Università degli Studi di Padova

Facoltà di Ingegneria

Corso di Laurea in Ingegneria delle Telecomunicazioni

Tesi di Laurea Specialistica

State of the art 3d technologies and MVV end to end system design

Relatore: Ch.mo Prof. Gianfranco Cariolaro

Correlatore: Ing. Luca Carniato

Laureando: Gabriele Longhi

Padova, 13 Aprile 2010

e-mail: lg.longhi@alice.it

Dedico questa tesi ai miei genitori
a mia sorella e al CdSA,
supporti fondamentali
durante la mia carriera universitaria.

L'autore Gabriele Longhi.

Prefazione

L'oggetto del presente lavoro di tesi è costituito dall'analisi e dalla recensione di tutte le tecnologie 3D: esistenti e in via di sviluppo per ambienti domestici; tenendo come punto di riferimento le tecnologie multiview video (MVV). Tutte le sezioni della catena dalla fase di cattura a quella di riproduzione sono analizzate. Lo scopo è di progettare una possibile architettura satellitare per un futuro sistema MVV televisivo, nell'ambito di due possibili scenari, *broadcast* o interattivo. L'analisi coprirà considerazioni tecniche, ma anche limitazioni commerciali.

Nella prima parte della tesi è presentata la panoramica delle attuali e future tecnologie 3D e MVV. Per quanto concerne la fase di cattura, essa sfrutta la visione stereoscopica umana e ne emula il funzionamento, a essa è strettamente legato il formato video scelto, poiché esso stabilisce il numero e il tipo di telecamere da utilizzare, nell'ambito dell'MVV i formati preferibili sono l'LDV, il DES e l'MVD, che sfruttano mappe di profondità ed eventualmente di occlusione per mandare un numero di viste minore di quelle rappresentate sul display. Le viste intermedie sono ricavate sfruttando algoritmi appositi (depth image based rendering) i quali sfruttano le informazioni di profondità e di occlusione.

Attualmente sono standardizzati due tipi di codifica video, volti all'MVV, l'MPEG-C, che supporta formati video con mappe di profondità, e l'MVC, contenuto in un'estensione dell'H264/AVC, che invece manda più viste simultaneamente e riduce il bitrate globale, sfruttando la correlazione tra viste adiacenti oltre a quella temporale già gestita dall'H.264. Un altro standard che sintetizzi i due precedenti è in fase di sviluppo, il 3Dvideo coding.

Per quanto riguarda l'utente finale, le maggiori osservazioni sono rivolte ai televisori, per impieghi MVV quelli più adatti sono i televisori auto stereoscopici, che permettono di rappresentare più viste sullo schermo simultaneamente, e non prevedono l'utilizzo di occhialini speciali, differentemente da quanto accade oggi nei cinema. I primi esemplari di questi televisori stanno raggiungendo il mercato, ma prima che lo conquistino definitivamente, sono necessari miglioramenti per quanto concerne la qualità di visualizzazione e i livelli di profondità, l'abbassamento dei costi e il proliferare di contenuti per questo nuovo servizio.

Nella seconda parte è riportato un possibile scenario futuro per il sistema MVV *end to end*.

Due differenti scenari sono possibili, uno *broadcast*, sviluppo di canali televisivi 3D o MVV, e uno interattivo, sistemi di videoconferenza.

I due servizi presentano requisiti diversi, da una parte è fondamentale la qualità dell'immagine, dall'altra la necessità dell'eseguire la trasmissione in tempo reale comporta vincoli sul ritardo della trasmissione, mentre la qualità non deve necessariamente essere così elevata.

Non esistono oggigiorno sistemi di videoconferenza end to end 3D o MVV, sono solo in fase di studio alcuni sistemi funzionanti, per cui non è stato possibile eseguire alcuna simulazione per questo servizio. Il formato riservato alle videoconferenze prevede di mandare due viste e due mappe di profondità, ognuna a un quarto della risoluzione, la qualità ne risente ma vi è un consistente guadagno in termini di bit rate. I requisiti di *delay* prevedono che i satelliti GEO, non siano adeguati allo scopo, mentre potendo deviare il traffico sui satelliti LEO o MEO, essi sarebbero rispettati.

Invece, per quanto concerne lo scenario *broadcast*, si è sviluppato un flusso video su quattro livelli, che assicura allo stesso tempo retrocompatibilità a servizi e apparecchiature 2D/3D esistenti, e scalabilità del servizio. Il bit-rate del più alto livello di qualità del servizio richiede un bit-rate stimato intorno ai 30 Mb/s, circa quattro volte quello necessario per un odierno canale HD.

Risultano quindi necessari molti sviluppi in primo luogo delle capacità dei satelliti e delle tecniche di codifica prima che questo servizio sia commercialmente fruibile e vendibile.

I tempi previsti per l'affermazione di questa tecnologia sono lunghi e si aggirano intorno ai 10-15 anni.

Il lavoro è stato una delle principali attività svolte durante un tirocinio effettuato presso la società vicentina Open-Sky. Nello specifico l'oggetto della tesi è parte di un progetto indetto dall'Agenzia spaziale Europea (ESA), eseguito in collaborazione con le società HHI e Astrium, rispettivamente tedesca e francese.

Il progetto, come stadi successivi, prevedeva una fase di test e simulazioni sul canale satellitare e sulla percezione dell'utente finale, in cascata a questa vi era un'analisi commerciale riguardante, principalmente, i costi e la complessità dell'intera architettura e una possibile pianificazione di come e in che ambiti diffondere il sistema. Questa parte del lavoro è riportata nella tesi di *Christian Fiocco*.

Contents

Introduction

This thesis describes the work made during an internship in Open-Sky. The main activities covered an ESA project about multi video view technologies (MVV).

In the last years cinema environments have focused their interests in 3D movies, the increase of 3D content is growing the user interest in 3D, so it's expected that it will also gain a large home usage.

3D cinema and 3D television have been studied for decades, history of 3D display is almost as old as conventional cinemas and television, but only in the last years the interest in these contents is significantly growing, users want a more involving experience, from the other side content producers, distributors and equipment providers consider it a new business opportunity.

It's expected that in the next years interest in 3D will progressively increase, more and more cinema are equipping with 3D technology, and also on producer side a lot of investments are being done.

The best way to reach mass market for home environment seems to be the MVV technology, as far as 3D appears a too intrusive experience for comfortable home usage. The most issue relies on special glasses, assumed acceptable for cinema environment, where it is seen more as an event, but unacceptable for home, where the user has a different approach to the service.

In this scenario 3D@Sat project falls, it's a project indicted by ESA (European Space Agency), its aim is to conduct a study on “3D scalable multi-view Satellite services”. This thesis covers the first half of the project in which a state of the art of 3D MVV/FVV technologies is performed and some possible future MVV scenarios are presented. The following phases of the project concern a simulation phase divided into a satellite channel simulation and in subjective tests on human 3D perception, followed by a commercial analysis on MVV future development and on the impact of the whole new architecture. These last phases are discussed in the work of Christian Fiocco, another trainee in Open-sky.

In the project Open Sky exploited its large experience in telecommunication services, linking commercial requirements to technical capabilities. My role, in the name of Open Sky, covered the state of the art review as concerns displays, STBs, interconnections and encoding techniques; the overview on video formats has been performed in collaboration with HHI, the architecture of the whole chain for future MVV systems and the two examined scenarios have been carried out jointly with HHI suggestions. As regards the capture phase, this work extends considerations made in the project by HHI, that has great experience in this field. Finally as regards next phases the transmission analysis requested collaboration of the three partners, while the simulation section had been performed by HHI and Astrium. The last step, commercial analysis it's all up to Open-Sky.

The thesis is structured in 7 chapters. After a brief introduction of the work, in the second chapter are presented 3D and MVV with a brief history of 3D and is reported the work proposal of the 3D@Sat project with a little description of the companies involved in it.

In following chapters the whole chain of a video transmission from the shooting to the display phase is analyzed.

In chapter 2 are examined the human visual system (HVS) and then the capture and the post production phases. Moreover are presented the production and the delivery format, that describes parameters of the cameras and of a simplified virtual camera setup where video and depth information are collected. These parameters are then used for any post processing correction as regards, as examples, lens distortions corrections or rectification procedure.

In chapter 3 the lens focus on 3D video formats that describe how the 3D scene is captured and encoded. As a matter of fact the selected format, defines from one side number and typology of the cameras and on the other side the encoding technique needed for the transmission. In the same chapter are reviewed different MVV coding techniques, from the one nowadays standardized as MPEG-C Part 3 and MVC extension of H264/AVC, to future standard, still under research, as 3DVC, that would be a synthesis of the two already available, supporting a multi video plus depth format.

In chapter 4 a deep 3D technologies review is performed, focusing on those now available and on future development expected. The review covers displays, set top boxes, interconnection cables and representations.

After the whole state of the art 3D review in chapter 5 are introduced the future possible scenarios, divided in broadcaster and interactive. The requirements of those scenarios are reported, especially as regards bandwidth constraints, moreover a possible structure of the video stream is shown, illustrating the layered structure, fundamental for backward compatibility and for scalability aspects.

Finally in the last chapter are briefly introduced the next phases of the project that won't be treated in this work.

Chapter 1

Introduction to 3D and MVV

1.1 3D Cinema and Television Revolution

In cinema and television environments there are more and more difficulties in figuring out new ideas or original contents, therefore 3D cinema and 3D television may be seen as the easiest solution for those writers, directors and producers that want reach a large market without having real ideas or contents to sell.

This is not quite true, for someone maybe 3D is only a modern trend, instead behind it there is a great technological innovation; furthermore the future prospectives are to give something completely new compared to traditional 2D experience. 3D isn't toys land, it will be a tough challenge to preview where 3D experience would really offers a relevant and amazing gap for everyday user; if its contents will be sport live events, movies, concert in opera, documentaries or something else more, is all to discover yet.

Finally 3D applications aren't limited only to cinema and television, they can also be extended to medical environment or military simulations or video conference sessions. Other fields of application have constantly to be found and developed.

1.1.1 Stereoscopic 3D

Nowadays commercial 3D is based on stereoscopic techniques, a left and a right images are displayed on the screen then using special glasses each viewer's eye receives only one image, finally the brain merges the two received images giving the 3D perception. In further sections will be reviewed different techniques to achieve this effect.

3D should be best called as 2.5D or pseudo 3D, because 3D isn't truly reproduced, but our brain creates only a perception of it.

3D cinema and 3D-television systems have been investigated for decades, the firsts attempts are quite old as conventional 2D cinema systems.

Since its invention about 250 films and television programs have been produced in 3D mode.
Technology for creating 3D films has been around for a long time, otherwise technology to view them has a different story.

History of 3D can be divided in 5 great eras:

1890-1946

In these period a few films, with small budget, are shot to explore the secrets of stereo production. In 1890 William Friese-Grene designed the first movie process, it comprised two films projected side by side on screen, then the viewer had to use a stereoscope to converge the two images. From this first approach a number of other attempts have been made in the years, in 1915 the first anaglyph method was developed, by Edwin S. Porter and William E. Waddell, using red-green filtering.

1950-1960

In 1952 began the first "golden era" of 3D, with the release of the first color stereoscopic movie "Bwana Devil" by Arch Obler. The film had been shoot in natural vision, it premises the projection of a dual strip in Polaroid filters. In 1953 Columbia and Warner Bros. released the first 3D movies with stereophonic sound, respectively "Man in the dark" and "House of Wax". One of the most famous film of these years is "Robot Monster", that has been written in almost an hour, and filmed in two weeks.

In these years a number of major studios, as Walt Disney, 20th Century Fox, Universal International, got involved in producing 3D movies and more than 60 films, including "Hondo" with john Wayne and "Dial M for murder" by Alfred Hitchcock, have been released.

After a few years began 3D decline due to:

- High complexity equipments needed (two projectors synchronized, polarized glasses, silver screens and special lenses)

- Poor viewing conditions in most theaters (Screen's silver protection made the vision very directional, and sideline seats were unusable)

1973 to 1985: The Renaissance

In the 80s 3D came back and a number of movies had been released, as "Jaws 3D" , "Coming at Ya!" and "Friday the 13th - Part 3". Nevertheless the cardboard glasses weren't able to resurrect 3D and it disappeared once again.

1986 to 2000: The Revolution

3D cinema came into its own thanks to invention of IMAX 3D format and new screening technology; 3D still had too high shooting costs therefore was used only for specialized productions.

2001 to today: The second golden age

The real grown in 3D field is due to computer animation technology, digital cameras and 3D home theater.

The first product of these new era is "Ghosts of the Abyssis" by James Cameron dated 2003, a documentary film, made with new HD video cameras, not more using films.

From this moment interest in 3D is growing, a lot of movies are been produced and more and more people is approaching it. Content producer, equipment providers and distributors approach to 3D as a new business opportunity.

The market of 3D cinema is expected to continue growing over the next years, since more and more cinema are equipping for it and more and more 3D contents are been made.

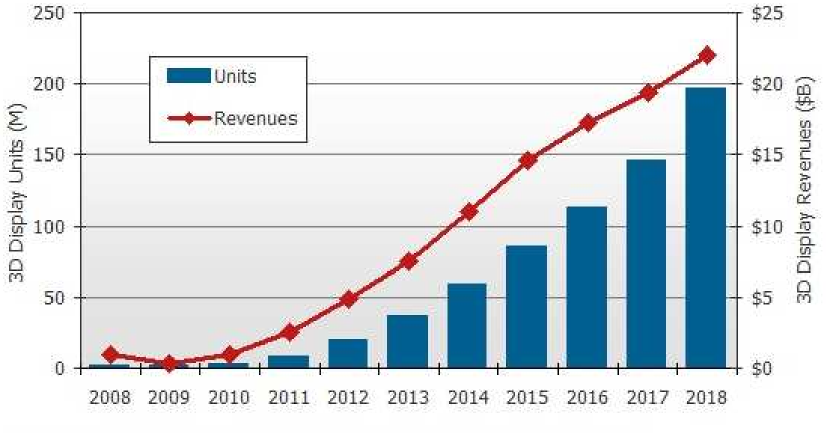

Figure 1.1: 3D forecast displays

Source: 3D Display Technology and Market Forecast Report

3D home environments go hand to hand with 3D cinema, as its success will grown as request for home entertainment is going to increase.

Nowadays there are a lot of ways for the contents to reach home, 3D-DVD/blueray, Internet, 3DTV broadcast, there had also been great progress in home display and there are different technologies available, from the conventional stereoscopic display to auto stereoscopic multi view devices, able to manage more views at a time, to volumetric displays able to create hologram but not yet ready for commercial use.

These improvements, together with Hollywood's focus on 3D content, have stimulated R&D and standardization in this area.

The great success, that 3D has reached in cinemas, with a lot of success titles till the last great project by James Cameron "Avatar" (today the movie with best box office in the world of all time), has also hit television contents producers and broadcasters. Relevant companies are more and more interested in 3D, Sony will produce, in 2010, 85 live football matches, 25 from the next world cup in South Africa. Recently, in partnership with Sony, British sky broadcaster has offered, as a test, the premiership football match Arsenal-Manchester and, as a greater event, the inaugural six nations rugby match England-Wales that got a lot of success among lucky viewers. Moreover other six nations matches will be offered in the next weeks.

Whereas these events have been exclusive for cinemas, the BSkyB and other broadcasters as ESPN idea is to create real 3D television channel providing live sport events, live concert and movies, obviously all in 3D mode.

Besides to BSkyB and Espn also other broadcaster are moving towards a 3D television channel, as japan Direct-TV and japan Sky, or Russian channels.

Finally blue/ray consortium, that is pushing hard on 3D and has licensed a patent for 3D format, will be soon ready to provide a lot of 3D contents.

An other interesting question is "Where is 3D going?" looking both technologies and possible fields of application.

As a matter of fact if 3D would be exploited only as a new amazing trend it will die again, as already happened in the past.

1.2 Multi View Video

It's expected that conventional stereoscopic 3D, nowadays common in 3D cinema, hardly would spread home environment. The greatest constraint is the need of wear glasses, this is considered acceptable in a cinema where it is seen as an event and it is necessary only for a limited time. In a home environment, a more comfortable scenario is requested by users. For this reason displays that wouldn't need wear glasses, are expected to be the best candidate for home usage, this devices are the auto - stereoscopic displays, which can support more than two views, bringing to till 9 views displayed at a time on the screen in a column interleaved spatial multiplex, reaching 180 degrees parallax. Display more then two views at a time is called multi video view (MVV), and leads to different scenarios for 3D future. From each different points of view, each user can see a slightly translated image; in a first approximation, the effect would be that moving his head, the viewer can see behind objects on the scene, giving perception to be much more inside the scene himself.

These devices exploit optical principles as diffraction, refraction and reflection to steer directly towards the user's eyes the images.

At a first sight, transmission of 9 views at a time requests too much bandwidth, instead this isn't the most relevant constraint, as a matter of fact using DIBR (Depth image based rendering) techniques 9 views can be obtained starting from only 3 views shoot. The intermediate views are generated through interpolation of the starting images, that must be in video plus depth format. The mechanism is briefly described in figure 1.3.

In later sections all this architecture will be deeply described.

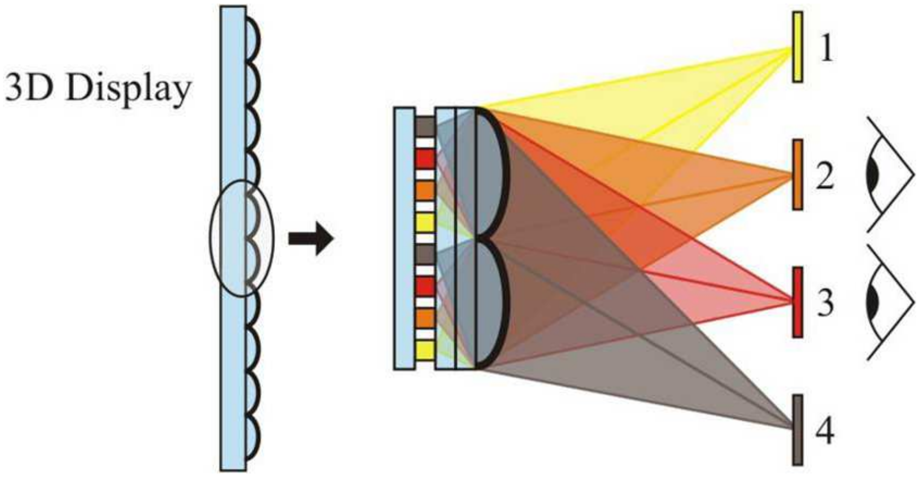

Figure 1.2: Multi view video principle

1.3 Future of 3D

Till now MVV is still a futuristic format, auto - stereoscopic displays are available, but only as prototypes. Moreover the shooting techniques are different for stereoscopic 3D and for MVV, therefore different contents are needed and there is still a lot to do.

Nowadays the most effort regards common stereoscopic 3D.

MVV is expected to reach home in 15-20 years and till now only theorical scenarios can be drawn.

Other prospectives exist for 3D, mainly display technologies will drive it, light - field and volumetric ones will give more and more immersive experiences.

Especially volumetric devices will completely change 3D ideas, bringing holography at end user home.

1.4 3D@Sat project

The 3D@Sat is a project announced by ESA (European Space Agency), its aim is to pave the way to a completely new 3D television system. The whole chain from the capture phase to the final 3D MVV/FVV signal display had been studied. This investigation relies on satellite 3D MVV/FVV video networks architecture and performances.

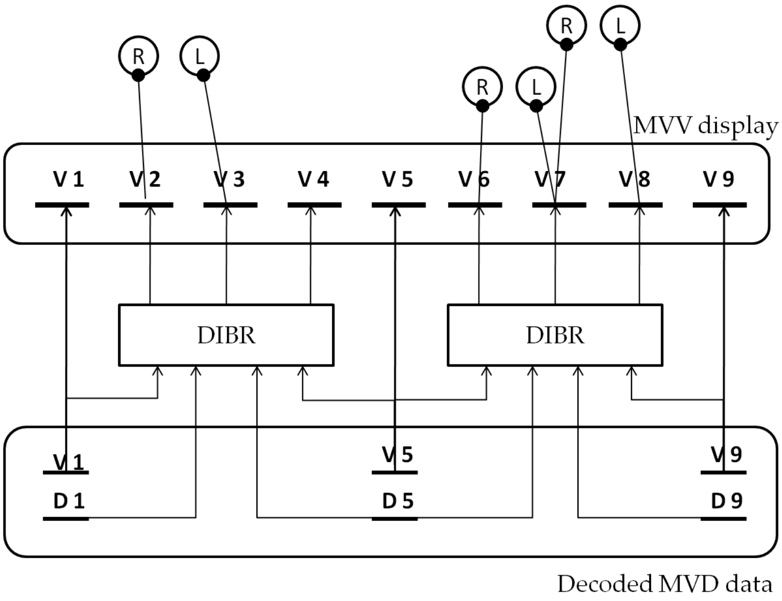

Figure 1.3: DIBR technique

1.4.1 Project aims

Nowadays MVV is still a far away solution, nor at display side, neither at encoding side, technologies are fully developed. The target of the project is to figure out possibles future scenarios for these technologies. The interest in 3D is constantly growing thanks to recent cinema involving in 3D movies, both at user side and content producers one. In the future there will be a lot of contents able to satisfy daily user necessities. For these reasons bring 3D at home is seen as a possible great market. Nowadays stereoscopic displays are next to hit the market and give a first sample of 3D home perception. For these displays it will be tough to exploit the market, as already mentioned, for some simple constraints, mainly the need of wear glasses, not a comfortable accessory, and the still scarce content's production.

A different 3D perception is offered by the multi view video format, that is able to provide till nine 3D views creating a 180 degrees scene. Moreover MVV doesn't require to wear glasses, thanks to auto-stereoscopic displays.

The work plan of the project is shown in figure 1.4.

The first step of the project is to describe the whole chain from capture, through encoding techniques, and satellite transmission, to final visualization. The state of the art of 3D technologies and devices have been analyzed. Different scenarios have

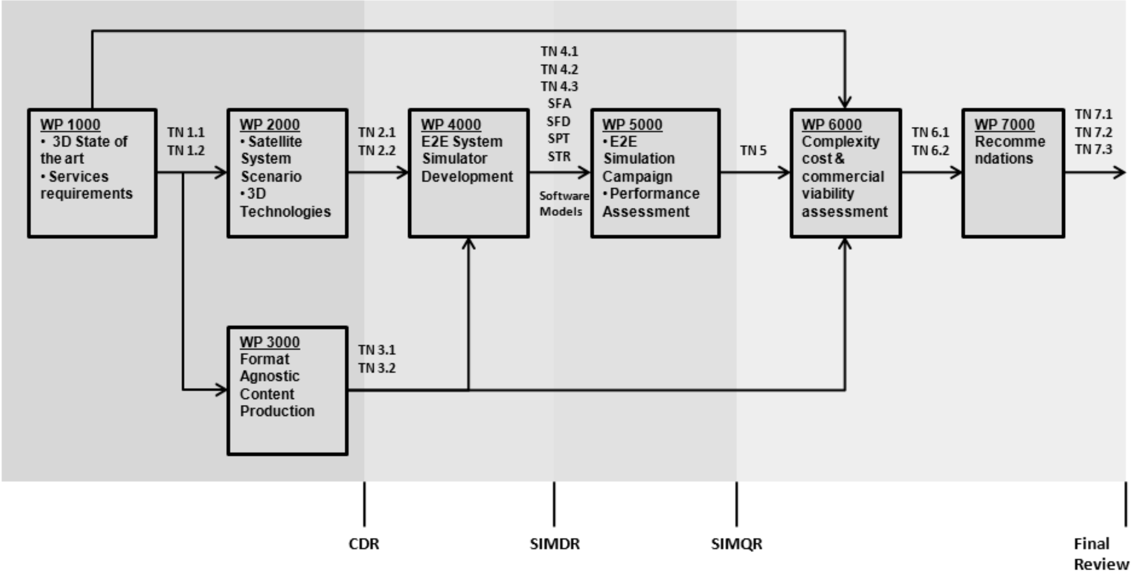

Figure 1.4: Work stages

been established taking in account bandwidth, quality and commercial aspects from one side, and scalability and backward compatibility from other side.

The second step is a simulation campaign. It is divided in two stages, the first is a channel transmission simulation performed by the Astrium Satem tool, provided by Astrium, that simulate the channel at IP level, the second stage is a perceptive test, related to user experience, and involves the display visualization of the stream. Due to lack of MVV displays and decoders technology a MVV real test isn't feasible, therefore it has been adopted a off time test for the MVV visualization on a prototype auto-stereoscopic display by Philips.

As regards the channel simulation the main parameters are the bandwidth needed and the impact of fading on final video quality, instead for perceptive tests the main factor, to be estimated, is the difference in video quality between various representation format as video plus depth or common stereo or video plus depth plus occlusion or others possibilities.

The last stage is a costs/benefits and a SWOT analysis, looking also to future fields, where these technologies may meet success, and to type of contents that might be interesting if seen in MVV format.

Due to impossibility on developing a prototype of the overall architecture, the main aims of the projects were:

- Review of current available devices (display, encoding format, 3D representations...)
- Draw every possible scenarios without looking at nowadays constraints
- Draw possible scenarios that can be the direct derives of currently available technologies
- Analyze advantages, drawbacks and constraint related on MVV respect to 3D through simulation campaign and tests
- Analyze fields of applications and outline the way for future MVV spread

ESA

The promoter of 3D@Sat project is ESA, European Space Agency. ESA is an international organization with 18 member states and is Europe gateway to space. Its mission is to shape the development of Europe's space capability and ensure that investment in space continues to deliver benefits to the citizens of Europe and the world. ESA aims to find out more about Earth, its immediate space environment, our Solar System and the Universe, as well as to develop satellite-based technologies and services, and to promote European industries. ESA also cooperate with space organization outside Europe.

The companies involved in the 3D@Sat project are Frauhnofer Heinz-Harald Institute, Astrium and Open Sky.

Heinz-Harald Institute

Fraunhofer - Gesellschaft was founded in the 1949, in the large conference hall of Bavarian Ministry of the Economy. The aim was to develop new structure after the war's destruction and to spur reconstruction of the economy. It undertakes applied research of direct utility to private and public enterprise and of wide benefit to society. The hearth of research and development work carried out by HHI are innovations for the digital future as regards both state of the art communications systems and digital media and services. It is a leading research institute in the fields of Mobile Broadband Communications, Photonic Networks and Systems, and

Electronic Imaging for Multimedia. The Image Processing Department of this institute has a long experience in signal, image, and video processing. Another field of research at HHI is Ultra High Resolution video systems. Main topics are further video and audio-coding / transmission (Video over IP), the 2D and 3D image processing, the image reproduction in virtual environments, tele - immersive systems, auto - stereoscopic displays, mixed reality displays, man-machine-interactions and information and video - retrieval. In particular HHI is chairing the standardization of the payload of SVC in the IETF and responsible for editing the specifications for the video parts of DVB-AVC.

Astrium

Astrium employs 15,000 men and women in five countries: France, Germany, the UK, Spain and the Netherlands. Guaranteeing Europe's access to space as the established leader in space transportation, satellite systems and services, Astrium has for over 40 years been dedicated to discover all about the space. It develops great european space projects as Ariane, the International Space Station, Envisat, Mars Express and Skynet 5. It is a mission with a consistent commitment, to offer customers the best possible solutions in the market, with high levels of, quality, cost-efficiency and schedule adherence.

Astrium is a wholly owned subsidiary of EADS, a global leader in aerospace, defense and related services.

EADS Astrium's recognition of the need to be visionary in the fast-moving world of communications has also influenced its research and development activities in the past 11 years, and significant progress has been made to acquire a complete capability of communication system turnkey supplier. In this frame, EADS Astrium has developed a significant expertise of end-to-end satellite telecommunication system engineering. As part of its missions, the satellite telecommunication system department has been and will be involved in the standardization of DVB-S2, DVB-RCS and DVB-RCS NG.

Open-sky

Open-sky is a commercial operator that provides Satellite Internet access, push services, streaming channels and bi-directional Satellite IP services.
It's specialized on satellite value added services mainly based on the IP broadcast

technology.
Open-sky mainly deals about: tooway that provides an Internet access for consumer with Satellite bidirectional channel available in Italy and Europe. Bidirectional networks with DVB-RCS terminals for business market. Communication services with Bidirectional Networks (voice and videoconferencing) VideoSat, an exclusive platform for Content Delivery.
An innovative set top box combined with the Open-Sky push service VideoSat is used for :

- V4DL project, ESA funded project for CME (Continuous Medical Education)
- Push Television Digital Cinema services
- Delivery and management of LIVE events
- Delivery and management of PUSH contents

Open-Sky leads the digital cinema revolution and the new live 3D at cinema.

DIGITAL-CINEMA

Open-Sky is supporting the digital cinema revolution by providing equipments and running the service. Open-Sky provides a turnkey service to cinemas and content providers.

First commercial services are:

- Live event: delivery of full HD content for live event at cinema
- PUSH content: delivery of movies and any other type of content by means of a PUSH transmission. Storage is done on a local HDD.

Open-Sky exploits its specific experience in multimedia broadcasting collaborating with ESA in the ISIDE project. Based on this solution Open-Sky has developed with some partner companies a first complete end to end technical platform for LIVE 3D events. The platform is made of the following components:

- 3D cameras
- 3D encoder-satellite-decoder

- Integration with 3D cinema projector

Thanks to its experience, as regards cinema environment, Open-Sky is now developing a 3DTV service and product. Specifically Open-Sky is involved with ESA and SKYLOGIC (Eutelsat group) in a ”Stereoscopic Broadcasting” project, which aim is to provide a first test and demo channel with 3D TV service. All these activities represent a base for further studying the long term requirements and solutions for a 3DTV system and related standards.

Chapter 2

Human visual system and 3D creation phase

In a 3D television system many phases are required between the capture and the final visualization on the display.

The whole chain is presented in fig. 2.1:

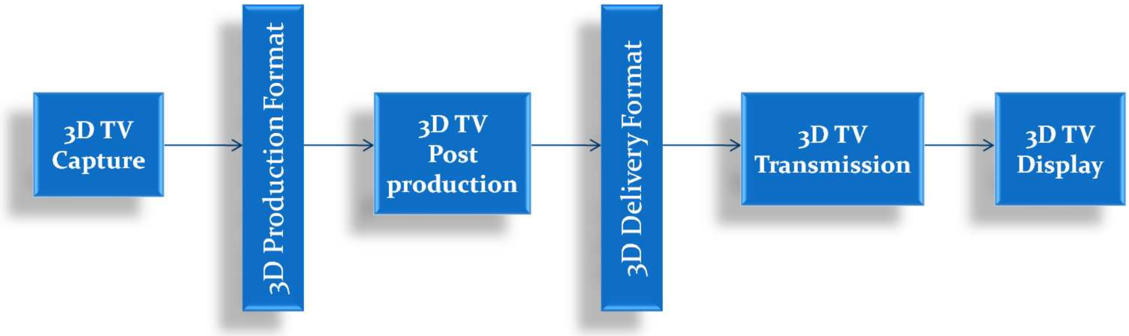

Figure 2.1: 3D TV processing chain

The first stage in a 3D television system is the capture one, which is much different from the corresponding 2D procedure, since a stereo representation needs a double image and therefore two synchronized cameras. After the capture there is the 3D production format, that refers to parameters of real camera, capturing devices and describes the interface between the capture and the post production. The post production phase hasn't an approach too far from the 3D one and is a section of the film making process, it mainly consists in adding soundtracks, or special effects or editing picture.

The 3D delivery format is between post production and transmission and describes a simplified virtual camera set up.

The 3D transmission refers to the physical data transmission to the user. Finally there is the 3D display, that comprises decoding and final visualization of the video stream.

The aim of this chapter is to analyze the stages from the capture to the 3D delivery format, the following parts of the end to end chain will be discussed in later chapters.

2.1 3D capture phase

The easiest way to reproduce 3D perception is the stereo approach, that emulates the human visual system. The capture of a 3D video stream is a very complex and interesting phase, two different cameras have to be set up for each view, and the mark is on the joint configuration, the arrangement and the calibration, further than on each camera configuration. Moreover MVV services require a multiple camera shooting phase that further increases setup complexity.

Firstly will be described the human visual system to understand which process have to be emulate and to perceive its main constraints and troubles. Then for the same reason, will be analyzed the basis of stereo perception.

Finally the real 3D capture will be analyzed looking at the geometry of stereo capturing, the distortions introduced with this approach and the generic rules that are applied to overcome them.

2.1.1 The human visual system

The human visual system (HVS) is able to perceive depth, interpreting several depth cues. Those cues can be divided in two main categories: monocular ones that brings information obtained with a single eye as relative size, linear perspective or motion parallax), binocular ones that need both eyes to get those type of information, as retinal disparity, stereopsis or convergence. In addition to this two categories there is a third one: the inferred cues that refer to depth information, regarding those regions that aren't within double vision coverage, estimated by brain. These last cues are due to front position of eyes, that allows a great definition of the image for a short angle range, therefore the regions outside the viewing angle are partially inferred by brain, instead other animals (as plant eating ones), that have eyes more far each other, have a less definite image but for a greater angle range.

Moreover how the cues have to be interpreted is mainly a knowledge derived from everyday experience. In each scene there are a lot of depth cues, and the whole of them brings brain to the depth perception.

Binocular cues most influence depth perception only for distances below ten meters, therefore they have to be take in serious account for 3DTV systems.

Cues come from horizontal eyes separation, the interocular distance is the distance between left and right eye, and it is around 64mm on an average. Each eye leads to a specific perspective view of the scene, the same point of the observed scene is projected in different position on the left and right retina, thus providing retinal disparities.

Brain deduces from these disparities the relative distances between objects in the scene and the spatial structure of it, in this way it's able to merge the two images, thus creating the three dimension perception.

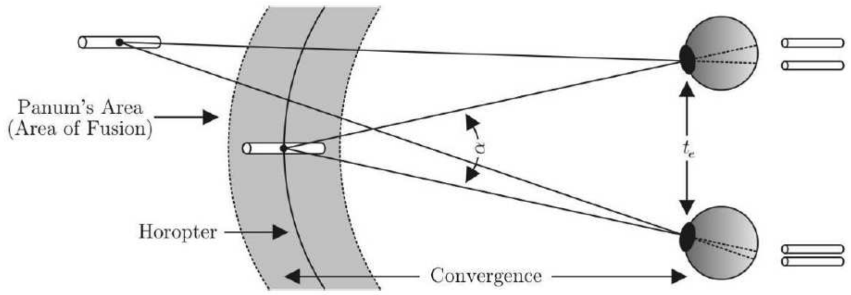

Figure 2.2: Principle of stereoscopic fusion and retinal disparity

Figure 2.2 shows the mechanism of image fusion. Observing a scene, eyes rotate till their optical axes converge on the same point in the scene, when this point is in the horopter area, no retinal disparity is created, because they are projected on the same point in both retinas. Instead, when the intersection is outside the horopter zone (also called Vieth-Muller circle and defined by nodal and convergence points of both eyes), retinal disparities are created, thus giving depth information of the observed environment.

Disparities, that don't exceed a certain magnitude and land on the Panum's fusional area (the region around horopter), can be fused in a three dimensional image; instead points outside this area can't be fused and give rise to the phenomenon of diplopia, leading to a halve image perception.

Disparities in front of the horopter are said to be crossed, vice versa those behind the horpoter are called uncrossed.

Accommodation and convergence pander to users, avoiding the perception of those double images. Accordingly to the rotation of the optical axes, eyes, changing the shape of eyes' lenses, focus on the object of interest, thus providing a sharp and clear perception of it, furthermore regions outside the Panum's area are automatically blurred, hiding the double image effect.

2.1.2 Geometrical basis of stereo perception

Almost all displays and projectors are called plano-stereoscopic devices, because they are based on the same basic principle, left and right images are reproduced on the same planar screen.

Therefore the perception of the binocular depth cues comes from the spatial distances in the screen between corresponding left and right image points. This distance is called parallax P and leads to retinal disparities.

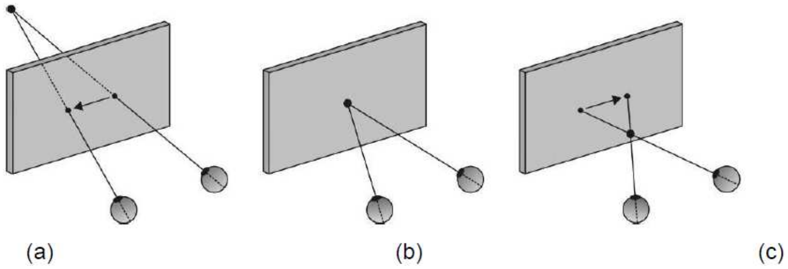

Figure 2.3: Different types of parallax: positive, zero and negative

As depicted in fig 2.1.2, three cases are possible:

- Zero Parallax: left and right image points lie on the same point on the screen. The resulting 3D point is perceived on the screen, this situation is referred as *zero parallax setting (ZPS).*

- Positive Parallax: also referred to as uncrossed parallax, it happens when the right image point is placed on the screen more right then the relative left image point. In this case the points is perceived behind the screen, in the so called *screen space*.

Furthermore, if the points' distance is equal to interocular distance (t_e),the point is projected at infinity, so this represents also the maximum limit for positive parallax.

- Negative Parallax: also referred to as negative parallax, it happens when left image point is placed on the screen more right of the correspondent right image point. The 3D point is perceived in front of the screen in the so-called *viewer space*.

Horizontal amount and type parallax aren't the only factors that influence the binocular cues perception, also other parameters as viewing distance contributes to it.

The distance at which the object is perceived (Z_v) can be easily estimated by:

$$Z_v = \frac{Z_D \cdot t_e}{t_e - P} \tag{2.1}$$

In the ZPS condition (P=0), the object is seen directly on the screen (Z_D refers to the screen distance from the viewer), $Z_v = Z_D$.

Furthermore in positive parallax condition, that is if P ¿ 0, $Z_v > Z_D$, the object is perceived behind the screen, otherwise in negative parallax condition, that is if $P < 0$, $Z_v < Z_D$, the object is perceived in the viewer's space.

Another relevant parameter is the maximum parallax range ΔP_{rel}, that is the maximum parallax allowed within which the left and right images are still fused. It relies on screen distance Z_D, screen width W_D and maximum parallax angle $\Delta\alpha_{max}$:

$$\Delta P_{rel} = \frac{Z_D}{W_D} \cdot \Delta\alpha_{max} \tag{2.2}$$

Parallax is unlimited in real world scenes, instead in stereo reproduction it has to, because eyes accommodate at the screen surface; simultaneously they converge depending on horizontal parallax, thus producing a conflict between accommodation and convergence that is the main cause for eye-strain, confusion and loss of stereopsis. For this reason $\Delta\alpha$ must be kept within certain limits, that means that 3D world scene must be reproduced close to the screen surface as depicted in fig 2.1.2.

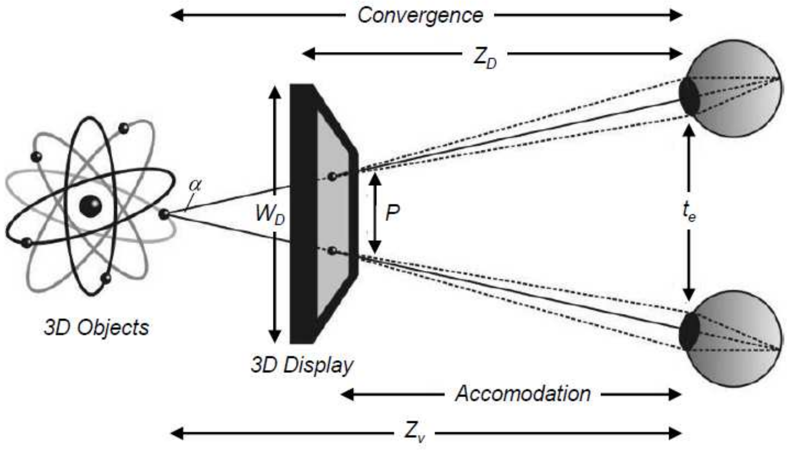

Figure 2.4: Conflict between accommodation and convergence

This consideration leads to the definition of $\Delta\alpha_{max}$, in literature various choices can be found. A common choice is to set the maximum parallax angle equal to 70 arc minutes, this yields to a common reference for stereo reproduction parallax of $\frac{1}{30}$, choosing a medium width display and a common viewing distance. ΔP_{rel} is an important dimensionless parameter that describes, in a first approximation, the whole 3D processing chain. As a matter of fact, changing display size or viewer's distance, the ratio $\frac{Z_D}{W_D}$ move in a range from 1 to 4, considering home theater, mobile 3D devices, flat screen displays or large projectors. Maximum parallax range is comprised between $\frac{1}{50}$ and $\frac{1}{12}$.

2.1.3 Geometry of stereo capturing

The stereo reproduction format requires that two different views have to be captured simultaneously by a stereo camera. They have to reproduce the same depth cues that human eyes would have perceived seeing in real time the scene, therefore the structure of the camera setup is strictly similar to the HVS. The inter-axial distance of the two camera lenses is the equivalent of the inter-ocular eyes distance. Furthermore camera has to continuously converge on different planes to choose which part of the scene has to be reproduced on the screen in the display phase. Convergence configures as a fundamental aspect and leads to two possible configuration:

- toed-in setup: a convergence point is chosen by a joint inward rotation of the two cameras
- parallel setup: a convergence plane is fixed by a shift of the sensor targets.

This two approaches are depicted in fig 2.5.

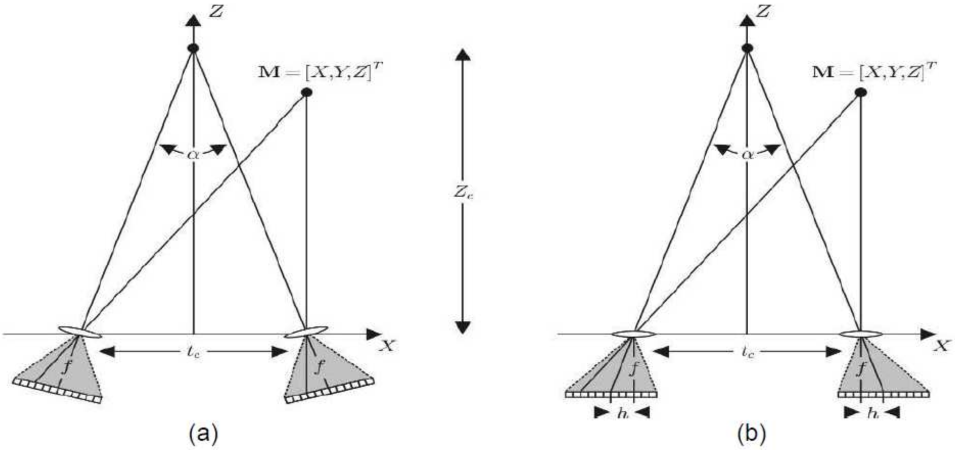

Figure 2.5: Camera setup: a) toed-in setup b) parallel setup

At a first sight the toed-in setup seems to be the best solution because it best fits at the HVS. Instead the parallel setup is preferable, the reason is that is most relevant the two images to lay on the same planar screen then the eyes focus on a specific part of the scene. In the parallel configuration this is achieved, thus providing a higher final stereoscopic image quality.

Moreover the toed-in setup provides a distortion between image planes on the display and image planes of capturing camera. The distortion produces not only horizontal parallax, that is the base for the 3D perception, but also vertical parallax that is one of the most relevant causes for eye-strain and so it is undesired. Otherwise the parallel approach only provides horizontal parallax that is requested for the stereoscopic effect.

Finally the toed in setup can be easily converted in a parallel configuration through a rectification process.

Parallel approach has itself some constraints, the most issue refers to camera convergence. As a matter of fact the optical axes of the camera intersect at infinity, this trouble is solved by the sensor shift of the cameras, they shift horizontally in inverse directions by the same distance, thus the optical rays intersect in the desired convergence point of the image.

Disparities for the parallel configuration belong to camera focal length (F), interaxial distance (t_c), the distance of the convergence plane from the camera basis (Z_{conv}) and to the position of the point (Z) and come out from the relation:

$$d = t_c \cdot F \cdot (\frac{1}{Z_{conv}} + \frac{1}{Z}) = h - t_c \cdot \frac{F}{Z} \tag{2.3}$$

Equation 2.3 shows an analogy with equation 2.1, that referred to HVS, as a matter of fact if $Z = Z_{conv}$, that means the object is on the convergence plane, that is ZPS condition, there is no disparity (d=0). Instead in the case of $Z > Z_{conv}$, that means the object is behind the convergence plane, it leads to positive disparity, finally if $Z < Z_{conv}$ the point is in front of the convergence plane and there is negative disparity.

In previous sections it has be seen that at display side, all scene must be reproduced around the screen surface to allow the two images fusion. Furthermore for the best perceiving quality is rather preferred the overall scene to be behind the screen surface. For this reason the parameter h, that refers to the sensor shift, becomes fundamental. As an example if there isn't sensor shift (h=0), the convergence plane goes to infinity and the whole scene appears in front of the scene, producing great distortion effects. It's clear that the sensor shift has to be used to adjust how the 3D scene is distributed around the display surface.

The last parameter that has to be analyzed is the inter axial camera distance t_c, it is close to the inter ocular distance presented in the previous sections, but some more remarks are necessary. Usually t_c and t_e have not the same value, inter axial distance belongs to the depth structure of the scene and to the ratio between sensor width and focal length.

Specifically t_c follows the equation 2.4.

$$t_c = \frac{\Delta d}{F \cdot (1/Z_{near} - Z_{far})} = \frac{W_s \cdot \Delta P_{rel}}{F \cdot (1/Z_{near} - Z_{far})} \tag{2.4}$$

In the equation Z_{near} and Z_{far} symbolize respectively near and far clipping plane.

Moreover ΔP_{rel} must not exceed the display, for this reason t_c have to be chosen properly.

2.1.4 Stereo distortions

Display parallax is obtained by the disparity multiplied by a parameter S_M ($P = S_M d$).

Stereo depth reconstruction isn't linear, hence produces distortions, that mainly belongs on this parameter (S_M) and the sensor shift.

From this relation and from equations 2.1 and 2.3 the relation between perceived depth Z_v and real object depth Z becomes:

$$Z_v = \frac{Z_D \cdot t_e}{t_e - P} = \frac{Z_D \cdot t_e \cdot Z}{S_M F t_c - Z(S_M \cdot h - t_e)} \tag{2.5}$$

From the equation yields that depth reproduction is linear only when parallax is equal to interocular distance. Otherwise stereoscopic distortions come about in two possible ways, the foreground objects are more elongated then background ones if ($S_M h > t_e$) or vice versa.

It's clear that right linear depth reproduction is achieved only if point at infinity are placed properly in the parallel viewing axes.

The sensor shift and the convergence plane distance, since $Z_{conv} = \frac{t_c F}{h}$, play a key role and setting them properly distortions can be avoided
($h = \frac{t_e}{S_M} = \frac{W_S \cdot t_e}{W_D}$), but even if sensor width is known during the capture phase, this is not for the display width that can only be supposed. Therefore the ideal case can be achieved only for the targeted display.

Finally Z_{conv} and the scene setup must coincide with the selected parameters as regards the sensors shift h, the interaxial distance t_c and the camera focal length F.

2.1.5 Conventional stereo reproduction rules

Some rules have to be respected to achieve stereoscopic viewing comfort. These issues don't belong to scene structure or final display or viewing conditions but only refers to some technical aspects of the capturing devices.

- **Temporal Synchronization**: The temporal synchronization between left and right cameras has to be very precise to obtain an acceptable 3D quality. These errors especially for objects in motion induce horizontal and vertical alignments errors that lower comfort viewing creating fault disparities.

- **Colorimetric symmetry**: Differences in the two views as regards luminance, color and contrast have to be avoided. These asymmetries may cause binocular rivalry and, then, lower comfort viewing. The most constraint is given by contrast differences, while color and luminance variations are more tolerated.

  Some colorimetric adjustments can be done on the fly during the capture, or during post production phase with a grading process.

  Impact of these asymmetries isn't fully investigated yet, but as a common concern: the two views have to be identical, so far as if displayed in side by side format, no difference could be noticed.

- **Geometric Symmetries**: Errors in image height, size or scale, or non linear distortions between left and right images can occur due to differences in focal lengths or skewed camera sensors or optical lenses with notable radial distortions. When these asymmetries overcome a certain threshold they lead to visual strain. This effect occurs especially in the toed-in configuration, that for this reason requires a rectification process. Also parallel setup suffers these troubles, in case of a non perfect parallel camera axes setup.

  Geometric distortions can be corrected manually during post production phase or automatically, using calibration data, directly during the capture phase.

Other issues rely on display properties and viewing conditions and are partially correlated with knowledge of the content produced.

As farther seen before, the sensor shift h and the inter axial distance t_c are the most relevant parameters, a good choice for them can be made taking in account desired ZPS condition and allowed parallax range ΔP_{rel}.

- **Inter axial distance camera**: As described in equation 2.4, interaxial distance have to be chosen according to a target display, to achieve a certain parallax range, that, in turn, depends on the maximum parallax angle and on viewer conditions, but in last analysis also, and in a strong measure, by display technology.

Almost all current stereoscopic reproduction techniques lead to a certain amount of crosstalk, that, in turn, produces ghosting effects and can induce undesired headache. Therefore the upper parallax limit for a display can be much lower then the perceptual limit provided by equation 2.2, especially in the case of auto-stereoscopic displays.

Furthermore t_c always relies on scene structure and it is used to be lower then inter ocular eyes distances t_e. Usually it is in a range from 30mm to 70 mm, except for particular cases as longshots or macros.

- **Sensor Shifts and Convergence Plane**: The sensor shift (h) is the best parameter to minimize 3D distortions and to control the distribution of the scene on viewer and screen space. Especially in the case of the parallel setup, in which the shift is independent from t_c.

  As already seen if $h = t_e/S_M$ then 3D distortions are minimal, to estimate the shift some assumptions on display width and on its ratio with the sensor width are needed.

  In addition the content itself defines some constraints on the h value. As an example, scene, reproduced in the viewer space, mustn't be cut on the borders. The drawback provided is a conflict on the two depth cues, because binocular cues suggest the object to be in front of the screen, instead monocular views produce stereo framing inducing that the object is behind the screen, because it is obscured by the screen surround.

  This situation has to be managed, this can be done in the post production phase, as a matter of fact modifying opportunely the sensor shift, the whole scene can be shifted behind the screen.

  Finally h choice derives from a trade-off between distortions reduction and contents constraints, to avoid stereo framing and keep content close to the screen surface.

2.2 3D Production Format

The 3D production format interfaces the capture phase with the post production one. It precisely refers to the real camera and other capturing devices. It summarizes parameters, used in the capture phase, that have to be known during next post production phase.

For further considerations a specific camera is chosen with defined rectilinear lenses. Other parameters are defined through a calibration procedure.

As an useful shrewdness in case of image transformation (rectification or distortions process) a new camera description should be added, comprising updated camera's parameters.

Camera parameters can be divided in three groups:

- **Geometrical parameters**: they describe the 3D points projection in the image and are divided in extrinsic that refers to camera position and orientation and in intrinsic that refers to fixed camera property as focal length and radial distribution.

- **Geometric uncertain parameters**: they describe expectations or estimations of those undefined geometrical parameters.

- **Sensor related parameters**: they describe some specific properties related to the sensor and to the capture settings, as exposition time, sensor information, iris, focus distance, capture time and color space used.

Depending on the processing purpose only one of the previous groups can be used in this phase.

For 3D reconstruction only geometrical parameters are needed, in addition can be exploited also information regarding the uncertainty.

Whereas for image processing, as multi camera color equalization procedures, only sensor related parameters are needed.

2.3 3DTV post production

The post production phase is between the production and the delivery formats. In this phase all image processes are performed, to remove any colorimetric or geometric asymmetries.

It doesn't present too much differences from the respective 2D procedure. The most relevant aspects that specifically rely only on 3D post procedures are related, as seen in previous sections, to sensor shift (h) and inter axial distance (t_c).

Image distortions and shift in the depth dimension can be done starting from them, as an example, especially in a stereo format, adjustments of the two images, that have to be identical (as concern luminance, color and contrast), have to be done, or the whole scene must be shifted to be all inside the stereo window, that is the sum of screen and viewer area.

Furthermore there are also other aspects of these phase that are identical to the 2D one, these regards adding soundtracks, or special effects and also other picture editing.

2.4 3DTV delivery format

The delivery format interfaces post production phase with transmission one and can be seen as an extension of production format, whereas it represents a simplified camera setup with in addition occlusion information and a more simple camera geometry description.

Without any information regarding display and viewer configuration, that may change from each user system to another, only an idealized, simple and generic format can be produced with relative constraints on 3D display adaptation and future 3D services functionalities.

Through a rectification process the toed-in approach can be traduced in a parallel setup, therefore delivery format can be estimated only for the parallel configuration, taking in account the possibility to extend the double view system to multi view functionalities, thus supporting various multi-view video representation formats as MVD, LDV and DES (they will be described in farther chapters).

The framework of the stream relies on the number of views transmitted N that isn't fixed but is defined each time by production process and provider's preferences. Each streams is composed by a video stream with a per-pixel depth map, eventually there can be another layer for occlusion information with a related depth map.

The structure is presented in fig 2.4, where in addition the camera line position, also referred as base line (BL), the inter axial distance (t_c), the near and far clipping plane (Z_{near} and Z_{far}) and the depth plane (Z_{conv}) where there is ZPS (zero parallax settings) are indicate.

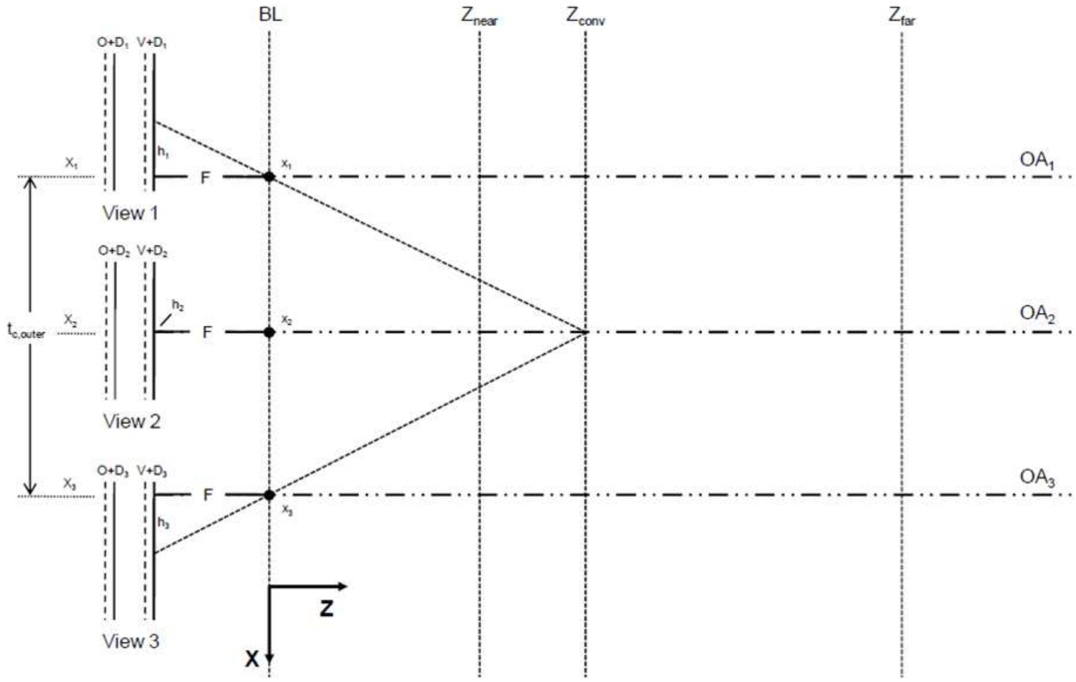

Figure 2.6: Framework of 3DTV delivery format for a 3 views stream

It is expected all views to have the same focal lengths and all lens distortions to have been corrected in the post production phase.

Final depth point values come out from a relation between Z_{near}, Z_{far} and Z_{conv}.

$$D_i(u, v) = 255 \frac{1/Z_i(u, v) - 1/Z_{far}}{1/Z_{near} - 1/Z_{far}} \tag{2.6}$$

Where the coefficient 255 is due to the gray scale structure of the depth map. Z_i is the depth position of the point in the real scene, therefore the zero disparity plane distance is obtained replacing it with Z_{conv}.

As already seen, from Z_{conv} it is also possible to obtain the sensor shift h:

$$h_i = t_{c,outer} \cdot x_i \cdot F/Z_{conv} \tag{2.7}$$

In the case of a toed-in approach the shift h is directly provided by the rectification process, otherwise in the parallel setup h is manually added during the post producing phase through shifting, cropping and scaling images.

Another relevant parameter necessary for the interpretation of the depth information is the relative disparity range Δd_{rel} between the outer cameras, thus it leads to maximum parallax range:

$$\Delta d_{rel} = \frac{\Delta d}{W_s} = \Delta P_{rel} \tag{2.8}$$

2.5 Shooting issues

The most intuitive constraints relies on the increase complexity and duration of the cameras setup, due to geometric rules and restrictions. Moreover as the number of camera increases the setup complexity grows itself. Formats including depth and occlusion maps, lead to error prone and very complex procedures and algorithms, especially for those setups that require more then two cameras for the acquisition. There are tools that make capture setup faster, as stereo image analyzer, or rigs mechanics that provide motorization and remote correction, but these procedures are still very complex and more research and studies are needed.

Moreover software able to check stereoscopic rules violations can help in the setting up procedures.

Some generic assumptions must be respected in any setup, for narrow shooting areas the horizontal rig size mustn't exceed twice the single camera width, this constraint is a big issue that creates problems for long shot scene.

Costs on 3D shooting side are increased respect a 2D shooting setup, in a first analysis they are doubled, because are required the double number of cameras and operators, but in a deep analysis other relevant figures become necessary as stereographers, and more equipments as stereo image analyzer that increase the costs.

Finally in a 3D setup and moreover in a MVV system costs and complexity increases in the setup phase, these must be limited and research is needed.

Chapter 3

3D Video Formats

A 3D video system fundamental aspect refers to 3D video formats that defines from one side the capture procedures and on other side how final video stream is encoded and transmitted.

Many approaches are available and under investigation, from the common and more simple stereoscopic one to video plus depth or video plus depth and plus occlusion one.

All these rely on different algorithms and have different advantages and constraints, as concern complexity, efficiency and functionality.

The analyzed system relies on satellite communication, therefore a relevant consideration concerns the constraint in bandwidth required for the transmission that greatly impacts on the costs of the whole chain.

At first sight in a multi view configuration, if each view is treated independently, the needed bandwidth increases linearly with the number of the views. Using a multi view coding technique, as MVC, that exploits interview correlation, bandwidth needed can be reduced, usually the gain is around 20%, but it is strictly related to the contents, furthermore the bandwidth still remains pretty high.

Video plus depth is another approach, depth data can be compressed very efficiently. Recently MPEG developed a corresponding standard MPEG-C Part 3, which allows the possibility to obtain the stereo effect starting from a single-eye view, through the depth image based rendering (DIBR). It obtains the same quality video as common stereo format, with much less bit-rate. It presents some issues with parts of the scene that are covered in the original view and can't be properly reproduced in the virtual view; to solve this problems occlusion maps are used, they get information about the covered parts of the scene, therefore the virtual view are closer to the real scene captured.

In the sequent sections all currently 3D video formats are more deeply presented.

3.1 Conventional Stereo Video CSV

The most simple approach is the simulcast, left and right image are encoded and sent independently, no depth information is required.

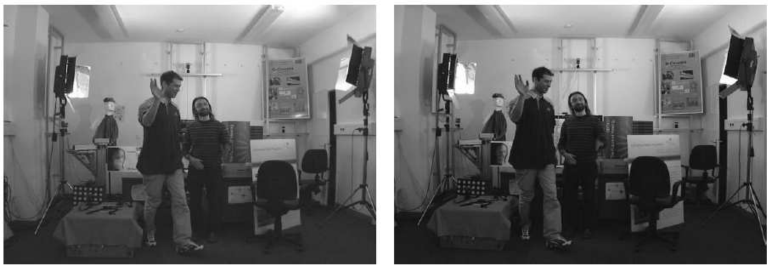

Figure 3.1: Conventional stereo video

After capture phase, that is made by two separate cameras, there may be some processing phases as rectification, normalization or color correction. This format has the advantage of low complexity, full spatial resolution, computation and processing delay reduced to minimum and backward compatibility to conventional 2D devices.

On an other hands the drawback is that inter view correlation isn't exploited, therefore the code efficiency is low.

3.2 Asymmetrical Stereo Video

Some studies on binocular suppression theory demonstrate that if one of the two views is low passed, the brain is able to fill missing information, so the perceived overall image quality is dominated by the higher quality image, and it is comparable to the case where both images are not low-passed.

Asymmetrical stereo video exploits this brain capabilities obtaining a little gain in bandwidth sending the second view in a lower resolution or quantizing it more coarsely. Different mixed solutions can be used, keeping the left view at full resolution, and down-sampling the right one to half or quarter resolution; in this way, a 3D stream can be obtained with an additive amount of 25-30% bitrate respect to an usual 2D-HD stream.

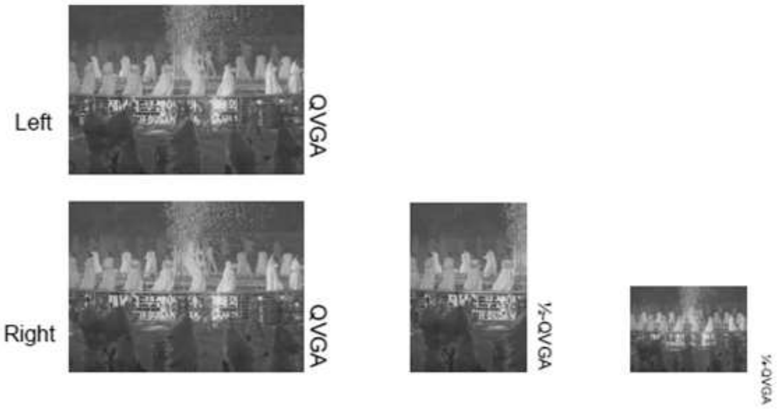

Figure 3.2: Asymmetrical stereo video

Further studies as regards how this phenomenon extends to multi view video have to be carried out.

3.3 Interleave Stereo Format

Another approach is to use one of the existent interleave video format, it can be a time multiplexing one as a frame interleaving, or a spatial multiplexing one as side by side or over under field as showed in figure 3.3.

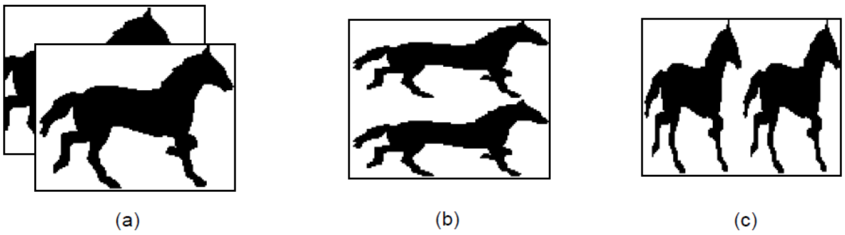

Figure 3.3: Interleave stereo formats

These approaches reduce by half the bit-rate required, anyway there is a further signaling information to notify the decoder of the use of interleaving.

Therefore backward compatibility to 2D devices isn't possible.

Moreover spatial multiplexing formats reduce by half the spatial resolution, and images are "squeezed" to fit to the screen. It's needed a small overhead to signal the left and right views. In the time multiplexed formats, can be introduced interleaving prediction, that, in turn, must be signaled.

3.4 Video plus Depth

Another format is 2D video plus depth (V+D), a conventional 2D stream video is encoded with in addition a per pixel depth map. The depth map is a gray scale image, the depth range is quantized with 8 bit, value 255 is associated to the closest point and 0 to the most far. The depth is restricted to a Z_{near} and Z_{far}, that refers to the minimum and maximum distance of the matching 3D point in the scene to the camera.

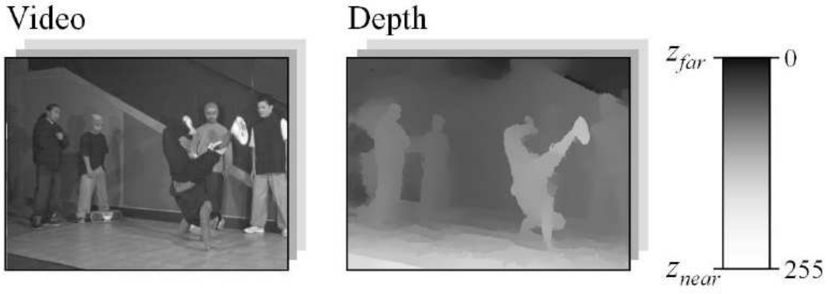

Figure 3.4: Video plus depth format

Depth information can be put in the luminance channel of the video signal, setting the chrominance to a constant value, allowing any state of the art video Codec to easily manage the stream and guarantying backward compatibilities with legacy 2D devices.

Starting from V+D format a stereo pair can be obtained using 3D warping at the decoder.

Moreover this format suits with auto stereoscopic displays that generates more then two views through depth information and allows head motion parallax viewing within practical limits.

Additional depth data can be efficiently compressed, due to its specific statistics, that is less structured and more smoothly than color data. Approximately depth data can be encoded at good quality with addition of 10-20% of the overall bitrate. These ratios, between color and depth weigth, have to be confirmed with latest video codec H.264/AVC.

This representation doesn't need any information regarding coding format, it works both with H264/AVC and MPEG - 2. It only needs a high level syntax that allows the decoder to manage two incoming video streams as video or depth and information regarding Z_{near} and Z_{far} plane. MPEG-C Part 3 implements this format, representing auxiliary video and supplemental information, the main drawbacks regards that it is only capable of rendering a small depth range and it has difficulties in managing occlusions.

3.5 Multi-view Stereo

Multi-view Stereo (MVS) is a straight away extension for conventional stereo video, where more than two cameras are set up along a common baseline using parallel stereo geometry. In a first analysis the views can be encoded independently through CSV.

Best coding efficiency can be achieved exploiting temporal/interview prediction, MPEG-2 has already provided a standard for multi-view video coding, more then ten years ago. In H.264/AVC have recently been added a stereo SEI message to implements prediction as depicted in figure 3.5.

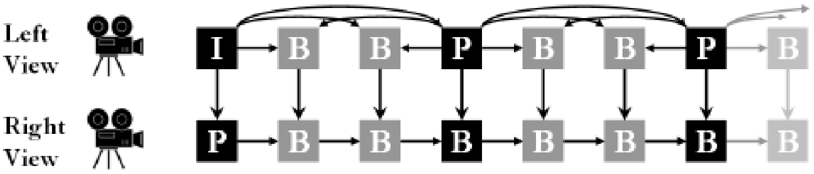

Figure 3.5: Stereo video coding with temporal/interview prediction

If there are more than two views to be sent, the stereo prediction stream can be extended to multi-view video coding (MVC) that is provided since 2008 by MPEG-ITU in an extension of H.264/AVC, nowadays it is the best efficient multi view video coding available and it can support an arbitrary number of views.

3.6 Multiple video plus depth

With the spreading of 3D video applications and the developing of auto stereoscopic multi-view displays, formats that provide a large number of views with an horizontal continuum perception are under research by standardization bodies, like MPEG. Since from one side MVC doesn't provide the rendering of a continuum of output views and the bigger is the view's number the bigger is the bitrate needed, on the other side V+D representation allows only a limited number of views around the available one, due to artifacts that increase greatly as increases the distance between the virtual viewpoint and the camera position.

MVD can be seen as a synthesis of these two formats, providing more V+D streams.

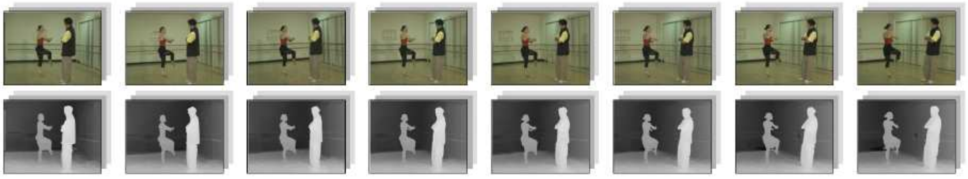

Figure 3.6: MVD scheme

MVD requires a high complex process, the views have to be generated during capture and post-production and their depth maps have to be estimated, then the streams have to be encoded and transmitted. At the receiver side, the streams have to be decoded and the virtual inter view have to be rendered.

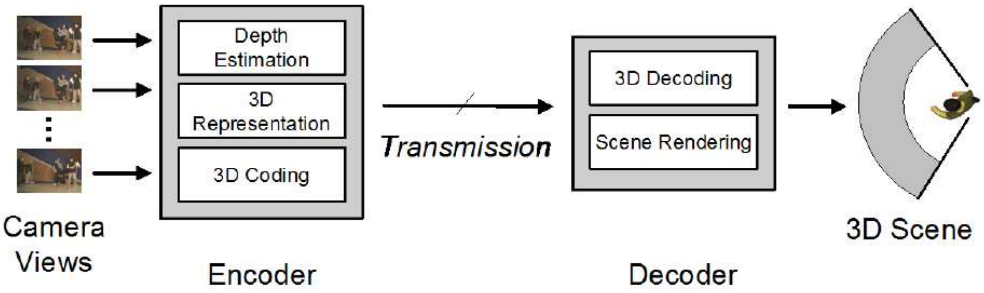

Figure 3.7: MVD processing chain

Transmitting all views directly using conventional MVC requires too much bandwidth and it wouldn't be a flexible approach, since the number of the views would be fixed not fitting on different displays. Moreover, exploiting depth image based rendering (DIBR), less views need to be transmitted then the final number displayed on the screen.

DIBR is an algorithm that, as shown in fig. 3.8, allows the extrapolation of intermediate views between two outer views. In the case of the figure starting from 3 views, 9 can be extrapolated, exploiting depth data. The procedure increases complexity at receiver side and is error prone but the bandwidth gain achieved is outstanding.

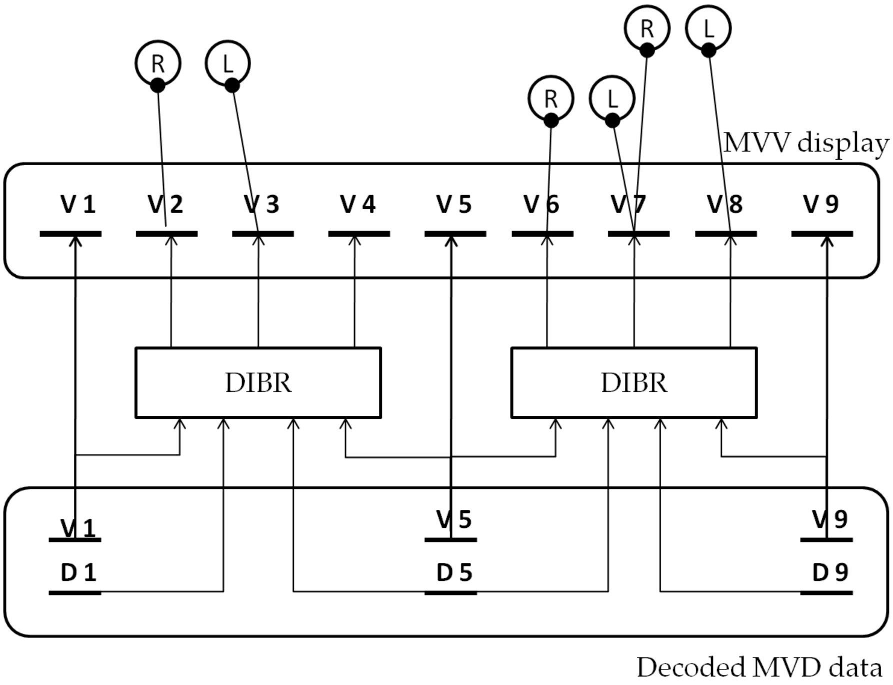

Figure 3.8: DIBR technique

3.7 Layered depth video

An alternative to MVD is offered by layered depth video (LDV). It recovers the V+D idea, adding an additional occlusion layer associated with the relative depth map. The occlusion layer brings data about the part of the scene that is covered by foreground object.

LDV comes from MVD, by warping the video image in left and right images, using depth information, than the occluded regions are obtained with other cameras in the shooting stage and collects data about scene parts covered by other objects in the scene. Only the occluded parts of the scene are finally transmitted, eventually with in addition depth information of occlusion, these streams are highly compressed.

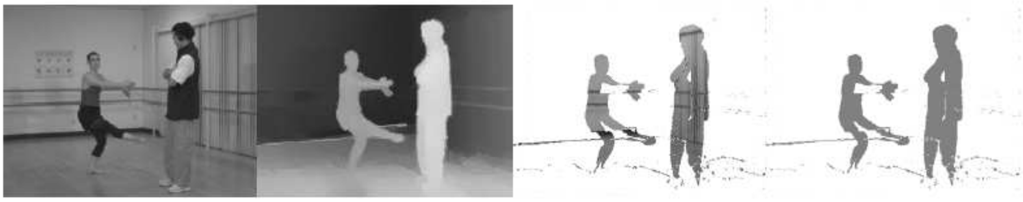

Figure 3.9: LDV structure

In this way LDV transmits less data and so is more efficient than MVD, but it requires a high complexity processing for generating the occlusion layer and to reconstruct the stereo pair images then it is more error prone than MVD.

A variation of LDV is LDV-R that consists in a LDV stream plus the video image of another views without any depth or occlusion information for the second view. The second view is sent as a residual image, not as a whole image. In this way the final reconstruction image quality is improved with only a minimum bandwidth added.

3.8 Depth enhanced stereo

Depth Enhanced Stereo (DES) is a trade-off between LDV and MVD formats, but it is restricted to stereo images. It sends a two view image with in addition a depth map and an occlusion map for both left and right view, than is let at the receiver the generation of the virtual views.

It is straight compatible with stereo display, reducing complexity processing for these devices, instead as regards MVV displays the complexity is the same achieved by LDV format, due to the same scheme of the stream.

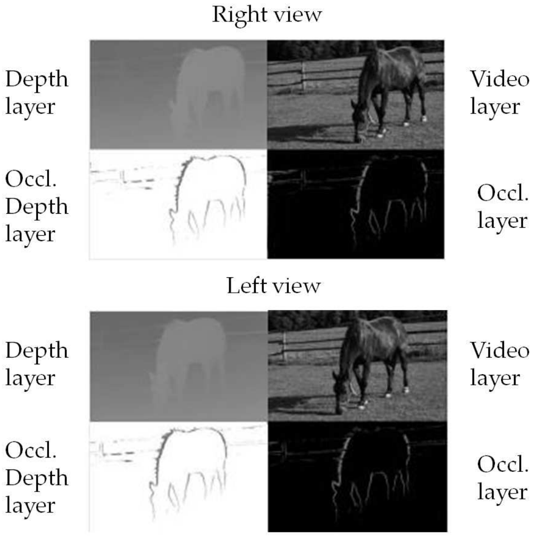

Figure 3.10: DES structure

3.9 Comparison of 3D video formats

The advantages of V+D approach over MVC is more coding efficiency, virtual view rendering, 3D display adaptation and user interactivity. The cost of this advantages is the increased processing complexity. At sender side algorithms to estimate depth maps are very complex and error prone, at receiver side has to be performed the synthesis of the stereo views, and also of the virtual ones.

LDV permits more distance between the virtual views generated than V+D format, thanks to the occlusion information. Moreover the more compression of depth and occlusion data allows a higher coding efficiency.

In contrast on this MVD formats have great redundant information from multiple views and depth maps. This brings to less efficient compression, but at the same time assures more robustness during virtual view rendering. Besides, it can apply multi-texture rendering to most of all display views, increasing the quality of the rendered images especially in some complex scene with lightning effects (mirrors, reflecting surfaces, shadows, etc.) or containing transparencies (water, smoke, glass, etc.). Using MVC some of the redundant data is discarded with a relative bandwidth gain. Also using MVC the more views are encoded the more increase the bit-rate, but MVD format allows the rendering of more virtual views than LDV.

Researches and studies are oriented on a multi video plus depth format, using DIBR alghothims bandwidth weight of the MVD stream is further reduced. MVD is flexible and generic, it supports any type of 3D displays from conventional two views stereoscopic to arbitrary large number views. On the other hand it needs further research as concern multi-view capture, depth estimation/generation, efficient compression of depth data, parametrization of system (number input views), transmission and rendering.

Finally DES is a trade-off between MVD and LDV. DES has a stereo approach, it's tied up to the left right-images, because of this it doesn't need view interpolation for standard stereo projections and displays, and it assures better quality for this system than LDV and MVD approach. The drawback is that it needs more bit-rate compared to LDV.

Thinking about a MVV/FTV scenario then, nowadays, the LDV format seems to be the best one, it's more flexible between number of views transmitted and number of virtual views interpolated, it needs less bit-rate with the same quality and it assures total compatibility with legacy 2D systems. It's drawbacks are its high complexity and error prone processing to produce the occlusion maps.

The real future scenario is expected to have MVV/FTV systems alternate with standard 3D television and legacy 2D systems, so a scalable efficient coding technique is needed.

The best candidate for this scenario might be DES, which has the main advantages of LDV format, has best image quality respect to LDV as concern about 3D stereo systems, only with a small more necessity of bandwidth.

3.10 Multi-view video coding standards

Nowadays are available two standards for video coding: MPEG C - Part 3 and MVC, in addition there is an MPEG ongoing project on 3DVC that would synthesize them.

3.10.1 MPEG-C Part 3

A video plus depth approach is implemented in MPEG-C part 3 specification (ISO/IEC 23002-3) that had been standardized in 2007. It enables simple stereoscopic application, supports video plus depth from which a second view is generated. The standard, more than just depth, specifies an Auxiliary Video Data format which consists in an array of N-bit values, associated to the pixel of a regular video stream. Then this data are compressed as conventional luminance signals. This means that depth maps are encoded as 2D video sequences, therefore it is necessary that receiver could distinguish these two types of data, to reconstruct correctly the 3D view and to avoid 2D display to show depth maps, instead of 2D video. This is achieved with a further signaling functionality.

The advantages of V+D is backward compatibility with legacy 2D systems, independence both on display and capture technology and finally good compression efficiency (low overhead).

3.10.2 Multi-view Video Coding (MVC)

The main goal of Multi-view Video Coding is to provide consistent compression efficiency compared to simulcast encoding. It combines temporal and interview prediction reducing the overall bit-rate.

Encoding and decoding each view of a multi-view stream separately can easily be done with H.264/AVC standard Codec. This simple solution provides only temporal dependencies without exploiting interview redundancy. In MVC images are predicted from temporal neighbors and from spatial neighbors in adjacent views.

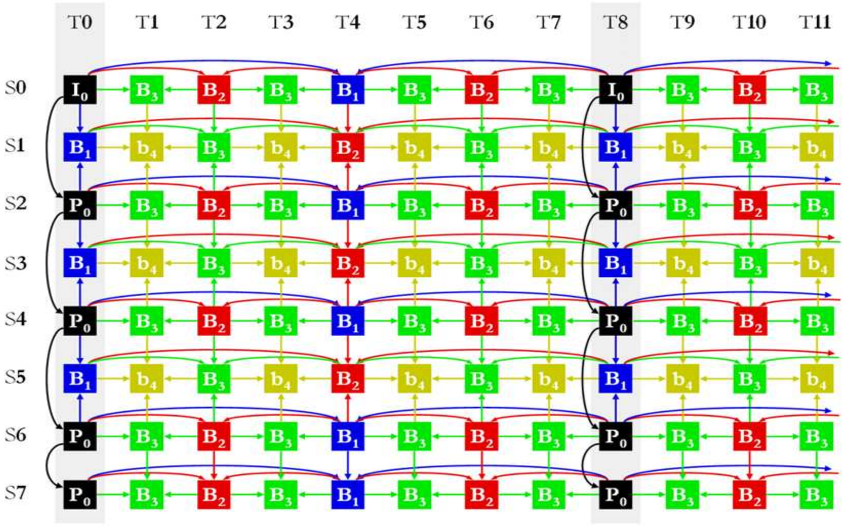
T0 T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11
S0 I_0 B_3 B_2 B_3 B_1 B_3 B_2 B_3 I_0 B_3 B_2 B_3
S1 B_1 b_4 B_3 b_4 B_2 b_4 B_3 b_4 B_1 b_4 B_3 b_4
S2 P_0 B_3 B_2 B_3 B_1 B_3 B_2 B_3 P_0 B_3 B_2 B_3
S3 B_1 b_4 B_3 b_4 B_2 b_4 B_3 b_4 B_1 b_4 B_3 b_4
S4 P_0 B_3 B_2 B_3 B_1 B_3 B_2 B_3 P_0 B_3 B_2 B_3
S5 B_1 b_4 B_3 b_4 B_2 b_4 B_3 b_4 B_1 b_4 B_3 b_4
S6 P_0 B_3 B_2 B_3 B_1 B_3 B_2 B_3 P_0 B_3 B_2 B_3
S7 P_0 B_3 B_2 B_3 B_1 B_3 B_2 B_3 P_0 B_3 B_2 B_3

Figure 3.11: MVC temporal/interview prediction scheme

Source: [1]

This format is mainly designed for multi-view auto stereoscopic displays that support more views at the same time. It's bit-rate relies on scene complexity resolution and camera arrangement, but with the increase of views number the rate reduction might be not enough to respect the channel constraints.

MVC has 3 picture types. Intracoded I-picture are decoded independently, since they are encoded without use of motion compensation. P-picture are predicted only using previous predicted frames, finally B-pictures are predicted from both future and past decoded frames. Each frame is encoded with group-of-pictures (GOP), which always starts with an I-picture and followed by B or P-pictures. A consistent coding gain is achieved replacing I-pictures with B or P-pictures.

The prediction structure of the first scene (S0 in figure 3.11) is the same as simulcast coding, it provides only temporal prediction, and so is called base view. All other views have I-pictured replaced with P or B-pictures, for the rest of the GOP the prediction scheme remains the same of the base view. Moreover synchronization and random access is provided by all key-pictures coded in intra-mode. In this way views cannot be decoded independently, therefore if errors happen in a view these strike again to other views leading to the loss of the all data.

The prediction structure lead to inter camera redundancy, therefore it's needed a trade-off in memory, delay, computation and coding efficiency.

MVC requires more buffering before the pictures can be displayed in order, therefore are possible simplified prediction schemes, which reduce the amount of buffer necessary. In the simplified scheme of fig.3.12 right view pictures are only predicted from other right view images, or left I-pictures.

This simplified scheme is called KS_IPP. In another scheme, called KS_PIP, it is possible to choose as reference image a central view, instead of S0, in this way not one but two views are directly predicted from a I-picture, this might improve coding gain.

All multi video sequences can be combined in a single uncompressed video stream, as presented in figure 3.13, then the stream is used as input of the H.264/AVC encoder.

After the decoding phase, a reordering is applied to the decoded pictures, to separate individual views.

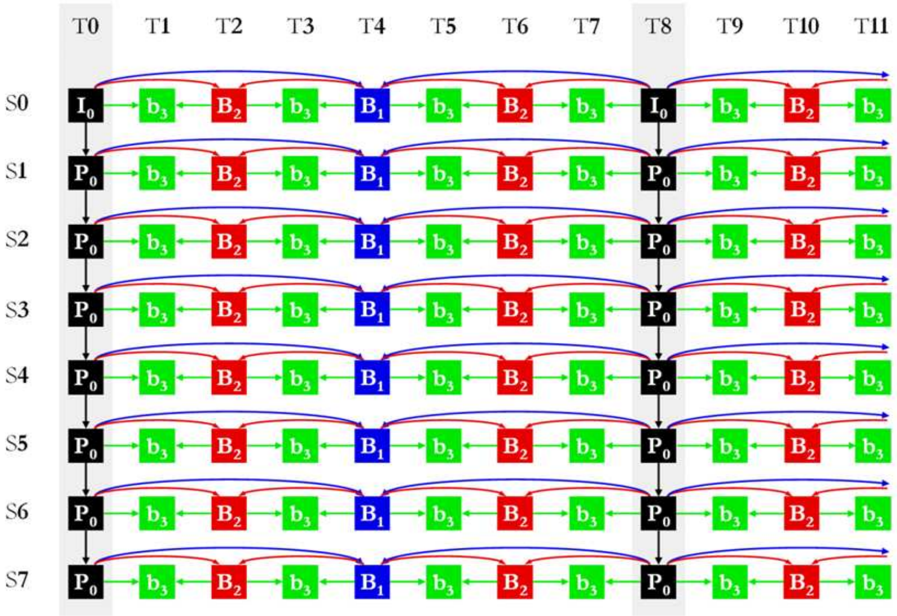
T0 T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11
S0 I_0 b_3 B_2 b_3 B_1 b_3 B_2 b_3 I_0 b_3 B_2 b_3
S1 P_0 b_3 B_2 b_3 B_1 b_3 B_2 b_3 P_0 b_3 B_2 b_3
S2 P_0 b_3 B_2 b_3 B_1 b_3 B_2 b_3 P_0 b_3 B_2 b_3
S3 P_0 b_3 B_2 b_3 B_1 b_3 B_2 b_3 P_0 b_3 B_2 b_3
S4 P_0 b_3 B_2 b_3 B_1 b_3 B_2 b_3 P_0 b_3 B_2 b_3
S5 P_0 b_3 B_2 b_3 B_1 b_3 B_2 b_3 P_0 b_3 B_2 b_3
S6 P_0 b_3 B_2 b_3 B_1 b_3 B_2 b_3 P_0 b_3 B_2 b_3
S7 P_0 b_3 B_2 b_3 B_1 b_3 B_2 b_3 P_0 b_3 B_2 b_3

Figure 3.12: MVC simplified scheme

Source: [1]

From an analysis of these different scheme it results that the best gain is obtained with the standard and more complex approach, otherwise KS_IPP and KS_PIP are simpler but have less coding gain. A simulcast approach assures worst performance as regard coding gain, as it can be easily supposed.

The large amount of coding gain is achieved through temporal prediction. Otherwise interview prediction introduces, in turn, an additional gain.

MVC extension of H.264/AVC

The widely deployed ITU-T H.264/AVC standard has been extended, in his version 4, to support MVC. This extension includes techniques for improved coding efficiency, reduced decoding complexity and new functionalities for multi-view operations.

This standard doesn't require any changes to lower-level syntax and is compatible with single-layer AVC hardware, it requires only small changes at high-level syntax, as concern specific view dependency.

It is based on High Profile of H.264/AVC, it uses mainly hierarchical B-pictures, CABAC (Context-adaptive binary arithmetic coding, an entropy coding method)

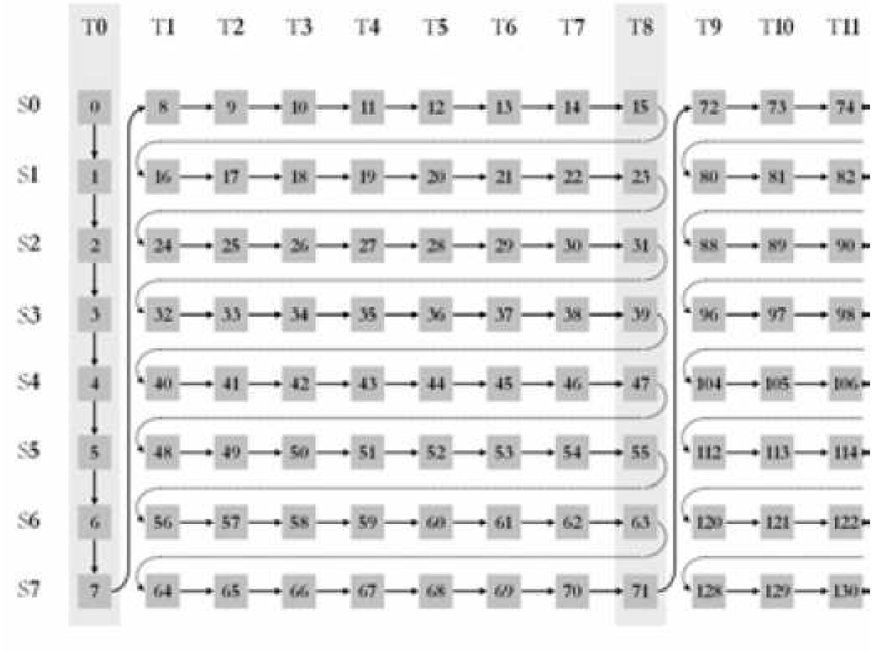
T0 T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11
S0 0 8 9 10 11 12 13 14 15 72 73 74
S1 1 16 17 18 19 20 21 22 23 80 81 82
S2 2 24 25 26 27 28 29 30 31 88 89 90
S3 3 32 33 34 35 36 37 38 39 96 97 98
S4 4 40 41 42 43 44 45 46 47 104 105 106
S5 5 48 49 50 51 52 53 54 55 112 113 114
S6 6 56 57 58 59 60 61 62 63 120 121 122
S7 7 64 65 66 67 68 69 70 71 128 129 130

Figure 3.13: Frame interleaving for compression of H.264/AVC

Source: [1]

and disparity compensation between the frames of different cameras.

The encoder generates a single stream from N temporally synchronized video streams. The decoder receives the stream and decode the N video signals.

The 2D video of an MVC bit-stream forms a base level for the MVC stream and can be reconstructed by a standard H.264/AVC decoder, allowing backward compatibility.

The generation of the output views requires control of the available decoder resources. In this version there are features as marks of the reference pictures, support for view switching, a bit-stream structuring, view scalability signaling, supplemental enhancement information (SEI) and parallel decoding (SEI).

A first version of this extension had been standardized in July 2008.

3.10.3 3D Video Coding (3DVC)

3DVC is a standard under research at MPEG, it synthesizes V+D approach of MPEG C - Part 3 and MVC extension of the H.264/AVC.

It implements a multi video plus depth format, more video plus depth views are encoded exploiting temporal and spatial redundancy in mvc streams.

Aim of 3DVC consists in supporting a large variety of 3D displays types and sizes, from stereoscopic ones of different sizes and baselines and more sophisticated multi-view ones, with variable number of views.

It exploits depth image based rendering, to reduce the transmitted views, as its expected by MVD format.

To obtain horizontal parallax narrow acquisition angles are needed, furthermore rectification process may be not necessary, if necessary they have to be performed only at encoder side, and not at the decoder, thus lowering complexity and so costs of home side devices.

The idea of 3DVC is to code separately video and depth information of the same view in two different MVC streams. 3DVC is an ongoing MPEG activity, a standard is expected in 2010/2011.

Chapter 4

Review of 3D Technologies

Before providing a possible MVV scenario, becomes fundamental an analysis of the currently available 3D technology as concern display, set top boxes (STB), interconnections and representations and of their future developments.

There is a great effort in research regarding each of these technologies, therefore there is a continue development and new devices or techniques are rising.

Different types of 3D displays are being studied, from available stereoscopic displays, and prototypes of auto stereoscopic ones to more far solutions as light-field or volumetric displays. 3D STB technologies are at early stages but some private demos has been shown during last IBC 2009 in Amsterdam. Both different type of displays and STBs currently available have been reviewed, looking at compatibility with different coding techniques and also on the interconnection standards to take in account their capability to interconnect future MVV TV set to STBs.

Also different 3D representations have been presented starting from object based representations to image based ones, underlying the advantages and the drawbacks of each one.

4.1 Review of 3D Displays

In the past there wasn't much confidence about the possibility that 3D systems for home applications would have achieved success. The need of wear glasses and the low image quality were seen as great obstacles for these systems. In cinema applications instead, wearing glasses was more accepted because it is necessary only for a short period and is seen more as an event.

Recently the interest in 3D video is growing fast, more and more 3D films are produced and consequently more and more cinemas are equipping with 3D technologies.

This is creating much popularity of 3D also at the user side and new technologies as 3D-DVD/blue ray, 3DTV broadcasting, and Internet will provide home spread of 3D, and so there is a lot of research and development on 3DTV.

The idea is to provide the best left-right view separation with the minimum cost for image quality.

In the future a TV set that will have received a 3D standard image shall display it according to its display technology.

A great effort in drawing a 3D display map is steadily done by 3D@home Consortium, that formed in the 2008 with the mission to speed the commercialization of 3D into homes worldwide, facilitate development of 3D standards and draw road-maps for the entire 3D industry.

The classification presented mainly follows 3D@home display's vision.

All possible 3D displays can be divided in four main categories[1]:

- stereoscopic,
- auto - stereoscopic,
- light-field
- volumetric

Currently only stereoscopic and auto - stereoscopic devices are available on the market for home user applications.

Stereoscopic displays are designed for stereo 3D video, only a view is supported and they need wearing special glasses; auto stereoscopic displays don't need to use any type of special glasses, most of them support multi view video; light-field displays are similar to the auto stereoscopic ones the difference consists in the way the light is projected by each pixel; finally there are the volumetric ones, these devices are completely different, they project the image on a white plane that rotates on itself, creating a volumetric image in the space, this is different from MVV as is nowadays referred to, because it isn't possible to represent scenarios but only characters.

It could be 10 years before high quality commercial products will be available at competitive pricing.

[1] Source: 3D@Home Consortium

Typology	Manufacturer
Projection Systems	Viewsonic, Projection Design
Flat Panels	Hyundai, JVC, Sony, Samsung, LG

Table 4.1: Stereoscopic display categories

In the next sections will be described more in depth each type of these displays.

4.1.1 Stereoscopic displays

Stereoscopic displays are nowadays ready to reach the market, many manufacturers have commercial products, the most constraints in their diffusion are lack of contents and prices till quite high.

These devices present on the screen two images, one intended for the right eye and one for the left eye; then using some special glasses each user's eye receives only an image and the brain create the 3D perception.

Projection systems

In projectors there are two basic ways to reproduce 3D: using a dual stack (two projectors), or a single one. For each of these structure there are multiple technologies to reproduce the final 3D perception.

Both setting up can be configured as front or rear projection: front projection is a high - end solution, instead rear projectors are the most common and currently adopted solutions. Nowadays a lot of these systems, mainly by Samsung or Mitsubishi have already reached end - user house, and can support 3D vision using active glasses, but most of the owner maybe don't even know about this possibility.

The dual stack approach has several advantages over the single projector. Mainly they yield much more light in 3D mode. Moreover they support all projection technologies (PdP, LcOS and DLP), instead single projectors support only DLP systems. As can be easily expected the drawbacks are added costs and alignment issues. Technology is moving on and so alignment issues will be soon solved, instead costs will remain.

Finally single projector 3D systems are more attractive than dual stack ones, but there will always be a gap between the two setup, also using the same type of projector.

A common drawback for both setup, mainly in home theater setup, where illumination is usually low, is the loss in light that will be consistent going from a 2D to a 3D, typically only 15 - 20 % of the light in 2D. Thus the projectors need a lot more light in the 2D mode to assure a comfortable 3D viewing.

Flat panels: LCD and PDP solutions

There are two main approaches for these displays, with a third one that could arrive in the market in a few years.

- 120Hz Active shutter glasses
- micro pol Patterned retarder
- Active retarder

The 120 Hz active shutter glasses is the simplest system, but nowadays various LCD displays at 120 or even at 240 Hz available on the market can't support 3D. They are designed to receive an input signal at 30 or 60 Hz and they can't manage a left-right 120 Hz input. Moreover in a 240 Hz display the incoming signal at 60 Hz is interpolated and 3 more frames are reproduced at 240 Hz refresh rate. In a 3D configuration the approach is different, the input signal is at 120 Hz because it contains a right and a left image for each frame, therefore in a 240Hz environment, the left frame is displayed in the first sub-frame then it is kept for the second one allowing the shutter of left eye glasses to be opened and then the same happens for the respective right frame for other two sub-frames.

A good separation in the timing of left and right images reproduction is a fundamental issue to lower and possible avoid crosstalk and ghosting effects, on the other side this reduces the light output of the screen.

In plasma displays the image update doesn't work as in the LCD systems, but more likely DLP projectors using bit planes. Panasonic has demonstrate the feasibility of these approach in large display till 103 inch.

The second approach is called x-pol or micro-pol, there are different techniques to attend this approach:

- Polarized glasses, space based division (row or column interleaved)

- Active field interleaved frames or Active DLP checker-board frame interleaved
- Spectral Frequency separation (Infitec like)

The idea of the space based division is to add a patterned retarder sheet, through a lamination process, over a 2D LCD screen. Each row of the sheet is aligned to the rows of the LCD and is composed by retarders or polarizers. In this way odd and even rows present opposite polarizations. Finally the image on the screen has the left image on the odd rows and the right one on the even rows, or vice versa. The last step belongs to the user that has to wear passive glasses to separates the images.

On the same idea are based active DLP checker-board frame interleaved and active field interleaved frames, the only difference relies on the pattern of the interleaved polarized pixels that can assume a checkerboard structure or may refers to whole fields.

The last typology of micro polarizator is spectral frequency separator, that are mainly produced by Infitec. It uses interference filters to make a spectral frequency separation.

The basic colors of left and right images are projected at slightly different wavelengths, then glasses with selective interference filters separate the two images.

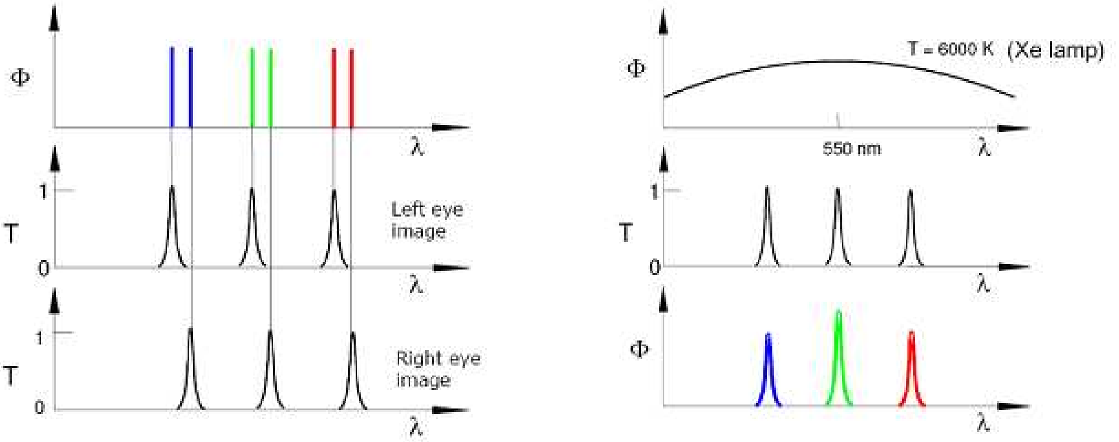

Figure 4.1: Infitec spectral frequency separation

A typical set of frequencies is:
Left eye: Red 629nm, Green 532nm, Blue 446nm
Right eye: Red 615nm, Green 518nm, Blue 432nm

These methods allow displays to work at 60 Hz and therefore an eventual greater refresh rate has the same advantage that in 2D television, with the improvement of motion blur.

Linear interleaved format provides only half the vertical resolution respect the page flipping system, moreover the larger is the screen the more this effect is perceived. Actually the drawback isn't so great thanks to the powerful human brain, that is able, merging the two images, to fill any missing information and also thanks to the encoding techniques that can quite overlap this issue.

Text and icons must be displayed separately from the image, to avoid tiresome distortions.

The third approach is a hybrid of the previous and it is expected to be ready for the market in two or three years.

A series of retarder bar is applied on the screen, each bar is activated as the row, underneath it, is updated, till all rows are done updating. Therefore the lines under the activated bars are in one circular polarization state while the rows under the unactivated bars are in the related orthogonal polarization state, so with passive glasses the two images are filtered and divided.

With the active retarder approach both eyes can simultaneously receive light from the display, therefore the light-passing duty cycle is greater compared to a shutter glasses setup, where only one eye at a time receive it, and this is achieved using passive glasses. (These systems are developed mainly by LG electronics).

Active glasses are nowadays expensive (from 50 to over 600 €), therefore aren't expected to spread the market, instead passive polarized glasses, that provide less final perceiving quality, are pretty inexpensive and supplementary costs are added at the display. These costs are expected to decrease due to new x-pol materials, lamination investments and inserting x-pol inside the LCD panels.

Also development of OLED (organic LED television) would come towards 3D, and they wouldn't have any constraint as regards 120 or 240 Hz page flipping or active retarder systems. Till now OLED have scalability issues, only Sony has introduced a 11-inch television, but it has also recently suspended its activities around it.

4.1.2 Auto stereoscopic displays

Auto-stereoscopic displays are the best candidates for a future home usage. They emit more views at the same time but the user only sees an image at a time.

Currently available displays support from two to ten different views at a time, these views are presented on the screen in a column interleaved spatial multiplex. More over if consecutive views are stereo pairs and arranged properly, than motion parallax viewing can be supported. In the future it's expected that the number of different views of a multi view display will increase.

These devices don't need the user to wear special glasses; they project left and right images in a specific point in the space so the user's eyes receive independently left and right views. This possibility is achieved using an eye tracking system which automatically adjusts the two images on user's eyes following their movements. As an alternative, through a preliminary set-up of the user's position and of distance between his eyes, a passive system can be implemented.

These displays exploit optical principles as diffraction, refraction, reflection and occlusion to drive the left-right images to the user, and they can be divided in two main categories, depending on which technology they are built on, parallel barrier or lenticular lens. These techniques are the cheaper one and provide an high final image quality. A relevant issue of these devices remains the necessity of convertibility from 3D to 2D and vice versa, thus providing high quality images both using stereo, multi view or 2D visualization.

Technology	Manufacturer
Lenticular	Alioscopy, Philips, LG, Samsung-Magn., Spatial view
Parallax Barrier	Tridelity, Samsung-Magn., Newsight
Others (Multiple Projectors, Fast Projectors/LCD, Head/eye tracked FPD)	Newsight, Visureal, Seereal

Table 4.2: MVV displays categories

Lenticular lens displays

Lenticular display presents on the screen a vertical interlaced image, obtained merging left and right images, then in front of the screen there is a sheet of half cylindrical lenses that deviates images orientating them towards viewer's eyes.

Under each tiny lens are interlaced from 10 to 15 separate images therefore is required high precision manufacturing.

The lenticular sheet can be easily activated or not applying a specific voltage to it, in this way it can be easily go back to a 2D visualization. The process is simple: each lenticular lens is filled with crystal liquid and an electrode (Indium Thin Oxide) with whom the voltage is applied. Outside the lens there is a replica with the same refractive index of the liquid crystal in the lens. This allows to activate or not the lenticular lens depending on the applied voltage. As a matter of fact with the appropriate voltage, refractive index is very close between the LC and the replica, making the lenticular lens transparent.

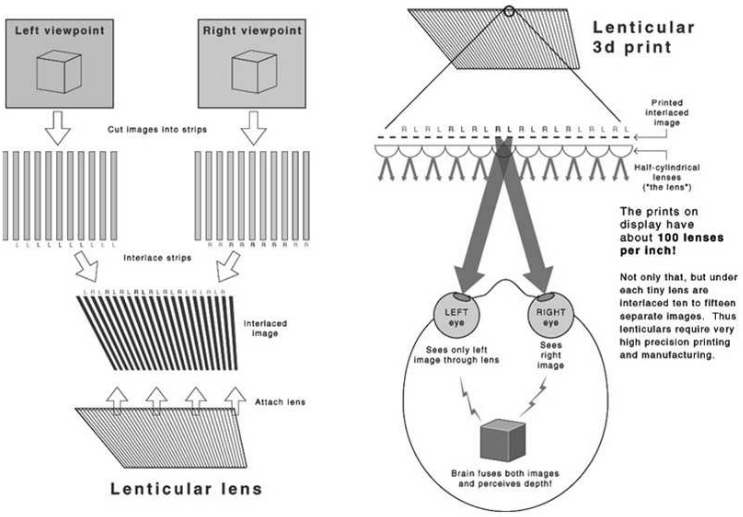
Left viewpoint
Right viewpoint
Cut images into strips
L L L L L L L L L L
R R R R R R R R R R
Interlace strips
L R L R L R L R L R L R L R L R L R L R
Interlaced image
Attach lens
Lenticular lens
Lenticular 3d print
Printed interlaced image
R L R L R L R L R L R L R L R L R L R L
Half-cylindrical lenses ("the lens")
The prints on display have about **100 lenses per inch!**
Not only that, but under each tiny lens are interlaced ten to fifteen separate images. Thus lenticulars require very high precision printing and manufacturing.
LEFT eye
RIGHT eye
Sees only left image through lens
Sees right image
Brain fuses both images and perceives depth!

Figure 4.2: Lenticular lens display

Source: [2]

Parallel barrier displays

These displays consist of a flat panel and a parallax barrier. They display on the screen more images side by side in vertical stripes, then the barrier, that is placed at a certain distance from the panel and has a series of white slit and black strip, splits them, allowing each viewer's eye to see only the right or left image. These displays have to produce high quality images with little or even without crosstalk between left and right images. It means that left eye has to see only left image, and right eye only the right one.

The quality of the stereo viewing is described by the stereo image quality factor, $Q = \frac{S_s}{S_w}$ where S_s is the area of a left or right image that the viewer can see at a specified position and S_w is the area of the left-eye or right-eye image displayed on the screen. Q=1 is the best for viewing stereo images Q=0 is the worst.

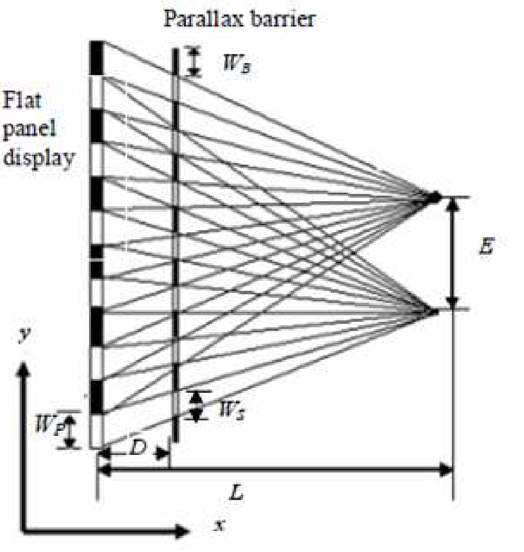

Figure 4.3: Parallax barrier display

Source: [3]

Finally the prerogative of a 3D auto stereoscopic displays to be competitive with common 2D ones are:

- Large screen comparable to 2D ones
- 2D/3D convertibility
- High quality image without or with little crosstalk between the left/right images

In a first analysis is clear that reproducing at the same time on the screen more views the resolution greatly decreases by the number of the views presented on the screen. The views are alternated in vertical stripes than the vertical resolution decreases, this, as a consequence, reduces the final overall image quality.

The first samples of these displays are reaching the market, also if they are at the beginning and a lot of research and development are needed, before achieving high quality devices, another restrain regards that MVV contents need a special shooting setup and nowadays there aren't true MVV pure contents available, then auto stereoscopic displays are used only for stereo videos without recurring to any special glasses.

It is expected that before auto - stereoscopic displays would hit the mass market they would achieve some improvements as regards width and visibility zone, the overall image quality and the display resolution in 2D mode.

4.1.3 Light-field displays

Light - field displays are usually referred as "incoherent holography", they are similar to auto stereoscopic lenticular lens ones; as lenticular ones, they offer the possibility to reproduce more than a view on the screen without need to wear glasses, the main difference refers to the way each pixel projects light in the environment.

Each pixel projects different beams of light on different directions. This structure brings to a more defined image, seen at different point of view by the user. Each pixel can be used for various views of the scene, instead in the standard auto - stereoscopic devices, each pixel provides light only for a specific view.

This effect is achieved using multiple projectors or a projection screen with a sheet of micro lenses. Projectors generate an array of pixels at controlled intensity and color, otherwise integral lenses orientate different color light rays towards various directions, providing a view-dependent effect. Lenticular auto stereoscopic displays use arrays of cylindrical lenses, instead light-field display as shown in fig 4.4 employs an integral lens that provide horizontal and vertical parallax. Next to this technology there are other techniques to achieve the light - field effect, as distributed views.

With these techniques is reproduced the same light-field, that would emanate from the original scene, moreover vertical and horizontal parallax are assured, through a large range of viewing angles.

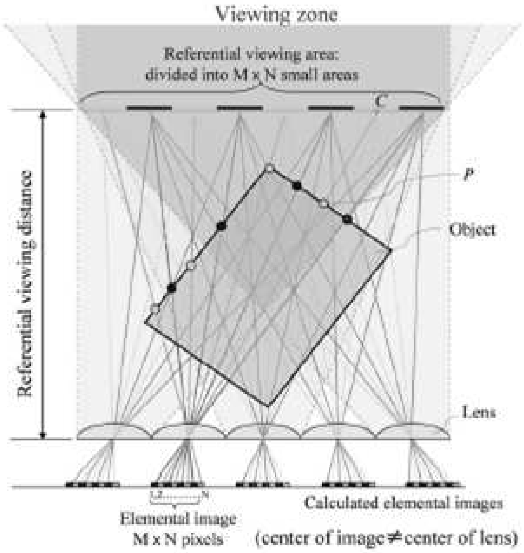
Viewing zone
Referential viewing area: divided into M x N small areas
C
P
Object
Referential viewing distance
Lens
1,2..........N
Elemental image M x N pixels
Calculated elemental images
(center of image≠center of lens)

Figure 4.4: Light-field principle

Source: [4]

4.1.4 Volumetric displays

Volumetric devices form a representation of the object in three physical dimensions, displaying points of light within a certain volume; they achieve a 360 ◦ view, but without providing a closed distanced background.

Instead of pixels they use voxels, unity of volume representing a signal intensity value or color. Each voxel emits visible light from the region in which it appears.

There are a number of techniques to create the 3D image, one of these is using a rotating panel. A projector lights a blank sheet that rotates quickly and continuously on itself, this rotation creates the effect of a real 3D object. These devices allow a wide field of view, as a matter of fact full parallax is achieved (both horizontal and vertical), and support multiple simultaneous observers, eventually size of the reproduced image varies from few centimeters to a meter. Moreover they solve the rivalry between accommodation and convergence, as already seen a great issue for 3D applications.

Otherwise in these devices displayed surfaces appears transparent objects lose consistency. The most constraint devices is given by the necessity of a full 360 ◦ view of the scene, that isn't supported teorically by MVV which is limited to 180 ◦ due to costs and complexity in shooting and encoding.

Another drawback has been anticipated before, with these displays it isn't possible to represent any type of background.

There isn't till now an official taxonomy about volumetric displays, a preliminary division can be done between virtual image displays and real image ones.

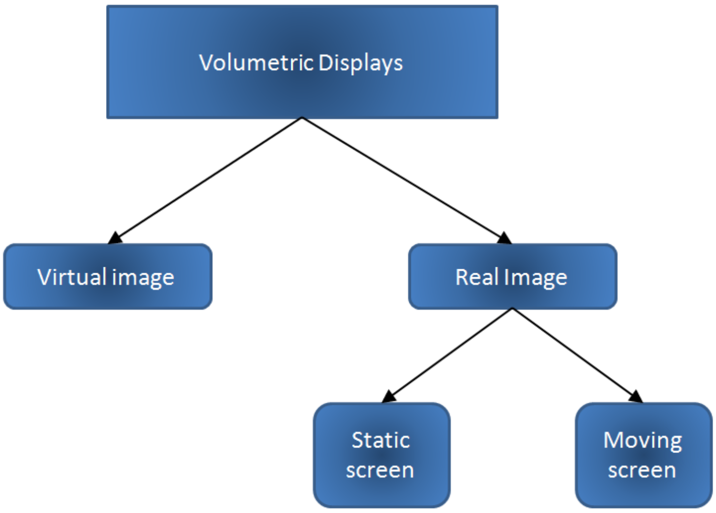

Figure 4.5: Swept volumetric classification

The first ones exploit deformable mirrors to create virtual depth planes, each mirror focal length varies with rotation or in time. The slice images are reproduced sequentially and using lens or mirrors the virtual image is projected through the space.

Real image devices display views on a screen with actual depth and divide itself in two other categories: moving or static screen. Moving screens exploit rotating plane or moving screen to produce series of slices in sequence. Otherwise static screens don't have any moving parts, two main methods are possible: light can be "piped" to the individual voxel position, or images are projected onto a series of stacked parallel screens that are sequentially rendered opaques.

Both moving and static screens can use different image space subsystem, moreover there are more techniques to generate and activate voxels. Image space creation subsystem defines the technique that creates the transparent volume within which the images have to be displayed. There are two approaches: swept volume display units, which creates a space for the image exploiting the rotation or translation of a blank surface, or static volume display motion, in which a static material includes the image space.

Moving screen displays rely on the human persistence of vision, the human brain fuse the last time series of 3D region, creating a unique 3D perception.

There are many techniques to create this effect, one of these is using a rotating panel. A projector lights a blank sheet that rotates quickly and continuously on itself, the image displayed slightly changes as the surface rotates, the whole rotation creates the effect of a real 3D object.

Static volume devices are the most direct form of volumetric display. An addressable volume of active voxels is created, the elements are transparent in the on state, and opaque or luminous in the off state. Activating them or not they display a solid image in the space. Several static displays exploit laser light to bring visible radiation within a liquid, gas or solid.
A more complex approach exploits a two-step up conversion illuminating a glassy earth-doped material intersecting two infrared laser beams of different wavelengths.

Voxel generation and voxel activation system

The voxel generation subsystem refers to the technique used to generate the visible voxels.

There can be a passive approach or an active one. In the passive system the planar surface contains a persistence phosphor and voxel's activation is achieved by one or more electron beams that etch on it, these devices are defined beam addressed.

As an alternative the rotating surface can be composed of an array of optoelectronics elements, and each voxel generation center could originate multiple voxels during each cycle of motion, by the application of a suitable electrical stimulus.

Moving screen display,based on passive voxel generation, are directly beam addressed, instead systems employing active array of discrete voxel generation centers are directly addressed.

Static volume could employ both passive or active voxel generation techniques. Both of them are beam addressed systems, voxels generation employs a two step excitation process, within the region of two intersecting beam sources through a passive gaseous or non gaseous medium. The last type of system employ a particle cloud or suspension for the voxel generation.

Light - field and volumetric displays are very futuristic and nowadays there is a lot of research regarding them, but there aren't devices ready yet neither in the form of prototypes.

4.2 Set Top Boxes

Set Top Boxes have two main features: provide the reception of satellite feed and optional interaction features, convert the video stream in the format supported by the specific display they serve. As the migration to a complete new system is always a tough race, it is needed that new 3D STBs are still able to provide the 2D-HD service, on the other side the old HD STBs must be able to manage the new MVV stream extracting old 2D HD video stream for old 2D displays.

Nowadays 3D STB technologies are at early stages, all of them still rely on the side by side or row interleaved image codification techniques. Some decoder, through an image manipulation module, are able to convert a side by side or over under coded image in all the output formats supported by today display sets:

- Side by side
- Checker-board
- Row interleaved
- Time interleaved
- Field interleaved

Some demos implementations from Sagem and NDS has been shown during last IBC 2009 in Amsterdam. NDS' prototypes, in particular, presents major improvements on the interface and image manipulation.

Till now there aren't samples, that provide advanced DVB or MPEG decoding features able to decode 3D streams as H264/MVC.

Some implementations provide a real interoperability between display sets and broadcasted channels, through a 3D stereoscopic interface that provides an exciting user experience, with a multiple layers menu that allows an advanced and amazing navigation among programs and channels.

Switch from 3D to 2D and vice versa requires the knowledge of the display type and the display mode through HDMI negotiation but this feature is not forecasted even in HDMI 1.4 according to released specs, STBs give the advantage to allow the 2D operational mode, displaying pixels received from HDMI and managing transition from mono to stereoscopic adapting the GUI accordingly.

Another important feature regards closed caption area, they are usually 2D characters generated by the decoder and overlayed to the image. It's up to the user to activate or hide them. These features are regulated by ETSI EN 300 743 and can be divided in two categories:

- Bitmap based
- Character based

The most simple approach is to apply the pixels directly as they are on the 3D scene, thus showing them on the zero plane, but this isn't always so comfortable. There are investigations regarding dynamic depth adjustment to avoid the viewer to go back and forth with eye convergence and accommodation.

4.3 Interconnections

Another relevant aspect regarding 3D devices refers to interconnection cables between displays, STBs, decoders and other devices. The relevance of transporting such 3D information is as great as the visualization and transmission phases themselves.

Recently have been released HDMI 1.4 and display port 1.2, two new specifications that best suit with a lot of stereo displaying techniques, from interleaved frame to side by side.

4.3.1 HDMI 1.4

Figure 4.6: HDMI cable

HDMI 1.4 presents a number of improvements respect to precedent 1.3.

Specifically as regards 3D aspects it supports up to a full HD resolution for a 3D stream 1920x1080p using almost any 3D technique now available for stereoscopic displays such as:

- Full side-by-side

- Half side-by-side
- Frame alternative
- Field alternative
- Line alternative
- Single view + Depth layer
- Single view + Depth layer + Occlusion layer + Occlusion Depth layer

Jointly to 3D specs also other new functionalities are introduced in the 1.4 released, as the ethernet support, the audio return channel and more color spaces supported. Moreover the resolution achieved by this release has increased to 4096x2160 at 24 Hz, or 3840x2160 at 24, 25 or 30 Hz.

4.3.2 Display port 1.2

Display port 1.2 is a competitor of HDMI 1.4, the most specification are similar in both cables, it supports all 3D techniques now available for stereoscopic displays and allows Ethernet interfaces.

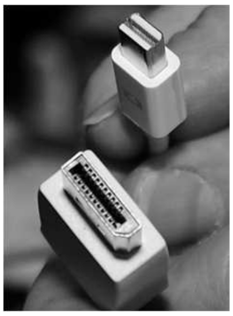

Figure 4.7: Display port cable

Display port supports till four independent channels at a time, and has doubled bandwidth, respect old 1.1 release, up to 21.6 Mb/s.

It achieves up to a 3D double stream 2560x1600 resolution at 120 Hz, or to a full-HD 3D resolution at 240 Hz, this is allowed by the four possible channels supported, with which is able to guide from one single monitor with a 3840x2400 resolution at 60 Hz, or 4 independent displays at full HD resolution at 60 Hz.

4.4 3D representation technologies

At the state of the art there are many techniques for 3D representation, they differs from the amount and type of geometry and texture used. All different 3D representations techniques can be classified in a continuous way between image-based or geometry based representations. The choice of the type of representation is fundamental for the design of any 3D/MVV/FVV system. On one hand it sets the requirements for acquisition and multi-view signal processing, on the other hand it determines the rendering algorithms, interactivity and if necessary compression and transmission. For instance, image-based representations require a dense camera setting, more scene have to be captured for a good rendering of the virtual views. Otherwise geometry-based representations need complex and error prone image processing algorithms such as 3D geometry reconstruction and object segmentation.

4.4.1 Geometry-based representations

Geometry-based representations have typical usage in 3D computer graphics, in the most cases, such as surface-based representations, they are based on meshes of polygons. Scene is built segmenting the image into a collection of surface patches, whose position, orientation and shape in 3D space is estimate. Scene geometry is reconstructed by a set of meshes that reproduce the real object using 3D geometry surfaces, without any redundancy. Typically a mesh is a set of triangles.

A texture map is associated to these surfaces, and also other attributes as appearance properties (opacity, reflectance, specular lights, etc.) may be assigned to enhance the realism of the model.

Instead a method that doesn't use meshes, is point-based representation, in which points are samples of the surface, and describe surface's 3D geometry and surface reflectance properties. It doesn't need any information regarding explicit connectivity, topology and texture or bumping maps, and is an efficient alternative to mesh based approaches. Point-based methods specify implicitly connection information through the interrelation among points. This technique is suitable for setting where geometry model changes frequently thanks to the ease of insertion, deletion and repositioning of point samples. The drawback is that each point requires more geometric information than image pixels.

Geometric-based approaches are commonly used in applications such as computer games, Internet, TV, movies. Especially if scenes are computer generated, the achievement performances are excellent. The available technologies for production and rendering is highly optimized and the current PC graphics card are able to render highly complex scene with excellent quality as regards reproduction of motion, accuracy of the texture, spatial resolution, levels of detail and refresh rate.

The main drawback is that content creation requires high costs and human assistance. The algorithms used are extremely complex, and then computation of 3D scene models is often limited only to foreground objects. Moreover they are error prone; therefore the geometric model may have some errors and these strikes again on the rendered images.

4.4.2 Image-based representations

Image-based representations don't use any 3D geometry. Virtual intermediate views are generated through interpolation of natural camera views. The main advantage of this type of representation is a potential high quality of virtual view synthesis without any 3D reconstruction, but to achieve this is required a large number of natural camera view images. If a sparse camera setting is used, interpolation and occlusion artifacts will appear in the synthesized images, affecting quality. Otherwise large number of camera implies a great amount of image data to be processed. Examples of image-based representations are light-field representations, where an intuitive description of the view-dependent appearance of the scene is offered by the parameterized light field.

They have to cope with an extremely complexity of data acquisition or they have to execute simplifications but reducing the interactivity.

Between these two methods there are a number of techniques that use more or less both approaches.

Pseudo-3D representations don't use explicit 3D models, but depth or disparity maps. This map assigns a depth value to each sample of the image. Then together with the 2D views the depth map creates a 3D-like representation. This representation is more flexible and needs less dense sampling of the real scene; otherwise it contains redundancy of transmitted data when multiple views are captured. Methods closer to geometry based representation use view-dependent geometry and/or view dependent texture. For example surface light-fields combine the idea of light fields with an explicit 3D model and achieve great results especially for shiny object under complex lighting conditions.

Instead, volumetric representations, which recur to voxel, can be used instead of a complete 3D mesh model; they use an octree structure, which permits an almost linear time access. The visualization has low quality especially when rendering viewpoint is close to the surface because voxels are rendered as cube.

Another type of mixed representation is object-based representation, which have powerful and promising approach. A 3D object or a scene is considered as a collection of images, called key frames taken from a specific viewpoints and then could be generate every other views of it. Each key frame is captured individually and synthesized through a mesh, and then the scene is generated collecting all single objects. This approach allows to store in memory different objects and to avoid repetitive rendering of the same object, reducing the overall cost computation. But at the state of the art models created through meshes in this way look still artificial and plastic.

4.4.3 Texture mapping

A texture mapping is necessary to achieve a final 3D good quality image in mesh based representations. It's a method for adding detail, surface texture, or color to a computer-generated graphic or 3D model. A texture map is applied to the surface of a shape or polygon. It can be performed also multi texture, which consists in using more than one texture for each polygon, for instance, may be used a light map to light a surface instead of recalculate that lighting every time the surface is rendered.

Another multi-texture technique is bump mapping, which assigns a particular texture to directly control the facing direction of a surface for the lighting calculations, in this way very complex surface, such as tree bark or rough concrete get a very good appearance.

4.5 Final considerations on state of the art 3D technologies

Displays

The first prerogative of any television system is to provide the best final user experience and such a system must support the most various end-user devices available.

Cinema 3D contents are spreading so it's expected that also home - user interest about 3D will grown, and stereoscopic displays should be the first step to bring 3D to the home, when their prizes will be reasonable. They provide only two views, so,

Technology	Eye-wear required	MVV Support	MASS Market availability forecast
Light Field	No	MVV	Long
Volumetric	No	Possible MVV	Very Long
Auto-stereoscopic	No	MVV	Medium
Stereoscopic	Yes	No MVV	Short

Table 4.3: Final resume of display technologies

eventually, the only way to support MVV service is using an additional STB with head tracking functionalities.

In the short period stereoscopic displays present a concrete possibility to have market, because the auto-stereoscopic are too much expensive and most of them still only in form of prototypes.

Nowadays, stereoscopic devices are almost ready their prizes are decreasing progressively, but not many devices are available in Europe or they are in pre-production. Moreover it's difficult to believe that many users would accept wearing glasses in a home environment.

The drawbacks of auto-stereoscopic displays are that they, giving nine or more multiple views to the user, reduce the resolution of the single view by the number of the provided views. In the 3D mode the resolution loss is partially compensated by the brain reconstruction of the scene, instead in 2D the resolution loss is present and is widely perceived by the user.

For these reasons are necessary some image quality improvements, as regards display resolution, especially in 2D mode, and the width of the visibility zone, before these devices could hit the mass market.

Multi view auto - stereoscopic displays best suits to future scenarios and in the long period they would overcome stereoscopic ones for home environment.

In synthesis the main aspects for 3D home systems are: no need of glasses, MVV support and mass market forecast availability.

STBs & Interconnection

3D STBs are fundamental for the development of MVV home environment systems, the expected scenario leads to a multitude of different technologies displays from legacy 2D devices and any possible 3D stereoscopic displays to auto stereoscopic ones, each device supports a different input format. It's not feasible to send a specific video format for each screen, a generic one would be defined, then will be up to the STBs to receive the stream and encode it in the way supported from the specific display.

As seen nowadays 3D STBs are able to receive as input only side by side or over under format, and are able to convert these streams in any different 2D-3D techniques supported. In the future the streams may still rely on a side by side format or not. Adopting other 3D video formats as MVD or LDV or DES, and different encoding techniques as MVC or future 3DVC or so on, the STBs input stream will be slightly different and an old STBs would need some updates to be compatible to such streams.

It is not assured that an MVV transmission will have too much different possibilities, due too high bandwidth requested, then finally only one transmission format maybe will overcome the others, thus let suppose that a standard transmission format would be established helping the decoder side and uniforming the display technologies.

As concern interconnection, they will have to support transmission of each possible format stream, nowadays they are able to support a lot of stereoscopic types from spatial/temporal interleaved to side by side. In a MVV system they'll have to support multiple video stream input also. A future cable will have to support such every type of transmission 2D, 3D and MVV. It isn't expected that the MVV addictive functionality wouldn't be a tough challenge.

Chapter 5

Future MVV scenarios

On a basis analysis both stereo and multi view video formats require more views or, generally, more information than a common 2D video stream. As regard the stereo format, a trick can be exploited; as a matter of fact reducing the image resolution by half and transmitting it in side by side or interleaved format, it has the same bit rate weight of a standard HD channel, the only matter relies on the necessary signaling of the different type of stream and on the final display that must be able to manage such a stream.

As concerns MVV, instead, on an empirical assumption, the bitrate increases proportionally as the number of views grows up. The multiple views of the same scene, statistically have a lot of redundancy, since they refers to the same scene at the same time, this have to be exploited to reduce the overall amount of the stream bitrate.

Multi view video coding techniques are been developed to exploit this interview redundancy.

As already seen in the 3D video formats section, nowadays there are two multi-view encoding standard: MPEG-C and H.264/AVC with MVC extension.

H.264/AVC has been extended to MVC by ITU-T, and MPEG-C is a ISO/IEC standardization performed in 2007. Further is under investigation 3D video coding another standard that synthesizes video plus depth and multiple redundancy views approach.

AVC streams take advantage only of temporal redundancy, instead multi-view video coding (MVC) exploits also spatial redundancy, giving an addictive gain as regards bitrate. For transmission constraints, also for satellite communications, the least a stream weight the best it is.

On transmission segment not only bandwidth constraints have to be considered, also aspects as scalability and coverage must be analyzed. In order to this for the possible MVV stream a layered structure has been chosen and different scenarios for different weather conditions, have be taken in exam, using different code modulations.

Before any possible analysis the different scenarios have to be defined, as a matter of fact, the MVV stream can be used both for a broadcast scenario, a television transmission, or for an interactive scenario, as a video conference. Obviously the requirements for these two scenarios are slightly different.

5.1 3D TV Broadcast scenario

The 3D television broadcast scenario describes a possible future television environment that provides 2D, 3D stereo and MVV services. The advent of MVV services would be as a revolution and all equipments will be changed from shooting phase to end user displays.

The first aspect of this revolution, still before the definition of the scenario itself, regards the migration from 2D to 3D and MVV. As it is happening nowadays for the transfer to digital terrestrial television, it is always a tough race to develop a technological revolution. For this reason the basic idea is to provide a backward compatible service to existent 2D one. The concept is the same that have been used in the past, going from black and white television to color television, older equipments were still able to manage the new video streams in this way the passage have been done in a long period, without creating constraints to end user.

This backward compatibility requires that the MVV stream keeps a reference 2D-HD layer. The best suitable solution for the broadcast scenario is to assure any possible video input stream from a stereo camera one or a depth camera or a multi camera setup or old 2D camera streams, and transmit them with a format that support all of them, then at the receiver side decode the streams as requested by the devices locally available, using, if necessary, DIBR techniques. This scenario is depicted in fig 5.1.

In the fig. 5.1 different scenarios are possible, it is expected that shooting phase would be performed in different ways, and the following format will accord to it. The encoding would be the same for each format then at the receiver a STB must be able to decode those different input streams and make them usable for every type of display device, from old 2D equipments to stereo 3D to auto stereoscopic displays,

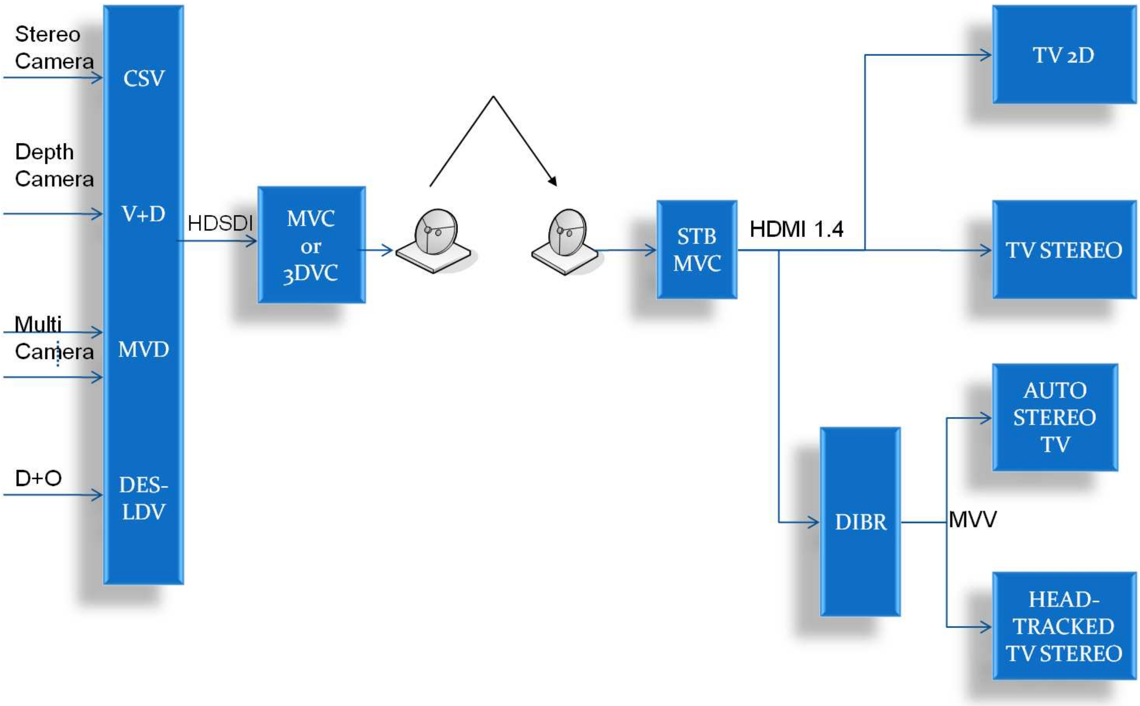

Figure 5.1: Future MVV Broadcast scenario

providing each type of video service.

The broadcast scenario focus on stereoscopic 3D and MVV, because other solutions seen before as volumetric or light field displays seem a too far solution till now, neither suitable for home usage because of too low spatial and temporal resolution, mechanics (scanning mirrors and rotating diffusers) too much complex and so too expensive for home devices.

In order to keep end user devices costs low, necessary for the development of 3D home systems, the re-rendering phase have to be left as much as possible to the encoder side, thus leaving all the preprocessing actions as lens correction, normalization of intrinsic parameters and rectification of convergent camera views, to capture or post producing phases. In this way at the receiver side only simple configuration steps are left, as parallax scaling and head motion parallax, that also gives a more immersive perception to the user, who, slightly moving his head, can see behind the scene objects.

Otherwise the interaction is limited, users can only control the depth reproduction, that in a 3D systems becomes what contrast or color adjustments are for a 2D video.

The depth image based rendering, that is up to the receiver side, is limited by specific display properties, (number of views supported), but it can also adds to the user some more interaction, he can adapt the 3D reproduction to his own preferences, as an example choosing less views to be decoded.

Auto-stereoscopic displays are necessary to get the MVV effect and at the same time old 2D equipments and conventional stereoscopic displays are able to support such a stream, thus providing a complete backward compatibility.

A further more functionality can be achieved using eye tracking auto stereoscopic displays, in this way instead of displaying all views at the time, such lowering the single view resolution, it's displayed only a view at the time at full resolution, and this changes as the viewer position changes.

As seen in section 2 the display width and the viewing distance are necessary to establish the interaxial camera distance, the value selected in the pre-producing phase might not be the same of the specific display at the receiver side. Depth data can help, in this case, to adapt to the specific situation and to the user preferences the views reproduced.

In synthesis it's evident that parameters bear on the final visualization are many, previous considerations become assumptions that have to be made in order to allow 3D video and delivery formats to be as generic as possible, to adapt to different viewing conditions and displays. The main assumptions are backward compatibility, usage of plano-stereoscopic devices, separating re rendering procedures as much as possible to pre processing phases, limited interaction to number and depth of views, finally end user will see 3DTV almost in a static position, moving its head a side to see behind the objects (more or less 50 cm horizontal moving).

5.1.1 Broadcast scenario: transmission stream and bit rate

The overall broadcast scenario have been presented, for the purpose of the project it is necessary to provide a possible scheme of the particular stream that have to be transmit, and of the bandwidth requested, to analyze if the whole system is affordable. As a matter of fact the main drawback of MVV is referred to much higher data to be transmitted respect standard 2D standard.

The analysis starts from backward compatibility assumptions, that is correlated with scalability aspects. If scalability is a goal to achieve isn't so sure; in a first analysis it can be assumed so.

These considerations lead to a layered structure, which could allow different levels also for 3D stream, providing at the same time a conventional stereo approach or more complex MVV formats.

This structure is easily obtained from DES format description, as depicted in fig 5.2.

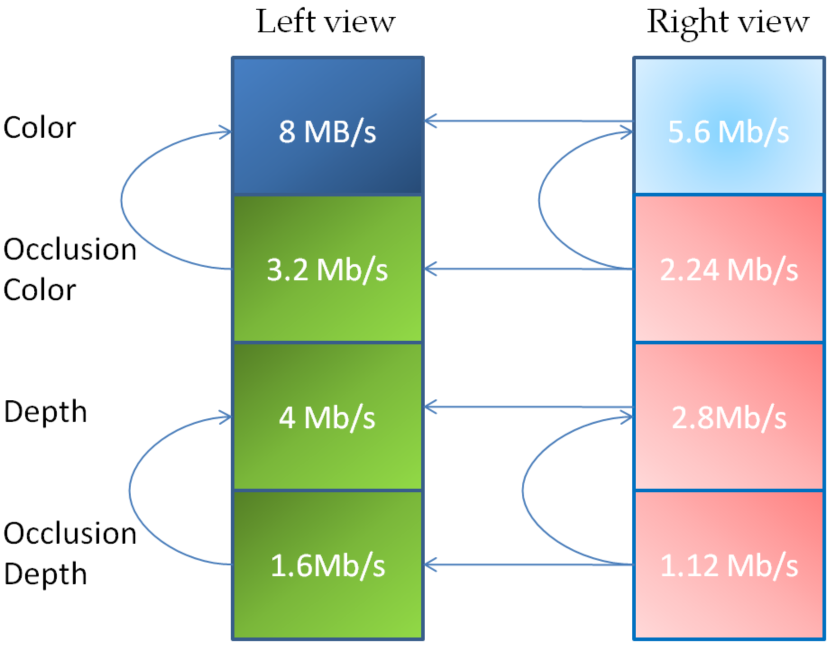

Figure 5.2: Layer provided by DES format

In fig. 5.2 is presented the DES format with the bitrate weight of each part. To exploit interview redundancy video stream and depth stream must be separate, leading to two independent MVC streams. The arrows in the figure represent the interview dependencies, it means, left video view is the base view, the right video view is encoded depending on the left view, the occlusion information is encoded starting from the left view, finally the right occlusion have redundancy both on left occlusion and right video image.

For depth information the route is the same, stating left and right depth views, respectively as left and right video views.

Four different layers can be obtained from the whole stream, choosing the sections of the DES structure:

- 2D HD stream
- 3D conventional 2-view stereo HD = 2D left view + 2D right view

- 3D Layered Depth video = single 2D left view + 1 occlusion color + 1 depth + 1 occlusion depth
- 3D Full-DES (Depth enhanced stereo) = All

The first level represents the base layer that allows the 2D conventional visualization, the second layer provides the conventional 3D stereoscopic and is obtained transmitting the two base views. The third level follows the LDV scheme and is achieved using the only left view plus its occlusion color map, and its depth and occlusion depth map.

Finally the last level is the one that provides the best final 3D image quality, it is achieved transmitting all streams, the two video views with their occlusion information, and the two depth maps with their occlusion information.

Only the third and the fourth level allows a MVV visualization, thanks to depth information, that gives the possibility to create other views next to the one directly encoded.

Mixing, opportunely, the various level different formats can be obtained, from a base 2D-HD stream, to LDV or DES, or Conventional 3D. The various combinations are shown in table 5.1:

3D format	Description	Video bit-rate (Mbit/s)	Audio kbit/s	Raw Data input (2.3 % encapsul. overhead)
Full DES	All components	28.56	384	29.61 Mbit/s
CVS	2 Color views	13,6	384	14.31Mbit/s
Mono V+D	1View +1 Depth	12	384	12.67 Mbit/s
2view MVD	2 view + 2 Depth	20.4	384	21.26 Mbit/s
LDV	1 occl. + 1 occl. depth	16.8	384	17.58 Mbit/s

Table 5.1: Different level combination

The encapsulation overhead refers to the overhead introduced if a GSE (Generic stream encapsulation) encapsulation is used. This encapsulation may be used to make an IP stream compatible to MPEG2-TS with all its capabilities, as segmentation and recovery mechanisms.

As a matter of fact, there are two possibilities for the structure of the transmitted stream: it can be sent as a MPEG2-TS over DVB, that is compatible with existent architecture as regards decoder and STB, or as an IP stream encapsulated over GSE, that is able to emulate MPEG2-TS streams, but goes towards an all-IP architecture where Internet satellite transmission and so on are compatible each other.

At a first sight it's evident that a whole DES stream requires too much bandwidth for nowadays satellite's capacity, due to high costs and capacity itself. Improvements in coding techniques or in satellite's capacity have to be made before this system can hit the mass market.

The goal of 3D multi-view scalable video coding is to transmit video content by different layers on the same carrier, enabling to reproduce it on each display, with the best possible quality, depending on the transmission conditions. The transmission of a base layer and enhancement layers allows to use unequal loss protection, using different FEC protection or modulation or power level.

Some considerations have to be made to best configure the scenario. The stream structure presented has the advantage to be backward compatible both to old 2D equipments and to stereoscopic 3D environments, this is surely from one side a great goal to achieve.

On another side, if scalability capabilities have to be exploited to assign different availabilities to each layer, is not an issue. As regards differents availabilities broadcasters and end users requirements collide, as a matter of fact if, from broadcaster side, using lower availability for higher level means bandwidth gain and then reduced costs, from the user side having differents availabilities isn't so acceptable. In a first analysis for two main reasons, firstly a MVV service is expected to cost to the user more then a 3D or a 2D service, then, if the user pays for a MVV service, he would be assured to see it, secondly the switching from MVV to 3D and especially from 3D to 2D back and forth can be very tiresome and can lead to eyestrain.

It can't be assumed that bad weather conditions might block a MVV transmission, and in variable conditions the service falls and restarts continuously, the arguing is still in progress and a final trade-off will be achieved only when practically a MVV system will be developed.

5.2 Interactive scenario

The interactive scenario describes a possible future 3D video conference service. It differs from the broadcast scenario, because its requirements are different, first of all the transmission is bilateral, constraints on delay are strict and less final quality image is acceptable. Nowadays there isn't a commercial 3D videoconferencing system available nor it is expected in the near future. Therefore the analysis covers only theorical aspects and constraints, waiting for future developments.

It is assumed a three sites environment, with a one to one site transmission and two users at a site, then are necessary 4 video plus 4 depth maps streams to each remote site. This scheme is an ongoing project, called 3D presence, that has the aim to develop a system prototype.

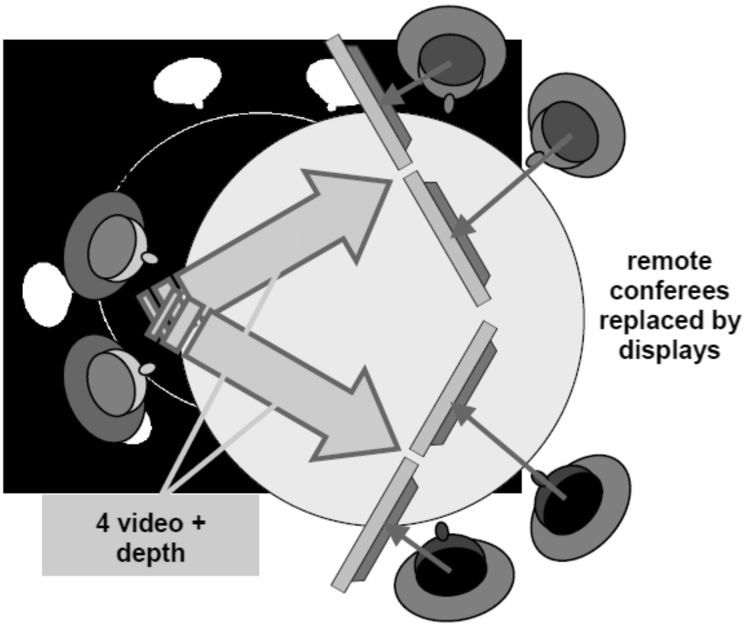

Figure 5.3: Video conference scenario

The overall bandwidth requested for such many streams is great, but taking in account that a low quality is acceptable, bit rate can be lowered transmitting the streams at a quarter resolution, in this way in a single HD video stream are collected two video and two depth layer each of size (960x540) as depicted in fig 5.4.

This solution allows to reduce the overall bandwidth to a quarter, leading to the sequent upper bound:

$$\begin{aligned} 20.4/4 &= 5.1 \text{ Mb/s} \quad \text{for a 2 video plus 2 depth maps} \\ (20.4*2)/4 &= 10.2 \text{ Mb/s} \quad \text{for a 4 video plus 4 depth maps} \end{aligned} \tag{5.1}$$

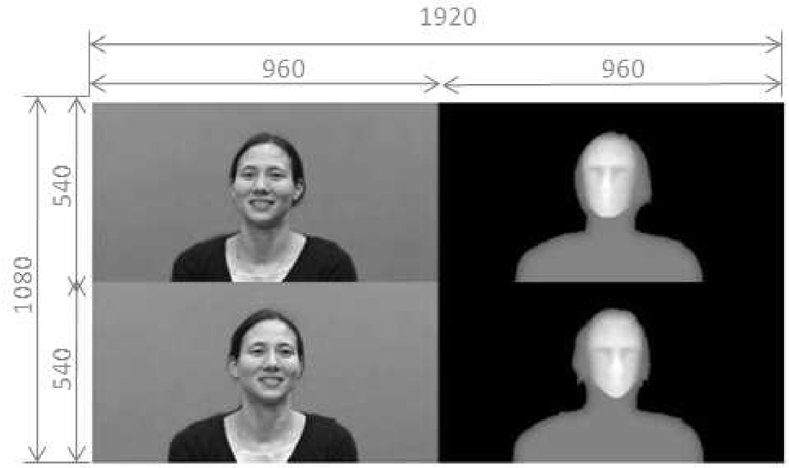

Figure 5.4: Video conference stream

The last main aspect regards the delay of the transmission, due to service necessities both video processing and transmission have to be performed with low delay. From ITU-T reccomendation (ITU-T Rec. G.114) for end to end transmission time in one direction delay for application establishes:

- 0 to 150 ms is acceptable for many user applications
- 150 to 400 ms is acceptable if delay influence on the performance is known
- over 400 ms is unacceptable

For a GEO satellite round trip times typically is around 600 ms, without evaluating the image compressing and decompressing delays; this value is quite higher than the 400 ms requested by ITU, then they aren't suitable for this scenario. LEO/MEO satellites would be able to support these applications mainly thanks to their nearer orbits. The round trip time could be around 140 ms within ITU reccomendation.

5.3 Future MVV systems spread

For interactive scenario bandwidth necessities and costs are still the main constraints, that make this service no more feasible and therefore studies on this side have been stopped for the related project. Also if delay requirements can be achieved, in the next future, with recurring to LEO/MEO satellites.

For broadcast scenario the migration will need a long time and a great revolution in devices and procedures on each side of the whole chain. A lot of research is still needed especially as regards end user equipments, shooting procedures, encoding techniques and real time encoders. Furthermore satellite capabilities have to increase lowering constraints on these side.

Nowadays more and more broadcasters are involving in 3D television channels. Open-Sky itself is providing, in collaboration with Eutelsat, as a part of an ESA project, a 3D demo television channel. From this demo channel can be performed an in-depth analysis of the gap requested to achieve MVV end to end systems.

The general scheme of the 3D channel provided by Open-Sky is shown in fig 5.5.

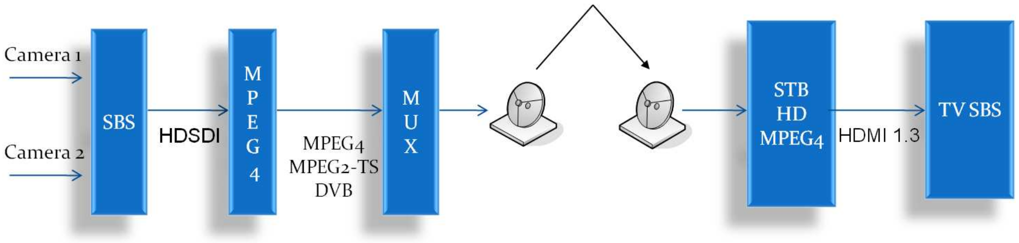

Figure 5.5: 3D today scenario

The capture is performed by two cameras, the video streams are encoded in a side by side format, reducing by half the resolution. Then the stream is transmitted as a standard HD channel, this allows no extra bandwidth requirement. Finally at the receiver side the stream is decoded and the final visualization is left to a 3D display that support side by side format, such a stream hasn't backward compatibility to old 2D equipments, if not only reducing the final quality 2D image.

This architecture has been designed in order to modify existent 2D systems as least as possible. Specifically, the stream is transmitted and received as a standard 2D HD channel, then, if an end-user has 2D display, he would see the 3D channel in side by side format and the conventional 2D streams without drawbacks.

Comparing this architecture to MVV one, depicted in fig 5.1, furthermore considerations are possible.

These designs have the same philosophy for some aspects, each possible stream can be encoded, transmitted and displayed in order to the final end-user equipments. Further requirements rely, mainly, on the transmission segment. The encoding format must be able to collect all different services (2D, 3D and MVV) and has to provide a stream for the decoder that can be reproduced in order to the desired or needed service (2D, 3D stereo or MVV).

The best service (MVV) obviously requires a complete update of end user devices, but other consumers without interest in this service would get 2D or 3D stereo services as before without noticing any difference. This philosophy is pretty different from digital revolution, that recently involved Italy consumers. In this way all added costs are let to broadcaster and to content producers (especially for shooting procedures), instead at user side everyone would chose the preferred service buying the adapt device for it.

It is not sure that MVV service would spread the market then a commercial analysis is necessary to realize if provide such a product would be feasible to the more costs and complexity requested by the MVV architecture, as regards bandwidth requirements and capture added costs and other already seen added costs.

Chapter 6

Further project phases

6.1 Simulation Campaign

After the review of all 3D technologies and the design of a possible MVV scenario as regards a broadcast scenario and an interactive scenario, 3D@Sat project goes on with a simulation stage and with a commercial analysis of the MVV designed system.

These phases will be deeply described in *Christian Fiocco*'s thesis that reports the work performed in the second part of the project.

The performed simulation involved both transmission channel and final quality visualization subjective tests on nowadays displays.

As concern channel simulation, Astrium provides the SATEM tool, a tool able to simulate the performances of the satellite link, evaluating different weather conditions, fading and various MODCOD selected for the transmission, in a first analysis.

On display side, the work is more difficult because of the lack of MVV displays available and moreover of real time MVV encoders. The simulation in this branch is attempted in two parts, the first is in real-time, and will regards only the base stereo layer, using a stereoscopic display, the second will be in off-time using an auto-stereoscopic displays and prerecorded and decoded streams. The MVV effect will be simulate putting different people at different angles, and showing them different views in different time.

6.1.1 Channel simulation

The channel simulation has been performed at IP level, the SATEM tool emulates DVB-S2 satellite link behavior through a data-gathering campaign.

The simulation evaluates bitrate, delay, jitter and packet loss parameters, varying traffic conditions, fading, MODCODs and layer of the stream sent.

6.1.2 Decoding simulation

For the real time display simulation, have been used a stereoscopic display, where have been reproduced a trailer of a race car, this isn't relevant for an MVV display, but it had been necessary to make a comparison between 3D and MVV.

The real time simulation is more strict linked to MVV project. Using an auto-stereoscopic display, the MVV effect has been reproduced and some subjective tests have been performed to viewers.

6.2 Commercial analysis

The final commercial analysis aim is to describe how each scenario hit the market, drawing a road map of future MVV possible systems and their impact on overall architecture, from satellite capacities to end user devices needed, from capturing new technologies costs and complexity to contents production and application fields.

The analysis will go through these steps:

- Network complexity, the overall additional complexity of the whole structure, especially for those aspects regarding transmission and for Esa interest on satellite segment.

- Costs impact: impact of increased costs on each side of the chain, from shooting to transmission and visualization.

- SWOT analysis : Strengths, weaknesses, opportunities, and threats: an analysis of the product made before it reaches the mass market.

- Development & future outlet: looking at backward compatibility, reference market expected and road map on how to spread this technology.

Chapter 7

Conclusions

7.1 Resume MVV end to end system and 3D state of the art

The whole end to end MVV architecture designed follows the scheme depicted in fig 5.1. The main effort of the framework is to be backward compatible with legacy devices, reuse of existent equipments is pretty affordable. Thus allowing reduced costs and complexity. The second main effort relies on transmission segment, where bandwidth constraints are quite consistent; MVV streams comprise much more bandwidth then usual 2D HD channel. Therefore efficient coding techniques with high compression data, exploiting also interview and temporal redundancy, are needed.

In the outlined architecture backward compatibility and reuse of existent devices is quite assured. In this way older equipments still provide old services without any drawback.

On shooting side costs increase as the service goes towards stereo or MVV services, some improvements are needed to simplify shooting procedures, till now too long and complex as the camera number and type (depth or occlusion cameras) increase. Stereographers are figures nowadays becoming pretty important for camera configurations, moreover some devices, that can help these phase, are under development, as stereo image analyzer, furthermore some improvements on rigs mechanics are necessary.

The parameters, that influence the quality perception of the user, change, in the 3D era depth scaling and zero parallax plane become as current color or brightness adjustment. These information has to be taken in account in the production and delivery formats, that must be as generic as possible, in order to be supported by the most disparates possible displays, but have to follow some common guidelines to support the generic video stream generated.

Capture phase is also tied up by the video format chosen for the video stream. The selected format influences number and type of cameras. Various formats are possible in order to achieve best quality image and lowest stream bitrate.

The main approaches rely on multi video capture or on video plus depth join with DIBR techniques. MVS requires high shooting costs and complexity, stream bitrate are quite high and can be partially reduced exploiting interview correlation, recurring to MVC encoding.

Otherwise going through a multiple video plus depth approach bit rate decreases due to less number of views shoot, but depth map estimation increases encoding, shooting and decoding complexity.

Exploiting DIBR techniques less views must be transmitted to achieve the final desired number of views, thus consistently reducing bitrate.

Furthermore LDV and DES still increase the overall complexity, thus providing high bitrate gain; manage occlusion maps brings to error prone decoding, but they achieve the best final image quality.

MVD, LDV and DES seem to be the best video format for MVV contents, the final choice hinges on the trade off allowed between complexity, bandwidth and final image quality.

Nowadays MVC and MPEG-C are the two standard encoding techniques available, MVC, that is implemented as an extension of H264/AVC, exploits temporal and interview redundancy, achieving consistent bitrate gain, it still provides base layer for a 2D stream and standard stereoscopic 3D streams.

Otherwise, MPEG-C implements video plus depth, it doesn't exploit interview redundancy, but through depth map it lowers the number of views to be shoot and transmitted, respect to MVC streams.

The next step comprises a synthesis of these two standards, it's under studies the 3D video coding that will support multi video plus depth stream exploiting at the same time interview redundancy and depth information.

For the project developed, a DES format has been chosen because it leads to a layered structure, that provides backward compatibility and, if necessary, scalability.

Backward compatibility is required in order to avoid a complete revolution of end user equipments, just performed for the migration to digital television, furthermore because it's expected that conventional 2D contents will continue in parallel to 3D and MVV services. Scalability issue is still under discussion, before the service is provided, it can be a consistent advantage to reduce bitrate, in order to the desired quality service, instead once the service has been bought, it must be assured, for two main reason: user strain and commercial requirements.

Anyway the video stream proposed in the project is able to support, if wanted, both scalability and backward compatibility.

The stream is encoded in two MVC streams, one for the video images and one for the depth maps, eventually the two streams are separate. Four levels are established: a base layer for old 2D streams, a step up level for conventional stereoscopic 3D, the third level is for LDV format 1 views plus one depth and occlusion map. Finally the highest DES level, 2 views with 2 depth and occlusion maps. As seen, the highest level has a high bitrate of 29.6 Mb/s, that isn't feasible for nowadays satellite costs and constraints.

Further improvements on satellite capabilities and on encoding techniques are needed to lower the overall bitrate or to increase bandwidth available.

As concern end user side, development and research are always in progress, MVV can be obtained through eye tracked stereoscopic displays, or auto-stereoscopic displays, instead light-field and volumetric devices are still more far solutions.

Nowadays a number of stereoscopic devices are been produced and more and more are reaching market at lower prices. These devices implement many technologies but they seem still not affordable for home environments, due to the wear glasses requirements.

Auto-stereoscopic devices seem more feasible for home environment, they provide more than a view at a time, moreover through optical principles as diffraction, refraction, reflection and occlusion they drive the left-right images to the user, without glasses requirement. Firsts samples are ready for the market, but their prizes are still high and most of them aren't available in Europe yet. Further improvements on width and visibility zone, on the overall image quality and on the display resolution in 2D mode are expected, before these devices would hit the mass market.

Light-field and volumetric displays are still a far solution, they are still only in form of prototypes therefore they are not considered for the MVV architecture designed. Especially volumetric units will completely change the final video perception, and they would give a completely different experience.

As regards interconnections, HDMI 1.4 and display port 1.2, last released, are ready for every possible stereoscopic devices, not much work is needed to support MVV services.

Finally STBs are still in their infancy, a lot of improvements are needed, in order to allow a complete inter compatibility to every display type and incoming video stream. They would manage a generic input video stream and decode it in order to 2D, stereoscopic or different auto-stereoscopic displays.

Nowadays 3D is left to side by side format, as provided by the 3D demo Open Sky channel and by other broadcaster or content producers, thus providing stereoscopic 3D with the same bitrate required by an HD channel, but losing resolution and therefore quality. In the future this format is expected to be set aside, on behalf of other ones that will require much more bandwidth but also, best final quality image.

7.2 MVV prospectives

MVV isn't expected to reach mass market within 10-15 years the main reasons refer to too high costs on each side of the chain, end user, shooting and transmission, to the still low presence and quality of auto stereoscopic displays, to the too high bandwidth required for the stream, not supportable by today satellite capabilities and finally to lack of suitable contents.

A lot of research are needed on many aspects from depth estimation, to shooting procedures, from end user displays to encoding techniques.

Once the technology will be ready and the overall architecture would have an affordable prize, the issue will move towards what type of services would be adapt for MVV services. There are two directions: what can be shoot and displayed in multiple views and what would be interesting and cool for end user.

The scenarios go from pure commercial use to business one, both rely on different necessities and interests. MVV would be a very step over in television and video communication fields, or would only be a transient trend? This is the real question. In a first approximation in an auto stereoscopic display, reproduce N views at a time means that each view's resolution is scaled by a factor of N, it isn't an issue that this resolution loss could be acceptable to achieve a best immersive experience. From a technical point of view, steps made towards MVV would be useful also for conventional 2D and 3D video transmission and also for future other possible technologies. Otherwise fields of application and way to exploit MVV have to be found in order to assure future utility for it.

MVV trade off is between on one side increased costs and resolution loss, and, on the other side, on more immersive experience or new idea of video productions.

MVV leads to a powerful new tool in the hands of contents providers to give a completely new service to the user. More views at the time mean more scenes to be managed at the time. At a first sight the obvious application is give to the user a wide perception of the scene, moving himself, he can see other parts of the scene, as an example he can see behind an object or a character, feeling to really be in front of the scene. This regards movies or this type of service. As concern live events as concerts or sport matches, the perception of slightly different scenes, achieved looking from various angles, could give the perception to really be in the place where the event is performing. Than the question is, is this step up enough for the user?

Obviously other implementation can be drown for MVV, for example, taking back the idea of some DVDs, where more ends are possible, in a movie various ends can be displayed in accord to the viewpoint of the user, or, otherwise, in some views can be clues or details that can be hidden in other views due to occlusion objects in the scene.

Furthermore completely different videos could be seen by different users, at the same time, on the same television. The wife looks at a movie and, on the same time, the husband looks at the football match, obviously for these last services some great constraints come up as concerns audio streams. These applications figure out as completely different to conventional MVV philosophy but show the potential of these technologies.

Many possibilities are offered by MVV, the right ones have to be followed.

As regards business applications, there are other considerations, especially as regards medical and military environment MVV offers great applications. As an example, perform a medical operation connected with other experts in other parts of the world, that can see the patient and his wounds from different angles.

These are only some suggestions, before these steps, but also in parallel on these discussions, the technological architecture must be developed.

Furthermore not only contents change in a MVV system, as a matter of fact it leads to a completely new way of living television. From one side auto stereoscopic displays remove the necessity of wear glasses (respect to stereoscopic 3D) but on the other side, projecting left-right images in specific points of the viewer's space, coerce user to see television moving only in a limited range and in certain positions. Nowadays "live" television is pretty different, user can move around or lay down without having any drawbacks. Figuring that television is mainly used to relax, MVV introduce some constraints that can be not acceptable for the end user, not even for this more involving service.

The expectation is that MVV would effectively give a more immersive experience, and wouldn't be a flash in the pan, but till now it is also a quite far solution, as already said 10 or 15 more years are necessary for MVV to be ready, a lot of work still have to be performed.

Bibliography

[1] H. M. Ozaktas and L. Onural, *Three dimensional television.* Springer, 2008.

[2] J. Canny, “Design realization : lecture 27,” 2003.

[3] Q. Wang and alii, “Stereo viewing zone in autostereoscopic display based on parallax barrier.” IEEE transaction.

[4] R. Yang and alii, “Toward the light field display: Autostereoscopic rendering via a cluster of projectors.” IEEE transactions, 2008.

[5] O. Schreer, P. Kauff, and T. Sikora, *3D video communication.* Wiley, 2005.

[6] B. Lee and alii, “Status and prospects of autostereoscopic 3d display technologies.” IEEE transactions, 2007.

[7] C. Chinnock, “3d coming home in 2010.” online documentation, 2009.

[8] H. Liao, K. Nomura, and T. Dohi, “Long visualization depth autostereoscopic display using light field: Rendering based integral videography.” IEEE transactions, 2002.

[9] B. G. Blundell and A. J. Schwarz, “The classification of volumetric display systems: Characteristics and predictability of the image space.” IEEE transactions, 2002.

[10] T. Annen and alii, “Distributed rendering for multiview parallax displays.” IEEE transactions, 2005.

[11] D. E. Roberts, “History of lenticular and related autostereoscopic methods.” IEEE transactions, 2003.

[12] K. Balasubramanian, "On the realization of constraint-free stereo television." IEEE transactions, 2004.

[13] D. F. McAllister, "Display technology:stereo & 3d display technologies," 2001.

[14] B. Brubaker, "3d and 3d screen technology," 2008.

[15] P. Merkleand and alii, "Coding efficiency and complexity analysis of mvc prediction structures," 2007.

[16] P. Surmanand and alii, "A roadmap for autostereoscopic multi-viewer domestic tv displays." IEEE transactions, 2006.

[17] Y. Zhu and T. Zhen, "3d multi-view autostereoscopic display and its key technology." IEEE transactions, 2009.

[18] "Design realization lecture 27." Online available, 2003.

[19] "Avatar movie: history of 3d cinema." Online available http://www.telegraph.co.uk, 2009.

[20] "History of 3d." Online available http://www.21stcentury3d.com/historyof3d.html, 2009.

[21] "History of 3d." Online available http://www.sensio.tv, 2009.

[22] "Technical notes 1.1, 1.2, 2.1, 2.2, 3.1, 3.2 of 3d@sat project," 2009.

[23] D. N. Wood and alii, "Surface light fields for 3d photography," 2000.

[24] T. Funkhouser, "Overview of 3d object representations," 2002.

[25] K. Mueller and alii, "An overview of available and emerging 3d video formats and depth enhanced stereo as efficient generic solution." IEEE transactions, 2009.

[26] Y. Jeon and alii, "Analysis of efficient coding tools for multi-view and 3d video." IEEE transactions, 2009.

[27] H. Kwon and alii, "Avc based stereoscopic video codev for 3d dmb." IEEE transactions, 2005.

[28] D. Park and alii, “Lenticular stereoscopic imaging and displaying techniques with no special glasses.” IEEE transactions, 1998.

[29] Y. Zhu and T. Zhen, “Parallax polarizer barrier stereoscopic 3d display systems.” IEEE transactions, 2005.

[30] H. Yamanoue, “The differences between toed-in camera configurations and parallel camera configurations in shooting stereoscopic images.” IEEE transactions, 2006.

[31] J. E. Cutting, “How the eye measures reality and virtual reality,” 1997.

[32] C. Fehn, “Depth-image-based rendering (dibr), compression and transmission for a new approach on 3d-tv.” SPIE, proceedings of, vol. 5291, 2004.

[33] S. Pehlivan and alii, “End to end stereoscopic video streaming system.” IEEE International Conference on Multimedia and Expo, 2006.

Ringraziamenti

Prima di tutto voglio ringraziare il mio relatore Gianfranco Cariolaro, per la sua guida durante il periodo di tesi e per avermi presentato ad OpenSky, consentendomi di svolgere un tirocinio molto interessante.

Un ringraziamento speciale a Open Sky che ha reso possibile questo lavoro e mi ha regalato una fantastica esperienza.

Ringrazio tutti i colleghi dell'ufficio per avermi aiutato durante questo periodo, rendendo leggera ogni giornata lavorativa.

In particolare sento di dover ringraziare l'Ing. Luca Carniato, che mi ha insegnato molto in questi mesi e che oltre ad essere stato un ottimo tutor si è rivelato importante anche a livello personale.

Ringrazio la mia famiglia: i miei genitori e mia sorella che hanno sopportato in silenzio il mio nervosismo e mi hanno sostenuto in ogni momento di difficoltà.

Ringrazio Silvia Saccardo che, pur trovandosi a un oceano e un continente di distanza, ha saputo starmi vicino, sostenermi e incoraggiarmi più di chiunque altro.

Un'amica di cui semplicemente non posso fare a meno.

Ringrazio Cosetta Masi, un'amica come ne esistono poche, su cui so di poter sempre contare.

Ringrazio Roberto Forestan, l'unica persona sempre presente nei momenti piú bui, nonché ottima spalla nella maggior parte delle avventure da me intraprese.

Ringrazio il CdSA un gruppo di amici fantastici, negli anni sono stati un bastone solido, senza il quale avrei faticato a stare in piedi.

Ringrazio tutti i miei amici, che mi hanno regalato bellissimi momenti e mi fanno sempre sentire importante.