

VIDEO PROCESSING METHOD, VIDEO PROCESSING APPARATUS, SMART DEVICE, AND STORAGE MEDIUM

[0001] This application claims priority to Chinese Patent Application No. 202011239333.1,
5 entitled "VIDEO PROCESSING METHOD, VIDEO PROCESSING APPARATUS, SMART
DEVICE, AND STORAGE MEDIUM" and filed with the China National Intellectual Property
Administration on November 9, 2020, which is incorporated herein by reference in its entirety.

FIELD

10 **[0002]** This application relates to the technical field of computers, and in particular, to a
video processing method, a video processing apparatus, a smart device, and a computer-readable
storage medium.

BACKGROUND

15 **[0003]** At present, video encoding technology is widely used in scenarios such as a video
session or video on demand. For example, a session video involved in a video session scenario is
processed by the video encoding technology, and an on-demand video involved in a
video-on-demand scenario is processed by the video encoding technology, and the like. The video
encoding technology refers to a technology of compressing and encoding a video according to an
20 encoding mode. By compressing and encoding a video, the storage space of the video can be
effectively saved and the video transmission efficiency can be effectively improved.

[0004] In a video encoding process, to accurately determine an appropriate encoding mode
and to encode a video according to the encoding mode can accelerate the entire video encoding
process and improve the video encoding efficiency. Therefore, how to accurately determine an
25 encoding mode for video frame encoding in a video encoding process has become a hot spot in the
current research.

SUMMARY

[0005] Embodiments of the present disclosure provide a video processing method, apparatus, device, and storage medium, which can more accurately select an appropriate encoding mode to perform video frame encoding.

[0006] An aspect of the embodiments of the present disclosure provides a video processing method, performed by a smart device, including:

[0007] obtaining a target video frame from a to-be-encoded video and determining a to-be-encoded target data block from the target video frame;

[0008] obtaining a scene complexity of the target data block, and obtain data block indicator information of the target data block based on the scene complexity of the target data block;

[0009] dividing the target data block into N sub-data blocks, obtaining scene complexities of the N sub-data blocks, and obtaining sub-block indicator information based on the scene complexities of the N sub-data blocks, N being an integer greater than or equal to 2;

[0010] determining an encoding mode for the target data block according to the data block indicator information and the sub-block indicator information; and

[0011] encoding the target data block according to the determined encoding mode.

[0012] Another aspect of the embodiments of the present disclosure provides a video processing apparatus, including:

[0013] an obtaining unit, configured to obtain a target video frame from a to-be-encoded video and determine a to-be-encoded target data block from the target video frame; and

[0014] a processing unit, configured to obtain a scene complexity of the target data block, and obtain data block indicator information of the target data block based on the scene complexity of the target data block; divide the target data block into N sub-data blocks, obtain scene complexities of the N sub-data blocks, and obtain sub-block indicator information based on the scene complexities of the N sub-data blocks, N being an integer greater than or equal to 2; determine an encoding mode for the target data block according to the data block indicator information and the sub-block indicator information; and encode the target data block according to the determined encoding mode.

[0015] Another aspect of the embodiments of the present disclosure provides a smart device, the smart device including:

[0016] a processor, configured to implement a computer program; and

[0017] a memory, storing a computer program, the computer program, when executed by the processor, implementing the foregoing video processing method.

[0018] According to another aspect, an embodiment of the present disclosure provides a computer-readable storage medium, storing a computer program, the computer program, when executed by a processor, implementing the foregoing video processing method.

[0019] According to another aspect, an embodiment of the present disclosure provides a computer program product or a computer program, the computer program product or the computer program including computer instructions, the computer instructions being stored in a computer-readable storage medium. A processor of a computer device reads the computer instructions from the computer-readable storage medium, and executes the computer instructions, to implement the foregoing video processing method.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] To describe the technical solutions in the embodiments of the present disclosure or in the related art more clearly, the following briefly describes the accompanying drawings required for describing the embodiments. Apparently, the accompanying drawings in the following description show some embodiments of the present disclosure, and a person of ordinary skill in the art may still derive other accompanying drawings from these accompanying drawings without creative efforts.

[0021] FIG. 1a is a schematic diagram of a recursive division process of a video frame according to an embodiment of the present disclosure.

[0022] FIG. 1b is a schematic diagram of a division result of a video frame according to an embodiment of the present disclosure.

[0023] FIG. 2 is a schematic structural diagram of an encoder according to an embodiment of the present disclosure.

[0024] FIG. 3 is a schematic diagram of a data block correlated with a target data block according to an embodiment of the present disclosure.

[0025] FIG. 4 is a schematic architectural diagram of a video processing system according to an embodiment of the present disclosure.

[0026] FIG. 5 is a schematic flowchart of a video processing method according to an embodiment of the present disclosure.

[0027] FIG. 6 is a schematic flowchart of another video processing method according to an embodiment of the present disclosure.

[0028] FIG. 7 is a schematic flowchart of another video processing method according to an embodiment of the present disclosure.

5 **[0029]** FIG. 8 is a schematic structural diagram of a video processing apparatus according to an embodiment of the present disclosure.

[0030] FIG. 9 is a schematic structural diagram of a smart device according to an embodiment of the present disclosure.

10 **DESCRIPTION OF EMBODIMENTS**

[0031] The following clearly and completely describes the technical solutions in the embodiments of the present disclosure with reference to the accompanying drawings in the embodiments of the present disclosure. Apparently, the described embodiments are merely some but not all of the embodiments of the present disclosure. All other embodiments obtained by a person of ordinary skill in the art based on the embodiments of the present disclosure without creative efforts shall fall within the protection scope of the present disclosure.

[0032] The embodiments of the present disclosure relate to a cloud technology, and the embodiments of the present disclosure may implement a video processing process by using the cloud technology. The cloud technology is a hosting technology that unifies a series of resources such as hardware, software, and networks in a wide area network or a local area network to implement computing, storage, processing, and sharing of data. The cloud technology is a collective name of a network technology, an information technology, an integration technology, a management platform technology, an application technology, and the like based on an application of a cloud computing business mode, and may form a resource pool, which is used as required, and is flexible and convenient. The cloud computing technology becomes an important support. A background service of a technical network system requires a large amount of computing and storage resources, such as video websites, image websites, and more portal websites. As the Internet industry is highly developed and applied, each article may have its own identifier in the future and needs to be transmitted to a background system for logical processing. Data at different levels is separately processed, and data in various industries requires strong system support, which can only be implemented through cloud computing.

[0033] Cloud computing is a computing mode, in which computing tasks are distributed on a resource pool formed by a large quantity of computers, so that various application systems can obtain computing power, storage space, and information services according to requirements. A network that provides resources is referred to as a "cloud". For a user, resources in a "cloud" seem to be infinitely expandable, and can be obtained readily, used on demand, expanded readily, and paid for use. A basic capability provider of the cloud computing may establish a cloud computing resource pool (referred to as a cloud platform for short), commonly known as a Infrastructure as a Service (IaaS) platform, and deploy various types of virtual resources in the cloud computing resource pool for external customers to choose and use. The cloud computing resource pool mainly includes: a computing device (a virtualized machine, including an operating system), a storage device, and a network device. According to logical function division, a Platform as a Service (PaaS) layer may be deployed on an IaaS layer and a Software as a Service (SaaS) layer may be then deployed on the PaaS layer, or the SaaS may be directly deployed on the IaaS. The PaaS is a platform on which software (for example, a database, a web container, and the like) runs. The SaaS is a variety of types of service software (for example, a web portal, an SMS group sender, and the like). In general, the SaaS and the PaaS are upper layers relative to the IaaS. The cloud computing also refers to a delivery and use mode of an Internet Technology (IT) infrastructure and refers to obtaining a required resource through a network in an on-demand and easily scalable manner. Broadly defined cloud computing refers to a delivery and use mode of a service and refers to obtaining a required service through a network in an on-demand and easily scalable manner. Such the service may be IT and software and be correlated with the Internet, or may be another service. The cloud computing is a product of integration of conventional computer and network technologies such as grid computing, distributed computing, parallel computing, utility computing, network storage technologies, virtualization, and load balance.

[0034] The cloud computing may be used in the field of cloud conferencing, which is an efficient, convenient, and low-cost conference form based on the cloud computing technology. Users can quickly and efficiently share voice, data files, and videos with teams and customers around the world only by carrying out simple and easy-to-use operations through an Internet interface, and a cloud conference service provider helps the users to operate complex technologies such as data transmission and processing in the conference. At present, domestic cloud conferences are mainly focused on service content with a SaaS mode as the main body, including service forms such as a call, a network, and a video. Cloud-based video conferencing is referred to as cloud conferencing. In the era of the cloud conferencing, transmission, processing, and storage of data are all processed by computer resources of a video conferencing manufacturer. Users do not need to

purchase expensive hardware and install complicated software, and they can carry out an efficient remote conference only by opening a browser and logging into a corresponding interface. The cloud conference system supports multi-server dynamic cluster deployment and provides multiple high-performance servers, thereby greatly improving the conference stability, security, and availability. In recent years, video conferencing has been welcomed by many users because it can greatly improve the communication efficiency, continuously reduce the communication costs, and bring about upgrading of the internal management level, and has been widely used in various fields such as government, military, traffic, transportation, finance, operators, education, and enterprises. There is no doubt that after the use of the cloud computing, the video conferencing has a stronger attraction in terms of convenience, quickness, and ease of use, which may surely stimulate the arrival of a new upsurge of video conferencing applications.

[0035] The embodiments of the present disclosure relate to a video, which is a video frame sequence composed of consecutive video frames. Due to the visual persistence effect of human eyes, in a case that the video frame sequence is played at a specific speed, what we see is a video with consecutive actions. Due to the extremely high similarity between consecutive video frames, there is a great amount of redundant information inside each video frame and between consecutive video frames. Therefore, before storing or transmitting a video, the video usually needs to be encoded by a video encoding technology, to remove redundant information of the video in dimensions of space, time, and the like, so as to save the storage space and improve the video transmission efficiency.

[0036] The video encoding technology may also be referred to as a video compression technology, and refers to a technology of compressing and encoding each video frame in a video according to an encoding mode. Further, in a process of encoding a video, each video frame in the video needs to be recursively divided into data blocks of multiple sizes, and then each divided data block is then inputted to an encoder for encoding. The following describes a recursive division process involved in the embodiments of the present disclosure in combination with FIG. 1a. FIG. 1a is a schematic diagram of a recursive division process of a video frame according to an embodiment of the present disclosure. As shown in FIG. 1a, a video frame 10 is composed of multiple data blocks 101 of a first size. The method for recursively dividing the video frame 10 may include: (1) the data blocks 101 of the first size are directly inputted to an encoder for encoding, that is, the data blocks 101 of the first size are not divided and the data blocks of the first size are directly inputted to the encoder for encoding; (2) a data block 101 of the first size is divided into four data blocks 102 of a second size, the data blocks 102 of the second size are not further divided, and the data blocks 102 of the second size are inputted to the encoder for encoding; (3) a data block 102 of the second

size is further divided, four data blocks 103 of a same size, namely, a third size, are obtained after division, the data blocks 103 of the third size are not further divided, and the data blocks 103 of the third size are inputted to the encoder for encoding. By analogy, a data block 103 of the third size may be further divided, so that the video frame 10 is recursively divided into data blocks of various sizes, and a division result of the video frame 10 may be referred to FIG. 1b. FIG. 1b is a schematic diagram of a division result of a video frame according to an embodiment of the present disclosure. As shown in FIG. 1b, the divided video frame 10 is composed of data blocks of three sizes, namely, the data blocks 101 of the first size, the data blocks 102 of the second size, and the data blocks 103 of the third size.

[0037] In an actual video encoding process, division strategies of multiple data blocks included in a video frame are different, encoding modes for encoding the video frame are also different, and the encoding speed for the video frame also varies. The embodiments of the present disclosure provide a video processing solution, which provides two different data block division strategies, respectively, a top-down data block division strategy and a bottom-up data block division strategy. Both the two data block division strategies can determine an optimal division method for each data block in a video frame. The optimal division method may be understood as that the encoding quality is good and the encoding bit rate is low during encoding of data blocks obtained after dividing a data block according to the data block division method. The encoding bit rate may refer to an amount of data transmitted per unit time (for example, 1 second) in a case that a video data stream obtained after encoding is transmitted. A greater amount of data transmitted per unit time indicates a higher encoding bit rate. The following separately describes the two data block division strategies in detail.

[0038] (1) Top-down data block division strategy

[0039] In a process of encoding a video, each data block in a video frame is recursively divided from top to bottom. Further, the top-down recursive division may refer to a process of dividing from a maximum size of data blocks in a video frame to a minimum size of the data blocks in the video frame, until an optimal division method for the data blocks is found. For example, a maximum size of data blocks in a video frame is 64 px (pixel)×64 px and a minimum size of the data blocks in the video frame is 4 px×4 px. The maximum size of the data blocks being 64 px×64 px and the minimum size of the data blocks being 4 px×4 px are used for example purposes only and do not constitute a limitation to the embodiments of the present disclosure. For example, the maximum size of the data blocks may be 128 px×128 px and the minimum size of the data blocks may be 8 px×8 px, and the like. One 64 px×64 px data block may be divided into four 32 px×32 px

data blocks. A Rate-Distortion Cost (RDCost) of encoding each of the four 32 px×32 px data blocks is calculated, a sum of the RDCosts of the four 32 px×32 px data blocks is calculated (that is, a first sum RDCost of the four 32 px×32 px data blocks is calculated), and a RDCost of directly encoding the 64 px×64 px data block is calculated. In a case that the first sum RDCost is greater than the
5 RDCost of the 64 px×64 px data block, it is determined that a division size of the data block is 64 px×64 px, that is, there is no need to divide the 64 px×64 px data block; and in a case that the first sum RDCost is less than or equal to the RDCost of the 64 px× 64 px data block, it is determined that a division size of the data blocks is 32 px×32 px. Subsequently, a 32 px×32 px data block may further be divided into four 16 px×16 px data blocks for encoding, and by analogy, until an optimal
10 division method for the data blocks is found. By the top-down data block division strategy, usually the optimal division method can be determined accurately. For a data block having a low scene complexity, the optimal division size of the data block is large, and therefore for the data block having a low scene complexity, the optimal division method for the data block can be quickly determined by the top-down data block division strategy. For a data block having a high scene
15 complexity, the optimal division size of the data block is small, and therefore for the data block having a high scene complexity, the process of determining the optimal division method for the data block by the top-down data block division strategy requires a great amount of time and cost, further affecting the encoding speed of the data block.

[0040] The optimal division size of the data block may be understood as that the encoding
20 quality is good and the encoding bit rate is low during encoding of data blocks that are obtained after division in a case that the size of the data blocks obtained after division is the size. Generally, the optimal division size corresponding to a data block having a low scene complexity is large, and the optimal division size corresponding to a data block having a high scene complexity is small. The scene complexity of a data block may be measured by a method such as using spatial
25 information (SI) or temporal information. The spatial information may be used for representing an amount of spatial details of a data block, and a greater quantity of elements included in a data block indicates a higher value of the spatial information of the data block and indicates a higher scene complexity of the data block. For example, a data block A includes five elements (a cat, a dog, a tree, a flower, and a sun), a data block B includes two elements (a cat and a dog), the data block A
30 includes more elements than the data block B, and therefore, a value of spatial information of the data block A is higher than that of spatial information of the data block B and a scene complexity of the data block A is higher than that of the data block B. The temporal information may be used for representing an amount of temporal variation of a data block, and a higher degree of motion of a target data block in a target video frame which is currently processed relative to a reference data

block in a reference video frame of the target video frame indicates a greater value of the temporal information of the target data block and indicates a higher scene complexity of the target data block. The reference video frame is a video frame in the video frame sequence whose encoding order is before the target video frame, and a position of the target data block in the target video frame is the same as that of the reference data block in the reference video frame. For example, the target data block includes one element (for example, an automobile element), the reference data block also includes the element, and a greater displacement of the automobile element in the target data block relative to the automobile element in the reference data block indicates a higher degree of motion of the target data block in the target video frame relative to the reference data block in the reference video frame, indicates a greater value of the temporal information of the target data block, and indicates a higher scene complexity of the target data block.

[0041] (2) Bottom-up video frame division strategy

[0042] In a process of encoding a video, each data block in a video frame is recursively divided from bottom to top. Further, the bottom-up recursive division may refer to a process of dividing from a minimum size of data blocks in a video frame to a maximum size of the data blocks in the video frame, until an optimal division method for the data blocks is found. For example, a minimum size of data blocks in a video frame is 4 px×4 px and a maximum size of the data blocks in the video frame is 64 px×64 px. Each of the four 4 px×4 px data blocks is encoded separately, then an 8 px×8 px data block composed of four 4 px×4 px data blocks is encoded, a RDCost for encoding each of the four 4 px×4 px data blocks is calculated separately, and a sum of the RDCosts of the four 4 px×4 px data blocks is calculated (that is, a second sum of RDCosts of the four 4 px×4 px data blocks is calculated); and a RDCost for encoding an 8 px×8 px data block is calculated. In a case that the second sum RDCost is less than or equal to the RDCost of the 8 px×8 px data block, it is determined that a division size of the data block is 4 px×4 px; and in a case that the second sum RDCost is greater than the RDCost of the 8 px×8 px data block, it is determined that a division size of the data block is 8 px×8 px. Subsequently, a 16 px×16 px data block composed of four 8 px×8 px data blocks may be continuously encoded, and by analogy, until an optimal division method for the data block is found. By the bottom-up video frame division strategy, usually the optimal division method can be determined accurately. For a data block having a high scene complexity, the optimal division size of the data block is small, and therefore for the data block having a high scene complexity, the optimal division method for the data block can be quickly determined by the bottom-up video frame division strategy. For a data block having a low scene complexity, the optimal division size of the data block is large, and therefore for the data block having a low scene complexity, the process of determining the optimal division method for the data block through the

bottom-up video frame division strategy requires a great amount of time and cost, further affecting the encoding speed of the data block.

[0043] Hence, by the top-down data block division strategy and the bottom-up video frame division strategy, usually the optimal division method for each data block in a video frame can be determined accurately. The top-down data block division strategy is more suitable for a data block having a low scene complexity and the bottom-up video frame division strategy is more suitable for a data block having a high scene complexity.

[0044] On this basis, the embodiments of the present disclosure provide a further video processing solution, by which a scene complexity of a to-be-encoded target data block in a to-be-encoded video frame is obtained; the target data block is divided into multiple sub-data blocks and a scene complexity of each sub-data block is obtained; and according to the scene complexity of the target data block and the scene complexity of each of the multiple sub-data blocks obtained after division, an optimal division method for the target data block is determined, and further an encoding mode for the target data block is determined, so that the target data block may be encoded according to the determined encoding mode. The scene complexity in the embodiments of the present disclosure may be obtained by any scene complexity analysis method in the related art, for example, a Sum of Squared Error (SSE) method, a Sum of Absolute Difference (SAD) method, and the like. The optimal division method in the embodiments of the present disclosure refers to a data block division method which can improve the encoding speed while achieving a balance between the encoding quality and the encoding bit rate. By the video processing solution, an encoding mode suitable for a scene complexity of a target data block can be formulated according to the scene complexity of the target data block, effectively improving the encoding speed of the target data block, thereby improving the video encoding speed; in addition, the video processing solution is more universal and suitable for any video encoding scenario, and by the video processing solution, not only an encoding mode suitable for a data block having a low scene complexity can be determined, but also an encoding mode suitable for a data block having a high scene complexity can be determined.

[0045] Based on the foregoing description, the embodiments of the present disclosure provide an encoder, which may be configured to implement the foregoing video processing solution, and a video encoder may be an Alliance for Open Media Video 1 (AV1) standard encoder. FIG. 2 is a schematic structural diagram of an encoder according to an embodiment of the present disclosure. As shown in FIG. 2, an encoder 20 includes a scene complexity analysis module 201, an adaptive division decision module 202, and a division early termination module 203.

[0046] The scene complexity analysis module 201 may be configured to obtain a scene complexity of a to-be-encoded target data block in a to-be-encoded video frame, and obtain data block indicator information of the target data block based on the scene complexity of the target data block. The scene complexity analysis module 201 may further be configured to divide the target data block into N sub-data blocks, obtain scene complexities of the N sub-data blocks, and obtain sub-block indicator information based on the scene complexities of the N sub-data blocks, N being an integer greater than or equal to 2. In an actual encoding scenario, a value of N is generally 4, which is not limited in this embodiment of the present disclosure. Dividing the target data block into N sub-data blocks may refer to dividing the target data block into four sub-data blocks of a same size.

[0047] The adaptive division decision module 202 may be configured to determine an encoding mode for the target data block according to the data block indicator information and the sub-block indicator information. The encoding mode may include a first mode or a second mode, that is, the adaptive division decision module 202 may determine that the encoding mode for the target data block is the first mode or the second mode according to the data block indicator information and the sub-block indicator information. The first mode is an encoding mode in which a data block is divided into N sub-data blocks and each of the N sub-data blocks is encoded separately; and the second mode is an encoding mode in which a target data block is not divided and the target data block is directly encoded. In a case that the adaptive division decision module 202 cannot determine the encoding mode for the target data block according to the data block indicator information and the sub-block indicator information, the adaptive division decision module 202 may also be configured to obtain correlated block indicator information of M data blocks correlated with the target data block, and determine an attempt encoding order for the target data block according to the correlated block indicator information, M being a positive integer. In an actual encoding scenario, a value of M is generally 3, which is not limited in this embodiment of the present disclosure. FIG. 3 is a schematic diagram of a data block correlated with the target data block according to an embodiment of the present disclosure. As shown in FIG. 3, the M data blocks correlated with the target data block may be a data block 302 at a left position of a target data block 301, a data block 303 at a top position of the target data block 301, and a data block 304 at a left upper position of the target data block 301. The attempt encoding order may include a first attempt encoding order or a second attempt encoding order, that is, the adaptive division decision module 202 may determine that the attempt encoding order for the target data block is the first attempt encoding order or the second attempt encoding order according to the correlated block indicator information. The first attempt encoding order refers to an encoding order of preferably attempting

to encode the target data block according to the first mode and then attempting to encode the target data block according to the second mode; and the second attempt encoding order refers to an encoding order of preferably attempting to encode the target data block according to the second mode and then attempting to encode the target data block according to the first mode.

5 **[0048]** The division early termination module 203 is configured to set, in a case that the adaptive division decision module 202 cannot determine the encoding mode for the target data block, a division termination condition and determine the encoding mode for the target data block according to the set division termination condition. Further, in a case that the adaptive division decision module 202 determines that the attempt encoding order for the target data block is the first attempt encoding order, the division early termination module 203 obtains encoding information of N sub-data blocks obtained by encoding the target data block according to the first mode, and in a case that the encoding information of the N sub-data blocks meets a first division termination condition (that is, a fifth condition below), the division early termination module 203 determines that the encoding mode for the target data block is the first mode. In a case that the adaptive division decision module 202 determines that the attempt encoding order for the target data block is the second attempt encoding order, the division early termination module 203 obtains encoding information of the target data block obtained by encoding the target data block according to the second mode, and in a case that the encoding information of the target data block meets a second division termination condition (that is, a sixth condition below), the division early termination module 203 determines that the encoding mode for the target data block is the second mode.

15 **[0049]** The encoder 20 shown in FIG. 2 can formulate an encoding mode suitable for a scene complexity of a target data block according to the scene complexity of the target data block, effectively improving the encoding speed of the target data block, thereby improving the video encoding speed. In addition, the division early termination module 203 in the encoder 20 can determine the encoding mode for the target data block in a case that the encoding information meets a division termination condition, and terminate further division of the target data block, so that the encoding speed of the target data block is further improved while achieving a balance between the encoding speed of the target data block and the encoding bit rate, and that the video encoding efficiency is further improved. The video processing solution provided in the embodiments of the present disclosure and the specific application scenario of the encoder 20 provided in the embodiments shown in FIG. 2 are described below with reference to the video processing system shown in FIG. 4.

[0050] FIG. 4 is a schematic architectural diagram of a video processing system according to an embodiment of the present disclosure. As shown in FIG. 4, the video processing system includes P terminals (for example, a first terminal 401, a second terminal 402, and the like) and a server 403, P being an integer greater than 1. Any of the P terminals may be a device having a camera function such as a smartphone, a tablet computer, a laptop, a desktop computer, a smartspeaker, a smartwatch, and a smart wearable device, which is not limited herein. Any of the P terminals may support installation and operation of various applications, which may include, but are not limited to, social applications (for example, instant messaging applications, audio session applications, video session applications, and the like), audio and video applications (for example, audio and video on demand applications, audio and video players, and the like), gaming applications, and the like. The server may be an independent physical server, or may be a server cluster composed of multiple physical servers or a distributed system, and may further be a cloud server that provides cloud computing services. This is not limited in the embodiments of the present disclosure. The P terminals and the server 403 may be directly or indirectly connected in a wired or wireless communication manner. This is not limited in the present disclosure. The following describes the video processing solution provided in the embodiments of the present disclosure by an example of a video session scenario.

[0051] (1) Encoders 20 are deployed in the P terminals, respectively, and the P terminals perform video processing.

[0052] P users participate in a video session by using video session applications run on the P terminals. For example, a user 1 participates in the video session by using the video session application run on the first terminal 401, a user 2 participates in the video session by using the video session application run on the second terminal 402, and so on. The server 403 is configured to transmit a target video generated by the P terminals during the video session. The target video may include a session video generated by the P terminals during the video session. The first terminal 401 is any of the P terminals. A description is provided by an example in which the first terminal 401 processes the target video, and processing of the target video by the other terminals among the P terminals except the first terminal 401 are the same as the processing of the target video by the first terminal 401.

[0053] The target video is composed of multiple consecutive video frames, and each video frame included in the target video includes multiple to-be-encoded data blocks. The encoder 20 is deployed in the first terminal 401. The first terminal 401 calls the encoder 20 to obtain a scene complexity of each data block in each video frame included in the target video, determine an

encoding mode for each data block, and encode each data block according to the determined encoding mode, to finally obtain an encoded target video. The first terminal 401 may transmit the encoded target video to the server 403, and the server 403 transmits the encoded target video to the other terminals which participate in the video session; or, the first terminal 401 may directly
5 transmit the encoded target video to the other terminals which participate in the video session; and the other terminals receive the encoded target video, and parse and play the target video, to implement the video session in which the P terminals participate.

[0054] (2) The encoder 20 is deployed in the server 403, and the server 403 performs video processing.

10 **[0055]** P users participate in a video session by using video session applications run on the P terminals. For example, a user 1 participates in the video session by using the video session application run on the first terminal 401, a user 2 participates in the video session by using the video session application run on the second terminal 402, and so on. The server 403 is configured to process and transmit a target video generated by the P terminals during the video session. The target
15 video may include a session video generated by the P terminals during the video session.

[0056] The target video may be a session video generated by the first terminal 401 in a process of participating in the video session, and the first terminal 401 transmits the target video to the server 403. The target video is composed of multiple consecutive video frames, and each video frame included in the target video includes multiple to-be-encoded data blocks. The encoder 20 is
20 deployed in the server 403. The server 403 calls the encoder 20 to analyze a scene complexity of each data block in each video frame included in the target video, determine an encoding mode for each data block, and encode each data block according to the determined encoding mode, to finally obtain an encoded target video. The server 403 transmits the encoded target video to the other terminals which participate in the video session; and the other terminals receive the encoded target
25 video, and parse and play the target video, to implement the video session in which the P terminals participate.

[0057] In the embodiments of the present disclosure, each terminal or the server which participates in the video session may call the encoder to encode the target video generated during the video session, and the encoding mode for each data block included in the target video by each
30 terminal or the server is determined according to the scene complexity of each data block, which effectively improves the encoding speed of the data block, improves the encoding speed of the target video, and greatly accelerates the encoding speed of the target video while ensuring the quality of the target video, thereby improving the smoothness of the target video, improving the

session quality of the video session to an extent, optimizing the video session effect, and improving the user experience.

[0058] FIG. 5 is a schematic diagram of a video processing method according to an embodiment of the present disclosure. The video processing method may be performed by a smart device. The smart device may be a user terminal or a server. The user terminal may be a device having a camera function such as a smartphone, a tablet computer, or a smart wearable device. The smart device may be, for example, any terminal or a server in the video processing system shown in FIG. 4. The video processing method includes the following steps S501 to step S505:

[0059] S501: Obtain a target video frame from a to-be-encoded video and determine a to-be-encoded target data block from the target video frame. The to-be-encoded video is composed of multiple consecutive video frames, the target video frame is obtained from the to-be-encoded video, and the target video frame is any video frame in the to-be-encoded video. The target video frame includes multiple data blocks, there may be encoded data blocks and to-be-encoded data blocks in the multiple data blocks, the to-be-encoded target data block is determined from the target video frame, and the target data block is any to-be-encoded data block in the target video frame.

[0060] S502: Obtain a scene complexity of the target data block, and obtain data block indicator information of the target data block based on the scene complexity of the target data block. The data block indicator information may include, but is not limited to, at least one of the following: a distortion estimation parameter of the target data block, a spatial information parameter of the target data block, and a temporal information parameter of the target data block. The distortion (Dist) estimation parameter is obtained by estimating the degree of distortion during intra-frame prediction encoding or inter-frame prediction encoding. The distortion estimation parameter of the target data block may be used for measuring the degree of distortion of a reconstructed target data block compared to a raw target data block. The raw target data block refers to a target data block that is not encoded, and the reconstructed target data block refers to an encoded target data block obtained after performing intra-frame prediction encoding or inter-frame prediction encoding on the target data block. The spatial information parameter of the target data block may include a value of the spatial information obtained through calculation of the target data block; and the temporal information parameter of the target data block may include a value of the temporal information obtained through calculation of the target data block.

[0061] S503: Divide the target data block into N sub-data blocks, obtain scene complexities of the N sub-data blocks, and obtain sub-block indicator information based on the scene complexities of the N sub-data blocks, N being an integer greater than or equal to 2.

[0062] The sub-block indicator information may include N pieces of sub-block indicator data, an i^{th} sub-data block is any sub-block among the N sub-data blocks, i^{th} sub-block indicator data is any sub-block indicator data among the N pieces of sub-block indicator data, an i^{th} sub-data block corresponds to the i^{th} sub-block indicator data, and the i^{th} sub-block indicator data is obtained
5 based on the scene complexity of the i^{th} sub-data block, $i \in [1, N]$.

[0063] The i^{th} sub-block indicator data may include, but is not limited to, at least one of the following: a distortion estimation parameter of the i^{th} sub-data block, a spatial information parameter of the i^{th} sub-data block, and a temporal information parameter of the i^{th} sub-data block. The distortion estimation parameter of the i^{th} sub-data block may be used for measuring the degree
10 of distortion of a reconstructed i^{th} sub-data block compared to a raw i^{th} sub-data block. The raw i^{th} sub-data block refers to an i^{th} sub-data block that is not encoded, and the reconstructed i^{th} sub-data block refers to an encoded i^{th} sub-data block after performing intra-prediction encoding or inter-prediction encoding on the i^{th} sub-data block. The spatial information parameter of the i^{th} sub-data block may include a value of the spatial information obtained through calculation of the i^{th}
15 sub-data block; the temporal information parameter of the i^{th} sub-data block may include a value of the temporal information obtained through calculation of the i^{th} sub-data block.

[0064] S504: Determine an encoding mode for the target data block according to the data block indicator information and the sub-block indicator information.

[0065] In an embodiment, the data block indicator information and the sub-block indicator information may be inputted to a joint statistical model, the joint statistical model performs calculation on the data block indicator information and the sub-block indicator information, to obtain an output value that is calculated by the joint statistical model based on the data block indicator information and the sub-block indicator information, and the encoding mode for the target data block is determined according to the output value of the joint statistical model. The joint
20 statistical model may be obtained by training the data block indicator information of the data block of the determined encoding mode, and the sub-block indicator information of the N sub-data blocks obtained by dividing the data block of the determined encoding mode. In this embodiment, the joint statistical model may perform weighted calculation on the data block indicator information and the sub-block indicator information, to obtain the output value. A weighting factor may be obtained by
25 training relevant information of the data block of the determined encoding mode and sub-blocks thereof.

[0066] In a case that the output value of the joint statistical model meets a first condition, that is, the output value is greater than a first division threshold, it is determined that the encoding

mode for the target data block is a first mode. The output value of the joint statistical model meeting the first condition indicates that a correlation between the N sub-data blocks obtained by dividing the target data block is weak, and it is more inclined to divide the target data block into N sub-data blocks and encode each sub-data block separately. In a case that the output value of the joint statistical model meets a second condition, that is, the output value is less than a second division threshold, it is determined that the encoding mode for the target data block is a second mode. The output value of the joint statistical model meeting the second condition indicates that a correlation between the N sub-data blocks obtained by dividing the target data block is strong, and it is more inclined not to divide the target data block, but directly encode the target data block. The first division threshold and the second division threshold may be obtained during training of the joint statistical model.

[0067] S505: Encode the target data block according to the determined encoding mode. In a case that the encoding mode is the first mode, encoding the target data block according to the determined encoding mode may include encoding the target data block according to the first mode, and further may include dividing the target data block into the N sub-data blocks and inputting each sub-data block to an encoder for encoding. In a case that the encoding mode is the second mode, the encoding of the target data block according to the determined encoding mode may include encoding the target data block according to the second mode, and further may include directly inputting the target data block to an encoder for encoding.

[0068] In this embodiment of the present disclosure, the data block indicator information and the sub-block indicator information obtained based on the scene complexity of the target data block is inputted to the joint statistical model for calculation, and the encoding mode for the target data block is determined according to the output value that is calculated by the joint statistical model. The video processing solution provided in this embodiment of the present disclosure is suitable for any video scene, and can adaptively adjust the encoding mode for the target data block according to a scene complexity of a to-be-encoded target data block in any video, so as to determine the encoding mode suitable for the scene complexity of the target data block, which effectively improves the encoding speed of the target data block, further improves the video encoding speed, and can obtain an optimal balance between the encoding speed of the video and the encoding bit rate of the video.

[0069] FIG. 6 is a schematic diagram of another video processing method according to an embodiment of the present disclosure. The video processing method may be performed by a smart device. The smart device may be a user terminal or a server. The user terminal may be a device

having a camera function such as a smartphone, a tablet computer, or a smart wearable device. The smart device may be, for example, any terminal or a server in the video processing system shown in FIG. 4. The video processing method includes the following steps S601 to step S608:

[0070] S601: Obtain a target video frame from a to-be-encoded video and determine a to-be-encoded target data block from the target video frame.

[0071] S602: Obtain a scene complexity of the target data block, to obtain data block indicator information of the target data block based on the scene complexity of the target data block.

[0072] S603: Divide the target data block into N sub-data blocks, obtain scene complexities of the N sub-data block, and obtain sub-block indicator information based on the scene complexities of the N sub-data block, N being an integer greater than or equal to 2.

[0073] A process of step S601 in this embodiment of the present disclosure is the same as a process of step S501 in the embodiments shown in FIG. 5, a process of step S602 is the same as a process of step S502 in the embodiments shown in FIG. 5, a process of step S603 is the same as a process of step S503 in the embodiments shown in FIG. 5, and the specific processes may be referred to the description of the embodiments shown in FIG. 5, which are not repeated herein.

[0074] S604: Input the data block indicator information and the sub-block indicator information to a joint statistical model.

[0075] S605: Obtain an output value that is calculated by the joint statistical model based on the data block indicator information and the sub-block indicator information.

[0076] S606: In a case that the output value meets a third condition, obtain scene complexities of M data blocks correlated with the target data block, and obtain correlated block indicator information of the M data blocks based on the scene complexities of the M data blocks, M being a positive integer.

[0077] S607: Determine an encoding mode for the target data block according to the correlated block indicator information.

[0078] In S606 to S607, in a case that the output value of the joint statistical model meets the third condition, that is, the output value is less than or equal to a first division threshold, and the output value is greater than or equal to a second division threshold, it indicates that a correlation between the N sub-data blocks obtained by dividing the target data block is between strong and weak, there is no significant tendency to divide the target data block into the N sub-data blocks and then encode each sub-data block separately, nor is there a significant tendency to encode the target

data block directly instead of dividing the target data block. In this case, the M data blocks correlated with the target data block are determined, scene complexities of the M data blocks are obtained, to obtain the correlated block indicator information of the M data blocks based on the scene complexities of the M data blocks, and the encoding mode for the target data block is determined according to the correlated block indicator information of the M data blocks, M being a positive integer. The correlated block indicator information of the M data blocks may include a first quantity of data blocks correlated with the target data block that are among the M data blocks and that are each divided into multiple sub-data blocks for encoding.

[0079] In an embodiment, in a case that the first quantity meets a fourth condition, that is, the first quantity is greater than or equal to the first quantity threshold, it indicates that the scene complexity of the M data blocks correlated with the target data block is high, first the target data block is divided into N to-be-analyzed sub-data blocks (that is, the N sub-data blocks), and the scene complexities of the N to-be-analyzed sub-data blocks are obtained, to obtain to-be-analyzed sub-data block indicator information based on the scene complexities of the N to-be-analyzed sub-data blocks; and the encoding mode for the target data block is further determined according to the to-be-analyzed sub-data block indicator information. The first quantity threshold may be set according to an empirical value, for example, the quantity of data blocks correlated with the target data block is 3, and the first quantity threshold may be set to 2. For example, in FIG. 3, the three data blocks correlated with the target data block 301 are the data block 302, the data block 303, and the data block 304, the data block 302 and the data block 303 are divided into multiple sub-data blocks for encoding, and the data block 304 is not divided but is directly encoded, that is, the first quantity of data blocks correlated with the target data block that are among the three data blocks correlated with the target data block and that are each divided into multiple sub-data blocks for encoding is 2. In a case that the first quantity meets the fourth condition, it can be determined that the scene complexities of the three data blocks correlated with the target data block 301 are high, and first the target data block 301 is divided into N to-be-analyzed sub-data blocks. The to-be-analyzed sub-data block indicator information may include a second quantity of to-be-analyzed sub-data blocks that are among the N to-be-analyzed sub-data blocks and that meet a further division condition. Further, the to-be-analyzed sub-data block indicator information may include a second quantity of output values, among the N output values calculated based on the data block indicator information of each to-be-analyzed sub-data block among the N to-be-analyzed sub-data blocks and the sub-block indication information of the N sub-data blocks obtained by dividing each to-be-analyzed data block by the joint statistical model, that meets a first condition.

[0080] In an embodiment, in a case that the second quantity meets the fifth condition, that is, the second quantity is greater than or equal to the second quantity threshold, it may be determined that the encoding mode for the target data block is the first mode. In a case that the second quantity does not meet the fifth condition, that is, the second quantity is less than the second quantity threshold, the target data block is attempted to be encoded directly and then according to the encoding information of the target data block obtained by directly encoding the target data block, and according to the encoding information of N sub-data blocks obtained by dividing the target data block into the N sub-data blocks and encoding each sub-data block separately, the encoding mode for the target data block is determined. The second quantity threshold may alternatively be set according to an empirical value. For example, in a case that the target data block is divided into four to-be-analyzed sub-data blocks, the second quantity threshold may be set to 3. In a case that all the four to-be-analyzed sub-data blocks meet the further division condition, that is, the second quantity of to-be-analyzed sub-data blocks that are among the four to-be-analyzed sub-data blocks and that meet the further division condition is 4, and the second quantity meets the fifth condition, it can be determined that the encoding mode for the target data block is the first mode.

[0081] In an embodiment, in a case that the first quantity does not meet the fourth condition, that is, the first quantity is less than the first quantity threshold, it indicates that the scene complexity of the M data blocks correlated with the target data block is low, first the data block is not to be divided and the target data block is directly encoded, the encoding information of the target data block obtained by encoding the target data block is obtained, and the encoding information of the target data block may include an encoding distortion parameter of the target data block. The distortion estimation parameter of the target data block may include an estimated value of the degree of distortion in the encoding process of the target data block, and the encoding distortion parameter of the target data block may be used for measuring an actual degree of distortion of the encoded target data block (that is, the reconstructed target data block) compared to the target data block before encoding (that is, the raw target data block). Further, an encoding parameter is calculated based on the encoding distortion parameter of the target data block and a quantization parameter (QP) for quantifying the target data block, and the calculation process of the encoding parameter can be referred to the following Formula 1:

$$\text{Code} = \text{Dist}/\text{QP}^2 \quad \text{Formula 1}$$

[0082] In the Formula 1, Code represents the encoding parameter, Dist represents the encoding distortion parameter of the target data block, and QP represents the quantization parameter.

[0083] In an embodiment, in a case that the encoding parameter meets a sixth condition, that is, the encoding parameter is less than a third division threshold, it may be determined that the encoding mode for the target data block is the second mode. In another embodiment, in a case that the encoding parameter does not meet the sixth condition, that is, the encoding parameter is greater than or equal to the third division threshold, the target data block is divided into N sub-data blocks, each sub-data block is encoded separately, and according to the encoding information of the target data block obtained by directly encoding the target data block, and according to the encoding information of the N sub-data blocks obtained by dividing the target data block into the N sub-data blocks and encoding each sub-data block separately, the encoding mode for the target data block is determined. The third division threshold may be obtained during training of the joint statistical model.

[0084] In the foregoing two embodiments, the encoding information of the target data block may also include a first encoding rate distortion loss parameter of the target data block; and the encoding information of the N sub-data blocks may include a second encoding rate distortion loss parameter of the N sub-data blocks. The second encoding rate distortion loss parameter is calculated based on a third encoding rate distortion loss parameter of each of the N sub-data blocks, for example, the second encoding rate distortion loss parameter may be the sum of third encoding rate distortion loss parameters of all the N sub-data blocks. The first encoding rate distortion loss parameter is calculated based on an encoding bit rate and an encoding distortion parameter of the encoding mode for directly encoding the target data block. For example, the first encoding rate distortion loss parameter may be a ratio of the encoding bit rate to the encoding distortion parameter. The first encoding rate distortion loss parameter may be used for measuring the encoding effect of directly encoding the target data block, and a smaller first encoding rate distortion loss parameter indicates a better encoding effect of the target data block. The good encoding effect of the target data block may include that the encoding bit rate of the target data block is low in a case that the degree of distortion of the encoded target data block is lower than that of the target data block before encoding. Similarly, the second encoding rate distortion loss parameter may be used for measuring the encoding effect of dividing the target data block into N sub-data blocks and encoding each sub-data block separately. A smaller second encoding rate distortion loss parameter indicates a better effect of encoding the target data block. In a case that the first encoding rate distortion loss parameter is greater than or equal to the second encoding rate

distortion loss parameter, it is determined that the encoding mode for the target data block is the first mode; and in a case that the first encoding rate distortion loss parameter is less than the second encoding rate distortion loss parameter, it is determined that the encoding mode for the target data block is the second mode.

5 **[0085]** S608: Encode the target data block according to the determined encoding mode.

[0086] A process of step S608 in this embodiment of the present disclosure is the same as a process of step S505 in this embodiment shown in FIG. 5, and a specific process may be referred to the description of the embodiments shown in FIG. 5, which are not repeated herein.

[0087] In the embodiments of the present disclosure, the scene complexity of the data block correlated with the target data block may also reflect the scene complexity of the target data block, and therefore, in the embodiments of the present disclosure, the encoding mode for the target data block is jointly determined according to the scene complexity of the target data block (that is, the output value of the joint statistical model) and the scene complexity of the data block correlated with the target data block (that is, the correlated block indication information). The video processing solution provided in this embodiment of the present disclosure has universality, and can adaptively adjust the encoding mode for the target data block according to a scene complexity of a to-be-encoded target data block in any video, so as to determine the encoding mode suitable for the scene complexity of the target data block, which effectively improves the encoding speed of the target data block, further improves the video encoding speed, and can effectively improve the video encoding speed without losing the video encoding efficiency and an optimal balance between the video encoding speed and the video encoding bit rate.

10
15
20

[0088] FIG. 7 is a schematic diagram of another video processing method according to an embodiment of the present disclosure. The video processing method may be performed by a smart device. The smart device may be a user terminal or a server. The user terminal may be a device having a camera function such as a smartphone, a tablet computer, or a smart wearable device. The smart device may be, for example, any terminal or a server in the video processing system shown in FIG. 4. The video processing method includes the following steps S701 to step S720:

25

[0089] S701: Obtain a target video frame from a to-be-encoded video and determine a to-be-encoded target data block from the target video frame.

[0090] S702: Obtain a scene complexity of the target data block, and obtain data block indicator information of the target data block based on the scene complexity of the target data block.

30

- [0091]** S703: Divide the target data block into N sub-data blocks, obtain scene complexities of the N sub-data blocks, and obtain sub-block indicator information of the N sub-data blocks based on the scene complexities of the N sub-data block.
- [0092]** S704: Input the data block indicator information and the sub-block indicator information to a joint statistical model for calculation, to obtain an output value of the joint statistical model.
- [0093]** S705: Determine whether the output value meets a first condition. In a case that the output value meets the first condition, step S706 is performed; and in a case that the output value does not meet the first condition, step S707 is performed.
- [0094]** S706: In a case that the output value meets the first condition, determine that the encoding mode for the target data block is a first mode. After step S706 is completed, step S720 is performed.
- [0095]** S707: In a case that the output value does not meet the first condition, determine whether the output value meets a second condition. In a case that the output value meets the second condition, step S708 is performed; and in a case that the output value does not meet the second condition, step S709 is performed.
- [0096]** S708: In a case that the output value meets the second condition, determine that the encoding mode for the target data block is a second mode. After step S708 is completed, step S720 is performed.
- [0097]** S709: In a case that the output value does not meet the second condition, determine whether the output value meets a third condition. In a case that the output value meets the third condition, step S710 is performed.
- [0098]** S710: In a case that the output value meets the third condition, determine a first quantity of data blocks correlated with the target data block that are among the M data blocks correlated with the target data block and that are each divided into multiple sub-data blocks for encoding.
- [0099]** S711: Determine whether the first quantity meets a fourth condition. In a case that the first quantity meets the fourth condition, step S712 is performed; and in a case that the first quantity does not meet the fourth condition, step S716 is performed.
- [0100]** S712: In a case that the first quantity meets the fourth condition, divide the target sub-data block into N sub-data blocks and then encode each sub-data block separately.

[0101] S713: Determine a second quantity of sub-data blocks that are among the N sub-data blocks and that meet a further division condition.

[0102] S714: Determine whether the second quantity meets a fifth condition. In a case that the second quantity meets the fifth condition, it is determined that the encoding mode for the target data block is the first mode, and step S720 is performed. In a case that the second quantity does not meet the fifth condition, step S715 is performed.

[0103] S715: Skip dividing the target data block, directly encode the target data block, and determine the encoding mode for the target data block according to the encoding information. Further, the encoding information of the target data block obtained by directly encoding the target data block is obtained, the encoding information of N sub-data blocks obtained by dividing the target data block into the N sub-data blocks and encoding each sub-data block separately is obtained, and the encoding mode for the target data block is determined according to the encoding information of the target data block and the encoding information of the N sub-data blocks. A specific process may be referred to the embodiments shown in FIG. 6. After step S715 is completed, step S720 is performed.

[0104] S716: In a case that the first quantity does not meet the fourth condition, skip dividing the target data block, directly encode the target data block, and determine an encoding distortion parameter of the target data block.

[0105] S717: Obtain an encoding parameter through calculation based on the encoding distortion parameter and a quantization parameter.

[0106] S718: Determine whether the encoding parameter meets a sixth condition. In a case that the encoding parameter meets the sixth condition, it is determined that the encoding mode for the target data block is the second mode, and step S720 is performed; and in a case that the encoding parameter does not meet the sixth condition, step S719 is performed.

[0107] S719: Divide the target sub-data block into N sub-data blocks, encode each sub-data block separately, and determine the encoding mode for the target data block according to the encoding information. Further, the encoding information of the target data block obtained by directly encoding the target data block is obtained, the encoding information of N sub-data blocks obtained by dividing the target data block into the N sub-data blocks and encoding each sub-data block separately is obtained, and the encoding mode for the target data block is determined according to the encoding information of the target data block and the encoding information of the

N sub-data blocks. A specific process may be referred to the embodiments shown in FIG. 6. After step S719 is completed, step S720 is performed.

[0108] S720: Encode the target data block according to the determined encoding mode. After the target data block is encoded according to the determined encoding mode, another to-be-encoded data block other than the target data block may be determined from the to-be-encoded video frame, the encoding mode for the to-be-encoded data block may be determined according to the scene complexity of the to-be-encoded data block, and the to-be-encoded data block may be encoded according to the determined encoding mode. The process for determining the encoding mode for the to-be-encoded data block according to the scene complexity of the to-be-encoded data block is the same as the process for determining the encoding mode for the target data block according to the scene complexity of the target data block.

[0109] In the embodiments of the present disclosure, the scene complexity of the data block correlated with the target data block may also reflect the scene complexity of the target data block to an extent, and therefore, in the embodiments of the present disclosure, the encoding mode for the target data block is jointly determined according to the scene complexity of the target data block (that is, the output value of the joint statistical model) and the scene complexity of the data block correlated with the target data block (that is, the block indication information). In the embodiments of the present disclosure, six determining conditions are set for performing a multi-perspective comprehensive analysis on the scene complexity of the target data block, so that the finally determined encoding mode for the target data block has a high degree of adaptation with the scene complexity of the target data block, which effectively improves the encoding speed of the target data block, further improves the video encoding speed, and can effectively improve the video encoding speed without losing the video encoding quality and video encoding efficiency while ensuring the definition and smoothness of the encoded video.

[0110] FIG. 8 is a schematic structural diagram of a video processing apparatus according to an embodiment of the present disclosure. A video processing apparatus 80 according to this embodiment of the present disclosure may be disposed in a smart device. The smart device may be a smart terminal or a server. The video processing apparatus 80 may be configured to perform the corresponding steps in the video processing method shown in FIG. 5, FIG. 6, or FIG. 7, and the video processing apparatus 80 includes the following units:

[0111] an obtaining unit 801, configured to obtain a target video frame from a to-be-encoded video and determine a to-be-encoded target data block from the target video frame; and

- [0112]** a processing unit 802, configured to obtain a scene complexity of the target data block, and obtain data block indicator information of the target data block based on the scene complexity of the target data block; divide the target data block into N sub-data blocks, obtain scene complexities of the N sub-data blocks, and obtain sub-block indicator information based on the scene complexities of the N sub-data blocks, N being an integer greater than or equal to 2; determine an encoding mode for the target data block according to the data block indicator information and the sub-block indicator information; and encode the target data block according to the determined encoding mode.
- [0113]** In an embodiment, the data block indicator information includes one or more of: a distortion estimation parameter of the target data block, a spatial information parameter of the target data block, and a temporal information parameter of the target data block.
- [0114]** The sub-block indicator information includes N pieces of sub-block indicator data, where i^{th} sub-block indicator data among the N pieces of sub-block indicator data includes one or more of: a distortion estimation parameter of an i^{th} sub-data block in the N sub-data blocks, a spatial information parameter of the i^{th} sub-data block, and a temporal information parameter of the i^{th} sub-data block, $i \in [1, N]$.
- [0115]** In an embodiment, the processing unit 802 is further configured to:
- [0116]** input the data block indicator information and the sub-block indicator information to a joint statistical model;
- [0117]** obtain an output value that is calculated by the joint statistical model based on the data block indicator information and the sub-block indicator information; and
- [0118]** determine the encoding mode for the target data block according to the output value.
- [0119]** In an embodiment, the processing unit 802 is further configured to:
- [0120]** in a case that the output value meets a first condition, determine that the encoding mode is a first mode; and
- [0121]** divide the target data block into the N sub-data blocks and input each sub-data block to an encoder for encoding, where
- [0122]** the output value meeting the first condition includes: the output value being greater than a first division threshold.
- [0123]** In an embodiment, the processing unit 802 is further configured to:

- [0124]** in a case that the output value meets a second condition, determine that the encoding mode is a second mode; and
- [0125]** input the target data block to an encoder for encoding, where
- [0126]** the output value meeting the second condition includes: the output value being less than a second division threshold.
- [0127]** In an embodiment, the processing unit 802 is further configured to:
- [0128]** in a case that the output value meets a third condition, obtain scene complexities of M data blocks correlated with the target data block, and obtain correlated block indicator information of the M data blocks based on the scene complexities of M data blocks, M being a positive integer;
- [0129]** determine the encoding mode for the target data block according to the correlated block indicator information, where
- [0130]** the output value meeting the third condition comprises: the output value being less than or equal to a first division threshold and the output value is greater than or equal to a second division threshold.
- [0131]** In an embodiment, the processing unit 802 is further configured to:
- [0132]** obtain the correlated block indicator information, the correlated block indicator information including: a first quantity of data blocks correlated with the target data block that are among the M data blocks and that are each divided into multiple sub-data blocks for encoding;
- [0133]** in a case that the first quantity meets a fourth condition, divide the target data block into N to-be-analyzed sub-data blocks;
- [0134]** obtain scene complexities of the N to-be-analyzed sub-data blocks, and determine to-be-analyzed sub-data block indicator information of the N to-be-analyzed sub-data blocks based on the scene complexities of the N to-be-analyzed sub-data blocks; and
- [0135]** determine the encoding mode for the target data block according to the to-be-analyzed sub-data block indicator information, where
- [0136]** the first quantity meeting the fourth condition includes: the first quantity being greater than or equal to a first quantity threshold.
- [0137]** In an embodiment, the processing unit 802 is further configured to:

[0138] obtain the to-be-analyzed sub-data block indicator information, the to-be-analyzed sub-data block indicator information including: a second quantity of to-be-analyzed sub-data blocks that are among the N to-be-analyzed sub-data blocks and that meet a further division condition;

[0139] in a case that the second quantity meets a fifth condition, determine that the encoding mode is a first mode; and

[0140] in a case that the second quantity does not meet the fifth condition, input the target data block to an encoder for encoding and determine the encoding mode for the target data block according to encoding information obtained through encoding for the target data block, where

[0141] the second quantity meeting the fifth condition includes: the second quantity being greater than or equal to a second quantity threshold; and the second quantity not meeting the fifth condition includes: the second quantity being less than the second quantity threshold.

[0142] In an embodiment, the processing unit 802 is further configured to:

[0143] in a case that the first quantity does not meet the fourth condition, input the target data block to an encoder for encoding; and

[0144] determine the encoding mode for the target data block according to encoding information obtained through encoding for the target data block, where

[0145] the first quantity not meeting the fourth condition includes: the first quantity being less than the first quantity threshold.

[0146] In an embodiment, the encoding information of the target data block includes an encoding distortion parameter of the target data block; and the processing unit 802 is further configured to:

[0147] obtain an encoding parameter through calculation based on the encoding distortion parameter of the target data block and a quantization parameter of the target data block;

[0148] in a case that the encoding parameter meets a sixth condition, determine that the encoding mode is a second mode;

[0149] in a case that the encoding parameter does not meet the sixth condition, divide the target data block into the N sub-data blocks and input each of the N sub-data blocks to the encoder for encoding; and

[0150] determine the encoding mode for the target data block according to the encoding information of the target data block and encoding information obtained through encoding for the N sub-data blocks, where

[0151] the encoding parameter meeting the sixth condition includes: the encoding parameter being less than a third division threshold; and the encoding parameter not meeting the sixth condition includes: the encoding parameter being greater than or equal to the third division threshold.

5 **[0152]** In an embodiment, the encoding information of the target data block also includes a first encoding rate distortion loss parameter of the target data block; and the encoding information of the N sub-data blocks includes a second encoding rate distortion loss parameter of the N sub-data blocks, the second encoding rate distortion loss parameter being calculated based on a third encoding rate distortion loss parameter of each of the N sub-data blocks; and the processing unit
10 802 is further configured to:

[0153] in a case that the first encoding rate distortion loss parameter is greater than or equal to the second encoding rate distortion loss parameter, determine that the encoding mode for the target data block is the first mode; and

[0154] in a case that the first encoding rate distortion loss parameter is less than the second
15 encoding rate distortion loss parameter, determine that the encoding mode for the target data block is the second mode.

[0155] In an embodiment of the present disclosure, the units in the video processing apparatus 80 shown in FIG. 8 may be separately or wholly combined into one or several other units, or one (or more) of the units herein may further be divided into multiple units of smaller functions.

20 In this way, same operations can be implemented, and implementation of the technical effects of the embodiments of the present disclosure is not affected. The foregoing units are divided based on logical functions. In an actual application, a function of one unit may also be implemented by multiple units, or functions of multiple units are implemented by one unit. In other embodiments of the present disclosure, the video processing apparatus 80 may also include other units. During an
25 actual application, the functions may also be cooperatively implemented by other units and may be cooperatively implemented by multiple units. According to another embodiment of the present disclosure, a computer program (including program code) that can perform the steps in the corresponding methods shown in FIG. 5, FIG. 6, or FIG. 7 may be run on a general computing device, such as a general-purpose computer, which includes processing elements and storage
30 elements such as a central processing unit (CPU), a random access memory (RAM), and a read-only memory (ROM), to construct the video processing apparatus 80 shown in FIG. 8, and implement the video processing method in the embodiments of the present disclosure. The computer program may be recorded on, for example, a computer-readable storage medium, and loaded in any terminal

(for example, the first terminal 401 or the second terminal 402, and the like) or the server 403 of the video processing system shown in FIG. 4 through the computer-readable storage medium, and run therein.

[0156] In the embodiments of the present disclosure, by performing the scene complexity analysis on the to-be-encoded target data block in the target video frame, to obtain an analysis result of the entire to-be-encoded target data block and an analysis result after dividing the to-be-encoded target data block into multiple sub-data blocks, and selecting the encoding mode based on these analysis results, an appropriate encoding mode can be determined accurately for the to-be-encoded target data block, which can effectively improve the encoding speed of the target data block and improve the video encoding efficiency.

[0157] Referring to FIG. 9, FIG. 9 is a schematic structural diagram of a smart device according to an embodiment of the present disclosure. The smart device at least includes a processor 901 and a memory 902. The processor 901 and the memory 902 may be connected by a bus or in another manner.

[0158] The processor 901 may be a central processing unit (Central Processing Unit, CPU). The processor 901 may further include a hardware chip. The foregoing hardware chip may be an application-specific integrated circuit (ASIC), a programmable logic device (PLD), or the like. The PLD may be a field-programmable gate array (FPGA), a generic array logic (GAL), or the like.

[0159] The memory 902 may include a volatile memory such as a random access memory (RAM); the memory 902 may also include a non-volatile memory such as a flash memory or a solid-state drive (SSD); and the memory 902 may alternatively include a combination of the foregoing types of memories.

[0160] The memory 902 is configured to store a computer program. The computer program includes program instructions. The processor 901 is configured to execute the program instructions. The processor 901 (or referred to as a central processing unit, CPU) is a computing core and control core of the smart device 90, which is adapted to implement one or more computer instructions, and specifically, adapted to load and execute one or more computer instructions to implement corresponding method processes or corresponding functions.

[0161] The smart device 90 may be any terminal (for example, the first terminal 401 or the second terminal 402, and the like) or the server 403 in the video processing system shown in FIG. 4. The memory 902 stores a computer program, which includes one or more computer instructions. One or more computer instructions are loaded and executed by the processor 901, to implement the

corresponding steps in the method embodiments shown in FIG. 5, FIG. 6 or FIG. 7. In a specific implementation, the computer instructions in the memory 902 are loaded by the processor 901 to perform the following steps:

[0162] obtaining a target video frame from a to-be-encoded video and determining a to-be-encoded target data block from the target video frame;

[0163] obtaining a scene complexity of the target data block, and obtain data block indicator information of the target data block based on the scene complexity of the target data block;

[0164] dividing the target data block into N sub-data blocks, obtaining scene complexities of the N sub-data blocks, and obtain sub-block indicator information based on the scene complexities of the N sub-data blocks, N being an integer greater than or equal to 2;

[0165] determining an encoding mode for the target data block according to the data block indicator information and the sub-block indicator information; and

[0166] encoding the target data block according to the determined encoding mode.

[0167] In an embodiment, the data block indicator information includes one or more of: a distortion estimation parameter of the target data block, a spatial information parameter of the target data block, and a temporal information parameter of the target data block.

[0168] The sub-block indicator information includes N pieces of sub-block indicator data, where i^{th} sub-block indicator data among the N pieces of sub-block indicator data includes one or more of: a distortion estimation parameter of an i^{th} sub-data block in the N sub-data blocks, a spatial information parameter of the i^{th} sub-data block, and a temporal information parameter of the i^{th} sub-data block, $i \in [1, N]$.

[0169] In an embodiment, the computer instructions in the memory 902 are loaded by the processor 901 to perform the following steps:

[0170] inputting the data block indicator information and the sub-block indicator information to a joint statistical model;

[0171] obtaining an output value calculated by the joint statistical model based on the data block indicator information and the sub-block indicator information; and

[0172] determining the encoding mode for the target data block according to the output value.

[0173] In an embodiment, the computer instructions in the memory 902 are loaded by the processor 901 to perform the following steps:

[0174] in a case that the output value meets a first condition, determining that the encoding mode is a first mode; and

[0175] dividing the target data block into the N sub-data blocks and inputting each sub-data block to an encoder for encoding, where

5 **[0176]** the output value meeting the first condition includes: the output value being greater than a first division threshold.

[0177] In an embodiment, the computer instructions in the memory 902 are loaded by the processor 901 to perform the following steps:

10 **[0178]** in a case that the output value meets a second condition, determining that the encoding mode is a second mode; and

[0179] inputting the target data block to an encoder for encoding, where

[0180] the output value meeting the second condition includes: the output value being less than a second division threshold.

15 **[0181]** In an embodiment, the computer instructions in the memory 902 are loaded by the processor 901 to perform the following steps:

[0182] in a case that the output value meets a third condition, obtaining scene complexities of M data blocks correlated with the target data block, and obtaining correlated block indicator information of the M data blocks based on the scene complexities of M data blocks, M being a positive integer;

20 **[0183]** determining the encoding mode for the target data block according to the correlated block indicator information, where

[0184] the output value meeting the third condition includes: the output value being less than or equal to a first division threshold and greater than or equal to a second division threshold.

25 **[0185]** In an embodiment, the computer instructions in the memory 902 are loaded by the processor 901 to perform the following steps:

[0186] obtaining the correlated block indicator information, the correlated block indicator information including: a first quantity of data blocks correlated with the target data block that are among the M data blocks and that are each divided into multiple sub-data blocks for encoding;

30 **[0187]** in a case that the first quantity meets a fourth condition, dividing the target data block into N to-be-analyzed sub-data blocks;

[0188] obtaining scene complexities of the N to-be-analyzed sub-data blocks, and determine to-be-analyzed sub-data block indicator information of the N to-be-analyzed sub-data blocks based on the scene complexities of the N to-be-analyzed sub-data blocks; and

[0189] determining the encoding mode for the target data block according to the to-be-analyzed sub-data block indicator information, where

[0190] the first quantity meeting the fourth condition includes: the first quantity being greater than or equal to a first quantity threshold.

[0191] In an embodiment, the computer instructions in the memory 902 are loaded by the processor 901 to perform the following steps:

[0192] obtaining the to-be-analyzed sub-data block indicator information, the to-be-analyzed sub-data block indicator information including: a second quantity of to-be-analyzed sub-data blocks that are among the N to-be-analyzed sub-data blocks and that meet a further division condition;

[0193] in a case that the second quantity meets a fifth condition, determining that the encoding mode is a first mode; and

[0194] in a case that the second quantity does not meet the fifth condition, inputting the target data block to an encoder for encoding and determining the encoding mode for the target data block according to encoding information obtained through encoding for the target data block, where

[0195] the second quantity meeting the fifth condition includes: the second quantity being greater than or equal to a second quantity threshold; and the second quantity not meeting the fifth condition includes: the second quantity being less than the second quantity threshold.

[0196] In an embodiment, the computer instructions in the memory 902 are loaded by the processor 901 to perform the following steps:

[0197] in a case that the first quantity does not meet the fourth condition, inputting the target data block to an encoder for encoding; and

[0198] determining the encoding mode for the target data block according to encoding information obtained through encoding for the target data block, where

[0199] the first quantity not meeting the fourth condition includes: the first quantity being less than the first quantity threshold.

[0200] In an embodiment, the encoding information of the target data block includes an encoding distortion parameter of the target data block; and in a case that computer instructions in the memory 902 are loaded by the processor 901, the following steps are performed:

[0201] obtaining an encoding parameter through calculation based on the encoding distortion parameter of the target data block and a quantization parameter of the target data block;

[0202] in a case that the encoding parameter meets a sixth condition, determining that the encoding mode is a second mode;

[0203] in a case that the encoding parameter does not meet the sixth condition, dividing the target data block into the N sub-data blocks and inputting each of the N sub-data blocks to the encoder for encoding; and

[0204] determining the encoding mode for the target data block according to the encoding information of the target data block and encoding information obtained through encoding for the N sub-data blocks, where

[0205] the encoding parameter meeting the sixth condition includes: the encoding parameter being less than a third division threshold; and the encoding parameter not meeting the sixth condition includes: the encoding parameter being greater than or equal to the third division threshold.

[0206] In an embodiment, the encoding information of the target data block also includes a first encoding rate distortion loss parameter of the target data block; and the encoding information of the N sub-data blocks includes a second encoding rate distortion loss parameter of the N sub-data blocks, the second encoding rate distortion loss parameter being calculated based on a third encoding rate distortion loss parameter of each of the N sub-data blocks; and in a case that computer instructions in the memory 902 are loaded by the processor 901, the following steps are performed:

[0207] in a case that the first encoding rate distortion loss parameter is greater than or equal to the second encoding rate distortion loss parameter, determining that the encoding mode for the target data block is the first mode; and

[0208] in a case that the first encoding rate distortion loss parameter is less than the second encoding rate distortion loss parameter, determining that the encoding mode for the target data block is the second mode.

[0209] In the embodiments of the present disclosure, by performing the scene complexity analysis on the to-be-encoded target data block in the target video frame, to obtain an analysis result

of the entire to-be-encoded target data block and an analysis result after dividing the to-be-encoded target data block into multiple sub-data blocks, and selecting the encoding mode based on these analysis results, an appropriate encoding mode can be determined accurately for the to-be-encoded target data block, which can effectively improve the encoding speed of the target data block and improve the video encoding efficiency.

[0210] According to an aspect of the present disclosure, a computer program product or a computer program is provided, the computer program product or the computer program including computer instructions, the computer instructions being stored in a computer-readable storage medium. A processor of a computer device reads the computer instructions from the computer-readable storage medium, and executes the computer instructions, so that the computer device performs the video processing method provided in the various optional implementations.

[0211] A person of ordinary skill in the art may understand that all or some of the steps of the methods in the foregoing embodiments may be implemented by a program instructing relevant hardware. The program may be stored in a computer-readable storage medium. The computer-readable storage medium may include: a flash disk, a read-only memory (ROM), a random access memory (RAM), a magnetic disk, or an optical disc.

[0212] The contents disclosed above are merely exemplary embodiments of the present disclosure, but not intended to limit the scope of the present disclosure. A person of ordinary skill in the art can understand all or a part of the procedures for implementing the foregoing embodiments, and any equivalent variation made according to the claims of the present disclosure shall still fall within the scope of the present disclosure.

CLAIMS

What is claimed is:

1. A video processing method, performed by a smart device, comprising:

obtaining a target video frame from a to-be-encoded video and determining a to-be-encoded
5 target data block from the target video frame;

obtaining a scene complexity of the target data block, and obtaining data block indicator
information of the target data block based on the scene complexity of the target data block;

dividing the target data block into N sub-data blocks, obtaining scene complexities of the N
sub-data blocks, and obtaining sub-block indicator information based on the scene complexities of
10 the N sub-data blocks, N being an integer greater than or equal to 2;

determining an encoding mode for the target data block according to the data block indicator
information and the sub-block indicator information; and

encoding the target data block according to the determined encoding mode.

2. The method according to claim 1, wherein the data block indicator information comprises
15 one or more of: a distortion estimation parameter of the target data block, a spatial information
parameter of the target data block, and a temporal information parameter of the target data block;
and

the sub-block indicator information comprises N pieces of sub-block indicator data, wherein i^{th}
sub-block indicator data among the N pieces of sub-block indicator data comprises one or more of:
20 a distortion estimation parameter of an i^{th} sub-data block among the N sub-data blocks, a spatial
information parameter of the i^{th} sub-data block, and a temporal information parameter of the i^{th}
sub-data block, $i \in [1, N]$.

3. The method according to claim 1, wherein the determining an encoding mode for the target
data block according to the data block indicator information and the sub-block indicator information
25 comprises:

inputting the data block indicator information and the sub-block indicator information to a
joint statistical model;

obtaining an output value that is calculated by the joint statistical model based on the data
block indicator information and the sub-block indicator information; and

30 determining the encoding mode for the target data block according to the output value.

4. The method according to claim 3, wherein in a case that the output value meets a first condition, the determining the encoding mode for the target data block according to the output value comprises:

determining that the encoding mode is a first mode; and

5 the encoding the target data block according to the determined encoding mode comprises:

dividing the target data block into the N sub-data blocks and inputting each of the N sub-data blocks to an encoder for encoding, wherein

the output value meeting the first condition comprises: the output value being greater than a first division threshold.

10 5. The method according to claim 3, wherein in a case that the output value meets a second condition, the determining the encoding mode for the target data block according to the output value comprises:

determining that the encoding mode is a second mode; and

the encoding the target data block according to the determined encoding mode comprises:

15 inputting the target data block to an encoder for encoding, wherein

the output value meeting the second condition comprises the output value being less than a second division threshold.

20 6. The method according to claim 3, wherein in a case that the output value meets a third condition, the determining the encoding mode for the target data block according to the output value comprises:

obtaining scene complexities of M data blocks correlated with the target data block, and obtaining correlated block indicator information of the M data blocks based on the scene complexities of the M data blocks, M being a positive integer; and

25 determining the encoding mode for the target data block according to the correlated block indicator information, wherein

the output value meeting the third condition comprises the output value being less than or equal to a first division threshold and greater than or equal to a second division threshold.

7. The method according to claim 6, wherein the determining the encoding mode for the target data block according to the correlated block indicator information comprises:

30 obtaining the correlated block indicator information, the correlated block indicator information

comprising: a first quantity of data blocks correlated with the target data block that are among the M data blocks and that are each divided into a plurality of sub-data blocks for encoding;

in a case that the first quantity meets a fourth condition, dividing the target data block into N to-be-analyzed sub-data blocks;

5 obtaining scene complexities of the N to-be-analyzed sub-data blocks, and determine to-be-analyzed sub-data block indicator information of the N to-be-analyzed sub-data blocks based on the scene complexities of the N to-be-analyzed sub-data blocks; and

determining the encoding mode for the target data block according to the to-be-analyzed sub-data block indicator information, wherein

10 the first quantity meeting the fourth condition comprises the first quantity being greater than or equal to a first quantity threshold.

8. The method according to claim 7, wherein the determining the encoding mode for the target data block according to the to-be-analyzed sub-data block indicator information comprises:

obtaining the to-be-analyzed sub-data block indicator information, the to-be-analyzed sub-data
15 block indicator information comprising: a second quantity of to-be-analyzed sub-data blocks that are among the N to-be-analyzed sub-data blocks and that meet a further division condition;

in a case that the second quantity meets a fifth condition, determining that the encoding mode is a first mode; and

in a case that the second quantity does not meet the fifth condition, inputting the target data
20 block to an encoder for encoding and determining the encoding mode for the target data block according to encoding information obtained through encoding for the target data block, wherein

the second quantity meeting the fifth condition comprises the second quantity being greater than or equal to a second quantity threshold; and the second quantity not meeting the fifth condition comprises: the second quantity being less than the second quantity threshold.

25 9. The method according to claim 7, wherein in a case that the first quantity does not meet the fourth condition, the determining the encoding mode for the target data block according to the correlated block indicator information further comprises:

inputting the target data block to an encoder for encoding; and

determining the encoding mode for the target data block according to encoding information
30 obtained through encoding for the target data block, wherein

the first quantity not meeting the fourth condition comprises: the first quantity being less than the first quantity threshold.

10. The method according to claim 8 or 9, wherein the encoding information of the target data block comprises an encoding distortion parameter of the target data block; and the determining the encoding mode for the target data block according to encoding information obtained through encoding for the target data block comprises:

obtaining an encoding parameter through calculation based on the encoding distortion parameter of the target data block and a quantization parameter for quantizing the target data block;

in a case that the encoding parameter meets a sixth condition, determining that the encoding mode is a second mode;

in a case that the encoding parameter does not meet the sixth condition, dividing the target data block into the N sub-data blocks and inputting each of the N sub-data blocks to the encoder for encoding; and

determining the encoding mode for the target data block according to the encoding information of the target data block and encoding information of the N sub-data blocks obtained through encoding, wherein

the encoding parameter meeting the sixth condition comprises: the encoding parameter being less than a third division threshold; and the encoding parameter not meeting the sixth condition comprises: the encoding parameter being greater than or equal to the third division threshold.

11. The method according to claim 10, wherein the encoding information of the target data block further comprises a first encoding rate distortion loss parameter of the target data block; the encoding information of the N sub-data blocks comprises a second encoding rate distortion loss parameter of the N sub-data blocks, the second encoding rate distortion loss parameter being calculated based on a third encoding rate distortion loss parameter of each of the N sub-data blocks; and the determining the encoding mode for the target data block according to the encoding information of the target data block and encoding information of the N sub-data blocks comprises:

in a case that the first encoding rate distortion loss parameter is greater than or equal to the second encoding rate distortion loss parameter, determining that the encoding mode for the target data block is the first mode; and

in a case that the first encoding rate distortion loss parameter is less than the second encoding rate distortion loss parameter, determining that the encoding mode for the target data block is the second mode.

12. A video processing apparatus, comprising:

an obtaining unit, configured to obtain a target video frame from a to-be-encoded video and determine a to-be-encoded target data block from the target video frame; and

5 a processing unit, configured to obtain a scene complexity of the target data block, and obtain data block indicator information of the target data block based on the scene complexity of the target data block; divide the target data block into N sub-data blocks, obtain scene complexities of the N sub-data blocks, and obtain sub-block indicator information based on the scene complexities of the N sub-data blocks, N being an integer greater than or equal to 2; determine an encoding mode for the target data block according to the data block indicator information and the sub-block indicator
10 information; and encode the target data block according to the determined encoding mode.

13. A smart device, comprising:

a processor, configured to implement a computer program; and

a memory, storing a computer program, the computer program, when executed by the processor, implementing the video processing method according to any one of claims 1 to 11.

15 14. A computer-readable storage medium, storing a computer program, the computer program, when executed by a processor, implementing the video processing method according to any one of claims 1 to 11.