

ENCODING TECHNIQUES FOR OPTIMIZING DISTORTION AND BITRATE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the priority benefit of the United States Provisional Patent Application titled "ENCODING TECHNIQUE FOR OPTIMIZING DISTORTION AND BITRATE", filed on July 18, 2017 and having Serial Number 62/534,170, and this application claims the priority benefit of the United States Patent Application titled "ENCODING TECHNIQUE FOR OPTIMIZING DISTORTION AND BITRATE", filed on July 12, 2018 and having Serial Number 16/034,303. The subject matter of these related applications is hereby incorporated herein by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

[0002] Embodiments of the present invention relate generally to video encoding and, more specifically, to encoding techniques for optimizing distortion and bitrate.

Description of the Related Art

[0003] A video streaming service provides access to a library of media titles that can be played on a range of different endpoint devices. Each endpoint device may connect to the video streaming service under different connection conditions, including available bandwidth and latency, among others. In addition, each different device may include different hardware for outputting the video content to the end user. For example, a given endpoint device could include a display screen having a particular screen size and a particular screen resolution.

[0004] Typically, an endpoint device that connects to a video streaming service executes an endpoint application that determines, for a given media title in the video content library, an appropriate version of the media title to stream to the endpoint device. Each different version of a given media title is usually encoded using a different bitrate, and the different versions of the media title have resolutions, scaling factors, and/or other parameters typically associated with video content that differ from one another. During playback of the media title on the endpoint device, the endpoint application selects the appropriate version of the media title to stream to the endpoint device based on factors such as network conditions, the quality of the network connection, and the hardware specifications of the endpoint device.

[0005] As noted above, to prepare a media title for streaming in the manner described above, the media title is encoded using multiple different bitrates. In doing so, an encoding application performs individual, "monolithic" encodes of the entire media title, using a different set of encoding parameters for each encode. Each
5 different encode may be associated with a different quality metric that objectively indicates the level of distortion introduced into that encoded version of the media title via the encoding process. The quality metric associated with a given encode typically depends on the encoding parameters used to generate that encode. For example, an encode generated with a high bitrate compared to another encode could
10 have a higher quality metric compared to that other encode.

[0006] Encoding a media title with different encoding parameters typically requires different computational resources and different storage resources. For example, generating an encode with a high bitrate and high quality metric generally consumes more computational / storage resources than generating an encode with a low bitrate
15 and low quality metric. A conventional encoding application may select a given set of encoding parameters for generating a single monolithic encode in order to meet a particular target quality metric for that encode.

[0007] However, one problem with this approach is that not all portions of a media title require the same encoding parameters to meet a given target quality metric, yet
20 conventional encoding applications use the same encoding parameters for the entire media title. Consequently, a conventionally-encoded media title may consume excessive computational and storage resources to meet the target quality metric, despite some portions of the media title not needing those resources to meet the same metric. This inefficiency needlessly wastes computational resources and
25 storage resources.

[0008] As the foregoing illustrates, what is needed in the art is a more efficient technique for encoding video sequences.

SUMMARY OF THE INVENTION

[0009] One embodiment of the present invention sets forth a computer-
30 implemented method, including generating a first set of encoded chunks for a source video sequence, generating a first set of data points based on the first set of encoded chunks, performing one or more convex hull operations across the first set of data points to compute a first subset of data points that are optimized across at least two

metrics, computing a first slope value between a first data point included in the first subset of data points and a second data point included in the first subset of data points, and determining, based on the first slope value, that a first encoded chunk associated with the first data point should be included in a final encoded version of the source video sequence.

[0010] At least one technological improvement of the disclosed techniques relative to prior art is that performing optimization operations at the granularity of the encoded chunks reduces encoding inefficiencies associated with conventional encoding techniques. As a result, the final encoded version of the source video sequence can be streamed to endpoint devices with an increased visual quality for a target bitrate. Conversely, the final encoded version of the source video sequence can be streamed to endpoint devices with a reduced bitrate for a target visual quality.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] So that the manner in which the above recited features of the present invention can be understood in detail, a more particular description of the invention, briefly summarized above, may be had by reference to embodiments, some of which are illustrated in the appended drawings. It is to be noted, however, that the appended drawings illustrate only typical embodiments of this invention and are therefore not to be considered limiting of its scope, for the invention may admit to other equally effective embodiments.

[0012] Figure 1A illustrates a cloud computing environment configured to implement one or more aspects of the present invention;

[0013] Figure 1B is a more detailed illustration of the encoding engines of Figure 1A, according to various embodiments of the present invention;

[0014] Figure 2 illustrates how the encoding engines of Figure 1B cut a video sequence into shot sequences, according to various embodiments of the present invention;

[0015] Figure 3 illustrates how the encoding engines of Figure 1B process the shot sequences of Figure 2 to generate a dataset, according to various embodiments of the present invention;

[0016] Figure 4 is a more detailed illustration of the processing pipeline of Figure 3, according to various embodiments of the present invention;

[0017] Figure 5A is a graph of bitrate versus quality that is generated based on the dataset of Figure 3, according to various embodiments of the present invention;

[0018] Figure 5B is a graph of convex hull data points that is generated based on the dataset of Figure 3, according to various embodiments of the present invention;

5 [0019] Figure 6 illustrates how the encoding engines of Figure 1B generate the convex hull data points of Figure 5B, according to various embodiments of the present invention;

[0020] Figure 7 illustrates how the encoding engines of Figure 1B generate different versions of the video sequence of Figure 2 using a plurality of convex hulls,
10 according to various embodiments of the present invention;

[0021] Figures 8A – 8D illustrate in greater detail how the encoding engines of Figure 1B assemble chunks of video content into an encoded video sequence, according to various embodiments of the present invention;

[0022] Figure 9 is a graph of convex hull data points generated for the encoded
15 video sequences shown in Figures 8A-8D, according to various embodiments of the present invention;

[0023] Figure 10 is a flow diagram of method steps for assembling chunks of video content into an encoded video sequence, according to various embodiments of the present invention;

20 [0024] Figure 11 is a flow diagram of method steps for processing a resampled shot sequence to generate a set of data points, according to various embodiments of the present invention; and

[0025] Figure 12 is a flow diagram of method steps for generating a set of encoded video sequences, according to various embodiments of the present invention.

25 DETAILED DESCRIPTION

[0026] In the following description, numerous specific details are set forth to provide a more thorough understanding of the present invention. However, it will be apparent to one of skill in the art that the present invention may be practiced without one or more of these specific details. In other instances, well-known features have
30 not been described in order to avoid obscuring the present invention.

[0027] As discussed above, conventional encoding techniques suffer from specific inefficiencies associated with performing “monolithic” encodes of video sequences. These inefficiencies arise because conventional encoding techniques encode all portions of a video sequence with the same encoding parameters to meet a given quality metric, despite the fact that some portions of the video sequence could be encoded with different encoding parameters and still meet the same quality metric.

[0028] To address this issue, embodiments of the present invention include an encoding engine configured to encode different shot sequences within a source video sequence with different encoding parameters that optimize bitrate for a given level of distortion. When encoding a shot sequence, the encoding engine resamples the shot sequence to a range of different resolutions and then encodes each resampled sequence using a range of quality parameters. The encoding engine then upsamples each encoded sequence to the original resolution of the source video sequence and computes a quality metric for the resultant upsampled sequences. Based on the upsampled sequences and corresponding quality metrics for each shot sequence, the encoding engine generates different encoded versions of the source video sequence. Each such version is a composite of multiple shot sequences encoded with potentially different encoding parameters.

[0029] An advantage of this approach is that portions of the source video sequence needing specific encoding parameters to meet a given quality metric are encoded with precisely those specific encoding parameters. Further, other portions of the source video sequence can be encoded with other appropriately chosen encoding parameters. Accordingly, encoded versions of the source video sequence are generated in a more efficient manner.

System Overview

[0030] Figure 1A illustrates a cloud-computing environment configured to implement one or more aspects of the present invention. As shown, a system 100 includes a host computer 110 coupled to a computer cloud 130. Host computer 110 includes a processor 112, input/output (I/O) devices 114, and a memory 116 coupled together.

[0031] Processor 112 may be any technically feasible form of processing device configured to process data and execute program code. Processor 112 could be, for example, a central processing unit (CPU), a graphics processing unit (GPU), an

application-specific integrated circuit (ASIC), a field-programmable gate array (FPGA), any technically feasible combination of such units, and so forth.

[0032] I/O devices 114 may include devices configured to receive input, including, for example, a keyboard, a mouse, and so forth. I/O devices 114 may also include devices configured to provide output, including, for example, a display device, a speaker, and so forth. I/O devices 114 may further include devices configured to both receive and provide input and output, respectively, including, for example, a touchscreen, a universal serial bus (USB) port, and so forth.

[0033] Memory 116 may include any technically feasible storage medium configured to store data and software applications. Memory 116 could be, for example, a hard disk, a random access memory (RAM) module, a read-only memory (ROM), and so forth. Memory 116 includes a host encoding engine 118 and a database 120.

[0034] Host encoding engine 118 is a software application that, when executed by processor 112, performs an encoding operation with media content stored within database 120 and/or an external storage resource. Host encoding engine 118 is configured to interoperate with various cloud encoding engines discussed in greater detail below.

[0035] Computer cloud 130 includes a plurality of cloud computers 140(0) through 140(N). Any cloud computer 140 may be a physically separate computing device or a virtualized instance of a computing device. Each cloud computer 140 includes a processor 142, I/O devices 144, and a memory 146, coupled together. A given processor 142 may be any technically feasible form of processing device configured to process data and execute program code, including a CPU, a GPU, an ASIC, an FPGA, any technically feasible combination of such units, and so forth. A given set of I/O devices 144 may include devices configured to receive input, including, for example, a keyboard, a mouse, and so forth, similar to I/O devices 114 discussed above. Each memory 146 is a storage medium configured to store data and software applications, including cloud encoding engine 148 and database 150.

[0036] Cloud encoding engines 148(0) through 148(N) are configured to interoperate with host encoding engine 118 in order to perform various portions of an encoding operation. In general, host encoding engine 118 coordinates the operation of cloud encoding engines 148(0) through 148(N), and may perform tasks such as

distributing processing tasks to those engines, collecting processed data from each engine, and so forth. Persons familiar with cloud computing will understand that cloud encoding engines 148(0) through 148(N) may operate substantially in parallel with one another. Accordingly, host encoding engine 118 may perform complex encoding tasks efficiently by configuring cloud encoding engines 148 to perform separate tasks simultaneously. As a general matter, host encoding engine 118 and cloud encoding engines 148 represent different modules within a distributed software entity, as described in greater detail below in conjunction with Figure 1B.

[0037] Figure 1B is a more detailed illustration of the encoding engines of Figure 1A, according to various embodiments of the present invention. As shown, an encoding engine 160 includes host encoding engine 118 and cloud encoding engines 148(0) through 148(N). As a general matter, encoding engine 160 constitutes a distributed software entity configured to perform one or more different encoding operations via execution of host encoding engine 118 and cloud encoding engines 148. In particular, encoding engine 160 processes a source video sequence 170 to generate a set of encoded video sequences 180. Source video sequence 170 is a media title that may be included in a content library associated with a video streaming service. Each encoded video sequence 180 is a different version of that media title encoded with different (and potentially varying) encoding parameters.

[0038] To perform the encoding operation, encoding engine 160 preprocesses source video sequence 170 to remove extraneous pixels and then cuts source video sequence 170 into a plurality of shot sequences. Each shot sequence includes frames captured continuously from a given camera or point of capture. This procedure is discussed in conjunction with Figure 2. Encoding engine 160 then resamples each shot sequence into one or more different resolutions, and processes all resampled sequences to generate a dataset. The resampling process is discussed in conjunction with Figure 3. Generation of the dataset based on resampled sequences is discussed in conjunction with Figure 4. Encoding engine 160 then generates, based on the dataset, a convex hull of data points that maximize bitrate for a given level of distortion, as discussed in conjunction with Figures 5A-5B and Figure 6. Based on all convex hull points across all shot sequences, encoding engine 160 generates the set of encoded video sequences 180. These encoded video sequences optimize distortion and bitrate, as discussed in conjunction with

Figures 7-9. The encoding operation discussed in conjunction with Figures 3-9 is also presented as a series of steps in conjunction with Figures 10-11.

Optimizing Distortion and Bitrate

[0039] Figure 2 illustrates how the encoding engines of Figure 1B cut a video sequence into shot sequences, according to various embodiments of the present invention. As mentioned above in conjunction with Figures 1-2, encoding engine 160 is configured to perform an encoding operation to generate different encoded versions of source video sequence 170, where each different version minimizes distortion for a given bitrate and/or optimizes distortion and bitrate. A first step in that encoding operation is illustrated in Figure 2. As shown, a shot analyzer 200 is configured to process source video sequence 170 to generate shot sequences 220(0) through 220(P). Shot analyzer 200 is a software module included within encoding engine 160. In the example shown, source video sequence 170 includes frames of video having a 4096 x 2048 resolution, meaning a 4096 pixel width by 2048 pixel height. The resolution of source video sequence 170 generally corresponds to the particular distribution format associated with the video sequence.

[0040] Shot analyzer 200 generates each shot sequence 220 to have the same resolution as source video sequence 170. However, each shot sequence 220 includes a different sequence of video frames that corresponds to a different "shot." In the context of this disclosure, a "shot" may be a sequence of frames captured continuously from a single camera or virtual representation of a camera (e.g., in the case of computer animated video sequences). In generating shot sequences 220, shot analyzer 200 may also remove extraneous pixels from source video sequence 170. For example, shot analyzer 200 may remove pixels included in black bars along border sections of source video sequence 170.

[0041] Shot analyzer 200 may determine which frames of source video sequence 170 correspond to each different shot using many different techniques. For example, shot analyzer 200 could identify a set of sequential frames having a continuous distribution of pixel values that do not change significantly across a subset of two or more sequential frames. Alternatively, shot analyzer 200 could compare features present in each frame and identify sequential frames having similar features. Persons skilled in the art will understand that many techniques for parsing a source video sequence into separate shot sequence exist. Upon parsing source video

sequence 170 in this manner, encoding engine 160 processes each shot sequence 220 to generate a different dataset, as described below in conjunction with Figure 3.

[0042] Figure 3 illustrates how the encoding engines of Figure 1B process the shot sequences of Figure 2 to generate a dataset, according to various embodiments of the present invention. As shown, a resampler 300 processes a shot sequence 220 to generate resampled sequences 320(0) through 320(M). Each resampled sequence 320 has a different resolution, as is shown. Resampled sequence 320(0) has a resolution of 4096 x 2048, resampled sequence 320(1) has a resolution of 2048 x 1024, and resampled sequence 220(M) has a resolution of 256 x 144. The set of resampled sequences 320 corresponds to a resolution ladder 330 that is associated with the shot sequence 220.

[0043] Resampler 300 may generate resolution ladder 330 to include any distribution of resolutions. In practice, however, resampler 300 first generates resampled sequence 320(0) to have the same resolution as shot sequence 220 (or source video sequence 170), and then generates each subsequent resampled sequence 320(1) onwards to have a resolution that is a constant fraction of the previous resolution. In practice, the ratio between the resolution of a given resampled sequence 320(H) and a previous resampled sequence 320(H-1) is approximately 1.5.

[0044] However, in various embodiments a denser resolution ladder may be used, i.e. with a ratio between the resolution of a given resampled sequence 320(H) and a previous resampled sequence 320(H-1) of less than 1.5, such as 1.414 or 1.26, or a coarser resolution ladder, i.e. with a ratio between the resolution of a given resampled sequence 320(H) and a previous resampled sequence 320(H-1) of more than 1.5, such as 2.0 or 3.0. The density of resolution ladder 330 can also depend on the characteristics of the video shot, such that it can span the desired quality levels uniformly. Additional constraints, such as the amount of CPU one wants to spend in encoding a certain sequence, can be used to decide the density of resolution ladders.

[0045] Upon generating resolution ladder 330, encoding engine 160 then executes a set of parallel processing pipelines 340 to process each different resampled sequence 320. Each processing pipeline 340 generates, based on the resampled sequence 320 input thereto, a collection of data points 350. Processing pipeline 340(0) generates data points 350(0), processing pipeline 350(1) generates data

points 350(1), and so forth for all processing pipelines 340. Encoding engine 160 then combines all such data points 350 to generate a data set 360. Because encoding engine 160 performs this processing for all shot sequences 220(0) through 220(P), encoding engine 160 generates P different datasets 360. An exemplary
5 processing pipeline 340 is described in greater detail below in conjunction with Figure 4, and data set 360 is described further in conjunction with Figures 5A-5B.

[0046] Figure 4 is a more detailed illustration of the processing pipeline of Figure 3, according to various embodiments of the present invention. As shown, processing pipeline 340 receives a resampled sequence 320 and generates, via a set of parallel
10 sub-pipelines 450(0) through 450(L), data points 350. Each sub-pipeline 450 includes an encoder 400, a decoder 410, an upsampler 420, and a metric analyzer 430. Sub-pipeline 450(0) includes encoder 400(0), decoder 410(0), upsampler 420(0), and metric analyzer 430(0), sub-pipeline 450(1) includes encoder 400(1), decoder 410(1), upsampler 420(1), and metric analyzer 430(1), and so forth for all
15 sub-pipelines 450. Encoders 400 and decoders 410 within each sub-pipeline 450 may implement any technically feasible encoding / decoding algorithm(s), including advanced video coding (AVC), high-efficiency video encoding (HEVC), or VP9, among others.

[0047] During execution of processing pipeline 340, each encoder 400(0) through
20 400(L) first encodes resampled sequence 320 with a different quantization parameter (QP). Encoder 400(0) encodes resampled sequence 320 with QP=0, encoder 400(1) encodes resampled sequence 320 with QP=1, and encoder 400(L) encodes resampled sequence 320 with QP=L. Generally, the number of encoders L corresponds to the number of available QPs for the given algorithm implemented by
25 encoders 400. In embodiments where encoders 400 implement AVC encoding algorithm using the x264 implementation, encoders 400 may perform the encoding operation described using different constant rate factors (CRFs) instead of QPs. In various embodiments, encoders 400 may vary any encoding parameter beyond QP or CRF.

[0048] Importantly, the encoded resampled sequences generated by encoders 400
30 may ultimately be included within encoded video sequence 180 shown in Figure 2B. In the context of this disclosure, these encoded, resampled sequences may be referred to herein as “chunks.” A “chunk” generally includes a sequence of video

frames encoded with a particular set of encoding parameters. In practice, each chunk is resampled with a particular resolution and then encoded with a given QP.

Also, each chunk is generally derived from a given shot sequence. However, persons skilled in the art will understand that a “chunk” in the context of video

5 encoding may represent a variety of different constructs, including a group of pictures (GOP), a sequence of frames, a plurality of sequences of frames, and so forth.

[0049] Once encoders 400 encode resampled sequences 320 with the different QPs in the manner described, each sub-pipeline 450 proceeds in relatively similar fashion. Decoders 410 receive the encoded sequences and then decode those

10 sequences. Accordingly, each video sequence output via upsamplers 420(0) through 420(L) has the same resolution. However, those video sequences may have different qualities by virtue of being encoded with different QPs.

[0050] In one embodiment, upsamplers 420 upsample the decoded sequences to target resolutions that may be relevant to the display characteristics of a class of

15 endpoint devices. For example, a certain video may be delivered in 3840 x 2160 resolution, yet be intended to be consumed by a large number of displays in 1920 x 1080 resolution. Another class of endpoint devices, for example laptop computers, is expected to display the same video in 1280 x 720 resolution. Yet another class of endpoint devices, for example, tablet or smartphone devices, is expected to display

20 the same video in 960 x 540 resolution. The decoded sequences can be upsampled to all these target resolutions in order to assess quality, when considering one of these different classes of endpoint devices, correspondingly.

[0051] Metric analyzers 330 analyze the upsampled sequences to generate an objective quality metric (QM) for each sequence. Metric analyzers 330 could

25 implement, for example, a video multimethod assessment fusion (VMAF) algorithm to generate a VMAF score for each upsampled sequence, among other possibilities.

Although a multitude of video quality metrics, such as VMAF scores, can be

calculated at different target resolutions, it should be clear that, when comparing qualities among encodes performed at different resolutions, one needs to use the

30 same target resolution for resampling, after decoding. In the following discussion, we consider one such resolution for upsampling and quality metric calculation, for example the common HD resolution of 1920 x 1080.

[0052] Each metric analyzer 330 then generates a different data point 440 that includes the resolution of resampled sequence 320, the QP implemented by the respective encoder 400, and the computed QM. Thus, for each different QP, processing pipeline 340 generates a separate data point, shown as data point 440(0) through 440(L). Importantly, each data point 440 corresponds to a particular resampled / encoded version of a given shot sequence 220. As described in greater detail below, encoding engine 160 selects resampled / encoded versions of each shot sequence 220 for inclusion into encoded video sequences 180 based on the associated data points 400. Processing pipeline 340 collects all such data points 440 into data points 350, as also shown in Figure 3.

[0053] Referring back now to Figure 3, encoding engine 160 generates a different set of data points 350(0) through 350(M) for each different resampled sequence 320(0) through 320(M), and then collects these data points 350 to data set 360. Accordingly, data set 360 includes $M * L$ data points, because encoder 160 generates a data point in data set 360 for each combination of the M different resampled sequences 320 and the L different QPs. One does not necessarily need to use the same number of QPs or the same QP values for each resolution, but instead use a fully customized number of QPs and QP values that is suitable for each shot. Encoding engine 160 then performs a processing operation discussed below in conjunction with Figures 5A-5B to identify the particular data points within data set 360 that minimize distortion and/or bitrate.

Convex Hull Analysis

[0054] Figure 5A is a graph of bitrate versus quality that is generated based on the dataset of Figure 3, according to various embodiments of the present invention. As shown, a graph 500 includes a bitrate axis 510 and a quality metric (QM) axis 520. Graph 500 also includes quality curves 502, 504, and 506 plotted against bitrate axis 510 and QM axis 520. Each curve shown corresponds to a different resolution encoding for a particular shot sequence 220 and therefore may be derived from a particular set of data points 350, where each data point 440 in a given set corresponds to a particular combination of resolution, QP, and QM. Encoding engine 160 generates the data points included in curves 502, 504, and 506 by converting the resolution of each data point 440 to a given bitrate. Encoding engine

160 could, for example, divide the total number of bits needed for the given resolution by the length of the associated shot sequence 320.

[0055] Encoding engine 160 is configured to reprocess dataset 160 plotted in Figure 5A to replace QM with a distortion metric. Encoding engine 160 may
5 compute a given distortion metric by inverting a QM value, subtracting the QM value from a constant value, or performing other known techniques for converting quality to distortion. Encoding engine 160 then generates a convex hull based on the converted values, as discussed below in conjunction with Figure 5B.

[0056] Figure 5B is a graph of convex hull data points that is generated based on
10 the dataset of Figure 3, according to various embodiments of the present invention. As shown, graph 550 includes bitrate axis 560 and distortion axis 570. Encoding engine 160 plots distortion curves 552, 554, and 556 against bitrate axis 560 and distortion axis 570. Then, encoding engine 160 computes convex hull points 580 by
15 identifying points across all curves that form a boundary where all points reside on one side of the boundary (in this case, the right side of the boundary) and also are such that connecting any two consecutive points on the convex hull with a straight line leaves all remaining points on the same side. In this manner, encoding engine 160
may generate convex hull points 580 for each shot sequence 220. Persons skilled in the art will understand that many techniques for generating convex hulls are well
20 known in the field of mathematics, and all such techniques may be implemented to generate convex hull 580. In one embodiment, encoding engine 160 applies machine learning techniques to estimate convex hull points 580 based on various parameters of the associated source video sequence 170. In this manner, some of the computations discussed thus far may be streamlined and/or avoided entirely.
25 Encoding engine 160 performs the processing described in conjunction with Figures 5A-5B via a sequence of processing stages discussed below in conjunction with Figure 6.

[0057] Figure 6 illustrates how the encoding engines of Figure 1B generate the convex hull data points of Figure 5B, according to various embodiments of the
30 present invention. As shown, a distortion converter 600 and convex hull analyzer 620 cooperatively process dataset 360 to generate convex hull points 580. In operation, distortion converter 600 receives data set 360 and then converts the QM

values included in that dataset to distortion values. Then, convex hull analyzer 620 computes the convex hull for the dataset 360 to generate convex hull points 580.

[0058] In this manner, encoding engine 160 computes convex hull points 580 for each shot sequence 320 based on the associated dataset 360. Thus, encoding engine 160 generates P sets of convex hull points 580 based on the P different shot sequences 320. Again, each set of convex hull points 580 includes data points that describe, for one shot sequence, the distortion and bitrate for a particular resampled, encoded version of the shot sequence. That version is resampled with a given resolution and encoded with a given QP. Encoding engine 160 collects all convex hulls 580 generated for all P shot sequences 320 and then performs additional processing to generate encoded video sequences 180, as described in greater detail below in conjunction with Figure 7.

Assembling Encoded Video Sequences via Trellis Iteration

[0059] Figure 7 illustrates how the encoding engines of Figure 1B generate different versions of the video sequence of Figure 2 using a plurality of convex hulls, according to various embodiments of the present invention. As shown, a trellis iterator 700 receives convex hull points 580(0) through 580(P) and then iteratively updates a sequence trellis 710 to generate sequence RD points 720. Trellis iterator 700 is a software module included within encoding engine 160. Sequence trellis 710 is a data structure that is described in greater detail below in conjunction with Figures 8A-8D. Sequence RD points 720 include bitrate-distortion (RD) points generated for different combinations of resampled, encoded sequences.

[0060] Each sequence RD point 720 corresponds to a different encoded video sequence 180. Each encoded video sequence 180 includes a different combination of the resampled, encoded shot sequences discussed above. A streaming application 730 is configured to stream encoded video sequences 180 to an endpoint device based on sequence RD points 720. Each encoded video sequence 180 minimizes distortion (on average) across all shot sequences in the video sequence for a given average bitrate associated with the video sequence, as also discussed in greater detail below in conjunction with Figure 9. Trellis iterator 700 generates these different sequences using a technique described in greater detail below.

[0061] Figures 8A-8D illustrate in greater detail how the encoding engines of Figure 1B assemble chunks of video content into an encoded video sequence,

according to various embodiments of the present invention. As shown in Figures 8A-8D, a sequence trellis 710 includes a shot axis 800 and a bitrate axis 810.

Sequence trellis 710 also includes columns of convex hull points 580, where each column corresponds to a particular shot sequence. For example, the zeroth column included in sequence trellis 710 corresponds to convex hull points 580(0). Hull points within any column are ranked according to ascending bitrate (and, by construction, descending distortion). Hull points are also guaranteed to have negative slopes that – in magnitude – are decreasing as a function of bitrate.

[0062] For convenience, convex hull points 580 are individually indexed according to the following system. For a given point, the first number is an index of the shot sequence, and the second number is an index into the bitrate ranking of those hull points. For example, convex hull point 00 corresponds to the zeroth shot sequence and the zeroth ranked bitrate (in this case the lowest bitrate). Similarly, convex hull point 43 corresponds to the fourth shot sequence and the third ranked bitrate (in this case the highest ranked bitrate).

[0063] Each convex hull point included within trellis 710 corresponds to a different resampled, encoded version of a shot sequence 220, as described. Encoding engine 160 generates encoded video sequences 180 shown in Figure 2B by combining these resampled, encoded versions of shot sequences 220. Encoding engine 160 implements sequence trellis 710 to iteratively perform this combining technique.

[0064] Each of Figures 8A-8D illustrates a different version of sequence trellis 710 generated by trellis iterator 700 at a different iteration. Figure 8A illustrates sequence trellis 710(0) in an initial state. Here, trellis iterator 700 generates sequence 820(0) of convex hull points that includes hull points 00, 10, 20, 30, and 40. These initially selected hull points have the lowest bitrate encoding and highest distortion, and therefore reside at the bottom of the respective columns. Based on sequence 820(0), trellis iterator 700 generates an encoded video sequence 180 that includes the resampled, encoded shot sequences 220 associated with each of convex hull points 00, 10, 20, 30, and 40. Trellis iterator 700 also generates sequence RD point 720(0) based on that encoded video sequence 180.

[0065] Trellis iterator 710 then computes, for each convex hull point within sequence 820(0), the rate of change of distortion with respect to bitrate between the

convex hull point and the above-neighbor of the convex hull point. For example, trellis iterator 710 could compute the rate of change of distortion with respect to bitrate between nodes 00 and 01, 10 and 11, 20 and 21, 30 and 31, and 40 and 41. The computed rate of change for the convex hull point associated with a given resampled, encoded shot sequence 220 represents the derivative of the distortion curve associated with that shot sequence, taken at the convex hull point.

[0066] Trellis iterator 710 selects the derivative having the greatest magnitude, and then selects the above neighbor associated with that derivative for inclusion in a subsequent sequence 820. For example, in Figure 8B, trellis iterator 700 determines that the derivative associated with convex hull point 30 is greatest, and therefore includes convex hull point 31 (the above-neighbor of convex hull point 30) in sequence 820(1). Based on sequence 820(1), trellis iterator 700 generates an encoded video sequence 180 that includes the resampled, encoded shot sequences 220 associated with each of convex hull points 00, 10, 20, 31, and 40. Trellis iterator 710 then generates sequence RD point 720(1) based on that encoded video sequence 180. Trellis iterator 710 performs this technique iteratively, thereby ascending trellis 710, as shown in Figures 8C-8D.

[0067] In Figure 8C, trellis iterator 700 determines that the derivative associated with convex hull point 10 is greatest compared to other derivatives, and then selects convex hull point 11 for inclusion in sequence 820(2). Based on sequence 820(2), trellis iterator 700 generates an encoded video sequence 180 that includes the resampled, encoded shot sequences 220 associated with each of convex hull points 11, 10, 20, 31, and 40. Trellis iterator 700 also generates sequence RD point 720(2) based on that encoded video sequence 180. Trellis iterator 700 continues this process until generating sequence 820(T) associated with trellis iteration 710(T), as shown in Figure 8D. In this manner, trellis iterator 700 incrementally improves sequences 820 by selecting a single hull point for which bitrate is increased and distortion is decreased, thereby generating a collection of encoded video sequences 180 with increasing bitrate and decreasing distortion.

[0068] In one embodiment, trellis iterator 700 adds convex hull points prior to ascending trellis 710 in order to create a terminating condition. In doing so, trellis iterator 700 may duplicate convex hull points having the greatest bitrate to cause the rate of change between the second to last and the last convex hull point to be zero.

When this zero rate of change is detected for all shots, i.e. when the maximum magnitude of rate of change is exactly zero, trellis iterator 700 identifies the terminating condition and stops iterating.

[0069] Referring back now to Figure 7, trellis iterator 700 generates encoded video sequences 180 that correspond to the sequences 820 shown in Figures 8A-8D via the trellis technique described above. Because trellis iterator 700 generates sequences 820 in an ascending manner to reduce distortion and increase bitrate, encoded video sequences 180 span a range from high distortion and low bitrate to low distortion and high bitrate. Each sequence RD point 720 provides the distortion and bitrate for a given encoded video sequence 180, and these sequence RD points 720 can be plotted to generate a convex hull, as discussed below in conjunction with Figure 9.

[0070] Figure 9 is a graph of convex hull data points generated for the different versions of the video sequence shown in Figures 8A-8D, according to various embodiments of the present invention. As shown, a graph 900 includes a bitrate axis 910 and a distortion axis 920. Curve 930 is plotted against bitrate axis 910 and distortion axis 920. Curve 930 can be generated based on the collection of sequence RD points 720 corresponding to the encoded video sequences 180 generated via the trellis technique discussed above in conjunction with Figures 8A-8D. Accordingly, curve 930 represents distortion as a function of bitrate across all encoded video sequences 180. An exemplary sequence RD point 720 is shown, corresponding to a bitrate 912 and distortion 922.

[0071] Based on curve 930, streaming application 730 of Figure 7 is capable of selecting, for a given available bitrate, the particular encoded video sequence 180 that minimizes distortion for that bitrate. Streaming application 730 may select a single encoded video sequence 180 during streaming, or dynamically select between video sequences. For example, streaming application 730 could switch between encoded video sequences 180 at shot boundaries. With this approach, streaming application 730 may deliver a consistent quality video experience to the end user without requiring excessive bandwidth.

[0072] Encoding engine 160 may implement variations on the technique described above in order to reduce storage and computational complexity. In one embodiment, encoding engine 160 implements a "constrained" version of the above approach. Referring now to Figure 3, to implement the constrained version, encoding engine 160

only encodes resampled sequences 320 with a limited range of QP values.

Accordingly, instead of generating versions of resampled sequence 320 for all possible QP values, encoding engine 160 may select a desired range of QP values and then only encode resampled sequence 320 for that range of QP values.

- 5 Because higher QP values provide quality that is intolerably low, those higher QP values may be deemed unnecessary for encoding purposes. Likewise, because lower QP values require an unreasonable bitrate, those QP values may also be considered unnecessary. Accordingly, encoding engine 160 may constrain encoding to only the QP values that are likely to produce encodes that should actually be delivered to the end-user. In a further embodiment, encoding engine 160 fixes the number of different encodes generated per shot to a constant value. In situations where fewer encodes are generated for a given shot than the constant value, encoding engine 160 may replicate encodes in order to meet the constant value. In various embodiments, encoding engine 160 may limit the number of values for an encoding parameter used for encoding purposes in any technically feasible fashion.

- 15 **[0073]** For instance, encoding engine 160 may select a subset of possible values for the encoding parameter based on one or more efficiency-related criteria. For example, encoding engine 160 could select every other QP value, every third QP value, or every Nth QP value (where N is an integer that ranges from 2 to 10, inclusive) within a smaller range of QP values. In some alternate embodiments, statistics may be collected from multiple encodings of various shot sequences to determine the statistical probability that different QP values will be selected by encoding engine 160. Subsequently, the QP values that encoding engine 160 uses for encoding purposes may be limited based on the statistical probabilities. More specifically, encoding engine 160 may be constrained to use the QP values that encoding engine 160 is most likely to select based on the statistical probabilities.

- 25 **[0074]** In another embodiment, encoding engine 160 implements an "iterative" version of the above approach whereby encoding engine 160 performs multiple encoding passes to determine an encoding having a target bitrate or target distortion level. Initially, encoding engine 160 may perform a first pass using a constrained range of QP values such as that discussed above in conjunction with the "constrained" approach. Once encoding engine 160 has generated a convex hull of sequence RD points, such as that shown in Figure 9, encoding engine 160 then identifies the sequence RD point closest to the target bitrate or target distortion level.

Encoding engine 160 then identifies one or more nearby points on the convex hull and, based on the range of QPs associated with those points, performs additional encodes. In this manner, encoding engine 160 may iteratively refine the range of QPs used for encoding in order to target a particular bitrate or distortion.

5 [0075] In yet another embodiment, encoding engine 160 implements a “fixed quality” version of the above approach and limits the number of shot encodes that need to be stored and subsequently processed. With this approach, encoding engine 160 may produce shot encodes at predetermined, well-spaced quality intervals. Encoding engine 160 may then assemble these shot encodes into
10 complete encoded video sequences 180 having a fixed quality across the entire sequence. The number of shot encodes implemented per shot sequence is a configurable parameter that represents a tradeoff between quality and storage needs. In performing this technique, encoding engine 160 processes convex hull points 580 and then iteratively removes extraneous points until the remaining points represent
15 the desired number of shot encodes. Encoding engine 160 could, for example, iteratively remove convex hull points 580 having the smallest gap relative to adjacent convex hull points 580. This technique allows encoding engine 160 to maximize the minimum quality of shot encodes.

[0076] In other embodiments, encoding engine 160 implements a “min-max
20 optimization” version of the above approach. In such an implementation, encoding engine 160 selects a convex hull point for inclusion in a subsequent sequence 820 based on the distortion metrics or quality metrics instead of the derivative values. In particular, encoding engine 150 determines the convex hull point included in sequence 820(x) that has the maximum distortion metric (or maximum quality metric)
25 and then includes the above-neighbor of the selected convex hull point for inclusion in the subsequent sequence 820(x+1).

[0077] In related embodiments, when ascending sequence trellis 710 encoding engine 160 may tradeoff changes in slope between convex hull points 580 with actual quality value. In doing so, prior to selecting a convex hull point 580 for inclusion into a
30 subsequent sequence, encoding engine 160 may filter out shot sequences (and corresponding convex hull points 580) with a quality metric below a given threshold (or distortion metric above a given threshold). Only after constraining the available shot sequences and convex hull points in this manner does encoding engine 160

generate a subsequent encoded video sequence 180 based on comparing slope values of the remaining convex hull points 580. This approach may maximize both average quality and minimum quality.

[0078] With any of the approaches discussed thus far, encoding engine 160 may be configured to enforce specific constraints that limit encoding behavior. For example, encoding engine 160 could be configured to limit the distortion of encoded shot sequences to always fall beneath a maximum tolerable distortion level. However, adjustments to encoding engine 160 may be needed in order to allow compliance with more complex constraints. An example of a complex constraint is the video buffer verifier (VBV) constraint, which is known to those skilled in the art. The VBV constraint generally states that data should arrive with a relatively constant bitrate and be stored in a buffer having relatively constant size. This constraint helps to avoid buffer overflow and/or underflow, among other potential issues. More specific formulations of the VBV constraint are also known to those skilled in the art, including the VBV constant bit rate (CBR) constraint and the VBV variable bit rate (VBR) constraint, although discussion of these specific versions is omitted for brevity.

[0079] In one embodiment, encoding engine 160 may be configured to perform the trellis ascension discussed previously in conjunction with Figures 8A-8D in a manner that allows the final encoded video sequences 180 to comply with arbitrarily complex sets of constraints, including the VBV constraint discussed above. In doing so, encoding engine 160 analyzes not only the slope values between neighboring convex hull points 580 to select a new hull point for inclusion into a subsequent sequence, but also compliance of each possible subsequent sequence with one or more constraints (e.g., VBV CBR, VBV VBR, and so forth). Thus, for each convex hull point 580 that could be potentially included in a subsequent sequence, encoding engine 160 determines the degree to which that sequence complies with the constraints. Encoding engine 160 then selects convex hull points 580 that allow subsequent sequences to maintain compliance. This form of trellis ascension constitutes a “dynamic programming” approach, and may also represent a form of Viterbi solution to the specific problem of optimizing bitrate versus distortion.

[0080] In alternate embodiments, encoding engine 180 and streaming application 730 may cause encoded video sequences 180 to be delivered to endpoint devices in any technically feasible fashion. In the same or other embodiments, any amount and

type of the functionality associated with encoding engine 180 and streaming application 730 may be implemented in or distributed across any number of host computers 110, any number of cloud computers 140, any number of client computers (not shown), and any number of endpoint devices, in any technically feasible fashion.

5 **[0081]** For instance, in some embodiments, encoding engine 180 configures streaming application 730 to deliver metadata to client applications executing on endpoint devices. Metadata includes, without limitation, metrics associated with encoded video content at any level of granularity, such as bitrates and quality metrics associated with one or more encoded shot sequences and/or encoded video
10 sequences 180. The client applications may perform any type and amount of adaptive streaming operations based on the metadata in any technically feasible fashion.

[0082] In one scenario, a user configures a video player application to stream a movie to a laptop. Streaming application 190 transmits the metadata associated with four
15 different encoded video sequences 180(1-4) to the video player application. The metadata indicates that encoded video sequence 180(4) is associated with the highest bitrate and the highest visual quality, while encoded video sequence 180(1) is associated with the lowest bitrate and the lowest visual quality. At any given time, the video player application selects the encoded video sequence 180 that provides
20 the highest available visual quality during playback of the movie while avoiding playback interruptions due to rebuffering.

[0083] Based on an initial available bandwidth and the metadata, the video player application configures streaming application 730 to begin streaming encoded video sequence 180(4) to the video player application. In this fashion, the video player
25 application provides the highest available visual quality during playback of the movie. In general, because of internet traffic, especially during peak times during the day, connection conditions can change quickly and become quite variable. In the described scenario, after ten minutes of playback, the available bandwidth decreases dramatically. Based on the reduced bandwidth and the metadata, the video player
30 application configures streaming application 730 to dynamically switch between encoded video sequence 180(4) and encoded video sequence 180(1). At the next shot boundary, streaming application 730 begins streaming encoded video sequence 180(1) instead of encoded video sequence 180(4) to the video player application.

Although the video player application is no longer able to provide the highest available visual quality during playback of the movie, the video player application successfully avoids playback interruptions due to rebuffering.

[0084] Persons skilled in the art will understand that the techniques described thus far are applicable beyond video to audio as well. For example, the objective quality metric discussed above could provide a measure of audio quality. The remaining portions of the above techniques would proceed in otherwise similar fashion.

Procedures for Generating Encoded Video Sequences

[0085] Figure 10 is a flow diagram of method steps for assembling chunks of video content into an encoded video sequence, according to various embodiments of the present invention. Although the method steps are described in conjunction with the systems of Figures 1-9, persons skilled in the art will understand that any system configured to perform the method steps, in any order, is within the scope of the present invention.

[0086] As shown, a method 1000 begins at step 1002, where encoding engine 160 receives source video sequence 170. Source video sequence 170 includes a sequence of frames encoded in a native or "distribution" format. At step 1004, encoding engine 160 processes source video sequence 170 to remove superfluous pixels. Such pixels may reside in horizontal or vertical black bars residing adjacent to the actual content of the video sequence. At step 1006, encoding engine 160 cuts source video sequence 170 into shot sequences 220. Each shot sequence 220 includes a subsequence of frames captured from a particular camera or simulated camera (in the case of computer animated sequences).

[0087] The method then proceeds to step 1008. At step 1008, for each shot sequence 220, encoding engine 160 resamples the shot sequence M times to generate a resolution ladder 330 of resampled sequences 320, as shown in Figure 3. Each resampled sequence 320 has a different resolution. One resampled sequence 320 has the same resolution as the original video sequence.

[0088] The method then proceeds to step 1010. For each resampled sequence 320 in resolution ladder 330, encoding engine 160 processes the resampled sequence 320 via a processing pipeline 340 to generate data points 350. Specific processing steps executed by processing pipeline 340 are described in greater detail below in conjunction with Figure 11. Each data point 350 indicates, for a given

resampled sequence 320, the encoding resolution of the sequence, a quality metric for the sequence, and the QP value used to encode the sequence, as discussed in greater detail below in conjunction with Figure 11.

[0089] At step 1012, encoding engine 160 collects all data points 350 for all
5 resampled sequences 320 in resolution ladder 330 to generate a data set 360. Data set 360 corresponds to one shot sequence 220. Each data point in data set 360 corresponds to a different encoding and different resolution of the shot sequence. At step 1014, encoding engine 160 converts the quality metric associated with these data points to a distortion metric, and then generates convex hull points 580 for the
10 dataset, as shown in Figure 5B. Convex hull points 580 minimize distortion or bitrate across all resampled / encoded shot sequences.

[0090] At step 1016, encoding engine 160 collects all convex hull points 580 across all resolution ladders to generate a sequence trellis 710. The construction of an exemplary sequence trellis 710 is discussed in detail in conjunction with Figures
15 8A-8D. At step 1018, encoding engine 160 iteratively ascends the sequence trellis to generate a collection of encoded video sequences 180 and corresponding sequence RD points 720. An approach for ascending sequence trellis 710 is discussed in conjunction with Figure 12.

[0091] At step 1020, streaming application 730 selects an encoded video
20 sequence 180 for streaming based on the associated sequence RD point 720. In doing so, streaming application may select a particular sequence RD point 720 that minimizes distortion for a given available bitrate, and then stream the encoded video sequence 180 associated with that sequence RD point 720 to an endpoint device.

[0092] Figure 11 is a flow diagram of method steps for processing a resampled
25 shot sequence to generate a set of data points, according to various embodiments of the present invention. Although the method steps are described in conjunction with the systems of Figures 1-9, persons skilled in the art will understand that any system configured to perform the method steps, in any order, is within the scope of the present invention.

[0093] Encoding engine 160 implements a method 1100 to perform processing
30 associated with a given sub-pipeline 450 within a processing pipeline 340. Encoding engine 160 may execute multiple sub-pipelines 450 in parallel to implement a given processing pipeline 340, and may thus perform the method 1100 multiple times.

[0094] As shown, the method 1100 begins at step 1102, where encoding engine 160 encodes a resampled sequence 320 with a selected quantization parameter (QP). At step 1104, encoding engine 160 then decodes the encoded sequence and, at step 1106, upsamples the decoded sequence to the resolution associated with source video sequence 170. At step 1108, encoding engine 160 generates one or more quality metrics (QMs) for the upsampled sequence. At step 1110, encoding engine 160 generates a data point 440 that includes the resampled sequence resolution, the choice of quantization parameter (QP), and the quality metric (QM) generated for the encoded resampled video sequence.

[0095] Figure 12 is a flow diagram of method steps for generating a set of encoded video sequences, according to various embodiments of the present invention. Although the method steps are described in conjunction with the systems of Figures 1-9, persons skilled in the art will understand that any system configured to perform the method steps, in any order, is within the scope of the present invention.

[0096] As shown, a method 1200 begins at step 1202, where encoding engine 160 generates a sequence trellis 710 based on convex hull points 580 for all shot sequences 220. Sequence trellis 710, as discussed above in conjunction with Figures 8A-8D, includes individual columns of convex hull points 580, where each column corresponds to a particular shot sequence. Accordingly, an encoded version of source video sequence 170 may be constructed by collecting one encoded, resampled shot sequence 220 from each such column.

[0097] At step 1204, encoding engine 160 determines a sequence of convex hull points 580 having the lowest bitrate. At step 1206, encoding engine 160 designates the determined sequence as the "current sequence." At step 1208, encoding engine 160 generates an encoded video sequence based on the current sequence. In doing so, encoding engine 160 collects each resampled, encoded shot sequence 220 associated with the sequence of convex hull points 580 to construct an encoded version of the source video sequence 170. At step 1210, encoding engine 160 generates a sequence RD point 720 based on that encoded video sequence.

[0098] At step 1212, encoding engine 160 computes the magnitude of the slope between each convex hull point in the current sequence and the above-neighbor convex hull point. The "above-neighbor" of a given convex hull point resides immediately above the convex hull point and in the same column. At step 1214,

encoding engine 160 identifies the convex hull point and above-neighbor convex hull point with greatest slope magnitude relative to one another. At step 1216, encoding engine 160 generates a new sequence of convex hull points that replaces the convex hull point with the above-neighbor convex hull point. Finally, at step 1218, encoding engine 160 designates the new sequence as the "current sequence" and returns to step 1208. Encoding engine 160 may repeat the method 1200 until generating an encoded sequence 170 with maximum bitrate compared to other sequences, or until another terminating condition is met.

[0099] In this manner, encoding engine 160 "climbs" sequence trellis 710 by determining subsequent versions of the current sequence that maximally reduce distortion and bitrate compared to other versions. By ascending sequence trellis 710 in this manner, encoding engine 160 need not consider all possible combinations of all resampled, encoded shot sequences (also referred to herein as "chunks"). Accordingly, encoding engine 160 may conserve considerable computing resources while still determining a spectrum of encoded video sequences that optimizes distortion for a range of bitrates.

[0100] In sum, an encoding engine encodes a video sequence to provide optimal quality for a given bitrate. The encoding engine cuts the video sequence into a collection of shot sequences. Each shot sequence includes video frames captured from a particular capture point. The encoding engine resamples each shot sequence across a range of different resolutions, encodes each resampled sequence with a range of quality parameters, and then upsamples each encoded sequence to the original resolution of the video sequence. For each upsampled sequence, the encoding engine computes a quality metric and generates a data point that includes the quality metric and the resample resolution. The encoding engine collects all such data points and then computes the convex hull of the resultant data set. Based on all convex hulls across all shot sequences, the encoding engine determines an optimal collection of shot sequences for a range of bitrates.

[0101] At least one advantage of the techniques described herein is that the video sequence can be streamed to an end-user with the best available quality for a given bitrate. Conversely, for a given desired quality, the video sequence can be provided with the minimum possible bitrate.

[0102] 1. Some embodiments of the invention include computer-implemented method, comprising: generating a first set of encoded chunks for a source video sequence, generating a first set of data points based on the first set of encoded chunks, performing one or more convex hull operations across the first set of data points to compute a first subset of data points that are optimized across at least two metrics, computing a first slope value between a first data point included in the first subset of data points and a second data point included in the first subset of data points, and determining, based on the first slope value, that a first encoded chunk associated with the first data point should be included in a final encoded version of the source video sequence.

[0103] 2. The computer-implemented method of clause 1, wherein generating the first set of encoded chunks comprises: identifying within the source video sequence a first sequence of frames that is associated with a first point of capture, resampling the first sequence of frames at a plurality of different resolutions to generate a resolution ladder of resampled versions of the first sequence of frames, and encoding each resampled version of the first sequence of frames with a different encoding parameter to generate the first set of encoded chunks.

[0104] 3. The computer-implemented method of any of clauses 1 and 2, wherein generating the first set of data points comprises: decoding each encoded chunk in the first set of encoded chunks to generate a first set of decoded chunks, upsampling each decoded chunk in the first set of decoded chunks to a source resolution associated with the source video sequence to generate a first set of upsampled chunks, and generating a different data point for each upsampled chunk in the first set of upsampled data chunks.

[0105] 4. The computer-implemented method of any of clauses 1, 2, and 3, wherein a specific data point in the first set of data points is generated by: generating a specific objective quality metric for a specific upsampled chunk in the first set of upsampled chunks, converting the specific objective quality metric to a specific distortion metric, computing a bitrate for the specific upsampled chunk, combining the specific distortion metric and the bitrate to generate the specific data point.

[0106] 5. The computer-implemented method of any of clauses 1, 2, 3, and 4, wherein performing one or more convex hull operations across the first set of data points to compute the first subset of data points comprises: determining a first region

that includes the first set of data points, identifying a first boundary of the first region, wherein no data points in the first set of points reside on a first side of the first boundary, discarding any data points that do not reside along the first boundary, wherein each data point that resides along the first boundary optimizes the first metric
5 with respect to the second metric.

[0107] 6. The computer-implemented method of any of clauses 1, 2, 3, 4, and 5, wherein the first metric comprises distortion and the second metric comprises bitrate.

[0108] 7. The computer-implemented method of any of clauses 1, 2, 3, 4, 5, and 6, further comprising: generating a second set of encoded chunks for the source
10 video sequence, generating a second set of data points based on the second set of encoded chunks, performing one or more convex hull operations across the second set of data points to compute a second subset of data points that are optimized across the at least two metrics, and computing a second slope value between a third data point included in the second subset of data points and a fourth data point
15 included in the second subset of data points.

[0109] 8. The computer-implemented method of any of clauses 1, 2, 3, 4, 5, 6, and 7, wherein determining that the first encoded chunk associated with the first data point should be included in the final encoded version of the source video sequence comprises determining that the first slope has a greater magnitude than the second
20 slope.

[0110] 9. The computer-implemented method of any of clauses 1, 2, 3, 4, 5, 6, 7, and 8, further comprising determining that a second encoded chunk associated with the fourth data point should be included in another encoded version of the source video sequence based on determining that the second slope value is greater than
25 other slope values associated with other subsets of data points.

[0111] 10. The computer-implemented method of any of clauses 1, 2, 3, 4, 5, 6, 7, 8, and 9, wherein the first set of encoded chunks is associated with a first sequence of video frames captured continuously from a first point of capture, and a second set of encoded chunks is associated with a second sequence of video frames
30 captured continuously from a second point of capture.

[0112] 11. A non-transitory computer-readable medium storing program instructions that, when executed by a processor, configures the processor to perform the steps of: generating a first set of encoded chunks for a source video sequence,

generating a first set of data points based on the first set of encoded chunks,
performing one or more convex hull operations across the first set of data points to
compute a first subset of data points that are optimized across at least two metrics,
computing a first slope value between a first data point included in the first subset of
5 data points and a second data point included in the first subset of data points, and
determining, based on the first slope value, that a first encoded chunk associated with
the first data point should be included in a final encoded version of the source video
sequence.

[0113] 12. The non-transitory computer-readable medium of clause 11, wherein
10 the step of generating the first set of encoded chunks comprises identifying within the
source video sequence a first sequence of frames that is associated with a first point
of capture, resampling the first sequence of frames at a plurality of different
resolutions to generate a resolution ladder of resampled versions of the first sequence
of frames, and encoding each resampled version of the first sequence of frames with
15 a different encoding parameter to generate the first set of encoded chunks.

[0114] 13. The non-transitory computer-readable medium of any of clauses 11
and 12, wherein the step of generating the first set of encoded chunks comprises
generating a plurality of values for an encoding parameter based on a plurality of
possible values and a maximum number of encoded chunks, wherein the total
20 number of values included in the plurality of values is less than the total number of
possible values included in the plurality of possible values; and encoding a plurality of
resampled versions of a first sequence of frames based on the plurality of values for
the encoding parameter to generate the first set of encoded chunks.

[0115] 14. The non-transitory computer-readable medium of any of clauses 11,
25 12, and 13, wherein the step of generating the first set of data points comprises
decoding each encoded chunk in the first set of encoded chunks to generate a first
set of decoded chunks; upsampling each decoded chunk in the first set of decoded
chunks to a source resolution associated with the source video sequence to generate
a first set of upsampled chunks; and generating a different data point for each
30 upsampled chunk in the first set of upsampled data chunks.

[0116] 15. The non-transitory computer-readable medium of any of clauses 11,
12, 13, and 14, wherein the step of performing one or more convex hull operations
across the first set of data points to compute the first subset of data points comprises:

determining a first region that includes the first set of data points, identifying a first boundary of the first region, wherein no data points in the first set of points reside on a first side of the first boundary, including any data points that reside along the first boundary in the first subset of data points.

5 **[0117]** 16. The non-transitory computer-readable medium of any of clauses 11, 12, 13, 14, and 15, wherein the first metric comprises distortion and the second metric comprises bitrate.

10 **[0118]** 17. The non-transitory computer-readable medium of any of clauses 11, 12, 13, 14, 15, and 16, further comprising the steps of: generating a second set of encoded chunks for the source video sequence, generating a second set of data points based on the second set of encoded chunks, performing one or more convex hull operations across the second set of data points to compute a second subset of data points that are optimized across the at least two metrics, and computing a second slope value between a third data point included in the second subset of data points and a fourth data point included in the second subset of data points.

15 **[0119]** 18. The non-transitory computer-readable medium of any of clauses 11, 12, 13, 14, 15, 16, and 17, wherein determining that the first encoded chunk associated with the first data point should be included in the final encoded version of the source video sequence comprises determining that the first slope has a greater magnitude than the second slope.

20 **[0120]** 19. The non-transitory computer-readable medium of any of clauses 11, 12, 13, 14, 15, 16, 17, and 18, further comprising determining that a second encoded chunk associated with the fourth data point should not be included in another encoded version of the source video sequence based on determining that the second slope value is less than one or more other slope values associated with one or more other subsets of data points.

25 **[0121]** 20. The non-transitory computer-readable medium of any of clauses 11, 12, 13, 14, 15, 16, 17, 18, and 19, wherein the first set of encoded chunks is associated with a first shot sequence and a second set of encoded chunks is associated with a second shot sequence.

30 **[0122]** 21. Some embodiments include a system, comprising: a memory storing a software application, and a processor that is couple to the memory and, when executing the software application, is configured to: generate a first set of encoded

chunks for a source video sequence, generate a first set of data points based on the first set of encoded chunks, perform one or more convex hull operations across the first set of data points to compute a first subset of data points that are optimized across at least two metrics, compute a first slope value between a first data point
5 included in the first subset of data points and a second data point included in the first subset of data points, and determine, based on the first slope value, that a first encoded chunk associated with the first data point should be included in a final encoded version of the source video sequence.

[0123] 22. The system of clause 21, wherein, when executing the software
10 application, the processor is further configured to: generate the first set of encoded chunks, generate the first set of data points, perform the one or more convex hull operations, compute the first slope value, and determine that the first encoded chunk associated with the first data point should be included in the final encoded version of the source video sequence.

[0124] Any and all combinations of any of the claim elements recited in any of the claims and/or any elements described in this application, in any fashion, fall within the contemplated scope of the present invention and protection.

[0125] The descriptions of the various embodiments have been presented for purposes of illustration, but are not intended to be exhaustive or limited to the
20 embodiments disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the described embodiments.

[0126] Aspects of the present embodiments may be embodied as a system, method or computer program product. Accordingly, aspects of the present
25 disclosure may take the form of an entirely hardware embodiment, an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and hardware aspects that may all generally be referred to herein as a "module" or "system." In addition, any hardware and/or software technique, process, function, component, engine, module, or system
30 described in the present disclosure may be implemented as a circuit or set of circuits. Furthermore, aspects of the present disclosure may take the form of a computer program product embodied in one or more computer readable medium(s) having computer readable program code embodied thereon.

[0127] Any combination of one or more computer readable medium(s) may be utilized. The computer readable medium may be a computer readable signal medium or a computer readable storage medium. A computer readable storage medium may be, for example, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples (a non-exhaustive list) of the computer readable storage medium would include the following: an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing. In the context of this document, a computer readable storage medium may be any tangible medium that can contain, or store a program for use by or in connection with an instruction execution system, apparatus, or device.

[0128] Aspects of the present disclosure are described above with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the disclosure. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine. The instructions, when executed via the processor of the computer or other programmable data processing apparatus, enable the implementation of the functions/acts specified in the flowchart and/or block diagram block or blocks. Such processors may be, without limitation, general purpose processors, special-purpose processors, application-specific processors, or field-programmable gate arrays.

[0129] The flowchart and block diagrams in the figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods and computer program products according to various embodiments of the present disclosure. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of code, which comprises one or more executable instructions for implementing the specified logical function(s). It should

also be noted that, in some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts, or combinations of special purpose hardware and computer instructions.

[0130] While the preceding is directed to embodiments of the present disclosure, other and further embodiments of the disclosure may be devised without departing from the basic scope thereof, and the scope thereof is determined by the claims that follow.

The invention claimed is:

1. A computer-implemented method, comprising:
 - generating multiple sets of encoded chunks for a source video sequence, wherein each set of encoded chunks corresponds to a different shot sequence included in the source video sequence;
 - generating, for a first set of encoded chunks, a first set of data points based on the first set of encoded chunks;
 - performing, for the first set of data points, one or more convex hull operations across the first set of data points to compute a first subset of data points that are optimized across at least two metrics;
 - computing a first slope value between a first data point included in the first subset of data points and a second data point included in the first subset of data points;
 - comparing the first slope value against one or more other slope values to select a final slope value, wherein each of the one or more other slope values is computed in association with a different set of encoded chunks; and
 - determining that a first encoded chunk included in the set of encoded chunks associated with the final slope value should be included in a final encoded version of the source video sequence.
2. The computer-implemented method of claim 1, wherein generating the multiple sets of encoded chunks comprises:
 - identifying within the source video sequence a first sequence of frames that is associated with a first point of capture;
 - resampling the first sequence of frames at a plurality of different resolutions to generate a resolution ladder of resampled versions of the first sequence of frames; and
 - encoding each resampled version of the first sequence of frames with a different encoding parameter to generate the first set of encoded chunks.
3. The computer-implemented method of claim 1, wherein generating the first set of data points comprises:
 - decoding each encoded chunk in the first set of encoded chunks to generate a first set of decoded chunks;

upsampling each decoded chunk in the first set of decoded chunks to a source resolution associated with the source video sequence to generate a first set of upsampled chunks; and

generating a different data point for each upsampled chunk in the first set of upsampled data chunks.

4. The computer-implemented method of claim 3, wherein a specific data point in the first set of data points is generated by:

generating a specific objective quality metric for a specific upsampled chunk in the first set of upsampled chunks;

converting the specific objective quality metric to a specific distortion metric;

computing a bitrate for the specific upsampled chunk; and

combining the specific distortion metric and the bitrate to generate the specific data point.

5. The computer-implemented method of claim 1, wherein performing the one or more convex hull operations across the first set of data points to compute the first subset of data points comprises:

determining a first region that includes the first set of data points;

identifying a first boundary of the first region, wherein no data points in the first set of points reside on a first side of the first boundary; and

discarding any data points that do not reside along the first boundary, wherein each data point that resides along the first boundary optimizes a first metric included in the at least two metrics with respect to a second metric included in the at least two metrics.

6. The computer-implemented method of claim 5, wherein the first metric comprises distortion and the second metric comprises bitrate.

7. The computer-implemented method of claim 1, further comprising:

generating a second set of data points based on a second set of encoded chunks included in the multiple sets of encoded chunks;

performing one or more convex hull operations across the second set of data points to compute a second subset of data points that are optimized across the at least two metrics; and

computing a second slope value between a third data point included in the second subset of data points and a fourth data point included in the second subset of data points.

8. The computer-implemented method of claim 7, wherein determining that the first encoded chunk associated with the final slope value should be included in the final encoded version of the source video sequence comprises determining that the first slope has a greater magnitude than the second slope.

9. The computer-implemented method of claim 7, further comprising determining that a second encoded chunk associated with the fourth data point should be included in another encoded version of the source video sequence based on determining that the second slope value has a greater magnitude than other slope values computed in associated with other sets of encoded chunks.

10. The computer-implemented method of claim 1, wherein the first set of encoded chunks is associated with a first sequence of video frames captured continuously from a first point of capture, and a second set of encoded chunks is associated with a second sequence of video frames captured continuously from a second point of capture.

11. One or more non-transitory computer-readable media storing program instructions that, when executed by one or more processors, configure the one or more processors to perform the steps of:

generating multiple sets of encoded chunks for a source video sequence, wherein each set of encoded chunks corresponds to a different shot sequence included in the source video sequence;

generating, for a first set of encoded chunks, a first set of data points based on the first set of encoded chunks;

performing, for the first set of data points, one or more convex hull operations across the first set of data points to compute a first subset of data points that are optimized across at least two metrics;

computing a first slope value between a first data point included in the first subset of data points and a second data point included in the first subset of data points;

comparing the first slope value against one or more other slope values to select a final slope value, wherein each of the one or more other slope values is computed in association with a different set of encoded chunks; and

determining that a first encoded chunk included in the set of encoded chunks associated with the final slope value should be included in a final encoded version of the source video sequence.

12. The one or more non-transitory computer-readable media of claim 11, wherein the step of generating the multiple sets of encoded chunks comprises:

identifying within the source video sequence a first sequence of frames that is associated with a first point of capture;

resampling the first sequence of frames at a plurality of different resolutions to generate a resolution ladder of resampled versions of the first sequence of frames; and

encoding each resampled version of the first sequence of frames with a different encoding parameter to generate the first set of encoded chunks.

13. The one or more non-transitory computer-readable media of claim 11, wherein the step of generating the multiple sets of encoded chunks comprises:

generating a plurality of values for an encoding parameter based on a plurality of possible values and a maximum number of encoded chunks, wherein the total number of values included in the plurality of values is less than the total number of possible values included in the plurality of possible values; and encoding a plurality of resampled versions of a first sequence of frames based on the plurality of values for the encoding parameter to generate the first set of encoded chunks.

14. The one or more non-transitory computer-readable media of claim 11, wherein the step of generating the first set of data points comprises:
- decoding each encoded chunk in the first set of encoded chunks to generate a first set of decoded chunks;
 - upsampling each decoded chunk in the first set of decoded chunks to a source resolution associated with the source video sequence to generate a first set of upsampled chunks; and
 - generating a different data point for each upsampled chunk in the first set of upsampled data chunks.
15. The one or more non-transitory computer-readable media of claim 11, wherein the step of performing the one or more convex hull operations across the first set of data points to compute the first subset of data points comprises:
- determining a first region that includes the first set of data points; identifying a first boundary of the first region, wherein no data points in the first set of points reside on a first side of the first boundary; and
 - including any data points that reside along the first boundary in the first subset of data points.
16. The one or more non-transitory computer-readable media of claim 15, wherein a first metric included in the at least two metrics comprises distortion and a second metric included in the at least two metrics comprises bitrate.
17. The one or more non-transitory computer-readable media of claim 11, further comprising the steps of: generating a second set of data points based on a second set of encoded chunks included in the multiple sets of encoded chunks; performing one or more convex hull operations across the second set of data points to compute a second subset of data points that are optimized across the at least two metrics; and
- computing a second slope value between a third data point included in the second subset of data points and a fourth data point included in the second subset of data points.
18. The one or more non-transitory computer-readable media of claim 17,

wherein determining that the first encoded chunk associated with the final slope value should be included in the final encoded version of the source video sequence comprises determining that the first slope has a greater magnitude than the second slope.

19. The one or more non-transitory computer-readable media of claim 17, further comprising determining that a second encoded chunk associated with the fourth data point should not be included in another encoded version of the source video sequence based on determining that the second slope value has a smaller magnitude is less than one or more other slope values associated with one or more other subsets of data points.

20. The one or more non-transitory computer-readable media of claim 17, wherein the first set of encoded chunks is associated with a first shot sequence and a second set of encoded chunks is associated with a second shot sequence.

21. A system, comprising:

one or more memories storing a software application; and one or more processors that are coupled to the one or more memories and, when executing the software application, are configured to:

generate multiple sets of encoded chunks for a source video sequence, wherein each set of encoded chunks corresponds to a different shot sequence included in the source video sequence, generate, for a first set of encoded chunks, a first set of data points based on the first set of encoded chunks, perform, for the first set of data points, one or more convex hull operations across the first set of data points to compute a first subset of data points that are optimized across at least two metrics, compute a first slope value between a first data point included in the first subset of data points and a second data point included in the first subset of data points, compare the first slope value against one or more other slope values to select a final slope value, wherein each of the one or more other slope values is computed in association with a different set of encoded chunks, and determine that a first encoded chunk included in the set of encoded chunks associated with the final slope value should be included in a final encoded version of the source video sequence.