Omnidirectional/360-degree Image and Video Standardizations Status

Junichi Hara[1]

[1]RICOH Company, Ltd. 2-7-1 Izumi, Ebina, Kanagawa 243-0460 Japan
Keywords: omnidirectional, 360-degree, JPEG 360, OMAF, immersive media.

ABSTRACT

This presentation reports technical aspects of the omni-directional/360-degree image and video standardizations; ISO/IEC 19566-6 JPEG 360 and ISO/IEC 23090-2 Omni-directional Media Format (OMAF) international standards. And this also introduces functions of these next version omnidirectional picture standards that now are discussed in standardization meeting, and discusses its applications.

1 INTRODUCTION

A photograph is a scene that is consist of a set of light rays traveling to and from a shooting location, which is visualized by instantly cutting out based on the shooting direction, shooting range, and shooting time intended by the photographer. A motion picture is a continues scene that is consist of series of target pictures which shot at a very short time interval, and arranged with chronological ordering for conveying the flow of the videographer's intended scene. Thus, photograph or video viewer gets light stimuluses according to intention of the photographer or videographer.

As a new imaging system, 360-degree cameras have recently been released and gradually become popular. This 360-degree camera extends the shooting range to all directions from the limited range by conventional cameras, and captures a 360-degree photograph or video picture by instantly capturing a scene in all directions viewed from the shooting location. Therefore, since 360-degree photo and 360-degree video pictures hold scene data in all directions, viewers can freely select their own viewing direction. Therefore, this new imaging system provides viewers with three degrees of freedom (3DoF) at a stationary viewing position, unlike a conventional imaging system provides fixed viewing pictures.

In 2019, ISO/IEC 19566-6 and ISO/IEC 23090-2 were published as international standards that provide this 3DoF feature to pictures and intend to store omni-directional/360-dgree and virtual reality (VR) pictures, are called JPEG 360 and OMAF (abbreviation of omni-directional media format) respectively. These standards are still being developed and attempted to expand more functionalities, such as 3DoF+ that allows a very narrow movement of the viewing position range. The 3DoF+ is extension of 3DoF and realizes a use case of movement of the viewer's neck while sitting on a chair. As other functional enhancement attempts, they are to include support for more codecs and projection format types used in the standard. In addition, supporting six degrees of freedom (6DoF), which includes the three-dimensional freedom of the viewing location, is being addressed and it is now being studied as a future standardization target.

This presentation briefly reports the configuration of the spherical image and video system. Then it is reported that technical features and future plans of two omnidirectional/360-degree image and video standards, the ISO/IEC 19566-5:2019 JPEG 360 and the ISO/IEC 23090-2:2019 MPEG-I (MPEG Immersive media) OMAF respectively. After that, this discuss applications and finally summarize it.

2 Overview of Omnidirectional Imaging Systems

Fig.1 is a block diagram of an omnidirectional/360-degree picture system, showing the main processing units from data generation to display. The outline of each processing unit will be described below.

2.1 Acquisition or Creation block

This block generates image data required for subsequent processing. These images are, captured by multiple camera sensors which convert optical data into image data, or generated by computer. These multiple image data are converted into a 360-degree picture by post-processing. At this time, like as each image data

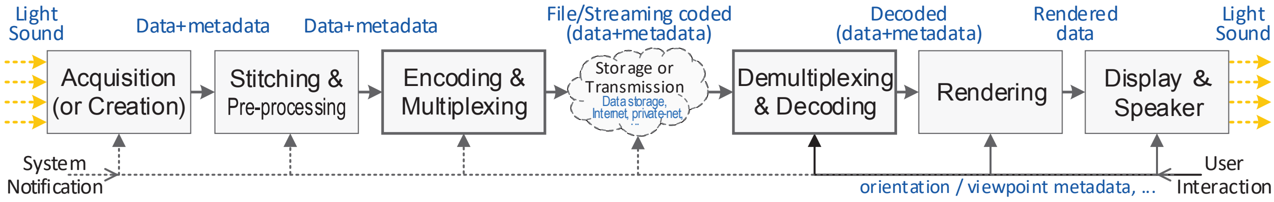


Fig. 1 Block diagram of omnidirectional/360-degree imaging system

element, metadata such as sensor arrangement and lens information are may also sent to subsequent processing blocks. A 360-degree picture may be generated by a compute, in this case the subsequent stitching process will be omitted.

2.2 Stitching and Pre-processing block

This block converts multiple images, acquired or generated by pre-processing block unit, to internal system formatted data express spherical picture compliance with such as post-processing blocks of coding and display. This process stitches multiple images onto a spherical picture. From the viewpoint of the interoperability of the system and the system implementation cost, this internal system format is ideally one kind. However, in the real systems, it is reasonable to prepare the minimum number of internal format models that satisfy the system requirements and convert them to alternative representation formats to be used in this system as needed. If it is efficient to use different formats for internal system data for the foreground and the background, it would be a solution to coexist a plurality of internal representation formats in the same scene. If the conventional codec, for 2-D pictures, is selected at the encoding, the input images are projected onto the 2-D picture. The current standards use this projected spherical picture for an internal system data.

2.3 Encoding and Multiplexing block

This block compresses the internal system data of the previous block output. This encoder should have effective coding performance on the internal system data while satisfying the functions required for systems such as random access, scalability and error resilience. Since this codec depends on the internal system data structure, the scheme may change for each data in the previous stage. After this processing, these coded data and metadata are multiplexed to media data with an appropriate file format.

If the system can use user interaction and/or system notification information, it may generate appropriate compressed data for displaying such as viewing direction area detail viewing or unknown object detail viewing.

2.4 Demultiplexing and Decoding block

This block generates the internal system data in the receiver/viewer site from the transmitted or stored media data. This block demultiplexing the packaged media data and decode it. It is also possible to generate pictures for display area by appropriate decoding parts of image data and metadata from the code based on response information from the user interaction.

2.5 Rendering block

This block converts the decoded data into an image format suitable for the display device in the subsequent stage. A head-mounted display (HMD) is often used for displaying this kind of images. The HMD displays a picture of the direction area of moving the face, and so outside image data of that area is not required for displaying.

In order to realize efficient rendering processing, it is necessary to skip processing this unnecessary image area data. By performing this ingenuity on the processing blocks before this Rendering block, only the image data necessary for the display area are processed sequentially from the preceding blocks, and then the efficiency of the entire system performance increases.

2.6 Display and Speaker block

This block displays and sounds the decoded media to viewers. The quality of displaying of this block affects the technical quality of each processing unit in the previous stage, as well as the user experience. In addition to using the HMD, various display devices can be used as this purpose devices such as a conventional display and a spherical dome display.

2.7 User Interaction

This provides system control information that can control the viewpoint, focal plane, and illumination conditions, which are generated by using user response actions such as the direction of viewing interest.

By sending these responses information up to the Acquisition or Creation block on the transmission side, it becomes possible to obtain more detailed display related data and instruct creation. In order to obtain a more natural immersive interaction, it is important to transmit response information without delay, to obtain necessary data instantaneously, and to display without a sense of incongruity.

2.8 System Notification

This can automatically and efficiently generate and send useful images that the system designer would like users to check by using information detected by the imaging system, such as important changes within the image. This is useful feature for such as surveillance camera applications. By sending this system information to the receiver side, the viewing system can intend users to check this important image.

3 JPEG 360: Omnidirectional image file format

The ISO/IEC 19566-6 JPEG 360 is a part of JPEG Systems family international standard that defines the metadata for omnidirectional/360-degree images. This standard uses ISO/IEC 19566-5 JUMBF as storing metadata into JPEG file format. The JPEG Systems is intended to prevent similar items redefinitions such as metadata and communication methods from being distributed in each standard.

The ISO/IEC 19566's part 1 File format and structure technical report and the part 2 Transport mechanisms and packaging standard have already been published. The part 4 Security and the part 7 JLINK standards are under discussion. The part 5 JUMBF and the part 6 JPEG 360 standards are also already completed, and these technical features extensions are still under discussion as technical amendments.

3.1 ISO/IEC 19566-5:2019 JUMBF

This international standard is an accumulation standard stores image-related and metadata in the box format used in JPEG 2000 and MPEG 4. JUMBF can store metadata in arbitrary formats such as XML (extensible markup language), JSON (JavaScript object notation), UUID box, codestream box, and other content box. The JUMBF Description box is located as the first inside box of the JUMBF box, and it describes the content type of content box which is located inside of the JUMBF box. The current content types which can be described are XML, JSON, UUID and codestream content types.

3.2 ISO/IEC 19566-6:2019 JPEG 360

This international standard is an 360-degree image file format standard. This standard defines minimum metadata necessary for 360-degre image composition, and the subsequent extended metadata set standard are now on discussion.

As shown in Fig. 2, this uses JUMBF standard, and the metadata of 360 degree images is described with XMP (eXtensible Metadata Platform) and soted in an XML box which is located inside the JUMBF box. These metadata elements are consist of image and viewport attributes, and uses ERP (equirectangular projection) coordinate system, as shown in Table 1 and Fig 3. When configuring an 360-degree image using with a JPEG file, it is possible to configure an JPEG backward compatibility image file, and then the JPEG viewer displays a normal JPEG image and the JPEG 360 viewer displays a 360-degree image.

3.3 New functionalities now being discussed

As expressed above, the development of the JPEG 360 standard has completed the minimum metadata definition that satisfies the basic and important use cases, and now the study of the extended JPEG 360 metadata set has begun. Extensions to store stereo images, to support new projection formats, and to store in non-JPEG image format such as JPEG 2000 are being studied and discussed. In addition, an extension of an image area related link information in JPEG 360 use cases is newly developing as the part 7 JLINK standard.

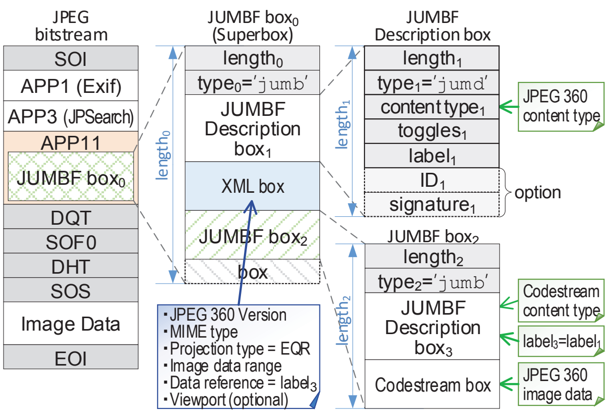
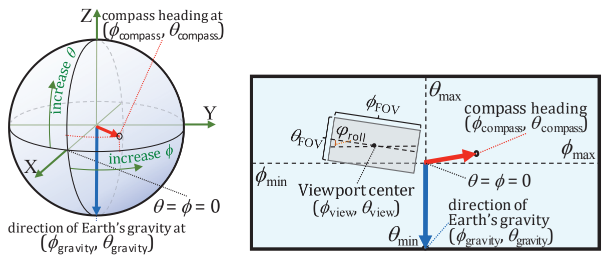


Fig. 2 JUMBF box structure for JPEG 360

Table 1 JPEG 360 metadata

JPEG 360 Element Name	Meaning	Default
JPEG360Metadata	JPEG 360 Metadata Container	n/a
JPEG360ImageMetadata	JPEG 360 image properties	n/a
JPEG360Version	JPEG 360 standard information	1
MediaType	Media data codestream type	image/jpg
ProjectionType	360 degree image projection format	Equirectangular
PhiMax	Longitude maximum value ϕ_{max}	+180
PhiMin	Longitude minimum value ϕ_{min}	-180
ThetaMax	Latitude maximum value θ_{max}	+90
ThetaMin	Latitude minimum value θ_{min}	-90
PhiGravity	Gravity direction (longitude) $\phi_{gravity}$	0
ThetaGravity	Gravity direction (latitude) $\theta_{gravity}$	-90
CompassPhi	Center direction (longitude) $\phi_{compass}$	0
CompassTheta	Center direction (latitude) $\theta_{compass}$	0
BoxReference	ID for the data container box	conventional
JPEG360ViewportMetadata	JPEG 360 Viewport properties	n/a
JPEG360ViewportNumber	Viewport number	0
ViewportPhi	Viewport center (longitude) ϕ_{view}	0
ViewportTheta	Viewport center (latitude) θ_{view}	0
ViewportPhiFOV	Viewport FOV (longitude) ϕ_{FOV}	75
ViewportThetaFOV	Viewport FOV (latitude) θ_{FOV}	100
ViewportRoll	Viewport roll (rotation angle) φ_{roll}	0



(1) Spherical coordinate (2) ERP coordinate

Fig. 3 Coordinate system for JPEG 360

4 OMAF: Omnidirectional video file format

The ISO/IEC 23090-2 MPEG-I OMAF is an omni-directional/360-degree video file format international standard, and it support a 3DoF feature on videos. The OMAF discussion was begun for standardizing a VR video format from mid of 2015 (officially Oct. 2015), it was started as ISO/IEC 23000-20 MPEG-A Part 20.

This OMAF was planned as three stages of standard development: Ver.1 standard realizes 3-D audio and basic framework, Ver.2 standard achieves additional features and AR support, and Ver.3 standard supports more advanced features. The currently published standard is the first version and plans to realize the next stage by technical amendments.

4.1 ISO/IEC 23090-2:2019 OMAF

This international standard was standardized with the goal of realizing following requirements; 2-D and 360-degree pictures and video with single/stereo/multiple viewpoint(es), 2-D and 3-D audio, projection system to realize spherical images and video, metadata for camera and user information, metadata for post-processing, media transfer/delivery standards of MMT and DASH, and ISO based media file format (ISOMBFF).

As shown in the Stitching and Pre-processing block process in Fig 1, the stitched spherical picture inputs are

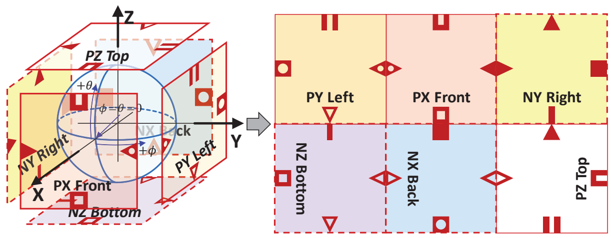


Fig. 4 OMAF cubemap parts placement onto 2-D plane

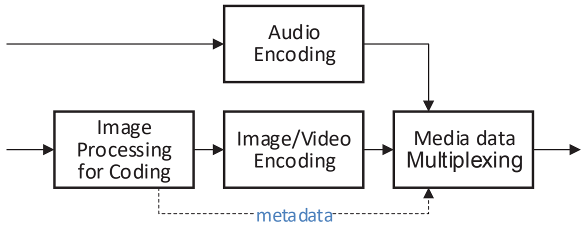


Fig. 5 Block diagram of OMAF encoding

converted onto 2-D picture with projection method. This OMAF Ver.1 uses equirectangular or cubemap projection method. The cubemap projection method projects one sphere picture onto cubic six wall square from inner-side, and then each squared picture is arranged onto one 2-D picture as illustrated in Fig. 4. Since fisheye picture is already a 2-D mapped half-sphere representation, the fisheye picture is no need to convert with this process.

These 2-D projected pictures, no fisheye image, are inputted into the image pre-processing block illustrated in Fig. 5, and they would be processed rotation and/or region-wise packing. Then, the processed video data will be encoded with H.264/AVC or H.265/ HEVC codec. Since the H.265/HEVC supports tile division features required for efficient transferring/delivering, to realize this purpose effectively, the pictures must be encoded by H.265/HEVC with tile division option. After these processes, the coded picture and audio data and metadata are multiplexed into an OMAF media data.

4.2 New functionalities now being discussed

The extensions currently proposed and discussed for the next version of the OMAF are new projection methods, effective video media transfer/delivery method, and late-binding function which achieves efficient VR streaming by acquiring, decoding and rendering. Above consideration of expansions is the OMAF Ver.2, and subsequence of this, OMAF Ver.3 is planned to support the newly coming codec currently developing new 360-degree image/video coding standard that aim to realize the 6DoF, which can change viewers' observation position with 3-D moving freedom.

According to the latest MPEG roadmap, the 3DoF+ video can be supported by 2021, and the 6DoF video and 6DoF audio codecs are predicated to be possible around 2022. Therefore, it is predicted that the Ver.3 will be completed at least after 2023.

5 Omnidirectional Picture Media Application

To improve the overall interactive performance, the system should utilize the user interaction information for obtaining data the only related with displaying from media data, as mentioned before. In other words, the coder should be able to generate media data that can instantly acquire only arbitrary display area data.

From a delivered data kinds perspective, these interactive applications can be classified into two types: accumulated media delivery and live media delivery. The accumulated media delivery system has no restriction on the media generation time; therefore, the system can perform coding without worrying about computing time. The live media delivery system needs to respond in real-time, so the latency can be a most important issue for this system, and its importance varies by application.

Applications interaction such as live concert delivery often affects no viewing target, so latency within several picture frames is no significant issue. Applications such as remote vehicle control need quick responses for control the vehicle, so this system requires a processing within single picture frame. Thus, these codecs are required coding time within each application's latency.

6 SUMMARY

This presentation reports a technical overview of omnidirectional/360-degree image and video systems, and technical features and future extensions of the international standards for omnidirectional/360-degree image and video file formats, ISO/IEC 19566-6:2019 JPEG 360 and ISO/IEC 23090-2:2019 MPEG-I OMAF. Currently standardized JPEG 360 and MPEG-I OMAF are the first version that achieve the minimum necessary functions, and the following extended version is currently under deliberation.

These discussion items might be going to be standardized within a few years, and will support new functions such as 3DoF+, which allows the viewers observation point to move slightly, and 6DoF which allows viewers observation point to move in space. In the near future, these expansions mostly might achieve to grow the global media market through the creation of new services and the expansion of application functions.

REFERENCES

[1] J. Hara, T. Ishikawa, O. Watanabe, and S. Ogawa, "Technical Trends of JPEG Family Standards Part 1," J. IIEEJ, Vol. 46, No. 3, pp.247-256 (2018).

[2] J. Hara, T. Ishikawa, O. Watanabe, and S. Ogawa, "Technical Trends of JPEG Family Standards Part 2," J. IIEEJ, Vol. 46, No. 4, pp. 469-478 (2018).

[3] "Presentation of MPEG Standardization Roadmap," ISO IEC/SC 29/WG 11 N18356, March (2019).

[4] K. Mizushima, et al., "Evaluation of Influence of Delay of Image Information on Steering Maneuver in Remotely Controllable Automated Driving System," J. JSAE, Vol.50, No.3, pp. 970-976 (2019).