



APSIPA

ASIA PACIFIC SIGNAL AND INFORMATION PROCESSING ASSOCIATION

Taipei Local Chapter

Workshop on Emerging and Immersive Visual Media: Hardware, Compression, and 3D/4D Representation

Date: June 8 (Monday), 2026

Time: 10:00 – 15:00 Hsinchu (UTC+8)

Location: Room EC340, Engineering Building 3 (工程三館), National Yang Ming Chiao Tung University (NYCU), Hsinchu

Organizer: Prof. [Wen-Hsiao Peng](#) (NYCU; Chair of [APSIPA Taipei Local Chapter](#))

Zoom link:

<https://us06web.zoom.us/j/87485546676?pwd=I9DfPP6SZoGr27mLtfj5cNpV3DoE7q.1>

Technical Program:

Time (UTC+8)	Topic	Speaker
10:00 – 10:05	Opening Remarks	Prof. Wen-Hsiao Peng Chair of APSIPA Taipei Chapter
10:05 – 10:55	Immersive Volumetric Video Experience with Spatial-Temporal Multi-Modal Gaussian Splatting	Dr. Guan-Ming Su (APSIPA Industrial DL) Dolby Laboratories
10:55 – 11:40	Results and Challenges of the Hardware Design for Video Coding	Prof. Luciano Agostini Universidade Federal de Pelotas
11:40 – 13:00	Lunch Break	
13:00 – 13:50	Post-Training Quantization for Overfitted Image Compression	Prof. Heming Sun (APSIPA DL) Institute of Science Tokyo
13:50 – 14:40	From Point Clouds to Gaussian Splatting: Neural Coding for Efficient 3D/4D Multimedia Representations	Prof. Jui-Chiu Chiang (APSIPA DL) National Chung Cheng Univ.

Sponsors: [APSIPA Taipei Local Chapter](#) and [H & J Global Chair Program \(華仁講座\)](#)

Contact: Prof. Wen-Hsiao Peng (wpeng@cs.nycu.edu.tw)



APSIPA

ASIA PACIFIC SIGNAL AND INFORMATION PROCESSING ASSOCIATION

Taipei Local Chapter

Talk #1: Immersive Volumetric Video Experience with Spatial-Temporal Multi-Modal Gaussian Splatting (Dr. Guan-Ming Su, Dolby Laboratories)

In the past few years, Gaussian Splatting has become the most promising volumetric video representation. For real-world scenarios, a great amount of research efforts has been conducted in many perspectives for better spatial-temporal multi-modal scene reconstruction and more effective deployment. In this talk, we will first introduce the fundamental representation of Gaussian Splatting, including attributes, construction, rendering, and optimization. Then, we will overview the recent development of Gaussian Splatting along the end-to-end ecosystem from content capture, content creation, content delivery, to content consumption. More specifically, on the content capture stage, we will address issues and solutions for multi-view cameras set up in both spatial and temporal domains to assist good quality model building. On the content creation stage, to enrich the multi-modal experience, learned audio, semantic, physics attributes from foundation models are embedded into the Gaussian Splatting primitives to enable the joint audio-visual-semantics-physics representation. Application of Gaussian Splatting on 2D video representation will be discussed. On the content delivery stage, instead of explicitly coding dozens of attributes per Gaussian, implicit methods such as using coarse geometry representations, 2D plane projection, and MLP to leverage the conventional 2D video codec to reach high photorealistic quality and streamable bit rate will be discussed. On the content consumption stage, relightable physics based rendered (PBR) Gaussian Splatting, multi-part Gaussian Splatting with joint, and LLM-assisted system can enable interactive 3D scene navigation and efficient physics-aware multi-modal rendering to improve user engagement. At the end of this talk, we will present the latest international standardization efforts and highlight future research trends.

Talk #2: Results and Challenges of the Hardware Design for Video Coding (Prof. Luciano Agostini, Universidade Federal de Pelotas)

This talk will present the main challenges in hardware design for video codecs and briefly discuss solutions developed by the speaker's research group to address them. The main focus will be on hardware architectures for VVC and AV1 encoders, but solutions targeting previous codecs, such as HEVC, AVC, and JPEG, will also be briefly covered. The key challenge in this area is the significant amount of data required to represent digital videos, combined with the high computational cost of modern codecs. Both aspects have become increasingly critical over the years, due to the emergence of new coding standards with progressively higher complexity and the continuous increase in requirements for spatial and temporal video resolutions. In this context, practical solutions for digital video processing typically rely on some type of hardware acceleration, from chipsets used in smartphones, through general-purpose processors and GPUs. The talk aims to present this scenario, discussing the challenges derived from this key challenge and some design alternatives for enabling the efficient hardware implementation of video codecs.

Talk #3: Post-Training Quantization for Overfitted Image Compression (Prof. Heming Sun, Institute of Science Tokyo)

Overfitted neural image codecs offer a promising path toward ultra-low-complexity decoding, making them well suited for resource-constrained hardware. However, fully realizing this potential requires an integer-only pipeline, which remains challenging due to their extreme sensitivity to post-training quantization (PTQ). We present an effective methodology for 8-bit weight, 16-bit activation quantization, developed through a three-stage investigation. (1) We first establish that a fine-grained, per-channel quantization scheme is the essential foundation for preserving model fidelity. (2) We then use Hessian analysis to identify strong weight dependencies as the primary bottleneck and apply a dependency-aware compensation algorithm to mitigate the resulting errors. (3) To further refine the compensation process, we integrate a principled, decreasing-importance order into the quantization sequence. Our final approach reduces the rate-distortion overhead to just +2.95% BD-Rate on the Kodak dataset and +2.78% BD-Rate on the CLIC dataset, establishing a practical methodology for this sensitive codec class.



APSIPA

ASIA PACIFIC SIGNAL AND INFORMATION PROCESSING ASSOCIATION

Taipei Local Chapter

Talk #4: From Point Clouds to Gaussian Splatting: Neural Coding for Efficient 3D/4D Multimedia Representations (Prof. Jui-Chiu Chiang, National Chung Cheng Univ.)

Next-generation interactive applications require efficient representation and compression frameworks to transmit immersive 3D/4D content. As 3D multimedia evolves from point clouds to view-synthesis paradigms such as 3D Gaussian Splatting (3DGS), a fundamental challenge remains consistent across these diverse formats: how to efficiently encode complex spatial and temporal data while faithfully reconstructing the plenoptic function of dynamic 3D scenes.

This talk introduces two end-to-end neural coding frameworks for efficient 3D and 4D multimedia representations. First, we present a unified point cloud compression framework with a single-encoder, dual-decoder architecture that jointly compresses geometry and attributes through a shared latent representation. The framework eliminates legacy recoloring bottlenecks and manual bit-allocation procedures, enabling efficient end-to-end optimization. Second, we present a rate-distortion-optimized framework for dynamic 4D Gaussian Splatting (4DGS). Built on a sparse anchor-based representation, the framework introduces learnable temporal-activation parameters to explicitly model the appearance and disappearance of Gaussian primitives over time, effectively handling occlusions without unnatural deformation artifacts. The framework further incorporates embedding-based deformation and lightweight INR-based hyperpriors for efficient compression.

By introducing advanced architectural priors and fully end-to-end data-driven optimization, these two works establish highly competitive processing efficiency and state-of-the-art rate-distortion performance for immersive spatial media representations.