(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 August 2014 (14.08.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/121571 A1

(51) International Patent Classification:
H04N 17/00 (2006.01)

(21) International Application Number:
PCT/CN2013/077082

(22) International Filing Date:
9 June 2013 (09.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/CN2013/071538
7 February 2013 (07.02.2013) CN

(71) Applicant *(for all designated States except US)*: **THOMSON LICENSING** [FR/FR]; 1-5 rue Jeanne d'Arc, F-92130 Issy-les-Moulineaux (FR).

(72) Inventors; and

(71) Applicants *(for US only)*: **LIAO, Ning** [CN/CN]; Room 507, Building 3, XueZhi Yuan, Xueqing Road, Haidian District, Beijing 100083 (CN). **CHEN, Zhibo** [CN/CN]; Room 1613 Yuehua Xuan, Beitaiping Zhuang Road, Haidian District, Beijing 100088 (CN). **ZHANG, Fan** [CN/CN]; Room 101, Block 3, Building 8, Polygardern, 1st Guanshan Road, Hongshan District, Wuhan, Hubei 430074 (CN).

(74) Agent: **LIU, SHEN & ASSOCIATES**; A0601, Huibin Building, No. 8 Beichen Dong Street, Chaoyang District, Beijing 100101 (CN).

(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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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,

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR CONTEXT-BASED VIDEO QUALITY ASSESSMENT

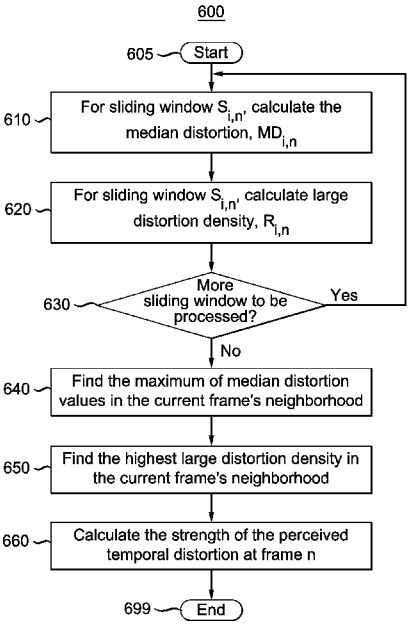

FIG. 6

(57) Abstract: Spatial distortion (i.e., when a frame is viewed independently of other frames in a video sequence) may be quite different from temporal distortion (i.e., when frames are viewed continuously). To estimate temporal distortion, a sliding window approach is used. Specifically, multiple sliding windows around a current frame are considered. Within each sliding window, a large distortion density is calculated and a sliding window with the highest large distortion density is selected. A distance between the current frame and the closest frame with large distortion in the selected window is calculated. Subsequently, the temporal distortion is estimated as a function of the highest large distortion ratio, the spatial distortion for the current frame, and the distance. In another embodiment, a median of spatial distortion values is calculated for each sliding window and the maximum of median spatial distortion values is used to estimate the temporal distortion.

WO 2014/121571 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

METHOD AND APPARATUS FOR CONTEXT-BASED VIDEO QUALITY ASSESSMENT

TECHNICAL FIELD

This invention relates to video quality measurement, and more particularly, to a method and apparatus for determining an objective video quality metric.

BACKGROUND

Different methods have been developed to estimate quality levels of individual frames in a video sequence. Given the estimated quality levels of individual frames in the video sequence, the methods of obtaining an objective overall video quality metric that matches well with the perceived quality of the video sequence has been studied under the umbrella of temporal pooling strategies.

SUMMARY

The present principles provide a method for estimating visual quality of a video sequence, comprising: accessing spatial distortion for frames in a plurality of sliding windows that include a current frame in the video sequence; determining a factor indicating at least one of a large distortion density and a representative artifact level for each sliding window responsive to respective spatial distortion for frames in the each sliding window; determining temporal distortion for the current frame responsive to the determined factor; and determining the visual quality of the video sequence responsive to the temporal distortion of the current frame as described below. The present principles also provide an apparatus for performing these steps.

The present principles provide a method for estimating visual quality of a video sequence, comprising: accessing spatial distortion for frames in a plurality of sliding windows that include a current frame in the video sequence; determining a factor indicating at least one of a large distortion density and a representative artifact level for each sliding window responsive to respective spatial distortion for frames in the each sliding window; determining a maximum of the factors for the plurality of sliding windows; determining temporal distortion for the current frame responsive to the maximum factor; and determining the visual quality of the video sequence responsive to the temporal distortion of the current frame 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 estimating visual quality of a video sequence according to the methods described above.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A is a pictorial example depicting spatial artifact levels for individual frames in an exemplary video sequence, and FIG. 1B is a pictorial example depicting perceived temporal quality for individual frames in the exemplary video sequence.

FIGs. 2A, 2B, and 2C are pictorial examples depicting sliding windows used in video quality modeling, in accordance with an embodiment of the present principles.

FIG. 3A is a pictorial example depicting spatial artifact levels for individual frames in another exemplary video sequence, and FIGs. 3B and 3C are pictorial examples depicting dominant distortion in a frame's neighborhood and estimated

temporal distortion, respectively, in accordance with an embodiment of the present principles.

FIG. 4A is a pictorial example depicting spatial artifact levels for individual frames in another exemplary video sequence, and FIGs. 4B and 4C are pictorial examples depicting the highest large distortion density in a frame's neighborhood and estimated temporal distortion, respectively, in accordance with an embodiment of the present principles.

FIG. 5A is a pictorial example depicting spatial artifact levels for individual frames in another exemplary video sequence, and FIGs. 5B and 5C are pictorial examples depicting the highest large distortion density in a frame's neighborhood and estimated temporal distortion, respectively, in accordance with an embodiment of the present principles.

FIG. 6 is a flow diagram depicting an exemplary method for modeling temporal distortion at frame n, in accordance with an embodiment of the present principles.

FIG. 7 is a block diagram depicting an exemplary video quality monitor, in accordance with an embodiment of the present principles.

FIG. 8 is a block diagram depicting an exemplary video processing system that may be used with one or more implementations.

DETAILED DESCRIPTION

In the present application, we use "spatial artifact" to denote artifact perceived in a picture in a video sequence when the picture is viewed independently of other

pictures in the video sequence, and use “temporal artifact” to denote artifact that is perceived in a picture of a video sequence when pictures in the video sequence are continuously displayed. Similarly, we use “spatial distortion” or “spatial quality” to denote distortion or quality perceived in a picture when the picture is viewed independently of other pictures in a video sequence, and use “temporal distortion” or “temporal quality” to denote distortion or quality that is perceived in a picture of a video sequence when pictures in the video sequence are continuously displayed.

When assessing spatial distortion, a picture is viewed independently of other pictures in the video sequence, for a period of time that is long enough for a viewer to recognize image content and distortion. This is different from assessing temporal distortion, wherein pictures are continuously displayed.

In the present application, the term “frame” is used interchangeably with the term “picture.” The terms “distortion” and “artifact” both refer to quality degradation in a frame, and are often used interchangeably.

Spatial distortion of frame n, denoted as ds(n), can be obtained by various image quality assessment methods, for example, but not limited to, a full-reference or no-reference method, and a method in a pixel domain or at a bitstream level. In one embodiment, spatial distortion can be estimated as the overall visible artifact level, caused by initial and/or propagated visible artifact, as disclosed in a commonly owned PCT application, entitled “Video quality assessment at a bitstream level” by N. Liao, Z. Chen, and K. Xie (PCT/CN2012/078766, Attorney Docket No. PA120027), the teachings of which are specifically incorporated herein by reference.

Spatial artifact in pictures needs to last for a period of time so that eyes can fix

on and recognize it as artifact. When the pictures are part of a video sequence and each is displayed only for a very short period of time (for example, a period of 1/frame_rate when the video is played in real time), the perceived video distortion at the time instant of frame n, i.e., temporal distortion at frame n, dt(n), can be quite different from spatial distortion of frame n, ds(n). This may be because temporal distortion dt(n) is affected by the content and distortion of its neighboring frames.

For an exemplary video sequence, FIG. 1A shows spatial artifact levels of individual frames in the video sequence, and FIG. 1B shows temporal quality of individual frames in the video sequence. More specifically, FIG. 1A shows spatial artifact levels of the frames when the exemplary video sequence suffers from packet losses. The spatial artifact may be sporadic in the video sequence, for example, the spatial artifact seen at frames 74, 77, 215, 261, and 262. The spatial artifact may also occur in a burst, such as the artifact seen around frames 106-111.

FIG. 1B shows temporal quality when the frames of the video sequence are displayed continuously, wherein score 100 corresponds to the best quality. Depending on the quality measurement method, the quality score may be at a different scale. The accurate curve of temporal quality may be obtained using a subjective test method, for example, but not limited to, Single Stimulus Continuous Quality Evaluation (SSCQE), as defined in ITU-R BT 500.

It can be seen from FIG. 1A that frames 74 and 77 have strong spatial artifact when these two frames are viewed independently. However, the artifact at these two frames becomes invisible when the video is displayed continuously, and thus, frames 74 and 77 are at the best quality level when viewed continuously as shown in FIG.

1B. Thus, strong spatial artifact may not always correspond to high temporal distortion. On the other hand, one frame of a video sequence may appear to have good quality when viewed independently of other frames, but may present very strong temporal distortion (for example, motion jitter) when the video sequence is displayed continuously. That is, small spatial distortion may not always correspond to small perceived temporal distortion (i.e., higher temporal quality).

Thus, spatial distortion (i.e., when a picture is viewed independently of other pictures in a video sequence) may be quite different from temporal distortion (i.e., when pictures are viewed continuously). The present principles provide a method and apparatus for accurately modeling temporal quality from spatial distortion for individual frames. Specifically, the present principles consider the context that affects how a viewer identifies temporal distortion, wherein the context includes, for example, but not limited to, the duration and the pattern of the distortion, and texture and object's motion that are recognized by a viewer via watching the neighboring frames.

In one embodiment, given the spatial distortion, the perceived temporal distortion can be modeled using a sliding window approach. As shown in FIGs. 2A-2C, a sliding window of L_0 frames that includes frame n (denoted as $S_{i,n}$) starts at frame $(n - i)$ and ends at frame $(n - i + L_0 - 1)$, $0 \leq i < L_0$. We may choose a window length that approximates the period of time that is needed for a viewer's eyes to recognize distortion. Accordingly, a sliding window includes $L_0 = \mathrm{round}(T_0 \times \mathrm{frame_rate})$ frames, where T_0 is a period of time that is needed for a viewer to recognize distortion, and frame_rate is the frame rate of the video sequence. Artifact existing outside the sliding windows is regarded as having little contribution to the visibility of the temporal artifact of a current frame.

We observe that the perceived temporal distortion of a current frame is mainly affected by frames with large distortion (i.e., distortion level exceeds a certain threshold) that are close by. In the present principles, we provide different embodiments to estimate the impact of such clustered large spatial distortion in a frame's neighborhood.

In one embodiment, we may use the maximum of median distortion values in the sliding windows to estimate the temporal distortion. Mathematically, for each sliding window $S_{i,n}$, we can calculate the median distortion of the frames in$S_{i,n}$as:

$$MD_{i,n} = \text{median}\{ds(j), \text{frame } j \in S_{i,n}\}. \quad (1)$$

A median function examines neighboring frames to decide whether or not the spatial distortion of the current frame is representative of its surroundings and rejects extreme distortion levels (outliers). That is, we may consider $MD_{i,n}$ as a representative distortion level for sliding window $S_{i,n}$. In other embodiment, we may use averaging or other filtering methods to obtain $MD_{i,n}$.

In all sliding windows that include frame n, the sliding window with larger distortion usually affects the visual quality more significantly. Thus, a maximum function can be used to identify the dominant distortion as the maximum of median distortion values among the sliding windows $\{S_{i,n}, 0 \le i < L_0\}$:

$$m_n = \max\{MD_{i,n},\ 0 \le i < L_0\}. \quad (2)$$

Perceived temporal distortion for the current frame is observed to be largely determined by the dominant distortion in its neighborhood, rather than by the spatial distortion strength of the current frame. For example, when frame n has severe spatial distortion ds(n) which does not last long enough to be recognized, the

perceived temporal distortion is mainly affected by distortion in its neighborhood. Thus, a minimum function can be used to set the temporal distortion for the current frame. That is, we may set temporal distortion$\mathrm{dt(n)}$ to the smaller one of the spatial distortion value and the maximum of median distortion values$\mathrm{m_n}$for frame n as:

$$\mathrm{dt(n) = \min\{m_n,\ ds(n)\}.} \tag{3}$$

FIG. 3A shows spatial artifact levels for frames in an exemplary video sequence, and FIGs. 3B and 3C show results after applying Eqs. (2) and (3), respectively. By using a median function, the maximum of median distortion values avoid extreme values as shown in FIG. 3B. Consequently, as can be seen from FIG. 3C, spatial artifact levels that are much larger than neighboring ones, for example, at frames 86, 125, and 166, are not present in the estimated temporal distortion. In addition, the estimated temporal distortion levels have smaller variations from frame to frame than the spatial distortion levels.

In another embodiment, to estimate temporal distortion, we may use the maximum value of the ratios of frames with large distortion (i.e., the distortion level exceeds a certain threshold) in all the sliding windows $\{\mathrm{S_{i,n}}, 0 \le \mathrm{i} < \mathrm{L_0}\}$.Since smaller distortion usually becomes invisible when the video sequence is played continuously, we only consider frames with large distortion. Mathematically, for each sliding window $\mathrm{S_{i,n}}$, we can calculate the ratio between the number of frames with large distortion in$\mathrm{S_{i,n}}$ and the total number of frames in $\mathrm{S_{i,n}}$as:

$$\mathrm{R_{i,n}} = \frac{\sum_j \mu(\mathrm{ds(j)})}{\mathrm{L_0}}, \mathrm{frame\ j} \in \mathrm{S_{i,n}}, \tag{4}$$

where $\mu(\mathrm{x}) = \begin{cases} 1, & \mathrm{x} \ge \mathrm{U} \\ 0, & \mathrm{otherwise} \end{cases}$, and U is a spatial distortion threshold. For ease of notation, we denote the ratio between the number of frames with large spatial

distortion in a sliding window and the total number of frames in the sliding window as a large distortion density for the sliding window.

We observe that a sliding window with a higher density of larger distortion usually affects the visual quality more significantly. Thus, we calculate the highest ratio among all the sliding windows as:

$$w_n = \max\{R_{i,n}, 0 \le i < L_0\}. \quad (5)$$

Then we can estimate the perceived temporal distortion at frame n as a weighted distortion:

$$dt(n) = w_n \times ds(n). \quad (6)$$

In addition, we observe that the distance between two nearest frames having large distortion also affects the perceived temporal distortion. That is, when two frames with large distortion are closer, the distortion becomes more visible to human eyes. Thus, the temporal distortion estimation may be improved by also considering the distance between frame n and a closest frame with large distortion. In one embodiment, Eq. (6) may be modified to become:

$$dt(n) = w_n \times ds(n)/f(dist(n)), \quad (7)$$

wherein dist(n) is the distance between frame n and the closest frame with large distortion in a sliding window corresponding to the highest large distortion density. If there is no other frame with large distortion in the corresponding sliding window, we set dist(n) to a very big value, for example, 1000. That is, when there is only one frame in the sliding window with large distortion, we consider the distortion as less visible and set dt(n) to a very small value. Different functions can be used to define f(dist(n)), for example, f(dist(n)) = dist(n), or f(dist(n)) = $(dist(n))^{\alpha}$, where α is a constant.

FIG. 4A shows spatial artifact levels for frames in an exemplary video sequence, and FIGs. 4B and 4C show results after applying Eqs. (5) and (7), respectively, wherein L_0 = 1second, U = (number of macro blocks per frame)/100. The values of L_0 and U may vary with configurations, for example, with the GOP length, video resolution, and the frame rate.FIG. 5A shows spatial artifact levels for frames in another exemplary video sequence, and FIGs. 5B and 5C show results after applying Eqs. (5) and (7), respectively.

From FIGs. 4B and 5B, we observe that when large distortion is more clustered, the ratio becomes higher. That is, the ratio effectively reflects whether and to which degree neighboring frames affect the perceived quality of the current frame. Consequently, as can be seen from FIGs. 4C and 5C, the duration of frames with clustered large distortion (for example, around frame 106 in FIG. 4A, around frame 105 and frame 200 in FIG. 5A) affects the estimated temporal distortion levels. That is, a shorter duration of distortion as shown in FIG. 4A results in estimated temporal distortion that is much smaller than the spatial distortion, while a longer duration of distortion as shown in FIG. 5A results in estimated temporal distortion that is in the same order as the spatial distortion. This is consistent with the human perception, where lasting distortion has more negative impact on the perceived quality.

As discussed above, we consider contexts provided in neighboring frames to estimate temporal distortion for frame n. In one embodiment, as shown in Eqs. (1)-(3), we use the median distortion value. In another embodiment, as shown in Eqs. (4)-(7), we use the large distortion density. In yet another embodiment, the median distortion value and the large distortion density approaches can be combined to

estimate the temporal distortion for frame n as:

$$\mathrm{dt}(\mathrm{n}) = \mathrm{w_n} \times \min\{\mathrm{m_n},\ \mathrm{ds}(\mathrm{n})\}. \quad (8)$$

In the above, we estimate temporal distortion based on the human vision property that eyes need a period of time that is long enough to recognize artifact. The temporal distortion may also be affected by other factors, for example, but not limited to, motion jerkiness. Consequently, the temporal distortion estimated as above may need to be adjusted to consider other factors. One exemplary combination can be $\mathrm{dt}'(\mathrm{n}) = \mathrm{dt}(\mathrm{n}) + \mathrm{c} \times \mathrm{dt}_2(\mathrm{n})$, where $dt_2(n)$is the distortion caused by motion jerkness.

Based on the temporal distortion for individual frames in a video sequence, $\{\mathrm{dt}(\mathrm{n})\}$, the overall perceived distortion of the video sequence can be derived using a temporal pooling strategy. For example, a simple average operation can be used, that is, $\text{overall distortion} = \frac{1}{\mathrm{N}}\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{dt}(\mathrm{n})$.

FIG. 6 illustrates an exemplary method 600 for modeling temporal distortion at frame n, according to the present principles. Method 600 starts at step 605. For sliding window $S_{i,n}$, it calculates the median distortion value at step 610, for example, using Eq. (1), and calculates the large distortion density at step 620, for example, using Eq. (4). It checks whether more sliding window needs to be processed at step 630. If yes, it returns the control to step 610. Otherwise, at step 640, it calculates the maximum of median distortion values in all sliding windows for frame n, for example, using Eq. (2). At step 650, it calculates the highest large distortion density in all sliding windows for frame n, for example, using Eq. (5). At step 660, it estimates the temporal distortion for frame n, for example, using Eq. (3),(6), or (8).

The distance between frame n and the closest frame with large distortion may be considered at step 660, for example, using Eq. (7).Method 600 ends at step 699.

In method 600, both the maximum of median distortion values and the highest large distortion density are used to estimate temporal distortion. In another embodiment, only the maximum of median distortion values is used to estimate the temporal distortion. That is, steps 620 and 650 are not needed, and step 660 estimates the temporal distortion based on the maximum of median distortion values, for example, using Eq. (3). In another embodiment, only the highest large distortion density is used to estimate the temporal distortion. That is, steps 610 and 640 are not needed, and step 660 estimates the temporal distortion based on the highest large distortion density, for example, using Eq. (6) or Eq. (7).

Method 600 or its variations may proceed in a different order of steps, for example, step 620 may be performed before step 610, step 650 may be performed before step 640.

The video quality modeling methods according to the present principles can be applied to measure video quality when a video sequence suffers from compression artifacts and/or packet losses. The present methods may be used in conjunction with spatial artifacts obtained at a bitstream level or a pixel level. That is, the present methods can be applied at both a bitstream level and a pixel level.

The present principles may be used in a video quality monitor to measure video quality. FIG. 7 depicts a block diagram of an exemplary video quality monitor 700. The input of apparatus 700 may include a transport stream that contains the bitstream. The input may be in other formats that contains the bitstream.

Demultiplexer 710 obtains packet layer information from the bitstream. Decoder 720 parses the input stream to obtain more information. Decoder 720 may or may not reconstruct the pictures. In other embodiments, the decoder may perform the functions of the demultiplexer.

Using the decoded information, the spatial artifact levels are estimated in spatial artifact level estimator730. Based on the estimated parameters, temporal distortion levels are estimated at temporal distortion estimator 740, for example, using method 600. A quality predictor 750 then pools temporal distortion levels for individual frames into a quality score for the video sequence. The quality predictor 750 may consider other types of artifacts and the property of human visual property.

The video quality monitor700 may be used, for example, in ITU-T P.NBAMS(parametric non-intrusive bitstream assessment of video media streaming quality) 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 (Mean Opinion 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 part or all of the video sequence, and the pixel information is

used for visual quality prediction in addition to parsing the bitstream information in order to improve the prediction accuracy.

Referring to FIG. 8, a video transmission system or apparatus 800 is shown, to which the features and principles described above may be applied. A processor 805 processes the video and the encoder 810 encodes the video. The bitstream generated from the encoder is transmitted to a decoder 830 through a distribution network 820. A video quality monitor may be used at different stages.

In one embodiment, a video quality monitor 840 may be used by a content creator. For example, the estimated video quality may be used by an encoder in 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 video based on the quality and charges the content accordingly.

In another embodiment, a video quality monitor850 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

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 bit rate and the bandwidth for channel protection should be considered.

In another embodiment, a video quality monitor 860 may be used by a user 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 user device may have access to several error concealment techniques. A video quality monitor can calculate quality metrics for different error concealment techniques and automatically choose which concealment technique to use based on the calculated quality metrics.

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. 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 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 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 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, 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, "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.

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 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 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 estimating visual quality of a video sequence, comprising:

accessing spatial distortion for frames in a plurality of sliding windows that include a current frame in the video sequence;

determining a factor indicating at least one of a large distortion density and a representative artifact level for each sliding window responsive to respective spatial distortion for frames in the each sliding window (610, 620);

determining temporal distortion for the current frame responsive to the determined factor (660); and

determining the visual quality of the video sequence responsive to the temporal distortion of the current frame.

2. The method of claim 1, further comprising:

determining a maximum of the factors for the plurality of sliding windows (640, 650), wherein the temporal distortion for the current frame is determined responsive to the maximum factor.

3. The method of claim 1, wherein the large distortion density is determined responsive to a ratio between the number of frames with large distortion in the each sliding window and the total number of frames in the each sliding window.

4. The method of claim 1, wherein the temporal distortion for the current frame is determined as a product of spatial distortion for the current frame and the maximum factor.

5. The method of claim 1, wherein the representative artifact level is determined responsive to a median function of the spatial distortion for the each sliding window.

6. The method of claim 1, wherein the temporal distortion for the current frame is determined as a smaller one of spatial distortion for the current frame and the maximum factor.

7. The method of claim 1, further comprising:

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

8. An apparatus (700, 800) for estimating visual quality of a video sequence, comprising:

a spatial artifact level estimator (730) accessing spatial distortion for frames in a plurality of sliding windows that include a current frame in the video sequence;

a temporal distortion estimator (740) determining a factor indicating at least one of a large distortion density and a representative artifact level for each sliding window responsive to respective spatial distortion for frames in the each sliding window and determining temporal distortion for the current frame responsive to the determined factors for the plurality of sliding windows; and

a quality predictor (750) determining the visual quality of the video sequence responsive to the temporal distortion of the current frame.

9. The apparatus of claim 8, wherein the temporal distortion estimator (740) determines a maximum of the factors for the plurality of sliding windows, wherein the temporal distortion for the current frame is determined responsive to the maximum factor.

10. The apparatus of claim 8, wherein the temporal distortion estimator (740) determines the large distortion density responsive to a ratio between the number of frames with large distortion in the each sliding window and the total number of frames in the each sliding window.

11. The apparatus of claim 8, wherein the temporal distortion estimator (740) determines the temporal distortion for the current frame as a product of the spatial distortion and the maximum factor.

12. The apparatus of claim 8, wherein the temporal distortion estimator (740) determines the representative artifact level responsive to a median function of the spatial distortion for the each sliding window.

13. The apparatus of claim 8, wherein the temporal distortion estimator (740) determines the temporal distortion for the current frame as a smaller one of the spatial distortion and the maximum factor.

14. The apparatus of claim 8, further comprising:

a video quality monitor (840, 850, 860) performing at least one of monitoring quality of a bitstream, adjusting the bitstream in response to the determined visual quality, creating a new bitstream based on the determined visual quality, adjusting parameters of a distribution network used to transmit the bitstream, determining whether to keep the bitstream based on the determined visual quality, and choosing an error concealment mode at a decoder.

15. A computer readable storage medium having stored thereon instructions for estimating visual quality of a video sequence, according to claims 1-7.

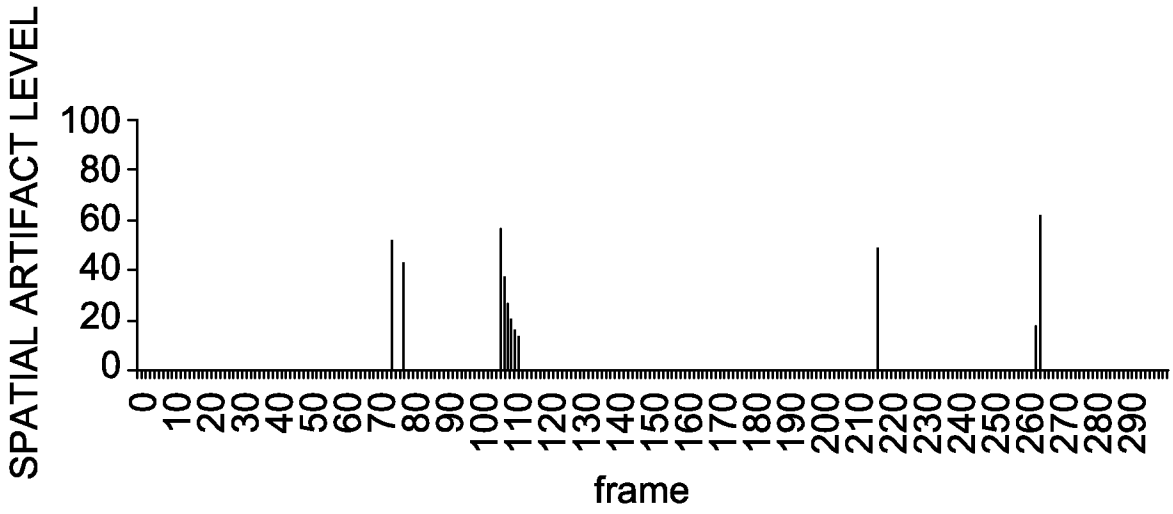
SPATIAL ARTIFACT LEVEL
100
80
60
40
20
0
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290
frame

FIG. 1A

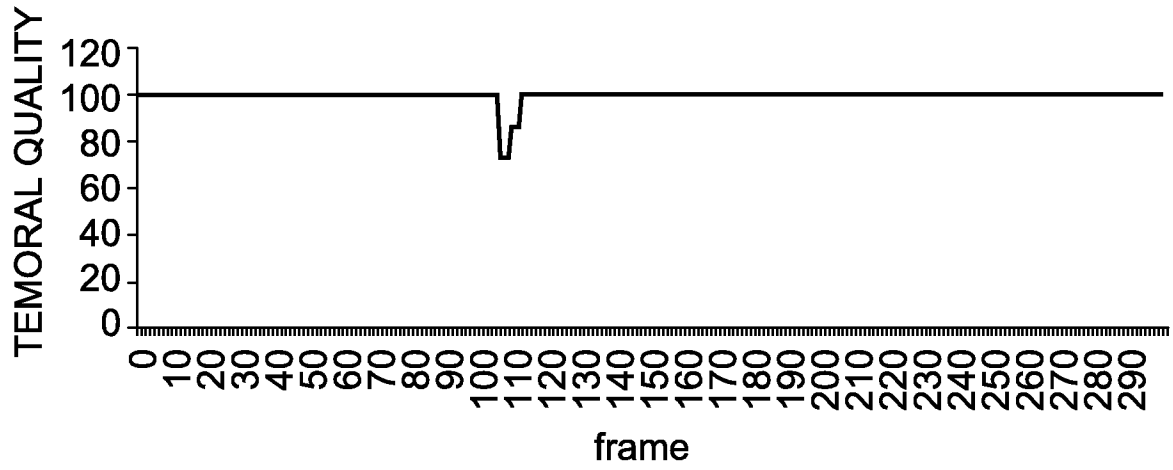
TEMORAL QUALITY
120
100
80
60
40
20
0
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290
frame

FIG. 1B

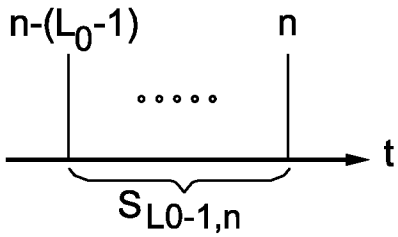

FIG. 2A

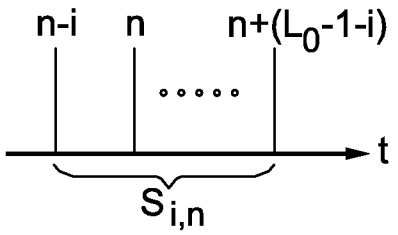

FIG. 2B

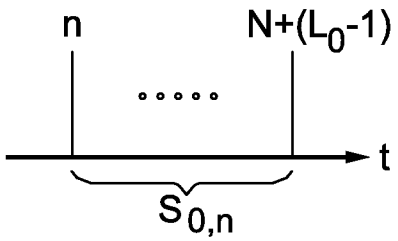

FIG. 2C

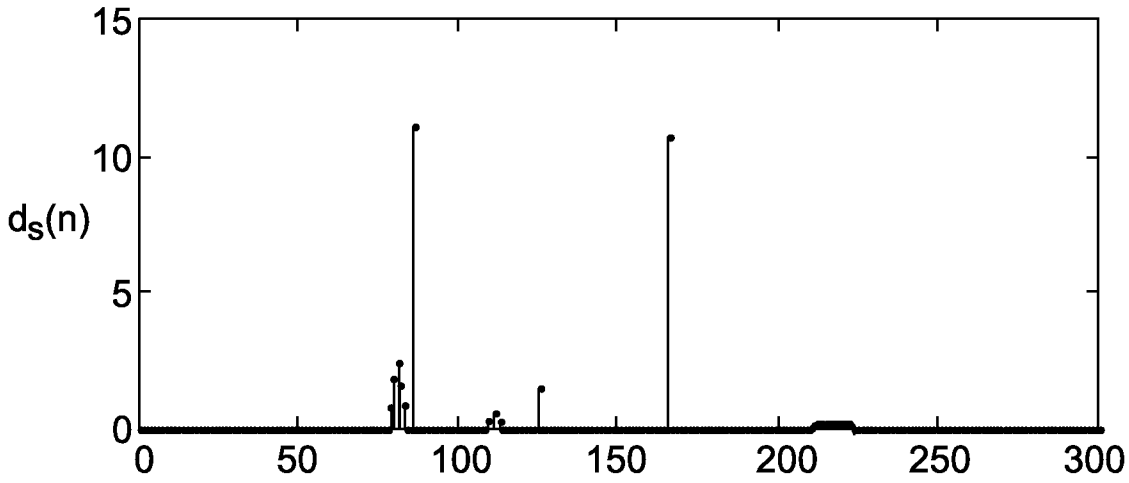

FIG. 3A

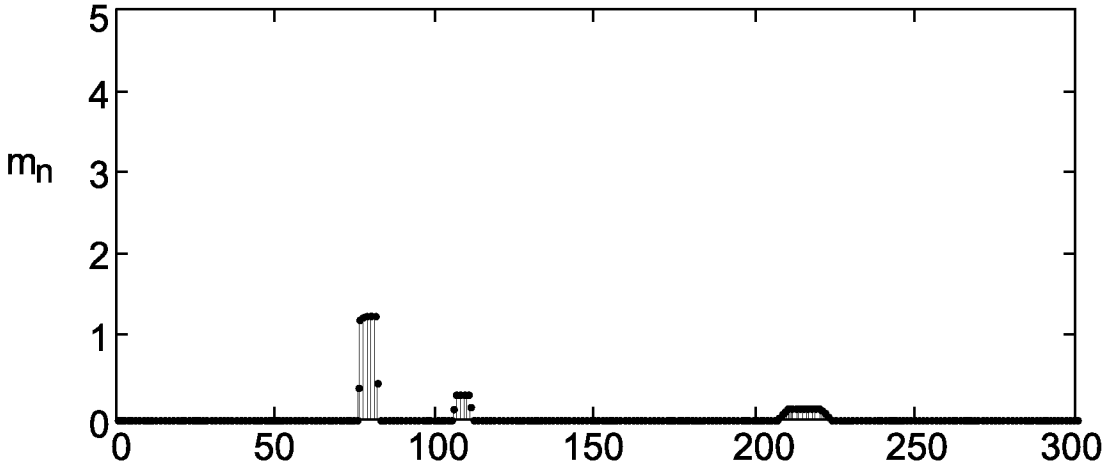

FIG. 3B

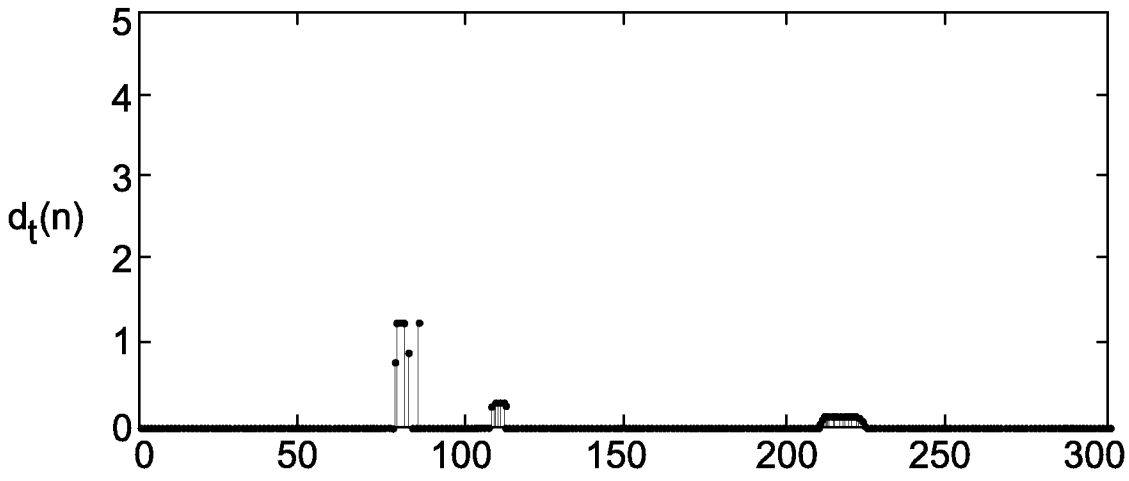

FIG. 3C

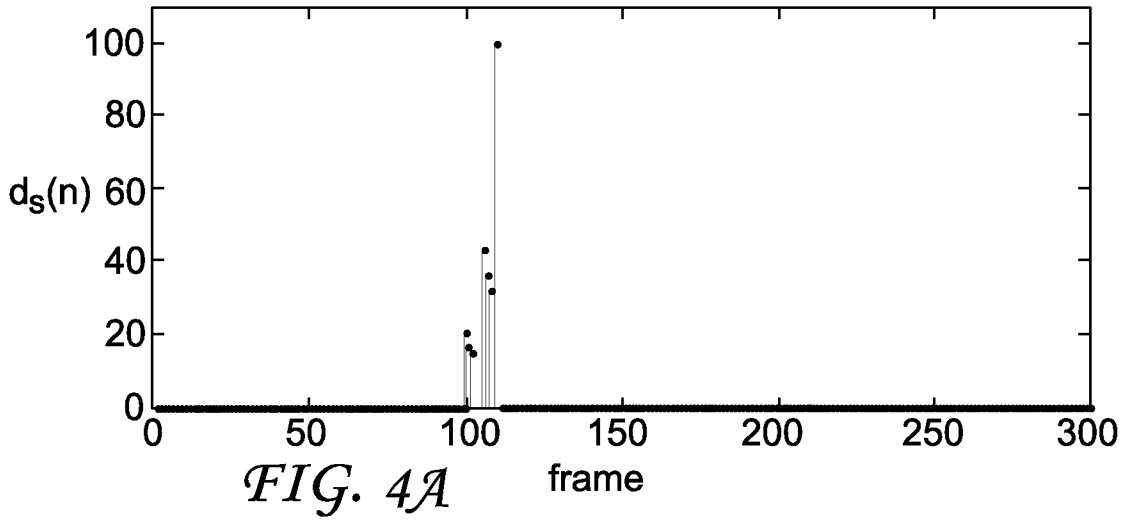

FIG. 4A

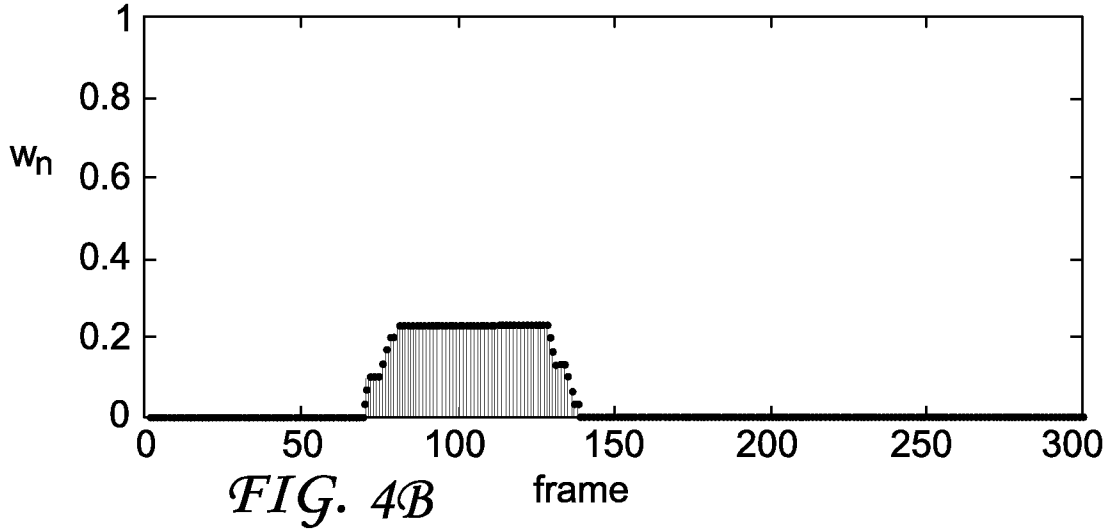

FIG. 4B

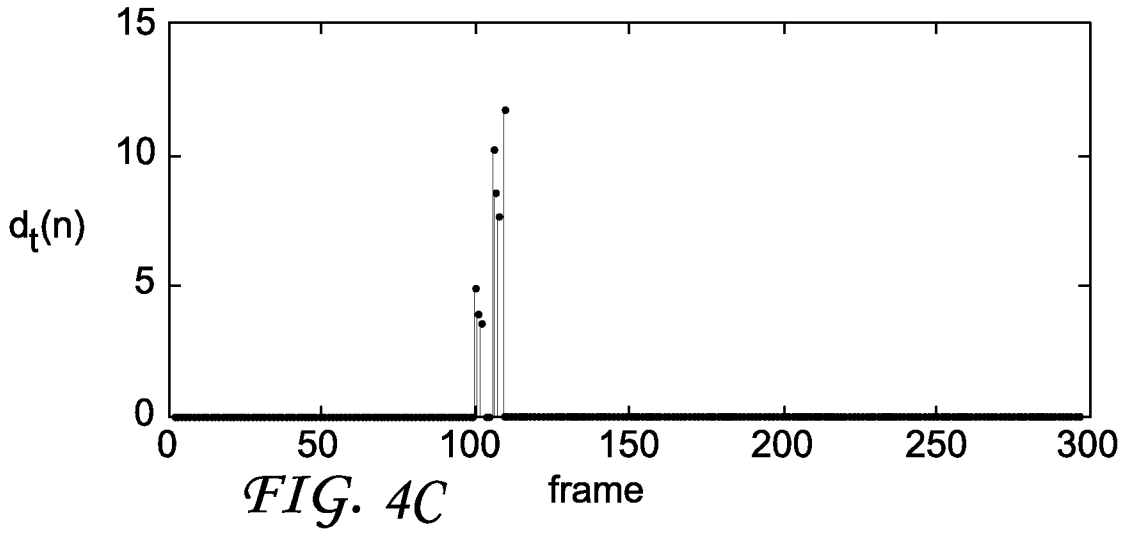

FIG. 4C

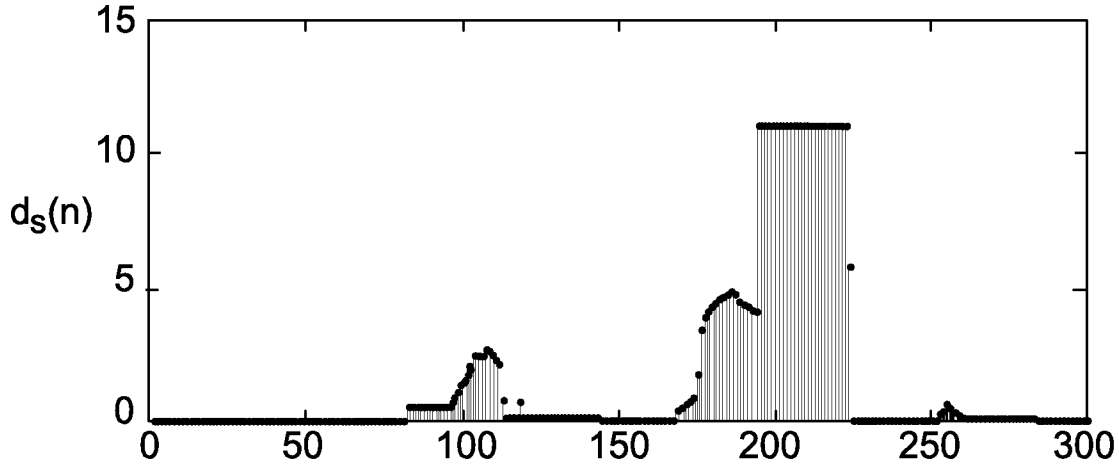

FIG. 5A

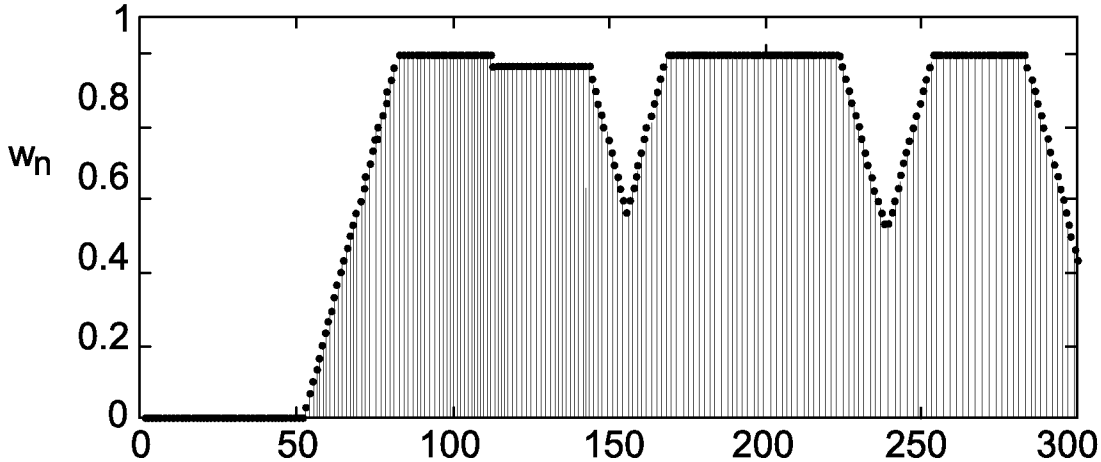

FIG. 5B

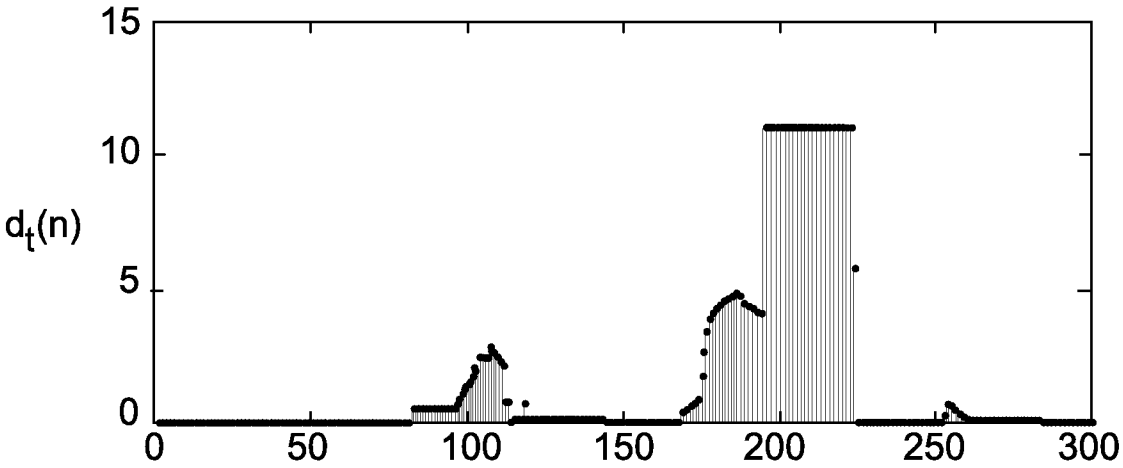

FIG. 5C

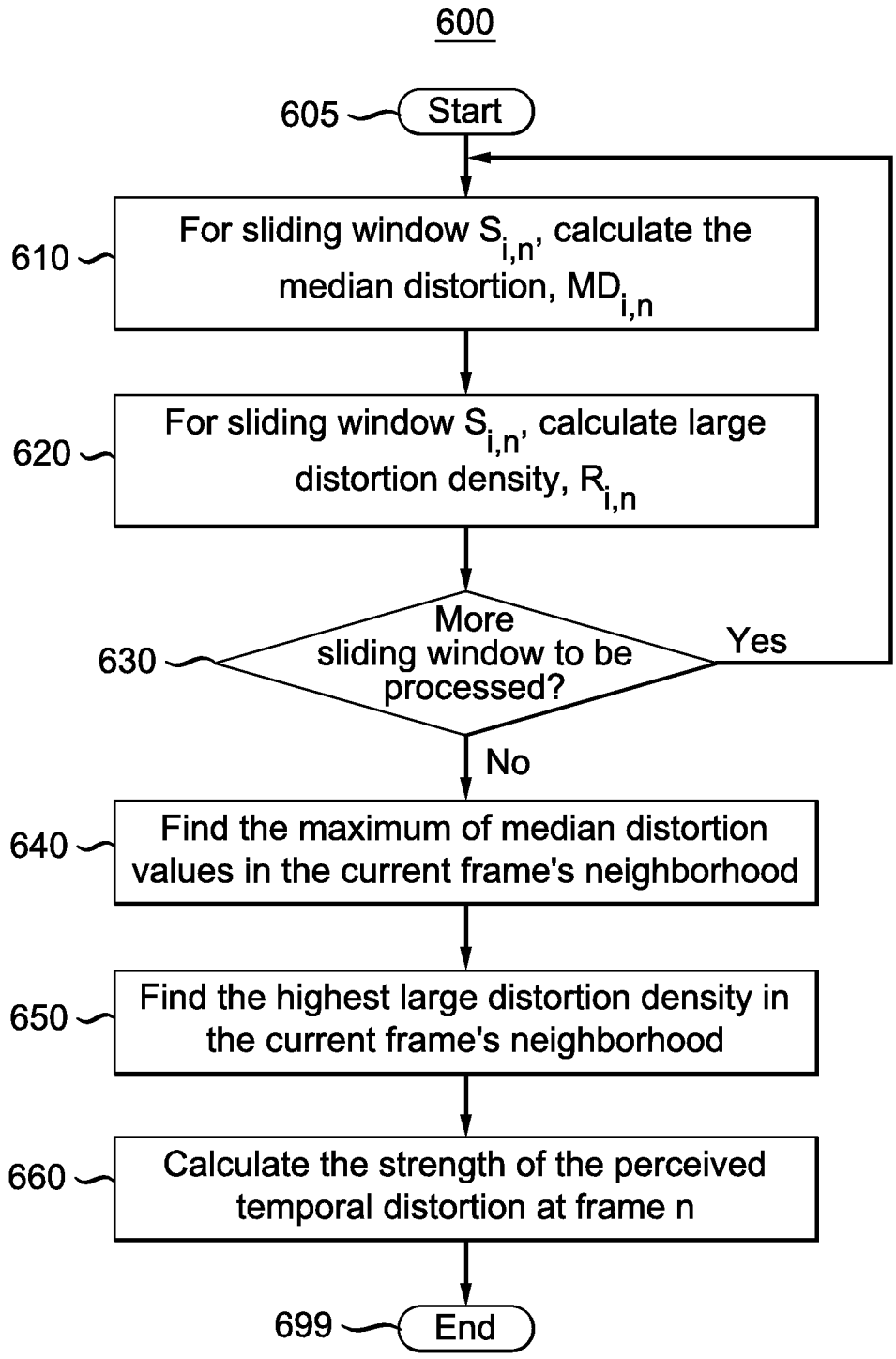
600
605
Start
610
For sliding window $S_{i,n}$, calculate the median distortion, $MD_{i,n}$
620
For sliding window $S_{i,n}$, calculate large distortion density, $R_{i,n}$
630
More sliding window to be processed?
Yes
No
640
Find the maximum of median distortion values in the current frame's neighborhood
650
Find the highest large distortion density in the current frame's neighborhood
660
Calculate the strength of the perceived temporal distortion at frame n
699
End

FIG. 6

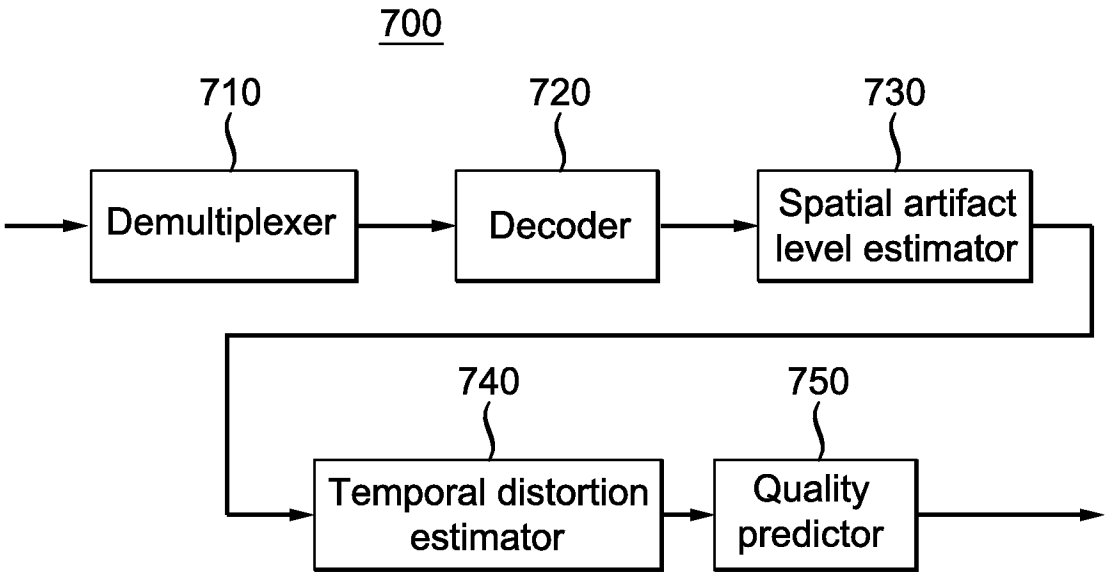
700
710
720
730
Demultiplexer
Decoder
Spatial artifact level estimator
740
750
Temporal distortion estimator
Quality predictor

FIG. 7

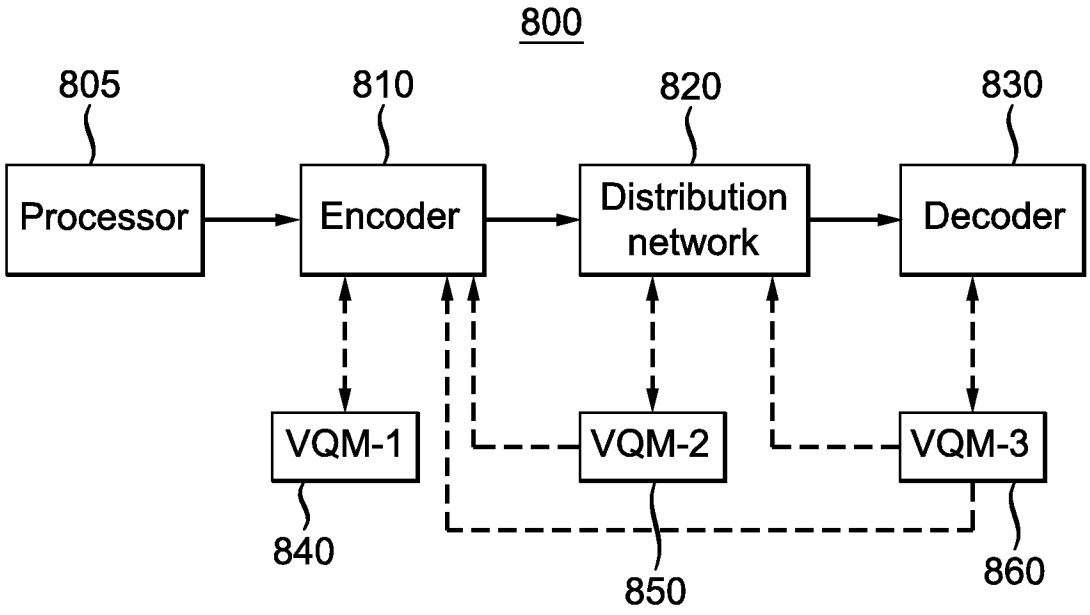
800
805
810
820
830
Processor
Encoder
Distribution network
Decoder
VQM-1
VQM-2
VQM-3
840
850
860

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/077082

A. CLASSIFICATION OF SUBJECT MATTER

H04N 17/00 (2006.01) i

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: H04N17/-; H04N7/-

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, CNPAT, CNKI: visual, quality, distortion, spatial ,temporal, video sequence, sliding window ,artifact level

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101695141 A(UNIV. ZHEJIANG) 14 April 2010 (14.04.2010) claim 1, description, paragraphs [0010], [0011], [0017], [0018]-[0020], [0027], and figures 1-2	1-15
A	CN 101572830 A (THOMSON LICENSING) 04 November 2009 (04.11.2009) the whole document	1-15
A	CN 101635846 A (HUAWEI TECHNOLOGIES CO., LTD.) 27 January 2010 (27.01.2010) the whole document	1-15
A	US 2007/0127819 A (SAMSUNG ELECTRONICS CO., LTD.) 7 June 2007 (07.06.2007) the whole document	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	Date of mailing of the international search report
22 October 2013 (22.10.2013)	**21 Nov. 2013 (21.11.2013)**
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 Gong,Jinling Telephone No. (86-10)62413269

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

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2013/077082

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101695141 A	14.04.2010	CN 101695141 B	23.05.2012
CN 101572830 A	04.11.2009	JP2009273127 A	19.11.2009
		US2009274390 A1	05.11.2009
		EP2114078 A1	04.11.2009
CN 101635846 A	27.01.2010	WO2010009637 A1	28.01.2010
		IN201004835 P2	18.02.2011
		US2011085605 A1	14.04.2011
		CN101635846 B	25.04.2012
		EP2296379 A1	16.03.2011
US 2007/0127819 A	07.06.2007	US7636454 B2	22.12.2009
		KR20070058983 A	11.06.2007