

SIGNALING IMPORTANT VIDEO INFORMATION IN NETWORK VIDEO STREAMING USING MIME TYPE PARAMETERS

[0001] This application claims the benefit of U.S. Provisional Application No. **62/477,350**, filed March 27, 2017, the entire contents of which are hereby incorporated by reference.

TECHNICAL FIELD

[0002] This disclosure relates to transport of encoded media data.

BACKGROUND

[0003] Digital video capabilities can be incorporated into a wide range of devices, including digital televisions, digital direct broadcast systems, wireless broadcast systems, personal digital assistants (PDAs), laptop or desktop computers, digital cameras, digital recording devices, digital media players, video gaming devices, video game consoles, cellular or satellite radio telephones, video conferencing devices, and the like. Digital video devices implement video compression techniques, such as those described in the standards defined by MPEG-2, MPEG-4, ITU-T H.263 or ITU-T H.264/MPEG-4, Part 10, Advanced Video Coding (AVC), ITU-T H.265 (also referred to High Efficiency Video Coding (HEVC)), and extensions of such standards, to transmit and receive digital video information more efficiently.

[0004] After video (and other media data) data has been encoded, the video data may be packetized for transmission or storage. The video data may be assembled into a video file conforming to any of a variety of standards, such as the International Organization for Standardization (ISO) base media file format and extensions thereof, such as AVC.

SUMMARY

[0005] In general, this disclosure describes techniques for signaling important video information related to high dynamic range (HDR), wide colour gamut (WCG) video, virtual reality / omnidirectional / 360-degree video, frame packed video, video with display orientation changes, video stored using the restricted scheme feature of the ISO base media file format (ISOBMFF), and video using other features that need special post-decoding rendering processing to provide a desirable visual experience. Specifically, various example MIME type parameters are described, which may expose

such important video information in high-level systems signaling message bodies, e.g., the dynamic adaptive streaming over HTTP (DASH) media presentation description (MPD) file (or other such manifest file), such that important video information can be conveniently accessed by application clients, such as DASH clients, to make content rejection/selection/accepting/requesting decisions. That is, DASH clients may use this information to select an appropriate set of media data (e.g., an appropriate DASH representation), where the media data may be considered “appropriate” when the client device is capable of decoding and rendering the media data (e.g., the client device includes a video decoder that is capable of decoding the media data included in the DASH representation).

[0006] In one example, a method of retrieving media data includes retrieving a manifest file specifying data for at least one representation of a media presentation, wherein the manifest file includes data that specifies one or more codecs for the at least one representation; extracting, from the manifest file, the data that specifies the one or more codecs, including: extracting a first element representing a sample entry type code of a track of the at least one representation, wherein the first element represents that the track includes video data stored using a restricted scheme; and extracting a second element representing a restricted scheme type code for the restricted scheme of the track; and retrieving media data of the at least one representation based on the first element and the second element.

[0007] In another example, a device for retrieving media data includes a memory configured to store media data; and one or more processors implemented in circuitry and configured to: retrieve a manifest file specifying data for at least one representation of a media presentation, wherein the manifest file includes data that specifies one or more codecs for the at least one representation; extract, from the manifest file, the data that specifies the one or more codecs, the data including a first element representing a sample entry type code of a track of the at least one representation, wherein the first element represents that the track includes video data stored using a restricted scheme, and a second element representing a restricted scheme type code for the restricted scheme of the track; and retrieve the media data of the at least one representation based on the first element and the second element.

[0008] In another example, a device for retrieving media data includes means for retrieving a manifest file specifying data for at least one representation of a media presentation, wherein the manifest file includes data that specifies one or more codecs

for the at least one representation; means for extracting, from the manifest file, the data that specifies the one or more codecs, including: means for extracting a first element representing a sample entry type code of a track of the at least one representation, wherein the first element represents that the track includes video data stored using a restricted scheme; and means for extracting a second element representing a restricted scheme type code for the restricted scheme of the track; and means for retrieving media data of the at least one representation based on the first element and the second element.

[0009] In another example, a computer-readable storage medium, such as a non-transitory computer-readable storage medium, has stored thereon instructions that, when executed, cause a processor to retrieve a manifest file specifying data for at least one representation of a media presentation, wherein the manifest file includes data that specifies one or more codecs for the at least one representation; extract, from the manifest file, the data that specifies the one or more codecs, including instructions that cause the processor to: extract a first element representing a sample entry type code of a track of the at least one representation, wherein the first element represents that the track includes video data stored using a restricted scheme; and extract a second element representing a restricted scheme type code for the restricted scheme of the track; and retrieve media data of the at least one representation based on the first element and the second element.

[0010] The details of one or more examples are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

[0011] FIG. 1 is a block diagram illustrating an example system that implements techniques for streaming media data over a network.

[0012] FIG. 2 is a block diagram illustrating an example set of components of a retrieval unit of FIG. 1 in greater detail.

[0013] FIG. 3 is a conceptual diagram illustrating elements of example multimedia content.

[0014] FIG. 4 is a block diagram illustrating elements of an example video file, which may correspond to a segment of a representation.

[0015] FIG. 5 is a flowchart illustrating an example method in accordance with the techniques of this disclosure.

DETAILED DESCRIPTION

[0016] In general, this disclosure describes techniques for signaling important video information related to high dynamic range (HDR), wide color gamut (WCG) video, virtual reality / omnidirectional / 360-degree video, frame packed video, video with display orientation changes, video stored using the restricted scheme feature of the ISO base media file format (ISOBMFF), and video using other features that need special post-decoding rendering processing to provide a desirable visual experience.

Specifically, various example MIME type parameters are described, which may expose such important video information in high-level systems signaling message bodies, e.g., the dynamic adaptive streaming over HTTP (DASH) media presentation description (MPD) file (or other such manifest file), such that important video information can be conveniently accessed by application clients, such as DASH clients, to make content rejection/selection/accepting/requesting decisions. That is, DASH clients may use this information to select an appropriate set of media data (e.g., an appropriate DASH representation), where the media data may be considered “appropriate” when the client device is capable of decoding and rendering the media data (e.g., the client device includes a video decoder that is capable of decoding the media data included in the DASH representation).

[0017] For example, this disclosure discloses several example methods of signaling important video information related to video stored using a restricted scheme, HDR/WCG video, VR/omnidirectional/360 video, frame packed video, and video with display orientation changes, such that the important video information can be conveniently accessed by application clients, such as DASH clients, to make content rejection/selection/accepting/requesting decisions. One or more of these methods may be performed independently, or in any combination.

[0018] In the context of this document, “important video information” includes video information that may be used for content selection, e.g., selection of a video track or a part thereof for consumption.

[0019] Video coding standards include ITU-T H.261, ISO/IEC MPEG-1 Visual, ITU-T H.262 or ISO/IEC MPEG-2 Visual, ITU-T H.263, ISO/IEC MPEG-4 Visual, ITU-T

H.264 or ISO/IEC MPEG-4 AVC, including its Scalable Video Coding (SVC) and Multiview Video Coding (MVC) extensions, and High-Efficiency Video Coding (HEVC), also known as ITU-T H.265 and ISO/IEC 23008-2, including its scalable coding extension (i.e., scalable high-efficiency video coding, SHVC) and multiview extension (i.e., multiview high efficiency video coding, MV-HEVC).

[0020] Both AVC and HEVC support frame packed video indicated by the frame packing arrangement SEI message. HEVC also support a different type of frame packed video indicated by the segmented rectangular frame packing arrangement SEI message. For such frame packed video, the decoder side should apply a special de-packing transformation to separate the components of the two views represented in the video bitstream prior to display.

[0021] AVC and HEVC also support video content for which the decoder side should apply a transformation of rotation and/or flipping to the cropped decoded picture prior to display, indicated by the display orientation SEI message. Such video is also referred to as video with display orientation changes.

[0022] The techniques of this disclosure may be applied to video files conforming to video data encapsulated according to any of ISO base media file format, Scalable Video Coding (SVC) file format, Advanced Video Coding (AVC) file format, Third Generation Partnership Project (3GPP) file format, and/or Multiview Video Coding (MVC) file format, or other similar video file formats.

[0023] File format standards include ISO base media file format (ISO/BMFF, ISO/IEC 14496-12), and other derived from the ISO/BMFF, including MPEG-4 file format (ISO/IEC 14496-15), 3GPP file format (3GPP TS 26.244) and file formats for AVC and HEVC families of video codecs (ISO/IEC 14496-15). Draft texts of ISO/IEC 14496-12 and 14496-15 are available at http://phenix.int-evry.fr/mpeg/doc_end_user/documents/111_Geneva/wg11/w15177-v6-w15177.zip and http://wg11.sc29.org/doc_end_user/documents/115_Geneva/wg11/w16169-v2-w16169.zip, respectively.

[0024] The ISO/BMFF is used as the basis for many codec encapsulation formats, such as the AVC file format, as well as for many multimedia container formats, such as the MPEG-4 file format, the 3GPP file format (3GP), and the DVB file format.

[0025] In addition to continuous media, such as audio and video, static media, such as images, as well as metadata can be stored in a file conforming to ISO/BMFF. Files structured according to the ISO/BMFF may be used for many purposes, including local

media file playback, progressive downloading of a remote file, segments for Dynamic Adaptive Streaming over HTTP (DASH), containers for content to be streamed and its packetization instructions, and recording of received real-time media streams.

[0026] A box is the elementary syntax structure in the ISOBMFF, including a four-character coded box type, the byte count of the box, and the payload. An ISOBMFF file consists of a sequence of boxes, and boxes may contain other boxes. A Movie box (“moov”) contains the metadata for the continuous media streams present in the file, each one represented in the file as a track. The metadata for a track is enclosed in a Track box (“trak”), while the media content of a track is either enclosed in a Media Data box (“mdat”) or directly in a separate file. The media content for tracks consists of a sequence of samples, such as audio or video access units.

[0027] The ISOBMFF specifies the following types of tracks: a media track, which contains an elementary media stream, a hint track, which either includes media transmission instructions or represents a received packet stream, and a timed metadata track, which comprises time-synchronized metadata.

[0028] Although originally designed for storage, the ISOBMFF has proven to be very valuable for streaming, e.g. for progressive download or DASH. For streaming purposes, the movie fragments defined in ISOBMFF can be used.

[0029] The metadata for each track includes a list of sample description entries, each providing the coding or encapsulation format used in the track and the initialization data needed for processing that format. Each sample is associated with one of the sample description entries of the track.

[0030] The ISOBMFF enables specifying sample-specific metadata with various mechanisms. Specific boxes within the Sample Table box (“stbl”) have been standardized to respond to common needs. For example, a Sync Sample box (“stss”) is used to list the random access samples of the track. The sample grouping mechanism enables mapping of samples according to a four-character grouping type into groups of samples sharing the same property specified as a sample group description entry in the file. Several grouping types have been specified in the ISOBMFF.

[0031] High dynamic range (HDR) and wide color gamut (WCG) information can be signaled using the ColourInformationBox, defined in clause 12.1.5 of the ISOBMFF specification. For example, the colour_type may be set equal to ‘nclx,’ which indicates that the most important HDR/WCG information is carried in the fields colour_primaries, transfer_characteristics, matrix_coefficients, and full_range_flag.

[0032] ISOBMFF specifies a restricted scheme design. The restricted scheme design in ISOBMFF is used for handling situations where a file author requires certain actions on the player or renderer, to enable players to simply inspect a file to find out such requirements for rendering a bitstream and stops legacy players from decoding and rendering files that require further processing. The mechanism applies to any type of video codec.

[0033] The mechanism is similar to the content protection transformation, where sample entries are hidden behind generic sample entries, 'encv,' 'enca,' etc., indicating encrypted or encapsulated media. The analogous mechanism for restricted video uses a transformation with the generic sample entry 'resv.' The method may be applied when the content should only be decoded by players that can present it correctly.

[0034] The restricted scheme is specified in clauses 8.15.1 to 8.15.3 of the ISOBMFF specification.

[0035] Clause 8.15.4 of the ISOBMFF specification defines a particular restricted scheme type for frame packed video.

[0036] Dynamic adaptive streaming over HTTP (DASH), specified in ISO/IEC 23009-1, is a standard for HTTP (adaptive) streaming applications. It mainly specifies the format of the media presentation description (MPD), also known as manifest, and the media segment format. The MPD describes the media available on the server and lets the DASH client to autonomously download the media version at the media time it is interested in.

[0037] A typical procedure for DASH based HTTP streaming includes the following steps:

- 1) A DASH client obtains the MPD of a streaming content, e.g., a movie. The MPD includes information on different alternative representations, e.g., bit rate, video resolution, frame rate, audio language, of the streaming content, as well as the URLs of the HTTP resources (the initialization segment and the media segments).
- 2) Based on information in the MPD and the DASH client's local information, e.g., network bandwidth, decoding/display capabilities and user preference, the client requests the desired representation(s), one segment (or a part thereof, e.g., a partial segment) at a time.
- 3) When the DASH client detects a network bandwidth change, the DASH client requests segments of a different representation with a better-matching

bitrate, ideally starting from a segment that starts with a random access point.

[0038] During an HTTP streaming “session,” to respond to a user request to seek backward to a past position or forward to a future position, the DASH client requests past or future segments starting from a segment that is close to the user-requested position and that ideally starts with a random access point. The user may also request to fast-forward the content, which may be realized by requesting data sufficient for decoding only the intra-coded video pictures or only a temporal subset of the video stream.

[0039] In HTTP streaming, such as DASH, frequently used operations include HEAD, GET, and partial GET. The HEAD operation retrieves a header of a file associated with a given uniform resource locator (URL) or uniform resource name (URN), without retrieving a payload associated with the URL or URN. The GET operation retrieves a whole file associated with a given URL or URN. The partial GET operation receives a byte range as an input parameter and retrieves a continuous number of bytes of a file, where the number of bytes correspond to the received byte range. Thus, movie fragments may be provided for HTTP streaming, because a partial GET operation can get one or more individual movie fragments. In a movie fragment, there can be several track fragments of different tracks. In HTTP streaming, a media presentation may be a structured collection of data that is accessible to the client. The client may request and download media data information to present a streaming service to a user.

[0040] In the example of streaming 3GPP data using HTTP streaming, there may be multiple representations for video and/or audio data of multimedia content. As explained below, different representations may correspond to different coding characteristics (e.g., different profiles or levels of a video coding standard), different coding standards or extensions of coding standards (such as multiview and/or scalable extensions), or different bitrates. The manifest of such representations may be defined in a Media Presentation Description (MPD) data structure. A media presentation may correspond to a structured collection of data that is accessible to an HTTP streaming client device. The HTTP streaming client device may request and download media data information to present a streaming service to a user of the client device. A media presentation may be described in the MPD data structure, which may include updates of the MPD.

[0041] A media presentation may contain a sequence of one or more Periods. Each period may extend until the start of the next Period, or until the end of the media presentation, in the case of the last period. Each period may contain one or more representations for the same media content. A representation may be one of a number of alternative encoded versions of audio, video, timed text, or other such data. The representations may differ by encoding types, e.g., by bitrate, resolution, and/or codec for video data and bitrate, language, and/or codec for audio data. The term representation may be used to refer to a section of encoded audio or video data corresponding to a particular period of the multimedia content and encoded in a particular way.

[0042] Representations of a particular period may be assigned to a group indicated by an attribute in the MPD indicative of an adaptation set to which the representations belong. Representations in the same adaptation set are generally considered alternatives to each other, in that a client device can dynamically and seamlessly switch between these representations, e.g., to perform bandwidth adaptation. For example, each representation of video data for a particular period may be assigned to the same adaptation set, such that any of the representations may be selected for decoding to present media data, such as video data or audio data, of the multimedia content for the corresponding period. The media content within one period may be represented by either one representation from group 0, if present, or the combination of at most one representation from each non-zero group, in some examples. Timing data for each representation of a period may be expressed relative to the start time of the period.

[0043] A representation may include one or more segments. Each representation may include an initialization segment, or each segment of a representation may be self-initializing. When present, the initialization segment may contain initialization information for accessing the representation. In general, the initialization segment does not contain media data. A segment may be uniquely referenced by an identifier, such as a uniform resource locator (URL), uniform resource name (URN), or uniform resource identifier (URI). The MPD may provide the identifiers for each segment. In some examples, the MPD may also provide byte ranges in the form of a *range* attribute, which may correspond to the data for a segment within a file accessible by the URL, URN, or URI.

[0044] Different representations may be selected for substantially simultaneous retrieval for different types of media data. For example, a client device may select an audio

representation, a video representation, and a timed text representation from which to retrieve segments. In some examples, the client device may select particular adaptation sets for performing bandwidth adaptation. That is, the client device may select an adaptation set including video representations, an adaptation set including audio representations, and/or an adaptation set including timed text. Alternatively, the client device may select adaptation sets for certain types of media (e.g., video), and directly select representations for other types of media (e.g., audio and/or timed text).

[0045] Virtual reality (VR) is the ability to be virtually present in a non-physical world created by the rendering of natural and/or synthetic image and sound correlated by the movements of the immersed user allowing to interact with that world. With the recent progress made in rendering devices, such as head mounted displays (HMD), and VR video (often also referred to as 360 degree video) creation, a significant quality of experience can be offered. VR applications including gaming, training, education, sports video, online shopping, video entertainment, and so on.

[0046] A typical VR system may include the following components and steps:

- 1) A camera set, which typically includes multiple individual cameras pointing in different directions and ideally collectively covering all viewpoints around the camera set.
- 2) Image stitching, where video pictures taken by the multiple individual cameras are synchronized in the time domain and stitched in the space domain, to be a spherical video, but mapped to a rectangular format, such as equi-rectangular (like a world map) or cube map.
- 3) The video in the mapped rectangular format is encoded/compressed using a video codec, e.g., H.265/HEVC or H.264/AVC.
- 4) The compressed video bitstream(s) may be stored and/or encapsulated in a media format and transmitted (possibly only the subset covering the area being seen by a user) through a network to a receiver.
- 5) The receiver receives the video bitstream(s) or part thereof, possibly encapsulated in a format, and sends the decoded video signal or part thereof to a rendering device.
- 6) The rendering device can be e.g., an HMD, which can track head movement and even eye move moment and rendering the corresponding part of the video such that an immersive experience is delivered to the user.

[0047] The Omnidirectional Media Application Format (OMAF) is being developed by MPEG to define a media application format that enables omnidirectional media applications, focusing on VR applications with 360 degree video and associated audio. OMAF specifies a list of projection methods that can be used for conversion of a spherical or 360° video into a two-dimensional rectangular video, followed by how to store omnidirectional media and the associated metadata using the ISO base media file format (ISOBMFF) and how to encapsulate, signal, and stream omnidirectional media using dynamic adaptive streaming over HTTP (DASH), and finally which video and audio codecs as well as media coding configurations can be used for compression and playback of the omnidirectional media signal.

[0048] OMAF is intended to be standardized as ISO/IEC 23000-20, and a draft specification, referred to as the OMAF Committee Draft (CD), is available at http://wg11.sc29.org/doc_end_user/documents/117_Geneva/wg11/w16636.zip.

[0049] Clause 7.1 of the OMAF CD defines a particular restricted scheme type for VR/omnidirectional/360 video, ‘odvd.’ The OMAF CD specifies that, when scheme_type is equal to ‘odvd,’ the Scheme Information box (‘schi’) needs to contain either a ProjectedOmnidirectionalVideoBox (‘povd’) or a FisheyeOmnidirectionalVideoBox (‘fovvd’). The OMAF CD specifies a ‘povd’ box that contains a ProjectionFormatBox, which carries the geometry_type and the projection_type. Per OMAF CD, the geometry_type may indicate a sphere geometry, for example, and the projection_type may indicate the equirectangular projection, a cubemap projection, or some other projection type. These information pieces are all important for content selection purposes.

[0050] The DASH specification includes the definition of the MPD attributes @mimeType and @codecs, both of which may be conveyed on the level of Adaptation Set, Representation, or Sub-Representation.

[0051] The @mimeType attribute is defined as follows in clause 5.3.7.2 of the DASH specification:

@mimeType	M	specifies the MIME type of the concatenation of the Initialization Segment, if present, and all consecutive Media Segments in the Representation.
-----------	---	---

[0052] Furthermore, in clause 7.3.1 of the DASH specification, the semantics of @mimeType attribute is further clarified as follows for ISOBMFF based media presentations:

“The @mimeType attribute of each Representation shall be provided according to RFC 4337. Additional parameters may be added according to RFC 6381.”

[0053] The @codecs attribute is defined as follows in clause 5.3.7.2 of the DASH specification:

@codecs	O	specifies the codecs present within the Representation. The codec parameters shall also include the profile and level information where applicable. For segment formats defined in this specification this element shall be present and the contents of this attribute shall conform to either the simp-list or fancy-list productions of RFC6381, Section 3.2, without the enclosing DQUOTE characters. The codec identifier for the Representation's media format, mapped into the name space for codecs as specified in RFC6381, Section 3.3, shall be used.
---------	---	---

[0054] Clause E of ISO/IEC 14496-15 defines the 'codecs' parameter for AVC, HEVC, and their extensions.

[0055] According to clause E of ISO/IEC 14496-15 and RFC 6381, the 'codecs' parameter is an optional MIME type parameter. However, ISO/IEC 14496-15 and RFC 6381 do not make it clear whether the 'codecs' parameter can be conveyed as part of the @mimeType attribute.

[0056] As specified in RFC 6381, the 'codecs' parameter is a single value or a comma-separated list of values, where each value consists of one or more dot-separated (e.g., period-delimited) elements. The name space for the first element is determined by the MIME type. The name space for each subsequent element is determined by the preceding element. For the ISOBMFF, the first element of a 'codecs' parameter value is a sample description entry four-character code.

[0057] Existing designs regarding signaling of HDR/WCG video, VR/omnidirectional/360 video, frame packed video, video with display orientation changes, and video stored using a restricted scheme may encounter the following issues:

- 1) There is a lack of a mechanism to indicate the use of a restricted scheme, as well as some important details of the used restricted scheme in MIME type parameters, for example for VR/omnidirectional/360 video and frame packed video. In addition, the following questions/issues are not clear.
 - a. How do DASH clients handle @mimeType containing unrecognized optional parameter? Ignore the unrecognized part and take the rest as if the unrecognized part did not exist? Or ignore the entire Adaptation Set/Representation/Sub-Representation (i.e., not trying to request/process the Adaptation Set/Representation/Sub-Representation containing this @mimeType attribute)?
 - i. The latter seems to make more sense.
 - ii. RFC 4337 / RFC 6381 are silent on this. This should be clearly specified somewhere, preferably in an update to RFC 6381 (which updates RFC 4337, by the way).
 - b. What should the 'codecs' parameter for video stored using a restricted scheme be?
 - c. Should a used restricted scheme be indicated by the 'codecs' parameter or a different/separate MIME type parameter? Is there any backward compatibility issue if the 'codecs' parameter is the same as if no restricted scheme was used while defining/having an additional optional MIME type parameter for indication of the used restricted scheme?
- 2) For video with display orientation changes, a special restricted scheme is missing, and also the first problem above applies.
- 3) There is a lack of a mechanism to include important video information for HDR/WCG video as part of the MIME type parameters.

[0058] FIG. 1 is a block diagram illustrating an example system 10 that implements techniques for streaming media data over a network. In this example, system 10 includes content preparation device 20, server device 60, and client device 40. Client device 40 and server device 60 are communicatively coupled by network 74, which may comprise the Internet. In some examples, content preparation device 20 and server

device 60 may also be coupled by network 74 or another network, or may be directly communicatively coupled. In some examples, content preparation device 20 and server device 60 may comprise the same device.

[0059] Content preparation device 20, in the example of FIG. 1, comprises audio source 22 and video source 24. Audio source 22 may comprise, for example, a microphone that produces electrical signals representative of captured audio data to be encoded by audio encoder 26. Alternatively, audio source 22 may comprise a storage medium storing previously recorded audio data, an audio data generator such as a computerized synthesizer, or any other source of audio data. Video source 24 may comprise a video camera that produces video data to be encoded by video encoder 28, a storage medium encoded with previously recorded video data, a video data generation unit such as a computer graphics source, or any other source of video data. Content preparation device 20 is not necessarily communicatively coupled to server device 60 in all examples, but may store multimedia content to a separate medium that is read by server device 60.

[0060] Raw audio and video data may comprise analog or digital data. Analog data may be digitized before being encoded by audio encoder 26 and/or video encoder 28. Audio source 22 may obtain audio data from a speaking participant while the speaking participant is speaking, and video source 24 may simultaneously obtain video data of the speaking participant. In other examples, audio source 22 may comprise a computer-readable storage medium comprising stored audio data, and video source 24 may comprise a computer-readable storage medium comprising stored video data. In this manner, the techniques described in this disclosure may be applied to live, streaming, real-time audio and video data or to archived, pre-recorded audio and video data.

[0061] Audio frames that correspond to video frames are generally audio frames containing audio data that was captured (or generated) by audio source 22 contemporaneously with video data captured (or generated) by video source 24 that is contained within the video frames. For example, while a speaking participant generally produces audio data by speaking, audio source 22 captures the audio data, and video source 24 captures video data of the speaking participant at the same time, that is, while audio source 22 is capturing the audio data. Hence, an audio frame may temporally correspond to one or more particular video frames. Accordingly, an audio frame corresponding to a video frame generally corresponds to a situation in which audio data and video data were captured at the same time and for which an audio frame and a video

frame comprise, respectively, the audio data and the video data that was captured at the same time.

[0062] In some examples, audio encoder 26 may encode a timestamp in each encoded audio frame that represents a time at which the audio data for the encoded audio frame was recorded, and similarly, video encoder 28 may encode a timestamp in each encoded video frame that represents a time at which the video data for encoded video frame was recorded. In such examples, an audio frame corresponding to a video frame may comprise an audio frame comprising a timestamp and a video frame comprising the same timestamp. Content preparation device 20 may include an internal clock from which audio encoder 26 and/or video encoder 28 may generate the timestamps, or that audio source 22 and video source 24 may use to associate audio and video data, respectively, with a timestamp.

[0063] In some examples, audio source 22 may send data to audio encoder 26 corresponding to a time at which audio data was recorded, and video source 24 may send data to video encoder 28 corresponding to a time at which video data was recorded. In some examples, audio encoder 26 may encode a sequence identifier in encoded audio data to indicate a relative temporal ordering of encoded audio data but without necessarily indicating an absolute time at which the audio data was recorded, and similarly, video encoder 28 may also use sequence identifiers to indicate a relative temporal ordering of encoded video data. Similarly, in some examples, a sequence identifier may be mapped or otherwise correlated with a timestamp.

[0064] Audio encoder 26 generally produces a stream of encoded audio data, while video encoder 28 produces a stream of encoded video data. Each individual stream of data (whether audio or video) may be referred to as an elementary stream. An elementary stream is a single, digitally coded (possibly compressed) component of a representation. For example, the coded video or audio part of the representation can be an elementary stream. An elementary stream may be converted into a packetized elementary stream (PES) before being encapsulated within a video file. Within the same representation, a stream ID may be used to distinguish the PES-packets belonging to one elementary stream from the other. The basic unit of data of an elementary stream is a packetized elementary stream (PES) packet. Thus, coded video data generally corresponds to elementary video streams. Similarly, audio data corresponds to one or more respective elementary streams.

[0065] Many video coding standards, such as ITU-T H.264/AVC and ITU-T H.265/High Efficiency Video Coding (HEVC), define the syntax, semantics, and decoding process for error-free bitstreams, any of which conform to a certain profile or level. Video coding standards typically do not specify the encoder, but the encoder is tasked with guaranteeing that the generated bitstreams are standard-compliant for a decoder. In the context of video coding standards, a “profile” corresponds to a subset of algorithms, features, or tools and constraints that apply to them. As defined by the H.264 standard, for example, a “profile” is a subset of the entire bitstream syntax that is specified by the H.264 standard. A “level” corresponds to the limitations of the decoder resource consumption, such as, for example, decoder memory and computation, which are related to the resolution of the pictures, bit rate, and block processing rate. A profile may be signaled with a `profile_idc` (profile indicator) value, while a level may be signaled with a `level_idc` (level indicator) value.

[0066] The H.264 standard, for example, recognizes that, within the bounds imposed by the syntax of a given profile, it is still possible to require a large variation in the performance of encoders and decoders depending upon the values taken by syntax elements in the bitstream such as the specified size of the decoded pictures. The H.264 standard further recognizes that, in many applications, it is neither practical nor economical to implement a decoder capable of dealing with all hypothetical uses of the syntax within a particular profile. Accordingly, the H.264 standard defines a “level” as a specified set of constraints imposed on values of the syntax elements in the bitstream. These constraints may be simple limits on values. Alternatively, these constraints may take the form of constraints on arithmetic combinations of values (e.g., picture width multiplied by picture height multiplied by number of pictures decoded per second). The H.264 standard further provides that individual implementations may support a different level for each supported profile.

[0067] A decoder conforming to a profile ordinarily supports all the features defined in the profile. For example, as a coding feature, B-picture coding is not supported in the baseline profile of H.264/AVC but is supported in other profiles of H.264/AVC. A decoder conforming to a level should be capable of decoding any bitstream that does not require resources beyond the limitations defined in the level. Definitions of profiles and levels may be helpful for interpretability. For example, during video transmission, a pair of profile and level definitions may be negotiated and agreed for a whole transmission session. More specifically, in H.264/AVC, a level may define limitations

on the number of macroblocks that need to be processed, decoded picture buffer (DPB) size, coded picture buffer (CPB) size, vertical motion vector range, maximum number of motion vectors per two consecutive MBs, and whether a B-block can have sub-macroblock partitions less than 8x8 pixels. In this manner, a decoder may determine whether the decoder is capable of properly decoding the bitstream.

[0068] In the example of FIG. 1, encapsulation unit 30 of content preparation device 20 receives elementary streams comprising coded video data from video encoder 28 and elementary streams comprising coded audio data from audio encoder 26. In some examples, video encoder 28 and audio encoder 26 may each include packetizers for forming PES packets from encoded data. In other examples, video encoder 28 and audio encoder 26 may each interface with respective packetizers for forming PES packets from encoded data. In still other examples, encapsulation unit 30 may include packetizers for forming PES packets from encoded audio and video data.

[0069] Video encoder 28 may encode video data of multimedia content in a variety of ways, to produce different representations of the multimedia content at various bitrates and with various characteristics, such as pixel resolutions, frame rates, conformance to various coding standards, conformance to various profiles and/or levels of profiles for various coding standards, representations having one or multiple views (e.g., for two-dimensional or three-dimensional playback), or other such characteristics. A representation, as used in this disclosure, may comprise one of audio data, video data, text data (e.g., for closed captions), or other such data. The representation may include an elementary stream, such as an audio elementary stream or a video elementary stream. Each PES packet may include a stream_id that identifies the elementary stream to which the PES packet belongs. Encapsulation unit 30 is responsible for assembling elementary streams into video files (e.g., segments) of various representations.

[0070] Encapsulation unit 30 receives PES packets for elementary streams of a representation from audio encoder 26 and video encoder 28 and forms corresponding network abstraction layer (NAL) units from the PES packets. Coded video segments may be organized into NAL units, which provide a “network-friendly” video representation addressing applications such as video telephony, storage, broadcast, or streaming. NAL units can be categorized to Video Coding Layer (VCL) NAL units and non-VCL NAL units. VCL units may contain the core compression engine and may include block, macroblock, and/or slice level data. Other NAL units may be non-VCL NAL units. In some examples, a coded picture in one time instance, normally presented

as a primary coded picture, may be contained in an access unit, which may include one or more NAL units.

[0071] Non-VCL NAL units may include parameter set NAL units and SEI NAL units, among others. Parameter sets may contain sequence-level header information (in sequence parameter sets (SPS)) and the infrequently changing picture-level header information (in picture parameter sets (PPS)). With parameter sets (e.g., PPS and SPS), infrequently changing information need not to be repeated for each sequence or picture, hence coding efficiency may be improved. Furthermore, the use of parameter sets may enable out-of-band transmission of the important header information, avoiding the need for redundant transmissions for error resilience. In out-of-band transmission examples, parameter set NAL units may be transmitted on a different channel than other NAL units, such as SEI NAL units.

[0072] Supplemental Enhancement Information (SEI) may contain information that is not necessary for decoding the coded pictures samples from VCL NAL units, but may assist in processes related to decoding, display, error resilience, and other purposes. SEI messages may be contained in non-VCL NAL units. SEI messages are the normative part of some standard specifications, and thus are not always mandatory for standard compliant decoder implementation. SEI messages may be sequence level SEI messages or picture level SEI messages. Some sequence level information may be contained in SEI messages, such as scalability information SEI messages in the example of SVC and view scalability information SEI messages in MVC. These example SEI messages may convey information on, e.g., extraction of operation points and characteristics of the operation points. In addition, encapsulation unit 30 may form a manifest file, such as a media presentation descriptor (MPD) that describes characteristics of the representations. Encapsulation unit 30 may format the MPD according to extensible markup language (XML).

[0073] Encapsulation unit 30 may provide data for one or more representations of multimedia content, along with the manifest file (e.g., the MPD) to output interface 32. Output interface 32 may comprise a network interface or an interface for writing to a storage medium, such as a universal serial bus (USB) interface, a CD or DVD writer or burner, an interface to magnetic or flash storage media, or other interfaces for storing or transmitting media data. Encapsulation unit 30 may provide data of each of the representations of multimedia content to output interface 32, which may send the data to server device 60 via network transmission or storage media. In the example of FIG. 1,

server device 60 includes storage medium 62 that stores various multimedia contents 64, each including a respective manifest file 66 and one or more representations 68A–68N (representations 68). In some examples, output interface 32 may also send data directly to network 74.

[0074] In some examples, representations 68 may be separated into adaptation sets. That is, various subsets of representations 68 may include respective common sets of characteristics, such as codec, profile and level, resolution, number of views, file format for segments, text type information that may identify a language or other characteristics of text to be displayed with the representation and/or audio data to be decoded and presented, e.g., by speakers, camera angle information that may describe a camera angle or real-world camera perspective of a scene for representations in the adaptation set, rating information that describes content suitability for particular audiences, or the like.

[0075] Manifest file 66 may include data indicative of the subsets of representations 68 corresponding to particular adaptation sets, as well as common characteristics for the adaptation sets. Manifest file 66 may also include data representative of individual characteristics, such as bitrates, for individual representations of adaptation sets. In this manner, an adaptation set may provide for simplified network bandwidth adaptation. Representations in an adaptation set may be indicated using child elements of an adaptation set element of manifest file 66.

[0076] Server device 60 includes request processing unit 70 and network interface 72. In some examples, server device 60 may include a plurality of network interfaces. Furthermore, any or all of the features of server device 60 may be implemented on other devices of a content delivery network, such as routers, bridges, proxy devices, switches, or other devices. In some examples, intermediate devices of a content delivery network may cache data of multimedia content 64, and include components that conform substantially to those of server device 60. In general, network interface 72 is configured to send and receive data via network 74.

[0077] Request processing unit 70 is configured to receive network requests from client devices, such as client device 40, for data of storage medium 62. For example, request processing unit 70 may implement hypertext transfer protocol (HTTP) version 1.1, as described in RFC 2616, “Hypertext Transfer Protocol – HTTP/1.1,” by R. Fielding et al, Network Working Group, IETF, June 1999. That is, request processing unit 70 may be configured to receive HTTP GET or partial GET requests and provide data of multimedia content 64 in response to the requests. The requests may specify a segment

of one of representations 68, e.g., using a URL of the segment. In some examples, the requests may also specify one or more byte ranges of the segment, thus comprising partial GET requests. Request processing unit 70 may further be configured to service HTTP HEAD requests to provide header data of a segment of one of representations 68. In any case, request processing unit 70 may be configured to process the requests to provide requested data to a requesting device, such as client device 40.

[0078] Additionally or alternatively, request processing unit 70 may be configured to deliver media data via a broadcast or multicast protocol, such as eMBMS. Content preparation device 20 may create DASH segments and/or sub-segments in substantially the same way as described, but server device 60 may deliver these segments or sub-segments using eMBMS or another broadcast or multicast network transport protocol. For example, request processing unit 70 may be configured to receive a multicast group join request from client device 40. That is, server device 60 may advertise an Internet protocol (IP) address associated with a multicast group to client devices, including client device 40, associated with particular media content (e.g., a broadcast of a live event). Client device 40, in turn, may submit a request to join the multicast group. This request may be propagated throughout network 74, e.g., routers making up network 74, such that the routers are caused to direct traffic destined for the IP address associated with the multicast group to subscribing client devices, such as client device 40.

[0079] As illustrated in the example of FIG. 1, multimedia content 64 includes manifest file 66, which may correspond to a media presentation description (MPD). Manifest file 66 may contain descriptions of different alternative representations 68 (e.g., video services with different qualities) and the description may include, e.g., codec information, a profile value, a level value, a bitrate, and other descriptive characteristics of representations 68. Client device 40 may retrieve the MPD of a media presentation to determine how to access segments of representations 68.

[0080] In particular, retrieval unit 52 may retrieve configuration data (not shown) of client device 40 to determine decoding capabilities of video decoder 48 and rendering capabilities of video output 44. The configuration data may also include any or all of a language preference selected by a user of client device 40, one or more camera perspectives corresponding to depth preferences set by the user of client device 40, and/or a rating preference selected by the user of client device 40. Retrieval unit 52 may comprise, for example, a web browser or a media client configured to submit HTTP GET and partial GET requests. Retrieval unit 52 may correspond to software

instructions executed by one or more processors or processing units (not shown) of client device 40. In some examples, all or portions of the functionality described with respect to retrieval unit 52 may be implemented in hardware, or a combination of hardware, software, and/or firmware, where requisite hardware may be provided to execute instructions for software or firmware.

[0081] Retrieval unit 52 may compare the decoding and rendering capabilities of client device 40 to characteristics of representations 68 indicated by information of manifest file 66. Retrieval unit 52 may initially retrieve at least a portion of manifest file 66 to determine characteristics of representations 68. For example, retrieval unit 52 may request a portion of manifest file 66 that describes characteristics of one or more adaptation sets. Retrieval unit 52 may select a subset of representations 68 (e.g., an adaptation set) having characteristics that can be satisfied by the coding and rendering capabilities of client device 40. Retrieval unit 52 may then determine bitrates for representations in the adaptation set, determine a currently available amount of network bandwidth, and retrieve segments from one of the representations having a bitrate that can be satisfied by the network bandwidth.

[0082] In general, higher bitrate representations may yield higher quality video playback, while lower bitrate representations may provide sufficient quality video playback when available network bandwidth decreases. Accordingly, when available network bandwidth is relatively high, retrieval unit 52 may retrieve data from relatively high bitrate representations, whereas when available network bandwidth is low, retrieval unit 52 may retrieve data from relatively low bitrate representations. In this manner, client device 40 may stream multimedia data over network 74 while also adapting to changing network bandwidth availability of network 74.

[0083] Additionally or alternatively, retrieval unit 52 may be configured to receive data in accordance with a broadcast or multicast network protocol, such as eMBMS or IP multicast. In such examples, retrieval unit 52 may submit a request to join a multicast network group associated with particular media content. After joining the multicast group, retrieval unit 52 may receive data of the multicast group without further requests issued to server device 60 or content preparation device 20. Retrieval unit 52 may submit a request to leave the multicast group when data of the multicast group is no longer needed, e.g., to stop playback or to change channels to a different multicast group.

[0084] Network interface 54 may receive and provide data of segments of a selected representation to retrieval unit 52, which may in turn provide the segments to decapsulation unit 50. Decapsulation unit 50 may decapsulate elements of a video file into constituent PES streams, depacketize the PES streams to retrieve encoded data, and send the encoded data to either audio decoder 46 or video decoder 48, depending on whether the encoded data is part of an audio or video stream, e.g., as indicated by PES packet headers of the stream. Audio decoder 46 decodes encoded audio data and sends the decoded audio data to audio output 42, while video decoder 48 decodes encoded video data and sends the decoded video data, which may include a plurality of views of a stream, to video output 44.

[0085] Video encoder 28, video decoder 48, audio encoder 26, audio decoder 46, encapsulation unit 30, retrieval unit 52, and decapsulation unit 50 each may be implemented as any of a variety of suitable processing circuitry, as applicable, such as one or more microprocessors, digital signal processors (DSPs), application specific integrated circuits (ASICs), field programmable gate arrays (FPGAs), discrete logic circuitry, software, hardware, firmware or any combinations thereof. Each of video encoder 28 and video decoder 48 may be included in one or more encoders or decoders, either of which may be integrated as part of a combined video encoder/decoder (CODEC). Likewise, each of audio encoder 26 and audio decoder 46 may be included in one or more encoders or decoders, either of which may be integrated as part of a combined CODEC. An apparatus including video encoder 28, video decoder 48, audio encoder 26, audio decoder 46, encapsulation unit 30, retrieval unit 52, and/or decapsulation unit 50 may comprise an integrated circuit, a microprocessor, and/or a wireless communication device, such as a cellular telephone.

[0086] Client device 40, server device 60, and/or content preparation device 20 may be configured to operate in accordance with the techniques of this disclosure. For purposes of example, this disclosure describes these techniques with respect to client device 40 and server device 60. However, it should be understood that content preparation device 20 may be configured to perform these techniques, instead of (or in addition to) server device 60.

[0087] Encapsulation unit 30 may form NAL units comprising a header that identifies a program to which the NAL unit belongs, as well as a payload, e.g., audio data, video data, or data that describes the transport or program stream to which the NAL unit corresponds. For example, in H.264/AVC, a NAL unit includes a 1-byte header and a

payload of varying size. A NAL unit including video data in its payload may comprise various granularity levels of video data. For example, a NAL unit may comprise a block of video data, a plurality of blocks, a slice of video data, or an entire picture of video data. Encapsulation unit 30 may receive encoded video data from video encoder 28 in the form of PES packets of elementary streams. Encapsulation unit 30 may associate each elementary stream with a corresponding program.

[0088] Encapsulation unit 30 may also assemble access units from a plurality of NAL units. In general, an access unit may comprise one or more NAL units for representing a frame of video data, as well audio data corresponding to the frame when such audio data is available. An access unit generally includes all NAL units for one output time instance, e.g., all audio and video data for one time instance. For example, if each view has a frame rate of 20 frames per second (fps), then each time instance may correspond to a time interval of 0.05 seconds. During this time interval, the specific frames for all views of the same access unit (the same time instance) may be rendered simultaneously. In one example, an access unit may comprise a coded picture in one time instance, which may be presented as a primary coded picture.

[0089] Accordingly, an access unit may comprise all audio and video frames of a common temporal instance, e.g., all views corresponding to time X . This disclosure also refers to an encoded picture of a particular view as a “view component.” That is, a view component may comprise an encoded picture (or frame) for a particular view at a particular time. Accordingly, an access unit may be defined as comprising all view components of a common temporal instance. The decoding order of access units need not necessarily be the same as the output or display order.

[0090] A media presentation may include a media presentation description (MPD), which may contain descriptions of different alternative representations (e.g., video services with different qualities) and the description may include, e.g., codec information, a profile value, and a level value. An MPD is one example of a manifest file, such as manifest file 66. Client device 40 may retrieve the MPD of a media presentation to determine how to access movie fragments of various presentations. Movie fragments may be located in movie fragment boxes (moof boxes) of video files.

[0091] Manifest file 66 (which may comprise, for example, an MPD) may advertise availability of segments of representations 68. That is, the MPD may include information indicating the wall-clock time at which a first segment of one of representations 68 becomes available, as well as information indicating the durations of

segments within representations 68. In this manner, retrieval unit 52 of client device 40 may determine when each segment is available, based on the starting time as well as the durations of the segments preceding a particular segment.

[0092] After encapsulation unit 30 has assembled NAL units and/or access units into a video file based on received data, encapsulation unit 30 passes the video file to output interface 32 for output. In some examples, encapsulation unit 30 may store the video file locally or send the video file to a remote server via output interface 32, rather than sending the video file directly to client device 40. Output interface 32 may comprise, for example, a transmitter, a transceiver, a device for writing data to a computer-readable medium such as, for example, an optical drive, a magnetic media drive (e.g., floppy drive), a universal serial bus (USB) port, a network interface, or other output interface. Output interface 32 outputs the video file to a computer-readable medium, such as, for example, a transmission signal, a magnetic medium, an optical medium, a memory, a flash drive, or other computer-readable medium.

[0093] Network interface 54 may receive a NAL unit or access unit via network 74 and provide the NAL unit or access unit to decapsulation unit 50, via retrieval unit 52. Decapsulation unit 50 may decapsulate elements of a video file into constituent PES streams, depacketize the PES streams to retrieve encoded data, and send the encoded data to either audio decoder 46 or video decoder 48, depending on whether the encoded data is part of an audio or video stream, e.g., as indicated by PES packet headers of the stream. Audio decoder 46 decodes encoded audio data and sends the decoded audio data to audio output 42, while video decoder 48 decodes encoded video data and sends the decoded video data, which may include a plurality of views of a stream, to video output 44.

[0094] FIG. 2 is a block diagram illustrating an example set of components of retrieval unit 52 of FIG. 1 in greater detail. In this example, retrieval unit 52 includes eMBMS middleware unit 100, DASH client 110, and media application 112.

[0095] In this example, eMBMS middleware unit 100 further includes eMBMS reception unit 106, cache 104, and server unit 102. In this example, eMBMS reception unit 106 is configured to receive data via eMBMS, e.g., according to File Delivery over Unidirectional Transport (FLUTE), described in T. Paila et al., “FLUTE—File Delivery over Unidirectional Transport,” Network Working Group, RFC 6726, Nov. 2012, available at <http://tools.ietf.org/html/rfc6726>. That is, eMBMS reception unit 106 may receive files via broadcast from, e.g., server device 60, which may act as a BM-SC.

[0096] As eMBMS middleware unit 100 receives data for files, eMBMS middleware unit may store the received data in cache 104. Cache 104 may comprise a computer-readable storage medium, such as flash memory, a hard disk, RAM, or any other suitable storage medium.

[0097] Local server unit 102 may act as a server for DASH client 110. For example, local server unit 102 may provide a MPD file or other manifest file to DASH client 110. Local server unit 102 may advertise availability times for segments in the MPD file, as well as hyperlinks from which the segments can be retrieved. These hyperlinks may include a localhost address prefix corresponding to client device 40 (e.g., 127.0.0.1 for IPv4). In this manner, DASH client 110 may request segments from local server unit 102 using HTTP GET or partial GET requests. For example, for a segment available from link <http://127.0.0.1/rep1/seg3>, DASH client 110 may construct an HTTP GET request that includes a request for <http://127.0.0.1/rep1/seg3>, and submit the request to local server unit 102. Local server unit 102 may retrieve requested data from cache 104 and provide the data to DASH client 110 in response to such requests.

[0098] In accordance with the techniques of this disclosure, encapsulation unit 30 may signal, and retrieval unit 52 may receive, important video information related to any or all of video data stored using a restricted scheme, HDR/WCG video, VR/omnidirectional/360 video, frame packed video, and video with display orientation changes, such that the important video information can be convenient accessed by application clients, such as DASH clients, to make content rejection/selection/accepting/requesting decisions. As noted above, important video information may include information that may be used for content selection, e.g., selection by retrieval unit 52 of a video track or a part thereof for consumption.

[0099] The techniques of this disclosure may overcome the problems noted above. For example, to address the first problem, encapsulation unit 30 and retrieval unit 52 may be configured to use a new format for the '@codecs' parameter, where the indication of the use of a restricted scheme is included. In one example, a value of the '@codecs' parameter is defined as follows, such that all information considered as important for the video is included in the 'codecs' parameter. In this example, a first element of the '@codecs' parameter is the sample entry type code of a track using a restricted scheme (that is, the track stores media data, such as video data, according to a restricted scheme), e.g., 'resv.' In this example, the second element is the restricted scheme type code, e.g., 'stvi' for frame packed video and 'odvd' for omnidirectional video.

[0100] Alternatively, for omnidirectional video, the second element is 'povd' for projected omnidirectional video or 'fovvd' for fisheye omnidirectional video.

Alternatively, for projected omnidirectional video, the second element indicates the projection type, e.g., 'erp' for the equirectangular projection or 'cmp' for the cubemap projection.

[0101] More information including important details of the particular type of restricted scheme may be included in subsequent elements of the manifest file. For example, if the second element is 'odvd,' the third element exists and is 'povd' for projected omnidirectional video or 'fovvd' for fisheye omnidirectional video. Alternatively, if the second element is 'odvd' and the third element is 'povd,' the fourth element exists and indicates the projection type, e.g., 'erp' for the equirectangular projection or 'cmp' for the cubemap projection. For example, the starting four elements of a value of the 'codecs' parameter of an equirectangular projected omnidirectional video would be "resv.odvd.povd.erp."

[0102] The above elements may be further followed by the usual elements of a value of the 'codecs' parameters as defined in clause E of ISO/IEC 14496-15. For example, a value of the 'codecs' parameter of an equirectangular projected omnidirectional video conforming to HEVC, progressive, non-packed, Main Profile, Main Tier, and Level 3.1 may be "resv.odvd.povd.erp.hev1.1.6.L93.B0."

[0103] In this manner, encapsulation unit 30 may signal, and retrieval unit 52 may receive, any or all of the values 'resv,' 'stvi,' 'odvd,' 'povd,' 'fovvd,' 'erp,' or 'cmp' as part of the '@codecs' parameter, as discussed above, e.g., within manifest file 66. Furthermore, retrieval unit 52 may determine one of representations 68 to retrieve, based on the values for the @codecs parameters signaled in manifest file 66 for representations 68, as well as capabilities of video decoder 48.

[0104] In a second alternative example, a value of the 'codecs' parameter is defined as follows, such that some of information considered as important for the video is included in the 'codecs' parameter while more details of the information considered as important for the video is included in a different MIME type parameter. In this example, the first element of the @codecs parameter is the sample entry type code of a track using a restricted scheme, i.e., 'resv.' In this example, the second element may be the restricted scheme type code, e.g., 'stvi' for frame packed video and 'odvd' for omnidirectional video. Alternatively, for omnidirectional video, the second element may be 'povd' for projected omnidirectional video or 'fovvd' for fisheye omnidirectional video.

[0105] The above two elements may further be followed by the usual elements of a value of the 'codecs' parameters as defined in clause E of ISO/IEC 14496-15. For example, a value of the 'codecs' parameter of an equirectangular projected omnidirectional video conforming to HEVC, progressive, non-packed, Main Profile, Main Tier, and Level 3.1 may be "resv.odvd.hev1.1.6.L93.B0." Alternatively, a value of the 'codecs' parameter of the above video may be "resv.povd.hev1.1.6.L93.B0." Alternatively, a value of the 'codecs' parameter of the above video may be "resv.erp.hev1.1.6.L93.B0."

[0106] In addition to the new format of the 'codecs' parameter discussed above, new optional MIME type parameters may be used that contain more details of the particular type of restricted scheme. The format of these optional MIME type parameters is similar to the 'codecs' parameter, i.e., they may be a single value or a comma-separated list of values, where each value in the comma-separated list includes one or more dot-separated (e.g., period-delimited) elements, and the name space for each element may be determined by the preceding element. As one example, an optional MIME type parameter 'odvdinfo' may contain more details of omnidirectional video. According to this example, for a value of 'odvdinfo,' the first element may be 'povd' for projected omnidirectional video or 'fov' for fisheye omnidirectional video, and in the former case of 'povd,' the second element exists and indicates the projection type, e.g., 'erp' for the equirectangular projection or 'cmp' for the cubemap projection. More elements can be added to contain more information. Alternatively, as another example, an optional MIME type parameter 'fpvdinfo' may contain more details of frame packed video. For example, 'fpvdinfo' may include elements corresponding to stereo_scheme and stereo_indication_type as defined in clause 8.15.4.2 of the ISO/BMFF specification.

[0107] As another example, to address the second problem (which may be performed in combination with any of the techniques above to address the first problem), encapsulation unit 30 and retrieval unit 52 may be configured to use a new restricted scheme type that indicates that a track of a representation carries video data with display orientation changes. For example, a 4-character code 'vdoc' may indicate that a corresponding track carries video with display orientation changes.

[0108] In one example, no further information is provided regarding the display orientation changes, and a SchemeInformationBox may be absent in a RestrictedSchemeInfoBox. In another alternative, whether either or both of rotation and flipping is further indicated by a new box contained in the SchemeInformationBox.

This new box may include, for example, a field named `display_orientation_change_type`, with value 0 indicating both rotation and flipping apply, value 1 indicating only rotation applies, and value 2 indicating only flipping applies. Thus, encapsulation unit 30 may set the value of the `display_orientation_change_type` field based on whether a corresponding track includes either or both of rotation and/or flipping, and retrieval unit 52 may determine whether a track of one of representations 68 includes display orientation changes, and if so, whether the changes include either or both of rotation and/or flipping, from the value of the field.

[0109] Furthermore, the new format for the 'codecs' parameter as defined in above may also be applied. For example, a value of the 'codecs' parameter of video with display orientation the changes conforming to HEVC, progressive, non-packed, Main Profile, Main Tier, and Level 3.1 may be "resv.vdoc.hev1.1.6.L93.B0." Similarly, as in some alternatives of the examples above, some more information, such as `display_orientation_change_type`, may be included in a third element of a value of the 'codecs' parameter, and the remaining elements may be pushed further down in order by one element.

[0110] As another example, to address the third problem (which may be performed in combination with any of the techniques above to address the first problem and/or the techniques above to address the second problem), encapsulation unit 30 and retrieval unit 52 may be configured to use an optional MIME type parameter 'hdrinfo,' which may contain important information of HDR/WCG video. The format of this optional MIME type parameters may be a single value or a comma-separated (e.g., period-delimited) list of values, where each value includes one or more dot-separated elements. For example, a value of the 'hdrinfo' parameter may contain four fields, in the form of "element1.element2.element3.element4," where the four elements 1 to 4 may be the hexadecimal representations of the fields `colour primaries`, `transfer characteristics`, `matrix coeffs`, and `full_range_flag`, respectively, as defined in clause 12.1.5 of the ISOBMFF specification.

[0111] FIG. 3 is a conceptual diagram illustrating elements of example multimedia content 120. Multimedia content 120 may correspond to multimedia content 64 (FIG. 1), or another multimedia content stored in storage medium 62. In the example of FIG. 3, multimedia content 120 includes media presentation description (MPD) 122 and a plurality of representations 124A–124N (representations 124). Representation 124A

includes optional header data 126 and segments 128A–128N (segments 128), while representation 124N includes optional header data 130 and segments 132A–132N (segments 132). The letter N is used to designate the last movie fragment in each of representations 124 as a matter of convenience. In some examples, there may be different numbers of movie fragments between representations 124.

[0112] MPD 122 may comprise a data structure separate from representations 124. MPD 122 may correspond to manifest file 66 of FIG. 1. Likewise, representations 124 may correspond to representations 68 of FIG. 2. In general, MPD 122 may include data that generally describes characteristics of representations 124, such as coding and rendering characteristics, adaptation sets, a profile to which MPD 122 corresponds, text type information, camera angle information, rating information, trick mode information (e.g., information indicative of representations that include temporal sub-sequences), and/or information for retrieving remote periods (e.g., for targeted advertisement insertion into media content during playback).

[0113] Header data 126, when present, may describe characteristics of segments 128, e.g., temporal locations of random access points (RAPs, also referred to as stream access points (SAPs)), which of segments 128 includes random access points, byte offsets to random access points within segments 128, uniform resource locators (URLs) of segments 128, or other aspects of segments 128. Header data 130, when present, may describe similar characteristics for segments 132. Additionally or alternatively, such characteristics may be fully included within MPD 122.

[0114] Segments 128, 132 include one or more coded video samples, each of which may include frames or slices of video data. Each of the coded video samples of segments 128 may have similar characteristics, e.g., height, width, and bandwidth requirements. Such characteristics may be described by data of MPD 122, though such data is not illustrated in the example of FIG. 3. MPD 122 may include characteristics as described by the 3GPP Specification, with the addition of any or all of the signaled information described in this disclosure.

[0115] Each of segments 128, 132 may be associated with a unique uniform resource locator (URL). Thus, each of segments 128, 132 may be independently retrievable using a streaming network protocol, such as DASH. In this manner, a destination device, such as client device 40, may use an HTTP GET request to retrieve segments 128 or 132. In some examples, client device 40 may use HTTP partial GET requests to retrieve specific byte ranges of segments 128 or 132.

[0116] In accordance with the techniques of this disclosure, MPD 122 may include any or all of the various example MIME type information discussed above. For example, MPD 122 may include an @codecs parameter as discussed above that may indicate, for example, a sample entry type code of a track using a restricted scheme, a restricted scheme type code, and additional information, e.g., for omnidirectional video, projection types, or the like. Additionally or alternatively, MPD 122 may include information indicating whether display orientation changes apply to one of representations 124A–124N, and if so, a type of display orientation change (e.g., either or both of rotation and/or flipping). Additionally or alternatively, the display orientation change information may be provided in header data 126, 130, and/or header data of any or all of segments 128, 132. Additionally or alternatively, MPD 122, header data 126, 130, and/or header data of any or all of segments 128, 132 may contain important information of HDR/WCG video, as discussed above.

[0117] FIG. 4 is a block diagram illustrating elements of an example video file 150, which may correspond to a segment of a representation, such as one of segments 114, 124 of FIG. 3. Each of segments 128, 132 may include data that conforms substantially to the arrangement of data illustrated in the example of FIG. 4. Video file 150 may be said to encapsulate a segment. As described above, video files in accordance with the ISO base media file format and extensions thereof store data in a series of objects, referred to as “boxes.” In the example of FIG. 4, video file 150 includes file type (FTYP) box 152, movie (MOOV) box 154, segment index (sidx) boxes 162, movie fragment (MOOF) boxes 164, and movie fragment random access (MFRA) box 166. Although FIG. 4 represents an example of a video file, it should be understood that other media files may include other types of media data (e.g., audio data, timed text data, or the like) that is structured similarly to the data of video file 150, in accordance with the ISO base media file format and its extensions.

[0118] File type (FTYP) box 152 generally describes a file type for video file 150. File type box 152 may include data that identifies a specification that describes a best use for video file 150. File type box 152 may alternatively be placed before MOOV box 154, movie fragment boxes 164, and/or MFRA box 166.

[0119] In some examples, a Segment, such as video file 150, may include an MPD update box (not shown) before FTYP box 152. The MPD update box may include information indicating that an MPD corresponding to a representation including video file 150 is to be updated, along with information for updating the MPD. For example,

the MPD update box may provide a URI or URL for a resource to be used to update the MPD. As another example, the MPD update box may include data for updating the MPD. In some examples, the MPD update box may immediately follow a segment type (STYP) box (not shown) of video file 150, where the STYP box may define a segment type for video file 150. FIG 7, discussed in greater detail below, provides additional information with respect to the MPD update box.

[0120] MOOV box 154, in the example of FIG. 4, includes movie header (MVHD) box 156, track (TRAK) box 158, and one or more movie extends (MVEX) boxes 160. In general, MVHD box 156 may describe general characteristics of video file 150. For example, MVHD box 156 may include data that describes when video file 150 was originally created, when video file 150 was last modified, a timescale for video file 150, a duration of playback for video file 150, or other data that generally describes video file 150.

[0121] TRAK box 158 may include data for a track of video file 150. TRAK box 158 may include a track header (TKHD) box that describes characteristics of the track corresponding to TRAK box 158. In some examples, TRAK box 158 may include coded video pictures, while in other examples, the coded video pictures of the track may be included in movie fragments 164, which may be referenced by data of TRAK box 158 and/or sidx boxes 162.

[0122] In some examples, video file 150 may include more than one track. Accordingly, MOOV box 154 may include a number of TRAK boxes equal to the number of tracks in video file 150. TRAK box 158 may describe characteristics of a corresponding track of video file 150. For example, TRAK box 158 may describe temporal and/or spatial information for the corresponding track. A TRAK box similar to TRAK box 158 of MOOV box 154 may describe characteristics of a parameter set track, when encapsulation unit 30 (FIG. 3) includes a parameter set track in a video file, such as video file 150. Encapsulation unit 30 may signal the presence of sequence level SEI messages in the parameter set track within the TRAK box describing the parameter set track.

[0123] MVEX boxes 160 may describe characteristics of corresponding movie fragments 164, e.g., to signal that video file 150 includes movie fragments 164, in addition to video data included within MOOV box 154, if any. In the context of streaming video data, coded video pictures may be included in movie fragments 164

rather than in MOOV box 154. Accordingly, all coded video samples may be included in movie fragments 164, rather than in MOOV box 154.

[0124] MOOV box 154 may include a number of MVEX boxes 160 equal to the number of movie fragments 164 in video file 150. Each of MVEX boxes 160 may describe characteristics of a corresponding one of movie fragments 164. For example, each MVEX box may include a movie extends header box (MEHD) box that describes a temporal duration for the corresponding one of movie fragments 164.

[0125] As noted above, encapsulation unit 30 may store a sequence data set in a video sample that does not include actual coded video data. A video sample may generally correspond to an access unit, which is a representation of a coded picture at a specific time instance. In the context of AVC, the coded picture include one or more VCL NAL units which contains the information to construct all the pixels of the access unit and other associated non-VCL NAL units, such as SEI messages. Accordingly, encapsulation unit 30 may include a sequence data set, which may include sequence level SEI messages, in one of movie fragments 164. Encapsulation unit 30 may further signal the presence of a sequence data set and/or sequence level SEI messages as being present in one of movie fragments 164 within the one of MVEX boxes 160 corresponding to the one of movie fragments 164.

[0126] SIDX boxes 162 are optional elements of video file 150. That is, video files conforming to the 3GPP file format, or other such file formats, do not necessarily include SIDX boxes 162. In accordance with the example of the 3GPP file format, a SIDX box may be used to identify a sub-segment of a segment (e.g., a segment contained within video file 150). The 3GPP file format defines a sub-segment as “a self-contained set of one or more consecutive movie fragment boxes with corresponding Media Data box(es) and a Media Data Box containing data referenced by a Movie Fragment Box must follow that Movie Fragment box and precede the next Movie Fragment box containing information about the same track.” The 3GPP file format also indicates that a SIDX box “contains a sequence of references to subsegments of the (sub)segment documented by the box. The referenced subsegments are contiguous in presentation time. Similarly, the bytes referred to by a Segment Index box are always contiguous within the segment. The referenced size gives the count of the number of bytes in the material referenced.”

[0127] SIDX boxes 162 generally provide information representative of one or more sub-segments of a segment included in video file 150. For instance, such information

may include playback times at which sub-segments begin and/or end, byte offsets for the sub-segments, whether the sub-segments include (e.g., start with) a stream access point (SAP), a type for the SAP (e.g., whether the SAP is an instantaneous decoder refresh (IDR) picture, a clean random access (CRA) picture, a broken link access (BLA) picture, or the like), a position of the SAP (in terms of playback time and/or byte offset) in the sub-segment, and the like.

[0128] Movie fragments 164 may include one or more coded video pictures. In some examples, movie fragments 164 may include one or more groups of pictures (GOPs), each of which may include a number of coded video pictures, e.g., frames or pictures. In addition, as described above, movie fragments 164 may include sequence data sets in some examples. Each of movie fragments 164 may include a movie fragment header box (MFHD, not shown in FIG. 4). The MFHD box may describe characteristics of the corresponding movie fragment, such as a sequence number for the movie fragment. Movie fragments 164 may be included in order of sequence number in video file 150.

[0129] MFRA box 166 may describe random access points within movie fragments 164 of video file 150. This may assist with performing trick modes, such as performing seeks to particular temporal locations (i.e., playback times) within a segment encapsulated by video file 150. MFRA box 166 is generally optional and need not be included in video files, in some examples. Likewise, a client device, such as client device 40, does not necessarily need to reference MFRA box 166 to correctly decode and display video data of video file 150. MFRA box 166 may include a number of track fragment random access (TFRA) boxes (not shown) equal to the number of tracks of video file 150, or in some examples, equal to the number of media tracks (e.g., non-hint tracks) of video file 150.

[0130] In some examples, movie fragments 164 may include one or more stream access points (SAPs), such as IDR pictures. Likewise, MFRA box 166 may provide indications of locations within video file 150 of the SAPs. Accordingly, a temporal sub-sequence of video file 150 may be formed from SAPs of video file 150. The temporal sub-sequence may also include other pictures, such as P-frames and/or B-frames that depend from SAPs. Frames and/or slices of the temporal sub-sequence may be arranged within the segments such that frames/slices of the temporal sub-sequence that depend on other frames/slices of the sub-sequence can be properly decoded. For example, in the hierarchical arrangement of data, data used for prediction for other data may also be included in the temporal sub-sequence.

[0131] In accordance with the techniques of this disclosure, MOOV box 154 may include one or more boxes (such as a SchemeInformationBox and/or a RestrictedSchemeInfoBox) that indicates whether either or both of rotation and/or flipping is applied to video data included in movie fragments 164, as discussed above. Additionally or alternatively, MOOV box 154 may contain important information of HDR/WCG video, as discussed above.

[0132] Likewise, video data of a track of video file 150 may be stored according to a restricted scheme, such as omnidirectional video, frame packed video, or the like. A manifest file, such as an MPD, may include data indicating the restricted scheme for the video data of the track, as discussed above.

[0133] FIG. 5 is a flowchart illustrating an example method in accordance with the techniques of this disclosure. The method of FIG. 5 is explained as being performed by retrieval unit 52 of client device 40 of FIG. 1, although it should be understood that other devices may be configured to perform this or a similar method.

[0134] Initially, retrieval unit 52 retrieves a manifest file (180). The manifest file may correspond to manifest file 66 of FIG. 1. The manifest file may be, for example, a media presentation description (MPD). In accordance with the techniques of this disclosure, the manifest file may include data that specifies one or more codecs for one or more representations, e.g., one or more of representations 68 of FIG. 1.

[0135] Retrieval unit 52 may then extract the data that specifies the one or more codecs. In particular, according to the techniques of this disclosure, retrieval unit 52 may extract a sample entry type code from the manifest file (182). The sample entry type code may be for a track of one of the representations corresponding to the manifest file. As discussed above, the sample entry type code may comprise “resv” to indicate that the track stores video data using a restricted scheme.

[0136] Retrieval unit 52 may then extract a restricted scheme type code from the manifest file (184). For example, if the video data of the track is stored using a frame packed video scheme, retrieval unit 52 may extract “stvi” from the manifest file for the track, indicating that the track stores video data using the frame packed video scheme. As another example, retrieval unit 52 may determine that the video data of the track is stored using an omnidirectional video scheme in response to extracting “odvd” from the manifest file for the track. As yet another example, “fovvd” may indicate that the video data of the track is stored using fisheye omnidirectional video data, “erp” for an equirectangular projection scheme, or “cmp” for a cubemap projection scheme.

[0137] Although FIG. 5 illustrates an example in which two elements are extracted, it should be understood that retrieval unit 52 may extract additional elements, e.g., in a period- or comma-delimited format. In some examples, retrieval unit 52 may extract a set of values for one or more @codecs parameters, where the values for the @codecs parameters may be respective period-delimited lists of elements. For example, retrieval unit 52 may extract “resv.odvd.povd.erp” for a track and determine that “resv” indicates that the track includes video data stored using a restricted scheme, “odvd” indicates that the video data is omnidirectional video data, “povd” indicates that the video data is projected omnidirectional video data, and “erp” indicates that the video data is equirectangular projected.

[0138] Furthermore, as discussed above, retrieval unit 52 may extract additional elements of a different MIME type parameter from the manifest file. The MIME type parameter may be, for example, “odvdinfo.” The additional elements may be, for example, “povd,” “fov,” “erp,” “cmp,” or the like. Additionally or alternatively, the MIME type parameter may be “fpvdinfo” to specify frame packed video information, such as a stereo scheme and/or a stereo indication type.

[0139] Retrieval unit 52 may then retrieve media data using the extracted codes (186). For example, retrieval unit 52 may retrieve video data of a scheme that is supported by, e.g., video decoder 48 and video output 44 (FIG. 1), and also avoid retrieving video data of schemes that are not supported by video decoder 48 and video output 44. For example, if video decoder 48 is capable of decoding, and video output 44 is capable of displaying, video data of an omnidirectional format, retrieval unit 52 may search the manifest file for a representation including a track storing video data using the omnidirectional video scheme, e.g., “resv.odvd.” Likewise, if video decoder 48 is not capable of decoding frame packed video data, retrieval unit 52 may avoid retrieving video data of a track indicated as having “resv.stvi” for the @codecs parameter in the manifest file.

[0140] In this manner, the method of FIG. 5 represents an example of a method including retrieving a manifest file specifying data for at least one representation of a media presentation, wherein the manifest file includes data that specifies one or more codecs for the at least one representation, extracting, from the manifest file, the data that specifies the one or more codecs, including extracting a first element representing a sample entry type code of a track of the at least one representation, wherein the first element represents that the track includes video data stored using a restricted scheme,

and extracting a second element representing a restricted scheme type code for the restricted scheme of the track, and retrieving data of the at least one representation based on the first element and the second element.

[0141] In one or more examples, the functions described may be implemented in hardware, software, firmware, or any combination thereof. If implemented in software, the functions may be stored on or transmitted over as one or more instructions or code on a computer-readable medium and executed by a hardware-based processing unit. Computer-readable media may include computer-readable storage media, which corresponds to a tangible medium such as data storage media, or communication media including any medium that facilitates transfer of a computer program from one place to another, e.g., according to a communication protocol. In this manner, computer-readable media generally may correspond to (1) tangible computer-readable storage media which is non-transitory or (2) a communication medium such as a signal or carrier wave. Data storage media may be any available media that can be accessed by one or more computers or one or more processors to retrieve instructions, code, and/or data structures for implementation of the techniques described in this disclosure. A computer program product may include a computer-readable medium.

[0142] By way of example, and not limitation, such computer-readable storage media can comprise RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage, or other magnetic storage devices, flash memory, or any other medium that can be used to store desired program code in the form of instructions or data structures and that can be accessed by a computer. Also, any connection is properly termed a computer-readable medium. For example, if instructions are transmitted from a website, server, or other remote source using a coaxial cable, fiber optic cable, twisted pair, digital subscriber line (DSL), or wireless technologies such as infrared, radio, and microwave, then the coaxial cable, fiber optic cable, twisted pair, DSL, or wireless technologies such as infrared, radio, and microwave are included in the definition of medium. It should be understood, however, that computer-readable storage media and data storage media do not include connections, carrier waves, signals, or other transitory media, but are instead directed to non-transitory, tangible storage media. Disk and disc, as used herein, includes compact disc (CD), laser disc, optical disc, digital versatile disc (DVD), floppy disk and Blu-ray disc where disks usually reproduce data magnetically, while discs reproduce data optically with lasers. Combinations of the above should also be included within the scope of computer-readable media.

[0143] Instructions may be executed by one or more processors, such as one or more digital signal processors (DSPs), general purpose microprocessors, application specific integrated circuits (ASICs), field programmable logic arrays (FPGAs), or other equivalent integrated or discrete logic circuitry. Accordingly, the term “processor,” as used herein may refer to any of the foregoing structure or any other structure suitable for implementation of the techniques described herein. In addition, in some aspects, the functionality described herein may be provided within dedicated hardware and/or software modules configured for encoding and decoding, or incorporated in a combined codec. Also, the techniques could be fully implemented in one or more circuits or logic elements.

[0144] The techniques of this disclosure may be implemented in a wide variety of devices or apparatuses, including a wireless handset, an integrated circuit (IC) or a set of ICs (e.g., a chip set). Various components, modules, or units are described in this disclosure to emphasize functional aspects of devices configured to perform the disclosed techniques, but do not necessarily require realization by different hardware units. Rather, as described above, various units may be combined in a codec hardware unit or provided by a collection of interoperative hardware units, including one or more processors as described above, in conjunction with suitable software and/or firmware.

[0145] Various examples have been described. These and other examples are within the scope of the following claims.

CLAIMS

1. A method of retrieving media data, the method comprising:
 - retrieving a manifest file specifying data for at least one representation of a media presentation, wherein the manifest file includes data that specifies one or more codecs for the at least one representation, the at least one representation including one or more media files including media data, the manifest file being separate from the at least one representation, the manifest file not including media data;
 - extracting, from the manifest file, the data that specifies the one or more codecs, including:
 - extracting a first element representing a sample entry type code of a track of the at least one representation, wherein the first element represents that the track includes video data stored using a restricted scheme; and
 - extracting a second element representing a restricted scheme type code for the restricted scheme of the track; and
 - retrieving the one or more media files including the media data of the at least one representation based on the first element and the second element.
2. The method of claim 1, wherein extracting the data that specifies the one or more codecs comprises:
 - extracting a first set of values for one or more @codecs parameters, wherein the values for the @codecs parameters comprise respective period-delimited lists of elements including the first element and the second element; and
 - extracting a second set of values comprising additional information for the @codecs parameters from a different MIME type parameter.
3. The method of claim 2, wherein the additional information comprises information indicating a particular type of restricted scheme.
4. The method of claim 2, wherein the different MIME type parameter comprises 'odvinfo,' and wherein extracting the second set of values comprises extracting at least one of 'povd,' 'fovvd,' 'erp,' or 'cmp.'

5. The method of claim 4, further comprising determining that the track of the at least one representation comprises video data including:

projected omnidirectional video data when the restricted scheme type code comprises 'povd;'

fish-eye omnidirectional video data when the restricted scheme type code comprises 'fov;'

equi-rectangular projected omnidirectional video data when the restricted scheme type code comprises 'erp;' or

cubemap projected omnidirectional video data when the restricted scheme type code comprises 'cmp.'

6. The method of claim 2, wherein the additional information comprises information indicating a particular type of frame packed video data.

7. The method of claim 6, wherein the different MIME type parameter comprises 'fpvinfo.'

8. The method of claim 6, wherein the particular type of frame packed video data comprises elements corresponding to stereo_scheme and stereo_indication_type.

9. The method of claim 1, further comprising:

receiving one or more values for a MIME type parameter 'hdrinfo' for the track, the values for the MIME type parameter including information for at least one of high dynamic range (HDR) or wide color gamut (WCG) video data for the track,

wherein retrieving the media data of the at least one representation comprises retrieving media data of the track based on the values for the MIME type parameter.

10. The method of claim 9, wherein the one or more values are a single value for the MIME type parameter 'hdrinfo.'

11. The method of claim 9, wherein the one or more values are a comma-separated list of values for the MIME type parameter 'hdrinfo,' each of the values in the comma-separated list comprising a single element or a dot-separated list of elements.

12. The method of claim 11, wherein at least one of the values in the comma-separated list comprises a dot-separated list of elements, the elements comprising hexadecimal representations of HDR or WCG fields, including colour_primaries, transfer_characteristics, matrix_coeffs, and full_range_flag.
13. The method of claim 1, wherein the manifest file comprises a dynamic adaptive streaming over HTTP (DASH) media presentation description (MPD).
14. The method of claim 1, wherein the manifest file specifies data for a plurality of representations including the at least one representation, including data that specifies one or more codecs for each of the plurality of representations.
15. The method of claim 14, further comprising selecting the at least one representation in response to determining that a client device includes a video codec conforming to the sample entry type code and the scheme type code.
16. The method of claim 1, wherein the representation comprises a plurality of segments, each of the segments comprising an individually retrievable file associated with a unique uniform resource locator (URL).
17. The method of claim 1, wherein retrieving the media data of the at least one representation comprises retrieving at least one of a segment of the representation using an HTTP GET request or a partial segment of the representation using an HTTP partial GET request.
18. The method of claim 1, wherein the sample entry type code comprises 'resv.'
19. The method of claim 1, wherein the restricted scheme type code comprises at least one of 'stvi,' 'odvd,' 'povd,' 'fov,' 'erp,' or 'cmp.'
20. The method of claim 19, further comprising determining that the track of the at least one representation comprises video data including:
 - frame packed video data when the restricted scheme type code comprises 'stvi;'
 - omnidirectional video data when the restricted scheme type code comprises 'odvd;'
 - projected omnidirectional video data when the restricted scheme type code comprises 'povd;'

fisheye omnidirectional video data when the restricted scheme type code comprises 'fovdi;'

 equirectangular projected omnidirectional video data when the restricted scheme type code comprises 'erp;,' or

 cubemap projected omnidirectional video data when the restricted scheme type code comprises 'cmp.'

21. The method of claim 1, wherein extracting the data that specifies the one or more codecs further comprises, in response to determining that the restricted scheme type code comprises 'odvdi,' extracting a third element representing a type of omnidirectional video data for the track.

22. The method of claim 21, further comprising determining that the track of the at least one representation comprises video data including:

 projected omnidirectional video data when the third element comprises a value of 'povdi;'
or

 fisheye omnidirectional video data when the restricted scheme type code comprises 'fovdi.'

23. The method of claim 21, wherein extracting the data that specifies the one or more codecs further comprises, in response to determining that the third element comprises a value of 'povdi,' extracting a fourth element that indicates a projection type.

24. The method of claim 23, further comprising, in response to determining that the video data of the track comprises projected omnidirectional video data, determining that the projected omnidirectional video data comprises:

 an equirectangular projection when the fourth element has a value of 'erp;,' or
 a cubemap projection when the fourth element has a value of 'cmp.'

25. The method of claim 1, wherein extracting the data that specifies the one or more codecs comprises extracting a plurality of elements following the first element and the second element.

26. The method of claim 1, wherein extracting the data that specifies the one or more codecs comprises extracting values for one or more @codecs parameters, wherein the values for the @codecs parameters comprise respective period delimited lists of elements.

27. The method of claim 1, further comprising extracting data that specifies that the track includes video data with display orientation changes, wherein the track is included in a media file of the at least one representation, and wherein retrieving the media data of the at least one representation comprises retrieving media data of the track based on the data that specifies that the track includes the video data with the display orientation changes.
28. The method of claim 27, wherein the data that specifies that the track includes the video data with the display orientation changes comprises a 'vdoc' value for a restricted scheme type for the track.
29. The method of claim 27, wherein extracting comprises extracting the value for the restricted scheme type from a RestrictedSchemeInfoBox for the track.
30. The method of claim 27, further comprising, in response to determining that the track includes the video data with the display orientation changes, extracting data indicating whether the display orientation changes include either or both of rotation or flipping from a SchemeInformationBox.
31. The method of claim 30, further comprising:
determining that the display orientation changes include both rotation and flipping when the SchemeInformationBox has a value of 0;
determining that the display orientation changes include rotation when the SchemeInformationBox has a value of 1; or
determining that the display orientation changes include flipping when the SchemeInformationBox has a value of 2.
32. A device for retrieving media data, the device comprising:
a memory configured to store media data; and
one or more processors implemented in circuitry and configured to:
retrieve a manifest file specifying data for at least one representation of a media presentation, wherein the manifest file includes data that specifies one or more codecs for the at least one representation, the at least one representation including one or more media files including media data, the manifest file being separate from the at least one

representation, the manifest file not including media data;

extract, from the manifest file, the data that specifies the one or more codecs, the data including a first element representing a sample entry type code of a track of the at least one representation, wherein the first element represents that the track includes video data stored using a restricted scheme, and a second element representing a restricted scheme type code for the restricted scheme of the track; and

retrieve the one or more media files including the the media data of the at least one representation based on the first element and the second element.

33. The device of claim 32, wherein to extract the data that specifies the one or more codecs, the one or more processors are configured to:

extract a first set of values for one or more @codecs parameters, wherein the values for the @codecs parameters comprise respective period-delimited lists of elements; and

extract a second set of values comprising additional information for the @codecs parameters from a different MIME type parameter.

34. The device of claim 33, wherein the different MIME type parameter comprises 'odvinfo,' and wherein extracting the second set of values comprises extracting at least one of 'povd,' 'fovvd,' 'erp,' or 'cmp,' and wherein the one or more processors are further configured to determine that the track of the at least one representation comprises video data including:

projected omnidirectional video data when the restricted scheme type code comprises 'povd;'

fish-eye omnidirectional video data when the restricted scheme type code comprises 'fovvd;'

equiangular projected omnidirectional video data when the restricted scheme type code comprises 'erp;' or

cubemap projected omnidirectional video data when the restricted scheme type code comprises 'cmp.'

35. The device of claim 33, wherein the additional information comprises information indicating a particular type of frame packed video data, wherein the different MIME type parameter comprises 'fpvinfo,' and wherein the particular type of frame packed video data

comprises elements corresponding to stereo_scheme and stereo_indication_type.

36. The device of claim 32, wherein the one or more processors are further configured to receive one or more values for a MIME type parameter 'hdrinfo' for the track, the values for the MIME type parameter including information for at least one of high dynamic range (HDR) or wide color gamut (WCG) video data for the track, wherein to retrieve the media data of the at least one representation, the one or more processors are configured to retrieve media data of the track based on the values for the MIME type parameter.

37. The device of claim 36, wherein the one or more values comprise one of a single value for the MIME type parameter 'hdrinfo' or a comma-separated list of values for the MIME type parameter 'hdrinfo,' each of the values in the comma-separated list comprising a single element or a dot-separated list of elements.

38. The device of claim 32, wherein the manifest file comprises a dynamic adaptive streaming over HTTP (DASH) media presentation description (MPD).

39. The device of claim 32, wherein the device comprises at least one of:
 an integrated circuit;
 a microprocessor; or
 a wireless communication device.

40. A device for retrieving media data, the device comprising:
 means for retrieving a manifest file specifying data for at least one representation of a media presentation, wherein the manifest file includes data that specifies one or more codecs for the at least one representation, the at least one representation including one or more media files including media data, the manifest file being separate from the at least one representation, the manifest file not including media data;

means for extracting, from the manifest file, the data that specifies the one or more codecs, including:

means for extracting a first element representing a sample entry type code of a track of the at least one representation, wherein the first element represents that the track includes video data stored using a restricted scheme; and

means for extracting a second element representing a restricted scheme type code for the restricted scheme of the track; and

means for retrieving the one or more media files including the media data of the at least one representation based on the first element and the second element.

41. A non-transitory computer-readable storage medium having stored thereon instructions that, when executed, cause a processor to:

retrieve a manifest file specifying data for at least one representation of a media presentation, wherein the manifest file includes data that specifies one or more codecs for the at least one representation, the at least one representation including one or more media files including media data, the manifest file being separate from the at least one representation, the manifest file not including media data;

extract, from the manifest file, the data that specifies the one or more codecs, including instructions that cause the processor to:

extract a first element representing a sample entry type code of a track of the at least one representation, wherein the first element represents that the track includes video data stored using a restricted scheme; and

extract a second element representing a restricted scheme type code for the restricted scheme of the track; and

retrieve the one or more media files including the media data of the at least one representation based on the first element and the second element.