(54) Title: A FINE GRANULARITY SCALABILITY SCHEME

(57) Abstract: A bit-plane coding scheme is described for the transform coefficient residues of the enhancement layer of a fine granularity scalability (FGS) scheme, e.g., as used in MPEG-4 video compression. In a first aspect, a determination is made as to the encoding of the bit planes of the enhancement layer residues either in a fixed length code manner or in a variable length code manner based on the expectation value of the residue values of the bit planes not encoded. In another aspect, the encoding is applied to either the absolute values of the residues, or to the differences between the absolute values and an average of the absolute values, together with the signs of the differences.
A Fine Granularity Scalability Scheme

The present invention relates to a fine granularity scalability (FGS) scheme, and relates particularly, but not exclusively, to the transmission of video data over a packet-switching network, such as streaming video over the Internet.

Background of the Invention

The optimization of video coding over a bit rate range has been widely studied, due to the fast growing number of network video applications, such as video streaming over the Internet or over a mobile telephone connection. The applications may for example relate to multiple-point-video-conferencing, video phones, Internet TV and HDTV.

Internet video streaming applications need to provide real-time delivery while compensating for lack of certainty in the Quality of Service of the delivery systems, with variations and unpredicability arising in bandwidth, delay variations and packet loss rates.

In order to achieve this, the MPEG-4 video compression standard has adopted Fine Granularity Scalability (FGS) for streaming applications.

FGS is composed of two layers: a base layer which is designed to meet the lower bound of the bit rate range, and an enhancement layer which meets the upper bound of the bit rate range. A receiver of the video signal (base layer and enhancement layer) may decode the base layer to provide a basic video signal, and may then decode varying amounts of the enhancement layer to improve the quality of the image.

Three types of technique were proposed for FGS: bit plane coding of the DCT residues of the enhancement layer, wavelet coding of image residues, and matching pursuit coding of image residues. Bit plane coding of the DCT residues was finally chosen by MPEG-4. In this scheme, the enhancement layer bit stream can be sliced
and packetized at the transmission time to satisfy the varying user bitrates. This makes FGS suitable for application with varying transmission bandwidth.

In order to improve the efficiency in the coding of the bit-planes, it is known to use the correlation between a bit-plane array to be encoded and bit-plane layers already encoded as well as the base layer ("Improved bit-plane coding for FGS video coding", Ralf Buschmann et al. ISO/IEC JTC1/SC29/WG11, MPEG99/M5561, December, 1999).

An aim of the present invention is to provide an improved coding of the bit-plane arrays of the residues of the enhancement layer of an FGS scheme.

**Summary of the Present Invention**

Viewed from a first aspect, the present invention provides a fine granularity scalability system for video compression, including means for encoding video data into a base layer bitstream of quantized coefficients and an enhancement layer bitstream of coefficient residues, and enhancement layer encoding means for encoding said enhancement layer bitstream in a bit-plane manner, characterized in that the enhancement layer encoding means determines whether to encode the remaining bit planes using fixed length code dependent upon an expectation value of the residue values of the bit planes not encoded.

An encoder according to the present invention can determine whether or not to use fixed length code for all remaining bit planes depending on the expectation value of the remaining values of the residues not encoded. This enables fixed length coding to be used where it will be more efficient than other coding schemes such as variable length encoding or the like. This in turn provides a more efficient encoding scheme for the enhancement layer and so the whole video sequence, and provides better quality video reconstruction.

Preferably, for a current bit plane to be encoded, the enhancement layer encoding means computes the partial residues that can be reconstructed from all of the encoded bit planes and skips those position which reach their maximum values determined by
using the knowledge of quantization parameters at the base layer, and then computes the average of the remaining value of residues at each position in order to determine whether or not to use fixed length coding for all the remaining bit planes.

Preferably, if (DL-1) is the significance level of the current bit plane to be encoded, then the bit planes of lesser significance are encoded via fixed length coding if:

\[
\frac{1}{64 - NM} \left( \tilde{R}_m - \sum_i R_{rec}(i) \right) \approx \frac{\text{Range}}{2},
\]

where \( \tilde{R}_m = \sum_i \text{ARed}(i) \), \( \tilde{R}_m \) being the sum of the residues value and ARed(i) the absolute values of the residues i); 
\( R_{rec}(i) \) are the partial residue values that can be reconstructed from the coded bit-planes; 
\( NM \) is the number of positions of the bit plane (DL-1) with \( R_{rec}(i) \) equal to the maximum value of the residues in the enhancement layer; and

\[\text{Range} = 2^{DL-1} - 1.\]

Preferably, the bit planes of lesser significance are encoded via fixed length coding if

\[
\frac{1}{64 - NM} \left( \tilde{R}_m - \sum_i R_{rec}(i) \right) \text{ is between 90% and 110% of Range/2.}
\]

In a preferred embodiment, the enhancement layer encoding means encodes the residues using either the absolute values of the residues, or the differences of the absolute values from an average of the absolute values.

This allows for further efficiencies in the encoding procedure by using whichever of the two representations of the residues is most efficient to encode.

In accordance with such an embodiment, the enhancement layer encoding means preferably includes two sub-encoders, a first sub-encoder for encoding the absolute
values of the residues $A\text{Red}(i)$, and a second sub-encoder for encoding the values $A\text{SeARed}(i)$ of each of the residues, where:

$$A\text{SeARed}(i) = |A\text{Red}(i) - R_m|,$$

and where $R_m = \left\lfloor \frac{1}{\text{Tot}} \sum_{i, \text{SIGN1}(i) \neq 0} A\text{Red}(i) + 0.5 \right\rfloor$,

$[a]$ being the largest integer less than or equal to a; $\text{SIGN1}(i)$ being the sign of each residue ($\text{SIGN1}(i) = -1$ indicating a negative sign and $\text{SIGN1}(i) = 1$ indicating a positive sign); and Tot denoting the total number of residues with $\text{SIGN1}(i) \neq 0$.

Preferably, the second sub-encoder is used to encode the enhancement layer when:

$$\frac{\max_i \{A\text{SeARed}(i)\}}{\max_i \{A\text{Red}(i)\}} \leq \frac{1}{6},$$

and the first sub-encoder is used otherwise.

Preferably, the second sub-encoder further defines a sign bit for each bit-plane array position as follows:

$$\text{SIGN2}(i) = \begin{cases} 0 & \text{A\text{Red}(i)} > R_m \\ 1 & \text{Otherwise} \end{cases}$$

The system preferably include a decoder for reconstructing the video data from the base layer bitstream and the enhancement layer bitstream, wherein, assuming that the lowest significant bit-plane array received by said decoder is the DL array, then the enhancement layer residues $\text{DEFF}(i)$ are reconstructed in accordance with the following:

$$\text{DEFF}(i) = \begin{cases} AD(i) + 2^{DL-3} + \min\{1, 2^{NFB-2}\} & AD(i) \neq 0; DL \geq 3 \\ AD(i) + \min\{1, 2^{NFB-2}\} & AD(i) \neq 0; DL = 2 \\ AD(i) & AD(i) \neq 0; DL = 1 \\ 0 & \text{Otherwise} \end{cases}$$
wherein $AD(i)$ is the partial value for the residue $i$ determined from the bit-planes received by said decoder.

Where encoding may use either the absolute values of the residues or their difference from an average value, the decoder preferably includes two sub-decoders, the first sub-decoder determining the enhancement layer residues in accordance with the above-mentioned method, and the second sub-decoder determining the enhancement layer residues $DEFF(i)$ in accordance with the following:

$$DEFF(i) = \begin{cases} R_m + (2 \times SIG2(i) - 1) \times AD(i) + 2^{DL-3} & DL \geq 3 \text{ and } SIGN(i) \neq 0 \\ R_m + (2 \times SIG2(i) - 1) \times AD(i) & DL < 3 \text{ and } SIGN1(i) \neq 0 \\ 0 & SIGN(i) = 0 \end{cases}$$

In both decoding methods, in order to obtain an enhanced video signal from the base and enhancement layer signals, the enhanced transform coefficients $ENH\_COEFF(i)$ are preferably obtained from the base layer reconstructed coefficients $BASE\_COEFF(i)$ and from the enhancement layer coefficient residues in accordance with:

$$ENH\_COEFF(i) = BASE\_COEFF(i) + SIGN1(i) \times DEFF(i).$$

The present invention also extends to a method of video compression in accordance with the above, and to encoders and decoders for use in such systems, as well as to software and hardware for implementing such systems.

The use of either the absolute values of the residues or the difference between the absolute values of the residues and an average of these values is inventive in itself, and, viewed from a further aspect, the present invention provides an FGS system for the encoding of video data including means for encoding the video data into a base layer of quantized transform coefficients and an enhancement layer of transform coefficient residues, and an enhancement layer encoding means for encoding the enhancement layer residues in a bit-plane manner, characterized in that the enhancement layer encoding means either encodes the residues using the absolute
values of the residues, or encodes the residues using the differences of the residue absolute values from an average of the absolute values.

Preferably, the encoder uses the second encoding procedure when:

\[
\max_i \left\{ \frac{ASeARe d(i)}{\max_i \{ARe d(i)\}} \right\} \leq \frac{1}{6}
\]

where ARed(i) is the absolute value of the ith coefficient residue, and ASeARe(i) is the absolute value of the ith coefficient residue minus an average of the absolute values of the coefficient residues, e.g. \( R_m \) as defined previously.

The present invention may be used in conjunction with the bit rate control scheme disclosed in the co-pending International PCT Patent Application filed in Singapore on 25 May 2001 and entitled “Bit Rate Control for Video Compression”. Thus, the bit rate control of the co-pending application may be used in the encoding of the base layer for the present FGS scheme.

**Brief Description of the Drawings**

The present invention will hereinafter be described in greater detail by reference to the attached drawings which show an example form of the invention. It is to be understood that the particularity of the drawings does not supersede the generality of the preceding description of the invention.

Figure 1 is a diagram of the structure of a typical network over which video streaming may be provided;

Figure 2 is a functional block diagram of an FGS encoder according to an embodiment of the present invention; and

Figure 3 is a functional block diagram of an FGS decoder according to an embodiment of the present invention.
Detailed Description of the Present Invention

Fig. 1 shows a typical Internet structure over which a video sequence may need to be transmitted from a source 1 to one or more receivers 2. Due to the amount of data in a video sequence, the data must be compressed, otherwise the required transmission bitrate would be unachievably high.

Thus, an encoder 3 is provided at the source 1 in order to compress the video data, and decoders 4 are provided at the receivers 2 in order to decode the data and reconstruct the video sequence. In between the encoder 1 and decoders 2, the compressed data is routed through various servers 5 and over what may be many different types of transmission channel 6 having various different characteristics, including variable bandwidth.

Various different encoding systems have been provided for the compression of video data, and, for example, MPEG video compression is often employed. The current MPEG standards are MPEG-1 and MPEG-2, which are similar in basic concept, and MPEG-4 which is able to provide a low-bandwidth multimedia format that can contain a mix of media (including recorded video images and sounds and their computer-generated counterparts), and uses the concept of "video objects" made up of a number of "video object planes" (VOPs) to transmit independent images of arbitrary shape.

In MPEG compression, a video sequence is broken into a number of Groups of Pictures, each of which comprises a number of picture frames. Each frame is broken into a series of slices, and each slice consists of a set of macroblocks comprising arrays of luminance pixels and associated chrominance pixels. The macroblocks are divided into 8x8 blocks for encoding. Each block undergoes a Discrete Cosine Transform (DCT) to provide an array of DCT coefficients that are then quantized to force various of the coefficients (generally higher frequency coefficients) to zero so as to reduce the amount of data to be transmitted. Quantization is carried out by multiplying the DCT coefficient array by a quantization matrix, each value in the matrix being scaled by a quantization parameter. The matrix and quantization
parameter can be altered on a frame-by-frame and/or block-by-block basis to alter the amount of compression. The quantized coefficients then undergo further encoding to compress the transmission data still further.

The frames in a Group of Pictures comprise an Intra-frame (I frame), which is spatially compressed (in accordance with the above method), and Inter-frames (P and/or B frames) which are also temporally compressed in a motion-compensated prediction manner. Thus, each P frame in a sequence is predicted from the frame preceding it, and each B frame is predicted from preceding and succeeding frames.

MPEG-4 also includes a Video Object layer between the frame layer and macroblock layer for specifying different independent objects within a scene.

In order to optimize video quality over a bit-rate range, e.g. in video-streaming to a number of receivers 2 over channels 6 having different bandwidth capabilities, MPEG-4 also provides a Fine Granularity Scalability (FGS) scheme in which the coding of the video data is provided by a base layer and an enhancement layer, the base layer being designed to meet the lower bound of the bit rate range and the enhancement layer meeting the upper bound of the bit-rate range.

The base layer is coded as discussed above, whilst the enhancement layer takes the reconstructed DCT coefficients of the base layer and subtracts them from the original coefficients to provide a residue that is then encoded and transmitted with the base layer. The receivers 2 of the data decode the base layer to provide a video signal based on the lowest bit rate range, and can improve the quality by decoding various amounts of the enhancement layer.

As can be seen in Fig. 2, an encoder 3 takes an input video signal, and performs a DCT transform on the signal at block 1. The DCT coefficients are then quantized at block 2 using a quantization parameter and matrix suitable for the lower bit rate of the bit rate range over which the video data is to be transmitted. After further compression at block 3 (e.g. using variable length coding of the bit-planes), the quantized coefficients are transmitted as a base layer bitstream. Blocks 4-8 provide
standard processing features to allow for Inter-frame coding using motion prediction where the referenced frame is stored in the frame memory at block 6.

In order to provide the enhancement layer, the quantized coefficients from block 2 are dequantized at block 4 and are provided to a subtractor 9. At subtractor 9, the dequantized coefficients are subtracted from the original DCT coefficients from block 1 to provide a set of coefficient residues. The residues are then coded at block 11, as discussed below in accordance with the present invention, before being transmitted with the base layer bitstream.

The base layer bitstream and enhancement layer bitstream are transmitted via the servers 5 and channels 6 to the decoders 4 of each of the receivers 2. As shown in Fig. 3, each decoder 4 decompresses the base layer bitstream at block 12, and dequantizes it at block 13. The resulting DCT coefficients are then transformed back to the pixel color space representation (of luminance and chrominance) at block 14. Standard processing for providing decoding of inter-frames are provided at blocks 15 and 16.

The enhancement layer bitstream is decompressed at block 17 using a process as described below in accordance with the present invention. The resulting residues are added to the dequantized base layer DCT coefficients at adder 18 before being transformed back to the pixel color space representation at block 19 to provide an enhanced video sequence. Various amounts of the enhancement layer bitstream may be received and decompressed depending on the characteristics of the delivery channel 6 to the decoder 4 and on the decoder 4 itself. The quality of the enhanced video will depend on the amount of the enhancement layer bitstream processed.

The present invention relates to the coding and decoding of the enhancement layer coefficient residues, and particularly to the encoding and decoding of the bit planes of the binary representations of these residues. The present invention uses a switched encoder which switches between two basic sub-encoders in accordance with a switching law described in detail below, as well as determining the appropriate manner in which to code the bit planes based upon, amongst other things, an expectation value of the coefficient residues.
The first sub-encoder encodes the absolute values ARed(i) of the residues Red(i) (1 ≤ i ≤ 64), whilst the second sub-encoder encodes the absolute values ASeARed(i) of the differences between the absolute values of the residues ARed(i) and an average value $R_m$ which is defined as follows:

$$R_m = \left\lfloor \frac{1}{\text{Tot}} \sum_{i, \text{SIGN}(i) \neq 0} \text{A{Red}(i)} + 0.5 \right\rfloor,$$

$\lfloor a \rfloor$ being the largest integer less than or equal to $a$; SIGN1(i) being the sign of each residue (SIGN1(i) = -1 indicating a negative sign and SIGN1(i) = 1 indicating a positive sign); and Tot denoting the total number of residues with SIGN1(i) ≠ 0.

The second sub-encoder also determines another sign bit, SIGN2(i), defined as follows:

$$\text{SIGN2}(i) = \begin{cases} 0 & \text{A{Red}(i)} > R_m \\ 1 & \text{Otherwise} \end{cases}.$$

Two sub-decoders are provided at the decoder 4 to compliment the sub-encoders, as discussed below.

The choice of which sub-encoder to use depends on which encoding method is determined to be most efficient, and the law for switching between the two sub-encoders uses the following inequality:

$$\max_i \{\text{ASeARed}(i)\} \leq \frac{1}{6} \max_i \{\text{A{Red}(i)}\}.$$ 

If this inequality holds, then less bits are required to encode the sign bit SIGN2(i) and ASeARed(i) as compared with the number of bits required to encode ARed(i) directly. Thus, coding efficiency is improved by encoding SIGN2(i) and ASeARed(i). On the other hand, if the inequality does not hold, then less bits are
required to encode ARed(i) directly as compared to the number of bits required to
encode the sign bit SIGN2(i) and ASeARed(i). Thus, the switching law is defined by:
If the above inequality holds, use sub-encoder 2, otherwise use sub-encoder 1.

After obtaining all of the DCT residues of e.g. a VOP, the maximum absolute of the
residues is found, and the minimum number of bit planes for coding the residues,
MNB, is decided based on this.

In the first sub-encoder, the 64 absolute values ARed(i) of the residues are zigzag
ordered into an array, and the bit planes for representing these 64 values are then
determined. Each bit plane comprises a 64-bit array, and each bit of the same array
comprises the same significant bit from each of the residues. This arranging of the
bit planes is well-known in the art.

The sub-encoder 1 then determines R_{\text{max}}(i) (1 \leq i \leq 64), the maximum value of the
residues, and initially sets DL (denoting the significance level of the bit plane to be
encoded) to the value of MNB + 1. Also, R_{\text{rec}}(i) (1 \leq i \leq 64), denoting the partial
residue values that would be reconstructed from the coded bit planes, are initially set
as 0.

Further, letting NM denote the number of positions with R_{\text{rec}}(i) = R_{\text{max}}(i), and:

\[
\text{Range} = 2^{DL-1} - 1;
\]

Then the enhancement layer sub-encoder 1 will encode the bit planes of the
significance 1, 2, ..., DL-1 using fixed length code if:

\[
\frac{1}{64 - NM} \left( \frac{\overline{R}_m - \sum_i R_{\text{rec}}(i)}{R_{\text{max}}} \right) \approx \frac{\text{Range}}{2},
\]

where \( \overline{R}_m = \sum ARed(i) \), \( \overline{R}_m \) is the sum of the residues.
This condition may be held to be true when \( \frac{1}{64 - NM} \left( \tilde{R}_m - \sum_i R_{rec}(i) \right) \) is about 90% to 110% of Range/2.

Otherwise, the encoder will encode the (DL-1)th bit plane using a variable length coding method, in which the bit plane is encoded using the symbols (RUN,EOP) - RUN being the number of consecutive 0's before a 1, and EOP (End-of-Plane) being a symbol as to whether there are any 1's left in the plane. If the (DL-1)th bit plane contains all 0's, a special symbol ALL-ZERO is formed to represent it. Also, the encoding skips positions satisfying \( R_{rec}(i) > R_{\max}(i) - 2^{DL-2} \) (as in “Improved bit-plane coding for FGS video coding”, Ralf Buschmann et al. ISO/IEC JTC1/SC29/WG11, MPEG99/M5561, December, 1999).

After the (DL-1)th bit plane is encoded, the term DL is reset as (DL-1), and a check is again made as to whether condition (4) is satisfied. The above process is then repeated until the least significant bit plane (LSB) is reached (or until the condition (4) is satisfied and the remaining bit planes are encoded using fixed length codes).

By making this determination, the present scheme is able to use fixed length encoding in situations where it is more efficient than variable rate encoding, and so provides a more efficient system.

The sign bit SIGN1(i) for each of the residues is encoded with one bit following the most significant bit of the residue as the coefficient residue’s MSB is encoded. A binary “0” denotes a positive difference and a binary “1” denotes a negative difference.

The encoded bitstream will also include information in a header of the form (R_m, SE, NFB), where SE = 0 indicates that Sub-encoder 1 has been used, SE = 1 indicates that Sub-encoder 2 has been used, and NFB is the total number of bit planes which are encoded by using fixed length code.
The second sub-encoder works in a similar manner to the first sub-encoder, except that the input of the sub-encoder is \text{AseARed}(i)$ rather than the absolute values \text{ ARed}(i)$ of the residues. Also, the second sign bit \text{SIGN2}(i)$ is determined as discussed above.

The corresponding FGS decoder 17 of Fig. 3 is composed of two sub-decoders to compliment the two sub-encoders and a corresponding switching law. In this case, the switching law is determined by the header information of the encoded bitstream received. Thus, if \text{SE} of $(R_{m}, \text{SE}, \text{NFB})$ is \text{"0\text{"}, then sub-encoder 1 was used to encode the information, and so sub-decoder 1 will be used in the decoder, whereas if \text{SE} is \text{"1\text{"}, then sub-encoder 2 was used to encode the information, and so sub-decoder 2 will be used in the decoder.

In general, the information received by the decoder will be a truncated version of the output bit stream from the FGS enhancement encoder. Thus, the decoder must first decode the available information \text{AD}(i)$, and must then reconstruct the complete DCT coefficients from this information as best as possible.

In order to decode the available information \text{AD}(i)$, the decoder first initializes the absolute values of all of the residues to zero. It then decodes the enhancement bitstream using a variable length decoding of the (RUN,EOP) symbols (and/or a fixed length decoding). These are translated into RUN numbers of consecutive 0's before a 1, and the filling of 0's to the end of the bit plane from the EOP symbols. If an ALL-ZERO symbol is decoded, the bit-plane is set to contain all 0's. Once this is achieved, the bits to the partial result of each absolute value of the residue are accumulated at their proper significant bit positions.

In sub-decoder 1, the sign bits, \text{SIGN1}(i)$ $(1 \leq i \leq 64)$, are initially set at 0, and are decoded from the enhancement layer bit stream immediately after the (RUN, EOP) code corresponding to the MSB of the non-zero residues. The partial absolutes of all of the residues are then obtained as:
\[
DEFF(i) = \begin{cases} 
AD(i) + 2^{DL-3} + \min \{1, 2^{NFB-2} \} & AD(i) \neq 0; DL \geq 3 \\
AD(i) + \min \{1, 2^{NFB-2} \} & AD(i) \neq 0; DL = 2 \\
AD(i) & AD(i) \neq 0; DL = 1 \\
0 & \text{Otherwise}
\end{cases}
\]

Sub-decoder 2 operates in a similar manner, and from the available information AD(i), the partial absolutes of all of the residues can be obtained as follows:

\[
DEFF(i) = \begin{cases} 
R_m + (2 \cdot \text{SIGN2}(i) - 1) \cdot (AD(i) + 2^{DL-3}) & DL \geq 3 \text{ and } \text{SIGN1}(i) \neq 0 \\
R_m + (2 \cdot \text{SIGN2}(i) - 1) \cdot AD(i) & DL < 3 \text{ and } \text{SIGN1}(i) \neq 0 \\
0 & \text{SIGN1}(i) = 0
\end{cases}
\]

For both sub-decoders, the enhanced DCT coefficients comprising the contribution from the base layer and the enhancement layer can then be determined as:

\[
\text{ENH_COEFF}(i) = \text{BASE_COEFF} + \text{SIGN1}(i) \cdot \text{DEFF}(i).
\]

Thus, the present embodiment determines whether it would be most efficient to code the bit planes with variable length or fixed length coding to provide improved efficiency. It also determines whether to encode the absolute values of the residues or to encode the absolute values of the differences between the residues and an average of the residues, together with a sign of the difference, depending again on which is most efficient.

It is to be understood that various alterations additions and/or modifications may be made to the parts previously described without departing from the ambit of the invention, and that, in the light of the teachings of the present invention, the FGS encoding scheme may be implement in software and/or hardware in a variety of manners.
Claims

1. A fine granularity scalability system for video compression, including means for encoding video data into a base layer bitstream of quantized coefficients and an enhancement layer bitstream of coefficient residues, and enhancement layer encoding means for encoding said enhancement layer bitstream in a bit-plane manner, characterized in that said enhancement layer encoding means determines whether to encode the remaining bit planes using fixed length code dependent upon an expectation value of the residue values of the bit planes not encoded.

2. The system of claim 1, wherein for the remaining bit planes to be coded, the enhancement layer encoding means determines the partial residues that may be reconstructed by the significant bit planes down to the current bit plane to be encoded, and determines the difference between the sum of the partial residues and the sum of the whole residues in order to determine whether or not to use fixed length coding for the whole of said remaining bit planes.

3. The system of claim 2, wherein said enhancement layer encoding means compares said difference with the maximum residue value obtainable from the encoded bit planes.

4. The system of any preceding claim, wherein, if DL-1 is the significance level of the current bit plane to be encoded, then the bit planes of lesser significance are encoded in a fixed length code manner if:

\[
\frac{1}{64 - NM} \left( \bar{R}_m - \sum_i R_{res}(i) \right) \approx \frac{Range}{2},
\]

where \( \bar{R}_m = \sum ARe\ d(i) \), \( \bar{R}_m \) being the sum of the residues and ARe\(d(i)\) the absolute values of the residues \(i\); \(R_{res}(i)\) are the partial residue values that can be reconstructed by the coded bit-planes;
NM is the number of positions i of the bit plane DL-1 with $R_{rec}(i)$ equal to the maximum value of the residues in the enhancement layer; and

$$Range = 2^{DL-1} - 1.$$ 

5. The system of claim 4, wherein the bit planes of lesser significance are encoded in a fixed length code manner if

$$\frac{1}{64 - NM} \left( \bar{R}_m - \sum R_{rec}(i) \right)$$

is between 90% and 110% of Range/2.

6. The system of any preceding claim, wherein said enhancement layer encoding means either encodes said residues using the absolute values of said residues, or encodes said residues using the differences of the said absolute values from an average of said absolute values.

7. The system of any preceding claim, wherein the enhancement layer encoding means includes two sub-encoders, a first sub-encoder for encoding the absolute values of the residues $ARe_d(i)$, and a second sub-encoder for encoding the values $ASeARe_d(i)$ of each of the residues, where:

$$ASeARe_d(i) = |ARe_d(i) - R_m|,$$

and where

$$R_m = \left| \frac{1}{Tot} \sum_{i, SIGN1(i) \neq 0} ARe_d(i) + 0.5 \right|,$$

$[a]$ being the largest integer less than or equal to a; SIGN1(i) being the sign of each residue (SIGN1(i) = -1 indicating a negative sign and SIGN1(i) = 1 indicating a positive sign); and Tot denoting the total number of residues with SIGN1(i) $\neq 0$.

8. The system of claim 7, wherein said second sub-encoder further defines a sign bit for each residue as follows:

$$SIGN2(i) = \begin{cases} 
0 & ARe_d(i) > R_m, \\
1 & Otherwise
\end{cases}.$$
9. The system of claim 7 or 8, wherein said second sub-encoder is used to encode the enhancement layer when:

$$\max_i \left\{ \frac{A Se A Re d(i)}{\max_i \{ A Re d(i) \} } \right\} \leq \frac{1}{6},$$

and said first sub-encoder is used otherwise.

10. The system of any of claims 1 to 5, including a decoder for reconstructing the video data from the base layer bitstream and the enhancement layer bitstream, wherein, assuming that the lowest significant bit-plane array received by said decoder is the DL array, then the enhancement layer residues are reconstructed in accordance with the following:

$$DEFF(i) = \begin{cases} AD(i) + 2^{DL-3} + \min \{1, 2^{NFB-2}\} & AD(i) \neq 0; DL \geq 3 \\ AD(i) + \min \{1, 2^{NFB-2}\} & AD(i) \neq 0; DL = 2 \\ AD(i) & AD(i) \neq 0; DL = 1 \\ 0 & Otherwise \end{cases}$$

wherein AD(i) is the partial value for the residue i determined from the bit-planes received by said decoder; and

NFB is the number of bit planes coded via fixed length coding.

11. The system of claims 7 to 10, wherein said decoder includes two sub-decoders, the first sub-decoder determining the enhancement layer coefficient residues in accordance with claim 10, and the second sub-decoder determining the enhancement layer residues in accordance with the following:

$$DEFF(i) = \begin{cases} R_m + (2 \cdot SIGN2(i) - 1) \cdot (AD(i) + 2^{DL-3}) & DL \geq 3 \ and \ SIGN1(i) \neq 0 \\ R_m + (2 \cdot SIGN2(i) - 1) \cdot AD(i) & DL < 3 \ and \ SIGN1(i) \neq 0 \\ 0 & SIGN1(i) = 0 \end{cases}$$

12. The system of claim 10 or 11, wherein enhanced transform coefficients ENH_COEFF(i) for reconstructing the video data are obtained from the base layer
reconstructed coefficients BASE_COEFF(i) and from the enhancement layer coefficient residues DEFF(i) in accordance with:

\[ \text{ENH}_\text{COEFF}(i) = \text{BASE}_\text{COEFF}(i) + \text{SIGN1}(i) \times \text{DEFF}(i). \]

13. An FGS system for the encoding of video data including means for encoding the video data into a base layer of quantized transform coefficients and an enhancement layer of transform coefficient residues, and an enhancement layer encoding means for encoding said enhancement layer in a bit-plane manner, characterized in that said enhancement layer encoding means either encodes said residues using the absolute values of said residues, or encodes said residues using the differences of the residue absolute values from an average of the absolute values.

14. A method of video compression using fine granularity scalability, including the step of encoding video data into a base layer bitstream of quantized coefficients and an enhancement layer bitstream of coefficient residues, and the step of encoding the enhancement layer bitstream in a bit-plane manner, characterized in that the method includes the step of determining whether to encode the remaining bit planes using fixed length code dependent upon an expectation value of the residue values of the bit planes not encoded.

15. Computer software for video compression using fine granularity scalability, including a software component for encoding video data into a base layer bitstream of quantized coefficients and an enhancement layer bitstream of coefficient residues, and a software component for encoding the enhancement layer bitstream in a bit-plane manner, characterized in that the software includes a component for determining whether to encode the remaining bit plane using fixed length code dependent upon an expectation value of the residue values of the bit planes not encoded.
Fig. 1

Fig. 2

Input Video

DCT

Q

Frame Memory

Motion Compensation

Motion Estimation

Frame Memory

VLC

Enhanced Bitstream

Bit Plane VLC

Base Layer Bitstream
Fig. 3
**INTERNATIONAL SEARCH REPORT**

**A. CLASSIFICATION OF SUBJECT MATTER**

| IPC | H04N7/26 | H04N7/30 |

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

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

| IPC | H04N |

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

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

EPO-Internal, INSPEC, COMPENDEX

**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>A</td>
<td>RADHA H M ET AL: &quot;The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP&quot; IEEE TRANSACTIONS ON MULTIMEDIA, MARCH 2001, IEEE, USA, vol. 3, no. 1, March 2001 (2001-03), pages 53-68, XP002191171 ISSN: 1520-9210 paragraph <em>00III</em>; figures 3,4 paragraph <em>0III</em></td>
<td>1-15</td>
</tr>
</tbody>
</table>

Date of the actual completion of the international search: 22 February 2002

Date of mailing of the international search report: 07/03/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Tx: 31 651 epo nl, Fax: (+31-70) 340-3016

Authorized officer: Georgiou, G
<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>A</td>
<td>EP 0 782 341 A (THOMSON CONSUMER ELECTRONICS) 2 July 1997 (1997-07-02) abstract; figure 1</td>
<td>1-15</td>
</tr>
<tr>
<td>A</td>
<td>EP 0 548 430 A (MATSUSHITA ELECTRIC IND CO LTD) 30 June 1993 (1993-06-30) column 1, line 11 - line 55; figure 8</td>
<td>1-15</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>EP 0782341</td>
<td>02-07-1997</td>
<td>CN 1158050 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 0782341 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 9200741 A</td>
</tr>
</tbody>
</table>