RATE CONTROL WITH PICTURE-BASED LOOKAHEAD WINDOW

BACKGROUND OF THE INVENTION

[0001] The present invention relates to compression coding of video signals, and more particularly to rate control with a picture-based lookahead window for dual-pass compression encoding/transcoding.

[0002] Two of the international standards for digital video compression are known as MPEG2 (H.262) and H.264 (MPEG 4 part 10). There are several other standards to which the invention might be applied, such as H.261, MPEG1 and H.263, but the following description of an embodiment of the invention refers mainly to MPEG2 and H.264 so we will not discuss the other standards.

[0003] An uncompressed video stream can be described as a consecutive series of pictures, or frames. An individual frame describes a particular setting at a particular instant in time. A scene is a series of frames describing the same setting at consecutive moments in time. The second frame of a scene shows the same setting as the first frame, slightly further in the future. MPEG standards take advantage of that repetition of information using what is known as temporal encoding. In accordance with an MPEG video compression standard, such as MPEG2, the encoder divides a video stream into sets of related pictures, known as groups of pictures (GOPs). Each frame within a GOP is labeled by the encoder as an intraframe, a predicted frame or a bi-directional frame. An intraframe (I type frame) is encoded using only information from within that frame. No temporal encoding is used to compress the frame. A predicted (P type) frame is encoded using information from within the frame and uses an earlier I frame or P frame as a reference for temporal compression. I and P frames are referred to as anchor frames. A bi-directional (B type) frame is encoded using information

from within the frame and can also use information from at least one earlier anchor frame and at least one later anchor frame. Within a GOP, the I frame is generally the most complex, followed by the P frames, and the B frames typically have the least amount complexity.

[0004] In one conventional implementation of MPEG2, each GOP (which is referred to herein as a standard GOP) has a period (N) of 15 pictures, or frames, and includes only one I type picture, which is the first picture in the GOP. The fourth, seventh, tenth and thirteenth pictures are P type pictures and the remaining ten pictures are B type pictures. Thus, each standard GOP is composed of 5 sub-groups of 3 pictures each. Each sub-group is made up of an anchor picture and two B type pictures. The spacing between anchor pictures, in this case 3, is known as the intra-period (M) of the GOP. Thus a standard GOP, in display order, will appear as:

I B B P B B P B B P B B P B B

[0005] In this conventional implementation of MPEG2, the standard GOP is closed, i.e. it does not make any predictions based on frames outside of the GOP.

[0006] A video stream is further sub-divided in MPEG standards by defining a frame as being made up of a series of macro-blocks (MBs). A macro-block contains all the information required to display an area of the picture representing 16x16 luminance pixels.

[0007] MPEG2 and H.264 specify the syntax of a valid bit stream and the ways in which a decoder must interpret such a bit stream to arrive at the intended output, the output being uncompressed digital video. The MPEG standards do not, however, specify an encoder. An encoder is defined as any device, hardware or software, capable of outputting a bit stream that will produce the desired output when input to an MPEG compliant decoder.

[0008] In a typical application of an encoder, an uncompressed video signal is input to the encoder and encoded in accordance with the applicable compression standard and the newly encoded signal is output from the encoder, received by a decoder and decoded for viewing. In order to accommodate variation in the rate at which data is received by the decoder, the decoder may include a buffer that receives the encoded data and provides data to the decoding process. The encoder must ensure that the encoded signal is being output at such a rate that the decoder may be continuously decoding the encoded signal and transmitting the decoded signal. If the encoder transmits the signal too slowly, there will be gaps in the data transmitted by the decoder as the decoder waits for data from the encoder. If the encoder transmits the signal too quickly, the decoder may not be able to keep up, causing buffer overflow at the decoder and unacceptable information loss. The process of managing an encoder's rate of transmission is known as rate control. The encoder keeps track of decoder buffer fullness by use of a virtual buffer.

[0009] A simplistic method of rate control would be to assign a set number of bits to each picture, or frame, of the video signal to be encoded. However this is not efficient, as the number of bits used to encode every picture in the video stream must then be large enough to accommodate the most complex frame possible, when in fact a simple frame, such as a picture of a blue sky, will require fewer bits to encode than a complex frame (a picture of cloudy sunrise at the horizon). A measure of a picture's complexity, prior to the picture being encoded, allows the encoder to make a better decision regarding how many bits should be used to encode the picture. This method can be further improved if the encoder has knowledge of the complexity of frames that it will be encoding in the future. Consider a video stream that begins showing a clear blue sky, and then pans down to show a

sunset. The initial frames will have a low measure of complexity and thus will be encoded using a relatively small number of bits. The later frames, however, contain much more complex information and will require a larger number of bits to encode. If, while deciding how many bits to allocate to the initial, simple, video frames, the encoder is made aware that it will soon need to allocate a large number of bits to encoding more complex frames, the encoder may further reduce the number of bits used to encode the simple frames and avoid or reduce the risk of overflowing a downstream decoder.

[0010] Another effective method of controlling the number of bits used to encode a picture is by modifying the quantization step size dynamically for each MB within the frame. For an MB of generally uniform color and intensity only a small number of possible pixel values are necessary and thus less bits will be required to describe it. The inverse is true for an MB containing a wide variety of color and intensity values since the encoder will have to describe wider range of pixel values. In accordance with this approach each MB is assigned a quantization scale factor (Mquant) that is used to modify quantization step size.

[0011] During the development of the MPEG2 standard it became necessary to design a generic rate control and quantization methodology with which to test bit-stream syntax, decoder designs and other aspects of the standard. This methodology was known as the Test Model and it evolved as the development of MPEG2 continued. The fifth and final version of the model (TM-5) was produced with the freezing of the MPEG2 standard. TM-5 is divided into three primary steps: (a) target bit allocation; (b) rate control; and (c) adaptive quantization.

a) Target Bit Allocation

[0012] As discussed above, the amount of bits allocated to a certain picture for encoding is a function of that picture's complexity, relative to other pictures. For a particular GOP, a complexity weight factor is assigned for each picture type (X_{I2}, X_{P2} and X_{B2} for I, P and B type pictures respectively). X_{I2}. X_{P2} and X_{B2} represent the complexity measure for I, P, B pictures and may be calculated as:

$$X_{I2} = S_{I2}Q_{I2}$$

$$X_{P2} = S_{P2}Q_{P2}$$

$$X_{B2} = S_{B2}Q_{B2}$$

where S_{I2}, S_{P2} and S_{B2} are the number of bits for each picture and Q_{I2}, Q_{P2} and Q_{B2} are the average quantization parameters for all MBs in each picture (see below).

[0013] In TM-5, bit allocation for a picture is targeted based on how much of the bit space allocated for the GOP is remaining, the type of picture being encoded, and the complexity statistics of recently encoded pictures of the same type. The target bit allocation is the number of bits TM-5 anticipates will be necessary to encode the frame.

b) Rate Control

[0014] If there is a difference between the target bit allocation (B_{tar}) and the actual number of bits required to encode a picture (B_{act}), than there is a risk of under or over filling TM-5's virtual buffer. The virtual buffer tracks the fullness of the decoder's buffer on an MB by MB basis while a picture is being encoded. Encoding all the MBs up to, but not including, the jth MB should use a certain portion of the total targeted bits. This portion is equal to B_{tar} multiplied by the number of MBs already encoded (j-1) and divided by the total number of MBs in the picture (MB_cnt). The number of bits actually generated by encoding up to, but not including,

the jth MB is equal to $B_{(j-1)}$.The delta between the targeted and generated number of bits represents a change in the fullness of the virtual buffer after each MB is encoded (d_j) and is calculated prior to encoding the jth MB.

$$d_j = d_0 + B_{(j-1)} - B_{tar}*(j-1)/MB_cnt$$

where d_0 equals the fullness of the virtual buffer at the beginning of the current picture.

[0015] If the virtual buffer begins to overflow, the quantization step size increases, leading to a smaller yield of bits for the subsequent MBs. Similarly, if the virtual buffer begins to underfill, the quantization step size is decreased, leading to a larger yield of bits for subsequent MBs. This measure of the virtual buffer fullness is used to generate the MB's reference quantization number (Q_j).

c) Adaptive Quantization

[0016] The macro-block quantization step size is further modulated as a function of spatial activity (act_j). The macro-block is divided into four 8x8 sub-blocks and the spatial activity is measured for each sub-block. The smallest of the four measurements is then normalized (N_act_j) against the average spatial activity (avg_act) of the previously encoded picture. The minimum spatial activity measurement is used because the quality of a macro-block is no better than its sub-block of highest visible distortion.

$$N_act_j = (2*act_j + avg_act)/(act_j + 2*avg_act)$$

[0017] The product of a MB's normalized spatial activity and its reference quantization parameter gives the MB's quantization scale factor ($Mquant_j$).

$$Mquant_j = Q_j*N_act_j$$

Generally, the encoding algorithms recommended by newer compression standards are more efficient, but they are

usually more complicated to implement. With the fast growth in computation speeds of central processing units (CPUs) and digital signal processing (DSP) chips, implementation of more and more sophisticated algorithms has become practically feasible. Video encoders/decoders (codecs) built based on newer standards eventually replace those built based on older standards in applications where specifications overlap, such as for bit-rate, resolution, etc. This replacement procedure takes a long period of time, since it is expensive to replace older video codecs with newer ones. Another reason older codecs continue to be used is that many video streams have already been compressed with the older algorithms, and may easily be decompressed by the older codecs. However where high coding efficiency is desired, there arises the mixed use of both older and newer codecs. In some applications it is desirable to re-transmit video streams compressed with an older codec at a new bit-rate that is lower than the older codec can achieve for the same video quality. Therefore to obtain higher compression efficiency a transcoder having mixed codecs (an older decoder and a newer encoder) is used. One good example is a transcoder that converts MPEG2 compressed video streams to H.264 compressed video streams.

[0018] It is recognized by the digital compression industry that dual-pass encoding with a lookahead window provides higher coding efficiency than single-pass encoding. For the emerging, more sophisticated, compression technologies, even single-pass encoding is expensive and the cost of dual-pass encoding is much higher than that of single-pass encoding. Using two sophisticated codecs for encoding/transcoding in a dual-pass architecture raises the cost of the encoder/transcoder by almost an order of magnitude over the older technology codecs.

[0019] What is desired is the ability to achieve higher coding efficiency for minimal cost in an encoder/transcoder using mixed codecs.

BRIEF SUMMARY OF THE INVENTION

[0020] In accordance with a first aspect of the invention there is provided a method of encoding frames of an uncompressed digital video stream, each frame having a level of complexity, said method comprising analyzing a first frame of the uncompressed digital video stream with a first algorithm to measure a first value of the first frame's complexity and to assign the first frame's picture type, estimating a second value of the first frame's complexity using the first measured value as a parameter, and encoding the first frame with a distinct second algorithm employing the second value of the first frame's complexity and the first frame's picture type as parameters.

[0021] In accordance with a second aspect of the invention there is a method of transcoding frames of a compressed digital video stream, each frame being encoded according to a first encoding algorithm and having a level of complexity, said method comprising decoding a first frame of the compressed digital video stream with a first decoding algorithm to produce a decoded version of the first frame and to measure a first value of the first frame's complexity and to determine the first frame's picture type, estimating a second value of the first frame's complexity using the first value as a parameter, and encoding the decoded version of the first frame with a distinct second encoding algorithm employing the second value of the first frame's complexity and the first frame's picture type as parameters.

[0022] In accordance with a third aspect of the invention there is an apparatus for encoding an uncompressed digital video input stream composed of a succession of frames, each

frame having a plurality of characteristics associated therewith, said apparatus comprising an extraction means for receiving the succession of frames of the uncompressed digital video input stream and employing a first method to obtain measured values for the plurality of characteristics of a frame of the input stream and to assign a picture type to the frame, a delay means for receiving the succession of frames of the input stream and outputting the frames in delayed fashion relative to the frames of the input stream, a value storage means for storing the measured values and the picture type of a frame in the delay means, and an encoding means for receiving a frame from the delay means and encoding the frame, the encoding means being responsive to a measured value stored in the value storage means for adjusting the size of the encoded version of the frame.

[0023] In accordance with a fourth aspect of the invention there is an apparatus for transcoding a compressed digital video input stream composed of a succession of encoded frames, each encoded frame having a plurality of characteristics associated therewith, said apparatus comprising a decoding means for receiving the succession of encoded frames of the compressed digital video input stream and employing a first method to obtain a succession of decoded frames and measured values for the plurality of characteristics of a decoded frame and to assign a picture type to the decoded frame, a delay means for receiving the succession of decoded frames of the input stream and outputting the decoded frames in delayed fashion relative to the encoded frames of the input stream, a value storage means for storing the measured values and the picture type of a decoded frame in the delay means, and an encoding means for receiving a decoded frame from the delay means and encoding the frame, the encoding means being responsive to a measured

value stored in the value storage means for adjusting the size of the encoded version of the frame.

[0024] An embodiment of the present invention provides rate control with a picture-based lookahead window for encoders/transcoders having mixed codecs in a dual-pass compressed video architecture. In a transcoder, where the input video signal is a compressed video signal, statistics are extracted by using a simple compression decoder to produce the statistics from the compressed video signal; and in an encoder, where the input video signal is an uncompressed video signal, statistics are extracted by using a simple compression encoder to generate the statistics from the uncompressed video signal. A trans-factor is calculated for a current picture based on previous pictures in a sliding "past" window to predict the complexity of the current picture, the trans-factor being a ratio of global complexity measures for the simple compression standard versus a sophisticated compression standard. Bits for the current picture are then allocated based on the complexity of future pictures in the lookahead or "future" window. If future pictures are difficult to encode, then less bits are allocated to the current picture, and vice versa. This is effective for a scene change. Because the lookahead window takes into account the statistics of future pictures, i.e., pictures that have not yet been compressed according to the sophisticated compression standard, a more reasonable bit allocation and better quality is achieved. After encoding the current picture according to the sophisticated compression standard, the actual bits, the picture complexity and the trans-factor for the encoded picture are updated as the past and lookahead windows are shifted by one picture, i.e., the encoded picture moves into the past window and out of the lookahead window as a new picture is loaded into the lookahead window.

[0025] The objects, advantages and other novel features of the present invention are apparent from the following detailed description when read in conjunction with the appended claims and attached drawings.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0026] Fig. 1 is a block diagram view of a dual-pass encoder/transcoder architecture implementing rate control with a picture-based lookahead window according to the present invention.

[0027] Fig. 2 is a flow chart view of a rate control algorithm according to the present invention.

[0028] Fig. 3 is a conceptual view of virtual sliding windows according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0029] Fig. 1 illustrates an encoder/transcoder having a simple compression decoder **12** for receiving and decoding a compressed video stream encoded according to a simple compression standard, such as MPEG2, to produce an uncompressed video signal and related statistics. Alternatively a simple compression encoder **14** receives an uncompressed video stream to generate related statistics. The statistics are input to a lookahead window module **18** for processing by a rate control algorithm, described below, while the uncompressed video signal in either configuration (transcoder or encoder) is input to a storage and delay module **16**. The storage and delay module is a buffer memory that receives, delays and outputs the uncompressed video stream. The lookahead window module contains the statistics for each picture in the storage and delay module **16**, for example, the number of bits for the picture, the picture type and the average quantization step size over all the macro-blocks for the picture. The lookahead window module **18**

generates bit allocation data from the statistics for use by a sophisticated compression encoder **24**, such as an H.264 encoder, in determining rate control for the sophisticated encoding process. The storage and delay module compensates for the time required for the lookahead window module **18** to generate bit allocation data.

[0030] The delayed uncompressed video stream from the storage and delay module **16** is input to an adaptive pre-filter **20** to produce a filtered uncompressed video stream. The filter may be a low-pass filter that attenuates high spatial frequencies in the images represented by the uncompressed video stream and thus serves to "blur" the uncompressed video stream so that it is easier to compress, i.e., is less complex and so requires fewer bits to compress. The strength of the filtering may depend on a threshold or cut-off frequency above which spatial frequency components are attenuated and on the degree to which high spatial frequencies are attenuated.

[0031] Both the delayed uncompressed video stream from the storage and delay module **16** and the filtered, uncompressed video stream are input to a switch **22** which selects one of the streams. The selected uncompressed video stream from the switch **22** and the bit allocation data from the look ahead window module **18** are input to the sophisticated compression encoder **24** to produce a compressed video stream according to a sophisticated compressed video standard, such as H.264 (MPEG4, Part 10). The sophisticated compression encoder **24** also provides a control signal to the adaptive pre-filter **20** and to the switch **22** that determines the "strength" of the filtering and which uncompressed video stream is to be encoded. The strength of the filtering may be implemented as different filtering levels or may be continuous. The adaptive pre-filter **20** may be switched off or set to a low strength for minimum filtering when the filtered uncompressed

video stream is not selected for encoding by the sophisticated compression encoder **24**.

[0032] By using a simple encoder/decoder **12/14** instead of a sophisticated encoder/decoder at the input, the implementation cost is reduced close to that of a single-pass sophisticated codec. However the information on complexity estimation for pictures in the lookahead window module **18** is not exactly the desired information for the sophisticated compression encoder **24**. For example, a P-type picture needs high bit-rate for motion compensation in simple (MPEG2) compression encoding if its corresponding original picture was recorded during light off/on/off transition time. On the other hand this P-type picture may be a simple picture for the sophisticated (H.264) encoder. Despite this deficiency, a correlation of picture complexity estimation can be made, based on both the simple and sophisticated compression standards. In most cases a picture or a group of pictures (GOP) that is relatively complicated/simple for the simple compression standard is also relatively complicated/simple for the sophisticated compression encoder **24**. The complexity statistics still indicate important relationships among pictures and macro-blocks (MBs), with the error being tolerable. Therefore compared to single-pass sophisticated coding, the pseudo dual-pass sophisticated coding is superior in video coding efficiency with only a slightly higher implementation cost.

[0033] The statistics of picture complexity are used for:

- estimation of a bit-rate target and selection of quantizer step sizes of macro-blocks for a current picture before second pass encoding; and
- control of the strength of the adaptive pre-filter **20** for a current GOP which includes the current picture before second pass encoding.

[0034] The larger the volume of statistics that are available to compute the bit allocation provided to the compression encoder **24**, the better the video quality performance of the encoder/transcoder. Therefore the storage and delay module **16** stores multiple uncompressed images. Each image of the uncompressed video stream will ultimately be encoded by the sophisticated compression encoder **24** as an I, P or B type picture. The type of picture (I, P or B) that a given uncompressed image will be encoded as is based on the statistics provided to the lookahead window module **18**. Therefore, even though the images stored in the module **16** are not encoded, it will be convenient to refer to these images as I, P or B type pictures. The number of images stored by the module **16** is limited by the size of the memory and an allowed maximum delay. A storage length corresponding to at least two GOPs of the input video signal is desired. For the purposes of this description, it is assumed that the storage and delay module is designed to contain two standard GOPs of fifteen pictures each.

[0035] The lookahead window module **18** sets a bit-rate target for the current picture being encoded based on the received statistics, which include picture types (I, P or B), picture size (in bytes), and average quantizer step sizes at picture levels.

[0036] The picture complexity in the sense of encoding is not the same for the two different compression standards. A P-type picture may be complicated and need high bit-rate for motion compensation in MPEG2 encoding if its corresponding original picture was recorded during flash light off/on/off transition time. On the other hand this P-type picture may be a simple picture to an H.264 encoder which is able to select one out of up to six reference pictures for motion prediction, and one of the references may be strongly correlated with this P-type picture, as indicated above.

[0037] In addition to setting the bit-rate target, the statistics of picture complexity obtained by the lookahead window module **18** may also be used for generating the control signal for the adaptive pre-filter **20** to control the strength of the low-pass filtering. If the rate control information indicates that the current picture is a difficult picture which needs more bit-rate to encode, the strength of the adaptive pre-filter **20** may be increased so that the picture is heavily low-pass filtered, i.e., becomes softer and easier to encode. The sophisticated compression encoder **24** employs the switch **22** to select either the delayed uncompressed video signal output from the storage and delay module **16** or the filtered video signal output by the adaptive pre-filter **20** based on the rate control information and on the virtual buffer fullness of the sophisticated compression encoder **24**. For example, if the virtual buffer is approaching full and the rate information indicates that the current picture for encoding requires more bits than are available in the virtual buffer, then the amount of pre-filtering is increased so that the virtual buffer does not overflow and the filtered uncompressed video is the video signal that is encoded. If there is no danger of virtual buffer overflow, then the current picture is slightly filtered or not filtered at all. In the latter event the uncompressed video signal from the storage and delay module **16** is used as the input for encoding. However, frequently and abruptly changing the filter strength and/or switching between the uncompressed video signal and the filtered uncompressed video signal within a GOP may lead to a motion compensating residue signal for P and B pictures. This is avoided by smoothly controlling the pre-filter **20** within a GOP. If a picture is filtered, any other pictures which use it as reference should also be filtered with at least the same filter strength.

[0038] The rate control algorithm used, by way of illustration, is based on the Test Model 5 (TM5) specification. TM5 takes a complexity measure to allocate target bits for each picture and then sets a quantization parameter for each MB based on the fullness of the virtual buffer. In the transcoder configuration all of the information about the input video signal is available from the encoded compressed video stream via the decoder **12**, especially the statistics about the complexity of the input content. In the encoder configuration all the information about the input video signal is available from the uncompressed video stream via the simple encoder **14**, especially the statistics about the complexity of the input content. The rate control algorithm includes two parts:

1. Take "past" statistics for complexity prediction.
2. Take "future" statistics for bit allocation.

Both processes are adaptive and a past sliding window and a future sliding window are maintained to update the statistics after each picture is encoded. Note that the past sliding window is located in the lookahead window **18** and the future sliding window is located in the sophisticated compression encoder **24**. Contrary to prior applications that used sliding windows which increment in terms of GOPs, the sliding windows of the present invention are picture-based, and move forward after encoding each picture.

[0039] The rate control algorithm has four steps: (a) statistics extraction; (b) complexity prediction; (c) bit allocation; and (d) statistics update.

(a) Statistics Extraction

[0040] When transcoding from an MPEG2 variable bit-rate (VBR) stream to an H.264 constant bit-rate (CBR) stream or encoding an uncompressed video stream to an H.264 CBR stream the following information is collected:

1. Average quantization parameters (quantization step size) for each picture.

2. Output bits for each picture.

3. Picture type (I, P, B) for each picture.

Items 1 and 2 are used for calculation of the input video's complexity, while item 3 records the picture type that is used by the sophisticated compression encoder **24**.

(b) Complexity Prediction

[0041] Complexity prediction is to predict the complexity of a current picture from a prior simple/sophisticated (MPEG2/H.264) complexity ratio and the input complexity of the current picture. In TM5 the current picture's complexity is predicted by that of the previous picture of the same type. In an embodiment of the invention, the current picture's complexity is predicted on the basis of the complexity of all pictures of the same type in the past window. However since the statistics are based on a simple encoding format an adjustment to the algorithm in the form of a scaling factor, referenced here as trans-factor, is introduced to take into account the difference between the sophistication of the two standards and/or the two bit-rates. The trans-factor is calculated as the average of previous simple/sophisticated ratios and is updated after the encoding of each picture. Because of the different properties of different picture types, the trans-factor is calculated independently for each picture type.

[0042] The complexity prediction algorithm has two steps:

[0043] 1. Calculate a current trans-factor for a current picture by averaging previous trans-factors

[0044] At the beginning of a video sequence to be encoded/transcoded there are three initial values for the trans-factor, corresponding to the three picture types (I, P, B) respectively. The average trans-factor over a past sliding

window is generally better than that of only one picture and takes into account pictures that have already been encoded by the sophisticated encoder **24** and are within the past window.

[0045] A GOP can be described as containing N_I I type pictures, N_P P type pictures and N_B B type pictures. For a standard GOP, as described above:

$N = 15$

$M = 3$

$N_I = 1$

$N_P = (N/M) - N_I = (15/3) - 1 = 4$

$N_B = N - N_I - N_P = 15 - 1 - 4 = 10$

[0046] The storage and delay module **16** contains W GOPs. For this discussion, it will be assumed that the module **16** is designed to store 2 standard GOPs. W_I, W_P and W_B represent the total number of I, P and B type pictures respectively in the storage and delay module **16**.

$W_I = WN_I = 2*(1) = 2$

$W_P = W(N_P) = 2*(4) = 8$

$W_B = W(N_B) = 2*(10) = 20$

[0047] The trans-factor T_{cur} of a picture in the uncompressed video stream is calculated by averaging the trans-factors of the previous W_{type} pictures of the same type (I, P or B), the number of previous trans-factors that are averaged being equal to the total number of pictures of that type in the storage and delay module **16** (W_I, W_P, W_B).

$$T_{Icur} = (1/W_I)\sum_{i=j-WI+1->j} T_I(j) = (1/2)\sum_{i=j-1->j} T_I(j)$$

$$T_{Pcur} = (1/W_P)\sum_{i=j-WP+1->j} T_B(j) = (1/8)\sum_{i=j-7->j} T_P(j)$$

$$T_{Bcur} = (1/W_B)\sum_{i=j-WB+1->j} T_B(j) = (1/20)\sum_{i=j-19->j} T_B(j)$$

where **j** is the picture number of the current picture.

[0048] For a storage and delay module 16 that contains two standard GOPs, as described above, two, eight or twenty previous trans-factors are averaged for an I, P or B type picture respectively.

[0049] 2. Predict current picture's complexity

[0050] For I and P type pictures, the current picture's sophisticated, or MPEG4, complexity (X_{I4}, X_{P4} or X_{B4}) is then predicted from the picture's simple, or MPEG2, complexity (X_{I2}, X_{P2} or X_{B2}) using the updated trans-factor (T_{cur}) to appropriately scale the simple complexity factor. The trans-factors for B type pictures are further adjusted by a weight factor (K_{B4}) to take into account the different quality requirements for different picture types. This weight factor has been empirically determined and is a function of the ratio of the current GOP's I type simple complexity and the average simple complexity of the GOP's B type pictures.

$$X_{I4} = X_{I2}/T_{Icur}$$

$$X_{P4} = X_{P2}/T_{Pcur}$$

$$X_{B4} = X_{B2}/(T_{Bcur}*K_{B4})$$

[0051] K_{B4} is larger for a well-predicted sequence, i.e. a sequence without fast motion, and is smaller for a sequence with fast motion. K_{B4} is set adaptively after encoding each GOP in accordance with the ratio X_I/X_B where X_I and X_B are the average simple complexity of all I and B pictures in the current GOP.

Table 1. Empirically Determined Values for K_{B4}

X_I/X_B	K_{B4}
>5.0	1.35
2.0-5.0	1.2
<2.0	1.1

[0052] In principle, the sophisticated complexity X_{P4} for a P type picture may also adjusted by a weight factor (K_{P4}), but it has been found that this is not necessary in practice.

c) Bit Allocation

[0053] Bit allocation may be based on GOP-layer and picture-layer. Picture-layer breaks the GOP boundary and performs better than GOP-layer. This is particularly effective for scene changes in the video signal. Bit allocation has two steps.

[0054] 1. Allocate target bits for current (kth) picture

[0055] The target size (T_W), in bits, for all the pictures currently referenced in the sliding lookahead window is calculated based on the number of pictures in the window (W_F), the constant bit rate (R), in bits per second, and the picture rate (F), in pictures per second.

$$T_W = W_F(R/F)$$

[0056] Then the targeted number of bits ($B_{4_tar}(k)$) to be allocated for the kth picture is calculated by multiplying T_W by the ratio of the current picture's complexity factor to the complexity factor of all the pictures in the sliding lookahead window.

$$B_{4_tar}(k) = T_W*(X_4(k)/\sum_{i=0->WF-1}X_4(k+i))$$

[0057] This calculation essentially identifies the proportion of the target size (T_W) that should be used for the current picture. The size of the current picture when encoded by the complex encoding algorithm is not permitted to be larger than the size of the current picture when encoded by the simple compression algorithm ($B_2(k)$). Thus, the size of the current picture when encoded is clamped to $B_2(k)$. If $B_{4_tar}(k)$ does exceed $B_2(k)$, the number of bits targeted for

the kth picture is still $B_{4_tar}(k)$. However it is already known that the smaller number $B_2(k)$ will be the upper limit of bits used when the encoding actually takes place. Thus there is a known surplus of bits in the target window. The target window size is then modified to take the extra bits into account.

$$T_W(k+1) = T_W(k) + B_{4_tar}(k) - B_2(k)$$

[0058] 2. Adaptive quantization and encoding (TM5)

[0059] Before encoding MB_j the fullness of the virtual buffer is computed for I, P, B independently:

$$d_j = d_0 + B_{j-1} - (T*(j-1))/MB_cnt$$

where B_j is the number of bits generated by encoding all MBs in the picture up to and including j, *MB_cnt* is the number of MBs in the picture, T is the constant bit rate (CBR) per picture, d_0 is the initial fullness of the virtual buffer and d_j is the fullness of the virtual buffer at MB_j. The reference quantization parameter Q_j is then computed for MB_j

$$Q_j = d_j*51/r$$

where the reaction parameter, r, is

$$r = 2*R/F$$

Adaptive quantization:

[0060] The spatial activity for MB_j from four luminance picture-organized sub-blocks (n = 1 . . . 4) and four luminance field-organized sub-blocks (n = 5 . . . 8) are computed using the original pixel values

$$act_j = 1 + \min(vblk_1, vblk_2, \ldots, vblk_8)$$

where

$$vblk_n = (1/64)*\sum_{k=1->64}(P_k^n - P_mean_n)^2$$

and

$$P_mean_n = (1/64)*\sum_{k=1->64}P_k^n$$

where P is the pixel gray level.

Then normalize act_j:

$$N_act_j = ((2*act_j) + avg_act)/(act_j + (2*avg_act))$$

where avg_act is the average value of act_j for the last picture to be encoded.

Then adjust $mquant_j$ as:

$$mquant_j = Q_j*N_act_j$$

The final value of $mquant_j$ is clipped to a range [1 . . . 51] and used for the quantization. Delta QP should be clipped to [-26,26], as defined by H.264 semantics. Then encode one MB with $mquant_i$ and repeat this step until all MBs of the current picture are encoded.

[0061] (d). Update picture complexity and trans-factor for just encoded picture

[0062] The picture complexity and trans-factor for the just encoded picture are updated and stored in the sliding past window for use with future pictures.

[0063] 1. Trans-factor is defined as the ratio of "global complexity measure" of corresponding simple and sophisticated compression standards pictures.

$$T_I[current_picture_SN] = X_{I2}/X_{I4}$$

$$T_P[current_picture_SN] = X_{P2}/X_{P4}$$

$$T_B[current_picture_SN] = X_{B2}/X_{B4}$$

where X_{I4}, X_{P4} and X_{B4} represent the complexity measure for the I, P, B picture of the output, sophisticated compression standard (H.264) stream:

$$X_{I4} = S_{I4}Q_{I4}$$

$$X_{P4} = S_{P4}Q_{P4}$$

$$X_{B4} = S_{B4}Q_{B4}$$

and the definitions of S_{I4}, S_{P4}, S_{B4}, Q_{I4}, Q_{P4} and Q_{B4} correspond to the definitions of the corresponding quantities under the simple compression standard (MPEG2).

[0064] 2. Since a picture's bit target (B_{4_tar}) is a function of the target size of the look-ahead window we adjust the target size of the look-ahead window, T_W, for the (k+1)th picture after the kth picture is encoded in order to account for any difference between the actual encoded size of the kth picture, S(k), and the average encoded picture size, R/F, where R is the constant bit rate and F is the frame rate.

$$T_W(k+1) = T_w(k) + R/F - S(k)$$

This adjustment is done in order to maintain the proper level of virtual buffer fullness and makes $T_w(k+1)$ a more reasonable target size of the lookahead window for calculating the bit target for the (k+1)th picture. If the input sequence is not infinitely long, at the end of the sequence the lookahead window size W_F and target size T_W both decrease. After encoding each picture the trans-factor, T_I, T_P or T_B, is updated and all the predicted complexity values in the future sliding window are updated as well.

[0065] As illustrated in Fig. 3 the trans-factor to be used in determining the complexity for the current picture to encode is based upon the average of the trans-factors for the same picture type in the past window, while the bit allocation for the current picture is based on the overall complexity of the pictures in the lookahead window. After each picture is encoded, then the actual sophisticated

standard complexity is determined and entered in the past window while the oldest one is shifted out. A new picture's statistics are loaded into the lookahead window to determine a new complexity for the window as the next picture to be encoded becomes the current picture.

[0066] Thus the present invention provides rate control with a picture-based sliding window to simplify transcoding/encoding from a simple compression standard to a sophisticated compression standard by extracting statistics for a video signal using the simple compression standard, by using the extracted statistics and virtual buffer fullness to control a lowpass pre-filter for the uncompressed video signal, and by encoding the filtered or unfiltered uncompressed video signal using a trans-factor which is the ratio of global complexity measures for the simple and sophisticated compression generated standards pictures with a sliding window on a picture-by-picture basis, updating the trans-factor and sliding window for each picture.

[0067] It will be appreciated that the invention is not restricted to the particular embodiment that has been described, and that variations may be made therein without departing from the scope of the invention as defined in the appended claims and equivalents thereof. Unless the context indicates otherwise, a reference in a claim to the number of instances of an element, be it a reference to one instance or more than one instance, requires at least the stated number of instances of the element but is not intended to exclude from the scope of the claim a structure or method having more instances of that element than stated.

CLAIMS

1. A method of encoding frames of an uncompressed digital video stream, each frame having a level of complexity, said method comprising:

analyzing a first frame of the uncompressed digital video stream with a first algorithm to measure a first value of the first frame's complexity and to assign the first frame's picture type,

estimating a second value of the first frame's complexity using the first measured value as a parameter, and

encoding the first frame with a distinct second algorithm employing the second value of the first frame's complexity and the first frame's picture type as parameters.

2. A method according to claim 1, further comprising:

analyzing the first frame with the second encoding algorithm to measure a third value of the first frame's complexity, and

storing the picture type and a complexity ratio of the first frame.

3. A method according to claim 2, comprising, prior to encoding the first frame with the second algorithm,

generating an encoded version of a past frame with the second algorithm, said encoded version of the past frame having a size, and

estimating a value of a future frame's complexity,

and wherein the step of encoding the first frame with the second algorithm further comprises:

determining a target size for the first frame using the size of the past frame and the estimated value of the future frame's complexity as parameters, and

encoding the first frame with the second algorithm employing the first frame's target size as a parameter.

4. A method according to claim 3, wherein the step of encoding the first frame with the second algorithm produces an encoded version of the first frame, and the step of determining the target size of the first frame comprises predicting the size of said encoded version of the first frame.

5. A method according to claim 4, wherein the first frame's complexity ratio is calculated by dividing the first value of the frame's complexity by the third value of the first frame's complexity.

6. A method according to claim 5, wherein the first frame is followed by a plurality of frames of the uncompressed video stream, and the method further comprises:

analyzing each of the plurality of frames of the uncompressed digital video stream with the first algorithm to measure first values of each frame's complexity and to assign each frame's picture type,

estimating a second value of each frame's complexity, using each frame's first value as a parameter, and

encoding each of the plurality of frames with the second algorithm employing the second value and picture type of each frame as parameters.

7. A method according to claim 6, further comprising:

analyzing each of the plurality of frames with the second encoding algorithm to measure a third value of the complexity for each frame, and

storing the picture type and complexity ratio for each of the plurality of frames.

8. A method according to claim 7, wherein the plurality of frames is followed by a second frame, and the method further comprises:

analyzing the second frame with the first algorithm to measure a first value of second frame's complexity and to assign the second frame's picture type,

estimating a second value of the second frame's complexity using the first value as a parameter, and

encoding the second frame with the second algorithm, employing the second value of the second frame's complexity and the second frame's picture type as parameters.

9. A method according to claim 8, comprising estimating the second frame's second value by dividing the second frame's first value by a translation factor.

10. A method according to claim 9 comprising calculating the second frame's translation factor by averaging the stored complexity ratios associated with a sub-set of the plurality of frames, the sub-set being the frames that have been encoded by the second algorithm and have the same picture type as the second frame.

11. A method according to claim 10, further comprising:

analyzing the second frame with the second encoding algorithm to measure a third value of the second frame's complexity, and

storing the picture type and complexity ratio of the second frame.

12. A method according to claim 1, wherein the step of estimating the first frame's second value comprises dividing the first value by a default value based on the first frame's picture type.

13. A method according to claim 1, further comprising:

receiving an unfiltered version of the first frame,

creating a filtered version of the first frame prior to encoding the first frame with the second algorithm, and

selecting either the unfiltered or filtered version of the first frame for encoding by the second algorithm.

14. A method of transcoding frames of a compressed digital video stream, each frame being encoded according to a first encoding algorithm and having a level of complexity, said method comprising:

decoding a first frame of the compressed digital video stream with a first decoding algorithm to produce a decoded version of the first frame and to measure a first value of the first frame's complexity and to determine the first frame's picture type,

estimating a second value of the first frame's complexity using the first value as a parameter, and

encoding the decoded version of the first frame with a distinct second encoding algorithm employing the second value of the first frame's complexity and the first frame's picture type as parameters.

15. A method according to claim 14, further comprising:

analyzing the decoded version of the first frame with the second encoding algorithm to measure a third value of the first frame's complexity, and

storing the picture type and a complexity ratio of the first frame.

16. A method according to claim 15, comprising, prior to encoding the decoded version of the first frame with the second algorithm,

generating an encoded version of a past frame with the second algorithm, said encoded version of the past frame having a size, and

estimating a value of a future frame's complexity,

and wherein the step of encoding the decoded version of the first frame with the second algorithm further comprises:

determining a target size for the decoded version of the first frame using the size of the past frame and the second value of the future frame's complexity as parameters, and

encoding the decoded version of the first frame with the second algorithm employing the first frame's target size as a parameter.

17. A method according to claim 16, wherein the step of encoding the first frame with the second algorithm produces an encoded version of the first frame, and the step of determining the target size of the decoded version of the first frame comprises predicting the size of said encoded version of the first frame.

18. A method according to claim 17, wherein the first frame's complexity ratio is calculated by dividing the first value of the frame's complexity by the third value of the first frame's complexity.

19. A method according to claim 18, wherein the first frame is followed by a plurality of frames of the compressed video stream, and the method further comprises

analyzing each of the plurality of frames of the compressed digital video stream with the first algorithm to produce a decoded version of each frame and to measure a first value of each frame's complexity and to determine each frame's picture type,

estimating a second value of each frame's complexity using each frame's first value as a parameter, and

encoding the decoded version of each of the plurality of frames with the second algorithm, employing the second value and picture type of each frame as parameters.

20. A method according to claim 19, further comprising:

analyzing the decoded version of each of the plurality of frames with the second algorithm to measure a third value of the complexity of each frame and

storing the picture type and complexity ratio for each of the plurality of frames.

21. A method according to claim 20, wherein the plurality of frames is followed by a second frame, and the method further comprises:

analyzing the second frame with the first algorithm to produce a decoded version of the second frame and to measure a first value of the second frame's complexity and to determine the second frame's picture type,

estimating a second value of the second frame's complexity using the first value as a parameter, and

encoding the decoded version of the second frame with the second algorithm employing the second value of the second frame's complexity and the second frame's picture type as parameters.

22. A method according to claim 21, comprising estimating the second frame's second value by dividing the second frame's first value by a translation factor.

23. A method according to claim 22, comprising calculating the second frame's translation factor by averaging the stored complexity ratios associated with a sub-

set of the plurality of frames, the sub-set being the frames that have been encoded by the second algorithm and have the same picture type as the second frame.

24. A method according to claim 23, further comprising:

analyzing the decoded version of the second encoded frame with the second algorithm to measure a third value of the second frame's complexity, and

storing the picture type and the complexity ratio of the second frame.

25. A method according to claim 14, wherein the step of estimating the first frame's second value comprises dividing the first value by a default value based on the first frame's picture type.

26. A method according to claim 14, further comprising:

creating a filtered version of the first frame from an unfiltered, uncompressed version of the first frame prior to encoding the first frame with the second algorithm, and

selecting either the unfiltered or filtered version of the first frame for encoding by the second algorithm.

27. An apparatus for encoding an uncompressed digital video input stream composed of a succession of frames, each frame having a plurality of characteristics associated therewith, said apparatus comprising:

an extraction means for receiving the succession of frames of the uncompressed digital video input stream and employing a first method to obtain measured values for the plurality of characteristics of a frame of the input stream and to assign a picture type to the frame,

a delay means for receiving the succession of frames of the input stream and outputting the frames in delayed fashion relative to the frames of the input stream,

a value storage means for storing the measured values and the picture type of a frame in the delay means, and

an encoding means for receiving a frame from the delay means and encoding the frame, the encoding means being responsive to a measured value stored in the value storage means for adjusting the size of the encoded version of the frame.

28. An apparatus according to claim 27, wherein the value storage means further comprises a means for manipulating a measured value of a first characteristic of a frame to derive estimated values for a second characteristic of the frame.

29. An apparatus according to claim 28, wherein the encoding means receives the picture type, a measured value of a characteristic and an estimated value of a characteristic associated with said frame and uses said picture type, said measured value and said estimated value as parameters for creating the encoded version of said frame.

30. An apparatus according to claim 29, wherein the value storage means will determine a predicted value of the size of an encoded version of a frame in the delay means and wherein the encoding means adjusts the size of the encoded version of the frame received from the delay means in response to the predicted value.

31. An apparatus according to claim 30, wherein the encoding means further comprises a means for measuring a value of a frame's complexity, a means for storing the

measured complexity value, and a means for storing the size of an encoded version of a frame.

32. An apparatus according to claim 31, further comprising:

a filtering means for receiving an unfiltered frame from the delay means and creating a filtered frame and

a switching means for receiving the unfiltered frame from the delay means, receiving the filtered frame from the filtering means and selectively transmitting either the unfiltered frame or the filtered frame to the encoding means.

33. An apparatus according to claim 32, wherein the switching means is responsive to the encoding means for selecting the unfiltered frame or the filtered frame for transmission to the encoding means.

34. An apparatus according to claim 27, wherein the extraction means comprises an encoding means for encoding the frames by a first encoding method and the encoding means for receiving the frames from the delay means encodes the frame by a second, distinct, encoding method.

35. An apparatus for transcoding a compressed digital video input stream composed of a succession of encoded frames, each encoded frame having a plurality of characteristics associated therewith, said apparatus comprising:

a decoding means for receiving the succession of encoded frames of the compressed digital video input stream and employing a first method to obtain a succession of decoded frames and measured values for the plurality of characteristics of a decoded frame and to assign a picture type to the decoded frame,

a delay means for receiving the succession of decoded frames of the input stream and outputting the decoded frames in delayed fashion relative to the encoded frames of the input stream,

a value storage means for storing the measured values and the picture type of a decoded frame in the delay means, and

an encoding means for receiving a decoded frame from the delay means and encoding the frame, the encoding means being responsive to a measured value stored in the value storage means for adjusting the size of the encoded version of the frame.

36. An apparatus according to claim 35, wherein the value storage means further comprises a means for manipulating a measured value of a first characteristic of a decoded frame to derive estimated values for a second characteristic of the decoded frame.

37. An apparatus according to claim 36, wherein the encoding means receives the picture type, a measured value of a characteristic and an estimated value of a characteristic associated with said decoded frame and uses said picture type, said measured value and said estimated value as parameters for creating the encoded version of said decoded frame.

38. An apparatus according to claim 37, wherein the value storage means will determine a predicted value of the size of an encoded version of a decoded frame in the delay means and wherein the encoding means adjusts the size of the encoded version of the decoded frame received from the delay means in response to the predicted value.

39. An apparatus according to claim 38, wherein the encoding means further comprises a means for measuring a value of a decoded frame's complexity, a means for storing the measured complexity value, and a means for storing the size of an encoded version of a decoded frame.

40. An apparatus according to claim 35, further comprising:

a filtering means for receiving an unfiltered decoded frame from the delay means and creating a filtered frame and

a switching means for receiving the unfiltered frame from the delay means, receiving the filtered frame from the filtering means and selectively transmitting either the unfiltered decoded frame or the filtered frame to the encoding means.

41. An apparatus according to claim 40, wherein the switching means is responsive to the encoding means for selecting the unfiltered decoded frame or the filtered frame for transmission to the encoding means.