 ISO/IEC JTC 1/SC 29/WG 04 N0537

**ISO/IEC JTC 1/SC 29/WG 04  
MPEG Video Coding   
Convenorship: CN**

**Document type:** Output Document

**Title:** Overview of lenslet video coding activities

**Status:** Approved

**Date of document:** 2024-08-16

**Source:** ISO/IEC JTC 1/SC 29/WG 04

**Expected action:** None

**Action due date:** None

**No. of pages:** 7 (without cover page)

**Email of Convenor:** yul@zju.edu.cn

**Committee URL:** <https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-4>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 04 MPEG VIDEO CODING**

**ISO/IEC JTC 1/SC 29/WG 04 N** **537**

**July 2024, Sapporo, Japan**

|  |  |
| --- | --- |
| **Title** | **Overview of lenslet video coding activities** |
| **Source** | **WG 04, MPEG Video Coding** |
| **Status** | **Approved** |
| **Serial Number** | **24268** |
| **Authors** | **Mehrdad Teratani (ULB), Xin Jin (Tsinghua SIGS)** |

# Introduction

This document aims to clarify all Dense Light Field activities briefly (from inputs/outputs of previous/current meetings), such that future activities of Dense Light Field and the Lenslet Video Coding AHG can be clarified. Updates of this document will be provided in every meeting.

# Current Activities

Dense Light Field (DLF) is an activity that addresses the requirements for immersive interaction using the content captured densely by camera arrays or by plenoptic cameras to create visual content with 3D impression. The whole activity covers the capture, data conversion, compression, rendering, and display, targeting dense Light Field representations and their dedicated codecs .

A Light Field is a representation of ray-space. Different devices can be used to capture and render a discretely sampled version of this Light Field. The goal of the Dense Light Field activities. However, has focused on the contents mainly captured by plenoptic camera – a novel type of a camera that can acquire the dense light field in 3D by micro-lens array (MLA) as an objective. The DLF activity followed by Lenslet Video Coding (LVC) AHG aims to develop coding standards for such a representation, mainly captured plenoptic camera, in a Lenslet format – the content captured through MLA. LVC group aims to explore its coding tools for other contents such as MIV and V-PCC, and vice versa. LVC group started working on a new end-to-end structures for compression of lenslet content using Implicit Neural Visual Representation (INVR) – based methods.

The following sections provide an overview of important documents during the past meetings, use cases, software tools, test material, exploration experiments, recommendations and future activities related to DLF and LVC. The terminology used in this context is summarized in Table 1.

DLF activities with focus on Lenslet Video Coding has been distinguished by MPEG-I-Visual AHG and the first AHG was established since 134th meeting (Ad Hoc Group on Lenslet Video Coding (LVC). These activities are distinguished after collecting specific uses cases for DLF, and providing coding evidences. LVC AHG will investigate coding technologies for compression of content captured by plenoptic camera or plenoptic camera array. Reference software are Reference Lenslet content Convertor (RLC) [3] [4] and Reference Plenoptic Virtual camera Convertor (RPVC).

DLF activities promotes codec agnostic tools and also has started collaboration with AhG on implicit neural visual representation (INVR). It is planned to update the LVC requirement document and also draft a call for proposal in the next meeting. The activities will continue productive collaboration and interaction with MPEG-I-Visual activities. All terminologies that are shared with or related to both “MPEG-I-Visual” and “Lenslet Video Coding / Dense Light Fields” are listed below in the Table 1.

*Table 1: Abbreviations used in MPEG-I-Visual, Dense Light Field and Lenslet Video Coding activities.*

|  |  |
| --- | --- |
| **Abbreviation** | **Informal Description** |
| Light Field | A conceptual representation of light, where in addition to luminance or color, also directional information is captured from each light ray emanating from a point in space. The concept of Light Field is often related to sampling of the Plenoptic function. |
| Dense Light Field | A Light Field with light rays densely packed into the volume of interest (the so-called field of view), typically captured with Plenoptic Cameras, or synthesized/raytraced from 3D visual media representations |
| Sparse Light Field | A coarsely sampled Light Field, e.g. captured with a discrete set of cameras |
| Light Field Camera System | A camera device/system where each pixel captures luminance/color and directional light information from the Plenoptic function, e.g. with a discrete set of cameras, a Plenoptic Camera, etc. |
| Plenoptic function | A mathematical description of the Light Field with up to 7 parameters (3 spatial position coordinates, 2 angular direction coordinates, the light wavelength/color, and time) |
| Plenoptic Camera | Light Field Camera where in addition to the luminance/color, the directional light information of the Plenoptic function is obtained through a Micro Lens Array (MLA), correctly refracting light to the underlying pixels |
| Plenoptic camera type 1.0 and 2.0 | Depending on the position of the MLA with respect to the main lens of a plenoptic camera and the imaging sensor (aka CCD), two types of plenoptic cameras are defined. Plenoptic 1.0 has spatial resolution because the focal length of the main lens is overlapped with MLA. On the other hand, in plenoptic 2.0, the focal point of the main lens is in front of the MLA so that the Lenslet content through MLA has both spatial and angular resolutions. |
| Lenslet Video | Video content captured through an MLA, or converted from dense multiview images to Lenslet format |
| RLC | The Reference Lenslet Content Converter used to convert the lenslet data format into its corresponding Multiview format (also called sub apertures). |
| PRVC | The Reference Plenoptic Virtual cameras Calibrator used to calibration the virtual cameras corresponding to multiview, after the Lenslet view is converted to multiview. |
| LVC | Lenslet Video Coding used to describe the coding tools that can deal with video contents in Lenelet format. |
| MV-LVC | Multiview extension of LVC, allowing the exploitation of redundancies over multiple Lenslet views |
| HEVC | High Efficiency Video Coding |
| MV-HEVC | Multiview extension of HEVC, allowing the exploitation of redundancies over multiple views |
| 3D-HEVC | 3D extension of HEVC, i.e. Multiview + Depth, where depth information is compressed, and improves the texture compression as well. |
| HTM | 3D-HEVC Test Model software. |
| HM | HEVC test Model software |
| VVC | Versatile Video Coding |
| VTM | VVC Test Model software |
| MV | MultiView, i.e. multiple views of the scene, typically in the order of a dozen of views |
| MVD | MultiView+Depth, adding depth to MultiView content |
| EE | Exploration Experiment |
| Calibration | In DLF, calibration is a set of parameters that define intrinsic and extrinsic parameters of the camera in different coordinate systems. Here the calibration stands for calibration parameters of a plenoptic camera, plenoptic camera array, and multiview views converted by RLC from Lenslet view(s). |
| Disparity | The displacement of a feature (typically all pixels) in the scene when viewed from one to another camera view. Disparity and Depth are inverse proportional to each other. Disparity is often used in the reference software, but a language abuse often wrongly refers to Depth instead of Disparity. |
| Depth Estimation | Estimation of depth for each visible point in the scene, by evaluating the Disparity between at least two adjacent camera views |
| View Synthesis | The process of synthesizing a virtual view from existing input camera views, typically by a disparity/depth-dependent interpolation process |
| Coding tool | Refers to the tools that has to be implemented in the pipeline of the compression tools such as HEVC |
| Codec agnostic tool | Refer to the tool that are implemented outside of the legacy codec, and achieve compression gain by changing the format of the media |
| Legacy codec | Refers to the compression techniques that has been standardized such as HEVC |
| DIBR | Depth Image-Based Rendering, where images are rendered based on depth information. It is the typical process used in image-based View Synthesis. |
| DERS | Depth Estimation Reference Software, estimating depth from camera views by methods similar to stereo matching |
| VSRS | View Synthesis Reference Software, synthesizing a virtual view from two existing input camera views |
| RVS | The Reference View Synthesizer used in 3DoF+ with an unlimited number of input reference views in ERP and/or perspective camera format. |
| Autostereoscopic Display | MultiView display, typically with a dozen of directional output views, providing a stereoscopic viewing experience without wearing 3D glasses. |
| SMV | Super-MultiView, i.e. MultiView with many captured and/or rendered views (several dozens to hundreds) |
| SMV display | An advanced Autostereoscopic Display device with several dozens to hundreds of directional output views, often restricted to providing horizontal parallax only stereoscopic viewing |
| Light Field Display | An Autostereoscopic Display device, more advanced than SMV displays, to render a Dense Light Field, typically supporting full parallax and correct eye accommodation at very high light ray densities. |
| MPEG-I Visual | Covers the Visual technologies of Immersive media in MPEG-I |
| MIV | Metadata for Immersive Video |
| AI | Artificial Intelligence |
| INVR | Implicit Neural Visual Representation |
| TMIV | Test Model for Immersive Video |
| 360 video | Panoramic video texture projected onto a virtual shape (often a sphere) surrounding the user’s head, out of which he/she visualizes a portion for an immersive video experience. |
| ERP | An Equi-Rectangular Projection maps the texture of a sphere to a rectangle, similar to mapping the earth surface to a planar world map. |
| DoF | Degrees of Freedom |
| 3DoF | 3 Degrees of Freedom, i.e. allow movements along head rotation axes |
| 3DoF+ | 3DoF with also small translational movements of the head within a restricted volume, typically a person sitting in a couch |
| 6DoF | 6 Degrees of Freedom, i.e. allow movements along 3 rotation axes and 3 translations. Without further specifications, 6DoF presumes that full freedom of movement through the scene is possible. |
| Omnidirectional 6DoF | A restricted form of 6DoF, or an extended form of 3DoF allowing – besides of unrestricted rotations – small translational movements of the body within a restricted volume, typically a person taking a few steps from a central position, with the ability to look all around (cf. omnidirectional).nwgn5797\_L |
| Windowed 6DoF | A restricted form of 6DoF where the user virtually views the scene from behind a (virtual) window, with any position allowing to still see at least part of the scene. |
| FTV | Free viewpoint TeleVision |
| FN | Free Navigation, i.e. the capability to create all views required (cf. view synthesis) to create a smooth, virtual walkthrough between successive viewing positions |
| Epipolar Line | The line on which a feature point in a first camera view will necessarily lie in another camera view, as a consequence of a physical/optical relationship between cameras. Only parallel cameras have horizontal Epipolar Lines |
| EPI | Epipolar Plane Image, i.e. an image composed of corresponding Epipolar Line sections over all input camera views |
| IVDE | Immersive Video Depth Estimation, software for inter-view consistent depth estimation using multiple perspective or/and omnidirectional views |
| VVS | Versatile View Synthesizer used in Windowed-6DoF with a large number of input reference views in ERP and/or perspective camera format. |
| Stereo Sweeping  View Sweeping | A method of sweeping from one view to the next (including virtual views) to evaluate the quality of the view synthesis |
| Perceptual metric | A quality metric incorporating the Human Visual System characteristics. This metric should correlate to subjective quality experiences |
| IV-PSNR | PSNR (adapted) for Immersive Video, objective quality metric for Immersive Video applications |
| WS-PSNR | Weighted to spherically uniform PSNR |

# Overviews

Documents that consist of overviews are:

* [WG04N0537] Overview of LVC Activities
* [WG04N086] MPEG-I-Visual activities

# Use case and Requirements

Document for the requirement of LVC/DLF is planned to be prepared with focus on LVC in the following meetings. The public requirement document will be released in MPEG146.

# Test Materials

Documents that consist of test materials are:

* [WG04N0541] Collection of datasets for lenslet video coding
* [WG04N0053] Overview of MPEG-I Visual Test Materials
* [WG04N0186] Call for Lenslet Video Coding Test Materials

# Exploration Experiments

Documents that consist of Exploration Experiments are:

* [WG04N0539] Common test conditions of lenslet video coding
* [WG04N0540] Manual of Reference Lenslet content Converter (RLC 3.0)

# Collection of coding evidences

An output document that will briefly summarizes all coding evidence for LVC:

* [WG04N0538] Collection of evidences on lenslet video coding

# Core Experiments

# Reference software and repositories

The original reference software used in DLF activates is RLC:

* [5] User manual of Reference Plenoptic Virtual cameras Calibrator (RPVC) ver. 0.5

MPEG Git repository: <http://mpegx.int-evry.fr/software/MPEG/Video/lvc/rs/rpvc>

Public access: https://gitlab.com/mpeg-dense-light-field/RPVC

Coordinators: Sarah Fachada, Daniele Bonatto, Mehrdad Teratani

(Université Libre de Bruxelles)

* [6] RLC3.0 with user manual conversion from Lenslet video to multiview video

MPEG Git repository: <https://mpegx.int-evry.fr/software/MPEG/Video/lvc/rs/rlc>

* Public access: https://gitlab.com/mpeg-dense-light-field/rlc

Coordinators: Sarah Fachada, Daniele Bonatto, Mehrdad Teratani

(Université Libre de Bruxelles)

NOTE: RLC3.0 will be released by the end of May 2024 in the same repository.

Some of the reference software in MPEG-I-Visual will also be used to evaluate technologies in DLF and LVC activities (cf. [WG 04N086]), such a DERS, RVS,

* [WG11N19143] DERS9.0 with its manual can be found on the MPEG Git repository:

<https://mpegx.int-evry.fr/software/MPEG/Explorations/6DoF/DERS>

Coordinators: Eduardo Juárez (Universidad Politécnica de Madrid)

* [WG11N18068] [2] Reference View Synthesizer (RVS) 4.0 can be found on the MPEG Git repository:

<http://mpegx.int-evry.fr/software/MPEG/Explorations/3DoFplus/RVS/-/tree/v4.0-dev>

Public Access : <https://gitlab.com/mpeg-i-visual/rvs>

Coordinators: Gauthier Lafruit (Université Libre de Bruxelles)

* [WG04N0013] IV-PSNR Software Manual, cf. the software available on the MPEG Git repository: <https://mpegx.int-evry.fr/software/MPEG/MIV/RS/IVPSNR>

Public access: <https://gitlab.com/mpeg-i-visual/ivpsnr>

Coordinators: Adrian Dziembowski (Poznań University of Technology)

# References

[1] G. Tech and Y. Chen and K. Müller and J. R. Ohm and A. Vetro and Y. K. Wang, Overview of the Multiview and 3D Extensions of High Efficiency Video Coding, IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, pp. 35–49, 2016.

[2] S. Fachada, D. Bonatto, A. Schenkel, G. Lafruit, “Depth Image-Based View Synthesis with Multiple Reference Views for Virtual Reality,” 3DTV-CON, Stockholm-Helsinki, June 2018.

[3] Mehrdad Teratani, Shu Fujita, Wenzhe Ouyang, Keita Takahashi, Toshiaki Fujii, "3D Imaging System Using Multi-Focus Plenoptic Camera and Tensor Display," 2018 International Conference on 3D Immersion (IC3D2018) in IEEE Xplore Digital Library, Brussels, Belgium, December 2018.

[4] Shu Fujita, Sho Mikawa, Mehrdad Teratan, Keita Takahashi, and Toshiaki Fujii, "Extracting Multi-View Images from Multi-Focused Plenoptic Camera, " International Workshop on Advanced Image Technology (IWAIT)/International Forum on Medical Imaging in Asia (IFMIA), in SPIE Digital Library, Singapore, January 2019.

[5] Sarah Fachada, Armand Losfeld, Takanori Senoh, Gauthier Lafruit, and Mehrdad Teratani, “A Calibration Method for Subaperture Views of Plenoptic 2.0 Camera Arrays,” in 2021 IEEE 23nd International Workshop on Multimedia Signal Processing, Tampere, Finland, October 2021.

[6] Daniele Bonatto, Sarah Fachada, Takanori Senoh, Jiang Guotai, Xin Jin, Gauthier Lafruit, Mehrdad Teratani. Multiview from micro-lens image of multi-focused Plenoptic camera. In 2021 International Conference on 3D Immersion (IC3D), Online, December 2021. IEEE.

[WG11N18068] Reference View Synthesizer (RVS) manual, ISO/IEC JTC1/SC29/WG11 MPEG2018/N18068, Macau SAR, CN, October 2018.

[WG11N19143] Description of DERS, ISO/IEC JTC1/SC29/WG11 MPEG2020/N19143, January 2020.

[WG04N0013] Software manual of IV-PSNR for Immersive Video, ISO/IEC JTC1/SC29/WG4 MPEG2020/ N0013, Online, October 2020.

[WG04N0053] Overview of MPEG-I Visual Test Materials, ISO/IEC JTC1/SC29/WG4 MPEG2020/ N0053, Online, January 2021.

[WG04N0086] Summary on MPEG-I Visual Activities, ISO/IEC JTC1/SC29/WG4 MPEG2021/ N0086, Online, April 2021.