
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202310.00426

Postprint: Technical Details of the “Panoramic View of the Winter Olympics” VR System for the 2022 Beijing Winter Olympics

Authors: Wang Nianhong, Sheng Xie, Zhang Jia, Wang Yu, Tang Junquan

Date: 2023-10-08T00:00:00+00:00

Abstract

CCTV.com’s “Panoramic View of the Winter Olympics” program launched during the 2022 Beijing Winter Olympics employed a fully self-developed VR system. This paper provides a detailed exposition of the Equi-Angular Pyramid (EAP) mapping mode applicable to 180° fisheye video sources from the Beijing Winter Olympics, video production solutions, a secure and reliable distribution and scheduling system based on the live broadcast center, player technology, page interaction and rendering techniques, as well as the CDRM copyright protection system.

Full Text

2022 Beijing Winter Olympics “Panoramic View of Winter Olympics” VR System: A Technical Deep Dive

Authors: Wang Nianhong, Xie Sheng, Zhang Jia, Wang Yu, Tang Junquan (CCTV.com, Beijing 100142)

Abstract: CCTV.com launched the “Panoramic View of Winter Olympics” program during the 2022 Beijing Winter Olympics, employing a self-developed, end-to-end VR system. This paper elaborates on the Equi-Angular Truncated Square Pyramid (EAP) projection mode tailored for the 180° fisheye video sources used at the Beijing Winter Olympics, along with the video production workflow, a secure and reliable distribution scheduling system based on a live broadcast center, player architecture, page interaction and rendering technologies, and the CDRM copyright protection system.

Keywords: Beijing Winter Olympics; VR System; EAP; Video Production and Live Streaming Scheduling

1. VR System Overview

A traditional VR system comprises four core components: video source, projection encoding, transmission, and terminal display. For the Beijing Winter Olympics, CCTV.com developed proprietary enhancements to each component (Figure 1 [Figure 1: see original paper]). To handle the 4K 180° fisheye video sources, we engineered a dedicated EAP mapping mode that preserves high-definition quality while reducing transmission bandwidth by over 50%. Additionally, we designed an efficient video production and transcoding solution to accelerate processing and ensure timeliness. In our distributed live broadcast center, we implemented a scheduling matrix that seamlessly switches to backup streams when the primary stream fails, guaranteeing VR live broadcast security. To deliver an exceptional viewing experience, our player integrates APIs for big data, advertising, and VDN, along with advanced page interaction features and spatial rendering technologies. Finally, to meet the stringent copyright protection standards of the Winter Olympics, we deployed our self-developed CDRM system to safeguard VR content.

2. EAP Mapping Mode

2.1 VR Mapping Mode Overview

VR mapping modes transform 360° spherical images into 2D planar images before encoding. Common approaches include Equi-Rectangular Projection (ERP), Cube Map Projection (CMP), Equi-Angular Cubemaps Projection (EAC), and Truncated Square Pyramid Projection (TSP), each with distinct advantages and limitations. ERP, the most widely adopted method, offers simplicity, fast computation, and intuitive 2D visualization, but suffers from non-uniform sampling and significant redundancy at the poles, making it inefficient for storage and transmission. CMP improves upon ERP's issues but still exhibits non-uniform sampling along face edges. In 2016, Facebook introduced TSP, which uses a truncated pyramid hexahedron to display 360° video, effectively highlighting the frontal view while de-emphasizing the other five faces. This approach significantly reduces storage requirements while maintaining high clarity in the primary viewing area. In 2017, Google proposed EAC, which further refined CMP but at the cost of increased computational overhead.

2.2 EAP Mapping Mode for 180° Fisheye Sources

The 2022 Beijing Winter Olympics VR live and on-demand services utilized 4K ultra-high-definition 180° fisheye video sources (Figure 2 [Figure 2: see original paper]), where the region of interest was the forward-facing area within the full 360° field of view, with side and rear areas rendered as black. Using CMP or EAC projection would treat all six faces equally, allocating only one-sixth of the total resolution to the critical frontal region. In contrast, TSP could allocate half the total resolution to the frontal area, dramatically enhancing clarity in

the primary region of interest. Building on TSP' s advantages and incorporating EAC' s uniform sampling characteristics, we developed the Equi-Angular Truncated Square Pyramid (EAP) mapping mode. EAP applies equi-angular transformation directly to the frontal view and adjacent left/right sides of the TSP structure, enabling more pixels within the same spatial area while reducing computational overhead compared to conventional cube-model conversion approaches.

Figure 3 [Figure 3: see original paper] illustrates the TSP hexahedral structure. Based on this foundation, EAP employs self-derived transformation formulas for mapping six 2D planes to 3D space. The derivation formulas for the front and right faces are as follows, where i, j represent pixel positions in the 2D image; x, y denote normalized plane coordinates ranging from $[0,1]$; and x', y', z' represent calculated coordinates in the 3D coordinate system. As shown in Figure 4 [Figure 4: see original paper], EAP' s frontal view incorporates equi-angular transformation for more uniform pixel distribution, while the side views utilize proprietary derivation algorithms combined with equi-angular transformation. Figure 5 [Figure 5: see original paper] displays a sample EAP-format VR video frame. During transcoding, we integrated our proprietary extreme-speed high-definition encoding optimization technology, enabling 2K (2880 \times 1440) output to achieve visual quality comparable to 4K while saving over 50% bandwidth.

3. VR Video Production System

3.1 VR Video Production Scheduling Workflow

To meet the massive VR video production demands of the Winter Olympics, CCTV.com' s Beijing Jike system and Wuxi video production system operated in tandem. After the Wuxi content team approved 4K ultra-high-definition 180 $^\circ$ fisheye clean feeds, they were transmitted via dedicated line to Beijing, where the Jike system ingested video metadata and cataloging information through custom APIs before routing to the transcoding resource pool (Figure 6 [Figure 6: see original paper]). To enhance VR transcoding efficiency, we deployed four additional high-performance transcoding servers and optimized the EAP projection computation using SIMD, precomputed lookup tables, and parallel algorithms, achieving a 7-8x improvement in production speed to fully satisfy the Olympics' demanding turnaround requirements.

3.2 Jike Video Transcoding Processing System Overview

The Jike system is CCTV.com' s self-developed video standardization platform that supports concurrent transcoding of various video formats into production-ready outputs. Built on a modern architecture with strict standard compliance, it offers high scalability, usability, maintainability, and stability. The system architecture is illustrated in Figure 7 [Figure 7: see original paper], with four key components detailed below.

3.2.1 Web Upload Employing a state-of-the-art frontend-backend separation design, the backend utilizes a Java architecture for high-concurrency video uploads. Restful APIs support uploading of VR and other encoded video formats, storing metadata and transcoding strategies in a backend database. The system accommodates batch uploads and large concurrent file transfers, saving files to private cloud storage. For the Winter Olympics, we developed new APIs to receive VR video metadata and cataloging information, with Wuxi's production system transmitting VR feeds via dedicated line to private cloud storage for cluster transcoding.

3.2.2 Video Transcoding We optimized texture mapping computation using SIMD, precomputed lookup tables, and parallel algorithms, significantly boosting transcoding efficiency. Users can configure transcoding strategies for each video, which are stored in the database alongside encoding parameters. We deployed an N-server transcoding cluster forming the Jike capacity pool for parallel processing, with task priorities ensuring critical videos are processed first. Each transcoding server fetches tasks from the database with lock-based contention protection to prevent duplicate processing. For long videos, we adopted Hadoop's Map/Reduce mechanism to split encoding tasks across machines for efficient processing. To prevent audio artifacts during slicing, we separated audio and video processing—splitting, encoding, and merging only the video streams before final audio-video multiplexing. This approach eliminates popping sounds at slice boundaries, solves slow transcoding for large files, and allows cluster scaling to continuously improve throughput.

3.2.3 Production System Scheduling After Jike's transcoding pool processes videos, a scanning transfer program migrates the transcoded files to the production system's internal storage while sending task completion messages to the MPC scheduling message queue. MPC retrieves these messages to perform multi-bitrate transcoding and packaging, then delivers the final transcoded files to downstream TMS systems.

3.2.4 Security Assurance All servers run Zabbix agents for secure operation monitoring, with flexible alerting mechanisms enabling rapid issue detection and resolution. We implemented highest-level password policies with regular rotation and promptly patched network security vulnerabilities. During the Winter Olympics, the system operated stably and securely 24/7, producing clear and smooth VR videos.

4. VR Live Streaming Scheduling System

To ensure high availability and prevent large-scale broadcast interruptions from single points of failure, CCTV.com built a VR live streaming scheduling system based on our newly constructed distributed live broadcast center (Figure 8 [Figure 8: see original paper]). This system incorporates primary/backup stream

ingestion and seamless switching technology (Figure 9 [Figure 9: see original paper]). We use encoders from different vendors to generate primary and backup streams (or multiple streams) that are ingested into the scheduling matrix via HTTP-FLV or RTMP protocols. The matrix connects to edge origin servers using HTTP-VFLV. When a single ingested stream experiences link anomalies, audio-video timestamp desynchronization, or timestamp errors, the system automatically switches to the backup stream within a single GOP duration and sends the new stream's encoding parameters to the edge origin, which initiates new segment slicing and delivers HLS segments to end users. Throughout this process, the scheduling matrix maintains continuous connection with the origin server, ensuring users experience no stuttering or flash interruptions, thereby significantly enhancing the viewing experience. Operationally, the system supports one-to-one and one-to-many stream scheduling for flexible distribution strategies to different customers, while sub-matrix partitioning enables clear business logic separation. These elegant designs substantially reduce system overhead and support scheduled tasks, lowering operational costs.

5. VR Player

5.1 API

CCTV.com's self-developed H5-VR player features a comprehensive API supporting multiple VR video formats, terminals, browsers, and video information retrieval methods. The player adapts to various VR projection formats including ERP, CMP, EAC, and EAP, with page integration parameters determining which format to play, meeting diverse streaming requirements.

5.1.1 Multi-Terminal and Multi-Browser Support The player supports PC platforms (Windows, Mac, Linux) and mobile platforms (Android, Huawei HarmonyOS, Apple iOS). On PC, it works with mainstream browsers including Chrome, Safari, Firefox, and Edge; on mobile, it supports system-native browsers, WeChat's X5 kernel, and other widely-used domestic browsers. We performed dedicated adaptations for each platform, significantly improving domestic compatibility and coverage. For greater versatility and lower maintenance costs, the player automatically detects PC or mobile environments and adjusts accordingly. Using identical parameters, it supports both live channel IDs and on-demand GUIDs, requesting the appropriate Video Delivery Network (VDN) for live or on-demand streams to retrieve corresponding video addresses, titles, and preview images.

5.1.2 Multiple Video Information Retrieval Methods Video metadata—including stream addresses, titles, and preview images—can be passed directly via parameters or retrieved indirectly from VDN, accommodating diverse business needs. The player also accepts control from VDN and advertising scheduling systems, integrates CCTV.com's big data statistics interface for real-time monitoring and quality adjustment, and employs a unified playback kernel for

easier page integration. Simple parameter settings enable mobile devices to play videos in windowed mode like PCs, supporting both horizontal and vertical full-screen playback.

5.2 Page Interaction

With 4K source video introduced for the Winter Olympics, we enhanced the VR player with virtual joysticks and gesture controls. PC users can drag with the mouse and zoom with the scroll wheel, while mobile users adjust viewing direction through multi-touch gestures and perform virtual camera push-pull operations, with automatic clarity adjustment when zooming to reveal fine details. Beyond conventional controls, we innovatively added a virtual joystick design enabling single-finger pan-tilt-zoom operations. PC users can similarly drag the virtual handle, offering smoother, more uniform movement than conventional dragging—particularly advantageous for sustained viewing. These technologies enable better presentation of 4K ultra-high-definition content on mobile screens. Additionally, the player features a mini radar indicator showing viewport size and direction, providing intuitive orientation for users prone to disorientation in VR environments.

5.3 Spatial Rendering and Display

EAP format's advantage is significantly lower resource and bandwidth consumption compared to other formats, though single-stream playback only provides clarity in one viewport while other areas appear blurred. To address this, we designed a clarity switching mechanism based on user viewing angle, ensuring high-definition playback for the user's current perspective while saving transmission costs and maximizing experience. Two implementation approaches are available: stream shunting and layered encoding.

5.3.1 Clarity Switching via Stream Shunting The server encodes the 4K ultra-high-definition 360° VR source into 4-6 separate 1080P or 2K video streams for different planes (Figure 10 [Figure 10: see original paper]). When a client user selects a viewing angle through dragging or gyroscope input and remains in that perspective beyond a threshold (typically 1-3 seconds), the player requests the corresponding video stream from the server and switches playback (Figure 11 [Figure 11: see original paper]). This method offers several benefits: the entire video stream switches without angular deviation or distortion; if the new viewport stream fails to load, the player continues with the previously successful stream; and switching speed depends on individual TS segment duration and network speed, with shorter segments enabling faster transitions—all configurable based on actual conditions.

5.3.2 Clarity Switching via Layered Encoding First, the 4K source is encoded as a base layer at standard definition, with residual layers compiled into different angular streams. The client initially pulls the base layer stream,

then fetches the corresponding residual layer based on the current viewing angle, and composites the two to produce high-definition video for that region. This approach enables efficient bandwidth utilization while maintaining high visual quality in the user's focus area.

6. Copyright Protection System

6.1 Copyright Protection Technology Overview

We compare three copyright protection technologies: anti-hotlinking, traditional encryption, and Digital Rights Management (DRM), before detailing our CDRM system.

6.1.1 Anti-Hotlinking Our first-generation anti-hotlinking solution dynamically generates tokens for m3u8 and TS file URLs that vary with time, IP address, user ID, and content ID, preventing direct URL reuse. However, this only deters casual hotlinking or basic downloading.

6.1.2 Traditional Encryption Mode The mainstream approach encrypts TS files and provides the decryption key URL via the #EXT-X-KEY tag in m3u8. The drawback is that anyone can access this key URL. While the key URL itself can be protected using anti-hotlinking methods, the fundamental vulnerability remains.

6.1.3 DRM System In a DRM system, the user system issues certificates to users or devices. The authentication system verifies the user certificate, then encrypts the content key using the user's public key for delivery to the client. The client decrypts the ciphertext with its private key to obtain the content key. This two-step process completes HLS-specified TS file decryption using high-strength asymmetric encryption that significantly increases key cracking difficulty.

6.2 CDRM Copyright Protection System

Based on DRM principles, CCTV.com's CDRM system workflow (Figure 12 [Figure 12: see original paper]) comprises encryption and decryption components.

6.2.1 Content Key Delivery The encryption system initiates an authentication request to the certification system, which verifies the encryption system's certificate chain. Upon successful authentication, the certification system encrypts the content key using the encryption system's public key and transmits it to the encryption system, which decrypts the ciphertext with its private key.

6.2.2 TS Fragment Encryption The encryption system independently encrypts audio and video TS fragments, encrypting H.264/H.265 NAL units with selectable algorithms and modes (DES, AES, MS1, etc.). The extensible encryption protocol interface supports arbitrary combinations of encryption methods, enabling real-time adjustments during live broadcasts. Reverse-engineering the encryption logic requires extensive analysis, substantially increasing cracking difficulty.

6.2.3 Content Key Acquisition End users log into the user system, which validates their legitimacy and issues certificates for the user and device. The certification system authenticates the user certificate chain, then encrypts the content key with the user's public key for delivery to the client, which decrypts it with its private key.

6.3 Player Terminal Copyright Protection

Encrypted video streams require real-time decryption and decoding at the client side, demanding both high speed and strong anti-cracking capabilities. Our H5 player decryption module employs LLVM technology, transmitting decryption code to the user page in hexadecimal format to prevent easy tracing and debugging. Furthermore, rather than using generic compilers, we developed a proprietary WASM flat obfuscation compiler to strengthen the decryption and copyright protection modules, achieving optimal balance between decryption speed and anti-cracking capability.

6.3.1 Bogus Control Flow Obfuscation As shown in Figure 13 [Figure 13: see original paper], the obfuscated program inserts a conditional statement composed of opaque predicates and an unreachable bogus basic block compared to the original single code block. Since the obfuscator knows the opaque predicate's value, original code execution is preserved, ensuring correctness. However, compilers and decompilers cannot evaluate the opaque predicate, forcing its retention and disrupting static analysis.

6.3.2 Control Flow Flattening Obfuscation Figure 14 [Figure 14: see original paper] demonstrates that after flattening, all basic blocks except the first are concentrated at the same hierarchy level. The calling sequence between blocks is determined by a main dispatcher, obscuring their relationships and increasing analysis difficulty.

The WASM files generated by our proprietary obfuscation compiler significantly increase decompilation difficulty while ensuring real-time decryption of ultra-high-definition content. Additionally, we obfuscated page JavaScript files to raise the bar for cracking attempts before reaching the WASM module.

Conclusion

CCTV.com' s self-developed VR system delivered a distinctive panoramic Winter Olympics viewing experience, establishing a technologically advanced, controllable platform for future high-quality VR programming. To achieve clearer visuals, lower costs, faster transcoding, and better interactive experiences, our video technology R&D team continues deep research, aiming for further breakthroughs in SIMD instruction optimization, GL model optimization, multi-layer encoding standard optimization, and player virtual scene technologies.

Author Biographies: - Wang Nianhong (1981-), male, from Xiaogan, Hubei, Engineer at CCTV.com, research interests: video players and codecs. - Xie Sheng (1982-), male, from Deyang, Sichuan, Engineer at CCTV.com, research interests: video transmission and encryption. - Zhang Jia (1983-), male, from Hengshui, Hebei, Engineer at CCTV.com, research interests: on-demand production and video encoding. - Wang Yu (1979-), male, from Chengdu, Sichuan, Engineer at CCTV.com, research interests: H5 audio-video playback. - Tang Junquan (1985-), male, from Bazhong, Sichuan, Engineer at CCTV.com, research interests: players and audio-video transmission.

(Responsible Editor: Chen Xuguan)

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.