**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 2**

**MPEG TECHNICAL REQUIREMENTS**

**ISO/IEC JTC 1/SC 29/WG 2** **N360**

**Rennes, France – April 2024**

|  |  |
| --- | --- |
| **Title** | **Use cases and requirements for Lenslet Video Coding (LVC)** |
| **Source**  **Status**  **Serial number** | WG2, MPEG Technical Requirements  Approved  23870 |
|  |  |
|  |  |

# Abstract

This document provides the finalized requirements and use cases of Lenslet Video Coding (LVC) for Dense Light Fields (DLF) activities.

The LVC activities are foreseen to be followed in two phases. In the first phase, the focus will be on the technologies that are compatible with existing coding tools; namely Codec Agnostic Tools. In the second phase, the LVC activity will be exploring the technologies that can be embedded in the existing codec; namely coding tools.

# Introduction

The Lenslet Video Coding (LVC) is part of the Dense Light Fields (DLF) activities according to the MPEG roadmap[1]. DLF can be captured by different devices such as a single camera on a moving gantry, array camera, or plenoptic camera. Among all, the content is captured by plenoptic cameras composed of micro-images. The micro-images are captured through a micro-lens array that is placed between the main-lens of the camera and its imaging sensor. Thus, the pixels within each micro-image correspond to one micro-lens, which form the lenslet effect in the captured so-called lenslet image. It is found that no matter what kind of acquisition device is used, the lenslet data format can be used as a general format for dense light fields.

This activity is designed to efficiently compress light field video content captured by real plenoptic cameras or synthetically generated lenslet video.

In this document, use cases and requirements are collected for those in phase one of the LVC activity where codec agnostic tools are applicable.

This document is organized as follows. In Section two, the use cases are presented in two categories. General requirements and particular requirements for corresponding uses are listed in Section three. The estimated timeline for this standard is tabulated in section four.

**Definition of the terms used in the following sections**

|  |  |
| --- | --- |
| **Source** | A term used to describe the video material or some of its *attributes* before encoding. |
| **Lenslet data** | Image captured through an array of micro-lens or image synthetically generated from the light field data captured by camera array or a moving gantry. |
| **Lenslet parameters** | Includes location of each micro-image in the Lenslet image, it may include other metrics such as, the distance between image sensor and micro-lens array, distance between micro-lens array and the main-lens, focal distance(s) of micro-lenses, and focal distance of the main lens |
| **Viewing angle** | The maximum angle where the viewer can watch 3D contents without any clipping problem, i.e. correct perspective and parallax in the radial direction can be provided |
| **3D field of view** | The 3D field of view defines the limits of the volume within which the light field microscope can capture and reconstruct a three-dimensional representation of a specimen, including the depth, width, and height of the imaged volume. |

# List of Use Cases

Lenslet Video Coding (LVC) is fundamentally needed due to the use cases of the following two categories:

1. Communication-oriented applications that need fast and low bit rate compression for glasses-free 3D large-screen communication, glasses-free 3D large-screen Live broadcast, holographic communications, industrial inspection and robotics.
2. Storage-oriented applications that store the lenslet data captured by the plenoptic camera, store the lenslet data for integral display and layered display.

## **Communication-oriented**

Communication-oriented applications that need fast and low bit rate compression for glasses-free 3D large-screen communication, glasses-free 3D large-screen live broadcast, holographic communications, industrial inspection and robotics.

### **Coding of synthetic and natural contents for glasses-free 3D (G3D) large-screen communication**

This use case reflects lenslet data captured by using light field acquisition devices, e.g., array cameras, plenoptic cameras. At the terminal, the dense light field data are presented by glasses-free 3D (G3D) large-screen display, which enables individuals or groups to view highly realistic 3D images from diverse angles and perspectives without specialized glasses. In this scenario, there would exist many persons watching at various angles simultaneously (dense viewpoints) and conducting real-time communication. The specific applications contain remote meeting, remote education and remote control in the future for instance. The involved data may contain the lenslet data, 2D image and calibration information (e.g., camera, gyroscope). The capture content has the characteristics of ultra-high resolution, high frame rate, larger depth of field, and larger field of view (FoV).

### **Coding of natural scenes for G3D large-screen live broadcast**

At present, G3D large screen displays become more and more popular and provided in the city's core commercial areas, large transportation hubs, cultural venues and other places. The contents may be captured by camera on a moving gantry, array cameras and contain large-scale scenes like live broadcasting of sports events, games, shows and TV programs. The lenslet data captured in such a scene allows for computationally rendering plenty of viewpoints (up to hundreds of view angles), the larger field of view (FoV), the higher resolution and the larger depth of field.

### **Coding natural scenes with diverse light field resolutions and light field parameters for holographic communications**

In future, holographic communication would become a widely-used application. Holographic technology's applications in military, education, displays, and medical fields would experience substantial growth [2]. Lenslet data can be captured from the real scenes and transmitted to the display side where it can be converted into a complex light field through phase estimation and diffraction propagation for rendering on a holographic 3D display. The parameters of the light field cameras and lenslet data with diverse spatial and angular resolutions should be coded.

### **Coding of industrial scenes with non-Lambertian characteristics for industrial Inspection**

This use case reflects lenslet data captured through industrial inspection using light field acquisition devices, especially the plenoptic cameras. The lenslet data captured in such a scene allows for computationally rendering multi-viewing angles, synthesizing focal stacks, flexibly controlling depth of field and achieving full volumetric reconstruction. They may contain the inspection images on the specular surfaces, transparent objects, or those under variable lighting conditions [3-6]. The appearance of the inspections may vary depending on the viewpoint or among the microimages.

### **Coding of roaming data with complex motion for SLAM in Robotics**

This use case reflects the utilization of Light field-based Simultaneous Localization and Mapping (SLAM), employing plenoptic cameras as monocular depth sensors. Specifically, it involves the acquisition of lenslet data as the robot traverses through a 3D environment. The content may undergo stochastic or dramatic motion variations, embodying a blend of global and local motion dynamics. The scenes typically exhibit Lambertian characteristics while accommodating the presence of non-Lambertian objects. The lenslet data captured in such scenes enable depth estimation for visual odometry [7], accurate pose detection in non-Lambertian scenes [8], scale recovery with monocular cameras [9], and real-time 3D reconstruction for robotic surgery [10]. These advancements enhance robot navigation, perception, and interaction, with significant applications in various domains.

## **Storage-oriented**

Storage-oriented applications focus on storing the lenslet data captured by the plenoptic camera, light field video surveillance system, storing the lenslet data for integral display, layered display and holographic display.

### **Coding of on-chip lenslet data for plenoptic cameras**

This use case reflects the on-chip storage scenario that is fundamentally needed in the plenoptic cameras [11-16] before transmitting the data. Plenoptic cameras can capture dense light fields by placing a micro-lens array between the main lens and sensor. The light field captured by a plenoptic 1.0 or 2.0 camera is stored in the on-chip memory as a lenslet image. Since the lenslet images suffer from vignetting and unknown translation and rotation, pre-processing is required in the local memory before transmission. The resolution of a lenslet frame and frame rate are relatively high. The data can be stored or transmitted directly.

### **Coding of natural scenes with large field of view for light field video surveillance system**

This use case reflects a large field of view 3D scene viewing and reconstructions enabled through light field video surveillance. Light field video surveillance uses integrated camera array [17-19], which provides the larger field of view (FoV), the higher resolution, the larger depth of field (DoF) and “seeing through” capabilities with the assistance of synthetic aperture imaging. The content may contain large scale scenes like city square, airports [20], railway stations, stadiums [21], big social activities [21], and so on [22].

### **Coding of natural scenes with diverse light field resolutions and light field parameters for glasses-free 3D Display**

This use case reflects dense light field viewing experience enabled through glasses-free 3D displays. Currently, several types of glasses-free 3D displays like integral displays [23-25], layered 3D displays [26-31] and holographic displays [32] are under development. Lenslet data can be directly projected to an integral display which attaches a lenslet array on a digital screen or a film to provide continuous 2D disparity. Since the display needs the metadata that takes up huge amounts of data, the intrinsic geometric information should be considered during the metadata compression. Lenslet data can be converted to focal stack or multi-layer attenuated maps for layered 3D display [26-31], which display the light field by illuminating the multi-layer attenuators using directional backlight. To feel the depth variation of the object, context-adaptive masks may need to be coded. lenslet data can be converted to complex light field by phase estimation and diffraction propagation to be displayed on holographic 3D display [32], which provides a high-precision 3D scene with continuous depth variation and avoids the vergence-accommodation conflict of 3D displays. To convert the lenslet data to the computer generated hologram, the camera parameters of the light field need to be coded to meet the holographic displays' etendue. Also, since different types of displays involve different light field resolution and spatial-angular arrangements, lenslet data with diverse spatial and angular resolutions need to be coded.

# Requirements

## **General requirements**

For communication- and storage-oriented applications, the common requirements are:

1. The specification shall support lenslet videos captured by plenoptic 1.0 and plenoptic 2.0 cameras and those be synthetically generated.
2. The specification shall support lenslet videos as input and output.
3. The specification shall include technology that efficiently uses existing standardized video codecs.
4. The specification shall support a substantial reduction in bitrate for lenslet video data compared to VVC, with no significant increase in encoding and decoding time.
5. The specification shall support lenslet frames with diverse spatial-angular resolutions.
6. The specification shall support highly efficient lossy compression with parameter control of bitrate for applications that need this.
7. The specification should enable content authoring in a manner that 2D parallax, volumetric focal stack are supported in rendering without annoying artefacts.

Note: Testing methodology for visual quality of experience of rendering with 2D parallax and volumetric focal stack should be developed and/or chosen during the standardization.

## **Scene characteristics**

1. The specification shall enable coding a dynamic scene that contains non-Lambertian surfaces.
2. The specification shall enable coding natural and application specific scenes with large field of view or diverse spatial-angular resolutions.
3. The specification shall enable handling of metadata including the lenslet parameters, virtual camera parameters and object related parameters (e.g. intrinsic and extrinsic camera parameters, the coordinate of lenslet center, the diameters of lenslet, the same object point distance under adjacent lenslet and virtual camera array parameters) at the bitstream level.
4. The specification shall enable coding a scene captured by plenoptic cameras.
5. The specification should enable coding a scene captured by camera array. The content captured by camera array shall be converted to lenslet video format.
6. The specification should enable coding a scene with time-varying extrinsic and/or intrinsic camera parameters.

## **Facilities for application**

### **Communication-oriented requirements**

1. The specification shall support a trade-off between efficiency and end-to-end delay.
2. The specification should support real-time encoding and decoding.
3. The specification should support low bit rate compression.
4. The specification should support QoE/QoS-based optimization (e.g., bit rate control, quality enhancement).

### **Storage-oriented requirements**

1. The specification should support real time encoding for nearly lossless compression.

# References

1. WG 02 MPEG Technical requirements , MPEG Roadmap after the MPEG 144 meeting, ISO/IEC JTC1/SC29/WG02 N00313, Oct. 2023, Hannover.
2. China Mobile Research Institute. “Development Trend of 6G Holographic Communication Service”, <https://www.gtigroup.org/Uploads/File/2023/01/13/u63c15d5e0201a.pdf>
3. J. Kang, X. Jin and Z. Chen, “Light Field Spatial Correlation Fusion for High Precision Defect Detection”, IEEE International Conference on Visual Communications and Image Processing, 2021.
4. Tan, Zu Puayen, et al. “Development of a high-speed plenoptic imaging system and application to marine biology PIV”, Measurement Science and Technology, 2019, 31(5): 4005-4017.
5. <https://raytrix.de/flow/>
6. <http://www.vommatec.com/list/57.html>
7. Zeller, Niclas, Franz Quint, and Uwe Stilla. "Depth estimation and camera calibration of a focused plenoptic camera for visual odometry." ISPRS Journal of Photogrammetry and Remote Sensing 118 (2016): 83-100.
8. Z. Zhou, T. Pan, S. Wu, H. Chang and O. C. Jenkins, “GlassLoc: Plenoptic Grasp Pose Detection in Transparent Clutter”, 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 4776-4783.
9. Zeller, N., Quint, F., & Stilla, U. (2018). Scale-awareness of light field camera based visual odometry. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 715-730).
10. Shademan, A., et al. “Plenoptic cameras in surgical robotics: Calibration, registration, and evaluation”, IEEE International Conference on Robotics & Automation IEEE, 2016, pp. 708-714.
11. Chen M, Ye M, Wang Z, Hu C, Liu T, Liu K, Shi J, Zhang X. “Electrically addressed focal stack plenoptic camera based on a liquid-crystal microlens array for all-in-focus imaging”, Optics Express, 2022, 30(19):34938-34955.
12. Johnson K, Clifford C, Thurow B S. “Development of a modular, high-speed plenoptic-camera for 3D flow-measurement[J]”, Optics express, 2019, 27(9): 13400-13415.
13. Zhang Q., et al. “A generic multi-projection-center model and calibration method for light field cameras[J]”, IEEE tran on pattern analysis and machine intelligence, 2018, 41(11): 2539-2552.
14. Xufu Sun, Xin Jin, Tingting Zhong, Pei Wang, “A New Self-designed Focused Plenoptic Camera and the Captured Video Sequence “Boys””, ISO/IEC JTC1/SC29/WG11 MPEG2019/ m46259, 2019.
15. Raytrix. <https://raytrix.de/products/>
16. Vomma. <http://www.vommatec.com/page/92/95.html>
17. <https://newatlas.com/argus-is-darpa-gigapixel-camers/26078/>
18. <http://www.typonteq.com/products/>
19. <http://www.visionsplend.com/index.aspx?lanmuid=86&sublanmuid=729>
20. <http://www.typonteq.com/portfolio_category/fenlei1/>
21. <http://www.typonteq.com/portfolio_category/fenlei4/>
22. <http://www.visionsplend.com/index.aspx?lanmuid=83&sublanmuid=699&id=663>
23. J. Arai, et al., “Integral three-dimensional television using a 33-megapixel imaging system,” Journal of Display Technology, 2010, 6(10): 422-430.
24. Xing Y, Lin X Y, Zhang L B, et al. Integral imaging-based tabletop light field 3D display with large viewing angle[J]. Opto-Electronic Advances, 2023: 220178-1-220178-12.
25. Okaichi N, Sasaki H, Kano M, et al. Integral three-dimensional display system with wide viewing zone and depth range using time-division display and eye-tracking technology[J]. Optical Engineering, 2022, 61(1): 013103-013103.
26. Wetzstein, Gordon, et al. “Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays”, Acm Transactions on Graphics, 2011, 30(4): 95-112.
27. D Kim, et al. “Hybrid multi-layer displays providing accommodation cues”, Optics Express, 2018, 26(13): 17170-17189.
28. Zhang, J., et al. “Unified Mathematical Model for Multilayer-Multiframe Compressive Light Field Displays Using LCDs”, IEEE Transactions on Visualization & Computer Graphics, 2018, 25(3): 3350-3360.
29. Kobayashi, Y., et al. “A 3-D Display Pipeline: Capture, Factorize, and Display the Light Field of a Real 3-D Scene”, Ite Transactions on Media Technology & Applications, 2017, 5(3): 88-95.
30. Panahpourtehrani, M. T., et al. “3D Imaging System using Multi-focus Plenoptic Camera and Tensor Display”, 2018 International Conference on 3D Immersion (IC3D) IEEE, 2018, 8: 1-7.
31. Hyeonseung Yu, Mojtaba Bemana, Marek Wernikowsi, et al. “A Perception-driven Hybrid Decomposition for Multi-layer Accommodative Displays”, IEEE Transactions on Visualization and Computer Graphics, 2019, 25(5): 1940-1950.
32. Shi, Liang, et al. “Near-eye light field holographic rendering with spherical waves for wide field of view interactive 3D computer graphics”, ACM Transactions on Graphics (TOG), 2017, 36(6), 1-17.
33. <https://m.boe.com/company/dynamic-f19ae11e60b7419db115eb6724a473fa>
34. <https://www.zx-real.com/lightFieldHolographicScreen>.