

D1.2

|  |  |  |  |
| --- | --- | --- | --- |
| Action | Name | Date | Approbation |
| Author |  |  |  |
| Approved by |  |  |  |
| Approved by | Project Coordinator  Issa Rakhodai |  |  |

History

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Version | Author | Modifications |
| 05/04/2011 | V01 | C. Celorrio | Initial Version |
| 09/05/2011 | V02 | C. Chamaret, X. Ducloux, J.Casal, P. Andrivon, D. Piepers, P. Perez, S. Degonde, M. De Geyter | Added Technicolor’s contribution. Added TVN’s contribution. Added SAPEC’s contribution. Added Phillips’ contribution. Added Vitec’s contribution. Added Pace’s contribution. Added ALU’s contribution. Added UPM’s contribution. Added VRT-medialab’s contribution. |
| 12/05/2011 | V03 | E. Paris, J. Vounchk | Added UNIZAR´s contribution, Added Telenet´s contribution |
| 17/05/2011 | V1.0 | C. Celorrio & I. Rakhodai, T. Gruszka | First Version |
| 15/09/2011 | V1.1 | C. Celorrio | Minor updates |
| 30/01/2012 | V1.2 | C. Celorrio | Update with Paris F2F Meeting Input |
| 24/04/2012 | V1.3 | C. Celorrio & I. Rakhodai, T. Gruszka | Update with Zaragoza F2F Meeting Input |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Table of contents

[1 Glossary 4](#_Toc327175117)

[2 Introduction 5](#_Toc327175118)

[3 JEDI Architecture 6](#_Toc327175119)

[3.1 Architecture Workflow 6](#_Toc327175120)

[3.2 Architecture Value Chain Overview 8](#_Toc327175121)

[3.3 Detailed Description of Architecture 9](#_Toc327175122)

[3.3.1 Capturing 9](#_Toc327175123)

[3.3.2 Processing 12](#_Toc327175124)

[3.3.3 Contribution 16](#_Toc327175125)

[3.3.4 Coding 19](#_Toc327175126)

[3.3.5 Distribution 20](#_Toc327175127)

[3.3.6 Decoding 23](#_Toc327175128)

[3.3.7 Rendering & Display 25](#_Toc327175129)

[3.4 JEDI Architecture Workflow and Use Cases 27](#_Toc327175130)

[3.4.1 JEDI Content Delivery Workflow 27](#_Toc327175131)

[3.4.2 GRAPHICS & Interactivity Workflow 34](#_Toc327175132)

[4 Conclusion 38](#_Toc327175133)

# Glossary

To assure coherent terminology and abbreviations across all documents inside the project, the specific terminology and abbreviations for this deliverable has also been aggregated into an internal report called JEDI Terminology and Abbreviations available on the internal project website.

# Introduction

The objective of this document is to introduce the architecture of the JEDI platform that will be latter be described in the work packages WP2, WP3 and WP4.

This deliverable aims at providing information that includes:

* Describing the high-level view of the JEDI Architecture,
* Defining the workflow that comprises the end to end chain in JEDI,
* Explaining the different processes that are performed in each step of the chain,
* Associating and instantiating the JEDI workflow according to the use cases enumerated in previous D1.1.

# JEDI Architecture

High level description of the JEDI Architecture and the workflow of elements associated to it.

All the different modules will appear with their inputs and outputs.

## Architecture Workflow

In order to take a closer look at the JEDI Architecture, we will firstly describe the whole chain of activities that can appear in JEDI.

Capturing

Processing

Rendering & Display

Distribution

Decoding

Coding

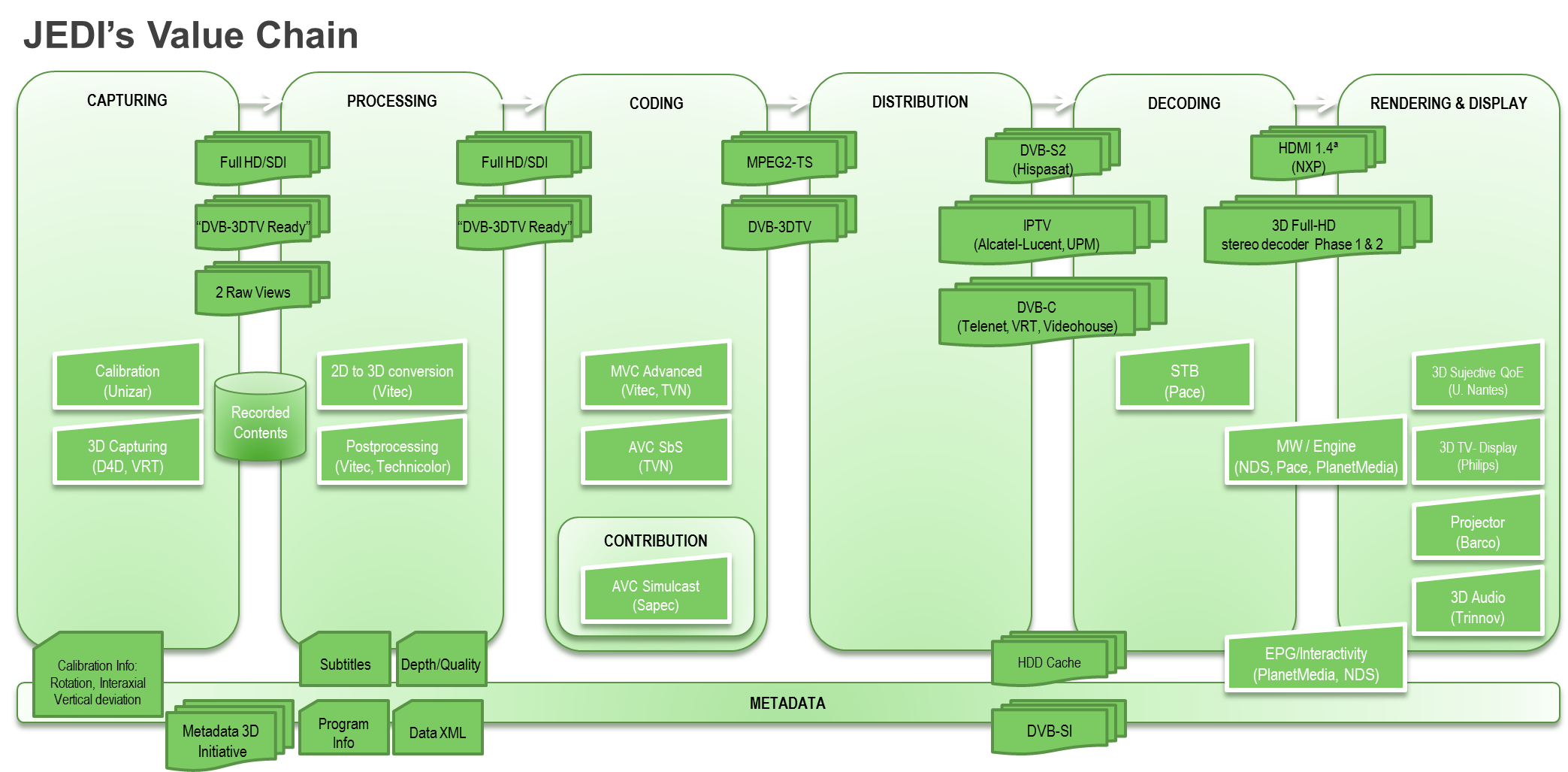
Contribution

JEDI Workflow.

The JEDI Architecture include the following activities over the end-to-end value chain:

1. **Capturing.** Process of obtaining the 3D scene and making it available to the system.
2. **Processing.** All the steps invloved in dealing with the digital format of the 3D capturing image, including post-production.
3. **Contribution.** Before being distributed to the end-user, transmission from capturing to production is usually needed in most of the scenarios. This transmission, known as contribution.
4. **Coding.** The 3D-stereo content, which has been previously captured and pre-processed, needs to be compressed in order to be transmitted to the end-user over existing networks.
5. **Distribution.** The distribution of the 3D streams comprises all the steps which are necessary to deliver the encododed content to the decoding system, with a sufficient degree of quality.
6. **Decoding.** Once it has been distributed, the 3D-stereo content is decoded. The decoding process consists of performing, to the extent possible, an inversion of each stage of the encoding process.
7. **Rendering & Display.** In the JEDI end to end chain, the 3D rendering part applies to the last part of the architecture. It includes all components that convert the received 3D streams in pictures displayed on TV or projectors

## Architecture Value Chain Overview



## Detailed Description of Architecture

### Capturing

The most common approach to professional broadcast and cinematic 3D capture is to use dedicated rigs and control systems that allow two standard HD cameras (one for each eye) to be used together. The most appropriate rig and camera layout varies according to the type of cameras used and the requirements of a particular project, studio, or location. The easiest way to shoot stereoscopic footage is by simply mounting two identical cameras next to each other. However, many rigs use mirrors or beamsplitters to pass images to the cameras, particularly where the camera lens or body is too wide to provide an appropriate offset distance (the distance between the lenses) for the scene. Furthermore, it is possible to shoot with parallel cameras or to let their lines of sight converge around the object of interest. The former requires reestablishing the depth plane in post-production, while the latter may demand for geometric corrections to solve possible keystoning problems.

Professional cameras may record in a variety of formats. Most of the time, compression (somewhere in the 50-100Mbps range) is applied when recording on the cameras storage medium. Often, it is also possible to output uncompressed footage via HD-SDI.

Special attention must be drawn for calibrating cameras all together as acquired videos streams will be processed. Quality processing and resulting user experience will depend on the calibration quality as this process stands the most upstream and then it will impact the rest of the processing chain.

* + - 1. Capture

|  |  |
| --- | --- |
| Capture | |
| Pre-condition and inputs | A subject to shoot |
| Process Description | Two image streams are captured to create a stereoscopic experience. It is absolutely necessary that all parameters of both cameras (white balance, aperture, focal length) except for their offset in horizontal position match |
| Outputs | 2 image streams, one for left and one for right eye. High bitrate compression may be applied |
| Related processes | All other processes are related, since they use 3D content |

* + - 1. 3D content acquisition: calibration modules

|  |  |
| --- | --- |
| Video streams calibration | |
| Pre-condition and inputs | Calibration modules inputs are couple of raw video streams. |
| Process Description | This module aims at preparing proper 3D video stream processing. Real-time calibration modules are cascaded. Indeed, it respectively stands for:   * Stereo geometric calibration, * Stereo rectification, * Colour calibration   of video streams coming directly from cameras. |
| Outputs | Outputs are calibrated raw video streams. |
| Related processes | This module is of utmost concern for correct depth map estimation and should be settled upstream every stream processing. |

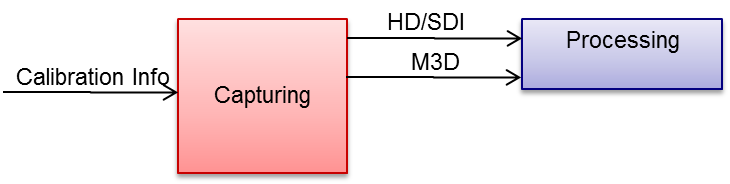
* + - 1. 3D rig calibration module

|  |  |
| --- | --- |
| 3D rig calibration | |
| Pre-condition and inputs |  |
| Process Description | This module calibrates the 3d rig for all stereoscopic parameters, this process is realized automatically. This process gives a perfect cameras synchronism  Optical sync (zoom and focus)  3D rig parameters sync  - Interaxial  - Convergence  - Vertical deviation  - Rotation between cameras |
| Otuputs | Stereoscopic video fully synchronized (zero position), L image captures the same image that the image R, for each of the zoom positions. |
| Related processes | This process is very important because if calibration of the rig is not correct, the two videos captured will have significant gaps between them. These gaps are impossible to correct in following processes. |

* + - 1. 3D capturing

|  |  |
| --- | --- |
| 3D capture process | |
| Pre-condition | -3D rig calibration  -Distance to the nearest point of filming and the farthest  -Type of display screen (dimensions) |
| Description | 3D video capture  From previous 3D rig calibration and the values ​​of screen size, closest point and distant of shooting, the convergence is calculated and interaxial appropriate for the recording of the scene and proceeds to 3D video capture. |
| Format | 2 Raw video streams , HD-SDI |
| Related processes | 3D rig calibration |

* + - 1. Interprocess formats



* Calibration Info
  + Rotation
    - Eliminates the rotation component introduced in the process of mounting cameras on the rig.
  + Interaxial
    - Corrects the horizontal offset between near and far objects.
  + Convergence
    - Corrects horizontal gaps between near and far objects.
  + Vertical deviation
    - Corrects all vertical deviations (offsets and gaps)
  + Other parameters
    - Focal length, Focus point, Max/Min Distance
* Video Contents
  + 2 Raw Views
* Video Formats
  + Full HD/SDI (Format YCbCr422)
  + MPEG-TS
* Metadata Formats
  + M3D: Metadata 3D Initiative
  + DVB-3DTV and DVB-SI
* Metadata Contents
  + Program/Info
    - Content providers
    - Channel owners
  + Subtitle
    - Content providers

### Processing

Once images are acquired from cameras rig and calibrated, a pre-processing toolbox computes image processing algorithms in order to enhance depth feeling according to user’s screen scaling and viewing distance and to adapt the content to specific 3D screen . Thus, user 3D experience is optimized according to its device rendering technology..Pre-processing toolbox: depth map computation.

* + - 1. Depth map computation

|  |  |
| --- | --- |
| Depth map computation | |
| Pre-condition and inputs | This module accepts raw calibrated video streams (at least two). Algorithm can be tweaked with user settings. |
| Process Description | Calibrated video streams couple are processed in order to perform in real-time an efficient depth map computation for each view. |
| Outputs | Output is the depth map associated with each input view. Depth maps are 8-bit coded on one channel and have the same resolution as the input views. |
| Related processes | Depth maps can help saliency estimation as well as dynamic adaptation of the convergence plane and adaptive depth-of-field processing modules. |

* + - 1. Pre-processing toolbox: 2D-to-3D conversion

|  |  |
| --- | --- |
| 2D-to-3D conversion | |
| Pre-condition and inputs | This module accepts any 2D raw video streams. Settings allow user to dynamically fine-tune depth parameters to improve immersion feeling. |
| Process Description | This module manages to internally compute depth map estimate thanks to depth cues retrieval. Then, Depth-Image Based Rendering (DIBR) process enables to generate a second view for S3D rendering. |
| Outputs | S3D compliant streams are the outputs of this module. Estimated depth map might be retrieved by user. |
| Related processes | As S3D compliant streams and depth maps are yielded, this module may be coupled with modules that aim at converting content to specific 3D screen (namely dynamic reframing, adaptive depth-of-field processing, dynamic adaptation of convergence plane, saliency estimation modules) |

* + - 1. Pre-processing toolbox: saliency estimation

|  |  |
| --- | --- |
| Saliency estimation | |
| Pre-condition and inputs | The inputs of the saliency estimation module are:   * the two views: not encoded, raw format * optionally the depth map estimated for the two views |
| Process Description | This processing computes automatically a saliency map in other words a grey level map where the white pixels describe the most attractive part of the video. The saliency estimation is also called visual attention model. |
| Outputs | The output may be:   * a saliency map: buffer of grey level values (8 bits). * a heat map: buffer 8 bits * text file describing the location of the N first saliency peaks |
| Related processes | The output of this module is used as an input for the following modules: dynamic reframing, dynamic adaptation of convergence plane, adaptive depth-of-field processing and automatic 3D subtitle insertion. |

* + - 1. Conversion to specific 3D screen: dynamic reframing

|  |  |
| --- | --- |
| Dynamic reframing | |
| Pre-condition and inputs | The inputs of the dynamic reframing are:   * the two views: not encoded, raw format * the saliency map for each view * optionally the depth map estimated for the two views |
| Process Description | This processing aims at cropping the most important part of the video source depending on the target aspect ratio. Typical use case is the conversion from theatre (2:35 aspect ratio) to TV screen (16:9 aspect ratio). Instead of applying black stripes or uncontrolled rescale of video source, this algorithm performs automatically the selection of an area and the temporal smoothing of this area for a natural visual effect. |
| Outputs | This technology may output:   * the two cropped views * cropping window location (center and size) for each view |
| Related processes | This technology may be coupled with the saliency map estimation, depth map estimation and correction of the depth distortion. |

* + - 1. Conversion to specific 3D screen: dynamic adaptation of convergence plane

|  |  |
| --- | --- |
| Dynamic adaptation of convergence plane | |
| Pre-condition and inputs | The inputs of this technology are:   * the two views: not encoded, raw format * Shift or distance between views * Disparity max/min in the sequence * Depth map between views * Saliency map for each view |
| Process Description | This process aims at adapting the shift between views in order to always set the main region-of-interest on zero parallax (screen plane). It changes the distance between views over time by applying light temporal filtering which is not discernible. |
| Outputs | The output of this algorithm is two synthesized views which have been shifted, cropped and rescaled. |
| Related processes | This technology may be coupled with the saliency map estimation, depth map estimation. |

* + - 1. Conversion to specific 3D screen: Adaptive depth-of-field processing

|  |  |
| --- | --- |
| Adaptive depth-of-field processing | |
| Pre-condition and inputs | The inputs of this technology are:   * the two views: not encoded, raw format * Depth map between views * Saliency map for each view |
| Process Description | This technology applies a depth-of-field processing on each, in other words it blurs the part of the view which are not in focus of interest. It creates artificial blur on non region-of-interest. Moreover, the applied blur quantity is related to the depth distance between the current pixel and the main region-of-interest which is not blurred at all. |
| Outputs | The output of this algorithm is two synthesized views which have been spatially purposely blurred. |
| Related processes | This technology may be coupled with the saliency map estimation, depth map estimation. |

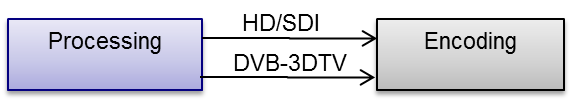
* + - 1. 3D content preparation: Automatic 3D subtitle insertion

|  |  |
| --- | --- |
| Automatic 3D subtitle insertion | |
| Pre-condition and inputs | The inputs of this technology are:   * the two views: not encoded, raw format * Shift or distance between views * Disparity max/min in the sequence * Depth map between views * Saliency map for each view |
| Process Description | This process provides an optimal depth value where a subtitle may be inserted within the content based on spatial and depth constraint. It set the subtitle on the same depth as the main region-of-interest and then reduces potential eye-strain linked to the convergence adaptation when switching between content and subtitle. |
| Outputs | It outputs depth value for inserting the subtitle per frame. |
| Related processes | This technology may be coupled with the saliency map estimation, depth map estimation. |

* + - 1. Video post-production

|  |  |
| --- | --- |
| Video post-production | |
| Pre-condition and inputs | 2 image streams (for left and right eye) in an appropriate format (depending on acquisition; camera format etc) |
| Process Description | The footage is imported, and some corrections concerning color or geometry may be performed. The footage is edited and combined into a complete programme |
| Outputs | A complete programme in an appropriate format for the remainder of the chain (coding etc) |
| Related processes | Capture, coding, distribution |

* + - 1. Interprocess formats



* Video Contents
  + 2 Raw Views
* Video Formats
  + Full HD/SDI (Format YCbCr422)
  + MPEG-TS
* Metadata Formats
  + DVB-3DTV and DVB-SI

### Contribution

Before being distributed to the end-user, transmission from capturing to production is usually needed in most of the scenarios. This transmission, known as contribution, is made with high quality/low delay 422 Profile encoding. Unlike distribution, being DVB compatible is not a requirement for contribution, as it is a point to point transmission usually done through private networks with guarnateed QoS.

The contribution phase covers the delivery of 3D contents from the place where the 3D image is created to the final Head Ends where the 3D image is encoding for transmission to final users. This delivery is typically done with maximum quality and minimum delay to be process for different operators, saving as much as possible the bandwidth needed.

* + - 1. 3D-full-HD stereo real time encoding for contribution

The aim of a contribution encoder is to meet a quality so that the encoding/decoding process can be considered as transparent as posible, that means, quality losses are imperceptible for the human eye after several encoding/decoding steps. This implies using high 422 encoding profiles with high transimision bitrates.

|  |  |
| --- | --- |
| 3D-full-HD-stereo encoder for contribution | |
| Pre-condition and inputs | inputs   * Two 1.5 Gbps HD-SDI (SMPTE 292M) video streams or one 3Gbps (SMPTE 425M) for HD stereoscopic video input |
| Process Description | Left and Right Views are encoded and decoded in a low delay/ high quality 422 profile for contribution links before production.   * For DVB-3D Phase1 frame packing compatible formats, high quality frame packing (quincunx) may be implemented before the contribution encoding process. * For DVB-3D Phase2 compatible formats, left and right views are encoded in the original resolution at high quality/low delay. After transmission they are decoded and synchronized to feed the production process. |
| Otuputs | The process outputs a MPEG2-TS containing one or two H264 422 video streams depending the DVB Phase compatibility. Each video stream may be set between 30 and 80 Mbps, so the complete MPEG2-TS bitrate may be between 70 and 200 Mbps. FEC is supported for IP network contributions. |
| Related processes | Output of this process is transmitted through a contribution link with guaranteed QoS. |

* + - 1. Contribution link

Private networks, with guaranteed bandwith and QoS, are usually used for contribution links to avoid network jitter, thus avoiding the use of big dejittering buffers that will result in an increase of the global delay in the contribution process.

Many different networks can be used for contribution links, for example IP, PDH or Satellite links.

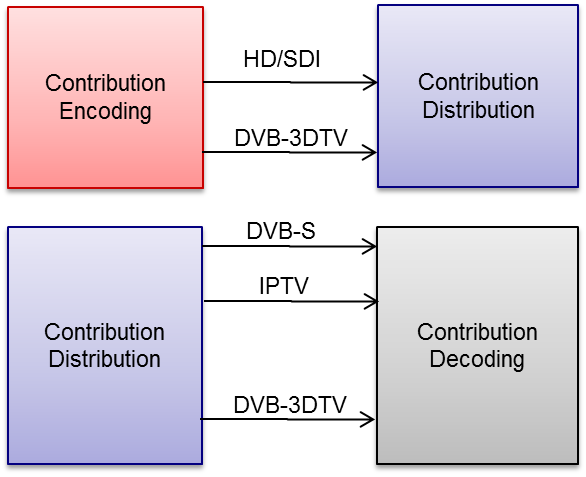
|  |  |
| --- | --- |
| Contribution links | |
| Pre-condition and inputs | MPEG-2 TS containing high quality encoded 3D video |
| Process Description | Depending on the network used for contribution, MPEG-2 TS content must be adapted in order to minimize transmission jitter |
| Otuputs | MPEG-2 TS containing high quality encoded 3D video |
| Related processes | Output of this process is used to feed the contribution decoder. |

* + - 1. 3D-full-HD stereo real time decoding for contribution

As there are still no standarized formats for 3D contribution encoding/decoding, the main objetive to meet in the decoder is to be compatible with the format of the encoder, beeing able to perform the opposite process to the one performed by the encoder, thereby achieving the transparency requirement previously mentioned.

|  |  |
| --- | --- |
| 3D-full-HD-stereo decoder for contribution | |
| Pre-condition and inputs | MPEG-2 TS containing high quality encoded 3D video |
| Process Description | The contribution decoder must able to decode and reconstruct the left and the right view   * Decoding one video stream and unpacking the left and right view to feed the production process. * Decoding two video streams and to feed the production process. |
| Otuputs | Two 1.5 Gbps HD-SDI (SMPTE 292M) video streams or one 3Gbps (SMPTE 425M) for HD stereoscopic video input |
| Related processes | Output of this process is used for production previous to distribution. |

* + - 1. Interprocess formats



* Contribution Encoding/Decoding
  + AVC Simulcast
* CONTRIBUTION Distribution
  + IPTV
  + DVB-S2
* CONTRIBUTION Metadata
  + DBV-3DTV

### Coding

The 3D-stereo content, which has been previously captured and pre-processed, needs to be compressed in order to be transmitted to the end-user over existing networks.

Moreover, content prepared beforehand in a pre-production chain may be encoded to feed non-linear distribution services like VoD or Blu Ray 3D production. High quality encoding is then very important (Blu Ray 3D) or special attention must be paid to obtain the best trade-off between high quality and bandwidth saving (VoD).The goal of JEDI project is to preserve the Full-HD resolution of both views to keep the highest definition of pictures, i.e. what is called 3D-stereo Phase 2 in DVB.

This section gives the technologies and processes which can be provided to address this goal.

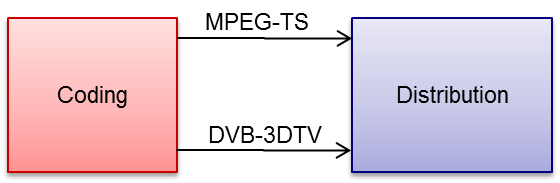
* + - 1. 3D-full-HD-stereo encoder for distribution

|  |  |
| --- | --- |
| Process Name : 3D-full-HD-stereo encoder for distribution | |
| Pre-condition and inputs | The inputs of this technology are :   * the left and right views in 1080p25 or 720p50 raw formats, carried over a 3G-SDI interface (SMPTE424M-2006) or over two separate HD-SDI interfaces (SMPTE292M-1998) * a logo as left and right views bitmaps |
| Process Description | The process encodes the two views, using MVC or MPEG4-AVC/H.264 compression depending on the available decoding process.  The Full-HD format is preserved for the two views (DVB-3D Phase 2). The foreseen encoding solution is 2D-HD service compatible but not DVB-3D Phase 1 compatible.  It can insert a logo in the left and right raw pictures before coding. |
| Outputs | The process outputs one single MVC or MPEG4-AVC/H.264 stream with a MPEG2-TS packetization. The MPEG2-TS stream can be encapsulated in RTP/UDP/IP packets if a transmission over IP networks is required.  The MPEG2-TS stream is at a Constant Bit rate which can be set between 10 and 20Mb:s |
| Related processes | 3D-full-HD-stereo decoder |

* + - 1. S3D video encoding for offline contribution

|  |  |
| --- | --- |
| S3D video encoder for offline contribution | |
| Pre-condition and inputs | Inputs are either two raw video streams either two video streams over two HD-SDI (SMPTE 292M). |
| Process Description | An encoder provides MVC Blu-Ray 3D compliant S3D 4:2:0 encoded streams with high quality. Another module offline encodes in MVC streams with *ad* hoc container to be stored for non-linear distribution (3D VoD) taking into account inter-view redundancy reduction in order to save bandwidth. |
| Outputs | Outputs are MVC stream contained in MPEG2-TS for providing IPTV CDN or VoD servers high quality feeds, and Blu-Ray 3D compliant files. |
| Related processes | 3D MVC compliant decoder (e.g. STB) and Blu-Ray 3D players. |

* + - 1. Interprocess formats



* Coding
  + AVC Side by Side (DVB ph1)
  + AVC Simulcast
  + MVC Simulcast (DVB ph2)
  + MVC advanced (offline)
  + MVC advanced (real-time)

### Distribution

The distribution of the 3D streams comprises all the steps which are necessary to deliver the encododed content to the decoding system, with a sufficient degree of quality.

To distribute a high-quality video stream over an IP network in real time, the network must provide capacities which allow the management of the possible quality issues. UDP is the reference transport protocol used to distribute video streams to a large number of subscribers in real time, due to its scalability (using multicast IP) and its low latency. On top of it, the network must at least provide a layer of control of the Quality of Service (i.e., protection against packet losses and control of jitter). As an enhancement of this property, the IPTV network systems will implement cross-layer analysis to generate a layer to control the Quality of Experience (i.e., protection against *the effect* of packet losses, jitter…).

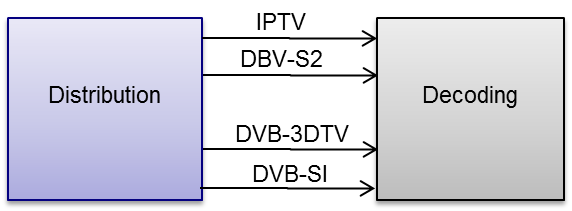
* + - 1. IPTV distribution

|  |  |
| --- | --- |
| IPTV distribution | |
| Pre-condition and inputs | The input of the IPTV distribution system is a video stream (i.e. a continuous flow), either generated in real-time from an distribution encoder or pre-generated and delivered as a file (working as a IPTV VoD server), with the following properties:  - MPEG-2 Transport Stream as transport layer, over (RTP)/UDP/IP, where RTP is optional, with up to 7 TS packets per IP packet (DVB-IPTV specifications can be used as a reference).  - MPEG-2 Transport Stream files are also supported.  - One or several video streams, encoded in H.264 and/or its extensions (MVC).  - One of several audio streams in MPEG-1 layer 2, AAC, AAC-HE, AC-3…  - Any other synchronized stream shall be included as PES within the MPEG-2 TS. |
| Process Description | IPTV distribution systems comprises all the modules required to deliver the content stream to the set-top-box. That includes:  - A preprocessing module as ingress to the IPTV network, which rewraps the MPEG-2 TS into a new RTP flow and includes video data to that flow (in the form of RTP header extensions) so that all the subsequent network processing is done at RTP level.  - Network elements which provide QoE enhancement layers: FEC generation, ARQ server, packet prioritization, quality monitoring and estimation…  - The IP transport layer itself (routing and transport elements). In laboratory scenarios, it can be replaced by network emulators.  - The home gateway subsystem (or a network device client attached to it), which operates as client to the QoE enhancement layer (i.e. FEC client, ARQ client, etc.). In some IPTV architectures, this element (of parts of it) is a part of the home network end device (set-top-box or similar). In JEDI, in order to simplify the architecture and adapt it to an hybrid transport scenario, these functionalities will be considered as the last step of the distribution system and will be part of the home gateway ecosystem. |
| Otuputs | The output of the IPTV distribution system is a video stream with the same requirements as the input (MPEG-2 TS/RTP/UDP/IP). Additionally,if FEC techniques are decided to be applied, an optional secondary RTP/UDP/IP stream will also be output. The aim of that stream is the protection of the video data against channel packet losses. |
| Related processes |  |

* + - 1. Distribution over cable via DVB-C

|  |  |
| --- | --- |
| DVB-C distribution | |
| Pre-condition and inputs | - ASI or IP from Sat. receiver  - SDI from broadcasters (VRT/ VTM/ VT4/ regional,…..)  - local encoding towards MPEG-2/H.264 for SD or H.264 for HD  - audio: Mpeg-1 Layer 2 and/or Dolby AC-3  - IP or ASI transport towards master headend and backup headend (=contribution of signals). |
| Process Description | - transparent feed-through towards multiplexer (combining x TV-channels together in a 51mbps pipe)  - re-encoding: to control bandwidth for stat.-muxed received services, to lower received bandwidth,…  - muxing: combining different TV/ Radio channels in pipes of 51mbps (equivalent digital bandwidth for RF-bandwidth of 8MHz and 256-QAM modulation)  - Conditional access: scrambling  - adding the necessary DVB-signalling: SI/PSI  - adding carousel streams holding applications/ upgrades of STB’s  - transport of IP-outputs of muxes over Gbe CWDM backbone (multiple 51mbps MPTS-streams) towards QAM-RF modulators in the different remote headends.  - RF-combining analog-TV and digital-TV QAM signals to fill up RF-spectrum HFC-network (adding Docsis RF-signals for IP) |
| Otuputs | - output STB:  CVBS via scart  digital via HDMI  component  - Reed-Salomon is used as FEC |
| Related processes |  |

* + - 1. Interprocess format



* Distribution
  + IPTV
  + DVB-S2

### Decoding

Once it has been distributed, the 3D-stereo content is decoded. The decoding process consists of performing, to the extent possible, an inversion of each stage of the encoding process.

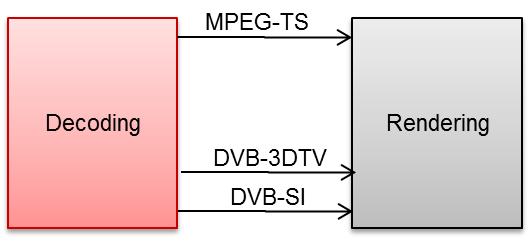
* + - 1. 3D Full-HD stereo decoder phase 1

|  |  |
| --- | --- |
| Process Name | |
| Pre-condition and inputs | Decoding process handle MPEG4-AVC/H264 [HP@L4.2](mailto:HP@L4.2). Handled 3D structure are Side-by-Side Half, Top-Bottom, Frame Packed  MPEG2-TS stream come through DVB-S/S2 satellite reception including DVB Phase 1 compliant signalling.  Alternate implementation get incoming MPEG2-TS through UDP/RTP/IP connection. |
| Process Description | STB get stream information from SI/PSI message and handle 3D metadata  STB decode incoming MPEG4-AVC/H264 stream up to [HP@L4.2](mailto:HP@L4.2)  STB Analyse 3DTV capability by parsing EDID through HDMI connection  STB will adapt incoming 3D video structure according to TV capability and will merge video with local User Interface / Graphics |
| Otuputs | STB will output HDMI 1.4a compliant signal according to 3DTV capabilities  Main output formats being:   * 1280x720p50 Side by Side Half * 1920x1080i50 Side by Side Half * 1280x720p25L25R Frame Packed * 1920x1080p25L25R Frame Packed * 1280x720p50 Top Bottom * 1920x1080p50 Top Bottom * 2D formats |
| Related processes | 3D-full-HD-stereo encoder for distribution |

* + - 1. 3D Full-HD stereo decoder phase 2

|  |  |
| --- | --- |
| Process Name | |
| Pre-condition and inputs | Decoding process handle MPEG4-MVC/H264.Frame Compatible Compatible and 2D compatible formats.  MPEG2-TS stream come through DVB-S/S2 satellite reception  Alternate implementation get incoming MPEG2-TS through UDP/RTP/IP connection. |
| Process Description | STB decode incoming MPEG4-MVC/H264 stream  STB Analyse 3DTV capability by parsing EDID through HDMI connection  STB will adapt incoming 3D video structure according to TV capability and will merge video with local User Interface / Graphics |
| Otuputs | STB will output HDMI 1.4a compliant signal according to 3DTV capabilities  Main output formats being:   * 1280x720p50 Side by Side Half * 1920x1080i50 Side by Side Half * 1280x720p25L25R Frame Packed * 1920x1080p25L25R Frame Packed * 1280x720p50 Top Bottom * 1920x1080p50 Top Bottom * 2D formats |
| Related processes | 3D-full-HD-stereo encoder for distribution |

* + - 1. Interprocess format



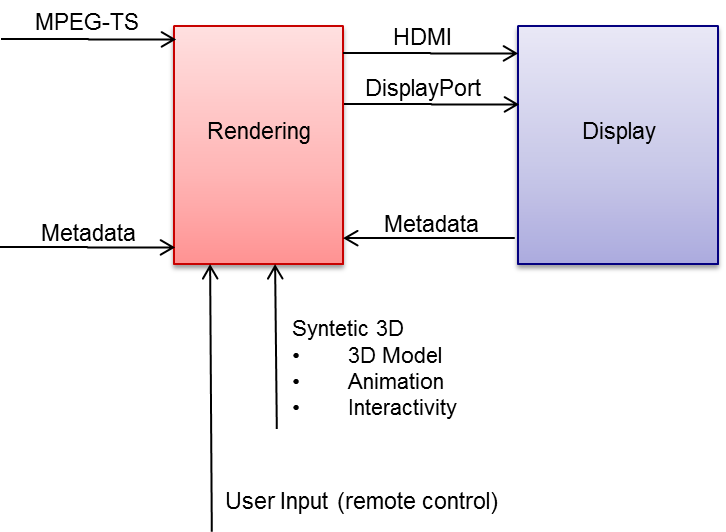
* Streams
  + MPEG2-TS
* Metadata Formats
  + DVB-3DTV
  + DVB-SI

### Rendering & Display

* + - 1. 3D TV Display

|  |  |
| --- | --- |
| 3D-TV | |
| Pre-condition and inputs | The HDMI input of the 3D television platform will be compliant to the HDMI1.4a specification. The following input formats should be supported :   * + Frame packed (double size): via HDMI     - Movie format : 1920x2205p@24/25/30Hz     - Game format : 1280x1470p@50/60Hz   + Broadcast formats: via HDMI and internal digital decoding     - Side by side and top-bottom (half HD) :       * 1920x1080i@ 50/60Hz       * 1920x1080p@ 24/25/30/50/60Hz   1280x720p@50/60Hz |
| Process Description | 3D platform and processing    The FE block will do the decoding, deinterlacing, cropping, scaling, adding the OSD and do the frame interleave and repeat. In the FPGA block 2D/3D conversion is done, 3D/3D conversion is done, dongle decoding is done, and PWM for the backlight is done. The BE will do the film judder removal, frame repeat ‘to 240Hz, retiming and 3D view mode scaling.  *Active 3d*  Extra possibilities are 2d/3d conversion and 3d/3d conversion (= depth scaling). |
| Otuputs | ·Stereoscopic 3D image. |
| Related processes |  |

* + - 1. Interprocess formats



* Video, Audio
  + MPEG-TS
* Formats
  + Modelling. (3DMAX, CINEMA4D)
    - 3DS
    - Collada
  + Animation (3DMAX, CINEMA4D)
    - 3DS
    - Collada
* Interactivity (Unity3-Mono)
  + Mono Scripts
  + User Input
  + Metadata
  + Depth Budget
  + Position/Range
  + Best Depth
  + Services XML (EPG…)
  + EDID (Video Format Supported…)

## JEDI Architecture Workflow and Use Cases

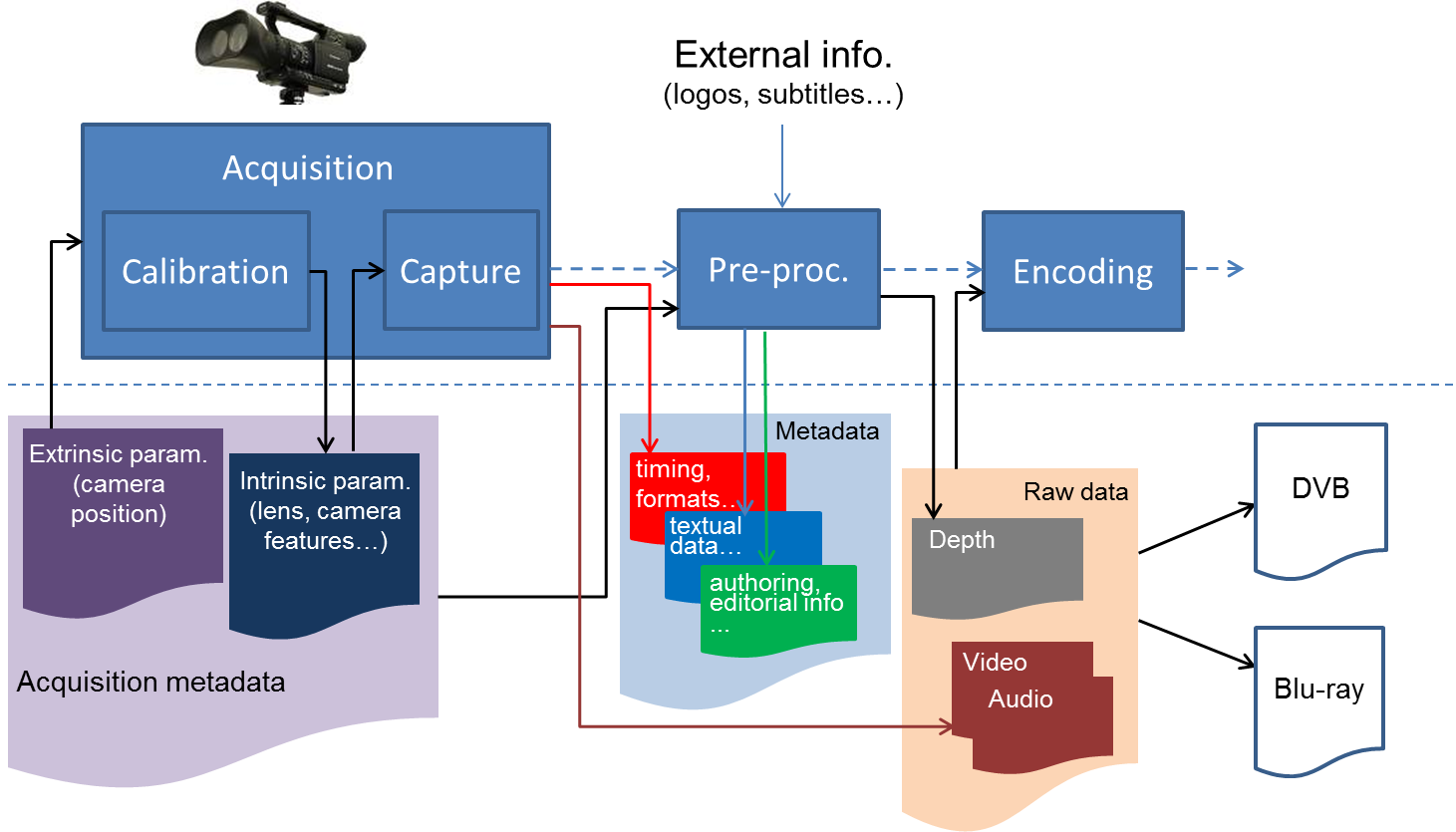
*Overview of the JEDI Architecture concerning the use cases defined in D1.1 and how the architecture adapts to them*

|  |  |
| --- | --- |
| Architecture Workflows | JEDI USE CASES (From D1.1) |
| **3.4.1 JEDI Content Delivery Workflow** | **CD1. 3DTV production workflow**  **CD2. BROADCAST 3DTV SERVICES**  **CD3. 3D graphics and objects insertion into the 3D uncompressed video stream**  **CD4. Contribution link for 3D emissions**  **CD5. Ensuring QoE on 3D IPTV networks** |
| **3.4.2 Presentation & Display based use cases** | **PD1. Transfert uncompressed 3D audiovisual content from a source (player, Set-Top-Box,…) to a display (TV, projector,…)** |
| **3.4.3 GRAPHICS & Interactivity Workflow** | **GO1. “3D Graphics Overlay”**  **GO2. Enriching user interfaces on a 3D TV** |

### JEDI Content Delivery Workflow

**CD1. 3DTV production workflow**

The goal of this use case is to create a working 3D production workflow for television from preproduction to distribution, thus allowing us to build up hands-on experience and expertise, especially in the domains of acquisition and postproduction.

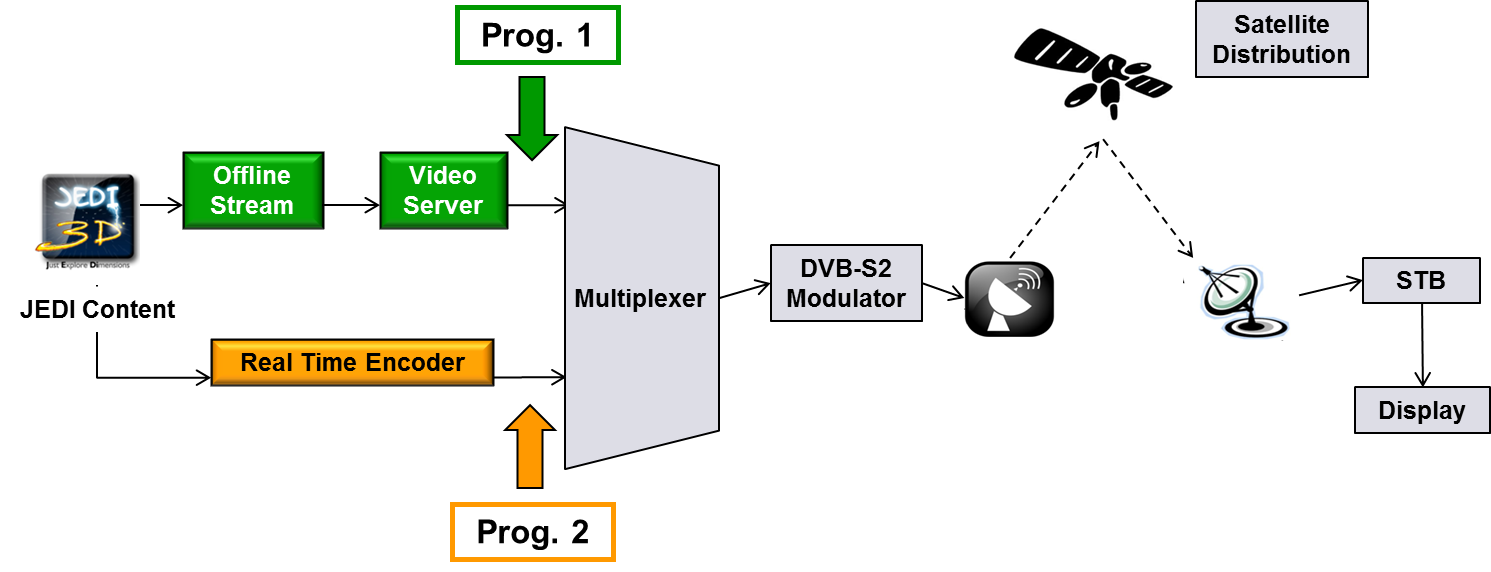
****

Relevant parts of this use case in the JEDI workflow:

|  |  |
| --- | --- |
| CD1. 3DTV production workflow | |
| **Capturing** | Several steps can be identified in such a 3D workflow. First of all, we identify an acquisition phase. The acquisition step consists of preproduction, where a 3D storyboard and depth script for a particular stereoscopic production/use case will be elaborated next to the “ordinary” 2D preproduction. Such a script contains notes about the targeted depth range (which will depend on the intended emotional impact, the set etc) or may even indicate the positions of the closest and farthest objects. After preproduction, the material can be shot in the actual acquisition phase. Here, it is important to learn which content works, which camera parameters enable good 3D (in a technical way; not causing headaches, but also artistically; as in how to create compelling stereoscopic footage) for the average consumer and which tools can deliver added value during the recordings of a 3D program. Next to dedicated 3DTV content, another important source of stereoscopic material will consist of 3D movies, adapted from cinema to TV. |
| **Processing** | Since the screen size has an important effect on the 3D experience (since the parallaxes and thus the depth effects are scaled together with the screen size), it has to be studied whether additional corrections to the content are necessary during this process, and whether these corrections are possible at all.  The next step is the postproduction phase, where the recorded assets are compiled into a finished program by means of video editing software. Here, some corrections may be applied to the stereoscopic media. The depth range may be shifted by a horizontal displacement of the left and right image in relationship to each other. Thus, the minimum and maximum values of horizontal parallax will be altered. Furthermore, some geometric corrections are possible; there exist tools to compensate for (unwanted) vertical parallax or keystone artifacts. Next to these 3D- specific operations, the typical editing operations such as cutting and applying transition effects will occur. In order to enable these, they have to be applied to both left and right videos, which need to be synchronized and kept together. Additionally, there is the aspect of adding e.g. subtitles or graphic overlays. During 3D postproduction, it is also important to limit the amount of format conversions. Therefore, it is important to choose a suitable format for editing, just like in 2D. Standardization efforts are to be observed; evolutions in exchange formats such as MXF & AAF may help to improve the efficiency of the resulting 3D workflow. Stereoscopic metadata in these formats would allow storing certain creative decisions such as the aforementioned horizontal displacement of left and right images in relationship to each other, thus enabling easy adaptations.  After postproduction, where a finished program is created, the product is ready to be coded and distributed. |

**CD2. BROADCAST 3DTV SERVICES**

The main objective of this use case is to define the specifications, techniques and systems involved in the whole 3DTV delivery chain.



Relevant parts of this use case in the JEDI workflow:

|  |  |
| --- | --- |
| CD2. BROADCAST 3DTV SERVICES RENDERING & DISPLY | |
| **Coding** | Two important 3DTV subsystems should be studied:  1) Frame compatible system: broadcasters can use an existing DVB HDTV broadcast channel and viewers can use an existing or suitably adapted receiver to receive the 3D-TV content.  2) Service compatible system: bradcasters can use 2D services plus “something“ to deliver 3DTV services.  JEDI Project will focus its work in the “service compatible system“ which is under study currently in different standardization organizations. Nevertheless, frame compatible system shall be considered in order to ensure the complete technological solution.  Network/Access network technologies: The implication of new encoding systems in the transmission chain shall be studied, new efficient delivering technologies should be defined for service compatible system (e.g Variable Coding Modulation +SVC). |
| **Distribution** | In JEDI project d a complete variety of configurations suitable for the project aims from the satellite network design point of view.  Obviously, DVB-S2 standard is the preferred technology as it is almost 30% more efficient than DVB-S for the same EIRP and bandwidth. Nevertheless, both systems are studied with the aim of comparing performances and offer different set-ups to the content provider.  In the 9MHz carrier case, even though, DVB-S provides a very robust link, the final bit rate for the allocated bandwidth would be not enough for an MVC service. DVB-S2 provides high performances in terms of bandwidth efficiency with also good quality of service results. The MCPC case provides more flexibility and enables the transmission of several channels with different charateristics over the same carrier, with better availability values and improving overall coverage. |
| **Decoding** | The Set-Top-Box is able to parse DVB Table in according with DVB-3DTV document A154 (Frame Compatible Plano-Stereoscopic 3DTV).  In a first step, the Set-Top-Box can only support AVC encoding format and is compliant with DVB-3DTV phase 1 & 2. |
| **Rendering & Display** | The Set-Top-Box is the main unit in JEDI chain: it is the link between the “3D” content, the broadcaster and the display. |

**Video**

**Audio**

**CD3. 3D graphics and objects insertion into the 3D uncompressed video stream**

The use case presented involves 3D graphical objects insertion into the uncompressed stream (e.g. channel logos, interactive application signalling objects, …).

Relevant parts of this use case in the JEDI workflow:

|  |  |
| --- | --- |
| CD3. 3D graphics and objects insertion into the 3D uncompressed video stream | |
| **Processing** | These 3D object are being processed at the post production level.  A set of metadata could be developped so as to offer enhancements to the graphics objects insertion.  In addition, the rendering control can be given to the end user through a set of configuration option, or no control at all can be allowed when the service provider requires so.  Such a mechanism offers a wider flexibility as compared to the technologies used today in that field.  As of today, graphics objects are being inserted into the video content either in the uncompressed domain, or in the compressed video content.  - advantage: no synch required at the rendering side  - disadvantage:  o no flexibility offered to the end-user,  o channel logos or other object insertion in the uncompressed video may create encoding issues (objects contours degradation, complexity introduced by hidden/revealed zones). The degradation of channel logos quality for ex is a great concern for broadcasters.  Solution: the graphics objects are part of a separate stream, along with metadata and synch tags. The resulted stream is being transmitted with the audio, video and other data streams.  At the rendering side, the user is being offered some display choices for such 3D graphics insertions (spatial positioning, formatting options, …). The quality of graphics objects is better preserved than in the initial case.  Optionally, some indications on the underlying technical aspects behind the use case and/or known technical/regulatory constraints that might affect the story and that must be acknowledged  - metadata standardisation  - synch mechanism  - behind logo insertion is the need for copyright content protection. If logo is transmitted along with the video stream, it is key to avoid the video stream being decoded and displayed without any logo. The logo stream should be used as a key to be able to decode the associated video and audio streams. |

**CD4. Contribution link for 3D emissions**

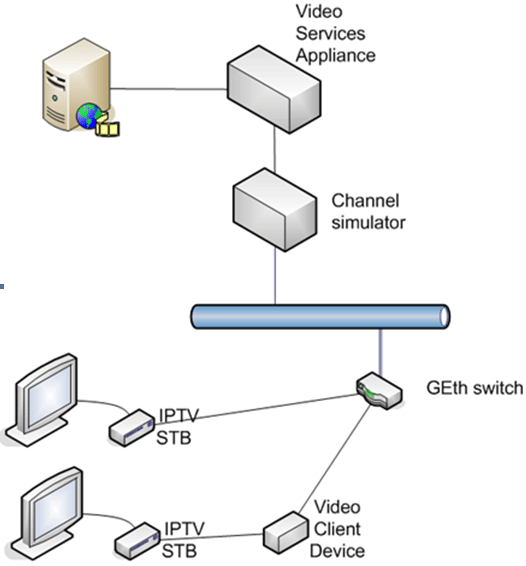
From the place where the 3D image is created to the final Head Ends where the 3D image is encoding for transmission to final users, normally is needed to send this contends with the maximum quality and minimum delay to be process for different operators, saving as much as possible the bandwidth needed.

Relevant parts of this use case in the JEDI workflow:

|  |  |
| --- | --- |
| CD4. Contribution link for 3D emissions | |
| **Contribution** | This use case study different ways and methods to send 3D contend over the network, analyzing mainly next 2 parameters:  - The Encoding/Decoding process.  - The latency in the encoder/decoder process  The main workflow (see FIGURE below) consists on transmit a 3D content in two different ways through a network:  - The first one is transmitted without encoding, with very high bandwidth required for the 3D contents. In this case normally fibber or cooper based in HD-SDI is required and will be used when the bandwidth is not a limitation.  - The second one is transmitted with Encoding / Decoding process. In this case, the required bandwidth is much lower, but the quality must be compared with the original 3D image, as well as the latency introduced in this process.  In case of Encoding / Decoding process, the latency will be higher, and this value will be more or less important depending of the contribution real application.  Also, the quality will be lower, but inside the contributions levels required, that will be dependant of the numbers of encoding/decoding process that will be present in the whole path.  at the “encoding”  In case of Encoding / Decoding process, the latency will be higher, and this value will be more or less important depending of the contribution real application.  at the “transmission”  Also, the quality will be lower, but inside the contributions levels required, that will be dependant of the numbers of encoding/decoding process that will be present in the whole path.  This point is the case of study. Diferent encoding methods will be evaluated, allways with the contribution quality in mind (4:2:2 profiles, etc)  For the use case, 2 transmissions ways will be required:  - High Bandwidth: Fiber/cooper for HD-SDI signals  - Low Bandwidth: DVB-ASI, SDH, IP, PDH.  at the “reception”  The same that for the transmission level |

**CD5. Ensuring QoE on 3D IPTV networks**

High bandwidth services such as 3D TV are challenging to deliver and may result in the end user noticing a reduced quality of experience (QoE).



Relevant parts of this use case in the JEDI workflow:

|  |  |
| --- | --- |
| CD5. Ensuring QoE on 3D IPTV networks | |
| **Distribution** | Despite the development and enhancement of transmission networks, the design and implementation of strategies and mechanisms of content transmission protection is still a very important matter, moreover taking into account the heterogeneous characteristics of the currently deployed networks.  A common technique used in content protection is FEC (Forward Error Correction), which added redundancy to the content goes from the result of simple XOR operations to more complex schemes such as Tornado or Raptor codes.  This use case proposes to ensure user QoE by adding Application Layer FEC (AL-FEC) protection to a 3D content.  The main workflow (see FIGURE below) consists on transmit a 3D content in two different ways through a noisy network: the first one is transmitted without protection, while FEC redundancy is added to the second one. At the other side of the noisy network two different devices (Set Top Boxes, if available) receive these transmissions. The one receiving content with FEC will be able to recover the errors caused in the network, while the other one will reproduce visible errors from both the Video/Audio and the depth information component.  AL-FEC correction capabilities are needed at the both ends, network (or headend) and final user end. In case that FEC correction enabled STB is not available, ALU could evaluate to provide a reference STB or PC prototype. |

### Presentation & Display based use cases

Display a 3D content stored on non-volatile support (e.g. BluRay Disc) with the highest resolution as possible, keeping high level of brightness, sharpness, color depth…

Display a 3D Live Event content broadcasted on the air or via cable, with limited processing (compression, buffering, decompression…) to keep the real time.

Both must use the same type of link to avoid various connections on the display.

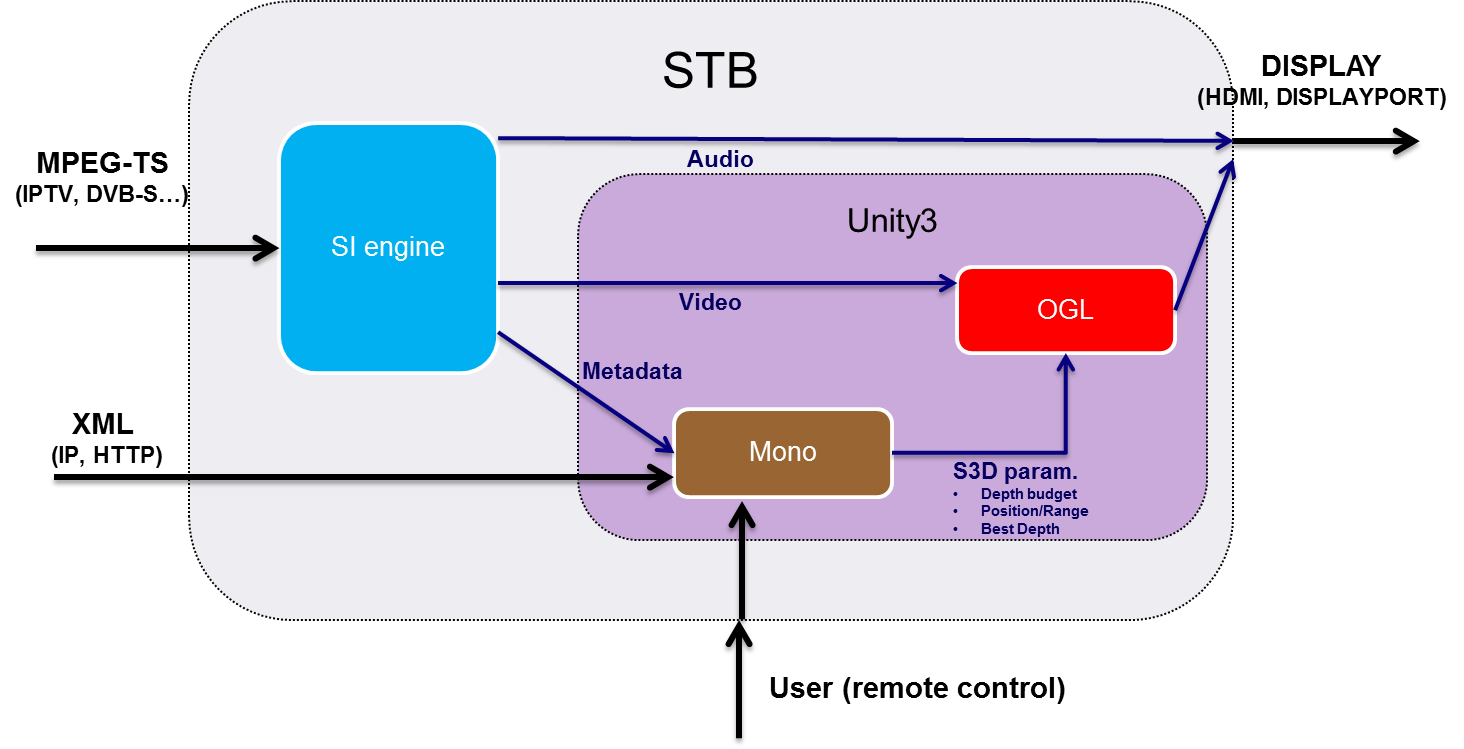
**PD1. Transfert uncompressed 3D audiovisual content from a source (player, Set-Top-Box,…) to a display (TV, projector,…)**

Relevant parts of this use case in the JEDI workflow:

|  |  |
| --- | --- |
| PD1. Transfert uncompressed 3D audiovisual content from a source (player, Set-Top-Box,…) to a display (TV, projector,…) | |
| **Decoding** | Display a 3D content stored on non-volatile support (e.g. BluRay Disc) with the highest resolution as possible, keeping high level of brightness, sharpness, color depth…  Display a 3D Live Event content broadcasted on the air or via cable, with limited processing (compression, buffering, decompression…) to keep the real time.  Both must use the same type of link to avoid various connections on the display.  Video On Demand; using high bandwidth network to download the content into an Hard Disk of a Media Box. At the end of the download, the content can be read as non-volatile support (see first point of description).  This use case requests high bandwidth connection between the player and the TV, to support a data stream at two times the rate of full HD data rate.  This use case requests constant bandwidth connection to warranty the bitrate and avoid freeze, black picture…    Specific information shall be exchanged:  • to indicate the 3D capability of the Sink  • to indicate the 3D structure of the datastream  Use 3D structure keeping the high quality of picture for each view, regardless the display technology. If the glasses need power (battery), the charging shall be easy (e.g. USB, docking station or contactless). Don’t need to remove glasses immediately when the content changes from 3D to 2D.  High bandwidth connection, dedicated for the video and audio (e.g. HDMI 1.4a or Display Port) If the synchro signal between the TV and the glasses is standardized, each user can bring its own glasses when visiting a friend. |
| **Rendering & Display** | Keep the possibility to transform 3D into 2D when there are not enough glasses for each viewer.  3D display must switch between 2D and 3D automatically depending of the content, without special action from the user, using information embedded into the stream |

### GRAPHICS & Interactivity Workflow

**GO1. “3D Graphics Overlay”**

****

The main goal of this use case is to show the user (“viewer”) 3D images over the real content. The context of this use case can be the insertion of vivid 3D advertisements, premium information/realtime statistics, games, widgets and user interfaces for services…

Relevant parts of this use case in the JEDI workflow:

|  |  |
| --- | --- |
| GO1. “3D Graphics Overlay” | |
| **Rendering & Display** | This is a basic use case that involves the generation of a 3D synthetic content that would be merged with real content and enables the removal or the replacement of 3D objects, or the reconstruction and the animation of 3D characters. The addition of 3D graphics needs to get some geometrical information about the 3D structure of the natural scene. This one can be provided by the depth maps previously computed on the input video streams.  Post-production techniques can be used to improve the user experience by taking into account the constraints of human perception. In order to ensure a good visualization comfort, it seems useful to simplify natural scenes so as to focus on the visual attention on the field of depth in the area where are standing the foreground objects or the characters in the scene.    A) Advertisement B) Interactivity: EPG, services, widgets…  This use case has mainly two variants, depending where the graphics overlay is generated and merged:  1) Before Transmission: The 3D graphics are generated and merged “unencoded“ with the real content before the transmision of the video.  2) STB 3D images generation: The 3D graphics are generated and overlayed by the Set Top Box, this means that the STB has 3D manipulation capabilities.  A blended aproach is the generation of 3D graphics content that is transmited and stored in the STB. The STB then does not have to generate the 3D objetcs but only overlay them. |

**GO2. Enriching user interfaces on a 3D TV**

3D video content has arrived to the 3D screen. Next we need to investigate how to deal with the user interface and the interaction of the users with the TV services. The main question to be answered is to identify the added value of 3D interfaces for an end user; is a 3D UI just a fancy add-on or does it really increase the value of the TV experience. The answer to that question will also define when the 3D UI has to be used, in conjunction with 3D video content, or also when the content is just 2D video (what will be the majority of the content for many years to come).

The use of 3D interfaces also is constrained by the real-life situation: a living room is not a lab, TV viewing typically happens in a multi-tasking environment, TV viewing is a multi-person activity, so users will not always have the glasses required. Moreover, we know that simplicity is key for a large scale adoption of new technologies, so the limitation of existing user habits must be taken into account.

Not-adapted, i.e. traditional 2D graphical user interfaces, that are generated by STB’s and TV sets appear messy on a 3D TV when the display is in 3D mode. Indeed, the frame-compatible image (e.g., a side-by-side video stream) sent to the TV is manipulated by the TV to be displayed with 3D visualization effects (see images below). This works well when the video consists of a left and right image, but does not work when the video is a normal 2D image, because the left and the right half of the image are completely mixed.

Relevant parts of this use case in the JEDI workflow:

|  |  |
| --- | --- |
| GO2. Enriching user interfaces on a 3D TV | |
| **Rendering & Display** | The full system composed by these stacks might be able to stream audio/video and play interactive application as supplied by JEDI distribution tool chain (cf. WP2) & expected by display device (cf WP3).  Authoring tools might be compatible & compelling to facilitate iterative process. For now, Unity3D seems to be the best candidate, able to import most of 3D design tools, such as 3Dsmax, cinema4D, Maya… along with exchange format such as those promoted by Kronos group (collada…).  We distinguish between basic 3D interfaces that actually ensure that the OSD (on-screen display) or the display provided by the STB (settop box) is displayed properly, regardless of the TV set’s display mode, and more advanced 3D interfaces that exploit the additional possibilities offered by a 3D display.  In both cases the following guidelines apply   * The 3D effects must support the user interface; it must provide additional value to the user interface. This additional value can be defined in multiple ways, ranging from providing a more comfortable user experience to end-users to a more inviting environment to drive additional media consumption. * The 3D effects must not be overwhelming, since (i) we know that between 10 and 20% of the population suffers to some extent from 3D effects that last too long or are of too high intensity and (ii) user interfaces for digital TV tend to contain a lot of textual information which is a specific challenge in 3D. * The 2D and 3D interfaces must be sufficiently comparable to allow consumers to seamlessly switch between 2D and 3D display modes |

# Conclusion

The JEDI Architecture for 3DTV has been introduced in this document. This architecture defines a workflow that comprises the whole end-to-end chain for 3D TV. The JEDI Workflow covers: (1) Capturing, (2) Processing, (3) Contribution, (4) Distribution, (5) Coding, (6) Decoding, (7) Rendering and Display.

Along these steps in the 3DTV chain, a list of processes have been identified that will be further investigated and developed in the following workpackages WP2, WP3 and WP4.

In order to link JEDI Architecture with a real context, the different use cases of the scenarios identified in D1.1 have been borrowed.