SYSTEM AND METHOD FOR EFFICIENT SCALABLE STREAM ADAPTATION

FIELD OF THE INVENTION

[0001] The present invention relates generally to media encoding, decoding, storage and transmission. More particularly, the present invention relates to scalable media encoding, decoding, storage and transmission.

BACKGROUND OF THE INVENTION

[0002] Multimedia applications include local playback, streaming or on-demand, conversational and broadcast/multicast services. Technologies involved in multimedia applications include, for example, media coding, storage and transmission. Media types include speech, audio, image, video, graphics and time text. Different standards have been specified for different technologies.

[0003] 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 and ITU-T H.264 (also know as ISO/IEC MPEG-4 AVC). In addition, new video coding standards are also being developed. For example, the development of a scalable video coding (SVC) standard is currently underway. This standard will become the scalable extension to H.264/AVC. The development of China video coding standards is also currently underway.

[0004] Scalable video coding can provide scalable video bit streams. A portion of a scalable video bit stream can be extracted and decoded with a degraded playback visual quality. A scalable video bit stream contains a non-scalable base layer and one or more enhancement layers. An enhancement layer may enhance the temporal resolution (i.e. the frame rate), the spatial resolution, or simply the quality of the video content represented by the lower layer or part thereof. In some cases, data of an enhancement layer can be truncated after a certain location, even at arbitrary positions. Each truncation position may include some additional data representing increasingly enhanced visual quality. Such scalability is referred to as fine-grained

(granularity) scalability (FGS). In contrast to FGS, the scalability provided by a quality enhancement layer that does not provide fined-grained scalability is referred to as coarse-grained scalability (CGS).

[0005] The scalable layer structure in the current draft SVC standard is characterized by three variables, referred to as temporal_level, dependency_id and quality_level, that are signaled in the bit stream or can be derived according to the specification. temporal_level is used to indicate the temporal scalability or frame rate. A layer comprising pictures of a smaller temporal_level value has a smaller frame rate than a layer comprising pictures of a larger temporal_level. dependency_id is used to indicate the inter-layer coding dependency hierarchy. At any temporal location, a picture of a smaller dependency_id value may be used for inter-layer prediction for coding of a picture with a larger dependency_id value. quality_level is used to indicate FGS layer hierarchy. At any temporal location and with identical dependency_id value, an FGS picture with quality_level value equal to QL uses the FGS picture or base quality picture (i.e., the non-FGS picture when QL-1 = 0) with quality_level value equal to QL-1 for inter-layer prediction.

[0006] Figure 1 shows a temporal segment of an exemplary scalable video stream with the displayed values of the three variables discussed above. It should be noted that the time values are relative, i.e. time = 0 does not necessarily mean the time of the first picture in display order in the bit stream. A typical prediction reference relationship of the example is shown in Figure 2, where solid arrows indicate the inter prediction reference relationship in the horizontal direction, and dashed block arrows indicate the inter-layer prediction reference relationship. The pointed-to instance uses the instance in the other direction for prediction reference.

[0007] As discussed herein, a layer is defined as the set of pictures having identical values of temporal_level, dependency_id and quality_level, respectively. To decode and playback an enhancement layer, typically the lower layers including the base layer should also be available, because the lower layers may be directly or indirectly used for inter-layer prediction in the coding of the enhancement layer. For example, in Figures 1 and 2, the pictures with (t, T, D, Q) equal to (0, 0, 0, 0) and (8, 0, 0, 0) belong to the base layer, which can be decoded independently of any enhancement

layers. The picture with (t, T, D, Q) equal to (4, 1, 0, 0) belongs to an enhancement layer that doubles the frame rate of the base layer; the decoding of this layer needs the presence of the base layer pictures. The pictures with (t, T, D, Q) equal to (0, 0, 0, 1) and (8, 0, 0, 1) belong to an enhancement layer that enhances the quality and bit rate of the base layer in the FGS manner; the decoding of this layer also needs the presence of the base layer pictures.

[0008] In the H.264/AVC standard, instantaneous decoding refresh (IDR) picture is defined, as follows. A coded picture in which all slices are I or SI slices cause the decoding process to mark all reference pictures as "unused for reference" immediately after decoding the IDR picture. After the decoding of an IDR picture, all following coded pictures in decoding order can be decoded without inter prediction from any picture decoded prior to the IDR picture. The first picture of each coded video sequence is an IDR picture.

[0009] The concept of an IDR picture is also used in the current draft SVC standard, wherein the definition is applicable to pictures with identical values of dependency_id and quality_level, respectively. In other words, an IDR picture is a coded picture in which the decoding of the IDR picture and all the following coded pictures in decoding order in the same layer (i.e. with the same values of dependency_id and quality_level, respectively, as the IDR picture) can be performed without inter prediction from any picture prior to the IDR picture in decoding order in the same layer. An IDR picture causes the decoding process to mark all reference pictures in the same layer as "unused for reference" immediately after decoding the IDR picture. It should be noted that as used herein for the context of the current draft SVC standard, the term "in the same layer" means that the decoded pictures are with the same values of dependency_id and quality_level, respectively, as the IDR picture. Either all pictures with an identical value of picture order count (i.e. at the same temporal location) but different values of dependency_id or quality_level, are coded as IDR pictures, or no picture for a specific value of picture order count is coded as IDR picture. In other words, either all pictures in an access unit (including all of the pictures with an identical value of picture order count) are IDR pictures or no picture in an access unit is an IDR picture.

[0010] Available media file format standards include ISO file format (ISO/IEC 14496-12), MPEG-4 file format (ISO/IEC 14496-14), AVC file format (ISO/IEC 14496-15) and 3GPP file format (3GPP TS 26.244). The SVC file format is currently under development by ISO/IEC MPEG and can be found at MPEG N7477, "VM Study Text for Scalable Video Coding (SVC) File Format", 73rd ISO/IEC MPEG meeting, Poznan, Poland, July 2005, incorporated by reference herein in its entirety.

[0011] One advantage of scalable coding compared to single-layer coding is that, with scalable coding, a single stream can meet different requirements of quality, bit rate, display size etc, while with single-layer coding, multiple streams must be used. Using multiple streams costs more store space and, in simulcast, more transmission bandwidth. In streaming applications, stream adaptation is needed when the capabilities of the transmission network or recipient(s) change compared to their earlier states, e.g. a change of transmission bandwidth. Gateways and other media-aware network elements (MANEs) could also perform stream adaptation. When a scalably coded file is played "locally" (i.e., the file resides in the same device as the decoder or resides in memory connected with a fast link to the decoding device), stream adaptation may be needed if the decoder shares computational resources with some other processing. For example, if decoding is performed on a general-purpose processor running a multi-process operating system, the decoder may be able to use full computational power at one time and decode all the scalable layers. At another time, however, it may only have a subset of the processor's computational power in its use to decode a subset of the available scalable layers. The adapted stream may have a changed bit rate, frame rate, and/or video resolution. With single-layer coding, stream adaptation can be performed through stream switching or transcoding. With single scalable stream, stream adaptation can be performed through layer switching.

[0012] In scalable coding, high-to-low layer switching can be performed at any location. However, the case is different for low-to-high layer switching, since decoding of the switch-to picture in the high layer typically requires the presence in the same layer of some previous pictures in decoding order.

[0013] For the current draft SVC standard, low-to-high layer switching can be performed at an IDR access unit (including IDR pictures). However, relying on an

IDR access unit either causes reduced coding efficiency due to frequent coding of IDR access units or non-prompt stream adaptation. Both of these issues are closely related to the end user experience. It is also theoretically possible to utilize the SP/SI picture coding or gradual decoding refresh technique to enable low-to-high layer switching. However, these techniques were designed for single-layer coding. Therefore, these techniques are not currently workable for scalable coding. Furthermore, even after these techniques are extended for use in scalable coding, their application will either result in additional coding constraints (in the form of equivalent to lower coding efficiency) or implementation complexity.

[0014] There is therefore a need for supporting simple and efficient low-to-high layer switching in scalable video coding. Furthermore, there is also a need for enabling the signaling of simple and efficient low-to-high layer switching in the file format level such that no parsing and analysis of the video bit stream is required to find the places for low-to-high layer switching, as parsing and analysis of stream could require complex computations.

SUMMARY OF THE INVENTION

[0015] The present invention provides for a system and method for the signaling of low-to-high layer switching points in the file format level to enable efficient scalable stream switching in streaming servers and local file playback. The present invention also discloses a system and method for the signaling of low-to-high layer switching points in the video bit stream, e.g., to enable intelligent forwarding of scalability layers in media-aware network elements or computationally scalable decoding in stream recipients.

[0016] The present invention enables simple and efficient low-to-high layer switching for scalable stream adaptation. The present invention eliminates the need for a streaming server to have to parse and analyze video bit streams extensively, which would otherwise require a significant computational and implementation burden. The present invention may be used in virtually any scalable media application.

[0017] These and other advantages and features of the invention, together with the organization and manner of operation thereof, will become apparent from the following detailed description when taken in conjunction with the accompanying drawings, wherein like elements have like numerals throughout the several drawings described below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] Figure 1 shows a temporal segment of an exemplary scalable video stream with the displayed values of the three variables temporal_level, dependency_id and quality_level;

[0019] Figure 2 is a typical prediction reference relationship for the temporal segment depicted in Figure 1;

[0020] Figure 3 is an overview diagram of a system within which the present invention may be implemented;

[0021] Figure 4 is a perspective view of an electronic device that can incorporate the principles of the present invention;

[0022] Figure 5 is a schematic representation of the circuitry of the electronic device of Figure 4; and

[0023] Figure 6 is an illustration of a common multimedia data streaming system in which the scalable coding hierarchy of the invention can be applied.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0024] With reference to Figure 6, a typical multimedia streaming system is described, which is one system for applying the procedure of the present invention.

[0025] A multimedia data streaming system typically comprises one or more multimedia sources 100, such as a video camera and a microphone, or video image or computer graphic files stored in a memory carrier. Raw data obtained from the different multimedia sources 100 is combined into a multimedia file in an encoder 102, which can also be referred to as an editing unit. The raw data arriving from the one or more multimedia sources 100 is first captured using capturing means 104 included in the encoder 102, which capturing means can be typically implemented as

different interface cards, driver software, or application software controlling the function of a card. For example, video data may be captured using a video capture card and the associated software. The output of the capturing means 104 is typically either an uncompressed or slightly compressed data flow, for example uncompressed video frames of the YUV 4:2:0 format or motion-JPEG image format, when a video capture card is concerned.

[0026] An editor 106 links different media flows together to synchronize video and audio flows to be reproduced simultaneously as desired. The editor 106 may also edit each media flow, such as a video flow, by halving the frame rate or by reducing spatial resolution, for example. The separate, although synchronized, media flows are compressed in a compressor 108, where each media flow is separately compressed using a compressor suitable for the media flow. For example, video frames of the YUV 4:2:0 format may be compressed using the ITU-T recommendation H.263 or H.264. The separate, synchronized and compressed media flows are typically interleaved in a multiplexer 110, the output obtained from the encoder 102 being a single, uniform bit flow that comprises data of a plural number of media flows and that may be referred to as a multimedia file. It is to be noted that the forming of a multimedia file does not necessarily require the multiplexing of a plural number of media flows into a single file, but the streaming server may interleave the media flows just before transmitting them.

[0027] The multimedia files are transferred to a streaming server 112, which is thus capable of carrying out the streaming either as real-time streaming or in the form of progressive downloading. In progressive downloading the multimedia files are first stored in the memory of the server 112 from where they may be retrieved for transmission as need arises. In real-time streaming the editor 102 transmits a continuous media flow of multimedia files to the streaming server 112, and the server 112 forwards the flow directly to a client 114. As a further option, real-time streaming may also be carried out such that the multimedia files are stored in a storage that is accessible from the server 112, from where real-time streaming can be driven and a continuous media flow of multimedia files is started as need arises. In such case, the editor 102 does not necessarily control the streaming by any means. The streaming

server 112 carries out traffic shaping of the multimedia data as regards the bandwidth available or the maximum decoding and playback rate of the client 114, the streaming server being able to adjust the bit rate of the media flow for example by leaving out B-frames from the transmission or by adjusting the number of the scalability layers. Further, the streaming server 112 may modify the header fields of a multiplexed media flow to reduce their size and encapsulate the multimedia data into data packets that are suitable for transmission in the telecommunications network employed. The client 114 may typically adjust, at least to some extent, the operation of the server 112 by using a suitable control protocol. The client 114 is capable of controlling the server 112 at least in such a way that a desired multimedia file can be selected for transmission to the client, in addition to which the client is typically capable of stopping and interrupting the transmission of a multimedia file.

[0028] When the client 114 is receiving a multimedia file, the file is first supplied to a demultiplexer 116, which separates the media flows comprised by the multimedia file. The separate, compressed media flows are then supplied to a decompressor 118 where each separate media flow is decompressed by a decompressor suitable for each particular media flow. The decompressed and reconstructed media flows are supplied to a playback unit 120 where the media flows are rendered at a correct pace according to their synchronization data and supplied to presentation means 124. The actual presentation means 124 may comprise for example a computer or mobile station display, and loudspeaker means. The client 114 also typically comprises a control unit 122 that the end user can typically control through a user interface and that controls both the operation of the server, through the above-described control protocol, and the operation of the playback unit 120, on the basis of the instructions given by the end user.

[0029] It is to be noted that the transfer of multimedia files from the streaming server 112 to the client 114 takes place through a telecommunications network, the transfer path typically comprising a plural number of telecommunications network elements. It is therefore possible that there is at least some network element that can carry out traffic shaping of multimedia data with regard to the available bandwidth or

the maximum decoding and playback rate of the client 114 at least partly in the same way as described above in connection with the streaming server.

[0030] The present invention provides for a system and method for the signaling of low-to-high layer switching points in the file format level to enable efficient scalable stream switching in streaming servers and local file playback. The present invention also discloses a system and method for the signaling of low-to-high layer switching points in the video bit stream, e.g., to enable intelligent forwarding of scalability layers in media-aware network elements or computationally scalable decoding in stream recipients.

[0031] A number of exemplary implementations of the present invention are described herein. It should be noted, however, that other variations of these implementations are also possible by, for example, combining subsets of the alternatives embodiments discussed herein.

[0032] A first implementation for signaling low-to-high layer switching points in the file format involves the use of signal low-to-high layer switching points as one type of scalability information. In this embodiment, a new box is defined for signaling the low-to-high layer switching points for each layer. The box is contained in the data structure containing the scalability information of each scalable layer, e.g. the ScalableLayerEntry() structure defined in the draft standard text of the SVC File Format. The box is defined as follows:

Box Types: 'lswp'
Container: ScalableLayerEntry
Mandatory: No
Quantity: Zero or One

[0033] This box provides a compact marking of the low-to-high layer switching points for an individual layer within the stream. The table is arranged strictly in increasing order of sample number. If the layer switching point box is not present for a layer, then for that layer, there are no additional low-to-high layer switching points other than the sync samples signaled in the sync sample box or shadow sync samples signaled in the shadow sync sample box.

```
class LayerSwitchingPointBox extends Box('lswp'){
        unsigned int(32) entry_count;
        int i,j;
        for (i=0; i < entry_count; i++) {
                unsigned int(32) sample_number;
                unsigned int(8) entry_count2;
                for (j=0; j < entry_count2; j++)
                        signed int(16) delta_layer_id;
        }
}
```

[0034] In this example, "entry_count" gives the number of entries in the following table. "sample_number" gives the sample number the sample whose information of low-to-high layer switching points for the current layer are signaled in the following table. "entry_count2" gives the number of entries in the following table.

[0035] "delta_layer_id" specifies the higher layer to which the stream can be switched from the current layer, or the lower layer from which the stream can be switched to the current layer, at the current sample with sample number equal to sample_number. A value equal to or greater than zero indicates that the stream can switch from the current layer to the higher layer with the layer identifier equal to (layerId + delta_layer_id + 1) at the current sample. A value smaller than zero indicates that the stream can switch from the lower layer with the layer identifier equal to (layerId + delta_layer_id) to the current layer at the current sample. The layerId for each layer is found from the data structure containing the scalability information of each scalable layer.

[0036] A second implementation for signaling low-to-high layer switching points in the file format involves the signaling of low-to-high layer switching points of all layers in a dedicated box. The box is defined as follows.

Box Types: 'lswp'
Container: Sample Table Box ('stbl')
Mandatory: No
Quantity: Zero or One

[0037] This box provides a compact marking of the low-to-high layer switching points for all the layers within the stream. The table is arranged strictly in increasing order of sample number. If the layer switching point box is not present, there are no additional low-to-high layer switching points other than the sync samples signaled in the sync sample box or shadow sync samples signaled in the shadow sync sample box.

```
class LayerSwitchingPointBox extends Box('lswp'){
        unsigned int(32) entry_count;
        int i,j,k;
        for (i=0; i < entry_count; i++) {
                unsigned int(32) sample_number;
                unsigned int(8) entry_count2;
                for (j=0; j < entry_count2; j++){
                        unsigned int(8) layer_id;
                        unsigned int(8) entry_count3;
                        for (k=0; k < entry_count3; k++)
                                signed int(16) delta_layer_id;
                }
        }
}
```

[0038] In this embodiment, "entry_count" gives the number of entries in the following table. "sample_number" gives the sample number the sample for which the information of low-to-high layer switching points is signaled in the following table. "entry_count2" gives the number of entries in the following table. "layer_id" gives the layer identifier of the layer for which the information of low-to-high layer switching points is signaled in the following table. "entry_count3" gives the number of entries in the following table.

[0039] "delta_layer_id" specifies the higher layer to which the stream can be switched from the current layer, or the lower layer from which the stream can be switched to the current layer, at the current sample with sample number equal to sample_number. A value not less than zero indicates that the stream can switch from

the current layer to the higher layer with the layer identifier equal to (layer_id + delta_layer_id + 1) at the current sample. A value less than zero indicates that the stream can switch from the lower layer to the current layer with the layer identifier equal to (layer_id + delta_layer_id) at the current sample. In this particular embodiment of the invention, it is also possible to put the loop of sample_number inside the loop of layer_id, instead of the current way of putting the sample_number loop inside the layer_id loop.

[0040] A third implementation for signaling low-to-high layer switching points in the file format involves the use of signaling track switching points. This embodiment may also be used simultaneously with any of the other above-identified embodiments. This embodiment applies to the case where independently-decodable combinations of scalable layers are combined to form independent tracks, typically by the use of referencing. In the ISO base media file format, the media data is represented as a set of network-independent tracks, whereas hint tracks contain network-dependent transport encapsulation information and associate media samples or portions thereof to transport packets (by references from the hint track to the media tracks). An extractor track contains references to a media track such that a subset of the coded pictures for an SVC access unit forms a sample in an extractor track. Therefore, each extractor track should represent a different subset of scalable layers, and each extractor track should be decidable without decoding other extractor tracks for the same media track.

[0041] In this particular implementation, the switching points between hint tracks or extractor tracks can be identified. Tracks that contain or refer to a sufficient subset of scalable layers enabling switching to at least one of the indicated switching point samples in this track are linked to this track by a track reference of type 'tswf' in the track reference container box. Tracks for which this track contains a sufficient subset of the scalable layers enabling switching to at least one of the indicated switching point samples in those tracks are linked to this track by a track reference of type 'tswt' in the track reference container box.

Box Type: 'tswd'

Container: Sample Table Box 'stbl'

Mandatory: No

Quantity: Zero or exactly one.

[0042] This box provides a compact marking of the low-to-high layer switching points between hint tracks or extractor tracks. The table is arranged strictly in increasing order of sample number. If the track switching point box is not present, then there are no additional low-to-high layer switching points other than the sync samples signaled in the sync sample box or shadow sync samples signaled in the shadow sync sample box.

```
class TrackSwitchingPointBox extends Box('tswp'){
        unsigned int(32) entry_count;
        int i,j;
        for (i=0; i < entry_count; i++) {
                unsigned int(32) sample_number;
                unsigned int(8) entry_count2;
                for (j=0; j < entry_count2; j++)
                        unsigned int(32) src_track_id;
                unsigned int(8) entry_count3;
                for (j=0; j < entry_count3; j++)
                        unsigned int(32) dest_track_id;
        }
}
```

[0043] In this scenario, "entry_count" gives the number of entries in the following table. "sample_number" gives the sample number of the switch-point sample. "entry_count2" gives the number of entries in the following table. "src_track_id" indicates the track identifier of the track containing the lower layer from which the stream can be switched to the layer contained in the current track. "entry_count3" gives the number of entries in the following table. "dest_track_id" indicates the track identifier of the track containing the higher layer to which the stream can be switched from the layer contained in the current track.

[0044] It should be noted that alternative syntax definitions are also possible. For example, the entry_count3 and the related loop for values of dest_track_id can be

omitted. Moreover, no loop for values of src_track_id are necessary if only one track is indicated in the 'tswf' track reference. It is also possible to use the TrackSwitchingPointBox syntax independently without the track references 'tswf' and/or 'tswt'. Furthermore, it is possible to reuse the information signaled in the media track using the first or the second alternative to obtain the low-to-high layer switching points in hint or extractor tracks, wherein the layer identifier is signaled for each hint or extractor track, e.g. in the track header or the sample entry.

[0045] The following is a discussion of alternative methods for the signaling of low-to-high layer switching points in the bit stream. A first such alternative involves the use of a new Network Abstraction Layer (NAL) unit type to indicate that a picture containing coded slices of that NAL unit type is an enhancement-layer instantaneous decoding refresh (EIDR) picture. An EIDR picture is a coded picture in which the decoding of the EIDR picture and all of the following coded pictures in decoding order in the same layer (i.e. with the same values of temporal_level, dependency_id and quality_level, respectively, as the EIDR picture) can be performed without inter prediction from any picture prior to the EIDR picture in decoding order in the same layer. An EIDR picture causes the decoding process to mark all reference pictures in the same layer as "unused for reference" immediately after decoding the EIDR picture. An EIDR picture will not have temporal_level, dependency_id and quality_level all equal to zero. It should be noted that, as used herein with the context of this alternative, "in the same layer" means that the decoded pictures are with the same values of temporal_level, dependency_id and quality_level, respectively, as the EIDR picture. It should be noted that the difference in the concept of being in the same layer compared to the latest draft SVC standard is that temporal_level is taken into account herein. By taking the temporal_level into account, low-to-high layer switching from a layer with a lower frame rate to a layer with a higher frame rate is enabled. Furthermore, the coding efficiency of the EIDR picture is improved, since reference pictures with higher values of temporal_level are automatically removed from the reference picture lists, such that some reference picture list reordering commands are saved or the values of the reference picture indices gets smaller, which improves the coding efficiency of entropy coding.

[0046] Because a lower layer cannot be dependent on a higher layer by nature of scalable coding, an EIDR picture cannot be predicted from a picture in the same layer and higher layers, but it can still be predicted from pictures from lower layers, including temporally lower layers. This enables layer switching from the next lower layer to the layer containing the EIDR picture at the location of the EIDR picture. For example, the new NAL unit type can be defined as in the following Table 1. The value 22 is used as the new NAL unit type, and the meaning of NAL unit type 20 is changed from "Coded slice of a non-IDR picture in scalable extension" to "Coded slice of a non-IDR and non-EIDR picture in scalable extension".

Table 1

nal_unit_type	Content of NAL unit and RBSP syntax structure	C
0	Unspecified	
1	Coded slice of a non-IDR picture slice_layer_without_partitioning_rbsp( )	2, 3, 4
2	Coded slice data partition A slice_data_partition_a_layer_rbsp( )	2
3	Coded slice data partition B slice_data_partition_b_layer_rbsp( )	3
4	Coded slice data partition C slice_data_partition_c_layer_rbsp( )	4
5	Coded slice of an IDR picture slice_layer_without_partitioning_rbsp( )	2, 3
6	Supplemental enhancement information (SEI) sei_rbsp( )	5
7	Sequence parameter set seq_parameter_set_rbsp( )	0
8	Picture parameter set pic_parameter_set_rbsp( )	1
9	Access unit delimiter access_unit_delimiter_rbsp( )	6
10	End of sequence end_of_seq_rbsp( )	7
11	End of stream end_of_stream_rbsp( )	8
12	Filler data filler_data_rbsp( )	9
13	Sequence parameter set extension seq_parameter_set_extension_rbsp( )	10
14..18	Reserved	
19	Coded slice of an auxiliary coded picture without partitioning slice_layer_without_partitioning_rbsp( )	2, 3, 4
20	Coded slice of a non-IDR and non-EIDR picture in scalable extension slice_layer_in_scalable_extension_rbsp( )	2, 3, 4
21	Coded slice of an IDR picture in scalable extension slice_layer_in_scalable_extension_rbsp( )	2, 3
22	Coded slice of an EIDR picture in scalable extension slice_layer_in_scalable_extension_rbsp( )	2, 3, 4
23	Reserved	
24..31	Unspecified	

[0047] This design is applicable to scalable layers coded using the scalable extension where temporal_level, dependency_id and quality_level are signaled. The H.264/AVC standard itself can be used to provide temporal scalability, which can be

signaled using sub-sequence information SEI messages. However, in this case, temporal_level is not signaled and therefore is inferred as zero for all of the temporal scalable layers according to the current draft SVC standard. To enable efficient low-to-high layer switching between temporal layers and not to break the H.264/AVC stream compatibility, a new SEI message is used to indicate that a picture is an EIDR picture. However, by using a SEI message, the reference management functionality of EIDR picture is impossible, i.e. an EIDR picture indicated by an SEI message can be used for low-to-high layer switching, but it does not cause the decoding process to mark all of reference pictures in the same layer as "unused for reference". This is due to the fact that SEI messages are optional to standard-compliant decoders. The SEI message is specified as follows.

EIDR picture SEI message syntax

eidr_pic(payloadSize) {	C	Descriptor
temporal_layer_num	5	ue(v)
}		

[0048] EIDR picture SEI message semantics. The semantics of the above is as follows. An EIDR picture SEI message, if present, is associated with a coded picture for which the temporal_level, dependency_id and quality_level are not present in the bit stream and therefore are all inferred as equal to zero. The target picture is also associated with a sub-sequence information SEI message, wherein the sub_seq_layer_num is signaled for indicating the temporal scalability hierarchy. The presence of an EIDR picture SEI message indicates that the target picture is an EIDR picture, in which the decoding of the picture and all of the following coded pictures in decoding order in the same layer (i.e. with temporal_level, dependency_id and quality_level equal to 0 and with the same value of sub_seq_layer_num as the EIDR picture) can be decoded without inter prediction from any picture in the same layer prior to the EIDR picture in decoding order.

[0049] The "temporal_layer_num" has the same semantics as the syntax element sub_seq_layer_num in sub-sequence information SEI message. The value is identical

to the sub_seq_layer_num in the sub-sequence information SEI message associated with the same target picture. It is also possible to exclude the syntax element temporal_layer_num from the syntax of the SEI message. However, if this is done, the parsing of the sub-sequence information SEI message is needed to know to which temporal scalable layer the EIDR picture belongs.

[0050] A second alternative implementation for signaling low-to-high layer switching points in the bit stream involves changing the definition of the IDR picture and relaxing the constraint for coding IDR pictures. In this implementation, the IDR is changed as follows. If any of the temporal_level, dependency_id and quality_level is not equal to zero, then an IDR picture is a coded picture in which the decoding of the IDR picture and all the following coded pictures in decoding order in the same layer (i.e, with the same values of temporal_level, dependency_id and quality_level, respectively, as the IDR picture) can be performed without inter prediction from any picture prior to the IDR picture in decoding order in the same layer. Furthermore, such an IDR picture causes the decoding process to mark all reference pictures in the same layer as "unused for reference" immediately after decoding the IDR picture. Otherwise, if the temporal_level, dependency_id and quality_level are all equal to zero, an IDR picture is a coded picture in which the decoding of the IDR picture and all the following coded pictures in decoding order can be performed without inter prediction from any picture prior to the IDR picture in decoding order. Such an IDR picture causes the decoding process to mark all reference pictures as "unused for reference" immediately after decoding the IDR picture.

[0051] The constraint for coding of IDR pictures (i.e., either all pictures in an access unit are IDR pictures or no picture in an access unit is an IDR picture) is relaxed as follows. If a picture having temporal_level, dependency_id and quality_level all equal to zero is an IDR picture, then all other pictures in the same access unit are IDR pictures. Due to this relaxed constraint for the coding of IDR pictures, it is possible to have IDR pictures in enhancement layers in an access unit where the base layer picture is a non-IDR picture. As can be observed in this second alternative, when any of temporal_level, dependency_id and quality_level are not equal to zero, the IDR picture definition is the same as the EIDR picture definition in the first alternative

method for the signaling of low-to-high layer switching points in the bit stream discussed above. Therefore all of the benefits of the EIDR picture are valid. The coding of an IDR picture as defined herein enables layer switching from the next lower layer to the layer containing the IDR picture at the location of the IDR picture.

[0052] This particular design is applicable to scalable layers coded using the scalable extension where the temporal_level, dependency_id and quality_level are signaled. The same method as in alternative 1 can be employed to enable efficient low-to-high layer switching between temporal layers coding using the H.264/AVC standard itself.

[0053] A third alternative implementation for signaling low-to-high layer switching points in the bit stream involves relaxing the constraint for coding IDR pictures and using SEI message for temporal scalable layers. In this alternative, the definition of IDR picture remains unchanged, but the constraint for coding IDR pictures is relaxed as follows.

[0054] If a picture having temporal_level, dependency_id and quality_level all equal to zero is an IDR picture, then all other pictures in the same access unit are IDR pictures. This relaxation enables the coding of IDR pictures capable of low-to-high layer switching between different values of dependency_id or quality_level. However, because the IDR picture definition is not changed, i.e., the temporal_level is not included in specifying a scalable layer, the problem of low-to-high layer switching between temporal layers remain unsolved.

[0055] To solve the problem of low-to-high layer switching between temporal layers, whether they are coded using the scalable extension or the H.264/AVC standard itself, the EIDR picture SEI message similar to that described for the first alternative discussed above is used to indicate that the target picture associated with the SEI message is an EIDR picture that can be used for low-to-high layer switching. Nevertheless, an EIDR picture indicated by an SEI message does not have the reference management functionality because SEI messages are optional to standard-compliant decoders.

[0056] The syntax of the SEI is the same as in the first alternative for signaling low-to-high layer switching points in the bit stream. The semantics is also the same when

the target picture is coded using the H.264/AVC standard itself. When the target picture is coded using the scalable extension, the semantics is changed as follows.

[0057] The presence of an EIDR picture SEI message indicates that the target picture is an EIDR picture, in which the decoding of the picture and all the following coded pictures in decoding order in the same layer (i.e. with identical values of temporal_level, dependency_id and quality_level as the EIDR picture) can be decoded without inter prediction from any picture in the same layer prior to the EIDR picture in decoding order. "temporal_layer_num" indicates the temporal_level value of the EDIR picture. It is also possible to exclude the syntax element temporal_layer_num from the syntax of the SEI message when the EIDR picture is coded using the scalable extension.

[0058] A fourth alternative for the signaling of low-to-high layer switching points in the bit stream involves using a SEI message for all scalable layers. In this alternative, the EIDR picture SEI message is similar to the message defined in the third alternative above is used to indicate that the target picture associated with the SEI message is an EIDR picture that can be used for low-to-high layer switching. However, in this particular embodiment, the SEI message is also used to solve the problem of low-to-high layer switching between different values of dependency_id or quality_level.

[0059] Figure 3 shows a system 10 in which the present invention can be utilized, comprising multiple communication devices that can communicate through a network. The system 10 may comprise any combination of wired or wireless networks including, but not limited to, a mobile telephone network, a wireless Local Area Network (LAN), a Bluetooth personal area network, an Ethernet LAN, a token ring LAN, a wide area network, the Internet, etc. The system 10 may include both wired and wireless communication devices.

[0060] For exemplification, the system 10 shown in Figure 3 includes a mobile telephone network 11 and the Internet 28. Connectivity to the Internet 28 may include, but is not limited to, long range wireless connections, short range wireless connections, and various wired connections including, but not limited to, telephone lines, cable lines, power lines, and the like.

[0061] The exemplary communication devices of the system 10 may include, but are not limited to, a mobile telephone 12, a combination PDA and mobile telephone 14, a PDA 16, an integrated messaging device (IMD) 18, a desktop computer 20, and a notebook computer 22. The communication devices may be stationary or mobile as when carried by an individual who is moving. The communication devices may also be located in a mode of transportation including, but not limited to, an automobile, a truck, a taxi, a bus, a boat, an airplane, a bicycle, a motorcycle, etc. Some or all of the communication devices may send and receive calls and messages and communicate with service providers through a wireless connection 25 to a base station 24. The base station 24 may be connected to a network server 26 that allows communication between the mobile telephone network 11 and the Internet 28. The system 10 may include additional communication devices and communication devices of different types.

[0062] The communication devices may communicate using various transmission technologies including, but not limited to, Code Division Multiple Access (CDMA), Global System for Mobile Communications (GSM), Universal Mobile Telecommunications System (UMTS), Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), Transmission Control Protocol/Internet Protocol (TCP/IP), Short Messaging Service (SMS), Multimedia Messaging Service (MMS), e-mail, Instant Messaging Service (IMS), Bluetooth, IEEE 802.11, etc. A communication device may communicate using various media including, but not limited to, radio, infrared, laser, cable connection, and the like.

[0063] Figures 4 and 5 show one representative mobile telephone 12 within which the present invention may be implemented. It should be understood, however, that the present invention is not intended to be limited to one particular type of mobile telephone 12 or other electronic device. The mobile telephone 12 of Figures 4 and 5 includes a housing 30, a display 32 in the form of a liquid crystal display, a keypad 34, a microphone 36, an ear-piece 38, a battery 40, an infrared port 42, an antenna 44, a smart card 46 in the form of a UICC according to one embodiment of the invention, a card reader 48, radio interface circuitry 52, codec circuitry 54, a controller 56 and a

memory 58. Individual circuits and elements are all of a type well known in the art, for example in the Nokia range of mobile telephones.

[0064] The present invention is described in the general context of method steps, which may be implemented in one embodiment by a program product including computer-executable instructions, such as program code, executed by computers in networked environments.

[0065] Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. Computer-executable instructions, associated data structures, and program modules represent examples of program code for executing steps of the methods disclosed herein. The particular sequence of such executable instructions or associated data structures represents examples of corresponding acts for implementing the functions described in such steps.

[0066] Software and web implementations of the present invention could be accomplished with standard programming techniques with rule-based logic and other logic to accomplish the various database searching steps, correlation steps, comparison steps and decision steps. It should also be noted that the words "component" and "module" as used herein, and in the claims, is intended to encompass implementations using one or more lines of software code, and/or hardware implementations, and/or equipment for receiving manual inputs.

[0067] The foregoing description of embodiments of the present invention have been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the present invention to the precise form disclosed, and modifications and variations are possible in light of the above teachings or may be acquired from practice of the present invention. The embodiments were chosen and described in order to explain the principles of the present invention and its practical application to enable one skilled in the art to utilize the present invention in various embodiments and with various modifications as are suited to the particular use contemplated.

WHAT IS CLAIMED IS:

1. A method of enabling low-to-high scalable layer switching in a video bit stream comprising a base layer and at least one enhancement layer, comprising:

encoding a plurality of low-to-high layer switching points within the video bit stream, wherein the plurality of low-to-high layer switching points are not at the location of instantaneous decoding refresh (IDR) access units.

2. The method of claim 1, further comprising using the encoded switching points to switch from lower layers in the video bit stream to higher layers in the video bit stream.

3. The method of claim 1, wherein the switching points are encoded by defining a NAL unit type to indicate that a picture containing coded slices of the particular NAL unit type is an enhancement-layer instantaneous decoding refresh (EIDR) picture.

4. The method of claim 1, wherein the switching points are encoded by defining a supplemental enhancement information (SEI) message to indicate that a picture is an EIDR picture.

5. The method of claim 1, wherein the switching points are encoded by defining an IDR picture to be a coded picture in which the decoding of the IDR picture and all the following coded pictures in decoding order can be performed without inter prediction from any picture prior to the IDR picture in decoding order, if temporal_level, dependency_id and quality_level values are all equal to zero, and, if any temporal_level, dependency_id and quality_level values are not equal to zero, an IDR picture is defined to be a coded picture in which the decoding of the IDR picture and all the following coded pictures in decoding order in the same layer can be performed without inter prediction from any picture prior to the IDR picture in decoding order having the same values of temporal_level, dependency_id and quality_level, respectively, as the IDR picture in the same layer, and wherein two

pictures are considered to be in the same layer if the two pictures have the same values of temporal_level, dependency_id and quality_level.

6. The method of claim 5, wherein all reference pictures having the same values of temporal_level, dependency_id and quality_level, respectively, as the IDR picture after decoding the IDR picture are immediately marked as "unused for reference" if any of the temporal_level, dependency_id and quality_level is not equal to zero.

7. The method of claim 5, wherein all reference pictures after decoding the IDR picture are immediately marked as "unused for reference" if the temporal_level, dependency_id and quality_level are all equal to zero.

8. The method of claim 1, wherein the switching points are encoded by, if a picture having temporal_level, dependency_id and quality level are all equal to zero is an IDR picture, defining all other pictures in the same access unit as IDR pictures.

9. The method of claim 8, wherein the switching points are encoded by, defining a supplemental enhancement information (SEI) message to indicate that a picture is an EIDR picture.

10. The method of claim 1, wherein the switching points are encoded by using a SEI message for all scalable layers to indicate that a target picture associated with the SEI message is an EIDR picture that can be used for low-to-high layer switching.

11. A method of enabling low-to-high scalable layer switching in a video bit stream comprising a base layer and at least one enhancement layer, comprising:

encoding a plurality of low-to-high layer switching points within a file format level of the video bit stream, wherein the plurality of low-to-high layer switching points are not at the location of instantaneous decoding refresh (IDR) access units.

12. The method of claim 11, further comprising using the encoded switching points to switch from lower layers in the video bit stream to higher layers in the video bit stream.

13. The method of claim 11, wherein the switching points are encoded by including a box in a data structure that contains scalability information for each scalable layer.

14. The method of claim 11, wherein the switching points are encoded by including a box indicating the switching points of all layers in the video bit stream.

15. The method of claim 11, wherein the switching points are encoded by including a box indicating the low-to-high switching points that exist between hint tracks or extractor tracks.

16. The method of claim 15, wherein information signaled in a media track is reused in order to obtain the low-to-high layer switching points that exist between the hint tracks or extractor tracks.

17. A computer program product for enabling low-to-high scalable layer switching in a video bit stream comprising a base layer and at least one enhancement layer, comprising:

computer code for encoding a plurality of low-to-high layer switching points within the video bit stream, wherein the plurality of low-to-high layer switching points are not at the location of instantaneous decoding refresh (IDR) access units.

18. The computer program product of claim 17, further comprising computer code for using the encoded switching points to switch from lower layers in the video bit stream to higher layers in the video bit stream.

19. The computer program product of claim 17, wherein the switching points are encoded by defining a NAL unit type to indicate that a picture containing

coded slices of the particular NAL unit type is an enhancement-layer instantaneous decoding refresh (EIDR) picture.

20. The computer program product of claim 17, wherein the switching points are encoded by defining a supplemental enhancement information (SEI) message to indicate that a picture is an EIDR picture.

21. The computer program product of claim 17, wherein the switching points are encoded by defining an IDR picture to be a coded picture in which the decoding of the IDR picture and all the following coded pictures in decoding order can be performed without inter prediction from any picture prior to the IDR picture in decoding order, if temporal_level, dependency_id and quality_level values are all equal to zero, and, if any temporal_level, dependency_id and quality_level values are not equal to zero, an IDR picture is defined to be a coded picture in which the decoding of the IDR picture and all the following coded pictures in decoding order in the same layer can be performed without inter prediction from any picture prior to the IDR picture in decoding order having the same values of temporal_level, dependency_id and quality_level, respectively, as the IDR picture in the same layer, and wherein two pictures are considered to be in the same layer if the two pictures have the same values of temporal_level, dependency_id and quality_level.

22. The computer program product of claim 21, wherein all reference pictures having the same values of temporal_level, dependency_id and quality_level, respectively, as the IDR picture after decoding the IDR picture are immediately marked as "unused for reference" if any of the temporal_level, dependency_id and quality_level is not equal to zero.

23. The computer program product of claim 21, wherein all reference pictures after decoding the IDR picture are immediately marked as "unused for reference" if the temporal_level, dependency_id and quality_level are all equal to zero.

24. The computer program product of claim 17, wherein the switching points are encoded by, if a picture having temporal_level, dependency_id and quality

level are all equal to zero is an IDR picture, defining all other pictures in the same access unit as IDR pictures.

25. The computer program product of claim 24, wherein the switching points are encoded by defining a supplemental enhancement information (SEI) message to indicate that a picture is an EIDR picture.

26. The computer program product of claim 17, wherein the switching points are encoded by using a SEI message for all scalable layers to indicate that a target picture associated with the SEI message is an EIDR picture that can be used for low-to-high layer switching.

27. A computer program product for enabling low-to-high scalable layer switching in a video bit stream comprising a base layer and at least one enhancement layer, comprising:

computer code for encoding a plurality of low-to-high layer switching points within a file format level of the video bit stream, wherein the plurality of low-to-high layer switching points are not at the location of instantaneous decoding refresh (IDR) access units.

28. The computer program product of claim 27, wherein the switching points are encoded by including a box in a data structure that contains scalability information for each scalable layer.

29. The computer program product of claim 27, wherein the switching points are encoded by including a box indicating the switching points of all layers in the video bit stream.

The computer program product of claim 27, wherein the switching points are encoded by including a box indicating the low-to-high switching points that exist between hint tracks or extractor tracks.

30. The computer program product of claim 31, wherein information signaled in a media track is reused in order to obtain the low-to-high layer switching points that exist between the hint tracks or extractor tracks.

31. An electronic device, comprising:

a processor; and

a memory unit operatively connected to the processor and including a computer program product for enabling low-to-high scalable layer switching in a video bit stream comprising a base layer and at least one enhancement layer, comprising:

computer code for encoding a plurality of low-to-high layer switching points within the video bit stream, wherein the plurality of low-to-high layer switching points are not at the location of instantaneous decoding refresh (IDR) access units.

32. The electronic device of claim 33, wherein the memory unit further comprises computer code for using the encoded switching points to switch from lower layers in the video bit stream to higher layers in the video bit stream.

33. The electronic device of claim 33, wherein the switching points are encoded by defining a NAL unit type to indicate that a picture containing coded slices of the particular NAL unit type is an enhancement-layer instantaneous decoding refresh (EIDR) picture.

34. The electronic device of claim 33, wherein the switching points are encoded by defining a supplemental enhancement information (SEI) message to indicate that a picture is an EIDR picture.

35. The electronic device of claim 33, wherein the switching points are encoded by defining an IDR picture to be a coded picture in which the decoding of the IDR picture and all the following coded pictures in decoding order can be performed without inter prediction from any picture prior to the IDR picture in decoding order, if temporal_level, dependency_id and quality_level values are all equal to zero, and, if any temporal_level, dependency_id and quality_level values are not equal to zero, an IDR picture is defined to be a coded picture in which the decoding of the IDR picture and all the following coded pictures in decoding order in the same layer can be performed without inter prediction from any picture prior to the IDR picture in

decoding order having the same values of temporal_level, dependency_id and quality_level, respectively, as the IDR picture in the same layer, and wherein two pictures are considered to be in the same layer if the two pictures have the same values of temporal_level, dependency_id and quality_level.

36. The electronic device of claim 37, wherein all reference pictures having the same values of temporal_level, dependency_id and quality_level, respectively, as the IDR picture after decoding the IDR picture are immediately marked as "unused for reference" if any of the temporal_level, dependency_id and quality_level is not equal to zero.

37. The electronic device of claim 37, wherein all reference pictures after decoding the IDR picture are immediately marked as "unused for reference" if the temporal_level, dependency_id and quality_level are all equal to zero.

38. The electronic device of claim 33, wherein the switching points are encoded by, if a picture having temporal_level, dependency_id and quality level are all equal to zero is an IDR picture, defining all other pictures in the same access unit as IDR pictures.

39. The electronic device of claim 40, wherein the switching points are encoded by defining a supplemental enhancement information (SEI) message to indicate that a picture is an EIDR picture.

40. The electronic device of claim 33, wherein the switching points are encoded by using a SEI message for all scalable layers to indicate that a target picture associated with the SEI message is an EIDR picture that can be used for low-to-high layer switching.

41. An electronic device, comprising:

a processor; and

a memory unit operatively connected to the processor and including computer program product for enabling low-to-high scalable layer switching in a

video bit stream comprising a base layer and at least one enhancement layer, comprising:

computer code for encoding a plurality of low-to-high layer switching points within a file format level of the video bit stream, wherein the plurality of low-to-high layer switching points are not at the location of instantaneous decoding refresh (IDR) access units.

42. The electronic device of claim 43, wherein the memory unit further comprises computer code for using the encoded switching points to switch from lower layers in the video bit stream to higher layers in the video bit stream.

43. The electronic device of claim 43, wherein the switching points are encoded by including a box in a data structure that contains scalability information for each scalable layer.

44. The electronic device of claim 43, wherein the switching points are encoded by including a box indicating the switching points of all layers in the video bit stream.

45. The electronic device of claim 43, wherein the switching points are encoded by including a box indicating the low-to-high switching points that exist between hint tracks or extractor tracks.

46. The electronic device of claim 47, wherein information signaled in a media track is reused in order to obtain the low-to-high layer switching points that exist between the hint tracks or extractor tracks.

47. An element of a streaming system, comprising:

a first component for receiving the video bit stream comprising at least two scalable layers;

a second component for identifying a low-to-high layer switching point within the video bit stream, wherein the low-to-high layer switching point is not at the location of any instantaneous decoding refresh (IDR) access units; and

a third component for switching from the transmission of a subset of scalable layers to a greater subset of scalable layers at the low-to-high layer switching point.

48. An element of a streaming system, comprising:

a first component for receiving the video bit stream stream comprising at least two scalable layers;

a second component for identifying a low-to-high layer switching point within a file format level of the video bit stream, wherein the low-to-high layer switching point is not at the location of instantaneous decoding refresh (IDR) access units; and

a third component for switching from the transmission of a subset of scalable layers to a greater subset of scalable layers at the low-to-high layer switching point.