



(51) International Patent Classification:

H04N 19/597 (2014.01) *H04N 19/182* (2014.01)
H04N 19/115 (2014.01) *H04N 19/167* (2014.01)
H04N 19/136 (2014.01)

(21) International Application Number:

PCT/US2018/028217

(22) International Filing Date:

19 April 2018 (19.04.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

201710294882.0 28 April 2017 (28.04.2017) CN

(71) Applicant: **MICROSOFT TECHNOLOGY LICENSING, LLC** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: **XU, Jizheng**; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US). **LI, Yiming**; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US).

(74) Agent: **MINHAS, Sandip S.** et al.; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,

(54) Title: IMAGE ENCODING

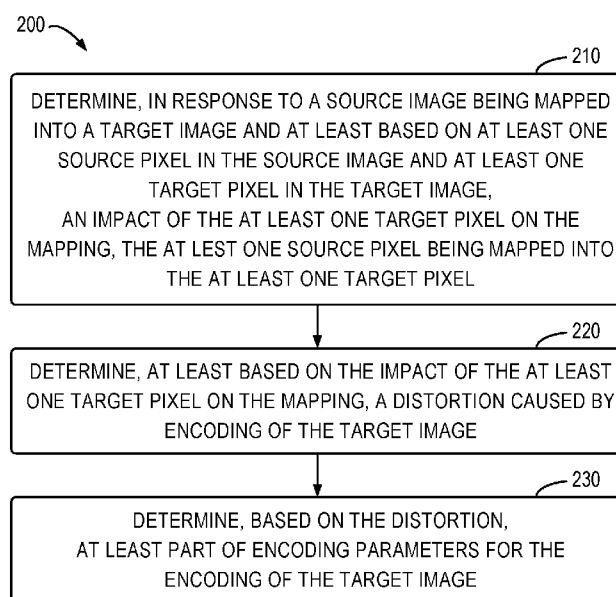


FIG 2

(57) Abstract: In the present disclosure, a method, a device and a computer program product for image and/or video encoding are proposed. According to the method, in response to a source image being mapped into a target image, an impact of at least one target pixel on the mapping is determined based on at least one source pixel in the source image and the at least one target pixel in the target image. The at least one source pixel is mapped into the at least one target pixel. Then, a distortion caused by encoding of the target image is determined based on the impact of the at least one target pixel on the mapping. After that, at least part of encoding parameters for the encoding of the target image is determined based on the distortion.

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*

IMAGE ENCODING

BACKGROUND

[0001] With the development of camera arrays, head-mounted displays and sensors to track head positions, the applications of omnidirectional video have emerged, aiming to provide users with visually immersive experiences. The omnidirectional video is generally composed of a plurality of spherical images. As the traditional image and/or video encoding system usually takes rectangular images as input, it is difficult for the omnidirectional video composed of a plurality of spherical images to be compressed. Therefore, to compress the omnidirectional video, it is usually necessary to map the spherical images into rectangular images. However, such mapping may introduce considerable distortion to the image. In addition, different parts of a moving object may have different geometric distortions, which may reduce the encoding performance significantly.

SUMMARY

[0002] According to the implementations of the subject matter described herein, there is provided an encoding method, an encoding device and a computer program product. Once a source image is mapped into a target image and a source pixel in the source image is mapped into a target pixel of the target image, the method determines an impact of the target pixel on the mapping based on the source and target pixels. The method also determines a distortion caused by encoding of the target image at least based on the determined impact. Furthermore, based on the determined distortion, the method determines at least part of encoding parameters for the encoding of the target image.

[0003] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] FIG. 1 shows a block diagram of a system 100 in which implementations of the subject matter as described herein can be implemented;

[0005] FIG. 2 shows a flowchart of an encoding method 200 according to implementations of the subject matter as described herein;

[0006] FIG. 3 shows a schematic diagram of mapping the spherical image into the rectangular image by equirectangular projection; and

[0007] FIG. 4 shows a block diagram of an example computing system/server 400 in which one or more implementations of the subject matter described herein can be implemented.

[0008] Throughout the drawings, the same or similar reference symbols are used to indicate the same or similar elements.

DETAILED DESCRIPTION

[0009] The subject matter described herein will now be discussed with reference to several example implementations. It should be understood these implementations are discussed only for the purpose of enabling those skilled persons in the art to better understand and thus implement the subject matter described herein, rather than suggesting any limitations on the scope of the subject matter.

[0010] As used herein, the term “includes” and its variants are to be read as open terms that mean “includes, but is not limited to.” The term “based on” is to be read as “based at least in part on.” The term “one implementation” and “an implementation” are to be read as “at least one implementation.” The term “another implementation” is to be read as “at least one other implementation.” The terms “first,” “second,” and the like may refer to different or same objects. Other definitions, explicit and implicit, may be included below. A definition of a term is consistent throughout the description unless the context clearly indicates otherwise.

[0011] FIG. 1 shows a block diagram of a system 100 in which implementations of the subject matter as described herein can be implemented. The system 100 may be used for processing an image and/or a video. Particularly, the system 100 may be used for processing an omnidirectional video. As shown in FIG.1, the system 100 in general may be divided into an encoding sub-system 110 and a decoding sub-system 120. It is to be understood that the structure and functionality of the system 100 are described only for the purpose of illustration without suggesting any limitations to the scope of the subject matter described herein. The subject matter described herein can be embodied with a different structure and/or functionality. Furthermore, some or all of the modules included in the system 100 can be implemented by software, hardware, firmware, and/or any suitable combination of the foregoing.

[0012] As shown in FIG. 1, the encoding sub-system 110 may include a capturing module 111, a stitching module 112, a mapping module 113 and an encoding module 114. In some implementations, these modules can be distributed on different devices. Alternatively, in some other implementations, these modules can be implemented on the

same device. In addition, some of these modules can be incorporated into other modules. For example, in some implementations, the mapping module 113 can be incorporated into the encoding module 114.

5 [0013] For example, the capturing module 111 may be implemented by a camera array which can cover all the directions of the light illuminating the camera array. The capturing module 111 may be used to capture video signals from different directions.

10 [0014] The stitching module 112 may receive image signals captured by the capturing module 111 from different directions and apply a stitching process to derive a source image. For example, the stitching module 112 may generate the source image from the image signals captured by the capturing module 111 from different directions. The source image may be a non-planar image. Particularly, the source image, for instance, may include a spherical image which records visual signals from all the directions observed from the center of the sphere. The stitching module 112 can also combine a plurality of spherical images as a spherical image sequence. The combined spherical image sequence is exactly an omnidirectional video signal.

15 [0015] The omnidirectional video signal represented by the spherical image sequence may treat video signals from every direction equally, which is beneficial to the subsequent video rendering. However, since the traditional image and/or video encoding system generally takes rectangular images as input, such omnidirectional video signal is difficult to be compressed. To compress and transmit the omnidirectional video signal, for example, a mapping module 113 may be utilized to map the source image in the omnidirectional video. For example, the mapping module 113 may map the source image into a target image. Particularly, the mapping module 113, for example, may map the spherical image in the spherical image sequence into a rectangular image.

20 [0016] The mapping module 113 may perform the above mapping in many ways. For example, in some implementations, to map the spherical image into a rectangular image, the mapping module 113 may perform the above mapping by means of equirectangular projection, which is widely adopted in head-mounted displays. As used herein, the term “equirectangular projection” may refer to a process by which longitudes of the sphere with a constant gap are mapped into vertical lines with a constant gap and latitude lines of the sphere with a constant gap are mapped into horizontal lines with a constant gap. For ease of description, implementations of the subject matter are described herein with equirectangular projection as an example. However, it should be understood that the subject matter described herein can also be embodied with other mapping methods. The

scope of the subject matter described herein is not limited in this aspect.

[0017] After the source image (such as the spherical image in the spherical image sequence) is mapped into the target image (such as the rectangular image), the encoding module 114 can be employed to compress the image and/or video signal into image and/or video bitstream. The compressed image and/or video bitstream, for example, can be utilized for transmission to different devices.

[0018] The decoding sub-system 120 may perform an inverse process of the above process performed by the encoding sub-system 110 to reconstruct the image and/or video signal from the received image and/or video bitstream. Particularly, when the received bitstream is a compressed omnidirectional video signal, the decoding sub-system 120 can reconstruct the omnidirectional video signal from the received video bitstream. As shown in FIG. 1, the decoding sub-system 120 may include a decoding module 121, an unmapping module 122 and a rendering module 123. In some implementations, these modules can be distributed on different devices. Alternatively, in some implementations, these modules can be implemented on the same device. Moreover, some of these modules can be incorporated into other modules. For example, in some implementations, the unmapping module 122 can be incorporated into the decoding module 121.

[0019] The decoding module 121 may decode the received image and/or video bitstream into the image and/or video. Then, the unmapping module 122 may perform an inverse process of the above process performed by the mapping module 113, and map the target image (such as the rectangular image) in the image and/or video back into the source image (such as the spherical image), thereby reconstructing the image and/or video signal. The rendering module 123 may display the reconstructed image and/or video signal to the corresponding display device (such as a head-mounted display).

[0020] As described above, during the processing of the image and/or video, the process for mapping the source image (such as the spherical image), which is probably non-planar, to the target image (such as the rectangular image) may introduce a considerable distortion to the image. Taking equirectangular projection as an example, a point at the north pole of the sphere may be mapped into the topest line of the rectangular image, while a point at the equator of the sphere may remain the same before and after mapping. Such a geometric distortion may become greater as a point becomes farther from the equator. Furthermore, the distortion may also impact the motion. For example, an object with a rigid motion may have a much more complicated motion after the mapping because different parts of the object may have different geometric distortions. As used herein, the term “rigid motion”

refers to a movement of an object in which the distance between any two points of the object remains constant before and after the movement. However, the traditional video encoding system is usually designed to be adapted for a linear motion (namely, the trajectory of the motion is linear) and/or a rigid motion. Therefore, such a geometric distortion can reduce the video encoding performance significantly.

[0021] In order to solve the problems above and one or more of other potential problems, according to the example implementations of the subject matter described herein, there is provided an encoding scheme, which can determine different impact of different points on the mapping process from the source image to the target image and can be adapted to any mapping methods (including but not limited to Equirectangular Projection) currently known or to be developed in the future. The scheme can adjust measurement for the distortion caused by encoding based on the determined impact, and determine at least part of encoding parameters for the encoding of the image based on the adjusted measurement for the distortion so as to improve the encoding performance. It is to be understood that the encoding scheme according to the implementations of the subject matter described herein can be used to encode an image and/or a video. Particularly, the encoding scheme can be used to encode an omnidirectional video to improve the encoding performance for the omnidirectional video.

[0022] FIG. 2 shows a flowchart of an encoding method 200 according to the implementations of the subject matter described herein. For example, in some implementations, method 200 can be implemented by the encoding sub-system 110 as shown in FIG. 1. It is to be understood that method 200 may further include additional actions not shown and/or omit the shown actions. The scope of the subject matter described herein is not limited in this aspect.

[0023] At block 210, the encoding sub-system 110 determines, in response to a source image being mapped into a target image and at least based on at least one source pixel in the source image and at least one target pixel in the target image, an impact of the at least one target pixel on the mapping. The at least one source pixel in the source image is mapped into the at least one target pixel in the target image.

[0024] As used herein, the term “source image” may include for example a planar image or a non-planar image of any type. Furthermore, the source image may be an image in a video. For example, the source image can be any of a plurality of spherical images composing of an omnidirectional video. As used herein, the term “target image”, for instance, may be an image of any type that the source image is mapped into. For instance,

when the source image is a spherical image, the target image may be a rectangular image. As used herein, the term “source pixel” may refer to a pixel in the source image while the term “target pixel” may refer to a pixel corresponding to the source pixel in the target image that the source image is mapped into. For the sake of illustration, in the following descriptions, a spherical image is taken as an example for the source image and a rectangular image is taken as an example for the target image. However, it is to be understood that this is only for the purpose of illustration, without suggesting any limitations to the scope of subject matter as described herein in any manner.

[0025] For example, FIG. 3 shows a schematic diagram of mapping the spherical image into the rectangular image by equirectangular projection. As shown in FIG. 3, a spherical image 310 may be mapped into a rectangular image 320. Particularly, a source pixel 311 in the spherical image 310 is mapped into a target pixel 321 in the rectangular image 320. Furthermore, it can be seen that a point at the north pole of the spherical image 310 may be mapped into the toppest line of the rectangular image 320 by equirectangular projection, while a relative distance between two points at the equator in the spherical image 310 remains the same before and after mapping. Actions of the method 200 will be described in further detail with reference to FIG. 3. For example, only for the purpose of illustration, the spherical image 310 is taken as an example for the source image and the rectangular image 320 is taken as an example for the target image in the following descriptions. However, it is to be understood that this is only for the purpose of illustration, without suggesting any limitations to the scope of subject matter as described herein in any manner.

[0026] As described above, although mapping the spherical image 310 into the rectangular image 320 may facilitate the encoding, such mapping may introduce a geometric distortion that significantly impacts the encoding performance, and result in different impact of different parts in the image on the geometric distortion.

[0027] Suppose that each of source pixels in the spherical image 310 is equally important, the respective target pixel in the rectangular image 320 may be of different significance. In other words, different target pixels in the rectangular image 320 may have different impact on the mapping.

[0028] In some implementations, the encoding sub-system 110 may determine the impact of the target pixel 321 on the mapping based on the number of pixels included in the source pixel 311 in the spherical image 310 (also referred as “first number”) and the number of pixels included in the target pixel 321 in the rectangular image 320 (also referred as “second

number”). For example, the impact may be represented as a weight associated with the target pixel 321. In some implementations, the weight may be quantified as, for example, a ratio between the first number and the second number. It is to be understood that this is only for the purpose of illustration, without suggesting any limitations to the scope of subject matter as described herein in any manner. In alternative implementations, any suitable calculation can be applied to the first and second numbers, and the calculated result can be used as a quantitative representation of the above impact.

[0029] For example, suppose the width of the rectangular image 320 derived from the spherical image 310 is of W pixels and the height thereof is of H pixels, a pixel (for example, target pixel 321) which is the j^{th} pixel of the rectangular image 320 in the horizontal direction and the i^{th} pixel in the vertical direction may be represented as $I[i, j]$, where $i \in [0, H-1]$ and $j \in [0, W-1]$. The weight associated with the target pixel $I[i, j]$ may be represented as $A[i, j]$. For example, a smaller $A[i, j]$ may indicate that the target pixel $I[i, j]$ corresponds to a smaller area of the spherical image (namely, less source pixels), and thus the target pixel $I[i, j]$ may have smaller impact on the distortion. Accordingly, a larger $A[i, j]$ may indicate that the target pixel $I[i, j]$ corresponds to a larger area of the spherical image (namely, more source pixels), and thus the target pixel $I[i, j]$ may have larger impact on the distortion.

[0030] In some implementations, for example, in the case of equirectangular projection, $A[i, j]$ may be determined based on the value of i as follows:

$$A[i, j] = \cos\left[\frac{(i+0.5)\pi}{H} - \frac{\pi}{2}\right] \quad (1)$$

[0031] Furthermore, each of source pixels in the spherical image 310 may have different importance. For instance, each of the source pixels in the spherical image 310 may have a degree of importance associated with the encoding. In some implementations, the degree of importance, for instance, may be represented as a weight reflecting the importance of the source pixel *per se*. For example, a higher weight may indicate that the source pixel has greater impact on the encoding, while a lower weight may indicate that the source pixel has smaller impact on the encoding. In this case, the encoding sub-system 110 may also determine the impact of the target pixel on the mapping based on the source pixel, the target pixel corresponding to the source pixel and the degree of importance associated with the source pixel (such as the weight reflecting the importance of the source pixel *per se*).

[0032] For example, the target pixel $I[i, j]$ (such as the target pixel 321) in the

rectangular image 320 may correspond to at least one source pixel (such as the source pixel 311) in the spherical image 310, and each of the at least one source pixel may have a respective weight reflecting its importance. In this case, in some implementations, the encoding sub-system 110 may determine the weight associated with the at least one source pixel based on a respective weight of each of the at least one source pixels (for example, by averaging or by other proper means). Suppose that the weight associated with the at least one source pixel is represented as $w[i,j]$ and the weight determined based on the number of pixels in the at least one source pixels and the number of pixels in the target pixel $I[i,j]$ is represented as $a[i,j]$ (for instance, according to equation (1)). In some implementations, the weight $A[i,j]$ associated with the target pixel $I[i,j]$ may be determined as, for instance, a product of $w[i,j]$ and $a[i,j]$.

[0033] In some implementations, for the purpose of encoding, the target image may be divided into a plurality of image blocks. If an image block is relatively small, it may be considered approximately that different target pixels in the image block have the same impact on the mapping. In other words, different target pixels in the image block may have the same weight. In this case, for example, the weight associated with one target pixel in the image block can be assigned to other pixels in the same image block to simplify some of the following-up processing.

[0034] In this manner, the implementations of the subject matter described herein can determine different impact of different pixels on the mapping during the mapping process from the source image to the target image in a quantitative manner. In addition, it can be seen that while determining the impact, the implementations of the subject matter described herein are not limited to a particular mapping method but can be adapted to any of mapping methods currently known or to be developed in the future.

[0035] Continuing to refer to FIG. 2, at block 220, the encoding sub-system 110 determines a distortion caused by encoding of the target image based on the impact of the at least one target pixel (such as the target pixel 321) on the mapping. In some implementations, the encoding sub-system 110 may determine the distortion by generating an error measuring the distortion and then estimate corresponding image and/or video encoding parameters based on the generated error in the subsequent actions.

[0036] In some implementations, the error for measuring the distortion caused by the encoding of the target image may include, for example, Sum of Squares for Error (SSE) between the rectangular image and the reconstructed image. As used herein, the term “reconstructed image” may refer to an image derived from at least in part applying an

inverse process of the encoding to the encoded target image. For instance, in image and/or video encoding, to provide a part of reference data (for example, to provide reference images for predicting encoding in the video encoding), the encoding module 114 may usually perform a reconstruction to the encoded image. The encoding sub-system 110 may, for example, obtain a reconstructed image of the target image from the encoding module 114 to calculate SSE between the target image and the reconstructed image.

[0037] In some implementations, the error for measuring the distortion caused by the encoding of the target image may further include Sum of Absolute Difference (SAD) or Sum of Absolute Transformation Difference (SATD) between the target image and the reconstructed image. For example, the encoding sub-system 110 may obtain the reconstructed image of the target image (such as from the encoding module 114) to calculate SAD or SATD.

[0038] Different errors for measuring the distortion caused by the encoding of the target image can be applied to different modules (such as movement estimation module, filtering module and/or so on) of the image and/or video encoder, for example, so as to estimate respective encoding parameters for different modules.

[0039] Suppose that the reconstructed pixel in the reconstructed image of the target image corresponding to the target pixel $I[i, j]$ is represented as $I'[i, j]$. In traditional solutions, SSE, SAD and SATD may be determined as follows, respectively:

$$SSE = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} (I[i, j] - I'[i, j])^2 \quad (2)$$

$$SAD = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} |I[i, j] - I'[i, j]| \quad (3)$$

$$SATD = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} |H \circ I[i, j] - H \circ I'[i, j]| \quad (4)$$

where $H \circ$ represents Hadamard transformation.

[0040] In the implementations of the subject matter described herein, taking into account different impact of different pixels in the target image on the mapping, the encoding sub-system 110 may determine the distortion based on the target image, the reconstructed image of the target image and different impact of different pixels in the target image on the mapping. For example, the encoding sub-system 110 may adjust distortion measurement of the traditional solutions based on different impact of different pixels in the target image on the mapping.

[0041] Specifically, in some implementations, the encoding sub-system 110 may generate, based on different impact of different pixels in the target image on the mapping,

Sum of Squares for Error (hereinafter referred to as “SSE”) between the target image and its reconstructed image, Sum of Absolute Difference (hereinafter referred to as “SAD”) and/or Sum of Absolute Transformation Difference (hereinafter referred to as “SATD”) for use in the estimation of image and/or video encoding parameters in the subsequent actions.

5 For example, SSE', SAD' and SATD' may be determined as follows, respectively:

$$\begin{aligned} SSE' &= \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} A[i,j](I[i,j] - I'[i,j])^2 \\ &= \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} (\sqrt{A[i,j]}I[i,j] - \sqrt{A[i,j]}I'[i,j])^2 \end{aligned} \quad (5)$$

$$SAD' = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} |\sqrt{A[i,j]}I[i,j] - \sqrt{A[i,j]}I'[i,j]| \quad (6)$$

$$SATD' = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} |H \circ (\sqrt{A[i,j]}I[i,j]) - H \circ (\sqrt{A[i,j]}I'[i,j])| \quad (7)$$

10 where $H \circ$ represents Hadamard transformation.

[0042] More generally, suppose that $D(I[i,j], I'[i,j])$ measures the distortion between the pixel $I[i,j]$ and its reconstructed pixel $I'[i,j]$ in traditional solution, the corresponding distortion $D'(I[i,j], I'[i,j])$ according to the implementations of the subject matter described herein may be represented as follows:

$$15 \quad D'(I[i,j], I'[i,j]) = D(\sqrt{A[i,j]}I[i,j], \sqrt{A[i,j]}I'[i,j]) \quad (8)$$

In this way, the implementations of the subject matter described herein can adjust the measurement of the distortion for the image and/or video encoding based on different impact of different pixels on the mapping.

[0043] At block 230, the encoding sub-system 110 determines at least part of encoding parameters for the encoding of the target image based on the distortion determined at block 220.

[0044] In image and/or video encoding, for example, for an image block in the target image to be encoded, there may exist a plurality of encoding parameters for the encoding of the image block. Examples of encoding parameters may include one or more of the following: a quantitation parameter (QP) corresponding to a respective quantitative step size and deciding the quantitative degree of the image and/or video encoding; an encoding mode for the image block (for example, intra-frame prediction or inter-frame prediction, as well as the prediction direction of intra-frame prediction and/or inter-frame prediction in video encoding); a manner for further dividing the image block into sub-image blocks; a type of transformation applied to the image block, such as discrete cosine transformation or

30

direction linear transformation; and the like.

[0045] In some implementations, a rate distortion optimization process may be applied by the encoding sub-system 110 to compare different values of encoding parameters, so as to determine at least part of encoding parameters with the optimal encoding performance.

5 In some implementations, the encoding sub-system 110 may compare different values of all of the encoding parameters to determine values of all of the encoding parameters with the optimal encoding performance. In some other implementations, the encoding sub-system 110 may also compare different values of part of the encoding parameters to determine values of part of encoding parameters with the optimal encoding performance. For
10 example, in some cases, one or more encoding parameters might have been determined. Therefore, the encoding sub-system may apply an optimization process to other encoding parameters other than the one or more encoding parameters.

[0046] As used herein, the term “rate distortion optimization process” refers to a process of determining encoding parameters with the optimal encoding performance based on the
15 rate distortion theory. The main targets of the rate distortion optimization process are: 1) to reduce distortion of the image and/or video at certain coding rate limits; and 2) to reduce encoding rate as much as possible with a certain degree of distortion being allowed.

[0047] In some implementations, the rate distortion optimization process may be transformed into a Lagrangian optimization process. Specifically, the Lagrangian
20 optimization process can be described as follows:

$$\text{minimize } \sum_{k=1}^N (D_k + \lambda R_k) \quad (9)$$

where N represents the number of image blocks in the target image to be encoded, D_k represents the encoding distortion corresponding to the k^{th} image block, λ represents a Lagrangian operator, and R_k represents the total number of bits for encoding the k^{th}
25 image block. Suppose that D is utilized to represent $\sum_{k=1}^N D_k$ and R is utilized to represent $\sum_{k=1}^N R_k$, the above equation (9) can be simplified as:

$$\text{minimize } (D + \lambda R) \quad (10)$$

according to equation (10), the Lagrangian optimization process can be interpreted as to determine encoding parameters such that the value of $(D + \lambda R)$ is minimum (namely, to
30 determine encoding parameters with the optimal encoding performance).

[0048] In some implementations, the encoding sub-system 110 may determine any of SSE' , SAD' and $SATD'$ according to equations (5)-(7) and use it as D in equation (10), so as to determine the encoding parameters with the optimal encoding performance.

More generally, in some implementations, the encoding sub-system 110 may also determine other distortion measurement based on equation (8) and use it as D in equation (10), thereby, so as to the encoding parameters with the optimal encoding performance.

5 [0049] In some implementations, as described above, if an image block is small, different pixels in the image block may be considered as having the same weight approximately. In this case, when SSE' is used as an error to measure the encoding distortion, the Lagrangian optimization process can be further simplified.

[0050] For example, suppose that the weight of the k^{th} image block is w_k , the distortion corresponding to the k^{th} image block determined based on the traditional solutions (for example, equation (2)) is SSE_k , and the distortion corresponding to k^{th} image block determined according to implementations of the subject matter described herein (for example, equation (5)) is SSE_k' . According to equations (2) and (5), it can be determined that SSE_k and SSE_k' may have a relationship as follows: $SSE_k' = w_k SSE_k$.

15 Thus, the above equation (9) can be transformed into:

$$\begin{aligned}
 & \text{minimize } \sum_{k=1}^N (SSE_k' + \lambda R_k) \\
 & = \text{minimize } \sum_{k=1}^N (w_k SSE_k + \lambda R_k) \\
 & = \text{minimize } \sum_{k=1}^N (SSE_k + \frac{\lambda}{w_k} R_k) \\
 & = \text{minimize } (SSE + \sum_{k=1}^N (\frac{\lambda}{w_k} R_k)) \tag{11}
 \end{aligned}$$

20 It can be seen from equation (11) that in some implementations of the subject matter described herein, the calculation of SSE may not need to be changed during the Lagrangian optimization process (namely, SSE may be calculated according to equation (2)), while only the Lagrangian operator may need to be changed from λ to $\frac{\lambda}{w_k}$, so as to determine the encoding parameters with the optimal encoding performance. In this way, the implementations of the subject matter described herein can determine, based on the adjusted distortion measurement, the optimal encoding parameters for the encoding of the image and/or video (such as the omnidirectional video) without making a big change to the current image and/or video encoder to improve the encoding performance.

[0051] In some implementations, the encoding parameter to be determined may include
 30 QP which corresponds to a respective quantitative step size and decides the quantitative

degree of the image and/or video encoding. In the image and/or video encoding, the Lagrangian operator λ and QP usually have the following relationship:

$$\lambda = 0.57 \times 2^{(QP-12)/3} \quad (12)$$

namely,

$$QP = 3 \log_2 \lambda + 14.43 \quad (13)$$

[0052] According to equation (11), for the k^{th} image block, its quantitation parameter QP' can be determined as follows:

$$QP' = 3 \log_2 \frac{\lambda}{w_k} + 14.43 = QP - 3 \log_2 w_k \quad (14)$$

It can be seen from equation (14) that in the implementations of the subject matter described herein, for the k^{th} image block, an offset amount of $-3 \log_2 w_k$ can be applied to QP to obtain the quantitation parameter QP' with the optimal encoding performance. In other words, the encoding sub-system 110 may determine, only based on the weight associated with the image block and the Lagrangian operator, the optimal quantitation parameter as the at least part of encoding parameters for the encoding of the target image, thereby further improving the encoding performance for the omnidirectional video.

[0053] Furthermore, in some implementations, the method 200 may further include actions not shown in FIG. 2. For example, before block 210, the encoding sub-system 110 (such as the mapping module 113) may map the source image into the target image. For example, the encoding sub-system 110 may map the spherical image 310 into the rectangular image 320 by means of equirectangular projection. After block 230, the encoding sub-system (such as the encoding module 114) may use the encoding parameters determined at block 230 to encode the target image. It is to be understood that this is merely for the purpose of simplification, without suggesting any limitations to the scope of subject matter as described herein in any manner.

[0054] It can be known from the above description that implementations of the subject matter described herein can determine different impact of different points on the mapping process from the source image to the target image and can be adapted to any mapping methods (including but not limited to equirectangular projection) currently known or to be developed in the future. The implementations of the subject matter described herein can adjust measurement for the distortion caused by the encoding based on the determined impact, and determine at least part of encoding parameters for the encoding of the image based on the adjusted measurement for the distortion so as to improve the encoding performance.

[0055] FIG. 4 shows a block diagram of an example computing system/server 400 in which one or more implementations of the subject matter described herein can be implemented. For example, in some implementations, the encoding sub-system 110 as shown in FIG. 1 can be implemented by the computing system/server 400. The computing system/server 400 as shown in FIG. 4 is only an example, which should not be construed as any limitation to the function and scope of use of the implementations of the subject matter described herein.

[0056] As shown in FIG. 4, the computing system/server 400 is in a form of a general-purpose computing device. Components of the computing system/server 400 may include, but are not limited to, one or more processors or processing units 400, a memory 420, one or more input devices 430, one or more output devices 440, storage 450, and one or more communication units 460. The processing unit 400 may be a real or a virtual processor and is capable of performing various processes in accordance with a program stored in the memory 420. In a multi-processing system, multiple processing units execute computer-executable instructions to increase processing power.

[0057] The computing system/server 400 typically includes a variety of machine readable medium. Such medium may be any available medium that is accessible by the computing system/ server 400, including volatile and non-volatile medium, removable and non-removable medium. The memory 420 may be volatile memory (e.g., registers, cache, a random-access memory (RAM)), non-volatile memory (e.g., a read only memory (ROM), an electrically erasable programmable read only memory (EEPROM), a flash memory), or some combination thereof. The storage 450 may be removable or non-removable, and may include machine readable medium such as flash drives, magnetic disks or any other medium which can be used to store information and which can be accessed within the computing system/server 400.

[0058] The computing system/server 400 may further include other removable/non-removable, volatile/non-volatile computing system storage medium. Although not shown in FIG. 4, a disk driver for reading from or writing to a removable, non-volatile disk (e.g., a “floppy disk”), and an optical disk driver for reading from or writing to a removable, non-volatile optical disk can be provided. In these cases, each driver can be connected to the bus by one or more data medium interfaces.

[0059] The memory 420 may include at least one program product having a set (e.g., at least one) of program modules that are configured to carry out the functions of various implementations described herein. For instance, when one or more modules in system 100

are implemented as software modules, they can be stored in the storage 420 and when accessed and operated by the processing unit 400, they can implement the function and/or method described herein, such as method 200.

5 [0060] The input unit(s) 430 may be one or more of various different input devices. For example, the input unit(s) 439 may include a user device such as a mouse, keyboard, trackball, etc. The communication unit(s) 460 enables communication over communication medium to another computing entity. Additionally, functionality of the components of the computing system/server 400 may be implemented in a single computing machine or in multiple computing machines that are able to communicate over
10 communication connections. Thus, the computing system/server 400 may operate in a networked environment using logical connections to one or more other servers, network personal computers (PCs), or another common network node. By way of example, and not limitation, communication media include wired or wireless networking techniques.

15 [0061] The computing system/server 400 may also communicate, as required, with one or more external devices (not shown) such as a storage device, a display device, and the like, one or more devices that enable a user to interact with the computing system/server 400, and/or any device (e.g., network card, a modem, etc.) that enables the computing system/server 400 to communicate with one or more other computing devices. Such communication may be performed via an input/ output (I/O) interface(s) (not shown).

20 [0062] Through the above depiction, it will be appreciated that implementations of the subject matter described herein can determine different impact of different points on the mapping process from the source image to the target image and can be adapted to any mapping methods (including but not limited to equirectangular projection) currently known or to be developed in the future. The implementations of the subject matter described
25 herein can adjust measurement for the distortion caused by the encoding based on the determined impact, and determine at least part of encoding parameters for the encoding of the image based on the adjusted measurement for the distortion so as to improve the encoding performance.

30 [0063] The functionally described herein can be performed, at least in part, by one or more hardware logic components. For example, and without limitation, illustrative types of hardware logic components that can be used include Field-Programmable Gate Arrays (FPGAs), Application-specific Integrated Circuits (ASICs), Application-specific Standard Products (ASSPs), System-on-a-chip systems (SOCs), Complex Programmable Logic Devices (CPLDs), and the like.

[0064] Program code for carrying out methods of the subject matter described herein may be written in any combination of one or more programming languages. These program codes may be provided to a processor or controller of a general purpose computer, special purpose computer, or other programmable data processing apparatus, such that the
5 program codes, when executed by the processor or controller, cause the functions/operations specified in the flowcharts and/or block diagrams to be implemented. The program code may execute entirely on a machine, partly on the machine, as a stand-alone software package, partly on the machine and partly on a remote machine or entirely on the remote machine or server.

10 [0065] In the context of this disclosure, a machine readable medium may be any tangible medium that may contain, or store a program for use by or in connection with an instruction execution system, apparatus, or device. The machine readable medium may be a machine readable signal medium or a machine readable storage medium. A machine readable medium may include but not limited to an electronic, magnetic, optical, electromagnetic,
15 infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples of the machine readable storage medium would include an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable
20 compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing.

[0066] Further, while operations are described in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable
25 results. In certain circumstances, multitasking and parallel processing may be advantageous. Likewise, while several specific implementation details are contained in the above discussions, these should not be construed as limitations on the scope of the subject matter described herein, but rather as descriptions of features that may be specific to particular implementations. Certain features that are described in the context of separate
30 implementations may also be implemented in combination in a single implementation. Conversely, various features that are described in the context of a single implementation may also be implemented in multiple implementations separately or in any suitable sub-combination.

[0067] Some implementations of the subject matter described herein are listed below.

[0068] In the first aspect, there is provided an encoding device. The device comprises a processing unit and a memory. The memory is coupled to the processing unit and stores instructions for execution by the processing unit. The instructions, when executed by the processing unit, cause the device to perform actions comprising: determining, in response to a source image being mapped to a target image and at least based on at least one source pixel in the source image and at least one target pixel in the target image, an impact of the at least one target pixel on the mapping, the at least one source pixel being mapped to the at least one target pixel; determining, at least based on the impact of the at least one target pixel on the mapping, a distortion caused by encoding of the target image; and determining, based on the distortion, at least part of encoding parameters for the encoding of the target image.

[0069] In some implementations, determining the impact comprises: determining the impact based on a first number of pixels in the at least one source pixel and a second number of pixels in the at least one target pixel.

[0070] In some implementations, the at least one source pixel is associated with at least one degree of importance on the encoding, and determining the impact comprises: determining the impact based on the at least one source pixel, the at least one target pixel and the at least one degree of importance.

[0071] In some implementations, the target image is associated with a plurality of image blocks. The actions further comprise: assigning the determined impact associated with the at least one target pixel to a pixel located in a same image block of the plurality of image blocks as the at least one target pixel.

[0072] In some implementations, determining the distortion comprises: obtaining a reconstructed image of the target image, the reconstructed image being derived from at least in part applying an inverse process of the encoding to the encoded target image; and determining the distortion based on the target image, the reconstructed image and the impact.

[0073] In some implementations, determining the distortion comprises: generating an error measuring the distortion.

[0074] In some implementations, generating the error comprises: generating, based on the impact, at least one of Sum of Squares for Error (SSE), Sum of Absolute Difference (SAD) and Sum of Absolute Transformation Difference (SATD) between the target image and the reconstructed image.

[0075] In some implementations, determining the at least part of the encoding parameters comprises: determining, based on the distortion, the at least part of the encoding parameters

through a Lagrangian optimization process.

5 [0076] In some implementations, determining the at least part of the encoding parameters further comprises: determining, based on the impact and a Lagrangian operator employed in the Lagrangian optimization process, a quantitation parameter as the at least part of the encoding parameters.

[0077] In some implementations, the source image is included in a video to be encoded by the device, the source image comprises a spherical image, the target image comprises a rectangular image and the actions further comprise: mapping the spherical image to the rectangular image by equirectangular projection before determining the
10 impact associated with the at least one target pixel.

[0078] In the second aspect, there is provided an encoding method. The method includes: determining, in response to a source image being mapped to a target image and at least based on at least one source pixel in the source image and at least one target pixel in the target image, an impact of the at least one target pixel on the mapping, the at least one
15 source pixel being mapped to the at least one target pixel; determining, at least based on the impact of the at least one target pixel on the mapping, a distortion caused by encoding of the target image; and determining, based on the distortion, at least part of encoding parameters for the encoding of the target image.

[0079] In some implementations, determining the impact comprises: determining the
20 impact based on a first number of pixels in the at least one source pixel and a second number of pixels in the at least one target pixel.

[0080] In some implementations, the at least one source pixel is associated with at least one degree of importance on the encoding, and determining the impact comprises: determining the impact based on the at least one source pixel, the at least one target pixel
25 and the at least one degree of importance.

[0081] In some implementations, the target image is associated with a plurality of image blocks. The method further comprises: assigning the determined impact associated with the at least one target pixel to a pixel located in a same image block of the plurality of image blocks as the at least one target pixel.

30 [0082] In some implementations, determining the distortion comprises: obtaining a reconstructed image of the target image, the reconstructed image being derived from at least in part applying an inverse process of the encoding to the encoded target image; and determining the distortion based on the target image, the reconstructed image and the impact.

[0083] In some implementations, determining the distortion comprises: generating an

error measuring the distortion.

5 [0084] In some implementations, generating the error comprises: generating, based on the impact, at least one of Sum of Squares for Error (SSE), Sum of Absolute Difference (SAD) and Sum of Absolute Transformation Difference (SATD) between the target image and the reconstructed image.

[0085] In some implementations, determining the at least part of the encoding parameters comprises: determining, based on the distortion, the at least part of the encoding parameters through a Lagrangian optimization process.

10 [0086] In some implementations, determining the at least part of the encoding parameters further comprises: determining, based on the impact and a Lagrangian operator employed in the Lagrangian optimization process, a quantitation parameter as the at least part of the encoding parameters.

[0087] In some implementations, the source image is included in a video to be encoded by the device, the source image comprises a spherical image, the target image
15 comprises a rectangular image and the method further comprises: mapping the spherical image to the rectangular image by equirectangular projection before determining the impact associated with the at least one target pixel.

[0088] In the third aspect, there is provided a computer program product which is tangibly stored on a non-transient machine-readable medium and comprising machine-
20 executable instructions. The machine-executable instructions, when executed by a device, causing the device to perform actions of the method according to the second aspect.

[0089] Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter specified in the appended claims is not necessarily limited to the specific features or acts described
25 above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

CLAIMS

1. An encoding device, comprising:

a processing unit;

a memory coupled to the processing unit and storing instructions for execution by the processing unit, the instructions, when executed by the processing unit, causing the device to perform actions comprising:

determining, in response to a source image being mapped into a target image and at least based on at least one source pixel in the source image and at least one target pixel in the target image, an impact of the at least one target pixel on the mapping, the at least one source pixel being mapped into the at least one target pixel;

determining, at least based on the impact of the at least one target pixel on the mapping, a distortion caused by encoding of the target image; and

determining, based on the distortion, at least part of encoding parameters for the encoding of the target image.

2. The device according to Claim 1, wherein determining the impact comprises:

determining the impact based on a first number of pixels in the at least one source pixel and a second number of pixels in the at least one target pixel.

3. The device according to Claim 1, wherein the at least one source pixel is associated with at least one degree of importance on the encoding, and determining the impact comprises:

determining the impact based on the at least one source pixel, the at least one target pixel and the at least one degree of importance.

4. The device according to Claim 1, wherein the target image is associated with a plurality of image blocks, and the actions further comprise:

assigning the determined impact associated with the at least one target pixel to a pixel in a same image block of the plurality of image blocks as the at least one target pixel.

5. The device according to Claim 1, wherein determining the distortion comprises:

obtaining a reconstructed image of the target image, the reconstructed image being derived from at least in part applying an inverse process of the encoding to the encoded

target image; and

determining the distortion based on the target image, the reconstructed image and the impact.

6. The device according to Claim 5, wherein determining the distortion comprises: generating an error measuring the distortion.

7. The device according to Claim 6, wherein generating the error comprises: generating, based on the impact, at least one of Sum of Squares for Error (SSE), Sum of Absolute Difference (SAD) and Sum of Absolute Transformation Difference (SATD) between the target image and the reconstructed image.

8. The device according to Claim 1, wherein determining the at least part of the encoding parameters comprises:

determining, based on the distortion, the at least part of the encoding parameters through a Lagrangian optimization process.

9. The device according to Claim 8, wherein determining the at least part of the encoding parameters further comprises:

determining, based on the impact and a Lagrangian operator employed in the Lagrangian optimization process, a quantitation parameter as the at least part of the encoding parameters.

10. The device according to Claim 1, wherein the source image is included in a video to be encoded by the device, the source image includes a spherical image, the target image includes a rectangular image, and the actions further comprise:

mapping the spherical image into the rectangular image via equirectangular projection before determining the impact associated with the at least one target pixel.

11. An encoding method, comprising:

determining, in response to a source image being mapped into a target image and at least based on at least one source pixel in the source image and at least one target pixel in the target image, an impact of the at least one target pixel on the mapping, the at least one source pixel being mapped into the at least one target pixel;

determining, at least based on the impact of the at least one target pixel on the mapping, a distortion caused by encoding of the target image; and

determining, based on the distortion, at least part of encoding parameters for the encoding of the target image.

12. The method according to Claim 11, wherein determining the impact comprises: determining the impact based on a first number of pixels in the at least one source pixel and a second number of pixels in the at least one target pixel.

13. The method according to Claim 11, wherein the at least one source pixel is associated with at least one degree of importance on the encoding, and determining the impact comprises:

determining the impact based on the at least one source pixel, the at least one target pixel and the at least one degree of importance.

14. The method according to Claim 11, wherein the target image is associated with a plurality of image blocks, and the actions further comprise:

assigning the determined impact associated with the at least one target pixel to a pixel in a same image block of the plurality of image blocks as the at least one target pixel.

15. The method according to Claim 11, wherein determining the distortion comprises:

obtaining a reconstructed image of the target image, the reconstructed image being derived from at least in part applying an inverse process of the encoding to the encoded target image; and

determining the distortion based on the target image, the reconstructed image and the impact.

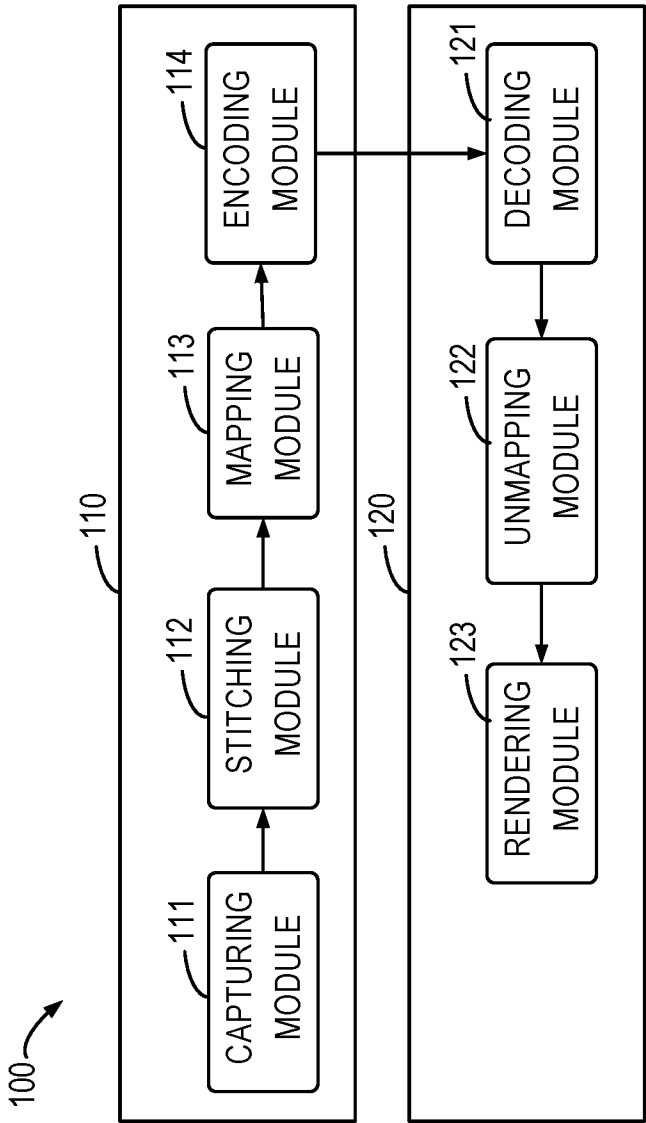
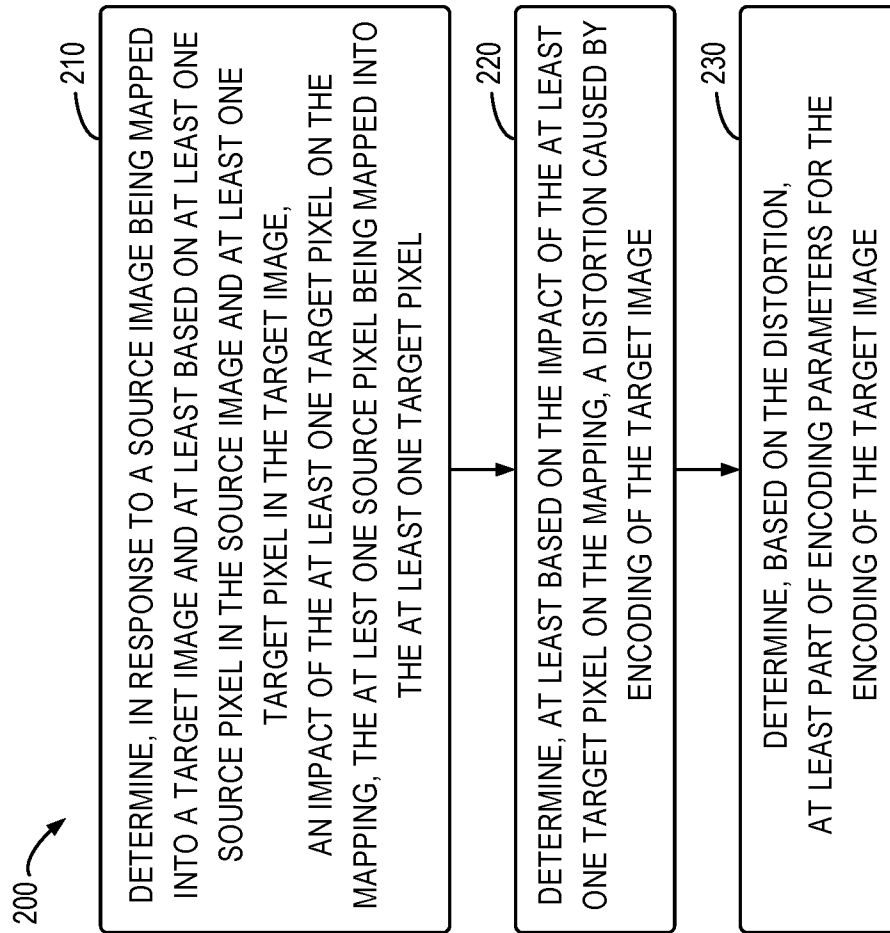


FIG 1

**FIG 2**

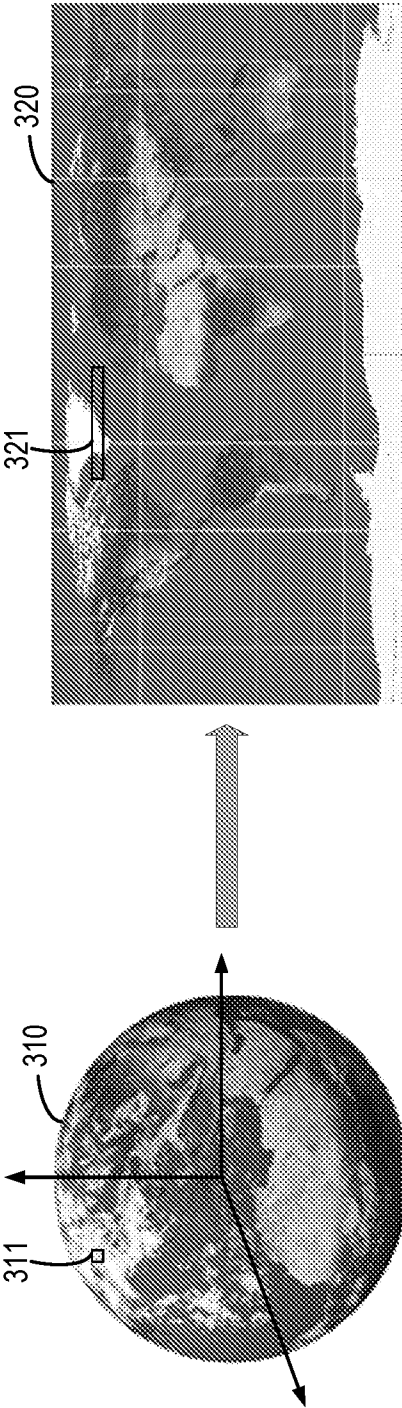


FIG 3

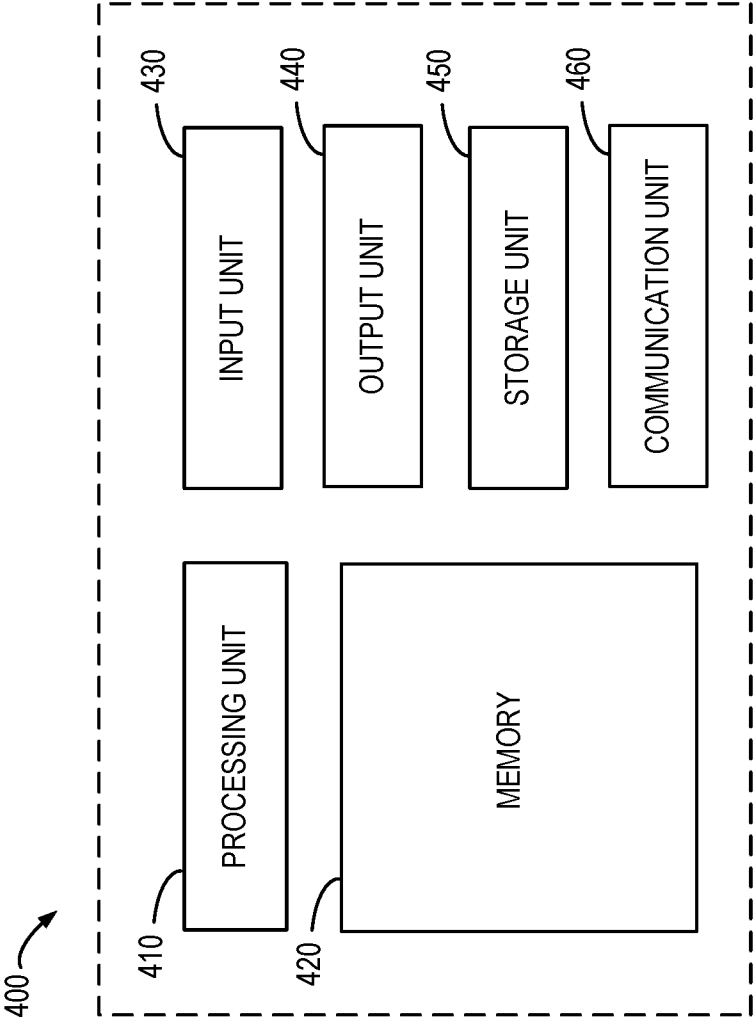


FIG 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/028217

A. CLASSIFICATION OF SUBJECT MATTER

INV. H04N19/597 H04N19/115 H04N19/136 H04N19/182 H04N19/167
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YULE SUN ET AL: "AHG8: Stretching ratio based adaptive quantization for 360 video", 6. JVET MEETING; 31-3-2017 - 7-4-2017; HOBART; (THE JOINT VIDEO EXPLORATION TEAM OF ISO/IEC JTC1/SC29/WG11 AND ITU-T SG.16); URL: HTTP://PHENIX.INT-EVRY.FR/JVET/, no. JVET-F0072, 30 March 2017 (2017-03-30), XP030150744, abstract section 1 ----- -/--	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

26 June 2018

Date of mailing of the international search report

04/07/2018

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

André, Thomas

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/028217

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	YULE SUN ET AL: "[FTV-AHG] WS-PSNR for 360 video quality evaluation", 115. MPEG MEETING; 30-5-2016 - 3-6-2016; GENEVA; (MOTION PICTURE EXPERT GROUP OR ISO/IEC JTC1/SC29/WG11), no. m38551, 27 May 2016 (2016-05-27), XP030066907, sections 1 and 2 -----	1-15
A	SULLIVAN G J; WIEGAND T: "Rate-distortion optimization for video compression", IEEE SIGNAL PROCESSING MAGAZINE., vol. 15, no. 6, 1 November 1998 (1998-11-01), pages 74-90, XP055480988, US ISSN: 1053-5888, DOI: 10.1109/79.733497 pages 78-79: Distortion Measures pages 80-87: Optimization Techniques -----	1-15
X,P	WO 2018/017599 A1 (VID SCALE INC [US]) 25 January 2018 (2018-01-25) paragraphs [0007] - [0009] paragraphs [0029] - [0045] -----	1-15
X,P	LI TIMING ET AL: "Spherical domain rate-distortion optimization for 360-degree video coding", 2017 IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO (ICME), IEEE, 10 July 2017 (2017-07-10), pages 709-714, XP033146709, DOI: 10.1109/ICME.2017.8019492 [retrieved on 2017-08-28] sections 1-3 -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2018/028217

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2018017599	A1	25-01-2018	NONE
