Title: FAST MACROBLOCK DELTA QP DECISION

Abstract: A system and method for encoding multimedia video is described. As video is encoded a quantization parameter is selected for each macroblock. As described herein, the quantization parameter for each macroblock may be selected by limiting the universe of all possible quantization parameters to a particular range of possible quantization parameter values. This increases the speed of video encoding by reducing the number of quantization parameters that are tested for each video macroblock.
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(H))

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(H))

**Published:**

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

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
This application claims priority to U.S. Provisional Patent Application No. 61/030,857, filed February 22, 2008, which is incorporated herein by reference in its entirety.

BACKGROUND

Field of the Invention

Present embodiments relate to multimedia image processing. More particularly, these embodiments relate to a system and method for adaptively controlling the digital bit rate and compression quality of digital video in a video encoder.

Description of the Related Art

Digital video capabilities can be incorporated into a wide range of devices, including digital televisions, digital direct broadcast systems, wireless communication devices, personal digital assistants (PDAs), laptop computers, desktop computers, digital cameras, digital recording devices, cellular or satellite radio telephones, and the like. These and other digital video devices can provide significant improvements over conventional analog video systems in creating, modifying, transmitting, storing, recording and playing full motion video sequences.

A number of different video encoding standards have been established for communicating digital video sequences. The Moving Picture Experts Group (MPEG), for example, has developed a number of standards including MPEG-1, MPEG-2 and MPEG-4. Other encoding standards include H.261/H.263, MPEG1/2/4 and the latest H.264/AVC.

Video encoding standards achieve increased transmission rates by encoding data in a compressed fashion. Compression can reduce the overall amount of data that needs to be transmitted for effective transmission of image frames. The H.264 standards, for example, utilize graphics and video compression techniques designed to facilitate video and image transmission over a narrower bandwidth than could be achieved without the compression. In particular, the H.264 standards incorporate video encoding techniques that utilize similarities between successive image frames, referred to as temporal or
interframe correlation, to provide interframe compression. The interframe compression techniques exploit data redundancy across frames by converting pixel-based representations of image frames to motion representations. In addition, the video encoding techniques may utilize similarities within image frames, referred to as spatial or intraframe correlation, in order to achieve intra-frame compression in which the spatial correlation within an image frame can be further compressed. The intraframe compression is typically based upon conventional processes for compressing still images, such as spatial prediction and discrete cosine transform (DCT) encoding.

[0006] To support the compression techniques, many digital video devices include an encoder for compressing digital video sequences, and a decoder for decompressing the digital video sequences. In many cases, the encoder and decoder comprise an integrated encoder/decoder (CODEC) that operates on blocks of pixels within frames that define the sequence of video images. In the H.264 standard, for example, the encoder of a sending device typically divides a video image frame to be transmitted into macroblocks comprising smaller image blocks. For each macroblock in the image frame, the encoder searches macroblocks of the neighboring video frames to identify the most similar macroblock, and encodes the difference between the macroblocks for transmission, along with a motion vector that indicates which macroblock from the reference frame was used for encoding. The decoder of a receiving device receives the motion vector and encoded differences, and performs motion compensation to generate video sequences.

[0007] The difference between the macroblocks is transformed and then quantized. A quantization parameter (QP) is used to perform the quantization and will thus determine the control bit rate and recovered frame quality. Quantization using a higher QP corresponds to lower bit rate and lower quality. Quantization using a lower QP corresponds to higher bit rate and higher quality. By adjusting the QP, different bit rates and degrees of quality can be realized.
SUMMARY OF THE INVENTION

[0008] In some embodiments, a system for encoding video is provided, the system comprising a storage, the storage comprising a video frame in turn comprising a macroblock. The system also includes a quantization module configured to select a range of quantization parameters for quantizing the macroblock, wherein the range is a subset of possible quantization parameters; a processor configured to determine a quantization parameter in the range which results in the optimal quantization of the macroblock; and an encoder configured to encode the macroblock using the determined quantization parameter.

[0009] In some embodiments, a system for encoding video is provided, comprising means for receiving a video frame comprising a macroblock for quantization; means for selecting a range of quantization parameters for quantizing the macroblock, wherein the range is a subset of possible quantization parameters; means for determining the quantization parameter in the range which results in the optimal quantization value for the macroblock; and means for encoding the macroblock using the determined quantization parameter.

[0010] In some embodiments, a method for encoding video is provided, comprising: receiving a video frame comprising a macroblock for quantization; selecting a range of quantization parameters for quantizing the macroblock, wherein the range is a subset of possible quantization parameters; determining the quantization parameter in the range which results in the lowest distortion value for the macroblock; and encoding the macroblock using the determined quantization parameter.

[0011] In some embodiments, a computer readable medium is provided, comprising a computer readable program code adapted to be executed to perform a method comprising: receiving a video frame comprising a macroblock for quantization; selecting a range of quantization parameters for quantizing the macroblock, wherein the range is a subset of possible quantization parameters; determining the quantization parameter in the range which results in the lowest distortion value for the macroblock; and encoding the macroblock using the determined quantization parameter.
BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The features, objects, and advantages of the disclosed embodiments will become more apparent from the detailed description set forth below when taken in conjunction with the drawings in which like reference characters identify correspondingly throughout and wherein:

[0013] FIG. 1 is a top-level block diagram of an encoding source device and a decoding receiving device as used in one embodiment of the invention.

[0014] FIG. 2 is a schematic diagram of a source device in one embodiment of the invention, implementing a QP Optimization Module as described in embodiments of the present invention.

[0015] FIG. 3 is a general block diagram illustrating an encoding system in which the QP Optimization Module is utilized.

[0016] FIG. 4 is a block diagram of the block encoding process using a Quantization Parameter.

[0017] FIG. 5 is a schematic diagram of the operational flow of the QP Optimization Module.

DETAILED DESCRIPTION

[0018] Embodiments of the invention include systems and methods for encoding multimedia video that is to be displayed to a user. In one embodiment, the systems and methods are adapted to control the encoded data rate for compressing the video. In some cases the video is transmitted wirelessly or through a wired network to a receiver for display. One step in compressing video relates to quantizing the transmitted video data. Quantization involves dividing the video data by a quantization parameter (QP) so as reduce the data size to a smaller, more manageable form.

Because not all segments of video data contain the same amount of information, different segments of the video may be quantized differently with different QPs. If the QP selected for a given section is too small, the compressed data will be of good quality, but won't be compressed to a high degree. This leads to generation of a high bit rate when the data is transmitted to a receiver. Conversely, if the QP selected is too large, the data will be compressed by a high degree, and the bitrate will be reduced as there are fewer bits to transmit. However, the quality of the video frames being transmitted will be relatively low since each frame will be represented by relatively
fewer bits than if the QP was smaller. In one embodiment, the invention provides a system and method for quickly and accurately selecting a proper QP for a given slice of a video. Particularly, embodiments of the invention provide intelligent means for selecting the ranges likely to contain a suitable QP for a given block of data.

[0019] One method for determining the proper QP for a macroblock is the "brute-force" method. The brute-force implementation comprised encoding the same macroblock multiple times using a range of QP values, and choosing the optimal QP as the QP value that offers the minimal coding cost. Different cost criteria may be used to determine the proper QP for a particular macroblock. One cost measure is known as the rate-distortion cost, expressed as a combination of rate and distortion,

\[ C = D + \lambda R \]

[0020] where \( D \) is the distortion between the original video block and reconstructed video block due to quantization, \( R \) is the rate (number of bits) used to encode the input video block, and \( \lambda \) is a fixed parameter. Both distortion \( D \) and rate \( R \) are functions of QP as 1) QP directly affects the degree of distortion between the original video block and the reconstructed video block (higher QP means higher distortion); and 2) directly relates to the number of bits needed to encode the quantization residual coefficients (higher QP means a lower bit rate). In one possible embodiment \( \lambda \) may similarly depend on QP and may comprise the formula:

\[ \lambda_{\text{mode}} = 0.85 \times 2^{(\text{QP-12})/3} \]

[0021] Depending on the application, the coefficients in this equation will be altered to specify the importance of the bit rate over distortion. Here, for example, with increasing QP the bitrate is given increasing relevance in selecting the optimal value.

[0022] As a result, the cost \( C \) is also a function of QP, \( C=C(QP) \). After multiple encodings of the current macroblock given a range of QP values, the best QP for the macroblock can be chosen as:

\[ \text{QP}_{\text{opt}} = \arg \min(C(QP)) \]
One improvement as described herein is to cache a default motion estimator for a specified QP and use this cached motion estimation when calculating the distortion for each of the QPs in the search range. This caching, although resulting in a less accurate measure of a potential QP’s cost, saves considerable time, as the motion estimation need not be calculated repeatedly for each potential QP.

Inter Mode Decision:

Embodiments contemplate techniques to increase the encoding speed of the macroblock QP decision process. For example, during mode decision (when the encoder chooses the best coding mode among a multitude of possible coding modes for a macroblock), if inter mode is considered, motion search is performed to decide the optimal macroblock partition and the optimal motion vectors for each partition (both in terms of optimizing a cost metric). The encoder may perform motion search only when the macroblock is encoded for the nominal QP (which may be given by the end user or decided by the quality control specifications for the device) and store the motion search results. In one embodiment, the encoder retrieves the saved motion information in subsequent encoding rounds (with different values of macroblock QP) instead of invoking motion search again. This ensures that motion search, which is a costly and time-consuming process, is performed for each macroblock only once. To further reduce encoding time needed to decide optimal macroblock QP, the following additional techniques may be used:

Intra Mode Decision

In an inter-coded slice (i.e., P- or B-slice), intra coding modes (spatial prediction) and inter coding modes (temporal prediction) are both allowed. Therefore, during mode decision for P- and B-slice macroblocks, intra coding modes may also be considered. Intra modes are usually chosen for macroblocks that represent new objects in the current frame and/or scene change. However, compared to inter mode decision (where motion search results are not significantly influenced by the value of macroblock QP), intra mode decision depends on the value of macroblock QP. For example, with regard to H.264/AVC, four intra coding modes are allowed: intra 4x4, intra 8x8, intra 16x16 and IPCM. In the former two modes, the macroblock is partitioned into smaller blocks (4x4 or 8x8) and predicted sequentially in the raster scan order. Therefore, when
the macroblock QP changes, the reconstructed blocks will change, therefore influencing the prediction and mode decision for subsequent blocks in the macroblock. In the optimal scheme, intra mode decision should be repeated using different QPs. However, this incurs long encoding time. To speed up intra mode decision, only one QP value may be used. Similar to inter mode decision, the intra mode decision results (which include the prediction mode) may be stored and re-used in subsequent encoding rounds with different QP values. While this may incur a performance loss due to non-optimal intra mode decision, the impact may remain limited as usually only a limited percentage of macroblocks in an inter slice (P or B-slice) are intra-coded. To further limit the impact of this non-optimal intra mode decision, the encoder may decide to perform intra mode decision multiple times for different QP values if and only if, during mode decision at the nominal QP, intra coding mode instead of inter coding mode is selected to be the best coding mode for the current macroblock.

**QP Range Restriction:**

[0026] Neighboring macroblocks usually are spatially correlated. This means the optimal QP values for neighboring macroblocks are usually similar. Depending on the optimal QP values for its already coded neighbors, a macroblock’ s search range for the optimal QP may be subjected to certain restrictions. The following pseudo code provides one possible implementation of these restraints:
Initialize currMBDeltaQP, e.g., currMBDeltaQP = 0

MB_CODING_LOOP_START:
currMBQP = sliceQP + currMBDeltaQP

// only consider currMBQP if it is within small range of predicted QP
Let QPpred be the predicted QP based on neighboring MBs' QPs
if (currMB is not a boundary MB and currMBQP is similar to QPpred) {
    encodejnacroblock(currMBQP)
    if (currQPCost < minQPCost) {
        bestMBQP = currMBQP
        minQPCost = currQPCost
    }
}

Proceed to next delta QP value, currMBDeltaQP = nextDeltaQPValue;
Go back to MB_CODING_LOOP_START until all delta QP values are tested.

[0027] Note that the pseudo code above provides a special condition for macroblocks that lie on the boundary of a video frame/slice. For these boundary macroblocks, one or more of their neighbors are not available. A variety of methods can be used to account for this condition, for example: 1) using a default frame/slice level QP for unavailable neighbors when calculating QP predictor; or 2) testing the full range of delta QP values. Option 2) has the advantage that it will alleviate the slow-start problem for non-boundary macroblocks by allowing the boundary macroblocks to search for the optimal QP value within a wider range.

[0028] Multiple methods may be used to calculate the QP predictor based on neighboring QPs. For example, an average QP of the left and the top neighbors may be used. Alternatively, an average QP of the left, the top, and the top-left neighbors may be used. Alternatively, an average QP of the left, the top, the top-left, and the top-right neighbors may be used. Various combinations are possible depending on the order in which the macroblocks are encoded.

[0029] Different methods to decide whether the current QP value is similar to QP predictor may be used for different types of slices. For example, a range of [QPpred-2, QPpred+1] may be applied on P-slice macroblocks, while a range of [QPpred-1, QPpred+2] may be applied on B-slice macroblocks.
[0030] Conditional restriction on QP search range helps reduce the number of encoding rounds performed for each macroblock, hence speeding up the encoding process. However, the imposed restriction may incur some performance loss compared to exhaustive search. The performance loss may be more severe for macroblocks in I- and P-slices as rate-distortion performance of I- and P-slices will be propagated beyond the current group of pictures (GOP) until the next random access point (also known as IDR or instant decoder refresh picture in H.264/AVC) is encountered. Therefore, an alternative is to loosen or not apply conditional QP restriction on I- and P-slice macroblocks.

[0031] FIG. 1 is a block diagram illustrating an example system 100 in which a source device 101 transmits an encoded sequence of video data over communication link 109 to a receive device 102. Source device 101 and receive device 102 are both digital video devices. In particular, source device 101 encodes and transmits video data using any one of a variety of video compression standards, including those discussed supra. Communication link 109 may comprise a wireless link, a physical transmission line, a packet based network such as a local area network, wide-area network, or global network such as the Internet, a public switched telephone network (PSTN), or combinations of various links and networks. In other words, communication link 109 represents any suitable communication medium, or possibly a collection of different networks and links, for transmitting video data from source device 101 to receive device 102.

[0032] Source device 101 may be any digital video device capable of encoding and transmitting video data. For example, source device 101 may include memory 103 for storing digital video sequences, video encoder 104 for encoding the sequences, and transmitter 105 for transmitting the encoded sequences over communication link 109. Memory 103 may comprise computer memory such as dynamic memory or storage on a hard disk. Receive device 102 may be any digital video device capable of receiving and decoding video data. For example, receive device 102 may include a receiver 108 for receiving encoded digital video sequences, decoder 107 for decoding the sequences, and display 106 for displaying the sequences to a user.

[0033] Example devices for source device 101 and receive device 102 include servers located on a computer network, workstations or other desktop computing devices, and mobile computing devices such as laptop computers. Other examples include digital
television broadcasting systems and receiving devices such as cellular telephones, digital
televisions, digital cameras, digital video cameras or other digital recording
devices, digital video telephones such as cellular radiotelephones and satellite radio
telephones having video capabilities, other wireless video devices, and the like.

[0034] In some cases, source device 101 and receive device 102 each include an
encoder/decoder (CODEC) (not shown) for encoding and decoding digital video data.
In that case, both source device and receive device may include transmitters and
receivers as well as memory and displays. Many of the encoding techniques outlined
below are described in the context of a digital video device that includes an encoder. It
is understood, however, that the encoder may form part of a CODEC.

[0035] Source device 101, for example, includes an encoder 104 that operates on blocks
of pixels within the sequence of video images in order to encode the video data into a
compressed format. For example, the encoder 104 of source device 101 may divide a
video image frame to be transmitted into macroblocks comprising a number of smaller
image blocks. For each macroblock in the image frame, encoder 104 of source device
101 searches macroblocks stored in memory 103 for the preceding video frame already
transmitted (or a subsequent video frame) to identify a similar macroblock, and encodes
the difference between the macroblocks, along with a motion vector that identifies the
macroblock from the previous frame that was used for encoding. Source device 101
may support programmable thresholds which can cause termination of various tasks or
iterations during the encoding process in order to reduce the number of computations
and conserve power.

[0036] The receiver 108 of receive device 102 receives the motion vector and the
encoded video data, and decoder 107 performs motion compensation techniques to
generate video sequences for display to a user via display 106. One skilled in the art will
readily recognize that rather than display the decoded data various other actions may be
taken including storing the data, reformatting the data, or retransmitting the decoded
data. The decoder 107 of receive device 102 may also be implemented as an
encoder/decoder (CODEC). In that case, both source device and receive device may be
capable of encoding, transmitting, receiving and decoding digital video sequences.

[0037] FIG. 2 is a block diagram illustrating an example source device 101,
in incorporate a video encoder 203 that compresses digital video sequences according to
the techniques described herein. Exemplary digital video device 101 is illustrated as a
wireless device, such as a mobile computing device, a personal digital assistant (PDA),
a wireless communication device, a radio telephone, and the like. However, the
techniques in this disclosure are not limited to wireless devices, and may be readily
applied to other digital video devices including non-wireless devices.

[0038] In the example of FIG. 2, digital video device 101 is configured to transmit
compressed digital video sequences via transmitter 202 and antenna 201. Video encoder
203 encodes the video sequences and buffers the encoded digital video sequences within
video memory storage 205 prior to transmission. Memory storage 205 may also store
computer readable instructions and data for use by video encoder 203 during the
encoding process. Memory 205 may comprise synchronous dynamic random access
memory (SDRAM), a hard disk, FLASH memory, electrically erasable programmable
read only memory (EEPROM), or the like. Video frames 204 are extracted from
memory 205 for encoding. As described in detail below, video encoder 203 implements
a QP Optimization Module (QPOM) 206 that is configured to optimize video encoding
performed by the video encoder 203. The QPOM 206 facilitates bit-rate control by
having instructions that determine what parameters will be used by the encoder 203
during the encoding process.

[0039] An exemplary data compression system 300 which incorporates a QPOM is
illustrated in FIG. 3. A video signal 305 is first presented to preprocessor 301 in
preparation for compression. Preprocessor 301 may serve a variety of purposes, or may
be excluded from the system altogether. Preprocessor 301 may, for example, format the
video signal 305 into components that are more easily processed by the compression
system. The output of the preprocessor 301 is presented to encoder 302. Encoder 302
quantizes the data that it has received then compresses the quantized coefficients. The
quantization performed is dependent on feedback quantization parameters 307 from the
QPOM 303. Together, encoder 302 and QPOM 303 may comprise an encoding system
308 or CODEC. QPOM 303 utilizes statistics characterizing the current encoded
segment of video to adaptively set the quantization parameters for encoding the next
segment of video. This process of adaptively setting the quantization parameters for a
video being encoded is described in more detail below. Once the data has been
quantized it is sent to a formatter 304 that formats an output bitstream 306 for
transmission. Formatter 304 takes the rate controlled data and assembles the data into a
formatted bit stream for transmission through a communications channel. In doing so,
Formatter 304 may append supplemental information to the data. For example, signals indicative of start of block, start of frame, block number, frame number, and quantization information may be appended to the data signal by formatter 304.

Block-based video coding is widely used in video coding standards such as H.261/H.263, MPEG1/2/4 and the latest H.264/AVC. In block-based video coding system, an input video frame is processed block-by-block. A commonly used block size is 16x16, which is also known as a macroblock. Figure 4 shows a block diagram of one example of block-based video encoding. One skilled in the art will readily recognize that additional features are possible. Figure 4 provides a generalized overview of the operation of the encoder and its use of the QP. For each input video block of data 401, the system generates a prediction block 413. The prediction block 413 is formed by a prediction module 410 that performs either a spatial prediction (prediction within the same frame using already neighbors) or temporal prediction (prediction across frames), and may be used in the subsequent encoding process 417. A residual block 402 is then calculated by subtracting the values of the prediction block from the values of the corresponding original video block from frame 401. The residual block 402 is then run through a transform module 403 which, for example, performs a discrete cosine transform (DCT) transformation on the residual block 402. Following transformation, the residual block is quantized using a chosen quantization parameter at a quantization module 404. A set of quantized transform coefficients 405 are then scanned into a 1-dimensional vector at a coefficient scanning module 406 and entropy coded at an entropy coding module 408 before being sent out as an encoded bitstream 409.

As shown by block 416, the quantization parameter is not only used to quantize the block, but also to determine quantization efficiency. The residual block 414 is reconstructed by de-quantizing 416 and inverting 415 the quantized residual block. This reconstructed block 412 is then stored 411 and used not only in determining the efficiency of the quantization parameter for this particular block, but also for the subsequent prediction 418 of future blocks.

Figure 5 illustrates a generalized flow diagram of one possible embodiment of the QPOM’s 303 operation process 500. Process 500 begins at a start state 501 by accepting the first macroblock. For the first macroblock, the predicted value for QP may be determined by a default method, for example, brute force optimization search. A brute force optimization search comprises iterating over all possible QP values,
applying a cost metric for each, and selecting the QP generating the optimal cost metric (as discussed supra). One skilled in the art will recognize that various alternative methods exist for selecting an initial predicted QP, such as a default based on the statistical properties of the video signal, or a global prediction - any one of these alternatives will suffice so long as the resulting QP provides reasonable quantization. For subsequent macroblocks, the Process 500 may select a predicted QP based on neighboring QPs or previous optimal QPs. If the current macroblock is at the corner of a frame, i.e. neighbors are unavailable, alternative neighbors or a default value may be used for the predicted QP. Once the predicted QP has been selected, the Process 500 identifies the range of QP to evaluate at a state 502. The Process 500 iterates through the QP range applying a cost metric to determine the efficiency of each QP within the range. As discussed above, the cost metric may comprise the rate-distortion cost. Once the optimal QP has been determined, process 500 inserts the predicted QP into the slice header at a state 506. This will be the QP used for this macroblock’s subsequent quantization by the system. Process 500 then determines whether more macroblocks are available for processing at a decision state 503. If macroblocks are available the process will continue with the next macroblock at a state 508. If macroblocks are not available, the process will end at an end state 509.

[0043] Normally, the range selected will simply extend by an integer offset less than the predicted QP and an integer offset higher than the predicted QP (i.e., +/- 3). However, the range selected may be biased above or beneath the QP and may depend on the type of frame being quantized. H.264, for example, comprise I, P, or B frames. I and P frames, are more frequently referenced by other frames than B frames, to recover image information. Accordingly, I and P frames should normally be quantized less, so that their information is not lost. Thus, upon recognizing an I or P frame, the QPOM may instead select a range biased to a lower QP - i.e., 2 less than the predicted QP and only 1 above. This will increase the chance that the optimal QP is lower than the predicted QP. In contrast, there will be occasions when an upward bias would be preferable, i.e., in systems requiring a high bitrate, with less emphasis on quality.

[0044] Thus, a novel and improved method and apparatus for encoding video has been described. Those of skill in the art would understand that the various illustrative logical blocks, modules, circuits, and algorithm steps described in connection with the embodiments disclosed herein may be implemented as electronic hardware, computer
software, or combinations of both. The various illustrative components, blocks, modules, circuits, and steps have been described generally in terms of their functionality. Whether the functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. Skilled artisans recognize the interchangeability of hardware and software under these circumstances, and how best to implement the described functionality for each particular application. As examples, the various illustrative logical blocks, modules, circuits, and algorithm steps described in connection with the embodiments disclosed herein may be implemented or performed with a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device, discrete gate or transistor logic, discrete hardware components such as, e.g., registers and FIFO, a processor executing a set of firmware instructions, any conventional programmable software module and a processor, or any combination thereof. The processor may be a microprocessor, but in the alternative, the processor may be any conventional processor, controller, microcontroller, or state machine. The software module could reside in RAM memory, flash memory, ROM memory, registers, hard disk, a removable disk, a CD-ROM, or any other form of storage medium known in the art. Those of skill would further appreciate that the data, instructions, commands, information, signals, bits, symbols, and chips that may be referenced throughout the above description are represented by voltages, currents, electromagnetic waves, magnetic fields or particles, optical fields or particles, or any combination thereof.

[0045] The previous description of the preferred embodiments is provided to enable any person skilled in the art to make or use the disclosed embodiments. The various modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without the use of the inventive faculty. Thus, the disclosed embodiments are not intended to be limited to the embodiments shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.
We Claim:

1. A system for encoding video, comprising:
   a storage containing a video frame comprising a macroblock;
   a quantization module configured to select a range of quantization
   parameters for quantizing the macroblock, wherein the range is a subset of
   possible quantization parameters;
   a module configured to determine a quantization parameter in the range
   which results in the optimal quantization of the macroblock; and
   an encoder configured to encode the macroblock using the determined
   quantization parameter.

2. The system of Claim 1, wherein the module configured to determine a
   quantization parameter comprises a cache of prior motion searches.

3. The system of Claim 1, wherein the module configured to determine a
   quantization parameter comprises a cache of intra mode decision results.

4. The system of Claim 1, wherein the storage comprises dynamic memory.

5. The system of Claim 1, wherein the storage comprises a hard disk.

6. The system of Claim 1, wherein the range of quantization parameters is
   selected based on at least the value of a predicted quantization parameter.

7. The system of Claim 6, wherein the predicted quantization parameter is
   selected based on at least the value of a previously determined quantization parameter.

8. The system of Claim 7, wherein the previously determined quantization
   parameter is the quantization parameter for a preceding macroblock.

9. The system of Claim 7, wherein the previously determined quantization
   parameter is determined using at least the average of the quantization parameters of at
   least two neighboring macroblocks.
10. The system of Claim 7, wherein the previously determined quantization parameter is determined using at least the average of the quantization parameters of at least three neighboring macroblocks.

11. The system of Claim 7, wherein the previously determined quantization parameter is determined using at least the average of the quantization parameters of at least four neighboring macroblocks.

12. The system of Claim 7, wherein the previously determined quantization parameter is determined using one or more neighboring macroblock's quantization parameter, and in the absence of one or more neighboring macroblocks, uses a default quantization parameter value.

13. The system of Claim 7, wherein all possible quantization parameter values are used when neighboring macroblocks are not available.

14. The system of Claim 7, wherein the module comprises a buffer storing the previously determined quantization parameter for each block.

15. The system of Claim 1, wherein the system is a cellular telephone.

16. The system of Claim 1, wherein the optimal quantization of the macroblock comprises the lowest rate-distortion cost for the macroblock.

17. A system for encoding video, comprising:
   means for receiving a video frame comprising a macroblock for quantization;
   means for selecting a range of quantization parameters for quantizing the macroblock, wherein the range is a subset of possible quantization parameters;
   means for determining the quantization parameter in the range which results in the optimal quantization value for the macroblock; and
   means for encoding the macroblock using the determined quantization parameter.
18. The system of Claim 17, wherein the means for selecting a range of quantization parameters comprises a cache of prior motion searches.

19. The system of Claim 17, wherein the means for selecting a range of quantization parameters comprises a cache of intra mode decision results.

20. The system of Claim 17, wherein the receiving a video frame means comprises an electronic storage.

21. The system of Claim 20, wherein the storage is a computer memory.

22. The system of Claim 17, wherein the range of quantization parameters is selected based on at least the value of a predicted quantization parameter.

23. The system of Claim 22, wherein the predicted quantization parameter is selected based on at least the value of a previously determined quantization parameter.

24. The system of Claim 23, wherein the previously determined quantization parameter is the quantization parameter for a preceding macroblock.

25. The system of Claim 23, wherein the previously determined quantization parameter is determined using at least the average of the quantization parameters of at least two neighboring macroblocks.

26. The system of Claim 23, wherein the previously determined quantization parameter is determined using at least the average of the quantization parameters of at least three neighboring macroblocks.

27. The system of Claim 23, wherein the previously determined quantization parameter is determined using at least the average of the quantization parameters of at least four neighboring macroblocks.

28. The system of Claim 23, wherein the previously determined quantization parameter is determined using one or more neighboring macroblock's quantization parameter, and in the absence of one or more neighboring macroblocks, uses a default quantization parameter value.
29. The system of Claim 23, wherein all possible quantization parameter values are used when neighboring macroblocks are not available.

30. The system of Claim 17, wherein the selecting means comprises at least a part of a quantization parameter optimization module.

31. The system of Claim 17, wherein the determining means comprises at least a part of a quantization parameter optimization module.

32. The system of Claim 17, wherein the encoding the macroblock means comprises a video encoder.

33. The system of Claim 17, wherein the optimal quantization of the macroblock comprises the lowest rate-distortion cost for the macroblock.

34. A method for encoding video, comprising the steps of:
   receiving a video frame comprising a macroblock for quantization;
   selecting a range of quantization parameters for quantizing the macroblock, wherein the range is a subset of possible quantization parameters;
   determining the quantization parameter in the range which results in the lowest distortion value for the macroblock; and
   encoding the macroblock using the determined quantization parameter.

35. The method of Claim 34, wherein the step of determining the quantization parameter includes the step of caching prior motion searches.

36. The method of Claim 34, wherein the step of determining the quantization parameter includes the step of caching intra mode decision results.

37. The method of Claim 34, wherein the range of quantization parameters is selected based on at least the value of a predicted quantization parameter.

38. The method of Claim 37, further comprising the step of selecting the predicted quantization parameter based on at least the value of a previously determined quantization parameter.
39. The method of Claim 38, wherein the previously determined quantization parameter is the quantization parameter for the preceding macroblock.

40. The method of Claim 38, wherein the previously determined quantization parameter is determined using at least the average of the quantization parameters for one of at least two neighboring macroblocks, at least three neighboring macroblocks, or at least four neighboring macroblocks.

41. The method of Claim 38, wherein the previously determined quantization parameter is determined using one or more neighboring macroblock’s quantization parameter, and in the absence of one or more neighboring macroblocks, uses a default quantization parameter value.

42. A system for encoding video, comprising:

   a storage containing a video frame comprising a macroblock;
   a quantization module configured to cache prior motion searches;
   a module configured to determine a quantization parameter which results in the optimal quantization of the macroblock, based at least in part on at least one of the cached prior motion searches; and
   an encoder configured to encode the macroblock using the determined quantization parameter.

43. A system for encoding video, comprising:

   a storage containing a video frame comprising a macroblock;
   a quantization module configured to cache intra-mode decision results;
   a module configured to determine a quantization parameter which results in the optimal quantization of the macroblock, based at least in part on at least one of the cached intra-mode decision results; and
   an encoder configured to encode the macroblock using the determined quantization parameter.

44. A computer readable medium comprising a computer readable program code adapted to be executed to perform a method comprising:

   receiving a video frame comprising a macroblock for quantization;
selecting a range of quantization parameters for quantizing the macroblock, wherein the range is a subset of possible quantization parameters; determining the quantization parameter in the range which results in the lowest distortion value for the macroblock; and encoding the macroblock using the determined quantization parameter.
START

Using predicted QP, identify range of QP to evaluate

Iterate through QP range applying cost metric to select most favorable QP

Set optimal QP value in slice header

Proceed to next Macroblock

Last macroblock?

NO

YES

END

FIG. 5
**International Search Report**

**International application No**
PCT/US2009/034800

**A. CLASSIFICATION OF SUBJECT MATTER**

INV. H04N7/26 H04N7/50

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

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

H04N

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

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

EPO-Internal, WPI Data, INSPEC, COMPENDEX, IBM-TDB

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category*</th>
<th>Citation of document with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>figures 1,2 paragraphs [0003], [0014], [0042], [0045] - [0048], [0053], [0064]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>abstract; figures 1,4,5 paragraphs [0021], [0029], [0034], [0037], [0050], [0052], [0053]</td>
<td></td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C. See patent family annex.

* 'A' document defining the general state of the art which is not considered to be of particular relevance
* 'E' earlier document 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

Date of the actual completion of the international search: 28 July 2009

Date of mailing of the international search report: 06/08/2009

Name and mailing address of the ISA/
European Patent Office, P B 5818 Patentlaan 2
NL - 2280 HV RIJWINKEL
Tel (+31-70) 340-2040, Fax (+31-70) 340-3016

Authorized officer: Streich, Sebastian
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>EP 1 727 371 A (THOMSON LICENSING [FR]) 29 November 2006 (2006-11-29) paragraphs [0016], [0019], [0022], [0051] - [0053], [0059] ; figure 4a</td>
<td>1-12, 14, 17-28, 30-32, 34-44</td>
</tr>
<tr>
<td>X</td>
<td>DO-KYOUNG KWON ET AL: &quot;Rate Control for H.264 Video With Enhanced Rate and Distortion Model s&quot; IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, vol . 16, no. 5, 1 May 2007 (2007-05-01), pages 517-529, PISCATAWAY, NJ, USA ISSN: 1051-8215 abstract; figure 2 page 518, right-hand column, line 17 - line 22 page 524, algorithm steps 2) and 3)</td>
<td>1-3,6-8, 13, 14, 16-24, 29-39, 42,43</td>
</tr>
<tr>
<td>X</td>
<td>LIANG-WEI LEE ET AL: &quot;On The Error Distribution And Scene Change For The Bit Rate Control Of MPEG&quot; IEEE 1993 INTERNATIONAL CONFERENCE ON CONSUMER ELECTRONICS, DIGEST OF TECHNICAL PAPERS., 8 June 1993 (1993-06-08) , pages 286-287, PISCATAWAY, NJ, USA ISBN: 978-0-7803-0843-5 abstract ; figure 1 equation 3</td>
<td>1-3,6-9, 12, 14, 17-25, 28, 30-32, 34-43</td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20060071289 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2006018552 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 2006223197 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2600455 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 101171842 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2008533855 T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RU 2345503 C1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2006333444 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20060122775 A</td>
</tr>
</tbody>
</table>