(54) Title: METHOD AND APPARATUS FOR ESTIMATING CONTENT COMPLEXITY FOR VIDEO QUALITY ASSESS-
MENT
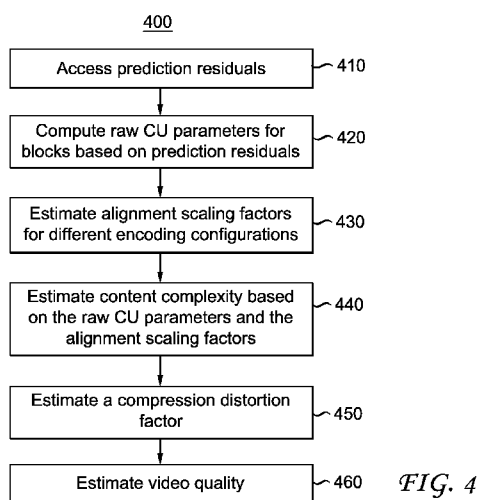
FIG. 4

(57) Abstract: To estimate content complexity of a video, energy of prediction residuals is calculated. The prediction residuals are
usually smaller when the video is less complex and more predictable. Scales of prediction residuals also depend on encoding config-
urations, for example, I pictures usually have larger prediction residuals than P and B pictures even when the contents are very simil-
ar and thus have similar perceived content complexity. To more closely reflect the content complexity, alignment scaling factors are
estimated for different encoding configurations. Based on the energy of prediction residuals and alignment scaling factors, an overall
content unpredictability parameter can be estimated to compute a compression distortion factor for the video. The compression dis-
tortion factor, combined with slicing and freezing distortion factors, can be used to estimate a video quality metric for the video.

METHOD AND APPARATUS FOR ESTIMATING CONTENT

COMPLEXITY FOR VIDEO QUALITY ASSESSMENT

CROSS-REFERENCE TO RELATED APPLICATIONS

5   This application claims the benefit of WO International Application No.

PCT/CN2012/080632, filed August 27, 2012.

TECHNICAL FIELD

This invention relates to video quality measurement, and more particularly, to

a method and apparatus for determining a video quality metric at a bitstream level.

10   The determined video quality metric can then be used, for example, to adjust

encoding parameters, or to provide required video quality at the receiver side.

BACKGROUND

Generally, the more textured or complex the video content is, the more

artifacts in the video content can be tolerated by human eyes. That is, when a video

15   content is viewed by human eyes, visual artifacts may be masked by the video

content itself. This property of human eyes is known as masking property or

masking effect. Thus, content complexity may be considered in visual quality

assessment.

In our previous work, we estimate a content unpredictability (CU) parameter to

20   indicate content complexity. In a commonly owned PCT application, entitled

"Method and apparatus for video quality measurement" by F. Zhang, N. Liao, K. Xie,

and Z. Chen (PCT/CN11/002096, Attorney Docket No. PA110061, hereinafter

"Zhang1"), the teachings of which are specifically incorporated herein by reference, we disclosed a method for predicting video quality using a quantization parameter (QP), which is adjusted by a correction function that depends on content unpredictability.

5   In another commonly owned PCT application, entitled "Video quality measurement" by F. Zhang, N. Liao, K. Xie, and Z. Chen (PCT/CN2011/082870, Attorney Docket No. PA110050, hereinafter "Zhang2"), the teachings of which are specifically incorporated herein by reference, we disclosed a method for estimating a compression distortion factor, a slicing distortion factor, and a freezing distortion

10   factor using parameters (for example, quantization parameters, content unpredictability parameters, ratios of lost blocks, ratios of propagated blocks, error concealment distances, motion vectors, durations of freezing, and frame rates) derived from a bitstream.

SUMMARY

15   The present principles provide a method for assessing video quality of a video included in a bitstream, comprising the steps of: determining a content unpredictability parameter responsive to prediction residuals of a block of a picture in the video; scaling the content unpredictability parameter responsive to the encoding configuration of the block; and determining a quality metric representative of video

20   quality of the video included in the bitstream in response to the scaled content unpredictability parameter as described below. The present principles also provide an apparatus for performing these steps.

        The present principles also provide a method for assessing video quality of a video included in a bitstream, comprising the steps of: determining a content

unpredictability parameter responsive to prediction residuals and encoding configuration of a block of a picture in the video, wherein the content unpredictability parameter is indicative of a degree of difficulty of predicting the block; scaling the content unpredictability parameter responsive to the encoding configuration, wherein a value of the scaled content unpredictability parameter is substantially same as a value of a second content unpredictability parameter for another block that has similar content; and determining a quality metric representative of video quality of the video included in the bitstream in response to the determined content unpredictability parameter as described below. The present principles also provide an apparatus for performing these steps.

The present principles also provide a computer readable storage medium having stored thereon instructions for assessing video quality of a video included in a bitstream, according to the methods described above.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A is a pictorial example depicting three original pictures to be encoded as I, P, and B pictures, respectively, FIG. 1B is a pictorial example depicting raw CU parameters corresponding to FIG. 1A, and FIG. 1C is a pictorial example depicting aligned CU parameters, generated in accordance with an embodiment of the present principles.

FIG. 2A is another pictorial example depicting another three original pictures to be encoded as I, P, and B pictures, respectively, FIG. 2B is a pictorial example depicting raw CU parameters corresponding to FIG. 2A, and FIG. 2C is a pictorial example depicting aligned CU parameters, generated in accordance with an embodiment of the present principles.

FIG. 3 is a pictorial example depicting raw CU parameter distributions for I, P, and B pictures, respectively.

FIG. 4 is a flow diagram depicting an example for estimating video quality based on content complexity, in accordance with an embodiment of the present principles.

FIG. 5 is a block diagram depicting an example of a video quality measurement apparatus that may be used with one or more implementations of the present principles.

FIG. 6 is block diagram depicting an example of a video processing system that may be used with one or more implementations of the present principles.

DETAILED DESCRIPTION

It is known that variance and entropy may be used to measure complexity of natural images. For a general signal sequence, variance quantifies contrast or energy, while entropy evaluates randomness or uncertainty. The work described in A. B. Watson, R. Borthwick, and M. Taylor, "Image quality and entropy masking," SPIE Proceedings, vol. 3016, 1997, suggests that another complexity-related characteristic apart from contrast and randomness, called learn ability, may be used to measure content complexity.

Learn ability may be considered to be similar to predictability. That is, the more likely a signal can be represented by neighborhood, the higher its predictability and learn ability are. In video coding, intra prediction and inter prediction are commonly used to remove redundancy in the video. Specifically, a current block may be predicted from a previously encoded block, and the difference between the

current block and the predicted block (i.e., prediction residual), rather than the current block itself, may be encoded. In general, when intra prediction or inter prediction work well, that is, when the current block can be predicted well from a previously encoded block, the prediction residual becomes small and/or sparse.

5   Therefore, energy of prediction residuals in video coding may be a good indicator of how difficult a video signal can be predicted (i.e., content unpredictability). In the present embodiments, we use content unpredictability to represent content complexity.

When the content predictability parameter is defined based on prediction

10   residuals, it can be computed as parameter to measure the degree of difficulty of predicting a video signal. Because the prediction residuals are available in the compressed domain, the content predictability parameter can be measured without full decoding or accessing decoded image pixels, and thus promote the computational efficiency of applications using CU parameters, such as QoE (quality

15   of experience) assessment and video indexing.

Most existing video compression standards, for example, H.264 and MPEG-2, use a macroblock (MB) as the basic encoding unit. Thus, the following embodiments use a macroblock as the basic processing unit. However, the principles may be adapted to use a block at a different size, for example, an 8x8 block, a 16x8 block, a

20   32x32 block, or a 64x64 block.

In one embodiment, for each macroblock, the CU parameter can be computed using the variance or energy of prediction residuals, which is theoretically equal to the sum of the squared de-quantized DCT coefficients minus the squared sum of de-quantized DC coefficients. When no prediction is performed on a macroblock, the

CU parameter can be computed as energy of pixels. That is, inverse DCT (IDCT) is not necessary for calculating the CU parameter. Given the QP and quantized DCT coefficients which are parsed from a bitstream, the de-quantized DCT coefficients can be estimated with sufficient precision in order to calculate CU parameters.

In an example, in Zhang1, when H.264/AVC is used for encoding, the CU parameter for the r-th macroblock can be calculated as:

$$CU_r = \frac{\left(0.625 \times 2^{QPr/6}\right)^2}{K_{MB}} \left[\sum_{l \subset r} \sum_{k \in l} DCT_{k,l}^2 - \left(\sum_{l \subset r} \frac{DCT_{0,l}}{|l|_r}\right)^2\right], \qquad (1)$$

where constant K_{MB} is the total number of pixels in a macroblock, i.e., 256 for H.264/AVC; $0.625 \times 2^{QPr/6}$ is an approximated formula of calculating a quantization step with respect to a QP value for H.264/AVC; l indexes the partitions in the r-th macroblock; $|l|_r$ denotes the total number of partitions therein, for example, 16 partitions for 4×4 block partitioning and 4 partitions for 8×8 block partitioning; non-negative k indexes the quantized DCT coefficients in the l-th partition of the r-th macroblock; and $DCT_{0,l}$ refers to the quantized DC coefficient in the l-th partition.

The above CU calculation can be applied to blocks (or Coding Unit/Prediction Unit in other standards, for example, HEVC) of any size and any partitions within a block. In the present application, a content unpredictability parameter calculated based on energy of prediction residuals, for example, as calculated in Eq. (1), is denoted as a raw CU parameter.

Prediction technique varies with encoding picture types. For example, I pictures support only intra-frame prediction and may have larger prediction residuals than P and B pictures. Due to inter-frame prediction, P pictures usually have smaller prediction residuals than I pictures. B pictures usually have smaller prediction

residuals than both I and P pictures because of bi-directional inter-frame prediction.

That is, different types of pictures contain prediction residuals of different numerical

scales, even if the corresponding pictures may have similar contents (for example,

the pictures belong to the same scene, or a similarity measure between the pictures

5   is smaller than a threshold). Therefore, the raw CU parameters, calculated as

prediction residual energy over pictures, depend on not only the video content but

also the picture type, and thus may not accurately reflect the complexity of pictures.

FIG. 1A shows three exemplary original pictures to be encoded as I, P, and B

pictures, respectively. Corresponding raw CU parameters are illustrated in FIG. 1B.

10   For visualization, CU parameter values are logarithmically mapped to pixel

brightness (i.e., the brightness of a pixel in FIG. 1B corresponds to the value of a CU

parameter at a macroblock in FIG. 1A). The brighter a pixel is, the larger the CU

parameter is. While the three original pictures are very similar in image content and

thus very similar in perceived content complexity, the I picture has the largest raw

15   CU parameters and the B picture has the smallest raw CU parameters. Similar

observations are also made in another example shown in FIGs.2A and 2B.

FIG. 3 illustrates exemplary distributions of macroblocks' raw CU parameter

values for different picture types. The probability distribution functions are

approximated by normalized histograms with a bin width of 15 (i.e., intervals of raw

20   CU parameter values). The raw CU parameters are calculated on all the

macroblocks from I, P and B pictures, respectively, in consecutive 250 frames of a

576-line video. Overall, the raw CU parameters are the largest for I pictures, and

smallest for B pictures. From I, P to B picture, the tail of the probability distribution

curve gets thinner. The legend of FIG. 3lists Γ distribution parameters, which are

fitted for each type of pictures separately. In this example, we observe that the

shape parameters (0.033 for I pictures, 0.041 for P pictures, and 0.049 for B

pictures) are quite similar, while the scale parameters (4916 for I pictures, 1086 for P

pictures, and 324 for B pictures) differ a lot.

5   From the examples shown in FIGs.1-3, we observe that even when the

perceived content complexity are very similar, the raw CU parameters may vary

significantly with picture types. In order to more accurately measure content

complexity, the present principles are directed to a method and apparatus for

improving the estimated content unpredictability parameters taking into account the

10   perceived variations described above. In particular, the raw CU parameters are

scaled to align with CU parameters obtained in another encoding configuration such

that the aligned CU parameters reflect the perceived content complexity more

closely. After CU parameter alignment, the value of the aligned CU parameter of a

current block in one encoding configuration is substantially the same as the value of

15   the aligned CU parameter of another block in another encoding configuration, when

the another block has similar content as the current block.

The proposed CU parameter alignment method is based on two general

assumptions as discussed in the following.

Assumption 1 Raw CU parameter values of macroblocks conform to independent

20   gamma distribution(Γdistribution). A random variable Y that is Γ-distributed with

shape parameter k and scale parameter θ, denoted as$Y \sim \Gamma(k, \theta)$, has a probability

distribution function:

$$f(y; k, \theta) = \frac{1}{\theta^k} \frac{1}{\Gamma(k)} y^{k-1} e^{-\frac{y}{\theta}} \quad \text{for } y \geq 0 \text{ and } k, \theta > 0. \qquad (2)$$

Assumption 1 is based on the observation that the raw CU parameter of a macroblock is defined as the variance of pixels or prediction residuals, as well as the following lemma:

Lemma: if x_1, x_2, ..., x_n are i.i.d. $\mathcal{N}(\mu, \sigma^2)$ random variables, their variance,

5   i.e., $\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{i=1}^{n}x_n\right)^2$, conforms to a special case of the Gamma distribution,

$\Gamma(\frac{n-1}{2}, 2\sigma^2)$.

Strictly speaking, neither adjacent pixels nor residuals are independently Gaussian distributed, therefore we make a careful approximation that their variance conforms to a generalized Γ distribution $\Gamma(k, 2\sigma^2)$ with a relaxed shape parameter k,

10   where k no longer equals $(n\text{-}1)/2$ exactly. Generally, the correlation and sparsity of adjacent pixels make k much smaller than $(n\text{-}1)/2$. Nevertheless, such correlation should be relatively consistent for the same visual content, despite of different encoding configurations. Therefore, we make the following assumption.

Assumption 2 Raw CU parameter distributions (of same or similar content), under

15   different configurations, have a constant shape parameter k and possibly different scale parametersθ.

Based on the above assumptions, the problem of CU alignment becomes estimation of scale parameters $\theta$for different configurations. Assuming there are two configurations X and Y, the estimation problem can be solved by the maximum

20   likelihood estimator of the distribution parameters on CU training sets $\{x\}_X$ and $\{y\}_Y$:

$$\arg\max_{k,\{\theta\}} \prod_{i=1}^{n_X} f(x_i; k, \theta_X) \prod_{i=1}^{n_Y} f(y_i; k, \theta_Y) \qquad (3)$$

where x and y denote the raw CU parameter values of the MBs under configurations X and Y, respectively, and n_X and n_Y are the total number of sets $\{x\}_X$ and $\{y\}_Y$, respectively.

Subsequently, we can align the MB-wise CU for each configuration and obtain the aligned MB-wise CU (denoted as $\widehat{CU}$), for example, as:

$$\widehat{CU}_i = \begin{cases} CU_i \theta_Y/\theta_X & \text{if MB}_i \text{ is configured with X} \\ CU_i & \text{if MB}_i \text{ is configured with Y} \end{cases} \quad . \tag{4}$$

In the following, we provide an exemplary embodiment for solving the estimation problem as described in Eq. (3). Referring to Eqs. (2) and (3), the log-likelihood is

$$\ell(k, \{\theta\}) = (k - 1)\left(\sum_i \ln x_i + \sum_i \ln y_i\right) - (n_X + n_Y)\ln\Gamma(k)$$
$$- \left(\sum_i x_i/\theta_X + \sum_i y_i/\theta_Y\right) - k(n_X\ln\theta_X + n_Y\ln\theta_Y).$$

Setting the derivative of log-likelihood with respect to θ_X to zero yields:

$$\frac{\partial\ell}{\partial\theta_X} = \sum_i x_i/\theta_X^2 - kn_X/\theta_X = 0,$$

and leads to the maximum likelihood estimator for θ_X:

$$\hat\theta_X = \sum_i x_i/(kn_X).$$

Similarly, the maximum likelihood estimator for θ_Y can be derived as:

$$\hat\theta_Y = \sum_i y_i/(kn_Y).$$

Thus, without estimating k, we have:

$$\hat\theta_Y/\hat\theta_X = \frac{n_X\sum_i y_i}{n_Y\sum_i x_i}. \tag{5}$$

That is, Eq. (5) provides a ratio between scale parameters θ for configurations X and Y, wherein the ratio is also referred to as an alignment scaling factor, to scale the raw CU parameters in order to align them between configurations X and Y.

In one example, when ITU-T P.NBAMS (Parametric Non-intrusive Bitstream Assessment of video Media Streaming quality) databases are used for training, we derive alignment scaling factors and estimate the content complexity of a video sequence as an average value of aligned CU parameters for all available MBs in all pictures of the video sequence. In this example, the content complexity is refined as a weighted combination of CUs:

$$z_c = \frac{1}{T}\left(\sum_{t\in\{I\}}\sum_{r\in t}\frac{CU_r}{20.6\times|r|_t} + \sum_{t\in\{P\}}\sum_{r\in t}\frac{CU_r}{3.52\times|r|_t} + \sum_{t\in\{B\}}\sum_{r\in t}\frac{CU_r}{|r|_t}\right) + c, \qquad (6)$$

where $t\in\{I\}, t\in\{P\}$ and $t\in\{B\}$ represent I pictures, P pictures and B pictures, respectively, CU_r is the raw content unpredictability parameter value of the r-th MB, T is the total number of pictures, c is a constant, and $|r|_t$ denotes the total number of partitions in the r-th MB. The constant 20.6 is the alignment scaling factor between I pictures and B pictures, and the constant 3.52 is the alignment scaling factor between P pictures and B pictures.

For the examples of FIGs.1A and 2A, the aligned CU parameters are illustrated in FIGs.1C and 2C, respectively. Compared with raw CU parameters shown in FIGs. 1B and 2B, the aligned CU parameters are in a more similar scale across different picture types. Thus, the aligned CU parameters may better reflect the content complexity.

In addition to picture types, other encoding configurations, for example, the compression standard in use, the maximum consecutive B pictures, GOP (group of picture) structure, instantaneous refresh rate, and the total number of prediction directions of intra prediction, may affect prediction efficiency and thereby affect the average raw CU parameter values of B, P or I pictures, especially when I frames are used infrequently (i.e., a long GOP is used). Moreover, raw CU parameter values

also depend on the video resolution and frame rate. For example, a higher resolution or frame rate often causes lower raw CU parameter values. Consequently, it is necessary to determine the alignment across different configurations. In one embodiment, depending on the encoding configurations,

5   different training sets can be used and thus alignment scaling factors can be estimated accordingly. For example, a power function with respect to the resolution and a power function with respect to the frame rate can be used for alignment, which may keep aligned CU parameters independent with the configurations and only related to the content.

10   The estimated content unpredictability parameter for the video sequence, for example, z_c in Eq. (6), can be used to replace Q_c in Eq. (1) of Zhang1 (PCT/CN11/002096) for measuring video quality, or CU_T in Eq. (3) of Zhang2 (PCT/CN2011/082870) for calculating the compression distortion factor. That is, the compression distortion factor may be calculated as:

15   $$d_c = (\log z_c)^{b_1} \times (51 - QP_T)^{b_2}, \tag{7}$$

where QP_T is an average QP for the video clip, and b_1 and b_2 are constants.

Combining the compression distortion factor and other distortion factors (for example, slicing distortion factor and freezing distortion factor), an overall video quality metric can be obtained for the video sequence. Since prediction residuals

20   are available in a bitstream, the video quality measurement according to the present principles may be implemented on a bitstream level.

Using z_c and other parameters, a quality metric may be calculated as:

$$q = \frac{MOS_{ub} - MOS_{lb}}{1 + \alpha\left(a_c x_c^{b_{c0}} z_c^{b_{c1}} R^{b_{c2}} F^{b_{c3}} + a_f x_f^{b_{f0}} z_f^{b_{f1}} + a_s x_s^{b_{s0}} z_s^{b_{s1}}\right)^{\beta}} + MOS_{lb}, \tag{8}$$

where output variable q is the predicted quality score; constants MOS_{ub} and MOS_{lb} are the upper bound and lower bound of MOS (Mean Opinion Score), i.e., 5 and 1, respectively; α, β, $\{a\}$ and $\{b\}$ are model parameters ($a_c = 1$ constantly); subscripts c, f and s indicate compression, freezing and slicing impairments respectively;

5   variables $\{x\}$ and $\{z\}$ are model factors and also generally termed as features, which are extracted from video data; R is width*height/256; and F is the frame rate.Specifically, $\{x\}$ and $\{z\}$ are respectively the key factor and the co-variate associated with each type of impairment, for example, x_c is the key factor for compression impairment and z_s is the co-variate for slicing impairment.

10   FIG. 4 illustrates an exemplary method 400 for measuring video quality based on content complexity according to an embodiment of the present principles. Method 400 starts at step 410, wherein prediction residuals are accessed, for example, from a bitstream. At step 420, raw CU parameters are calculated, for example, as energy of the prediction residuals as shown in Eq. (1). At step 430,

15   alignment scaling factors for different encoding configurations are estimated, for example, using Eq. (5). Based on the raw CU parameters and alignment scaling factors, aligned CU parameters are estimated at step 440 to estimate the content complexity. At step 450, a compression distortion factor may be calculated based on the aligned CU parameters and other encoding parameters (for example,

20   quantization parameters). Combining the compression distortion factor with slicing and/or freezing distortion factors, an overall video quality metric can be estimated at step 460, for example, using Eq. (8).

In addition to video quality measurement, the content unpredictability parameter can also be used in texture recognition, object recognition, or scene

recognition.

FIG. 5 depicts a block diagram of an exemplary video quality measurement apparatus 500 that can be used to generate a video quality metric. The input of apparatus 500 includes a transport stream that contains the bitstream. The input

5   may be in other formats that contains the bitstream. A receiver at the system level determines packet losses in the received bitstream.

Demultiplexer 510 parses the input stream to obtain the elementary stream or bitstream. It also passes information about packet losses to decoder 520. The decoder 520 parses necessary information, including QPs, picture types, transform

10   coefficients, and motion vectors for each block or macroblock, in order to generate parameters for estimating the quality of the video. The decoder also uses the information about packet losses to determine which macroblocks in the video are lost. Decoder 520 is denoted as a partial decoder to emphasize that full decoding is not performed, i.e., the video is not reconstructed.

15   Using the MB level QPs parsed from decoder 520, a QP parser 533 obtains average QPs for pictures and for the entire video clip. Using transform coefficients obtained from decoder 520, a transform coefficients parser 532 parses the coefficients. A content unpredictability parameter calculator 534 calculates the content unpredictability parameter for individual pictures and for the entire video clip,

20   based on transform coefficients, quantization parameters, and encoding configurations. Using information about which macroblocks are lost, a lost MB tagger 531 marks which MB is lost. Further using motion information, a propagated MB tagger 535 marks which MBs directly or indirectly use the lost blocks for prediction (i.e., which blocks are affected by error propagation). Using motion

vectors for blocks, an MV parser 536 calculates a motion parameter for MBs, pictures, and the entire video clip. Other modules (not shown) may be used to determine error concealment distances, durations of freezing, and frame rates.

A compression distortion predictor 540 estimates the compression distortion factor, a slicing distortion predictor 542 estimates the slicing distortion factor, and a freezing distortion predictor 544 estimates the freezing distortion factor. Based on the estimated distortion factors, a quality predictor 550 estimates an overall video quality metric.

When extra computation is allowed, a decoder 570 decodes the pictures. The decoder 570 is denoted as a full decoder and it will reconstruct the pictures and perform error concealment if necessary. A mosaic detector 580 performs mosaic detection on the reconstructed video. Using the mosaic detection results, the lost MB tagger 531 and the propagated MB tagger 535 update relevant parameters, for example, the lost block flag and the propagated block flag. A texture masking estimator 585 calculates texture masking weights. The texture masking weights can be used to weigh the distortions.

The video quality measurement apparatus500 may be used, for example, in ITU-T P.NBAMS standard, which works on video quality assessment models in two application scenarios, namely, IPTV and mobile video streaming, also called HR (High Resolution) scenario and LR (Low Resolution) scenario respectively. The difference between the two scenario ranges from the spatio-temporal resolution of video content and coding configuration to transport protocols and viewing conditions.

The input to the P.NBAMS VQM (Video Quality Model) is coded video bitstream with all transmission packet headers (UDP/IP/RTP or UDP/IP/RTP/TS).

The output is an objective MOS score. A major target application of P.NBAMS work is to monitor video quality in a set-top box (STB) or gateway. P.NBAMS mode 1 model only uses bitstream information, and mode 2 model may decode parts or all of the video sequence, and the pixel information is used for visual quality prediction in

5   addition to parsing the bitstream information in order to improve the prediction accuracy.

Referring to FIG. 6, a video transmission system or apparatus 600 is shown, to which the features and principles described above may be applied. A processor 605 processes the video and the encoder 610 encodes the video. The bitstream

10   generated from the encoder is transmitted to a decoder 630 through a distribution network 620. A video quality monitor or a video quality measurement apparatus, for example, the apparatus 500, may be used at different stages.

In one embodiment, a video quality monitor 640 may be used by a content creator. For example, the estimated video quality may be used by an encoder in

15   deciding encoding parameters, such as mode decision or bit rate allocation. In another example, after the video is encoded, the content creator uses the video quality monitor to monitor the quality of encoded video. If the quality metric does not meet a pre-defined quality level, the content creator may choose to re-encode the video to improve the video quality. The content creator may also rank the encoded

20   video based on the quality and charges the content accordingly.

In another embodiment, a video quality monitor 650 may be used by a content distributor. A video quality monitor may be placed in the distribution network. The video quality monitor calculates the quality metrics and reports them to the content distributor. Based on the feedback from the video quality monitor, a content

distributor may improve its service by adjusting bandwidth allocation and access control.

The content distributor may also send the feedback to the content creator to adjust encoding. Note that improving encoding quality at the encoder may not

5   necessarily improve the quality at the decoder side since a high quality encoded video usually requires more bandwidth and leaves less bandwidth for transmission protection. Thus, to reach an optimal quality at the decoder, a balance between the encoding bitrate and the bandwidth for channel protection should be considered.

In another embodiment, a video quality monitor 660 may be used by a user

10   device. For example, when a user device searches videos in Internet, a search result may return many videos or many links to videos corresponding to the requested video content. The videos in the search results may have different quality levels. A video quality monitor can calculate quality metrics for these videos and decide to select which video to store. In another example, the decoder estimates

15   qualities of concealed videos with respect to different error concealment modes. Based on the estimation, an error concealment that provides a better concealment quality may be selected by the decoder.

The implementations described herein may be implemented in, for example, a method or a process, an apparatus, a software program, a data stream, or a signal.

20   Even if only discussed in the context of a single form of implementation (for example, discussed only as a method), the implementation of features discussed may also be implemented in other forms (for example, an apparatus or program). An apparatus may be implemented in, for example, appropriate hardware, software, and firmware. The methods may be implemented in, for example, an apparatus such as, for

example, a processor, which refers to processing devices in general, including, for

example, a computer, a microprocessor, an integrated circuit, or a programmable

logic device. Processors also include communication devices, such as, for example,

computers, cell phones, portable/personal digital assistants ("PDAs"), and other

5   devices that facilitate communication of information between end-users.

Reference to "one embodiment" or "an embodiment" or "one implementation"

or "an implementation" of the present principles, as well as other variations thereof,

mean that a particular feature, structure, characteristic, and so forth described in

connection with the embodiment is included in at least one embodiment of the

10   present principles. Thus, the appearances of the phrase "in one embodiment" or "in

an embodiment" or "in one implementation" or "in an implementation", as well any

other variations, appearing in various places throughout the specification are not

necessarily all referring to the same embodiment.

Additionally, this application or its claims may refer to "determining" various

15   pieces of information. Determining the information may include one or more of, for

example, estimating the information, calculating the information, predicting the

information, or retrieving the information from memory.

Further, this application or its claims may refer to "accessing" various pieces

of information. Accessing the information may include one or more of, for example,

20   receiving the information, retrieving the information (for example, from memory),

storing the information, processing the information, transmitting the information,

moving the information, copying the information, erasing the information, calculating

the information, determining the information, predicting the information, or estimating

the information.

Additionally, this application or its claims may refer to "receiving" various pieces of information. Receiving is, as with "accessing", intended to be a broad term. Receiving the information may include one or more of, for example, accessing the information, or retrieving the information (for example, from memory). Further,

5 "receiving" is typically involved, in one way or another, during operations such as, for example, storing the information, processing the information, transmitting the information, moving the information, copying the information, erasing the information, calculating the information, determining the information, predicting the information, or estimating the information.

10 As will be evident to one of skill in the art, implementations may produce a variety of signals formatted to carry information that may be, for example, stored or transmitted. The information may include, for example, instructions for performing a method, or data produced by one of the described implementations. For example, a signal may be formatted to carry the bitstream of a described embodiment. Such a

15 signal may be formatted, for example, as an electromagnetic wave (for example, using a radio frequency portion of spectrum) or as a baseband signal. The formatting may include, for example, encoding a data stream and modulating a carrier with the encoded data stream. The information that the signal carries may be, for example, analog or digital information. The signal may be transmitted over a

20 variety of different wired or wireless links, as is known. The signal may be stored on a processor-readable medium.

CLAIMS

1. A method for assessing video quality of a video included in a bitstream, comprising the steps of:

determining (420) a content unpredictability parameter responsive to prediction residuals of a block of a picture in the video;

scaling (440) the content unpredictability parameter responsive to the encoding configuration of the block; and

determining (460) a quality metric representative of video quality of the video included in the bitstream in response to the scaled content unpredictability parameter.

2. The method of claim 1, wherein the content unpredictability parameter is indicative of a degree of difficulty of predicting the block, and wherein a value of the scaled content unpredictability parameter is substantially same as a value of a second content unpredictability parameter for another block that has similar content as the block.

3. The method of claim 1, wherein the encoding configuration includes at least one of a picture type and GOP (Group of Picture) structure.

4. The method of claim 1, wherein the step of determining a content unpredictability parameter comprises the steps of:

determining energy of the prediction residuals; and

scaling the energy of the prediction residuals responsive to the encoding configuration.

5. The method of claim 4, wherein a scaling factor used in the scaling step is determined on the basis that the prediction residuals follow a gamma distribution.

6. The method of claim 5, wherein respective scaling factors are determined for a plurality of encoding configurations, on the basis that a shape parameter of the gamma distribution for the encoding configuration is same as another shape parameter for another encoding configuration.

7. The method of claim 1, wherein a respective content unpredictability parameter is determined for each one of a plurality of blocks in the video, and wherein an overall content unpredictability parameter for the video is determined responsive the content unpredictability parameters for the plurality of blocks.

8. The method of claim 1, further comprising:

performing at least one of monitoring quality of the bitstream, adjusting the bitstream in response to the quality metric, creating a new bitstream based on the quality metric, adjusting parameters of a distribution network used to transmit the bitstream, determining whether to keep the bitstream based on the quality metric, and choosing an error concealment mode at a decoder.

9. The method of claim 1, wherein the step of determining the quality metric is further in response to at least one of a resolution and frame rate.

10. An apparatus (500) for assessing video quality of a video included in a
bitstream, comprising:

a decoder (520) configured to access prediction residuals for a block of a
picture of the video and determine encoding configuration of the block;

a content unpredictability parameter calculator (534) configured to

determine a content unpredictability parameter responsive to prediction
residuals of the block, and

scale the content unpredictability parameter responsive to the encoding
configuration of the block; and

a quality predictor (550) configured to determine a quality metric
representative of video quality of the video included in the bitstream in response to
the scaled content unpredictability parameter.

11. The apparatus of claim 10, wherein the content unpredictability parameter
is indicative of a degree of difficulty of predicting the block, and wherein a value of
the scaled content unpredictability parameter is substantially same as a value of a
second content unpredictability parameter for another block that has similar content
as the block.

12. The apparatus of claim 10, wherein the encoding configuration includes at
least one of a picture type and GOP (Group of Picture) structure.

13. The apparatus of claim 10, wherein the content unpredictability parameter
calculator determines energy of the prediction residuals and scales the energy of the
prediction residuals responsive to the encoding configuration.

14. The apparatus of claim 13, wherein the content unpredictability parameter calculator determines a scaling factor on the basis that the prediction residuals follow a gamma distribution.

15. The apparatus of claim 14, wherein respective scaling factors are determined for a plurality of encoding configurations, on the basis that a shape parameter of the gamma distribution for the encoding configuration is same as another shape parameter for another encoding configuration.

16. The apparatus of claim 10, wherein a respective content unpredictability parameter is determined for each one of a plurality of blocks in the video, and wherein an overall content unpredictability parameter for the video is determined responsive the content unpredictability parameters for the plurality of blocks.
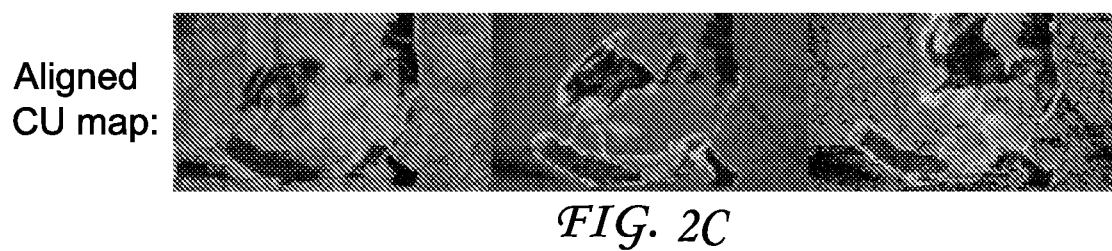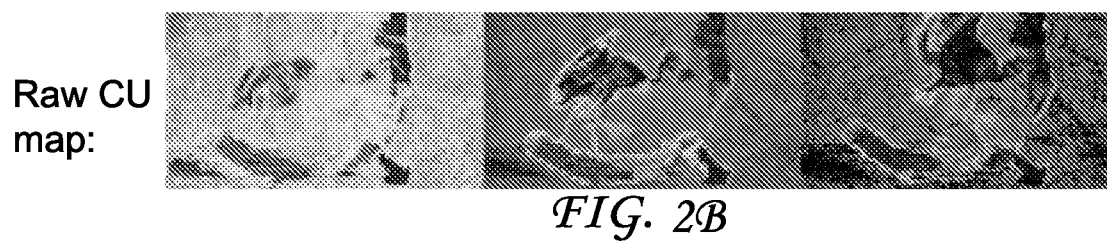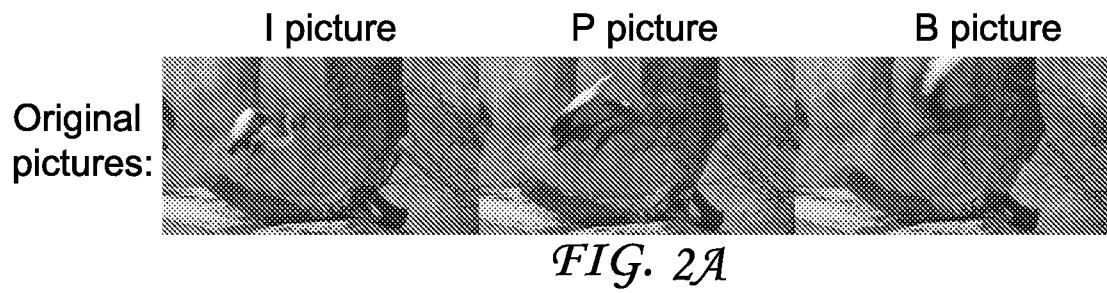
17. The apparatus of claim 10, further comprising:

a video quality monitor (640, 650, 660) configured to perform at least one of monitoring quality of the bitstream, adjusting the bitstream in response to the quality metric, creating a new bitstream based on the quality metric, adjusting parameters of a distribution network used to transmit the bitstream, determining whether to keep the bitstream based on the quality metric, and choosing an error concealment mode at a decoder.

18. The apparatus of claim 10, wherein the content unpredictability parameter calculator determines the quality metric further in response to at least one of a resolution and frame rate.

19. A computer readable storage medium having stored thereon instructions for assessing video quality of a video included in a bitstream, according to claims 1-9.
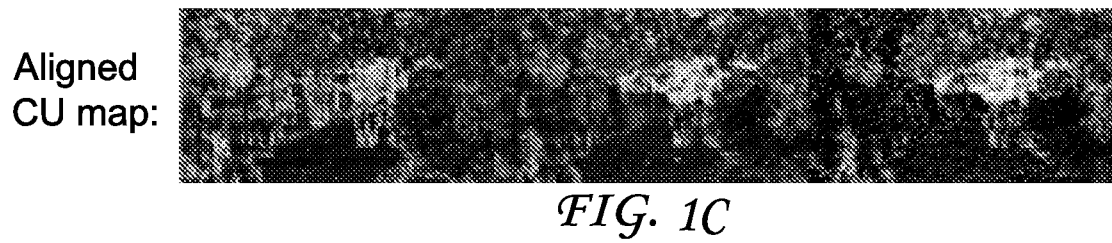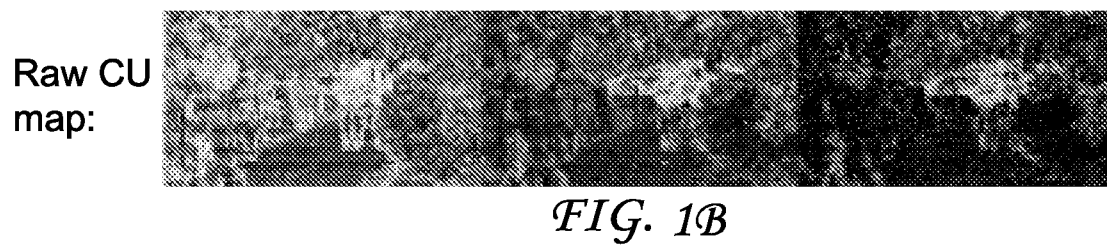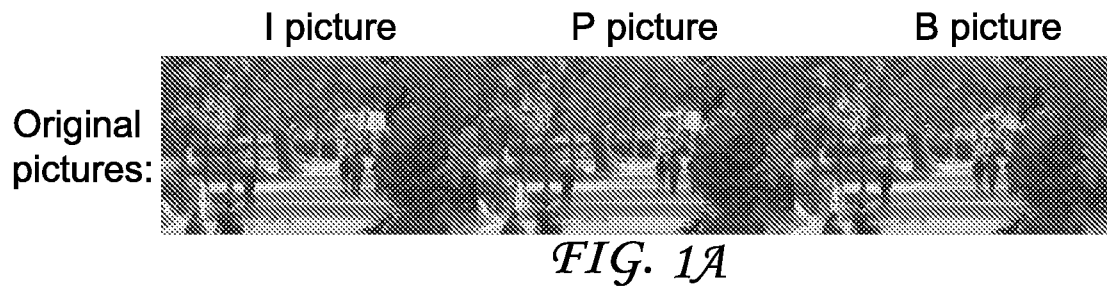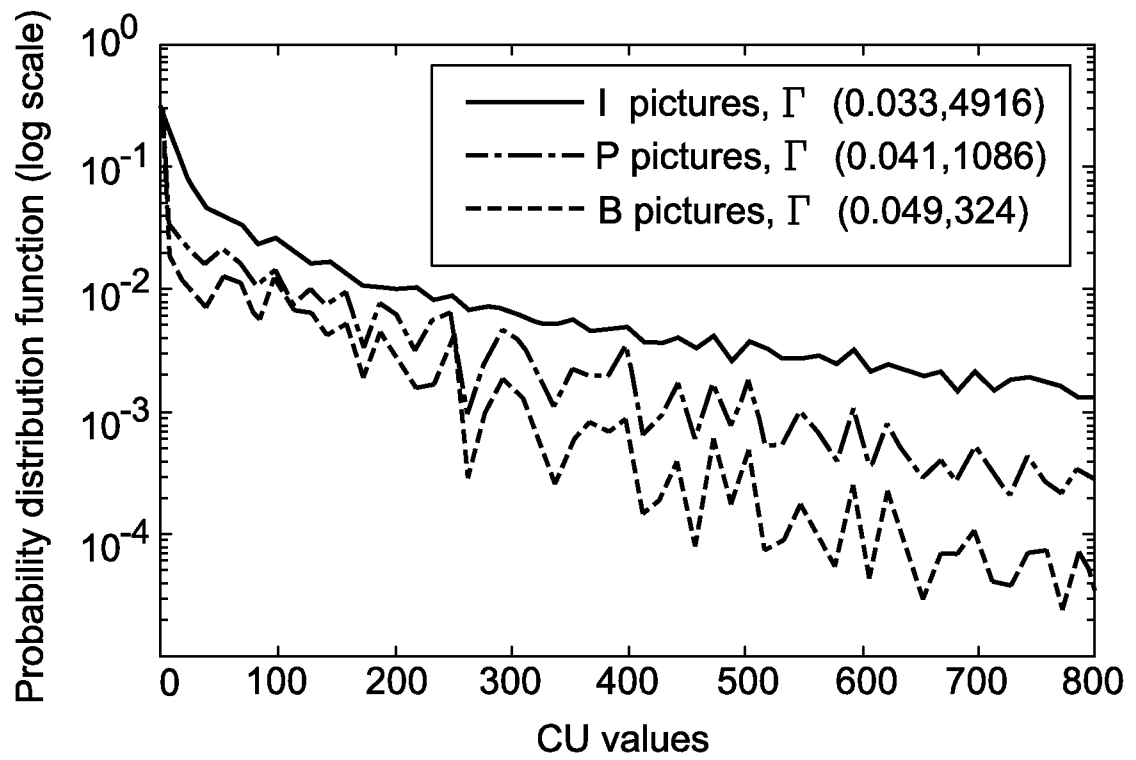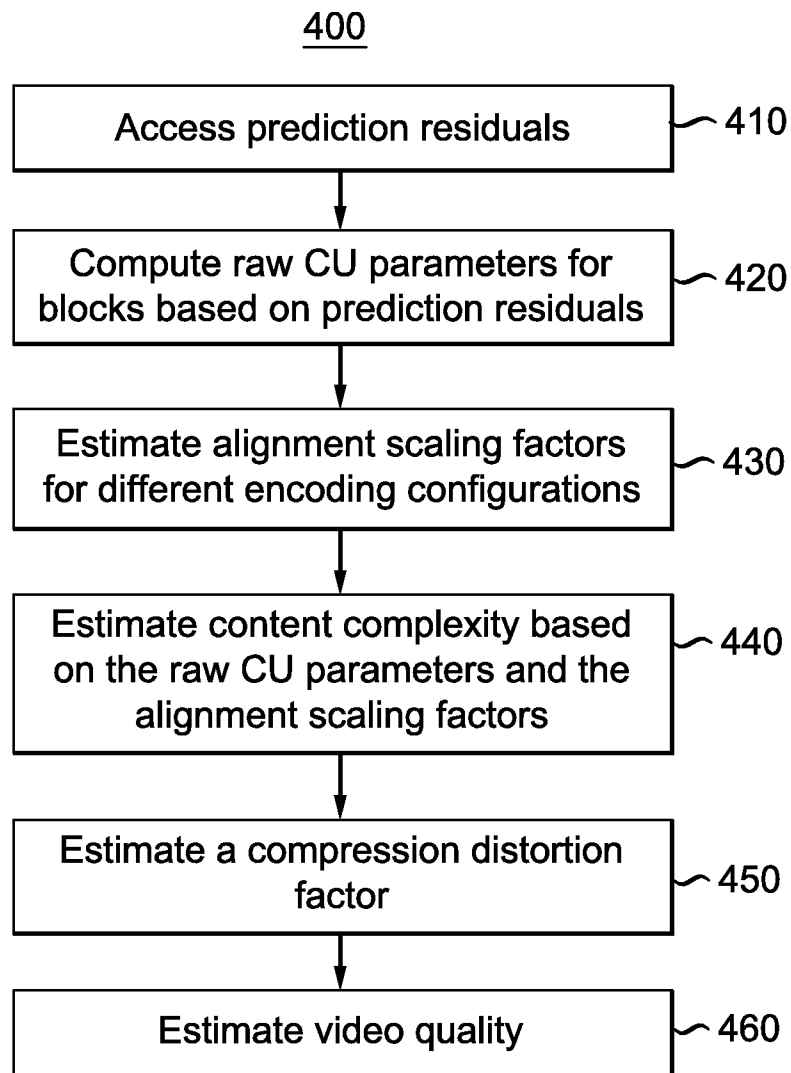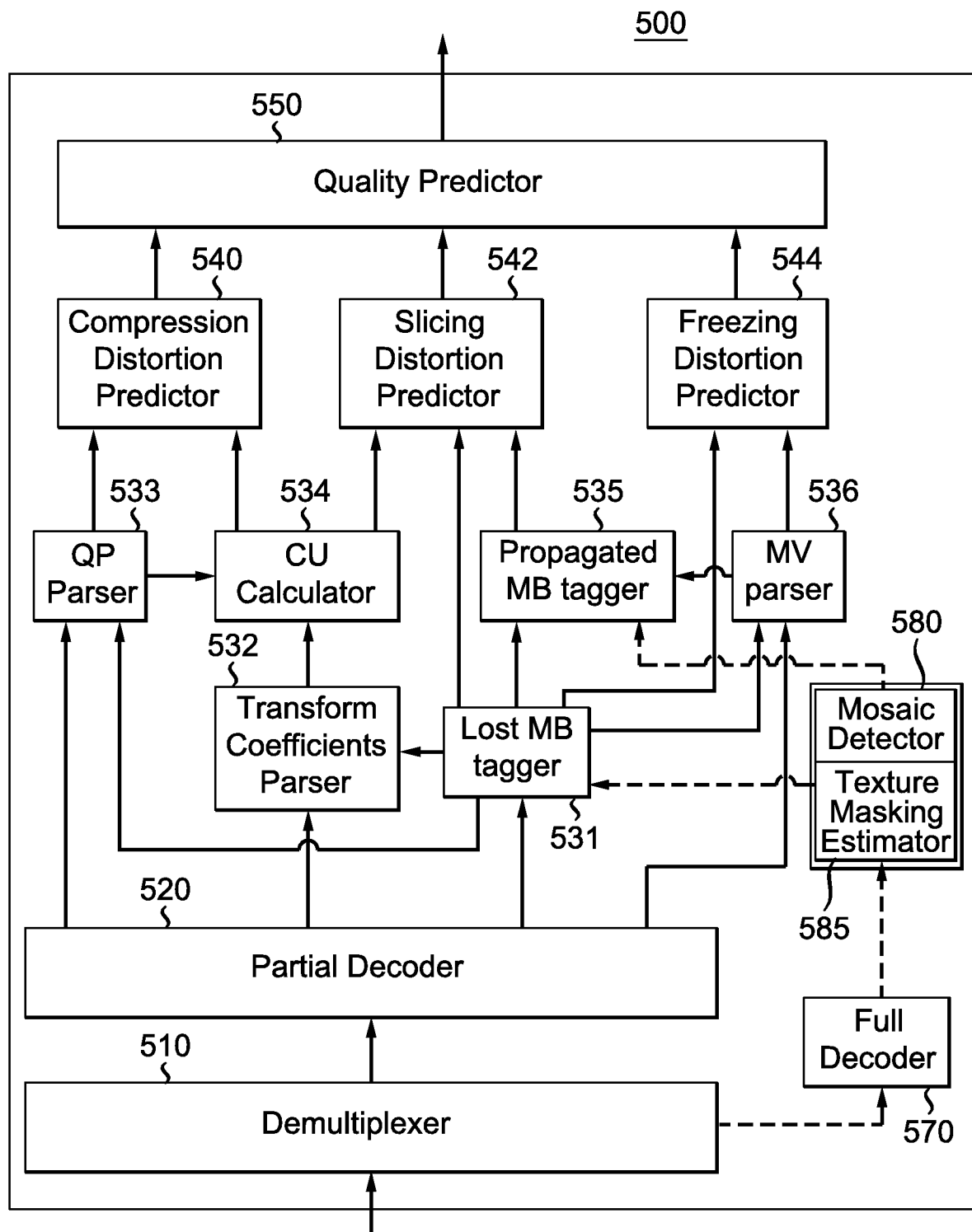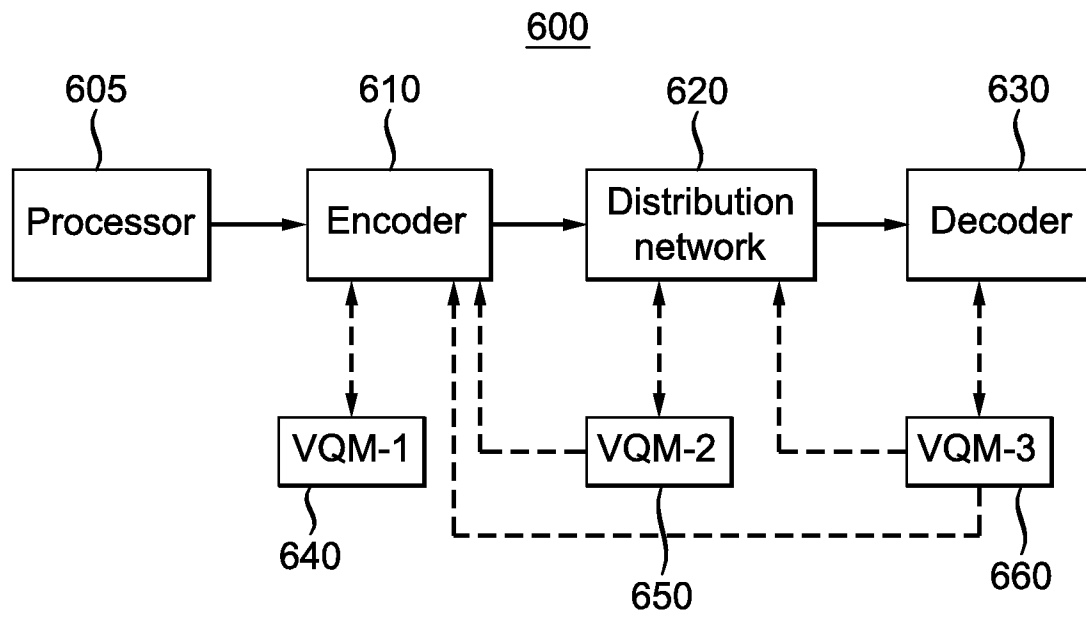
| | I picture | P picture | B picture |

FIG. 1A

FIG. 1B

FIG. 1C

| | I picture | P picture | B picture |

FIG. 2A

FIG. 2B

FIG. 2C

FIG. 3

<u>400</u>

```
┌─────────────────────────────────────┐
│      Access prediction residuals     │──── 410
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│      Compute raw CU parameters for   │──── 420
│      blocks based on prediction      │
│              residuals               │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│     Estimate alignment scaling       │──── 430
│  factors for different encoding      │
│           configurations             │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│  Estimate content complexity based   │──── 440
│  on the raw CU parameters and the    │
│       alignment scaling factors      │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│   Estimate a compression distortion  │──── 450
│                factor                │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│        Estimate video quality        │──── 460
└─────────────────────────────────────┘
```

FIG. 4

FIG. 5

FIG. 6

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

See the extra sheet

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)

IPC: H04N 17/-; H04N 7/-

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)

WPI, EPODOC, CNKI, CNPAT: encod+, decod+, frame?, video, residual?, configuration, unpredictability, parameter, prediction, content, estimate+, quality, metric

C.   DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 2144449 A1 (BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY) 13 January 2010(13.01.2010) description, paragraphs [0001] to [0007], figure 1	1-3, 8-12, 17-19
A	EP 2373049 A1 (BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY) 05 October 2011(05.10.2011) the whole document	1-19
A	US 2008/0198928 A1 (KABUSHIKI KAISHA TOSHIBA et al.) 21 August 2008(21.08.2008) the whole document	1-19
A	WO 2008/049445 A1 (FRAUNHOFER-GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG E.V.) 02 May 2008(02.05.2008) the whole document	1-19

☐ 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 29 August 2013(29.08.2013)	Date of mailing of the international search report **26 Sep. 2013 (26.09.2013)**
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer FENG,Ji Telephone No. (86-10)62413359

Form PCT/ISA /210 (second sheet) (July 2009)

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
EP 2144449 A1	13.01.2010	WO 2010/004238 A1	14.01.2010
		CN 102084660 A	01.06.2011
		JP 2011527544 A	27.10.2011
		KR 20110039214 A	15.04.2011
		US 2011/0102601 A1	05.05.2011
		CA 2726276 A1	14.01.2010
		EP 2294829 A1	16.03.2011
EP 2373049 A1	05.10.2011	WO 2011/121311 A1	06.10.2011
		US 2013/0016224 A1	17.01.2013
		EP 2553935 A1	06.02.2013
		CN 102918862 A	06.02.2013
US 2008/0198928 A1	21.08.2008	JP 2008205627 A	04.09.2008
WO 2008/049445 A1	02.05.2008	None	

INTERNATIONAL SEARCH REPORT

CLASSIFICATION OF SUBJECT MATTER

H04N 17/00(2006.01) i

H04N 7/26(2006.01) n

摘要

为了估计视频的内容复杂度，计算预测残差的能量。当视频较不复杂和更好预测时，预测残差通常较小。预测残差的尺度还取决于编码配置，例如即使当内容非常类似并且因此具有类似的感知到的内容复杂度时，I 画面也通常具有比 P 和 B 画面更大的预测残差。为了更准确地反映内容复杂度，对于不同编码配置估计对准缩放因子。基于预测残差的能量和对准缩放因子，可以估计总内容不可预测性参数，以计算视频的压缩失真因子。结合切片和冻结失真因子的压缩失真因子可以用于估计视频的视频质量度量。