

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 March 2007 (22.03.2007)

PCT

(10) International Publication Number
WO 2007/033242 A1

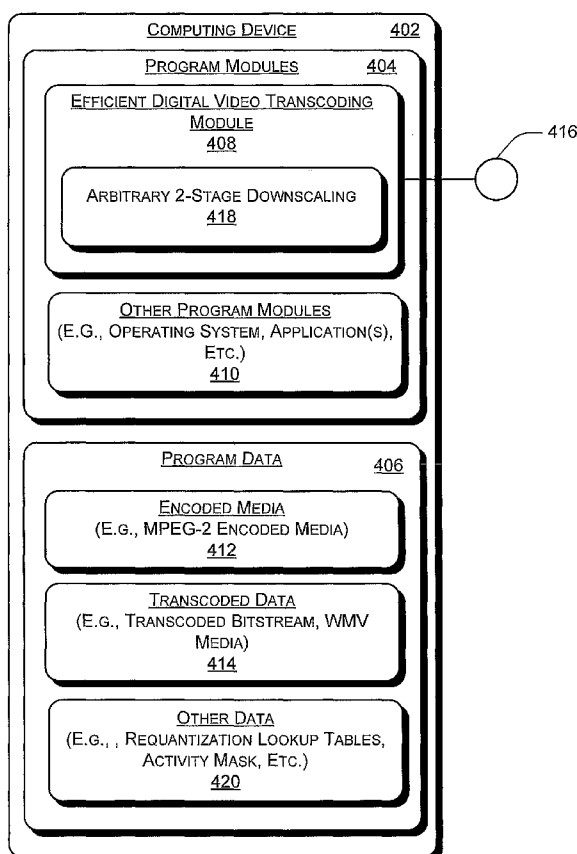
- (51) **International Patent Classification:**
H04N 7/24 (2006.01) *H04N 7/26* (2006.01)
- (21) **International Application Number:**
PCT/US2006/035640
- (22) **International Filing Date:**
13 September 2006 (13.09.2006)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
11/226,590 14 September 2005 (14.09.2005) US
- (71) **Applicant (for all designated States except US):** MICROSOFT CORPORATION [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).
- (72) **Inventors:** SHEN, Guobin; One Microsoft Way, Redmond, Washington 98052-6399 (US). LI, Shipeng; One

Microsoft Way, Redmond, Washington 98052-6399 (US). CAO, Wanyong; One Microsoft Way, Redmond, Washington 98052-6399 (US). HE, Yuwen; One Microsoft Way, Redmond, Washington 98052-6399 (US).

- (81) **Designated States (unless otherwise indicated, for every kind of national protection available):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States (unless otherwise indicated, for every kind of regional protection available):** ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, 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, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT,

[Continued on next page]

(54) **Title:** EFFICIENT INTEGRATED DIGITAL VIDEO TRANSCODING



(57) **Abstract:** Efficient integrated digital video transcoding is described. In one aspect, an integrated transcoder receives an encoded bitstream. The integrated transcoder transcodes the encoded bitstream by partially decoding the encoded bitstream based on a first transform associated with a first media data format. The decoding operations generate an intermediate data stream. The integrated transcoder then encodes the intermediate data stream using a second transform associated with a second media data format. The first and second transforms are not the same.

WO 2007/033242 A1



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(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Efficient Integrated Digital Video Transcoding

BACKGROUND

[0001] Digital video content is typically generated to target a specific data format. A video data format generally conforms to a specific video coding standard or a proprietary coding algorithm, with a specific bit rate, spatial resolution, frame rate, etc. Such coding standards include MPEG-2 and WINDOWS Media Video (WMV). Most existing digital video contents are coded according to the MPEG-2 data format. WMV is widely accepted as a qualified codec in the streaming realm, being widely deployed throughout the Internet, adopted by the HD-DVD consortium, and currently being considered as a SMPTE standard. Different video coding standards provide varying compression capabilities and visual quality.

[0002] Transcoding refers to the general process of converting one compressed bitstream into another compressed one. To match a device's capabilities and distribution networks, it is often desirable to convert a bitstream in one coding format to another coding format such as from MPEG-2 to WMV, to H.264, or even to a scalable format. Transcoding may also be utilized to achieve some specific functionality such as VCR-like functionality, logo insertion, or enhanced error resilience capability of the bitstream for transmission over wireless channels.

[0003] Fig. 1 shows a conventional Cascaded Pixel-Domain Transcoder (CPDT) system, which cascades a front-end decoder to decode an input bitstream with an encoder that generates a new bitstream with a different coding parameter set or in new format. One shortcoming of this conventional transcoding architecture is that its complexity typically presents an obstacle for practical

deployment. As a result, the CPDT transcoding architecture of Fig. 1 is typically used as a performance benchmark for improved schemes.

[0004] Fig. 2 shows a conventional cascaded DCT-domain transcoder (CDDT) architecture, simplifying the CPDT architecture of Fig. 1. The system of Fig. 2 limits functionality to spatial/temporal resolution downscaling and coding parameter changes. CDDT eliminates the DCT/IDCT processes implemented by the CPDT transcoder of Fig. 1. Yet, CDDT performs MC in the DCT domain, which is typically a time-consuming and computationally expensive operation. This is because the DCT blocks are often overlapped with MC blocks. As a result, the CDDT architecture typically needs to apply complex and computationally expensive floating-point matrix operations in order to perform MC in the DCT domain. Additionally, motion vector (MV) refinement is typically infeasible utilizing the CDDT architecture.

SUMMARY

[0005] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the detailed description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0006] In view of the above, efficient integrated digital video transcoding is described. In one aspect, an integrated transcoder receives an encoded bitstream. The integrated transcoder transcodes the encoded bitstream by partially decoding the encoded bitstream based on a first set of compression techniques associated with a first media data format. The decoding operations generate an intermediate data stream. The integrated transcoder then encodes the intermediate data stream

using a second set of compression techniques associated with a second media data format. The first and second sets of compression techniques are not the same.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] In the Figures, the left-most digit of a component reference number identifies the particular Figure in which the component first appears.

[0008] Fig. 1 shows a conventional Cascaded Pixel-Domain Transcoder (CPDT) system, which cascades a front-end decoder to decode an input bitstream with an encoder to generate a new bitstream with a different coding parameter set or in new format.

[0009] Fig. 2 shows a conventional cascaded DCT-domain transcoder (CDDT) architecture, simplifying the CPDT architecture of Fig. 1.

[0010] Fig. 3 shows an exemplary non-integrated pixel-domain transcoding split-architecture to transcode MPEG-2 to WMV, according to one embodiment. More particularly, this split-architecture provides a conceptual basis for efficient integrated digital video transcoding.

[0011] Fig. 4 shows an exemplary system for efficient integrated digital video transcoding, according to one embodiment.

[0012] Fig. 5 shows an exemplary simplified close-loop cascaded pixel-domain transcoder, according to one embodiment.

[0013] Fig. 6 shows an exemplary simplified closed-loop DCT-domain transcoder, according to one embodiment.

[0014] Fig. 7 shows an exemplary merge operation of four 4x4 DCT blocks into one 8 x 8 DCT block, according to one embodiment. This merge operation is performed during efficient video content transcoding.

[0015] Fig. 8 shows an exemplary architecture for a simplified DCT-domain numeral 2:1 resolution downscaling transcoder, according to one embodiment.

[0016] Fig. 9 shows an exemplary merge operation of four 4x4 DCT blocks into one 8 x 8 DCT block for interlace media for 2:1 spatial resolution downscaling transcoding operations, according to one embodiment.

[0017] Fig. 10 shows an exemplary simplified 2:1 downscaling transcoder architecture with full drift compensation, according to one embodiment.

[0018] Fig. 11 shows an exemplary standard virtual buffer verifier buffer (VBV) model for a decoder.

[0019] Fig. 12 shows a transcoder with arbitrarily spatial resolution downscaling, according to one embodiment.

[0020] Fig. 13 shows an exemplary procedure for efficient integrated digital video transcoding operations, according to one embodiment.

[0021] Fig. 14 shows an exemplary environment wherein efficient integrated digital video transcoding can be partially or fully implemented, according to one embodiment.

[0022] For purposes of discussion and illustration, color is used in the figures to present the following conventions. A blue solid arrow represents pixel domain signal with respect to real or residual picture data. A red solid arrow represents signal in the DCT domain. An orange dashed arrow represents motion information.

DETAILED DESCRIPTION

Overview

[0023] Systems and methods for efficient digital video transcoding are described below in reference to Figs. 4 through 14. These systems and methods utilize information in the input bitstream to allow an application to dynamically control error propagation, and thereby, selectively control speed and quality of video bitstream transcoding. This selective control allows an application to

seamlessly scale from close-loop transcoding (high-speed transcoding profile) to open-loop (high-quality transcoding profile) transcoding schemes. In contrast to conventional transcoding architectures (e.g., the CPDT of Fig. 1 and the CDDT of Fig. 2), the architectures for efficient digital video transcoding are integrated and that they combined different types of Discrete Cosine Transforms (DCTs) or DCT-like transforms into one transcoding module. The systems and methods for efficient video transcoding implement requantization with a fast lookup table, and provide fine drifting control mechanisms using a triple threshold algorithm.

[0024] In one implementation, where efficient digital video transcoding transcodes a bitstream data format (e.g., MPEG-2, etc.) to WMV, the high-quality profile transcoding operations support advanced coding features of WMV. In one implementation, high-speed profile transcoding operations implement arbitrary resolution two-stage downscaling (e.g., when transcoding from high definition (HD) to standard definition (SD)). In such two-stage downscaling operations, part of the downscaling ratio is efficiently achieved in the DCT domain, while downscaling ratio operations are implemented in the spatial domain at a substantially reduced resolution.

Exemplary Conceptual Basis

[0025] Fig. 3 shows exemplary non-integrated cascaded pixel-domain transcoding split-architecture 300 to convert MPEG-2 to WMV. This split-architecture is not integrated because separate modules respectively perform decoding and encoding operations. The split-architecture of Fig. 3 provides a conceptual basis for subsequent description of the integrated systems and methods for efficient digital video transcoding. TABLE 1 shows symbols and their respective meanings for discussion of Fig. 3.

TABLE 1

e_{i+1}	Error of frame ($i+1$) to be encoded by encoding portion of the transcoder;
$\hat{B}_i$	Reconstructed frame i by MPEG-2 decoder at original resolution;
$\tilde{B}_i$	Reconstructed frame i by the encoder at original resolution;
$\hat{b}_i$	Reconstructed frame i by the MPEG-2 decoder at reduced resolution;
$\tilde{b}_i$	Reconstructed frame i by the encoder at reduced resolution;
$\hat{r}_{i+1}$	Reconstructed residues of frame ($i+1$) by MPEG-2 decoder;
$\tilde{r}_{i+1}$	Reconstructed residues of frame ($i+1$) by the encoder
$MC_{mp2}(B, mv)$	Motion compensated prediction with reference picture B and motion vector mv by MPEG-2 decoder, on 16x16 block basis;
$MC_{vc1}(B, mv)$	Motion compensated prediction with reference picture B and motion vector mv by transcoder 308 (encoder), either on 16x16 or 8x8 block basis;
$MC'_{mp2}(b, mv)$	Motion compensated prediction with reduced resolution reference b and motion vector mv , using MPEG-2 filtering, on 8x8 or smaller block basis
$MC'_{vc1}(b, mv)$	Motion compensated prediction with reduced resolution reference B and motion vector mv , using transcoder 308 filtering, on 8x8 or smaller block basis;
MV	Motion vector in the original frame resolution
mv	Motion vector in the reduced frame resolution

[0026] For purposes of description and exemplary illustration, system 300 is described with respect to transcoding from MPEG-2 to WMV with bit rate reduction, spatial resolution reduction, and their combination. Many existing digital video contents are coded according to the MPEG-2 data format. WMV is widely accepted as a qualified codec in the streaming realm, being widely

deployed throughout the Internet, adopted by the HD-DVD Consortium, and currently being considered as a SMPTE standard.

[0027] MPEG-2 and WMV provide varying compression and visual quality capabilities. For example, the compression techniques respectively used by MPEG-2 and WMV are very different. For instance, the motion vector (MV) precision and motion compensation (MC) filtering techniques are different. In MPEG-2 motion precision is only up to half-pixel accuracy and the interpolation method is bilinear filtering. In contrast, in WMV, the motion precision can go up to quarter-pixel accuracy, and two interpolation methods namely bilinear filtering and bicubic filtering are supported. Moreover, there is a rounding control parameter involved in the filtering process. Use of WMV may result in up to a 50% reduction in video bit rate with negligible visual quality loss, as compared to an MPEG-2 bit rate.

[0028] In another example, transforms used by MPEG-2 and WMV are different. For instance, MPEG-2 uses standard DCT/IDCT and the transform size is fixed to 8x8. In contrast, WMV uses integer transforms (VC1-T) where the elements of transform kernel matrix are all small integers. Additionally, transform size can be altered using WMV from blocks to blocks using either 8x8, 8x4, 4x8 and 4x4. MPEG-2 does not support frame level optimization. Whereas, WMV supports various frame level syntaxes for performance optimization. WMV supports many other advanced coding features such as intensity compensation, range reduction, and dynamic resolution change, etc.

[0029] In view of the above, to provide bit rate reduction without resolution change, the filtering process bridging the MPEG-2 decoder and the WMV encoder shown in Fig. 3 is an all-pass filter (*i.e.*, not in effect). Therefore, the input to the encoder for frame ($i+1$) is expressed as:

$$e_{i+1} = \hat{r}_{i+1} + MC_{mp2}(\hat{B}_i, MV_{mp2}) - MC_{vc1}(\tilde{B}_i, MV_{vc1}) \quad (1)$$

[0030] In this implementation, WMV coding efficiency of Fig. 3 gains result from finer motion precision. In WMV, quarter-pixel motion precision is allowed beside the common half-pixel precision as in MPEG-2. Moreover, WMV allows better but more complex interpolation known as bicubic interpolation for MC filtering. Bilinear interpolation is used for MPEG-2 in the MC module (MC_{mp2}) for half-pixel MC. The bilinear interpolation method similar to that used in WMV with the exception that the MPEG-2 bilinear interpolation does not have rounding control. To achieve high speed, half-pixel motion accuracy can be implemented in the encoder portion. One reason for this is the lack of the absolute original frame (*i.e.*, bitstream input data (BS_IN) is already compressed). Thus, in this example, it is difficult to obtain a more accurate yet meaningful motion vector. On the other hand, the motion information obtained from MPEG-2 decoder (*i.e.* $MV_{vc1} = MV_{mp2}$) can be reused directly. Since there is no resolution change, there is no MV precision loss with this assumption. If the encoder is further restricted to use bilinear interpolation and force the rounding control parameter to be always off, then under the reasonable assumption that motion compensation is a linear operation and ignoring the rounding error (*i.e.*, $MC_{VC9} = MC_{mp2}$), Equation 1 is simplified as follows:

$$e_{i+1} = \hat{r}_{i+1} + MC_{mp2}(\hat{B}_i - \tilde{B}_i, MV_{mp2}) \quad (2)$$

According to Equation 2, the reference CPDT transcoder in Fig. 3 can be simplified. Such a simplified architecture is described below in reference to Fig. 5. Prior to describing the simplified architecture, an exemplary system for efficient digital video transcoding is first described.

An Exemplary System

[0031] Although not required, efficient digital video transcoding is described in the general context of computer-program instructions being executed by a computing device such as a personal computer. Program modules generally include routines, programs, objects, components, data structures, etc., that perform particular tasks or implement particular abstract data types. While the systems and methods are described in the foregoing context, acts and operations described hereinafter may also be implemented in hardware.

[0032] Fig. 4 shows an exemplary system 400 for efficient digital video transcoding. In this implementation, the operations of system 400 are described with respect to hybrid DCT and block-based motion compensation (MC) video coding schemes, upon which many video coding standards and proprietary formats are based. More particularly, system 400 is described with architectures, components, and operations used to transcode MPEG-2 to WMV. However, it can be appreciated that the architectures, components, and operations described for scalable complexity and efficiency transcoding embodied by system 400 for transcoding MPEG-2 to WMV can also be applied to other bitstream data format conversions besides MPEG-2 and WMV. For example, in one implementation, system 400 is utilized to transcode MPEG-2 bitstream to MPEG-4 bitstream and MPEG-4 bitstream data to WMV bitstream data, etc. In such alternate embodiments, the following described transcoding architectures of system 400 (including components and operations associated therewith), consider the type of bitstream data being decoded, encoded, and respective data formats.

[0033] In this implementation, system 400 includes a general-purpose computing device 402. Computing device 402 represents any type of computing device such as a personal computer, a laptop, a server, handheld or mobile computing device, etc. Computing device 402 includes program modules 404 and

program data 406 to transcode an encoded bitstream in a first data format (e.g. MPEG-2) to a bitstream encoded into a different data formats (e.g., WMV). Program modules 404 include, for example, efficient digital video transcoding module 408 ("transcoding module 408") and other program modules 410. Transcoding module 408 transcodes encoded media 412 (e.g., MPEG-2 media) into transcoded media 414 (e.g., WMV media). Other program modules 410 include, for example, an operating system and an application utilizing the video bitstream transcoding capabilities of transcoding module 408, etc. In one implementation, the application is part of the operating system. In one implementation, transcoding module 408 exposes its transcoding capabilities to the application via an Application Programming Interface (API) 416.

High-Speed Profile Transcoding

[0034] Fig. 5 shows an exemplary simplified integrated closed-loop cascaded pixel-domain transcoder without error propagation. For purposes of discussion and illustration, the components of Fig. 5 are described in reference to the components of Fig. 4. For instance, the architecture of Fig. 5 is representative of one exemplary architecture implementation of transcoding module 408 of Fig. 4. Referring to the architecture 500 Fig. 5, as compared to the architecture in Fig. 3, please note that this is an integrated architecture without separate encoder and decoder components. Additionally, please note that the MV refining motion estimation module is removed from the MC in MPEG-2 decoder. Additionally, MC in the WMV encoder is merged to a MC that operates on accumulated requantization errors. In this manner, the transcoding architecture of Fig. 5 significantly reduces computation complexity for high-speed transcoding of progressive and interlaced video data formats.

[0035] Please note that the WMV transform is different from the one used in MPEG-2. In MPEG-2, standard floating point DCT/IDCT is used whereas the integer transform, whose energy packing property is akin to DCT, is adopted in WMV. As a result, the IDCT in the MPEG-2 decoder and the VC1-T in WMV encoder do not cancel out each other. The integer transform in WMV is different from the integer implementation of DCT/IDCT. The integer transform in WMV is carefully designed with all the transform coefficients to be small integers. Conventional transcoders are not integrated to transcode a bitstream encoded with respect to a first transform to a second transform that is not the same as the first transform.

[0036] Equation 3 provides an exemplary transform matrix for 8x8 VC1-T.

$$T_8 = \begin{bmatrix} 12 & 12 & 12 & 12 & 12 & 12 & 12 & 12 \\ 16 & 15 & 9 & 4 & -4 & -9 & -15 & -16 \\ 16 & 6 & -6 & -16 & -16 & -6 & 6 & 16 \\ 15 & -4 & -16 & -9 & 9 & 16 & 4 & -15 \\ 12 & -12 & -12 & 12 & 12 & -12 & -12 & 12 \\ 9 & -12 & 4 & 15 & -15 & -4 & 16 & -9 \\ 6 & -16 & 16 & -6 & -6 & 16 & -16 & 6 \\ 4 & -9 & 15 & -16 & 16 & -15 & 9 & -4 \end{bmatrix} \quad (3)$$

[0037] Equation 3 in combination with equations 4 and 5, which are described below, indicate how two different transforms are implemented into a scaling component of transcoding module 408 (Fig. 4). In one implementation, the accuracy of VC1-T is 16-bit accuracy, which is very suitable for MMX implementation. As a result, the codec complexity can be significantly reduced.

[0038] Fig. 6 shows an exemplary simplified closed-loop DCT-domain transcoder. The architecture of Fig. 6 is representative of one exemplary architecture implementation of transcoding module 408 (Fig. 4). The architecture 600 of Fig. 6 is a simplified architecture as compared to the architecture 500 of Fig.

5. Referring to Fig. 6, let C_8 be the standard DCT transform matrix, B , the inverse quantized MPEG-2 DCT block, and b , the IDCT of B , then the MPEG-2 IDCT is calculated as follows:

$$b = C'_8 B C_8$$

Let $\hat{B}$ be the VC1-T of b , then $\hat{B}$ is calculated as:

$$\hat{B} = T_8 b T'_8 \circ N_{88}$$

where $\circ$ denotes element-wise multiplication of two matrices, and N_{88} is the normalization matrix for VC1-T transform which is calculated as follows:

$$N_{88} = c_8 \cdot c'_8$$

with

$$c_8 = [8/288 \ 8/289 \ 8/292 \ 8/298 \ 8/288 \ 8/289 \ 8/292 \ 8/298];$$

$\hat{B}$ is directly computed from B , using the following formula:

$$\hat{B} = T_8 (C'_8 B C_8) T'_8 \circ N_{88} \quad (4)$$

[0039] To verify that $T_8 C'_8$ and $C_8 T'_8$ are very close to diagonal matrices, if we apply the approximation, then Equation 4 becomes an element-wise scaling of matrix B . That is,

$$\hat{B} = B \circ S_{88} \quad (5)$$

where

$$S_{88} = \text{diag}(T_8 C'_8) \cdot \text{diag}(C_8 T'_8) \circ N_{88}$$

[0040] Equation 5 shows that the VC1-T in WMV encoder and the IDCT in MPEG-2 decoder can be merged. Consequently, the architecture in Fig. 5 can be further simplified to the one shown in Fig. 6. Detailed comparison reveals that the two DCT/IDCT modules are replaced by two VC1-T and inverse VC1-T modules. In one implementation, a simple scaling module is also added. Two switches are embedded along with and an activity mask in this architecture. These embedded

components, as described below, are used for dynamic control of the complexity of transcoding coding operations of transcoder 408 (Fig. 4). At this point, these components are connected. The 16-bit arithmetic property of the WMV transform lends itself to parallel processing for PC and DSP. In view of this, computation complexities are significantly reduced. Moreover, since all the elements of the scaling matrix, S_{88} , are substantially close in proximity with respect to one another, this computation, and one implementation, is replaced by a scalar multiplication.

[0041] Figs. 5 and 6 show exemplary respective closed-loop transcoding architectures, wherein a feedback loop is involved. In this implementation, the feedback loop, which includes VC-1 dequantization, VC-1 inverse transform, residue error accumulation and MC on the accumulated error, compensates for the error caused by the VC-1 requantization process. Requantization error is a main cause of the drifting error for bit-rate-reduction transcoders, such as that shown in Fig. 1. Although the transcoding architectures of Figs. 5 and 6 are not completely drift-free, even with error compensation, the drifting error is very small. This is because the remaining cause of drift error is the rounding error during motion compensation filtering. One merit of residue error compensation is that the architectures of Figs. 5 and 6 provide for dynamically turning on or off the compensation process, as described below with respect to TABLE 2. The transcoding architecture of Fig. 6 performs pure bit rate reduction transcoding from MPEG-2 to WMV such as SD to SD or HD to HD conversion in a substantially optimal manner.

[0042] More particularly, conventional cascaded transcoder architectures (e.g., the architectures of Figs. 1 and 2) lack complexity flexibility. With respect to computation savings, the most that such conventional architecture can achieve is through MV reuse and mode mapping. On the other hand, accumulated residue error compensation architectures, for example, the architecture of Fig. 6 (and the

architectures of Figs. 8 and 10, as described below) have built-in scalability in terms of complexity. TABLE 2 shows exemplary meanings of switches in Fig. 6.

TABLE 2
Exemplary Switches for Dynamic Control of Transcoding Speed and Quality

S_0	Block level	Error accumulation switch
S_1	Block level	Error update switch
S_2	Block level	Early skip block decision switch

[0043] After transcoding module 408 of Fig. 4 has implemented drift-free simplification, an application can dynamically trade-off between the complexity and the quality to accelerate transcoding speed. In this implementation, quality can be traded for speed, and vice versa. In other words, some drifting error may be allowed in the further simplified transcoder. With this strategy, the drifting error introduced in the faster method is limited and fully controllable. Based on this consideration, three switches (S_0 , S_1 , and S_2) are provided in the architectures of Figs. 6, 8, and 10. The switches are used only to the residue-error compensation based architectures. The switches selectively skip some time-consuming operations to reduce complexity substantially, while introducing only a small amount of error. The meanings of various switches are summarized in TABLE 2. Computational decisions associated with these switches are efficiently obtained according to criteria described below with respect to each switch.

[0044] Switch S_0 controls when requantization error of a block should be accumulated into the residue-error buffer. As compared to a standard reconstruction selector, the role of switch S_0 is improved by adopting a fast lookup table based requantization process and by providing a finer drifting control mechanism via a triple-threshold algorithm. As a result, all observations made

with respect to switch S_0 are considered. For example, in one implementation, the DCT domain energy difference may be utilized as the indicator.

[0045] Switch S_1 controls when the most time-consuming module, MC of the accumulated residue error. In one implementation, switch S_1 is on. A binary activity mask is created for the reference frame. Each element of the activity mask corresponds to the activeness of an 8x8 block, as determined by

$$Activity(Block_i) = \begin{cases} 1, & Energy(block_i) > Th \\ 0, & Energy(block_i) \leq Th \end{cases}$$

where $Energy(block_i)$ is the energy of the block in the accumulated residue-error buffer. In one implementation, $Energy(block_i)$ is calculated spatial domain or DCT domain. $Energy(block_i)$ can be approximated by the sum of absolute values. If the MV points to blocks belonging to the area of low activity, then MC of the accumulated residue error for that specific block is skipped.

[0046] Switch S_2 performs early detection to determine whether block error should be encoded. This is especially useful in transrating applications where the encoder applies a coarser quantization step size. In this implementation, if the input signal (the sum of the MC of accumulated residue error and the reconstructed residue from MPEG-2 decoder) is weaker than a threshold, then switch S_2 is turned off so that no error will be encoded.

[0047] In one implementation, thresholds for the switches S_0 , S_1 , and S_2 are adjusted such that earlier reference frames are processed with higher quality and at slower speed. This is because the purpose of the switches is to achieve a better trade-off between quality and speed, and because of the predictive coding nature.

High-Quality Profile Transcoder

[0048] If bit rate change is not significant or the input source quality is not very high, the architecture of Fig. 6 substantially optimizes bit rate reduction when converting MPEG-2 bitstreams to WMV bitstreams. On the other hand, input source may be of high quality and high quality output may be desired, also speed of transcoding may be a moderate requirement (*e.g.*, real-time). A high-quality profile transcoder, such as the cascaded pixel-domain transcoder (CDPT) of Fig. 3 with MV refinement, meets these criteria. With this architecture, we can turn on all the advanced coding features of the WMV encoder to ensure highest coding efficiency can be achieved.

Resolution Change

[0049] In conventional media transcoding systems there are generally three sources of errors for transcoding with spatial resolution downscaling. These errors are as follows:

- Downscaling: errors generated when obtaining a downsampled video. It is typically a hardwired choice when designing operations of the downscaling filter to make a trade-off between visual quality and complexity, especially when downscaling in the spatial domain.
- Requantization error: As with the pure bit rate reduction transcoding process, this is error due to the requantization with a coarser re-quantization step size.
- MV Error: Incorrect MV will lead to wrong motion compensated prediction. As a result, no matter how the requantization error is compensated, and no matter how high the bit rate goes, a perfect result is difficult to obtain if not re-computing the motion compensation based on the new MVs and modes. This is a problem for conventional systems that transcode B-frames, because WMV supports only one MV mode for B-frames. This could also be a problem if one

desires to perform optimization, which would lead to coding mode change, *e.g.*, from four-MV to one-MV mode. Moreover, the problem generally exists for chrominance components since they are typically compensated with a single MV. (This is not a problem for the described efficient digital video transcoding architectures when applied to P-frames. One reason for this is because WMV supports four-MV coding mode for P-frames).

The operations of transcoding module 408 (Fig. 4) address the last two sources of errors, as now described.

Requantization Error Compensation

[0050] Let D denote the down-sampling filtering. Referring to the architecture of Fig. 3, input to the VC-1 encoder for frame $(i+1)$ is derived as follows:

$$e_{i+1} = D(\hat{r}_{i+1}) + D(MC_{mp2}(\hat{B}_i, MV_{mp2})) - MC_{vc1}(\tilde{b}_i, mv_{vc1}) \quad (6)$$

Assume that $MC_{VC1} = MC_{mp2}$, $mv_{mp2} = mv_{vc1} = MV_{mp2}/2$. With the approximation that

$$D(MC_{mp2}(\hat{B}_i, MV_{mp2})) = MC'_{mp2}(D(\hat{B}_i), D(MV_{mp2})) = MC'_{mp2}(\hat{b}_i, mv_{mp2}) \quad (7),$$

Equation 6 is simplified to the following:

$$e_{i+1} = D(\hat{r}_{i+1}) + MC'_{mp2}(\hat{b}_i - \tilde{b}_i, mv_{mp2}) \quad (8)$$

[0051] The first term in Equation 8, $D(\hat{r}_{i+1})$, refers to the downscaling process of the decoded MPEG-2 residue signal. This first term can be determined using spatial domain low-pass filtering and decimation. However, use of DCT-domain downscaling to obtain this term results in a reduction of complexity and better PSNR and visual quality. DCT-domain downscaling results are substantially better than results obtained through spatial domain bi-linear filtering

or spatial domain 7-tap filtering with coefficients $(-1, 0, 9, 16, 9, 0, -1)/32$. In this implementation, DCT-domain downscaling retains only the top-left 4x4 low-frequency DCT coefficients. That is, applying a standard 4x4 IDCT on the DCT coefficients retained will result in a spatially 2:1 downsampled image (*i.e.*, transcoded media 414 of Fig. 4).

[0052] The second term in Equation 8, $MC'_{mp2}(\hat{b}_i - \tilde{b}_i, mv_{mp2})$, implies requantization error compensation on a downsampled resolution. In this implementation, the MC in MPEG-2 decoder and the MC in WMV encoder are merged to a single MC process that operates on accumulated requantization errors at the reduced resolution.

[0053] Fig. 7 shows an exemplary merge operation of four (4) 4x4 DCT blocks into one 8 x 8 DCT block. One practical issue remains. In DCT-domain downscaling, four 8x8 DCT (blocks, B_1 through B_4 in an MPEG-2 macroblock (MB) at the original resolution) are mapped to the four 4x4 sub-blocks of an 8x8 block of the new MB at the reduced resolution and still in DCT domain (*e.g.*, please see Fig. 7). In WMV, for P-frames and B-frames, the 4x4-transform type is allowed. As a result, nothing needs to be done further except the abovementioned scaling. However, for I-frames, only the 8x8-transform type is allowed. Thus, when dealing with I-frames, transcoding module 408 (Fig. 4) converts the four 4x4 low-frequency DCT sub-blocks into an 8x8 DCT block: $\hat{B}$. In one implementation, this is accomplished by inverse transforming the four 4x4 DCT sub-blocks back into the pixel domain, and then applying a fresh 8x8 VC1-T. In one implementation, and to reduce computation complexity, this is achieved in the DCT domain.

[0054] For example, let $\hat{B}_1$, $\hat{B}_2$, $\hat{B}_3$, and $\hat{B}_4$ represent the four 4x4 low-frequency sub-blocks of B_1 , B_2 , B_3 , and B_4 , respectively; C_4 be the 4x4 standard IDCT transform matrix; T_8 be the integer WMV transform matrix; and further let

$T_8 = [T_L, T_R]$ where T_L and T_R are 8x4 matrices. In this scenario, $\hat{B}$ is directly calculated from $\hat{B}_1$, $\hat{B}_2$, $\hat{B}_3$, and $\hat{B}_4$ using the following equation:

$$\hat{B} = (T_L C'_4) \hat{B}_1 (T_L C'_4)' + (T_L C'_4) \hat{B}_2 (T_R C'_4)' + (T_R C'_4) \hat{B}_3 (T_L C'_4)' + (T_R C'_4) \hat{B}_4 (T_R C'_4)'$$

After some manipulation, $\hat{B}$ is more efficiently calculated as follows:

$$\hat{B} = (X + Y)C' + (X - Y)D'$$

wherein

$$C = (T_L C'_4 + T_R C'_4) / 2$$

$$D = (T_L C'_4 - T_R C'_4) / 2$$

$$X = C(\hat{B}_1 + \hat{B}_3) + D(\hat{B}_1 - \hat{B}_3)$$

$$Y = C(\hat{B}_2 + \hat{B}_4) + D(\hat{B}_2 - \hat{B}_4)$$

In one implementation, both C and D of the above equation are pre-computed. The final results are normalized with N_{88} .

[0055] Fig. 8 shows an exemplary architecture 800 for a simplified DCT-domain numeral 2:1 resolution downscaling transcoder. In one implementation, transcoding module 408 of Fig. 4 implements the exemplary architecture 800. The switches in this architecture have the same functionality as those in Fig. 6, as described above in reference to TABLE 2. Referring to Fig. 8, and one implementation, the first two modules (MPEG-2 VLD and inverse quantization) are simplified as compared to what is shown in Fig. 6. This is because transcoding module 408 retrieves only the top-left 4x4 portion out of the 8x8 block.

[0056] Compared to a conventional drift-low transcoder with drifting error compensation in reduced resolution, the transcoders of Figs. 6 and 8 do not include a mixed block-processing module. This is because WMV supports Intra coding mode for 8x8 blocks in an Inter coded macroblock. In other words, an Intra MB at the original resolution is mapped into an Intra 8x8 block of an Inter MB at the

reduced resolution. In view of this, the MB mode mapping rule becomes very simple, as shown immediately below:

$$\text{mode_new} = \begin{cases} \text{INTRA} & \text{if all mode_orig} = \text{INTRA} \\ \text{SKIP} & \text{if all mode_orig} = \text{SKIP} \\ \text{INTER} & \text{otherwise} \end{cases}$$

Existing mixed block processing operations typically require a decoding loop to reconstruct a full resolution picture. Therefore, the removal of mixed block processing provides substantial computation savings as compared to conventional systems.

[0057] Simplified DCT-domain 2:1 resolution downscaling transcoding architecture 800 is substantially drifting-free for P-frames. This is a result of the four-MV coding mode. The only cause of drifting error, as compared with a CPDT architecture with downscaling filtering, is the rounding of MVs from quarter resolution to half resolution (which ensures $mv_{mp2} = mv_{ve1}$) and the non-commutative property of MC and downscaling. Any such remaining errors are negligible due to the low-pass downscaling filtering (e.g., achieved in the DCT domain or in the pixel domain).

[0058] Fig. 9 shows an exemplary merge operation of four 4x4 DCT blocks into one 8 x 8 DCT block for interlace media for 2:1 spatial resolution downscaling transcoding operations, according to one embodiment. 2:1 downscaling changes resolution of an original frame by two in both horizontal and vertical directions. In one implementation, this interlace process is implemented by transcoding module 408 of Fig. 4. More particularly, for interlace coded content, the top-left 8x4 sub-block in every MB is reconstructed by shortcut MPEG-2 decoder, both fields are smoothed by low pass filter in vertical direction, then one field is dropped before the WMV encoding process.

MV Error Compensation

[0059] Although WMV supports four MV coding mode, it is typically only intended for coding P-frames. As a result, system 400 (Fig. 4) implements the architecture of Fig. 6 when there are no B-frames in the input MPEG-2 stream or the B-frames are to be discarded during the transcoder towards a lower temporal resolution. One reason for this is that WMV allows only one MV per MB for B-frames. In such a scenario, transcoding module 408 (Fig. 4) composes a new motion vector from the four MVs associated with the MBs at the original resolution. Each of the previously mentioned MV composition methods is compatible. In one implementation, transcoding module 408 implements median filtering. As described, incorrect MV will lead to wrong motion compensated prediction. To make matters worse, no matter how the requantization error is compensated, and no matter how high the bit rate goes, perfect results are difficult to obtain if not re-doing the motion compensation based on the new MVs. Therefore, we provide an architecture that allows such motion errors to be compensated.

[0060] Again, referring to the architecture of Fig. 3, input to the VC-1 encoder for frame $(i+1)$, which is assumed to be a B-frame, is derived as follows:

$$e_{i+1} = D(\hat{r}_{i+1}) + D(\text{MC}_{\text{mp2}}(\hat{b}_i, MV_{\text{mp2}})) - \text{MC}_{\text{vc1}}(\tilde{b}_i, mv_{\text{vc1}}) \quad (9);$$

with the approximation that

$$D(\text{MC}_{\text{mp2}}(\hat{b}_i, MV_{\text{mp2}})) = \text{MC}'_{\text{mp2}}(D(\hat{b}_i), D(MV_{\text{mp2}})) = \text{MC}'_{\text{mp2}}(\hat{b}_i, mv_{\text{mp2}}) \quad (10)$$

[0061] Equation 9 is simplified to

$$e_{i+1} = D(\hat{r}_{i+1}) + \text{MC}'_{\text{mp2}}(\hat{b}_i, mv_{\text{mp2}}) - \text{MC}'_{\text{vc1}}(\tilde{b}_i, mv_{\text{vc1}}) \quad (11)$$

In view of Equation 11, the following is obtained:

$$e_{i+1} = D(\hat{r}_{i+1}) + \text{MC}'_{\text{mp2}}(\hat{b}_i, mv_{\text{mp2}}) - \text{MC}'_{\text{vc1}}(\tilde{b}_i, mv_{\text{vc1}})$$

$$\begin{aligned}
&= D(\hat{r}_{i+1}) + [MC'_{mp2}(\hat{b}_i, mv_{mp2}) - MC'_{vc1}(\hat{b}_i, mv_{vc1})] + MC'_{vc1}(\hat{b}_i, mv_{vc1}) - \\
&\quad MC'_{vc1}(\tilde{b}_i, mv_{vc1}) \\
&= D(\hat{r}_{i+1}) + [MC'_{mp2}(\hat{b}_i, mv_{mp2}) - MC'_{vc1}(\hat{b}_i, mv_{vc1})] + MC'_{vc1}(\hat{b}_i - \tilde{b}_i, mv_{vc1}) \quad (12)
\end{aligned}$$

[0062] The two terms in the square brackets in Equation 12 compensate for the motion errors caused by inconsistent MVs (*i.e.*, mv_{mp2} is different from mv_{vc1}) or caused by different MC filtering methods between MPEG-2 and WMV. The corresponding modules for this purpose are highlighted and grouped into a light-yellow block in Fig. 10.

[0063] Fig. 10 shows an exemplary simplified 2:1 downscaling transcoder architecture with full drift compensation, according to one embodiment. In one implementation, transcoding module 408 of Fig. 4 implements the exemplary architecture of Fig. 10. Referring to Equation 12, please note that $MC'_{mp2}(\hat{b}_i, mv_{mp2})$ is performed for all the 8x8 blocks that correspond to original Inter MBs, and $mv_{mp2} = MV_{mp2}/2$ with quarter pixel precision. The MV used in the VC-1 encoder is a single MV: $mv_{vc1} = \text{median}(MV_{mp2})/2$. Note that with respect to the motion-error-compensation module, the accuracy of mv_{vc1} can go to quarter-pixel level. The last term in Equation 12 compensates for the requantization error of reference frames. Since B-frames are not reference for other frames, they are more error tolerant. As a result, an application can safely turn off the error compensation to achieve higher speed. Again, such approximation is intended for B-frames only. Please note that MC for motion error compensation operates on reconstructed pixel buffers while the MC for requantization error compensation operates on accumulated residue error buffer.

[0064] As to the MC, Intra-to-Inter or Inter-to-Intra conversion can be applied. This is because the MPEG-2 decoder reconstructed the B-frame and the reference frames. In this implementation, this conversion is done in the mixed block-processing module in Fig. 10. Two mode composition methods are possible.

And one implementation, the dominant mode is selected as the composed mode. For example, if the modes of the four MBs at the original resolution are two bi-directional prediction mode, one backward prediction mode and one forward prediction mode, then bi-directional prediction mode is selected as the mode for the MB at the reduced resolution. In another implementation, the mode that will lead to the largest error is selected. In view of this example, suppose using the backward mode will cause largest error. In this scenario, the backward mode is chosen such that the error can be compensated. Results show that the latter technique offers slightly better quality as compared to the former mode selection technique.

[0065] An exemplary architecture according to Equation 12 is shown in Fig. 10. There are four frame-level switches specifically for this architecture, as shown in TABLE 3.

TABLE 3
Exemplary Frame-Level Switches

S_{IP}	Frame level	Switch to be closed for I- and P-frames only
S_P	Frame level	Switch to be closed for P-frames only
S_B	Frame level	Switch to be closed for B-frames only (= ! S_{IP})
$S_{IP/B}$	Frame level	Switch to be closed for I- and P-frames only if there are B-frames

[0066] The four frame-level switches ensure different coding paths for different frame types. Specifically, the architecture does not perform: residue-error accumulation for B-frames (S_{IP}), does not perform MV error compensation for I- and P-frames (S_B), and does not reconstruct reference frames if there is no B-frames to be generated ($S_{IP/B}$). Please note the frame-level switch S_B can be turned into block-level switch since the MV error needs to be compensated only when the corresponding four original MVs are significantly inconsistent.

[0067] More particularly, switch S_{IP} is closed only for I-frames or P-frames, Switch S_P is closed only for P-frames, and switch S_B is closed only for B-frames. The resulting architecture is not as complex as the reference cascaded pixel-domain transcoder of Fig. 3. One reason for this is that the explicit pixel-domain downscaling process is avoided. Instead, pixel-domain downscaling is implicitly achieved in the DCT domain by simply discarding the high DCT coefficients. This architecture has excellent complexity scalability achieved by utilizing various switches, as described above with respect to TABLE 2.

[0068] For applications that demand ultra-fast transcoding speed, the architecture of Fig. 10 can be configured into an open-loop one by turn off all the switches. This open-loop architecture can be further optimized by merging the dequantization process of MPEG-2 and the requantization process of WMV. The inverse zig-zag scan module (inside VLD) of MPEG-2 can also be combined with the one in WMV encoder.

Chrominance Components

[0069] With respect to chrominance components in MPEG-2 and in WMV, the MV and the coding mode of chrominance components (UV) are derived from those of luminance component (Y). If all the four MBs at the original resolution that correspond to the MB at the reduced resolution have consistent coding mode (*i.e.*, all Inter-coded or all Intra-coded), there is no problem. However, if it is not case, problems result due to different derivation rules of MPEG-2 and WMV. In MPEG-2, the UV blocks are Inter coded when the MB is coded with Inter mode. However, in WMV, the UV blocks are Inter coded only when the MB is coded with Inter mode and there are less than three Intra-coded 8x8 Y blocks. This issue exists for both P-frames and B-frames. Transcoding module 408 of Fig. 4 addresses these problems as follows:

- **Inter-to-Intra conversion:** When the Inter-coded MB has three Intra-coded 8x8 Y blocks (it is impossible for an Inter-coded MB to have all four 8x8 Y blocks Intra coded), the UV blocks are Intra coded. In this case, one MB at the original resolution is Inter-coded along with corresponding UV blocks. These UV blocks will be converted from Inter mode to Intra mode. Since the Human Visual System (HVS) is less sensitive to the chrominance signals, transcoding module 408 utilizes a spatial concealment technique to convert the 8x8 UV blocks from Inter to Intra mode. In one implementation, the DC distance is utilized as an indicator to determine the concealment direction. Concealment is achieved via a simple copy or any other interpolation method.
- **Intra-to-Inter conversion:** When an Inter-coded MB has one or two Intra-coded 8x8 Y blocks, transcoding module 408 inter-codes the UV blocks. In this scenario, there are one or two Intra-coded MBs among the four corresponding MBs at the original resolution. These UV blocks are converted from Intra mode to Inter mode. In this implementation, transcoding module 408 utilizes a temporal concealment technique called the zero-out method to handle these blocks, and thereby, avoid the decoding loop.

[0070] Using error concealment operations to handle mode conversion for chrominance component, error introduced into a current frame is negligible and can be ignored, although it may cause color drifting in subsequent frames. Drifting for the chrominance component is typically caused by incorrect motion. To address this and improve quality, in one implementation, transcoding module 408 uses reconstruction based compensation for the chrominance component (*i.e.*, always applying the light-yellow module for the chrominance component).

Rate Control

[0071] Fig. 11 shows an exemplary virtual buffer verifier buffer (VBV) model for a decoder. A decoder based on the VBV model of Fig. 11 will typically verify an existing MPEG-2 bitstream. In this implementation, if the video rate is decreased proportional to the input rate, then the transcoded WMV bitstream will automatically satisfy the VBV requirements. In view of this, the efficient digital video transcoding architecture of this specification makes the coded frame size proportional to the input frame size for all the frames. These novel architectures continually compensate for accumulated differences between the target frame size and the actual resultant frame size, and obtain, via training, a linear quantization step (QP) mapping rule for different bit rate ranges.

[0072] For high bit rate, there is an approximate formula between coding bits (B) and quantization step (QP) which is also used in MPEG-2 TM-5 rate control method.

$$B = S \cdot \frac{X}{QP} \quad (13)$$

where S is the complexity of frame, X is model parameters. Assuming the complexity of a frame remains the same for different codecs:

$$QP_{vc1} = \left(\frac{X_{vc1}}{X_{mp2}} \right) \cdot \left(\frac{B_{mp2}}{B_{vc1}} \right) \cdot QP_{mp2} = k \cdot \left(\frac{B_{mp2}}{B_{vc1}} \right) \cdot QP_{mp2}$$

where QP_{vc1} is the QP value used in WMV re-quantization, QP_{mp2} is QP value of MPEG-2 quantization, and k is the model parameter related to the target bit rate. In one implementation, the following linear model is utilized:

$$QP_{vc1} / QP_{mp2} = k \cdot (B_{mp2} / B_{vc1}) + t \quad (14)$$

The values of parameter k and t for low, medium and high bit rate cases are listed in TABLE 4 using the linear regression method.

TABLE 4
EXEMPLARY PARAMETER VALUES FOR LINEAR REGRESSION
METHODOLOGY

Frame Type	I frame		P frame		B frame	
Parameters	k	t	k	t	k	t
Low (<1Mbps)	0.612861	-0.194954	0.016081	3.128561	0.076037	2.264825
Med (<3Mbps)	0.314311	0.070494	0.041140	1.400647	0.207292	0.545977
High	0.682409	-0.248120	0.057869	1.115930	0.199024	0.441518

[0073] An exemplary detailed rate control algorithm based on Equation 14 is shown in TABLE 5, where the meanings of various symbols in the algorithm presented in TABLE 5 are defined in following TABLE 6.

TABLE 5
EXEMPLARY RATE CONTROL ALGORITHM

```

Initialize  $SumD = 0$ ;
While (MPEG-2 stream is not end)
{
    Step1:    Decode one MPEG2 frame and get  $B_{mp2}$  and  $QP_{mp2}$ ;
    Step2:  $B_{pred\_vc1} = B_{mp2} \cdot \frac{R_{vc1}}{R_{mp2}}$ 
             $B_{vc1} = B_{pred\_vc1} + SumD$ 
            If ( $B_{vc1} < 0$ ) then  $B_{vc1} = 1$ ;
             $QP_{vc1} = (k \cdot \frac{B_{mp2}}{B_{vc1}} + t) \cdot QP_{mp2}$ ;
            Round and Clip  $QP_{vc1}$  to [1, 31];
    Step3:    Encode this frame into WMV frame using  $QP_{vc1}$ ;
    Step4:    Obtain the actual coded WMV frame size  $B_{actual\_vc1}$ ;
            Update  $SumD$ :  $SumD = SumD + B_{pred\_vc1} - B_{actual\_vc1}$ ;
}

```

TABLE 6
DEFINITIONS OF SYMBOLS USED IN THE ALGORITHM OF TABLE 5

B_{mp2}	MPEG-2 frame size;
R_{mp2}	MPEG-2 stream bit rate;
R_{vc1}	Target WMV stream bit rate;
B_{pred_vc1}	WMV frame size predicted by the ratio of bit rate;
B_{vc1}	Expected WMV frame size to encode (new bit rate);
B_{actual_vc1}	Actual encoded WMV frame size;
$SumD$	Accumulated differences between the predicted and actual WMV frame size from beginning.

Arbitrarily Resolution Change

[0074] Conversion of contents from HD resolution to SD resolution, for example to support legacy SD receivers/players, is useful. Typical resolutions of HD format are 1920x1080i and 1280x720p while those for SD are 720x480i, 720x480p for NTSC. The horizontal and vertical downscaling ratios from 1920x1080i to 720x480i are 8/3 and 9/4, respectively. To keep the aspect ratio, the final downscaling ratio is chosen to be 8/3 and the resulting picture size is 720x404. Similarly, for 1280x720p to 720x480p, the downscaling ratio is chosen to be 16/9 and the resulting picture size is 720x404. Black banners are inserted to make a full 720x480 picture by the decoder/player (instead of being padded into the bitstream).

[0075] According to digital signal processing theory, a substantially optimal downscaling methodology for a downscaling ratio m/n , would be to first up sample the signal by n -fold (*i.e.*, insert $n-1$ zeros between every original samples), apply a low-pass filter (*e.g.*, a *sinc* function with many taps), and then decimate the resulting signal by m -fold. Performing such operations, any spectrum aliasing introduced by the down-scaling would be maximally suppressed. However, this

process would also be very computationally expensive, and difficult to implement with in real-time because the input signal is high definition. To reduce this computational complexity, a novel two-stage downscaling strategy is implemented.

[0076] Fig. 12 shows a transcoder with arbitrarily spatial resolution downscaling, according to one embodiment. In one implementation, transcoding module 408 of Fig. 4 implements architecture of Fig. 12. In one implementation, the arbitrary downscaling transcoder is a non-integrated transcoder, such as in Fig. 12. In another implementation, the following arbitrary downscaling transcoding operations, which are described below with respect to Fig. 12, are implemented in an integrated transcoder such as that shown in Figs. 5, 6, 8, and/or 10.

[0077] Referring to Fig. 12, system 1200 implements two-stage downscaling operations to achieve any arbitrary downscaling target. Results of the first stage downscaling are embedded into the decoding loop. This reduces the complexity of the decoding operations. For example, to achieve an $8/3$ downscale ratio, downscaling operations are first implemented to downscale by $2/1$. The results of this first stage downscaling are input into the decoding loop, wherein second stage downscaling is performed in the spatial domain. In this example, second stage downscaling operations downscale by $4/3$ to achieve an $8/3$ downscale ratio. In another example, a downscale ratio of $16/9$ is achieved by system 1200 by applying $4/3$ downscaling twice (in two stages). This two-stage downscaling methodology utilizes the previously discussed DCT-domain downscaling strategy, and then fully embeds the first stage downscaling results into the decoding loop. Since resolution is significantly reduced after the first stage downscaling, we can continue to apply the optimal downscaling method on the pixel-domain.

[0078] Referring to Fig. 12, please note that multiple MVs

$$\left(\text{between } \left\lfloor \frac{m}{n} \right\rfloor \times \left\lfloor \frac{m}{n} \right\rfloor \text{ and } \left\lceil \frac{m}{n} \right\rceil \times \left\lceil \frac{m}{n} \right\rceil\right)$$

are associated with a new MB (the MV scaling and filtering modules).

Exemplary Procedure

[0079] Fig. 13 illustrates a procedure 1300 for efficient digital video transcoding, according to one embodiment. In one implementation, transcoding module 408 of Fig. 4 implements the operations of procedure 1300. Referring to Fig. 13, at block 1302, the procedure receives an encoded bitstream (e.g., encoded media 412 of Fig. 4). At block 1304, the procedure partially decodes the encoded bitstream according to a first set of compression techniques associated with a first media data format (e.g., MPEG-2, MPEG-4, etc.). The partial decoding operations generate an intermediate data stream. The integrated transcoder does not perform full decoding. For example, in cases where the MC of the “conceptual” MPEG-2 decoder is merged with that of the WMV encoder, it is hard to describe the decoding operations as performing MPEG-2 decoding. At block 1306, if downscaling of the intermediate data stream is desired, the procedure downscales data associated with the encoded bitstream in a first stage of downscaling. The first stage of downscaling is implemented in the DCT domain of a decoding loop. At block 1308, if two-stage downscaling is desired, the procedure further downscales in the spatial domain the data that was downscaled in the DCT domain (see block 1306).

[0080] At block 1310, the data decoded according to the first set of compression techniques is encoded with a second set of compression techniques. In one implementation, procedure 1300 is implemented within a non-integrated transcoding architecture, such as that shown and described with respect to Figs. 12 and 14. In this implementation, the second set of compression techniques is the same as the first set of compression techniques. In another implementation,

procedure 1300 is implemented within an integrated transcoding architecture, such as that shown and described with respect to Figs. 5-11, and 14. In this other implementation, the second set of compression techniques is not the same as the first set of compression techniques. For example, in one implementation, the first set of compression techniques is associated with MPEG-2, and the second set of compression techniques is associated with WMV.

An Exemplary Operating Environment

[0081] Fig. 14 illustrates an example of a suitable computing environment in which efficient digital video transcoding may be fully or partially implemented. Exemplary computing environment 1400 is only one example of a suitable computing environment for the exemplary system 400 of Fig. 4, and is not intended to suggest any limitation as to the scope of use or functionality of systems and methods the described herein. Neither should computing environment 1400 be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in computing environment 1400.

[0082] The methods and systems described herein are operational with numerous other general purpose or special purpose computing system, environments or configurations. Examples of well-known computing systems, environments, and/or configurations that may be suitable for use include, but are not limited to personal computers, server computers, multiprocessor systems, microprocessor-based systems, network PCs, minicomputers, mainframe computers, distributed computing environments that include any of the above systems or devices, and so on. Compact or subset versions of the framework may also be implemented in clients of limited resources, such as handheld computers, or other computing devices. The invention is practiced in a networked computing environment where tasks are performed by remote processing devices that are linked through a communications network.

[0083] With reference to Fig. 14, an exemplary system providing efficient digital video transcoding architecture includes a general-purpose computing device in the form of a computer 1410 implementing, for example, initiator operations associated with computing device 102 of Fig. 1. Components of computer 1410 may include, but are not limited to, processing unit(s) 1418, a system memory 1430, and a system bus 1421 that couples various system components including the system memory to the processing unit 1418. The system bus 1421 may be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures. By way of example and not limitation, such architectures may include Industry Standard Architecture (ISA) bus, Micro Channel Architecture (MCA) bus, Enhanced ISA (EISA) bus, Video Electronics Standards Association (VESA) local bus, and Peripheral Component Interconnect (PCI) bus also known as Mezzanine bus.

[0084] A computer 1410 typically includes a variety of computer-readable media. Computer-readable media can be any available media that can be accessed by computer 1410, including both volatile and nonvolatile media, removable and non-removable media. By way of example, and not limitation, computer-readable media may comprise computer storage media and communication media. Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer-readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any

other medium which can be used to store the desired information and which can be accessed by computer 1410.

[0085] Communication media typically embodies computer-readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism, and includes any information delivery media. The term “modulated data signal” means a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example and not limitation, communication media includes wired media such as a wired network or a direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media. Combinations of the any of the above should also be included within the scope of computer-readable media.

[0086] System memory 1430 includes computer storage media in the form of volatile and/or nonvolatile memory such as read only memory (ROM) 1431 and random access memory (RAM) 1432. A basic input/output system 1433 (BIOS), containing the basic routines that help to transfer information between elements within computer 1410, such as during start-up, is typically stored in ROM 1431. RAM 1432 typically contains data and/or program modules that are immediately accessible to and/or presently being operated on by processing unit 1418. By way of example and not limitation, Fig. 14 illustrates operating system 1434, application programs 1435, other program modules 1436, and program data 1437.

[0087] The computer 1410 may also include other removable/non-removable, volatile/nonvolatile computer storage media. By way of example only, Figure 14 illustrates a hard disk drive 1441 that reads from or writes to non-removable, nonvolatile magnetic media, a magnetic disk drive 1451 that reads from or writes to a removable, nonvolatile magnetic disk 1452, and an optical disk

drive 1455 that reads from or writes to a removable, nonvolatile optical disk 1456 such as a CD ROM or other optical media. Other removable/non-removable, volatile/nonvolatile computer storage media that can be used in the exemplary operating environment include, but are not limited to, magnetic tape cassettes, flash memory cards, digital versatile disks, digital video tape, solid state RAM, solid state ROM, and the like. The hard disk drive 1441 is typically connected to the system bus 1421 through a non-removable memory interface such as interface 1440, and magnetic disk drive 1451 and optical disk drive 1455 are typically connected to the system bus 1421 by a removable memory interface, such as interface 1450.

[0088] The drives and their associated computer storage media discussed above and illustrated in Figure 14, provide storage of computer-readable instructions, data structures, program modules and other data for the computer 1410. In Figure 14, for example, hard disk drive 1441 is illustrated as storing operating system 1444, application programs 1445, other program modules 1446, and program data 1447. Note that these components can either be the same as or different from operating system 1434, application programs 1435, other program modules 1436, and program data 1437. Operating system 1444, application programs 1445, other program modules 1446, and program data 1447 are given different numbers here to illustrate that they are at least different copies.

[0089] A user may enter commands and information into the computer 1410 through input devices such as a keyboard 1462 and pointing device 1461, commonly referred to as a mouse, trackball or touch pad. Other input devices (not shown) may include a microphone, joystick, graphics pen and pad, satellite dish, scanner, etc. These and other input devices are often connected to the processing unit 1418 through a user input interface 1460 that is coupled to the system bus 1421, but may be connected by other interface and bus structures, such as a

parallel port, game port or a universal serial bus (USB). In this implementation, a monitor 1491 or other type of user interface device is also connected to the system bus 1421 via an interface, for example, such as a video interface 1490.

[0090] The computer 1410 operates in a networked environment using logical connections to one or more remote computers, such as a remote computer 1480. In one implementation, remote computer 1480 represents computing device 106 of a responder, as shown in Fig. 1. The remote computer 1480 may be a personal computer, a server, a router, a network PC, a peer device or other common network node, and as a function of its particular implementation, may include many or all of the elements described above relative to the computer 1410, although only a memory storage device 1481 has been illustrated in Figure 14. The logical connections depicted in Figure 14 include a local area network (LAN) 1481 and a wide area network (WAN) 1473, but may also include other networks. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets and the Internet.

[0091] When used in a LAN networking environment, the computer 1410 is connected to the LAN 1471 through a network interface or adapter 1470. When used in a WAN networking environment, the computer 1410 typically includes a modem 1472 or other means for establishing communications over the WAN 1473, such as the Internet. The modem 1472, which may be internal or external, may be connected to the system bus 1421 via the user input interface 1460, or other appropriate mechanism. In a networked environment, program modules depicted relative to the computer 1410, or portions thereof, may be stored in the remote memory storage device. By way of example and not limitation, Figure 14 illustrates remote application programs 1485 as residing on memory device 1481. The network connections shown are exemplary and other means of establishing a communications link between the computers may be used.

Conclusion

[0092] Although the above sections describe efficient digital video transcoding architectures in language specific to structural features and/or methodological operations or actions, the implementations defined in the appended claims are not necessarily limited to the specific features or actions described. Rather, the specific features and operations of the described efficient integrated digital video transcoding architecture are disclosed as exemplary forms of implementing the claimed subject matter.

For example, in one implementation, the described fast and high quality transcoding systems and methodologies, including transcoding, arbitrary sized downscaling, and rate reduction are used for MPEG-2 to MPEG-4 transcoding and MPEG-4 to WMV transcoding. For instance, the simplified closed-loop DCT-domain transcoder in Fig. 6 can be used to transcode MPEG-4 to WMV. One difference between MPEG-2 (IS-13818 Part.2) is that MPEG-2 only utilizes half pixel element (pel) MV precision and bilinear interpolation in MC; there is such a same mode (half pel bilinear) in WMV. However, MPEG-4 supports both half pel and quarter pel MV precision, as well as interpolation for quarter pel positions (different from that in WMV). To address this difference, when 1/2 pel MV is used by MPEG-4 video, then the transcoding process is the same as MPEG-2 to WMV transcoding, as described above. Additionally, when 1/4 pel MV is contained in MPEG-4 video, then error is introduced due to different interpolation methods in MC as described above with respect to Fig. 6. Additionally, the simplified 2:1 downscaling transcoder with full drift compensation described above with respect to Fig. 10 is applicable to MPEG-4 to WMV 2:1 downsized transcoding independent of change. Moreover, high quality transcoding, including

the above described rate reduction and arbitrarily downscaling transcoding operations of Fig. 12 are effective for MPEG-4 to WMV transcoding.

CLAIMS

1. A computer-implemented method comprising:
receiving, by an integrated transcoder, an encoded bitstream; and
transcoding, by the integrated transcoder, the encoded bitstream as follows:
 partially decoding, by the integrated transcoder, the encoded bitstream to generate an intermediate data stream, the encoded bitstream being encoded with a first set of compression techniques associated with the first media format; and
 encoding, by the integrated transcoder, the intermediate data stream using a second set of compression techniques to generate a transcoded bitstream, the second set of compression techniques corresponding to the second media format.
2. The computer-implemented method of claim 1, wherein the first media format is MPEG-2 and wherein the second media format is WMV.
3. The computer-implemented method of claim 1, wherein the first media format is MPEG-2 and wherein the second media format is MPEG-4.
4. The computer-implemented method of claim 1, wherein the integrated transcoder is a closed loop transcoder that prevents error propagation via proper error compensation.
5. The computer-implemented method of claim 1, wherein the integrated transcoder is an open loop transcoder that does not prevent error propagation.

6. The computer-implemented method of claim 1, wherein the integrated transcoder merges first and the second transforms respectively associated with the first and second media formats in a scaling component.

7. The computer-implemented method of claim 1, wherein transcoding further comprises dynamically turning on or off one or more operations associated with residue-error compensation to respectively increase quality or speed of the transcoding.

8. The computer-implemented method of claim 1, wherein transcoding implements residue-error compensation, and wherein transcoding further comprises:

responsive to determining that early reference frames are being processed, dynamically adjusting one or more operations to transcode the early reference frames with high quality and slow speed; and

wherein the operations comprise block requantization error accumulation, motion compensation of accumulated error, and early detection to determine whether a particular block is to be an encoded.

9. The computer-implemented method of 1, wherein transcoding implements residue-error compensation, and wherein transcoding further comprises one or more of:

responsive to a threshold based drifting control mechanism, dynamically turning on or off block requantization error accumulation into a residue-error buffer;

responsive to an evaluation of block activity, dynamically turning on or off motion compensation of accumulated error in the residue error buffer; and

responsive to a determination based on a sum of motion compensation accumulated residue error and reconstructed residue from the partially decoding, dynamically turning on or off early detection to determine whether the block is to be encoded.

10. The computer-implemented method of claim 1, wherein partially decoding further comprises:

implementing rate control with a coded frame size proportional to an input frame size for all frames; and

wherein the rate control provides continued compensation for accumulated differences between a target frame size and an actual resultant frame size.

11. The computer-implemented method of claim 1, further comprising:

discarding high Discrete Cosine Transform (DCT) coefficients to downscale the intermediate data stream in a DCT domain; and

wherein explicit pixel-domain downscaling is not performed.

12. The computer-implemented method of claim 1, wherein transcoding further comprises downscaling a block of the encoded bitstream by reconstruction of a full resolution picture from the encoded bitstream independent of mixed block processing and a corresponding decoding loop.

13. The computer-implemented method of claim 1, wherein transcoding further comprises implementing two-stage downscaling operations to obtain an arbitrary downscaling target ratio.

14. The computer-implemented method of claim 13, wherein the two-stage downscaling operations further comprise:

implementing first stage downscaling operations in a DCT domain decoding loop; and

performing second stage downscaling operations outside of the decoding loop in a pixel domain.

15. The computer-implemented method of claim 13, wherein the two-stage downscaling operations further comprise:

performing first stage downscaling operations to obtain a first result that is intermediate to the arbitrary downscaling target ratio;

inputting the first result into a decoding loop; and

implementing second stage downscaling operations in a spatial domain to obtain the arbitrary downscaling target ratio.

16. A computer-implemented method comprising:

transcoding, by an integrated transcoder, an encoded bitstream as follows:

partially decoding, by the integrated transcoder, the encoded bitstream to generate an intermediate data stream, the encoded bitstream being encoded with a first set of compression techniques associated with the first media format;

encoding, by the integrated transcoder, the intermediate data stream using a second set of compression techniques to generate a transcoded bitstream, the second transform corresponding to the second media format; and

wherein the integrated transcoder merges first and second transforms respectively associated with the first and second media formats in a scaling component.

17. The computer-implemented method of claim 16, wherein the first media format is MPEG-2 and wherein the second media format is WMV.

18. The computer-implemented method of claim 16, wherein transcoding further comprises:

dynamically turning on or off one or more operations associated with residue-error compensation to respectively increase quality or speed of the transcoding; and

wherein the one or more operations are toggled on or off based on one or more of the following:

a threshold based drifting control mechanism;

evaluation of block activity levels; and

a sum of motion compensation accumulated residue error and reconstructed residue from decoding operations.

19. The computer-implemented method of claim 16, further comprising:

downscaling data associated with the encoded bitstream in a DCT domain of a decoding loop; or

downscaling the data in a two-stage downscaling process, the two-stage downscaling process comprising:

downscaling, in a first stage, the data in the DCT domain; and

reducing resolution of downscaled results of the first stage in the spatial domain.

20. A computer-implemented method comprising:

partially decoding, by an integrated transcoder, an encoded bitstream to generate an intermediate data stream, the encoded bitstream being encoded with a first set of compression techniques associated with the first media format, the partially decoding comprising:

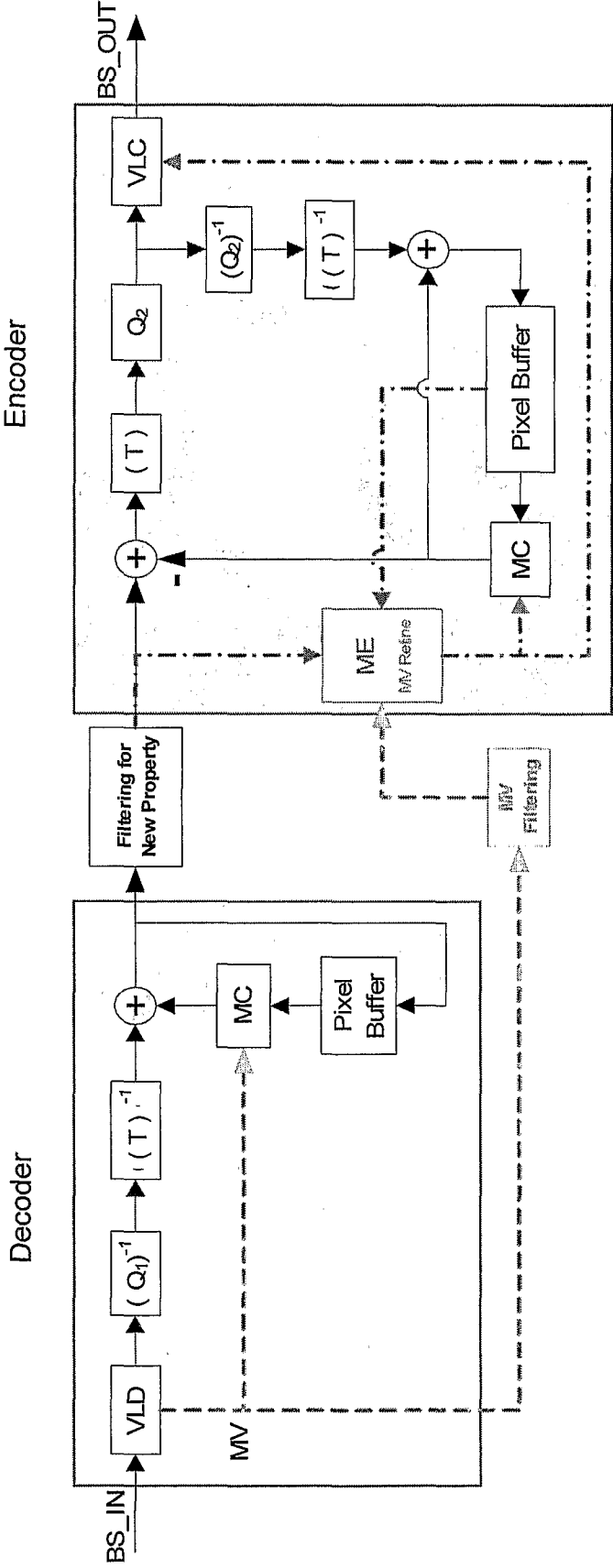
downscaling bitstream data in a DCT domain of a decoding loop;

determining whether to further reduce the bitstream data outside of the decoding loop in a spatial domain, the determining comprising:

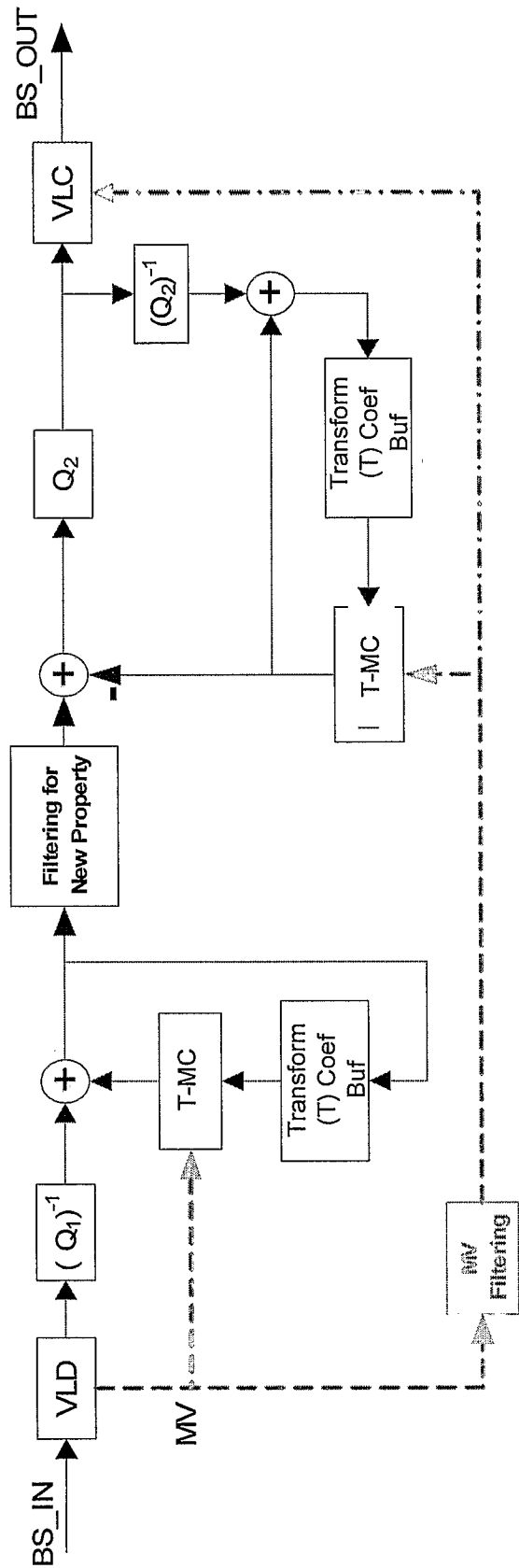
reducing resolution of DCT domain downsampled bitstream data in the spatial domain; or

allowing the DCT domain downsampled bitstream data to pass without further resolution reduction;

encoding, by the integrated transcoder, the intermediate data stream using a second set of compression techniques to generate a transcoded bitstream, the second set of compression techniques corresponding to the second media format.



Prior Art
Fig. 1



Prior Art
Fig. 2

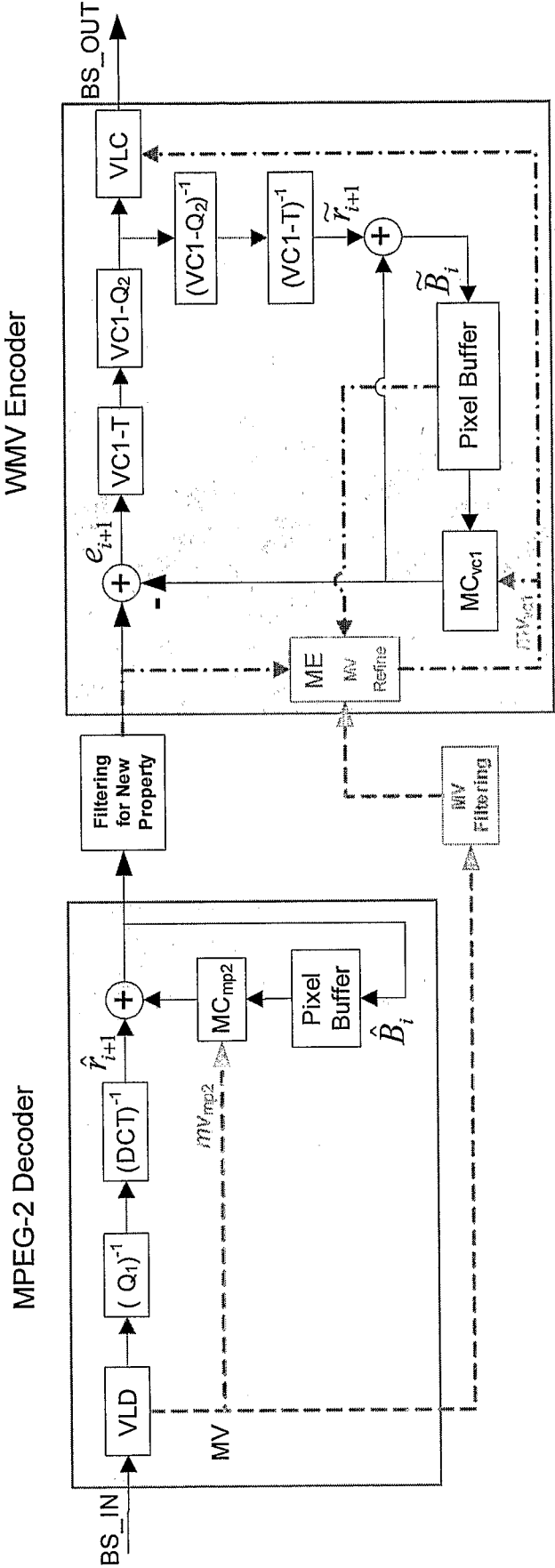


Fig. 3

4/13

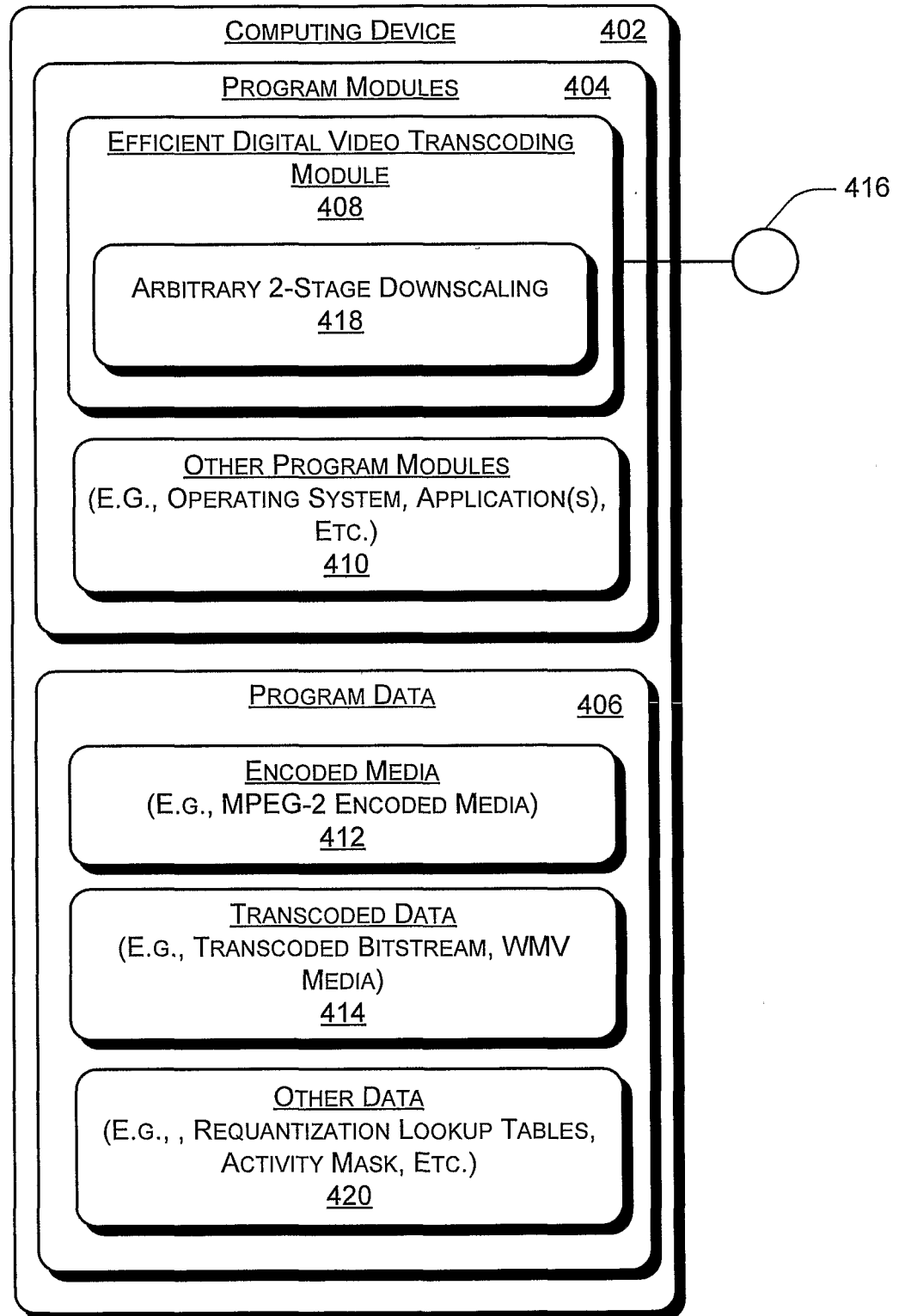
400

Fig. 4

5/13

500

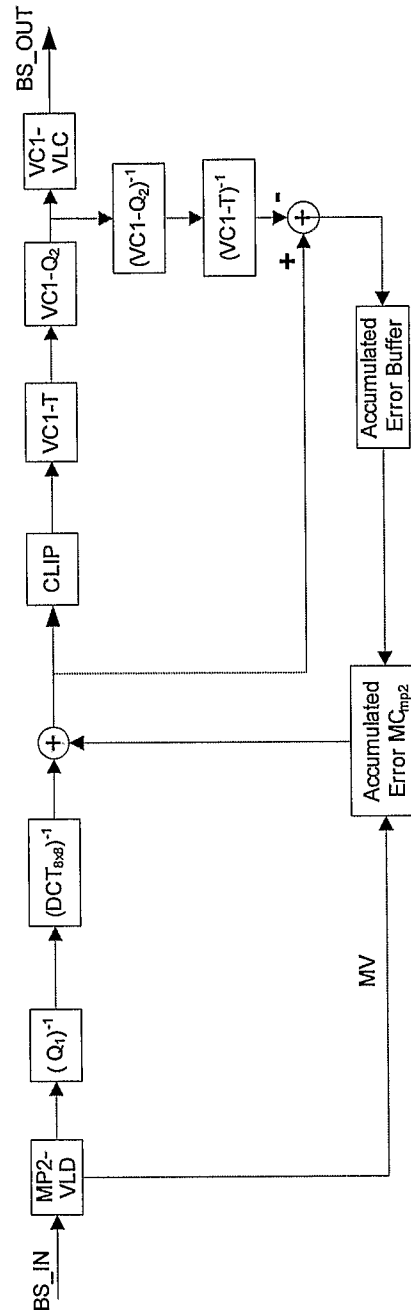


Fig. 5

6/13

600

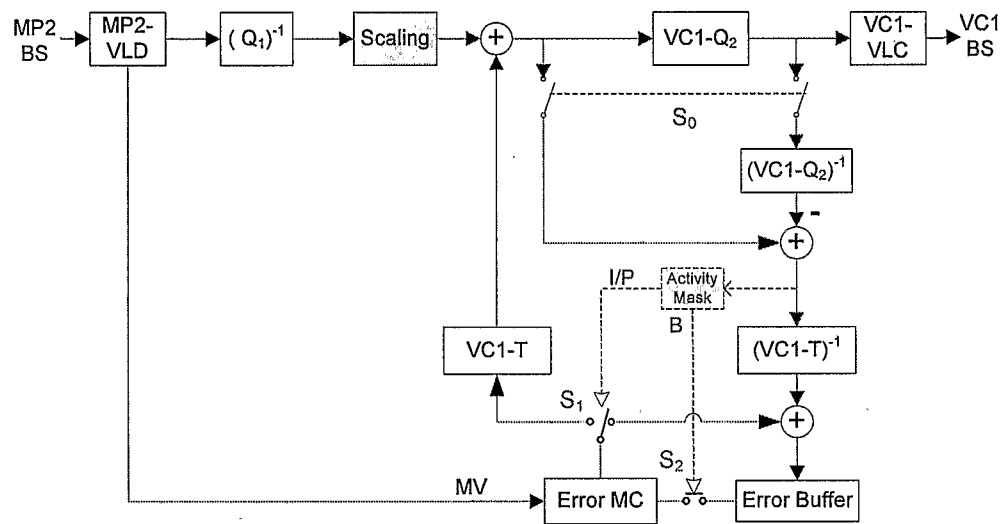


Fig. 6

7/13

700

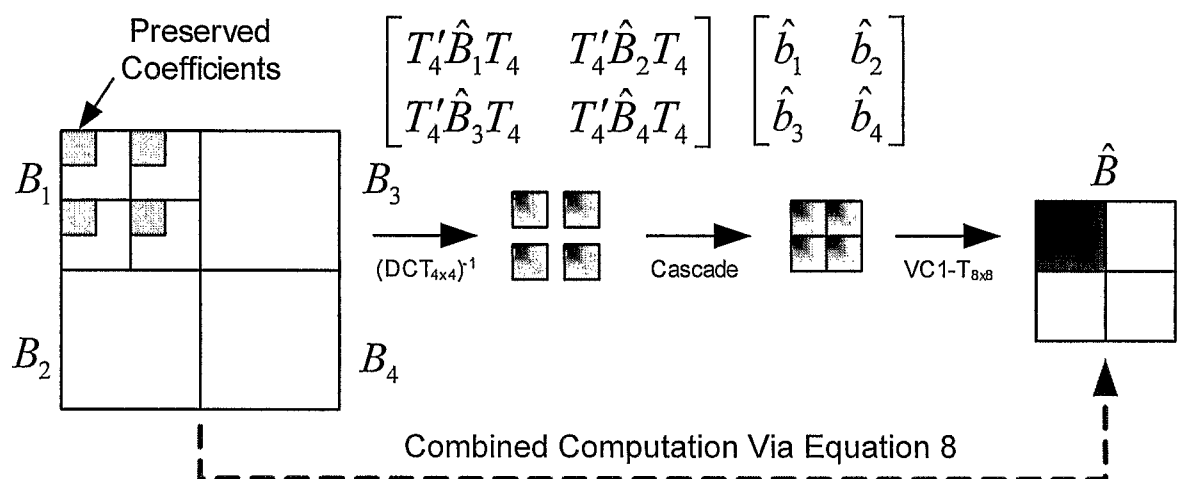


Fig. 7

8/13

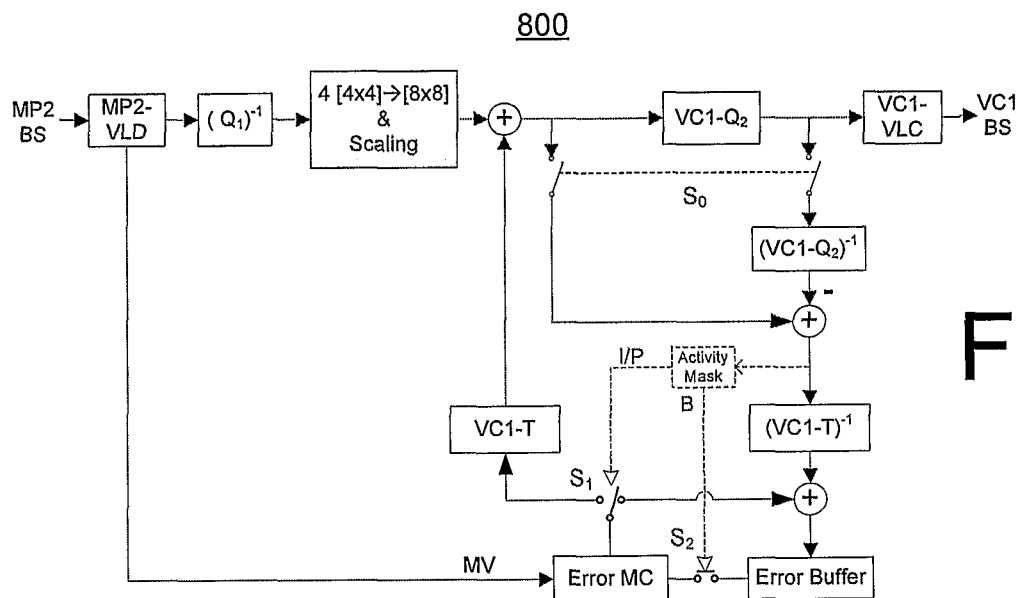


Fig. 8

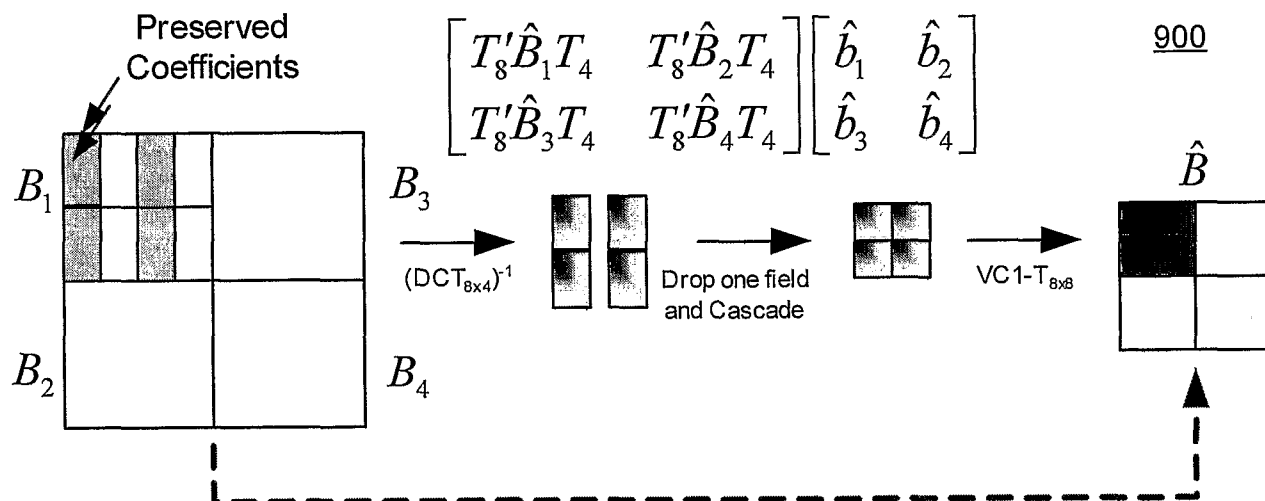


Fig. 9

9/13

1000

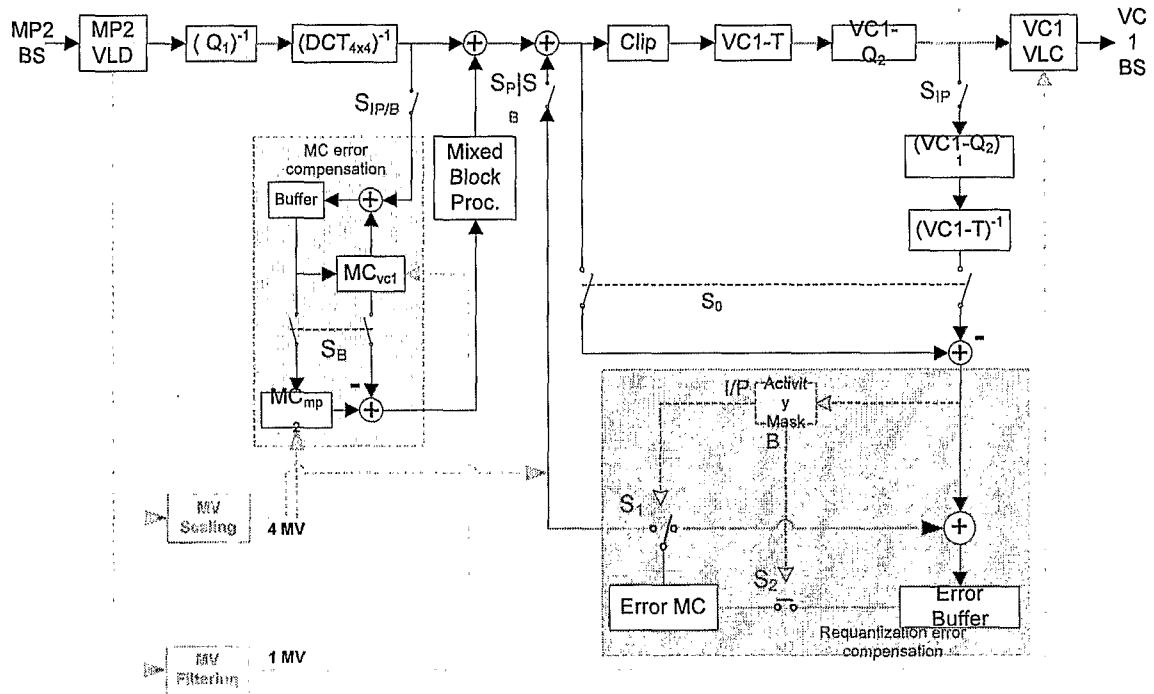


Fig. 10

10/13

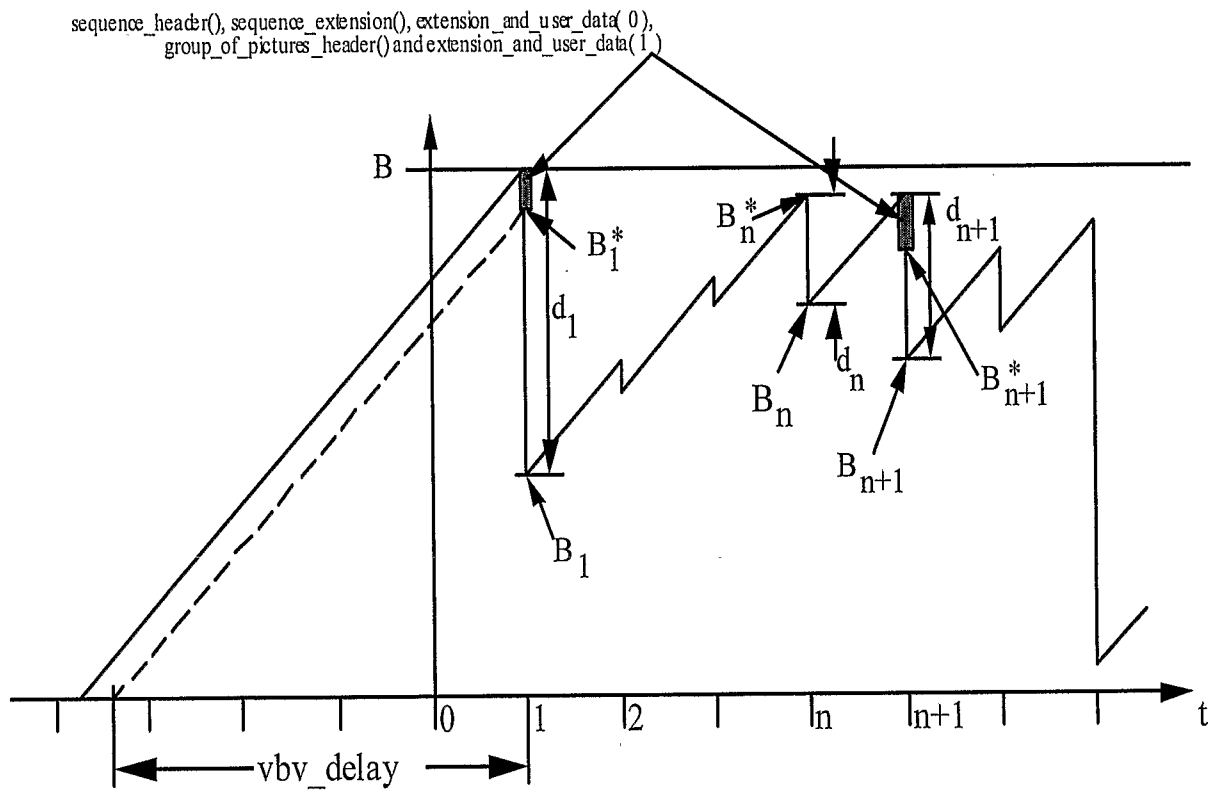
1100

Fig. 11

1200

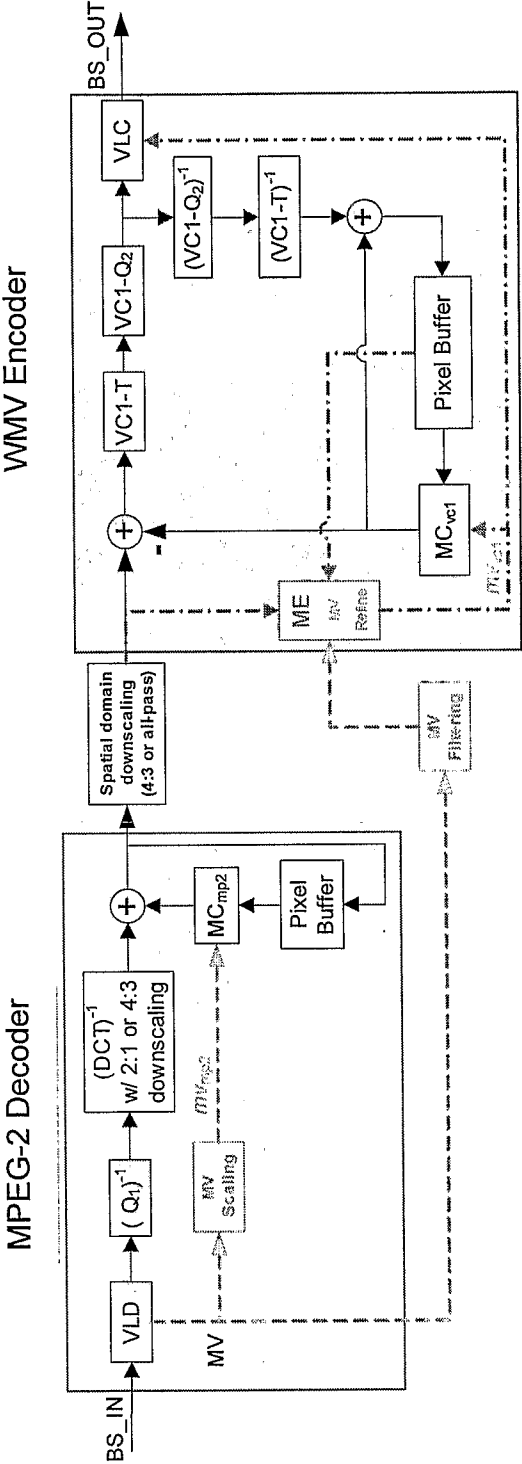


Fig. 12

12/13

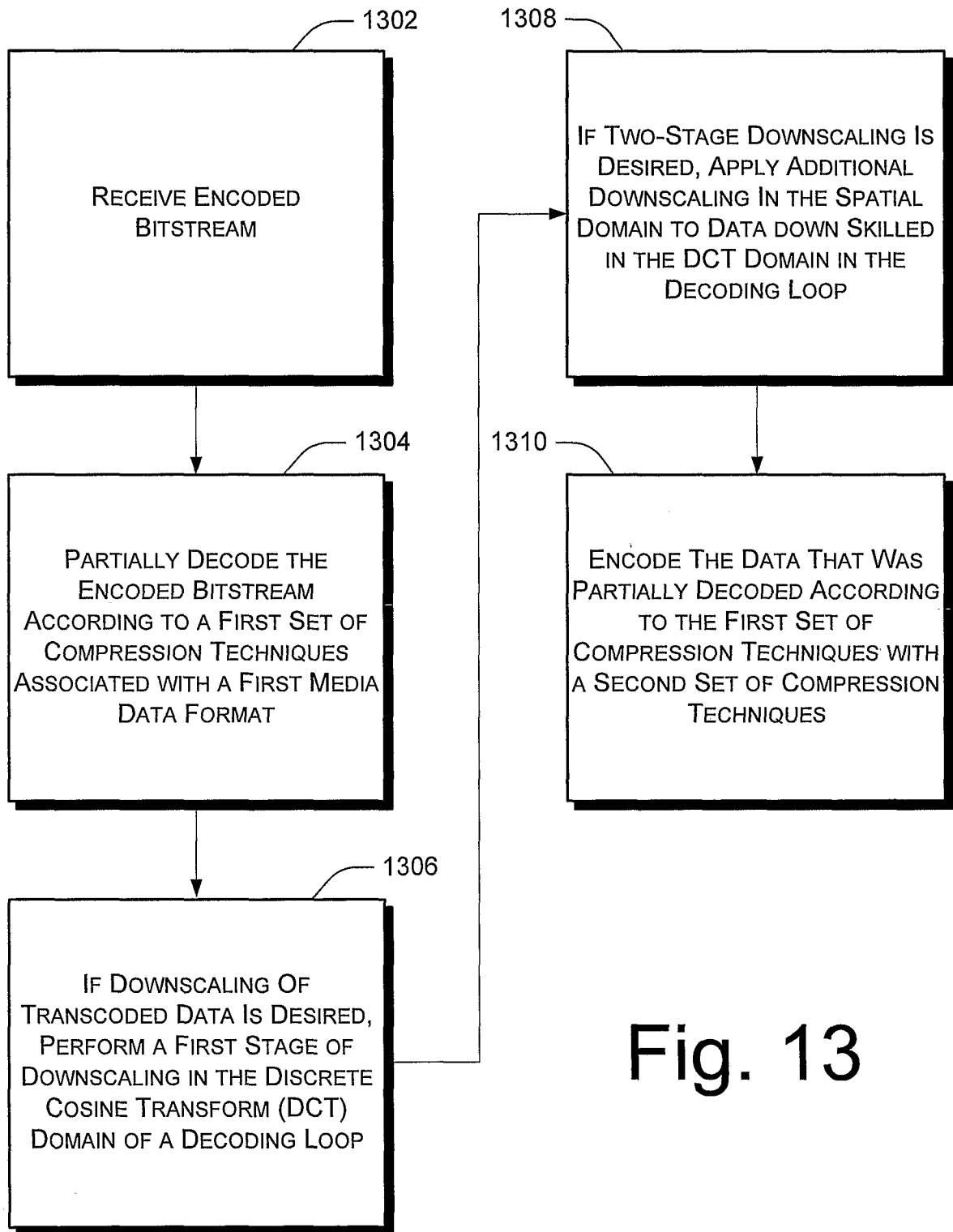
1300

Fig. 13

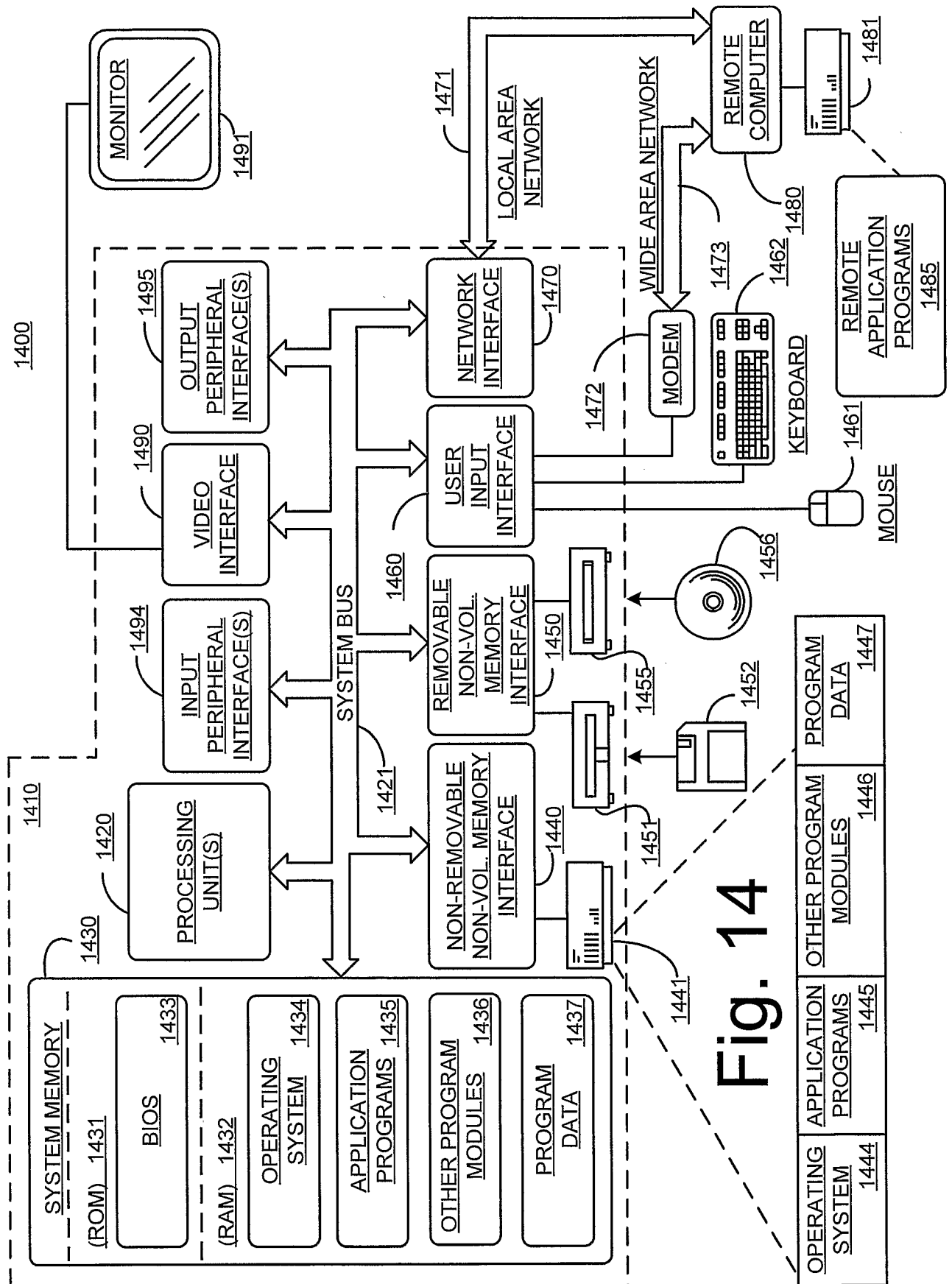


Fig. 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2006/035640**A. CLASSIFICATION OF SUBJECT MATTER****H04N 7/24(2006.01)i, H04N 7/26(2006.01)i**

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 8 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
KR : IPC as aboveElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKIPASS (KIPO internal), IEEE Xplore: "transcoder, MPEG, efficient, intermediate"**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	US 06647061 B1 (KRIT PANUSOPONE and XUEMIN CHEN) 11 November 2003 See abstract; claims 1-14; figure 3.	1-5 6-20
A	US 06393059 B1 (KENJI SUGIYAMA) 21 May 2002 See abstract; claims 2 and 3.	1-20
A	US6618442 B1 (GEORGE K CHEN et al.) 09 September 2003 See abstract; figure 6.	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

06 FEBRUARY 2007 (06.02.2007)

Date of mailing of the international search report

07 FEBRUARY 2007 (07.02.2007)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
920 Dunsan-dong, Seo-gu, Daejeon 302-701,
Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

KIM, Heung Soo

Telephone No. 82-42-481-5764



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2006/035640

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US06647061B1	11. 11. 2003	AT309676E	15. 11. 2005
		AU200167023A1	17. 12. 2001
		CA2410573A1	13. 12. 2001
		CA2522856A1	13. 12. 2001
		CN1210961C	13. 07. 2005
		CN1435056A	06. 08. 2003
		CN1681325A	12. 10. 2005
		DE60114812C0	15. 12. 2005
		EP0128770A2	05. 03. 2003
		EP0128770B1	09. 11. 2005
		EP0140120A2	24. 03. 2004
		EP0140120A3	03. 11. 2004
		JP16504739	12. 02. 2004
		JP2004504739T2	12. 02. 2004
		KR1020030014697	19. 02. 2003
		MXPA02012169A	06. 06. 2003
		TW527837B	11. 04. 2003
		US2004013195A1	22. 01. 2004
		US2004013195AA	22. 01. 2004
		W00195633A2	13. 12. 2001
		W0200195633A3	22. 08. 2002
US06393059B1	21. 05. 2002	CN1160965C	04. 08. 2004
		CN1237856A	08. 12. 1999
		JP11275592A2	08. 10. 1999
US6618442B1	09. 09. 2003	None	