as follows: A reduced GOP size always produces better random access ability and less bit-rate saving. In opposite, larger GOP size provides slower random access ability and better bit-rate saving.

Finally, the thesis examined the recent video codec HEVC while focusing on its multiview video extension MV-HEVC. MV-HEVC has been implemented, compared to MVC codec and evaluated through different datasets and common test conditions. The obtained test results show a significant compression efficiency of MV-HEVC compared to MVC, with a variable bit rate saving gain depending on the used multiview video sequences. Theory and concepts study of MV-HEVC in addition to experimental tests paved the way for our future research which is briefly presented in the following section.

## 6.3 Future Work

The thesis presented software-based solutions that improved the random access ability of the multiview video coding. In combination with the satisfactory obtained results in the research scope so far, some ideas for future research are to be executed in order to extend and develop the proposed designs to cover other coding requirements. Mandatory suggestions for future work as described as follows:

- Results of the proposed approaches have been only evaluated using objective evaluation by means of PSNR values. This could constantly lead to significant conclusions. However, subjective evaluation based on Mean Opinion Scores (MOSs) would affirm further the findings through live visual experiences.

- Two video resolutions have been addressed in this thesis (640x480 and 1024x768) according to the available MVC common dataset and test conditions. Coming dataset releases may include 4K UHD and 8K UHD multiview video sequences which will be considered to further validate the effectiveness of the proposed schemes. It is also important to note that the proposed predictions structures were not designed to handle arrangements with arbitrary cameras positioning. Therefore, improved designs that support different arrangement types might be subject for our future research.

- The thesis proposed software-based designs for speeding up the random access of the MVC. Implementing the proposed solutions in hardware environments is an important perspective to consider for future work. The hardware implementation will permit the evaluation of other MVC aspects such as the encoding time, resource consumption and the parallel processing functionality. A review of the candidate hardware architectures, in which MVC could be successfully implemented, is required to treat the costly encoding function such as motion and disparity estimations. For instance, considering Multi-Chip GPUs for offloading MVC encoding process is a suggested research area.

- Error resilience is another feature that is in the pool of our future works. The first step consists of evaluating the resulted bitstreams of the proposed multiview prediction schemes over error-prone networks of different topologies and conditions. The next step must be a thorough study of the already existing error resilience mechanisms in the conventional video coding to identify either possibility of technology adaptation or proposing innovative error robustness tools for the MVC and MV-HEVC, where an additional error propagation layer is introduced.

Last but not least, our ultimate future research goal is to develop an improved end-to-end multiview video system by exploiting the existing technologies and proposing cost-effective tools in the main three stages of the MVV systems: acquisition, coding and displaying. This research goal would not be possible without joining forces of other experts. The multiview end-to-end system will offer an open source platform and flexible prototypes that can be used for research and education purposes.

# Bibliography


[1] Sabra. A, The optics of Ibn al-Haytham. Warburg Institute, London, 1989.

[2] Cisco, Visual Networking Index: Forecast and Methodology, 2016–2021,2017.

[3] T. Okoshi, "Three-Dimensional Imaging Techniques", Academic Press, UK, 1976.

[4] Y. Kim et al., "Three-dimensional display technologies of recent interest: principles, status, and issues," Applied optics, 50(34), H87–115, 2011.

[5] Urey. H et al., "State of the art in stereoscopic and autostereoscopic displays," Proceedings of the IEEE, 99(4):540–555, 2011

[6] H.J. Choi, "Current status of stereoscopic 3D LCD TV technologies," 3D Research, 2(2):1-4, Nov. 2011.

[7] Lutz Goldmann et al., "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," Proc. SPIE, 7526, 7526 - 7526 – 11, 2010.

[8] W. J. Tam, F. Speranza et al., "Stereoscopic 3D-TV: Visual Comfort," IEEE Transactions on Broadcasting, 57(2), 335-346, June 2011.

[9] O. Cakmakci and J. Rolland, "Head-worn displays: a review," Journal of Display Technology, 2(3), 199–216, 2006.

[10] V. Ferrari, G. Megali et al, "A 3-D Mixed-Reality System for Stereoscopic Visualization of Medical Dataset," IEEE Transactions on Biomedical Engineering, (56)11, 2627-2633, Nov. 2009.

[11] Jangllin Chene et al, "Handbook of Visual Display Technology," Springer, 2016.

[12] G. E. Favalora, "Volumetric 3D Displays and Application Infrastructure," Computer, 38(8), 37-44, Aug. 2005.

[13] P. Hariharan, "Basics of holography", Cambridge University Press, 2002.

[14] Martin Richardson, John D Wiltshire," The hologram: principles and techniques," Hoboken, NJ: John Wiley & Sons, 2018.

[15] Aggoun. A et al., "Immersive 3D holoscopic video system," IEEE Multimedia, 20(1), 28–37, March 2013.

[16] Swash M. R., Aggoun, A. et al., "Holoscopic 3D image rendering for Autostereoscopic Multiview 3D Display," IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, (BMSB), 1–4, London, 2013.

[17] M. Domański, A. Dziembowski et al., "Poznan University of Technology test multiview video sequences acquired with circular camera arrangement, Poznan Team and Poznan Blocks sequences", ISO/IEC JTC1/SC29/WG11, Doc. MPEG M35846, Geneva, 2015.

[18] N. A. Dodgson, "Autostereoscopic 3D Display," IEEE Computer Society, 38(8), 31–36, Aug 2005.

[19] N. S. Holliman, N. A. Dodgson, G. E. Favalora and L. Pockett," Three-dimensional displays: A review and applications analysis," IEEE Transactions on Broadcasting, 57(2), 362–371, June 2011.

[20] Akamai Technologies, state of the internet, Q1 2017 report 10-1, 2017.

[21] A. H. Sadka, "Compressed Video Communications," Halsted Press, 2002.

[22] J. Ostermann et al.," Video coding with H.264/AVC: tools, performance, and complexity," IEEE Circuits and Systems Magazine. 4(1), 7– 28, 2004.

[23] G. J. Sullivan et al., "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Circuits and Systems Magazine. 22(12), 1649 – 1668, Dec. 2012.

[24] P. Merkle A. Smolic, K. Muller and T. Wiegand, "Efficient prediction structures for multiview video coding," IEEE Trans. Circuits Syst. Video Technol. 17(11), 1461–1473, Nov. 2007.

[25] G. Tech et al., "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding". IEEE Trans. Circuits Syst. Video Technol., 26(1), 35–49, 2016.

[26] S. Nasri, K. Khelil, N. Doghmane, "Enhanced view random access ability for multiview video coding," Journal Electronic Imaging 25(2), 023027, April 2016.

[27] S. Nasri, A. H. Sadka, N. Doghmane, K. Khelil, "Group of Pictures Effects on Proposed Multiview Video Coding Scheme", 40th International Conference on Telecommunications and Signal Processing, 548– 554, Barcelona, 2017.

[28] Charles Wheatstone, "Contributions to the physiology of vision – part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision," Philosophical Transactions of the Royal Society of London, 128:371–394, 1838.

[29] Charles Wheatstone, "Contributions to the physiology of vision – part the second. on some remarkable, and hitherto unobserved, phenomena of binocular vision," Philosophical Transactions of the Royal Society of London, 142:1–17, 1852.

[30] SPOTLIGHT, "Charles Wheatstone: the father of 3D and virtual reality technology", King's College London, 2016. [Online]. Available: spotlight.kcl.ac.uk.

[31] David Brewster, "The Stereoscope; its History, Theory, and Construction, with its Application to the Fine and Useful Arts and to Education," London, J. Murray, 1856.

[32] Stephen A. Benton, "Selected Papers on Three-Dimensional Displays," SPIE Optical Engineering Press, 2001.

[33] The Wachowski Brothers, "The Matrix (movie)". http://www.imdb.com/title/tt0133093/, 1999.

[34] James Cameron, "Avatar (movies)". http://www.imdb.com/title/tt0499549/, 2009.

[35] DisplaySearch, 3D Display Technology and Market Forecast Report, 2010.

[36] Project description in: www.bbc.co.uk/rd/projects/iview.

[37] O. Grau, G. A. Thomas et al.,"A Robust Free-Viewpoint Video System for Sport Scenes," 2007 3DTV Conference, 1-4, Kos Island, 2007.

[38] Laurent Lucas, Céline Loscos and Yannick Rémion, "Multiview Acquisition Systems. 3D Video: From Capture to Diffusion", Wiley-ISTE, Dec. 2013.

[39] R. B. Johnson and G. A. Jacobsen, "Advances in lenticular lens arrays for visual display," Proc. SPIE 5874, Current Developments in Lens Design and Optical Engineering VI, 587406, Aug. 2005.

[40] Cees van Berkel, John A. Clarke, "Characterization and optimization of 3D-LCD module design", Proc. SPIE 3012, Stereoscopic Displays and Virtual Reality Systems IV, May 1997.

[41] M. E. Lukacs, "Predictive coding of multi-viewpoint image sets", IEEE International Conference on Acoustics, Speech, and Signal Processing, 521-524, Japan, 1986.

[42] Dinstein et al., "On the com-pression of stereo images: Preliminary results", Signal Processing: Image Communications, (17)4, 373-382, Aug. 1989.

[43] M. G. Perkins, "Data compression of stereo pairs", IEEE Transactions on Communications, 40(4), 684-696, April 1992.

[44] ITU-T and ISO/IEC JTC 1, Final draft amendment 3, Amendment 3 to ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video), ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. N1366, 1996.

[45] ITU-T and ISO/IEC JTC 1, Generic coding of moving pictures and associated audio information Part 2: Video, ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video), 1994.

[46] Puri A., R. V. Kollarits, and B. G. Haskell, "Stereoscopic video compression using temporal scalability", Proc. SPIE Conf. Visual Communications and Image Processing, 2501, 745–756, 1995.

[47] J.-R. Ohm, "Stereo/Multiview Video Encoding Using the MPEG Family of Standards", Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems VI, San Jose, CA, Jan. 1999.

[48] ISO/IEC JTC1/SC29/WG11, Report on 3DAV exploration. Output doc. N5878, 65th MPEG meeting, Trondheim, Norway, Jul. 2003.

[49] ISO/IEC JTC1/SC29/WG11, Call for proposals on multi-view video coding. Output doc. N7327, 73rd MPEG meeting, Poznan, Poland, Jul. 2005.

[50] MPEG Video Sub-Group Chair (J.-R. Ohm), Submissions received in CfP on multiview video coding, ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. M12969, Bangkok, Thailand, 2006.

[51] MPEG Video and Test Sub-Groups, Subjective test results for the CfP on multi-view video coding, ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. N7799, Bangkok, Thailand, 2006.

[52] K. Muller et al., "Multiview coding using AVC", ISO/IEC JTC 1/SC 29/WG 11 (MPEG) Doc. M12945, Bangkok, Thailand, 2006.

[53] ISO/IEC JTC1/SC29/WG11, "Requirements on multiview video coding," Output doc. N8218, 77th MPEG meeting, Klagenfurt, Austria, Jul. 2006.

[54] Y. Chen, P. Pandit, and S. Yea, "WD 4 reference software for MVC," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-AD207, 2009.

[55] MPEG requirements sub-group, "requirements on multi-view video coding," July 2009.

[56] K. Ugur, H. Liu, J. Lainema, M. Gabbouj and H. Li, "Parallel Encoding - Decoding Operation for Multiview Video Coding with High Coding Efficiency," 3DTV Conference,1-4, Kos Island, 2007.

[57] H. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," IEEE Transactions on Circuits and Systems for Video Technology, 13(7), 598-603, July 2003.

[58] H. Nguyen and P. Duhamel. "Iterative joint source-channel decoding of variable length encoded video sequences exploiting source semantics International Conference on Image Processing, 5, 3221-3224, 2004.

[59] H. Nguyen, P. Duhamel, J. Brouet, and D. Rouffet. "Robust VLC sequence decoding exploiting additional video stream properties with reduced complexity,". IEEE International Conference on Multimedia and Expo, 375–378, Taipei, Taiwan, June 2004.

[60] "Advanced video coding for generic audiovisual services," ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC), 2013.

[61] Vetro, T. Wiegand and G. J. Sullivan," Overview of the Stereo and Multiview Video Coding Extensions of the H 264 / MPEG-4 AVC Standard," Proceedings of IEEE, 99(4), 626–642, April 2011.

[62] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Scalable Video Coding, 17, (9), 1103-1120, September 2007.

[63] N. Cheung and A. Ortega, "Distributed Source Coding Application to low-delay Free Viewpoint Switching in Multiview Video Compression", Picture Coding Symposium 2007, Portugal, 2007.

[64] Y. Liu, et. al., "Low-delay View Random Access for Multi-view Video Coding," IEEE International Symposium on Circuits and Systems, 997-1000, May 2007.

[65] Y. Zhang, G. Jiang, M. Yu, and Y. S. Ho, "Adaptive multiview video coding scheme based on spatio-temporal correlation analyses," ETRI Journal, 31(2), 151–161, Apr. 2009.

[66] Y. Yang, Q. Dai, J. Jiang, and Y-S. Ho, "Coding Order Decision of B Frames for Rate-Distortion Performance Improvement in Single-View Video and Multiview Video Coding IEEE Transactions on Image Processing, 19(8), 2029-2041, Aug 2010.

[67] L. Ma and F. Pan, "Efficient compression of multi-view video using hierarchical B pictures," International Conference on Multimedia and Ubiquitous Engineering, 118–121 Busan, 2008.

[68] Oka, S., Fujii, T., Tanimoto, M., "Dynamic ray-space coding using inter-view prediction" Proceeding of International Workshop on Advanced Image Technology, 19–24, Jan. 2005.

[69] Kazui K et al., "Response to Call for Evidence on multi-view video coding," Doc. M11596, 71th MPEG meeting, Hong Kong, China, Jan. 2005.

[70] ISO/IEC JTC1/SC29/WG11, "Survey of algorithms used for multi- view video coding (MVC)," Output doc. N6909, 71st MPEG meeting, Hong Kong, China, Jan. 2005.

[71] [Jun Xin, Emin Martinian, Anthony Vetro, " Multiview video decomposition and encoding." U.S. Patent 11015390, issued December 17, 2004.

[72] Girod B. Girod, A. M. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed Video Coding," Proceedings of the IEEE, 93, (1), 71-83, Jan. 2005.

[73] X. Guo et al., "Distributed multiview video coding," Proceeding of IS&T/SPIE International Conference on Visual Communications and Image Processing, 6077, 290–297, 2006.

[74] M. Ouaret et al., "Fusion-based multiview distributed video coding," Proceeding of ACM International Workshop on Video Surveillance and Sensor Networks, 139–144, 2006.

[75] Artigas, Xavier et al., "Comparison of different side infor- mation generation methods for multiview distributed video coding," Proceeding of the International Conference on Signal Processing and Multimedia Applications, Jul. 2007.

[76] Chuo-Ling Chang et al., "Inter-view wavelet compression of light fields with disparity-compensated lifting," Proceeding of IS&T/SPIE International Conference on Visual Communications and Image Processing, 14–22, 2003.

[77] W. Yang, Y. Lu, F. Wu, J. Cai, K. N. Ngan and S. Li, "4-D Wavelet-Based Multiview Video Coding," in IEEE Transactions on Circuits and Systems for Video Technology, 16 (11), 1385 –1396, Nov. 2006.

[78] Anthony Vetro et al., "Multi- view Video Model (JMVM) 8.0," Doc. JVT-AA207, 27th JVT meeting, Geneva, Switzerland, Apr. 2008.

[79] H. Yang et al., "CE1: Fine motion matching for motion skip mode in MVC," Doc. JVT-Z021, 26th JVT meeting, Antalya, Turkey, Jan. 2008.

[80] J. H. Kim et al., "New coding tools for illumination and focus mismatch compensation in multiview video coding," IEEE Transactions on Circuits and Systems for Video Technology 17 (11), 1519–1535, Nov. 2007.

[81] U. Fecker and A. Kaup, "Complexity evaluation of random access to coder multi-view video data," ISO/IEC JTC1/SC29/ WG11, N8019, 2006.

[82] Joint Video team, "JMVC reference software 8.5", 2011.

[83] Bekhouch and N. Doghmane, "Multiview video coding with an improved prediction structure for faster random access," Journal of Electronic Imaging 22(4), 043010, 2013.

[84] X. Lv, L. Ma, and J. Guo, "Multiview video coding scheme based upon enhanced random access capacity," International Journal of Computer Science Issues, 10(1), 285-289, Jan. 2013.

[85] ITU-R Recommendation BT.202-1, "Subjective methods for the assessment of stereoscopic 3DTV systems," Feb. 2015.

[86] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," Electronics Letters, 44(13), 800-801, June 2008.

[87] "Common test conditions for multiview video coding," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, July 2006.

[88] Gary J. Sullivan et al, "Standardized extensions of high efficiency video coding (HEVC)," IEEE Journal on Selected Topics in Signal Processing, 7(6), 1001–1016, Dec. 2013.

[89] J. M. Boyce et al., "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, 26(1), 20–34, Jan. 2016.

[90] J.-R. Ohm et al., "Comparison of the coding efficiency of video coding standards— Including High Efficiency Video Coding (HEVC)," IEEE Transactions on Circuits and Systems for Video Technology, 22(12),1669–1684, Dec. 2012.

[91] Frank Bossen et al, "HM software manual,": Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC 1/SC 29/WG 11, 21st Jan. 2018.

[92] MPEG Video and Requirements Group, Call for Proposals on 3D Video Coding Technology, document N12036, Geneva, Switzerland, Mar. 2011

[93] G. Tech et al., "MV-HEVC DraftText9," document JCT3V-I1002, Sapporo, Japan, Jul. 2014.

[94] High Efficiency Video Coding, document Rec. ITU-T H.265, Oct. 2014.

[95] M. M. Hannuksela, Y. Yan, X. Huang and H. Li, "Overview of the multiview high efficiency video coding (MV-HEVC) standard," IEEE International Conference on Image Processing (ICIP), 2154-2158, Quebec City, 2015.

[96] R. Sjoberg et al., "Overview of HEVC high-level syntax and reference picture management," IEEE Transactions on Circuits and Systems for Video Technology, 22(12), 1858–1870, Dec. 2012.

[97] V. Baroncini et al., "MV-HEVC Verification Test Report", Joint Collaborative Team on 3D Video Coding Extensions of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Doc. JCT3V-N1001, 2016.

[98] HM16.9 Software, https://hhi.fraunhofer.de/svn/svn_3DVCSoftware.