Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2737888 C 2012/08/07

(11)(21) 2 737 888

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(22) Date de dépôt/Filing Date: 2002/08/08

(41) Mise à la disp. pub./Open to Public Insp.: 2003/02/27

(45) Date de délivrance/Issue Date: 2012/08/07

(62) Demande originale/Original Application: 2 576 161

(30) Priorités/Priorities: 2001/08/09 (US60/311,436); 2001/11/30 (US60/319,018); 2002/05/02 (US10/139,036)

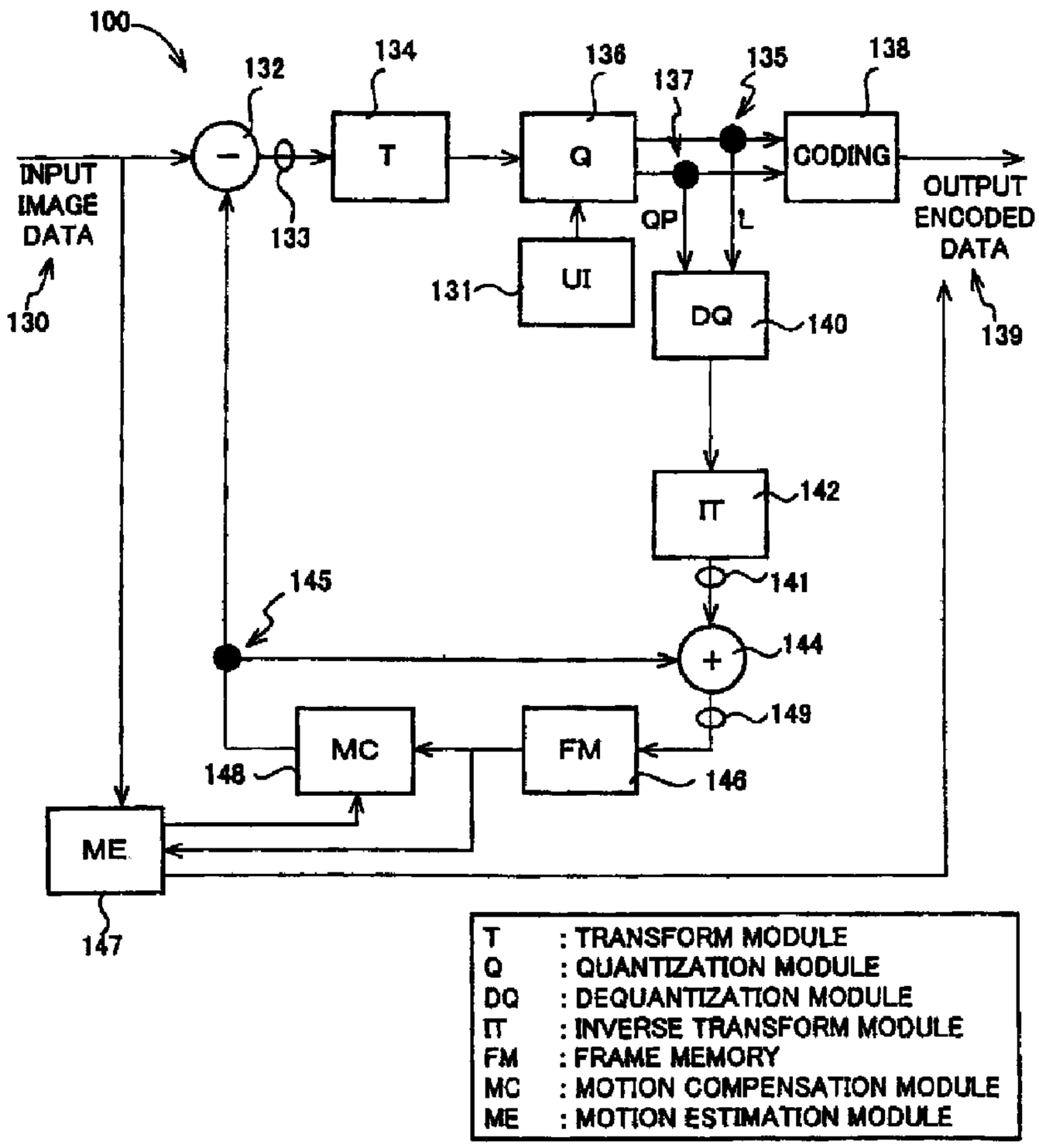
(51) Cl.Int./Int.Cl. *H04N 7/30* (2006.01), *H04N 7/50* (2006.01)

(72) Inventeur/Inventor:
KEROFSKY, LOUIS JOSEPH, US

(73) Propriétaire/Owner:
SHARP KABUSHIKI KAISHA, JP

(74) Agent: G. RONALD BELL & ASSOCIATES

(54) Titre : PROCEDE POUR LA QUANTIFICATION DE LA PROFONDEUR DE BITS
(54) Title: METHOD FOR REDUCED BIT-DEPTH QUANTIZATION



(57) Abrégé/Abstract:
A method is provided for the quantization of a coefficient. The method comprises: supplying a coefficient K; supplying a quantization parameter (QP); forming a quantization value (I) from the coefficient K using a mantissa portion (Am (QP)) and an

CA 2737888 C 2012/08/07

(11)(21) 2 737 888

(13) C

(57) Abrégé(suite)/Abstract(continued):
exponential portion ($x^{Ae(QP)}$). Typically, the value of x is 2. In some aspects of the method, forming a quantization value (I) from the coefficient K includes L = K*A(QP) = K*Am(QP)* ($2^{Ae(QP)}$). In other aspects, the method further comprises: normalizing the quantization value by 2^N as follows Ln = $L/2^N$ = K*Am(QP)/$2^{(N- Ae(QP))}$. In some aspects, forming a quantization value includes forming a set of recursive quantization factors with a period P, where A(QP+P) = A(QP)/x. Forming recursive quantization factors includes forming recursive mantissa factors, where Am(QP) = Am(QP mod P), and forming recursive quantization factors includes forming recursive exponential factors, where Ae(QP) = Ae(QP mod P) - QP/P.

ABSTRACT

A method is provided for the quantization of a coefficient. The method comprises: supplying a coefficient K; supplying a quantization parameter (QP); forming a quantization value (l) from the coefficient K using a mantissa portion (Am (QP)) and an exponential portion ($x^{Ae(QP)}$). Typically, the value of x is 2. In some aspects of the method, forming a quantization value (l) from the coefficient K includes $L = K*A(QP) = K*Am(QP)* (2^{Ae(QP)})$. In other aspects, the method further comprises: normalizing the quantization value by 2^N as follows $Ln = L/2^N = K*Am(QP)/2^{(N-Ae(QP))}$. In some aspects, forming a quantization value includes forming a set of recursive quantization factors with a period P, where A(QP+P) = A(QP)/x. Forming recursive quantization factors includes forming recursive mantissa factors, where Am(QP) = Am(QP mod P), and forming recursive quantization factors includes forming recursive exponential factors, where Ae(QP) = Ae(QP mod P) - QP/P.

METHOD FOR REDUCED BIT-DEPTH QUANTIZATION

This application is a divisional of Canadian Patent Application No. 2,576,161, filed on February 15, 2007 which is a divisional of Canadian Patent Application No. 2,454,626, filed on August 8, 2002.

The claims of the present application are directed to an image coding apparatus and method for encoding an image data.

The retention of any features which may be more particularly related to the parent application or a separate divisional thereof should not be regarded as rendering the teachings and claiming ambiguous or inconsistent with the subject matter defined in the claims of the divisional application presented herein when seeking to interpret the scope thereof and the basis in this disclosure for the claims recited herein.

FIELD OF THE INVENTION

This invention generally relates to video compression techniques and, more particularly, to a method for reducing the bit size required in the computation of video coding transformations.

BACKGROUND OF THE INVENTION

A video information format provides visual information suitable to activate a television screen, or store on a video tape. Generally, video data is organized in a hierarchical order. A video sequence is divided into group of frames, and each group can be composed of a series of single frames. Each frame is roughly equivalent to a still picture, with the still pictures being updated often enough to simulate a presentation of continuous motion. A frame is further divided into slices, or horizontal sections which helps system design of error resilience. Each slice is coded independently so that errors do not propagate across slices. A slice consists of macroblocks. In H.26P and Motion Picture Experts Group (MPEG)-X standards, a macroblock is made up of 16 x 16 luma pixels and a corresponding set of chroma pixels, depending on the video format. A macroblock always has an integer number of blocks, with the 8 x 8 pixel matrix being the smallest coding unit.

Video compression is a critical component for any application which requires transmission or storage of video data. Compression techniques compensate for motion by reusing stored information in different areas of the frame (temporal redundancy). Compression also occurs by transforming data in the spatial domain to the frequency domain. Hybrid digital video compression, exploiting temporal redundancy by motion compensation and spatial redundancy by transformation, such as Discrete Cosine Transform (DCT), has been adapted in H.26P and MPEG-X international standards as the basis.

As stated in US Patent 6,317,767 (Wang), DCT and inverse discrete cosine transform (IDCT) are widely used operations in the signal processing of image data. Both are used, for example, in the international standards for moving picture video compression put forth by the MPEG. DCT has certain properties that produce simplified and efficient coding models. When applied to a matrix of pixel data, the DCT is a method of decomposing a block of data into a weighted sum of spatial frequencies, or DCT coefficients. Conversely, the IDCT is used to transform a matrix of DCT coefficients back to pixel data.

Digital video (DV) codecs are one example of a device using a DCT-based data compression method. In the blocking stage, the image frame is divided into N by N blocks of pixel information including, for example, brightness and color data for each pixel. A common block size is eight pixels horizontally by eight pixels vertically. The pixel blocks are then "shuffled" so that several blocks from different portions of the image are grouped together. Shuffling enhances the uniformity of image quality.

Different fields are recorded at different time incidents. For each block of pixel data, a motion detector looks for the difference between two fields of a frame. The motion information is sent to the next processing stage. In the next stage, pixel information is transformed using a DCT. An 8-8 DCT, for example, takes eight inputs and returns eight outputs in both vertical and horizontal directions. The resulting DCT coefficients are then weighted by multiplying each block of DCT coefficients by weighting constants.

The weighted DCT coefficients are quantized in the next stage. Quantization rounds off each DCT coefficient within a certain range of values to be the same number. Quantizing tends to set the higher frequency components of the frequency

matrix to zero, resulting in much less data to be stored. Since the human eye is most sensitive to lower frequencies, however, very little perceptible image quality is lost by this stage.

The quantization stage includes converting the two-dimensional matrix of quantized coefficients to a one-dimensional linear stream of data by reading the matrix values in a zigzag pattern and dividing the one-dimensional linear stream of quantized coefficients into segments, where each segment consists of a string of zero coefficients followed by a non-zero quantized coefficient. Variable length coding (VLC) then is performed by transforming each segment, consisting of the number of zero coefficients and the amplitude of the non-zero coefficient in the segment, into a variable length codeword. Finally, a framing process packs every 30 blocks of variable length coded quantized coefficients into five fixed-length synchronization blocks.

Decoding is essentially the reverse of the encoding process described above. The digital stream is first deframed. Variable length decoding (VLD) then unpacks the data so that it may be restored to the individual coefficients. After inverse quantizing the coefficients, inverse weighting and an inverse discrete cosine transform (IDCT) are applied to the result. The inverse weights are the multiplicative inverses of the weights that were applied in the encoding process. The output of the inverse weighting function is then processed by the IDCT.

Much work has been done studying means of reducing the complexity in the calculation of DCT and IDCT. Algorithms that compute two-dimensional IDCTs are called "type I" algorithms. Type I algorithms are easy to implement on a parallel machine, that is, a computer formed of a plurality of processors operating simultaneously in parallel. For example, when using N parallel processors to perform a matrix multiplication on N x N matrices, N column multiplies can be simultaneously performed. Additionally, a parallel machine can be designed so as to contain special hardware or software instructions for performing fast matrix transposition.

One disadvantage of type I algorithms is that more multiplications are needed. The computation sequence of type I algorithms involves two matrix multiplies separated by a matrix transposition which, if N=4, for example, requires 64 additions and 48 multiplications for a total number of 112 instructions. It is well known by those skilled in the art that multiplications are very time-consuming for processors to perform

and that system performance is often optimized by reducing the number of multiplications performed.

A two-dimensional IDCT can also be obtained by converting the transpose of the input matrix into a one-dimensional vector using an L function. Next, the tensor product of constant a matrix is obtained. The tensor product is then multiplied by the one-dimensional vector L. The result is converted back into an N x N matrix using the M function. Assuming again that N=4, the total number of instructions used by this computational sequence is 92 instructions (68 additions and 24 multiplications). Algorithms that perform two-dimensional IDCTs using this computational sequence are called "type II" algorithms. In type II algorithms, the two constant matrices are grouped together and performed as one operation. The advantage of type II algorithms is that they typically require fewer instructions (92 versus 112) and, in particular, fewer costly multiplications (24 versus 48). Type II algorithms, however, are very difficult to implement efficiently on a parallel machine. Type II algorithms tend to reorder the data very frequently and reordering data on a parallel machine is very time-intensive.

There exist numerous type I and type II algorithms for implementing IDCTs, however, dequantization has been treated as an independent step depending upon DCT and IDCT calculations. Efforts to provide bit exact DCT and IDCT definitions have led to the development of efficient integer transforms. These integer transforms typically increase the dynamic range of the calculations. As a result, the implementation of these algorithms requires processing and storing data that consists of more than 16 bits.

It would be advantageous if intermediate stage quantized coefficients could be limited to a maximum size in transform processes.

It would be advantageous if a quantization process could be developed that was useful for 16-bit processors.

It would be advantageous if a decoder implementation, dequantization, and inverse transformation could be implemented efficiently with a 16-bit processor. Likewise, it would be advantageous if the multiplication could be performed with no more than 16 bits, and if memory access required no more than 16 bits.

SUMMARY OF THE INVENTION

The present invention is an improved process for video compression. Typical video coding algorithms predict one frame from previously coded frames. The error is subjected to a transform and the resulting values are quantized. The quantizer controls the degree of compression. The quantizer controls the amount of information used to represent the video and the quality of the reconstruction.

The problem is the interaction of the transform and quantization in video coding. In the past the transform and quantizer have been designed independently. The transform, typically the discrete cosine transform, is normalized. The result of the transform is quantized in standard ways using scalar or vector quantization. In prior work, MPEG-1, MPEG-2, MPEG-4, H.261, H.263, the definition of the inverse transform has not been bit exact. This allows the implementer some freedom to select a transform algorithm suitable for their platform. A drawback of this approach is the potential for encoder/decoder mismatch damaging the prediction loop. To solve this mismatch problem portions of the image are periodically coded without prediction. Current work, for example H.26L, has focused on using integer transforms that allow bit exact definition. Integer transforms may not normalized. The transform is designed so that a final shift can be used to normalize the results of the calculation rather than intermediate divisions. Quantization also requires division. H.26L provides an example of how these integer transforms are used along with quantization.

In the current H.26L Test Model Long-term (TML), normalization is combined with quantization and implemented via integer multiplications and shifts following forward transform and quantization and following dequantization and inverse transform. H.26L TML uses two arrays of integers A(QP) and B(QP) indexed by quantization parameter (QP), see Table 1. These values are constrained by the relation shown below in Equation 1.

Table 1 TML quantization parameters

QP	$A_{TML}(QP)$	$B_{TML}(QP)$
0	620	3881
1	553	4351
2	492	4890
3	439	5481
4	391	6154
5	348	6914
6	310	7761
7	276	8718
8	246	9781
9	219	10987
10	195	12339
11	174	13828
12	155	15523
13	138	17435
14	123	19561
15	110	21873
16	98	24552
17	87	27656
18	78	30847
19	69	34870
20	62	38807
21	55	43747
22	49	49103
23	44	54683
24	39	61694
25	35	68745
26	31	77615
27	27	89113
28	24	100253
29	22	109366
30	19	126635
31	17	141533

Equation 1 Joint Normalization/Quantization relation

$$A(QP)\cdot B(QP)\cdot 676^2 = 2^{40}$$

Normalization and quantization are performed simultaneously using these integers and divisions by powers of 2. Transform coding in H.26L uses a 4x4 block size and an integer transform matrix T, Equation 2. For a 4x4 block X, the transform coefficients K are calculated as in Equation 3. From the transform coefficients, the quantization levels, L, are calculated by integer multiplication. At the decoder the levels are used to calculate a new set of coefficients, K'. Additional integer matrix transforms followed by a shift are used to calculate the reconstructed values X'. The encoder is allowed freedom in calculation and rounding of the forward transform. Both encoder and decoder must compute exactly the same answer for the inverse calculations.

Equation 2 H.26L test model 8 transform matrix

$$T = \begin{pmatrix} 13 & 13 & 13 & 13 \\ 17 & 7 & -7 & -17 \\ 13 & -13 & -13 & 13 \\ 7 & -17 & 17 & -7 \end{pmatrix}$$

Equation 3 TML DCT_LUMA and IDCT_LUMA

$$Y = T \cdot X$$
$$K = Y \cdot T^{T}$$
$$L = (A_{TML}(QP) \cdot K)/2^{20}$$
$$K' = B_{TML}(QP) \cdot L$$
$$Y' = T^{T} \cdot K'$$
$$X' = (Y' \cdot T)/2^{20}$$

Where the intermediate result Y is the result of a new dimensional transform and the intermediate result Y' is the result of a one dimensional inverse transform.

The dynamic range required during these calculations can be determined. The primary application involves 9-bit input, 8 bits plus sign, the dynamic range required by intermediate registers and memory accesses is presented in Table 2.

Table 2 Dynamic range of TML transform and inverse transform (bits)

9-bit input	LUMA Transform	Inverse Transform
Register	30	27
Memory	21	26

To maintain bit-exact definitions and incorporate quantization, the dynamic range of intermediate results can be large since division operations are postponed. The present invention combines quantization and normalization, to eliminate the growth of dynamic range of intermediate results. With the present invention the advantages of bit exact inverse transform and quantization definitions are kept, while controlling the bit depth required for these calculations. Reducing the required bit depth reduces the complexity required of a hardware implementation and enables efficient use of single instruction multiple data (SIMD) operations, such as the Intel MMX instruction set.

The present invention provides a video decoding method for reconstruction of a sample from a quantized level L[i][j], comprising the steps a) inputting the quantized level L[i][j] (200); b) performing a combined dequantization and normalization of the quantized levels L[i][j] to obtain an integer transform coefficient K[i][j]; wherein the combined quantization and normalization step includes b1) inputting a quantization parameter QP (206); b2) defining a combined dequantization and normalization matrix B(QP)[i][j] eliminating the need to compensate for the normalization differences due to the different norms of the basis functions of an inverse integer transform to be carried out afterwards under step c); b3) representing said combined dequantization and normalization matrix B(QP)[i][j] in mantissa exponent format with a mantissa portion matrix Bm(QP)[i][j] being a function of the quantization parameter QP and an exponential portion matrix Be(QP)[i][j] being a function of the quantization parameter QP which is independent of i, j and QP in the range [0,P-1], where Bm(QP)[i][j] and Be(QP)[i][j] satisfy Bm(QP)[i][j]=Bm(QP mod P)[i][j] and Be(QP)[i][j]= Be(0) + QP/P where P is an integer and QP/P represents the integer obtained truncating towards zero the result of dividing QP by P; and b4) calculating the integer transform coefficient K[i][j] using K[i][j]=[L[i][j]*Bm(QP)[i][j]]

<< Be(QP)[i][j], where "<<" expresses a left-shift operation; c) performing inverse integer transformation of the integer transform coefficient K[i][j], where the basis functions of the inverse integer transformation have different norms obtaining an inverse transformed coefficient; and d) performing a further normalisation on the inverse transformed coefficient using a single scalar value 2^N obtaining the reconstructed sample.

According to the invention, there is provided a method for quantization which derives a quantized level (L) by quantizing a transform coefficient (K), comprising the steps of : inputting the transform coefficient; inputting a quantization parameter (QP); and deriving the quantized level, wherein: the quantized level is derived, by using a mantissa portion being a function of the quantization parameter (Am(QP)) and an exponential portion being a function of the quantization parameter (Ae(QP)), as L = [K*Am(QP)] >> Ae(QP), where ">>" expresses a right-shift operation, and wherein the function structure of the mantissa portion is Am(QP) = Am(QP mod P).

According to the invention, there is provided a method for dequantization which derives a transform coefficient (K) by dequantizing a quantized level (L), comprising the steps of: inputting the quantized level; inputting a quantization parameter (QP); and deriving the transform coefficient, wherein: the transform coefficient is derived, by using a mantissa portion being a function of the quantization parameter (Bm(QP)) and an exponential portion being a function of the quantization parameter (Be(QP)), as K = [L*Bm(QP)] << Be(QP), where "<<" expresses a left-shift operation, and wherein the function structure of the mantissa portion is Bm(QP) = Bm(QP mod P).

According to an aspect of the present invention, there is provided a video decoder for reconstruction of a sample from a quantized level, including a first means for deriving a transform coefficient; a second means for performing inverse transformation; and a third means for performing normalization, wherein the first means derives the transform coefficient by using a mantissa portion and an exponential portion, the second means derives a scaled sample by inverse transforming the transform coefficient, and the third means derives a reconstructed sample by normalizing the scaled sample using a constant normalization factor.

According to another aspect of the present invention, there is provided a video decoder for reconstruction of a sample from a quantized level, including a first means for deriving a transform coefficient; and a second means for performing inverse transformation, wherein the first means derives the transform coefficient by using a mantissa portion and an exponential portion, the second means derives a scaled sample by inverse transforming the transform coefficient, and the mantissa portion is a matrix with a raw represented by a quantization parameter and a column represented by a value related to the basis of the transformation.

According to a further aspect of the present invention, there is provided a video decoder for reconstruction of a sample from a quantized level L, including a first means for deriving a transform coefficient K; a second means for performing inverse transformation; and a third means for performing normalization, wherein the first means derives the transform coefficient, by using a mantissa portion Bm(QP) being a function of a quantization parameter QP and an exponential portion Be(QP) being a function of the quantization parameter, as K = [L*Bm(QP)] << Be(QP), where "<<" expresses a left-shift operation, the functions Bm(QP) and Be(QP) satisfy, Bm(QP) = Bm(QP mode P), and Be(QP) = B(0) + QP/P, where P in an integer, the second means derives a scaled sample by inverse transforming the transform coefficient, and the third means derives a reconstructed sample by normalizing said scaled sample using a constant normalization factor.

According to yet a further aspect of the present invention, there is provided A video decoder for reconstruction of a sample from a quantized level L, including a first means for deriving a transform coefficient K; and a second means for performing inverse transformation, wherein the first means derives said transform coefficient by using a mantissa portion and an exponential portion, the second means derives a scaled sample by inverse transforming said transform coefficient, and the mantissa portion is a matrix with a matrix element B(QP)[i][j] being a function of a quantization parameter QP, where i and j are horizontal basis or vertical basis of the transform coefficient.

In some aspects of the method, forming a quantization value (L) from the coefficient K includes:

$$L = K*A(QP)$$
$$= K*Am(QP)*(2^{Ae(Q\ P)}).$$

In other aspects, the method further comprises: normalizing the quantization value by 2^N as follows:

$$Ln = L/2^N$$
$$= K*Am(QP)/2^{(N-Ae(QP))}.$$

In some aspects, forming a quantization value includes forming a set of recursive quantization factors with a period P, where A(QP+P) = A(QP)/x. Therefore, forming a set of recursive quantization factors includes forming recursive mantissa factors, where Am(QP) = Am(QP mod P). Likewise, forming a set of recursive quantization factors includes forming recursive exponential factors, where Ae(QP) = Ae(QP mod P) - QP/P.

More specifically, supplying a coefficient K includes supplying a coefficient matrix K[i][j]. Then, forming a quantization value (L) from the coefficient matrix K[i][j] includes forming a quantization value matrix (L[i][j]) using a mantissa portion matrix (Am(QP)[i][j]) and an exponential portion matrix ($x^{Ae(QP)[i][j]}$).

Likewise, forming a quantization value matrix (L[i][j]) using a mantissa portion matrix (Am(QP)[i][j])) and an exponential portion matrix ($x^{Ae(QP)[i][j]}$) includes, for each particular value of QP, every element in the exponential portion matrix being the same value. Every element in the exponential portion matrix is the same value for a period (P) of QP values, where Ae(QP) = Ae(P *(QP/P)).

Additional details of the above-described method, including a method for forming a dequantization value (X1), from the quantization value, using a mantissa portion (Bm(QP)) and an exponential portion ($x^{Be(QP)}$), are provided below.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a flowchart illustrating the present invention method for the quantization of a coefficient.

Figs. 2 to 9 show embodiments of the present invention comprise systems and methods for video coding.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The dynamic range requirements of the combined transform and quantization is reduced by factoring the quantization parameters A(QP) and B(QP) into a mantissa and exponent terms as shown in Equation 4. With this structure, only the precision due to the mantissa term needs to be preserved during calculation. The exponent term can be included in the final normalization shift. This is illustrated in the sample calculation Equation 5.

Equation 4 Structure of quantization parameters

$$A_{proposed}(QP)=A_{mantissa}(QP)\cdot 2^{Aexponent(QP)}$$

$$B_{proposed}(QP)=B_{mantissa}(QP)\cdot 2^{Aexponent(QP)}$$

Equation 5 reduced bit depth LUMA Transform

$$Y = T\cdot X$$

$$K = Y\cdot T^{T}$$

$$L = (A_{mantissa}(QP)\cdot K)2^{20-Aexponent(QP)}$$

$$K' = T^{T}\cdot L$$

$$Y' = K'\cdot T$$

$$X' = (B_{mantissa}(QP)\cdot Y')/2^{20-Bexponent(QP)}$$

To illustrate the present invention, a set of quantization parameters is presented that reduce the dynamic range requirement of an H.26L decoder to 16-bit memory access. The memory access of the inverse transform is reduced to 16 bits. Values for $A_{mantissa}$, $A_{exponent}$, $B_{mantissa}$, $B_{exponent}$, $A_{proposed}$, $B_{proposed}$ are defined for QP=0-5 as shown in Table 3. Additional values are determined by recursion, as shown in Equation 6. The structure of these values makes it possible to generate new quantization values in addition to those specified.

Table 3 Quantization values 0-5 for TML

QP	$A_{mantissa}$	$A_{exponent}$	$B_{mantissa}$	$B_{exponent}$	$A_{proposed}$	$B_{proposed}$
0	5	7	235	4	640	3760
1	9	6	261	4	576	4176
2	127	2	37	7	508	4736
3	114	2	165	5	456	5280
4	25	4	47	7	400	6016
5	87	2	27	8	348	6912

Equation 6 Recursion relations

$$A_{mantissa}(QP + 6) = A_{mantissa}(QP)$$

$$B_{mantissa}(QP + 6) = B_{mantissa}(QP)$$

$$A_{mantissa}(QP + 6) = A_{mantissa}(QP)-1$$

$$B_{mantissa}(QP + 6) = B_{mantissa}(QP)+1$$

Using the defined parameters, the transform calculations can be modified to reduce the dynamic range as shown in Equation 5. Note how only the mantissa values contribute to the growth of dynamic range. The exponent factors are incorporated into the final normalization and do not impact the dynamic range of intermediate results.

With these values and computational method, the dynamic range at the decoder is reduced so only 16-bit memory access is needed as seen in Table 4.

Table 4 Dynamic range with low-bit depth quantization (QP>6)

8-bit	LUMA Transform	Inverse Transform
Register	28	24
Memory	21	16

Several refinements can be applied to the joint quantization/normalization procedure described above. The general technique of factoring the parameters into a mantissa and exponent forms the basis of these refinements.

The discussion above assumes all basis functions of the transform have an equal norm and are quantized identically. Some integer transforms have the property that different basis functions have different norms. The present invention technique has been generalized to support transforms having different norms by replacing the scalars A(QP) and B(QP) above by matrices A(QP)[i][j] and B(QP)[i][j]. These parameters are linked by a normalization relation of the form shown below, Equation 7, which is more general than the single relation shown in Equation 1.

Equation 7 Joint quantization/normalization of matrices

$$A(QP)[i][j] \cdot B(QP)[i][j] = N[i][j]$$

Following the method previously described, each element of each matrix is factored into a mantissa and an exponent term as illustrated in the equations below, Equation 8.

Equation 8 Factorization of matrix parameters

$$A(QP)[i][j] = A_{mantissa}(QP)[i][j] \cdot 2^{Aexponent(QP)[i][j]}$$

$$B(QP)[i][j] = B_{mantissa}(QP)[i][j] \cdot 2^{Bexponent(QP)[i][j]}$$

A large number of parameters are required to describe these quantization and dequantization parameters. Several structural relations can be used to reduce the number of free parameters. The quantizer growth is designed so that the values of A are halved after each period P at the same time the values of B are doubled maintaining the normalization relation. Additionally, the values of $A_{exponent}(QP)[i][j]$ and $B_{exponent}(QP)[i][j]$ are independent of i, j and (QP) in the range [0,P-1]. This structure is summarized by structural equations, Equation 9. With this structure there are only two parameters $A_{exponent}[0]$ and $B_{exponent}[0]$.

Equation 9 Structure of exponent terms

$$A_{exponent}(QP)[i][j] = A_{exponent}[0]\text{-}QP/P$$

$$B_{exponent}(QP)[i][j] = B_{exponent}[0]\text{-}QP/P$$

A structure is also defined for the mantissa values. For each index pair (i,j), the mantissa values are periodic with period P. This is summarized by the structural equation, Equation 10. With this structure, there are P independent matrices for $A_{mantissa}$ and P independent matrices for $B_{mantissa}$ reducing memory requirements and adding structure to the calculations.

Equation 10 Structure of mantissa terms

$$A_{mantissa}(QP)[i][j] = A_{mantissa}(QP\%P)[i][j]$$

$$B_{mantissa}(QP)[i][j] = B_{mantissa}(QP\%P)[i][j]$$

The inverse transform may include integer division that requires rounding. In cases of interest the division is by a power of 2. The rounding error is reduced by designing the dequantization factors to be multiples of the same power of 2, giving no remainder following division.

Dequantization using the mantissa values $B_{mantissa}(QP)$ gives dequantized values that are normalized differently depending upon qp. This must be compensated for following the inverse transform. A form of this calculation is shown in Equation 11.

Equation 11 Normalization of inverse transform I

$$K[i][j] = B_{mantissa}(QP\%P)[i][j] \cdot Level[i][j]$$

$$X = (T^{-1} \cdot K \cdot T)/2^{(N-QP/P)}$$

In Equation 11, Level[i][j] is the quantized version of the transform coefficients and is called as "quantization value". K[i][j] is the scaled version of the transform coefficients and is called as "dequantization value".

To eliminate the need for the inverse transform to compensate for this normalization difference, the dequantization operation is defined so that all dequantized values have the same normalization. The form of this calculation is shown in Equation 12.

Equation 12 Normalization of inverse transform II

$$K[i][j] = B_{mantissa}\,(QP\%P)[i][j] \cdot 2^{QP/P} \cdot Level[i][j]$$

$$X = (T^{-1} \cdot K \cdot T)/2^{N}$$

The power of 2 can be calculated by using left-shift operation and the dequantization value K[i][j] in Equation 12 will then be given as follows.

$$K[i][j] = [B_{mantissa} \cdot Level[i][j]] << (QP/P)$$

An example follows that illustrates the present invention use of quantization matrices. The forward and inverse transforms defined in Equation 13 need a quantization matrix rather than a single scalar quantization value. Sample quantization and dequantization parameters are given. Equation 14 and 16, together with related calculations, illustrate the use of this invention. This example uses a period P=6. In Equation 14, $A_{mantissa}$ is represented by Q and QP is represented by m. In Equation 16, $B_{mantissa}$ is represented by R and QP is represented by m.

Equation 13 transforms

$$T_{forward} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{pmatrix}$$

$$T_{reverse} = \begin{pmatrix} 2 & 2 & 2 & 1 \\ 2 & 1 & -2 & -2 \\ 2 & -2 & -2 & 2 \\ 2 & -1 & 2 & -1 \end{pmatrix}$$

Equation 14 quantization parameters

$$Q(m)[i][j] = M_{m.0} \text{ for } (i, j) = \{(0,0), (0,2), (2,0), (2,2)\}$$
$$Q(m)[i)[j] = M_{m.1} \text{ for } (i, j) = \{(1,1), (1,3), (3,1), (3,3)\}$$
$$Q(m)[i][j] = M_{m.2} \text{ otherwise}$$

$$M = \begin{bmatrix} 21844 & 8388 & 13108 \\ 18724 & 7625 & 11650 \\ 16384 & 6989 & 10486 \\ 14564 & 5992 & 9532 \\ 13107 & 5243 & 8066 \\ 11916 & 4660 & 7490 \end{bmatrix}$$

Equation 16 Dequantization parameters

$R(m)[i][j] = S_{m.0}$ for (i, j) = {(0,0), (0,2), (2,0), (2,2)}

$R(m)[i][j] = S_{m.1}$ for (i, j) = {(1,1), (1,3), (3,1), (3,3)}

$R(m)[i][j] = S_{m.2}$ otherwise

$$S = \begin{bmatrix} 6 & 10 & 8 \\ 7 & 11 & 9 \\ 8 & 12 & 10 \\ 9 & 14 & 11 \\ 10 & 16 & 13 \\ 11 & 18 & 14 \end{bmatrix}$$

The description of the forward transformation and forward quantization, Equation 18, are given below assuming input is in X, quantization parameter QP.

Equation 17 forward transform

$$K = T_{forward} \cdot X \cdot T^{T}_{forward}$$

Equation 18 forward quantization

$period = QP / 6$

$phase = QP - 6 \cdot period$

$Level[i][j] = (Q(phase)[i][j] \cdot K[i][j])/2^{(17+period)}$

The description of dequantization, inverse transform, and normalization for this example is given below, Equation 19 and 20.

Equation 19 Dequantization

$period = QP / 6$

$phase = QP - 6 \cdot period$

$K[i][j] = R(phase)[i][j] \cdot Level[i][j] \cdot 2^{period}$

Equation 20 IDCT and normalization

$$X' = T_{reverse} \cdot K \cdot T^{T}_{reverse}$$

$$X''[i][j] = X'[i][j]/2^{7}$$

Fig. 1 is a flowchart illustrating the present invention method for the quantization of a coefficient. Although this method is depicted as a sequence of numbered steps for clarity, no order should be inferred from the numbering unless explicitly stated. It should be understood that some of these steps may be skipped, performed in parallel, or performed without the requirement of maintaining a strict order of sequence. The methods start at Step 100. Step 102 supplies a coefficient K. Step 104 supplies a quantization parameter (QP). Step 106 forms a quantization value (L) from the coefficient K using a mantissa portion (Am(QP)) and an exponential portion ($x^{Ae(QP)}$). Typically, the exponential portion ($x^{Ae(QP)}$) includes x being the value 2.

In some aspects of the method, forming a quantization value (L) from the coefficient K using a mantissa portion (Am(QP)) and an exponential portion ($x^{Ae(QP)}$) in Step 106 includes:

$$L = K*A(QP)$$

$$= K*Am(QP)* (2^{Ae(QP)}).$$

Some aspects of the method include a further step. Step 108 normalizes the quantization value by 2^{N} as follows:

$$Ln = L/2^{N}$$

$$= K*Am(QP)/2^{(N-Ae(QP))}.$$

In other aspects, forming a quantization value in Step 106 includes forming a set of recursive quantization factors with a period P, where A(QP+P) = A(QP)/x. Likewise, forming a set of recursive quantization factors includes forming recursive mantissa factors, where Am(QP) = Am(QP mod P). Then, forming a set of recursive quantization factors includes forming recursive exponential factors, where Ae(QP) = Ae(QP mod P) - QP/P.

In some aspects, forming a quantization value includes forming a set of recursive quantization factors with a period P, where A(QP+P) = A(QP)/2. In other aspects, forming a set of recursive quantization factors includes forming recursive mantissa factors, where P = 6. Likewise, forming a set of recursive quantization factors includes forming recursive exponential factors, where P = 6.

In some aspects of the method, supplying a coefficient K in Step 102 includes supplying a coefficient matrix K[i][j]. Then, forming a quantization value (L) from the coefficient matrix K[i][j] using a mantissa portion (Am(QP)) and an exponential portion ($x^{Ae(QP)}$) in Step 106 includes forming a quantization value matrix (L[i][j]) using a mantissa portion matrix (Am(QP)[i][j]) and an exponential portion matrix ($x^{Ae(QP)[i][j]}$). Likewise, forming a quantization value matrix (L[i][j]) using a mantissa portion matrix (Am(QP)[i][j]) and an exponential portion matrix ($x^{Ae(QP)[i][j]}$) includes, for each particular value of QP, every element in the exponential portion matrix being the same value. Typically, every element in the exponential portion matrix is the same value for a period (P) of QP values, where Ae(QP) = Ae(P*(QP/P)).

Some aspects of the method include a further step. Step 110 forms a dequantization value (X1) from the quantization value, using a mantissa portion (Bm(QP)) and an exponential portion ($x^{Be(QP)}$). Again, the exponential portion ($x^{Be(QP)}$) typically includes x being the value 2.

In some aspects of the method, forming a dequantization value (XI) from the quantization value, using a mantissa portion (Bm(QP)) and an exponential portion ($2^{Be(QP)}$) includes:

$$X1 = L*B(QP)$$
$$= L*Bm(QP)*(2^{Be(QP)}).$$

Other aspects of the method include a further step, Step 112, of denormalizing the quantization value by 2^N as follows:

$$X1d = X1/2^N$$
$$= X1*Bm(QP)/2^N.$$

In some aspects, forming a dequantization value in Step 110 includes forming a set of recursive dequantization factors with a period P, where B(QP+P) = x*B(QP). Then, forming a set of recursive dequantization factors includes forming recursive mantissa factors, where Bm(QP) = Bm(QP mod P). Further, forming a set of recursive dequantization factors includes forming recursive exponential factors, where Be(QP) = Be(QP mod P) + QP/P.

In some aspects, forming a set of recursive quantization factors with a period P includes the value of x being equal to 2, and forming recursive mantissa factors includes the value of P being equal to 6. Then, forming a set of recursive dequantization factors includes forming recursive exponential factors, where Be(QP) = Be(QP mod P) + QP/P.

In some aspects of the method, forming a dequantization value (X1), from the quantization value, using a mantissa portion (Bm(QP)) and an exponential portion ($x^{Be(QP)}$) in Step 110 includes forming a dequantization value matrix (X1[i][j]) using a mantissa portion matrix (Bm(QP)[i][j]) and an exponential portion matrix ($x^{Be(QP)[i][j]}$). Likewise, forming a dequantization value matrix (X1[j][j]) using a mantissa portion matrix (Bm(QP)[i][j]) and an exponential portion matrix ($x^{Be(QP)[i][j]}$) includes, for each particular value of QP, every element in the exponential portion matrix being the same value. In some aspects, every element in the exponential portion matrix is the same value for a period (P) of QP values, where Be(QP) = Be(P*(QP/P)).

Another aspect of the invention includes a method for the dequantization of a coefficient. However, the process is essentially the same as Steps 110 and 112 above, and is not repeated in the interest of brevity.

A method for the quantization of a coefficient has been presented. An example is given illustrating a combined dequantization and normalization procedure applied to the H.26L video coding standard with a goal of reducing the bit-depth required at the decoder to 16 bits. The present invention concepts can also be used to meet other design goals within H.26L. In general, this invention has application to the combination of normalization and quantization calculations.

Embodiments of the present invention may be implemented as hardware, firmware, software and other implementations. Some embodiments may be implemented on general purpose computing devices or on computing devices

specifically designed for implementation of these embodiments. Some embodiments may be stored in memory as a means of storing the embodiment or for the purpose of executing the embodiment on a computing device.

Some embodiments of the present invention comprise systems and methods for video encoding, as shown in Figure 2. In these embodiments, image data 130 is subtracted from 132 with data representing prior video frames 145 resulting in a differential image 133, which is sent to a transform module 134. Transform module 134 may use DCT or other transform methods to transform the image. Generally, the result of the transform process will be coefficients (K), which are then sent to a quantization module 136 for quantization.

Quantization module 136 may have other inputs, such as user inputs 131 for establishing quantization parameters (QPs) and for other input. Quantization module 136 may use the transformation coefficients and the quantization parameters to determine quantization levels (L) in the video image. Quantization module 136 may use methods employing a mantissa portion and an exponential portion, however, other quantization methods may also be employed in the quantization modules 136 of embodiments of the present invention. These quantization levels 135 and quantization parameters 137 are output to a coding module 138 as well as a dequantization module (DQ) 140.

Output to the coding module 138 is encoded and transmitted outside the encoder for immediate decoding or storage. Coding module 138 may use variable length coding (VLC) in its coding processes. Coding module 138 may use arithmetic coding in its coding process. Output from coding module 138 is encoded data 139 which may be transmitted to the decoder or stored in the storage device.

Output from quantization module 136 is also received at dequantization module 140 to begin reconstruction of the image. This is done to keep an accurate accounting of prior frames. Dequantization module 140 performs a process with essentially the reverse effect as quantization module 136. Quantization levels or values (L) are dequantized yielding transform coefficients. Dequantization modules 140 may use methods employing a mantissa portion and an exponential portion as described herein.

The transform coefficients output from dequantization module 140 are sent to an inverse transformation (IT) module 142 where they are inverse transformed to a differential image 141. This differential image 141 is then combined with data from prior image frames 145 to form a video frame 149 that may be input to a frame memory 146 for reference to succeeding frames.

Video frame 149 may also serve as input to a motion estimation module 147, which also receives image data 130. These inputs may be used to predict image similarities and help compress image data. Output from motion estimation module 147 is sent to motion compensation module 148 and combined with output data from coding module 138, which is sent out for later decoding and eventual image viewing.

Motion compensation module 148 uses the predicted image data to reduce frame data requirements; its output is subtracted from input image data 130.

Some embodiments of the present invention comprise systems and methods for video decoding, as shown in Figure 3. A decoder of embodiments of the present invention may receive encoded data 150 to a decoder module 152. Encoded data 150 may comprise data that has been encoded by an encoder 100 such as that described with reference to Figure 2.

Decoding module 152 may employ variable length decoding methods if they were used in the encoding process. Other decoding methods may also be used as dictated by the type of encoded data 150. Decoding module 152 performs essentially the reverse process as coding module 138. Output from decoding module 152 may comprise quantization parameters 156 and quantization values 154. Other output may comprise motion estimation data and image prediction data that may be sent directly to a motion compensation module 166.

Typically, quantization parameters 156 and quantization values 154 are output to a dequantization module 158, where quantization values are converted back to transform coefficients. Dequantization module 158 may use methods employing a mantissa portion and an exponential portion as described herein. These coefficients are then sent to an inverse transformation module 160 for conversion back to spatial domain image data 161.

The motion compensation unit 166 uses motion vector data and the frame memory 164 to construct a reference image 165.

Image data 161 represents a differential image that must be combined with prior image data 165 to form a video frame 163. This video frame 163 is output 168 for further processing, display or other purposes and may be stored in frame memory 164 and used for reference with subsequent frames.

In some embodiments of the present invention, as illustrated in Figure 4, image data 102 may be sent to an encoder or encoding portion 104 for the various transformation, quantization, encoding and other procedures typical of video encoding as described above for some embodiments of the present invention. Output from the encoder may then be stored on any computer-readable storage media 106. Storage media 106 may act as a short-term buffer or as a long-term storage device.

When desired, encoded video data may be read from storage media 106 and decoded by a decoder or decoding portion 108 for output 110 to a display or other device.

In some embodiments of the present invention, as illustrated in Figure 5, image data 112 may be sent to an encoder or encoding portion 114 for the various transformation, quantization, encoding and other procedures typical of video encoding as described above for some embodiments of the present invention. Output from the encoder may then be sent over a network, such as a LAN, WAN or the Internet 116. A storage device such as storage media 106 may be part of a network. Encoded video data may be received and decoded by a decoder or decoding portion 118 which also communicates with network 116. Decoder 118 may then decode the data for local consumption 120.

In some embodiments of the present invention, as illustrated in Figure 6, a quantization method or apparatus comprises a mantissa portion 172 and an exponential portion 174. Quantization parameters 176 are input to both portions 172 & 174. A coefficient K 170 is input to the mantissa portion 172 where it is modified using the quantization parameter and other values as explained above. The result of this operation is combined with the result produced in the exponential portion using the quantization parameter thereby producing a quantization level or value L 178.

In some embodiments of the present invention, as illustrated in Figure 7, a quantization method or apparatus comprises a mantissa portion 182 and a shifting portion 184. Quantization parameters 186 are input to both portions 182 & 184. A

coefficient, K 180 is input to the mantissa portion 182 where it is modified using the quantization parameter and other values as explained above. The result of this operation is further processed in the shifting portion using the quantization parameter thereby producing a quantization level or value, L 188.

Some embodiments of the present invention, as illustrated in Figure 8, comprise a dequantization method or apparatus with a mantissa portion 192 and an exponential portion 194. Quantization parameters 196 are input to both portions 192 & 194. A quantization value, L 190 is input to the mantissa portion 192 where it is modified using the quantization parameter and other values as explained above. The result of this operation is further processed in the exponential portion using the quantization parameter thereby producing a coefficient, X1 198.

Some embodiments of the present invention, as illustrated in Figure 9, comprise a dequantization method or apparatus with a mantissa portion 202 and a shifting portion 204. Quantization parameters 206 are input to both portions 202 & 204. A quantization value, L 200 is input to the mantissa portion 202 where it is modified using the quantization parameter and other values as explained above. The result of this operation is further processed in the exponential portion using the quantization parameter thereby producing a coefficient, X1 208.

Some embodiments of the present invention may be stored on computer-readable media such as magnetic media, optical media, and other media as well as combinations of media. Some embodiments may also be transmitted as signals across networks and communication media. These transmissions and storage actions may take place as part of operation of embodiments of the present invention or as a way of transmitting the embodiment to a destination.

Other variations and embodiments of the invention will occur to those skilled in the art.

The embodiments of the present invention in which an exclusive property or privilege is claimed are defined as follows:

1. An image coding method for encoding an image data, comprising:

a transform step for integer transforming the image data for each of blocks into which the image is divided;

a quantization step for quantizing the integer transformed data to derive a quantized value L[i][j];

a dequantization step for dequantizing the quantized value L[i][j] to derive a transform coefficient K[i][j];

an inverse transformation step for inverse integer transforming the transform coefficient derived by the dequantization step; and

a normalization step for normalizing the inverse integer transformed transform coefficient by the inverse transformation step, wherein

the inverse integer transform is executed by an inverse transform matrix in which a different basis function has a different norm,

the dequantization step uses a mantissa portion matrix element Bm(QP)[i][j] being a function of quantization parameter QP and an exponential portion Be(QP) being a function of quantization parameter QP to derive said transform coefficient K[i][j] as follows:

K[i][j]=[L[i][j] x Bm(QP)[i][j]]<<Be(QP)

(where"<<" is a left shift operation),

the mantissa portion matrix element Bm(QP)[i][j] and the exponential portion Be(QP) are expressed by using constant values B and P respectively as follows:

Bm(QP)[i][j] = Bm(QP mod P)[i][j] and

Be(QP) = B + QP/P; and

the mantissa portion matrix element has a different value depending upon a basis and the normalization step normalizes the inverse integer transformed transform coefficient by a shift operation with use of 2^N (where N is a natural number).

2. An image coding apparatus for encoding an image data, comprising:

a transform means for integer transforming the image data for each of blocks into which the image is divided;

a quantization means for quantizing the integer transformed data to derive a quantized value L[i][j];

a dequantization means for dequantizing the quantized value L[i][j] to derive a transform coefficient K[i][j];

an inverse transformation means for inverse integer transforming the transform coefficient derived by the dequantization means; and

a normalization means for normalizing the inverse integer transformed transform coefficient by the inverse transformation means, wherein

the inverse integer transform is executed by an inverse transform matrix in which a different basis function has a different norm,

the dequantization means uses a mantissa portion matrix element Bm(QP)[i][j] being a function of quantization parameter QP and an exponential portion Be(QP) being a function of quantization parameter QP to derive said transform coefficient K[i][j] as follows:

$$K[i][j]=[L[i][j] \times Bm(QP)[i][j]]<<Be(QP)$$

(where "<<" is a left shift operation),

the mantissa portion matrix element Bm(QP)[i][j] and the exponential portion Be(QP) are expressed by using constant values B and P respectively as follows:

$$Bm(QP)[i][j] = Bm(QP \bmod P)[i][j] \text{ and}$$

$$Be(QP) = B + QP/P; \text{ and}$$

the mantissa portion matrix element has a different value depending upon a basis and the normalization means normalizes the inverse integer transformed transform coefficient by a shift operation with use of 2^N (where N is a natural number).

FIG.1

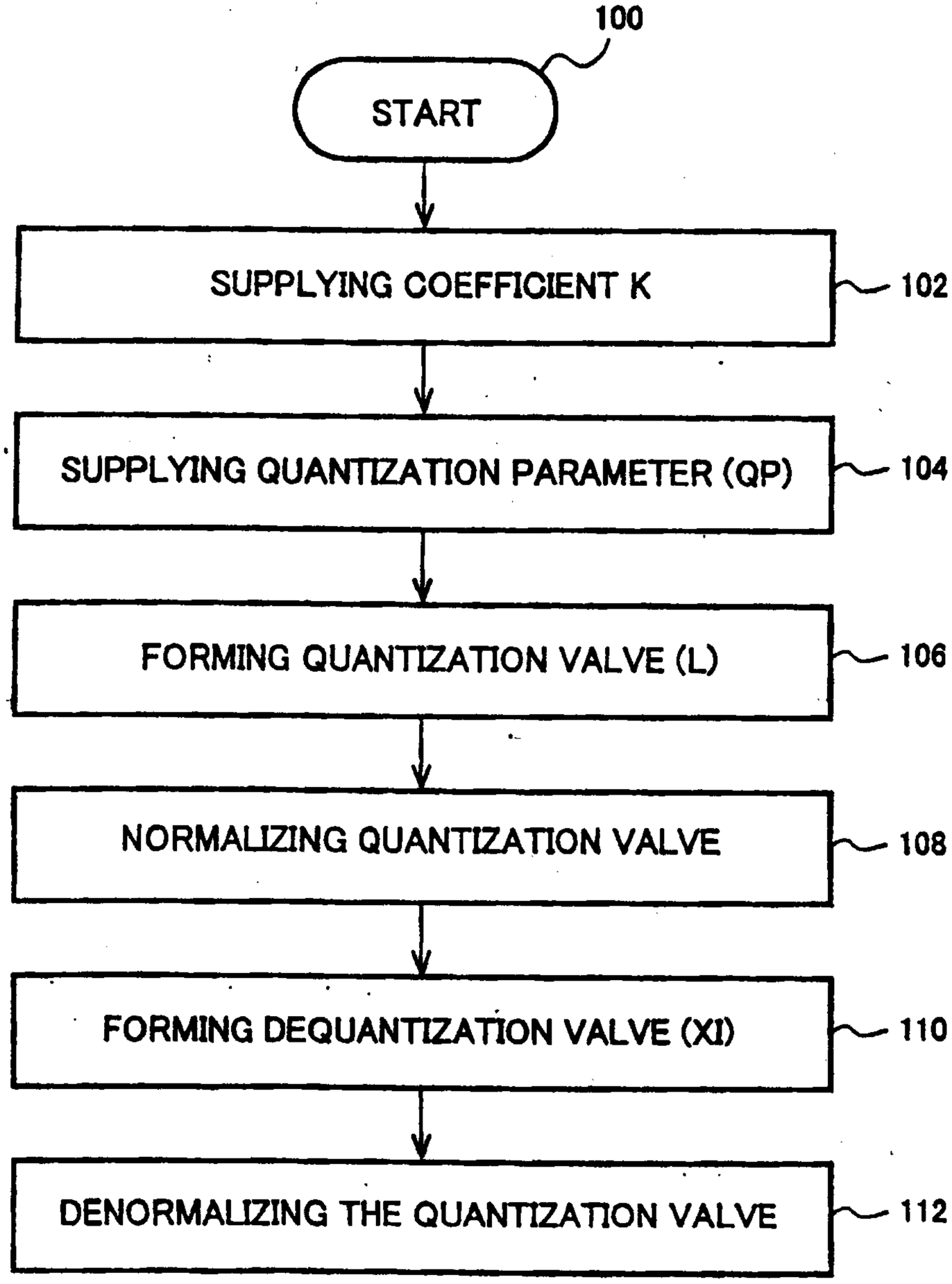

100
START
SUPPLYING COEFFICIENT K
102
SUPPLYING QUANTIZATION PARAMETER (QP)
104
FORMING QUANTIZATION VALVE (L)
106
NORMALIZING QUANTIZATION VALVE
108
FORMING DEQUANTIZATION VALVE (XI)
110
DENORMALIZING THE QUANTIZATION VALVE
112

FIG.2

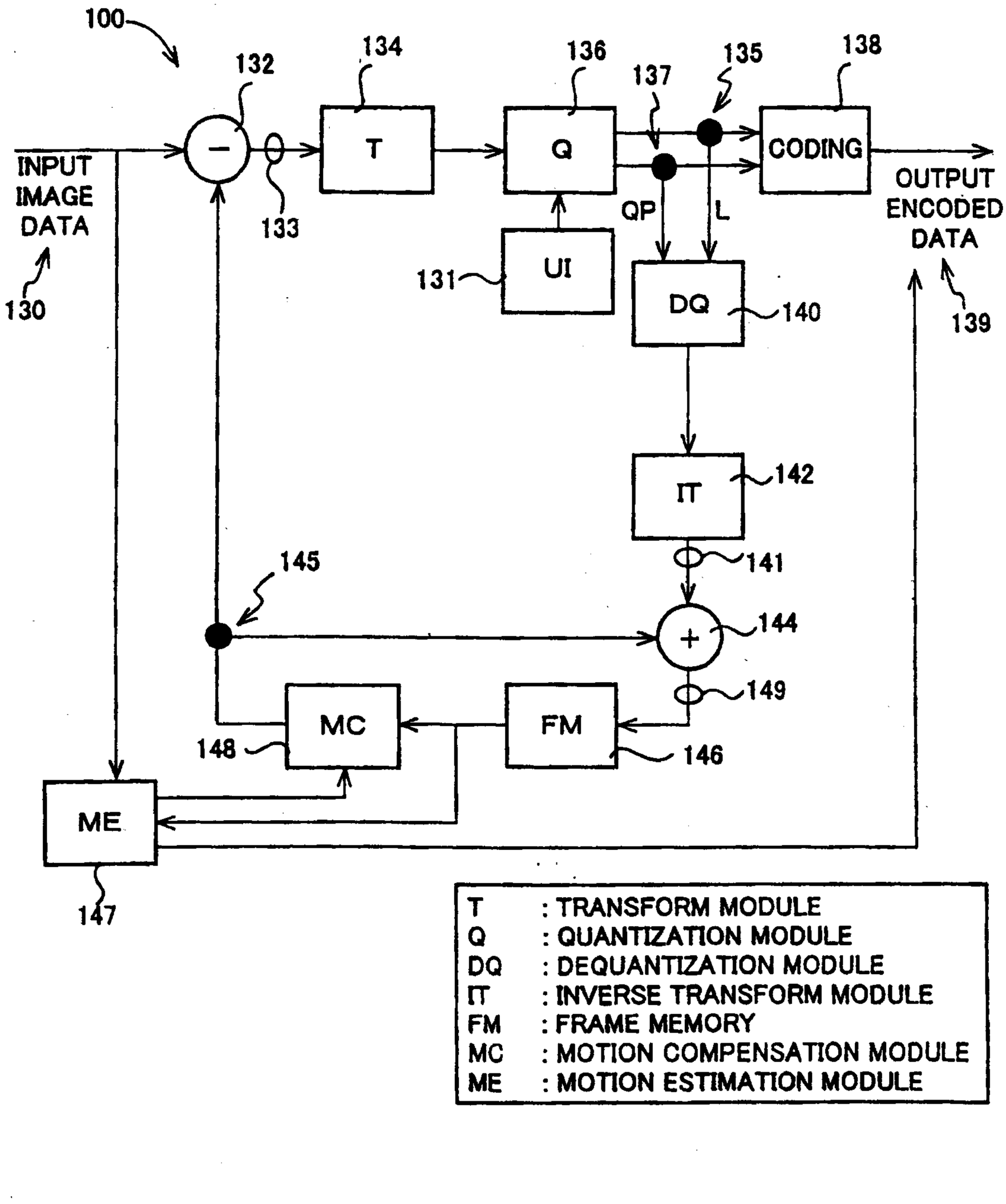
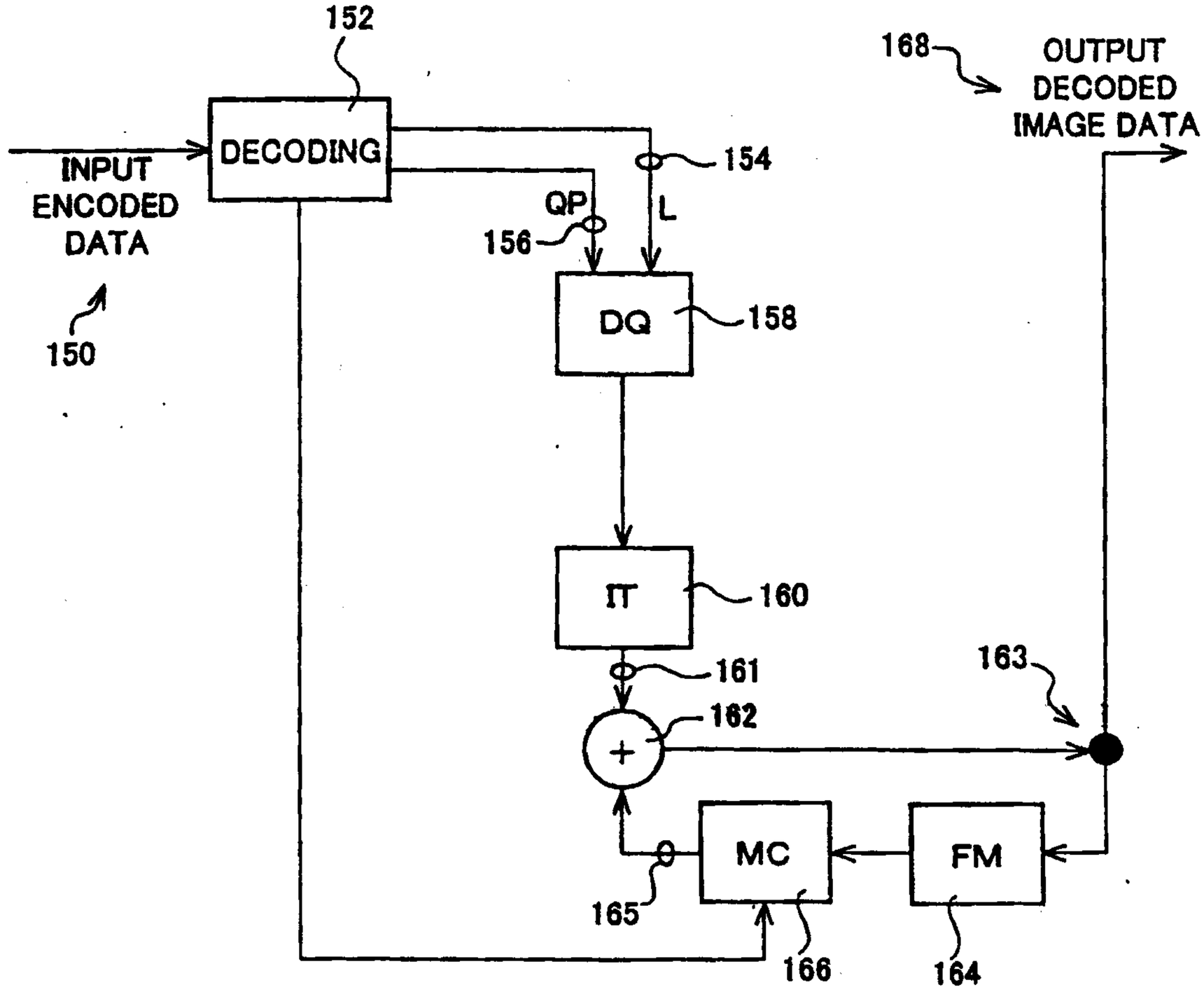

100
132
134
136
135
137
138
INPUT IMAGE DATA
130
133
T
Q
CODING
OUTPUT ENCODED DATA
139
131
UI
QP
L
DQ
140
142
IT
141
145
144
149
MC
148
FM
146
ME
147
T : TRANSFORM MODULE
Q : QUANTIZATION MODULE
DQ : DEQUANTIZATION MODULE
IT : INVERSE TRANSFORM MODULE
FM : FRAME MEMORY
MC : MOTION COMPENSATION MODULE
ME : MOTION ESTIMATION MODULE

FIG.3



DQ	: DEQUANTIZATION MODULE
IT	: INVERSE TRANSFORM MODULE
FM	: FRAME MEMORY
MC	: MOTION COMPENSATION MODULE

FIG.4

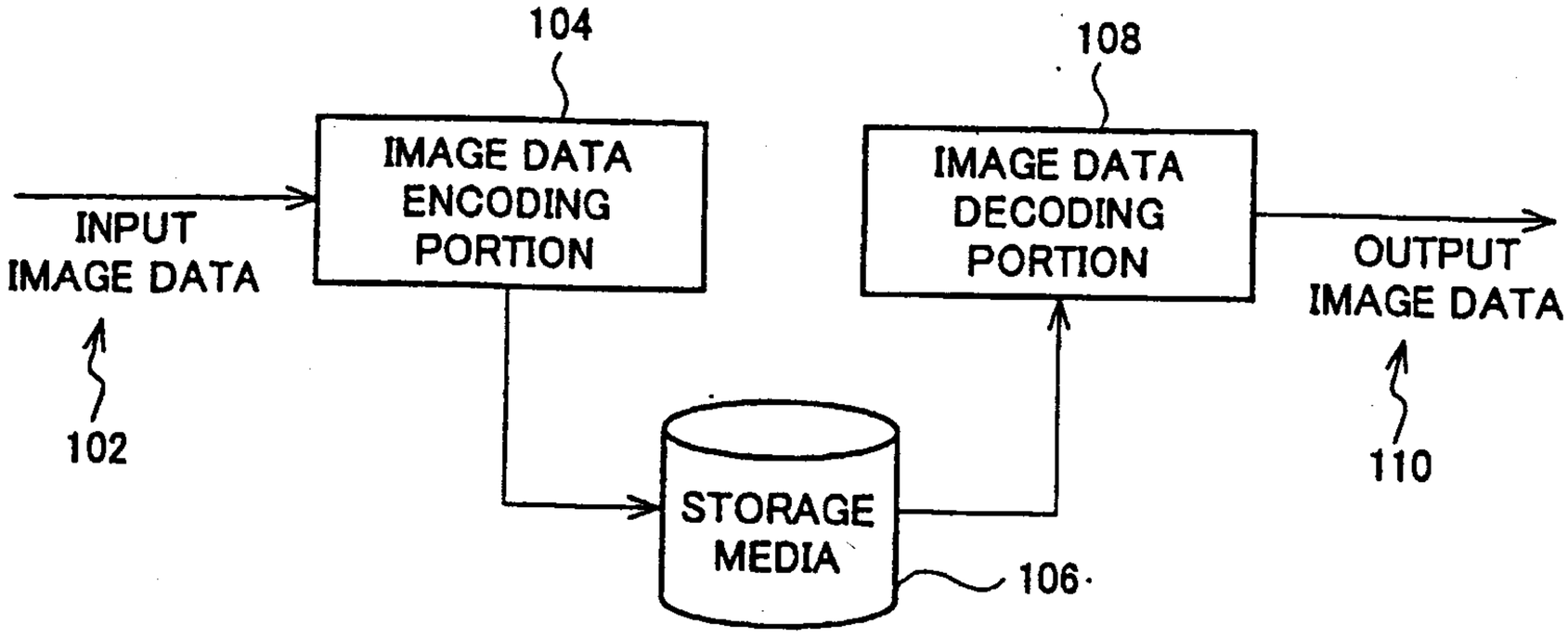

104
IMAGE DATA ENCODING PORTION
108
IMAGE DATA DECODING PORTION
INPUT IMAGE DATA
102
OUTPUT IMAGE DATA
110
STORAGE MEDIA
106

FIG.5

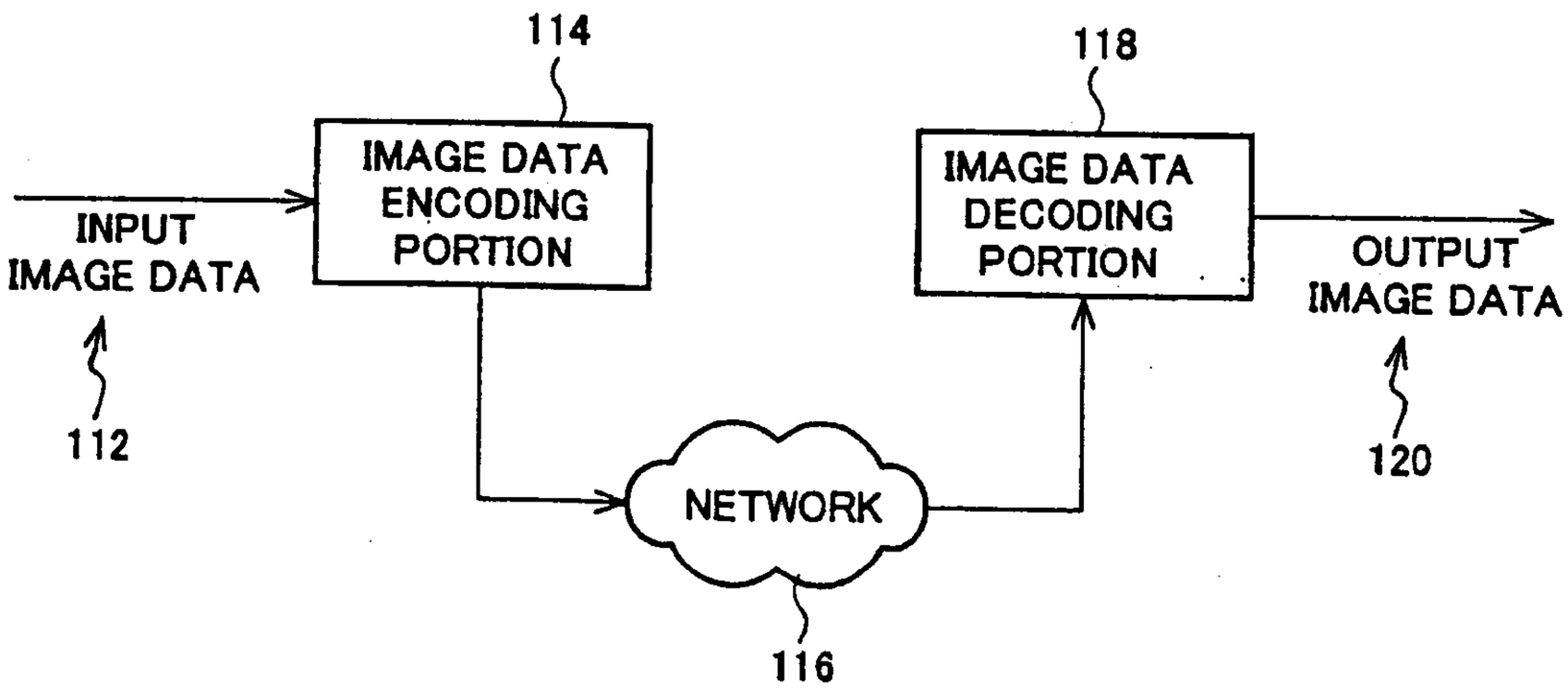

114
IMAGE DATA ENCODING PORTION
118
IMAGE DATA DECODING PORTION
INPUT IMAGE DATA
112
OUTPUT IMAGE DATA
120
NETWORK
116

FIG.6

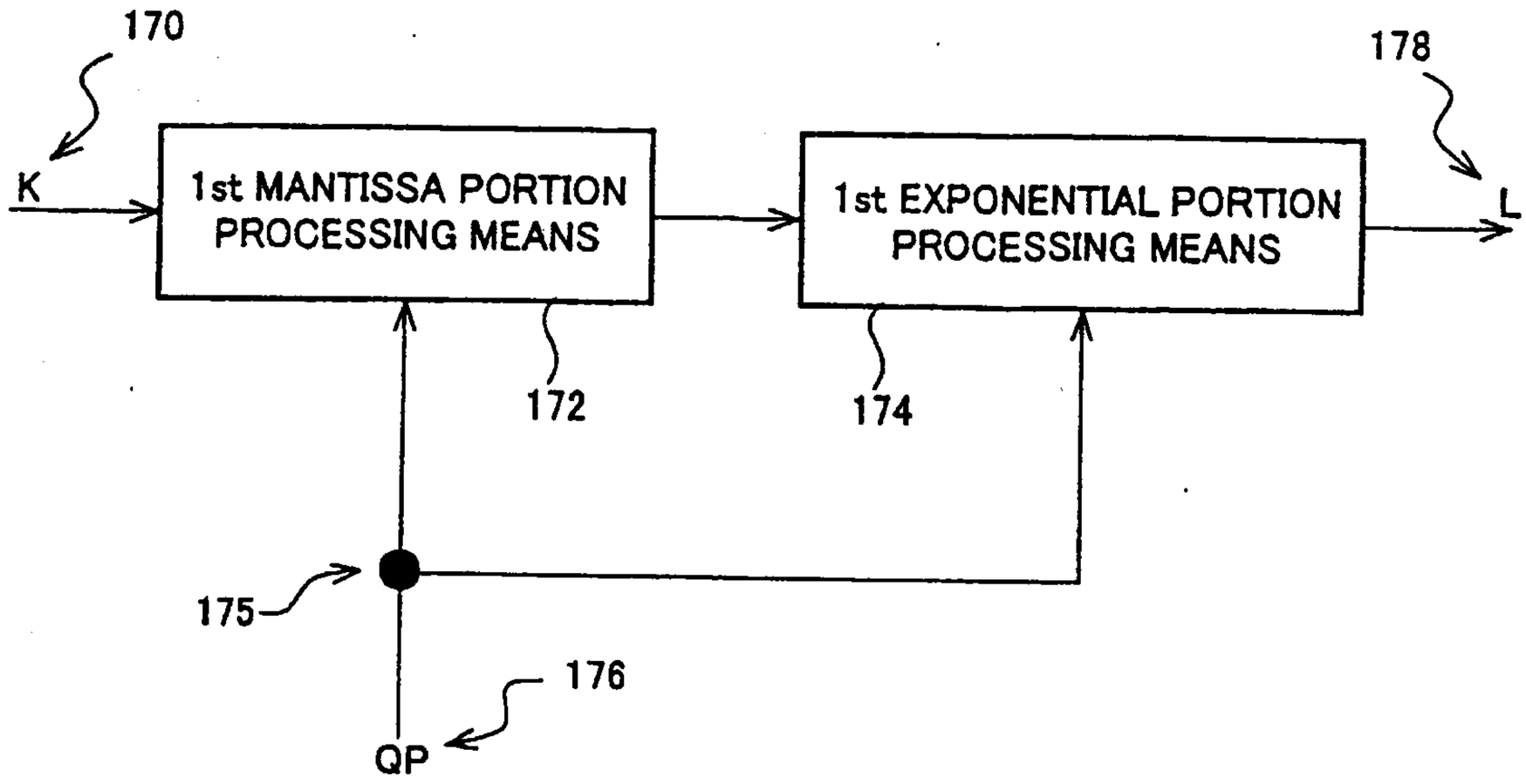

170
K
1st MANTISSA PORTION PROCESSING MEANS
172
1st EXPONENTIAL PORTION PROCESSING MEANS
174
178
L
175
176
QP

FIG.7

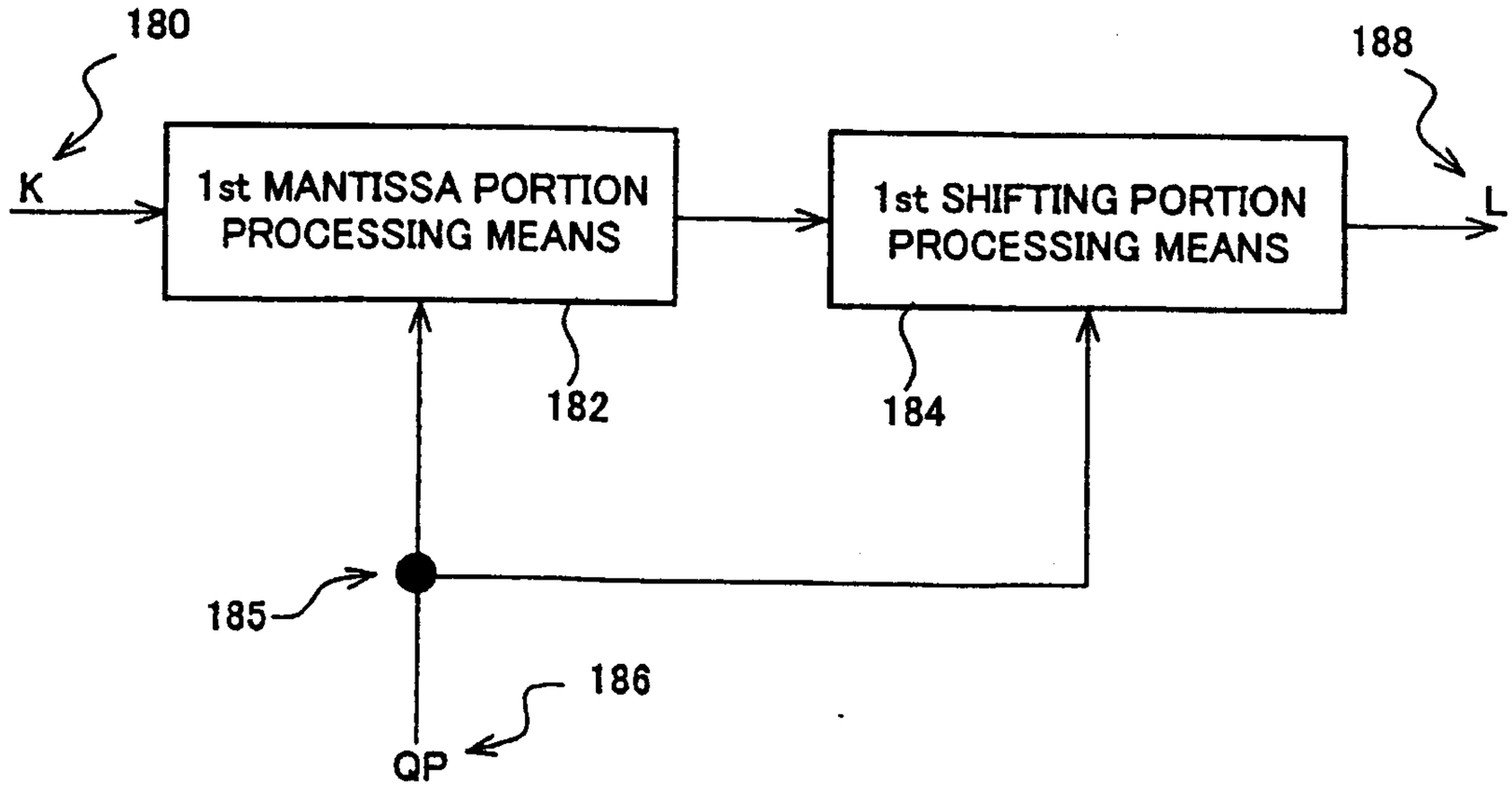

180
K
1st MANTISSA PORTION PROCESSING MEANS
182
1st SHIFTING PORTION PROCESSING MEANS
184
188
L
185
186
QP

FIG.8

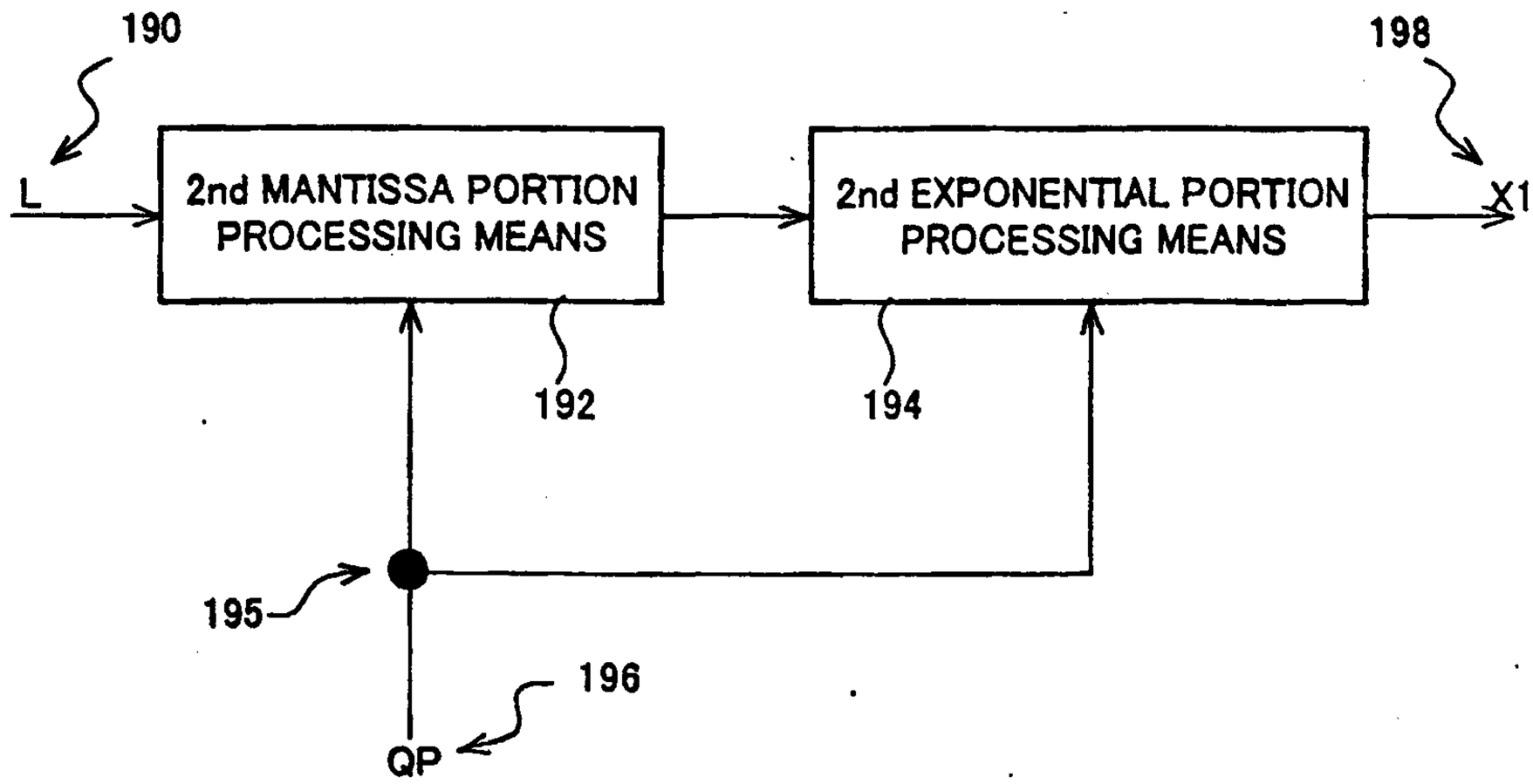
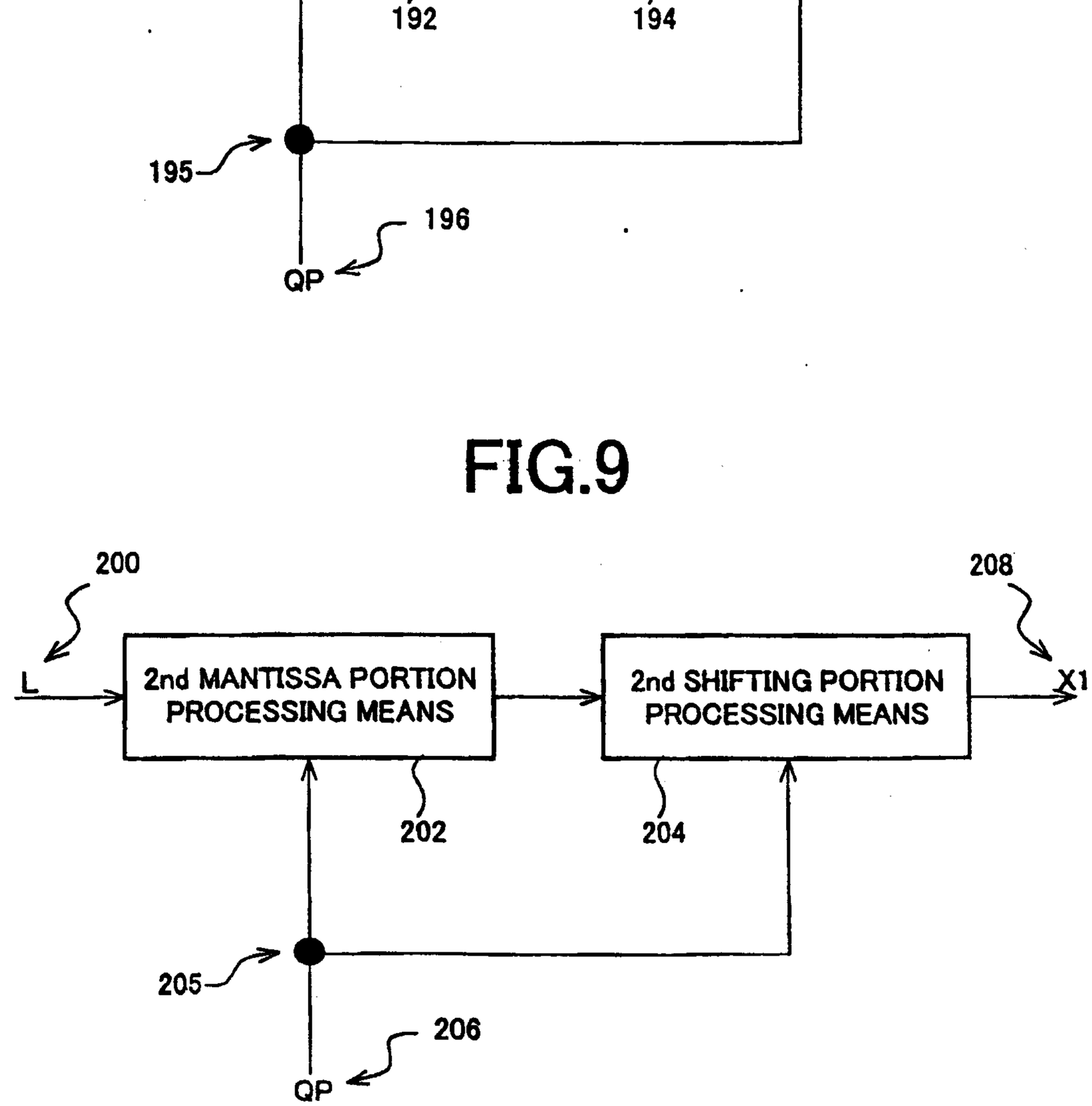
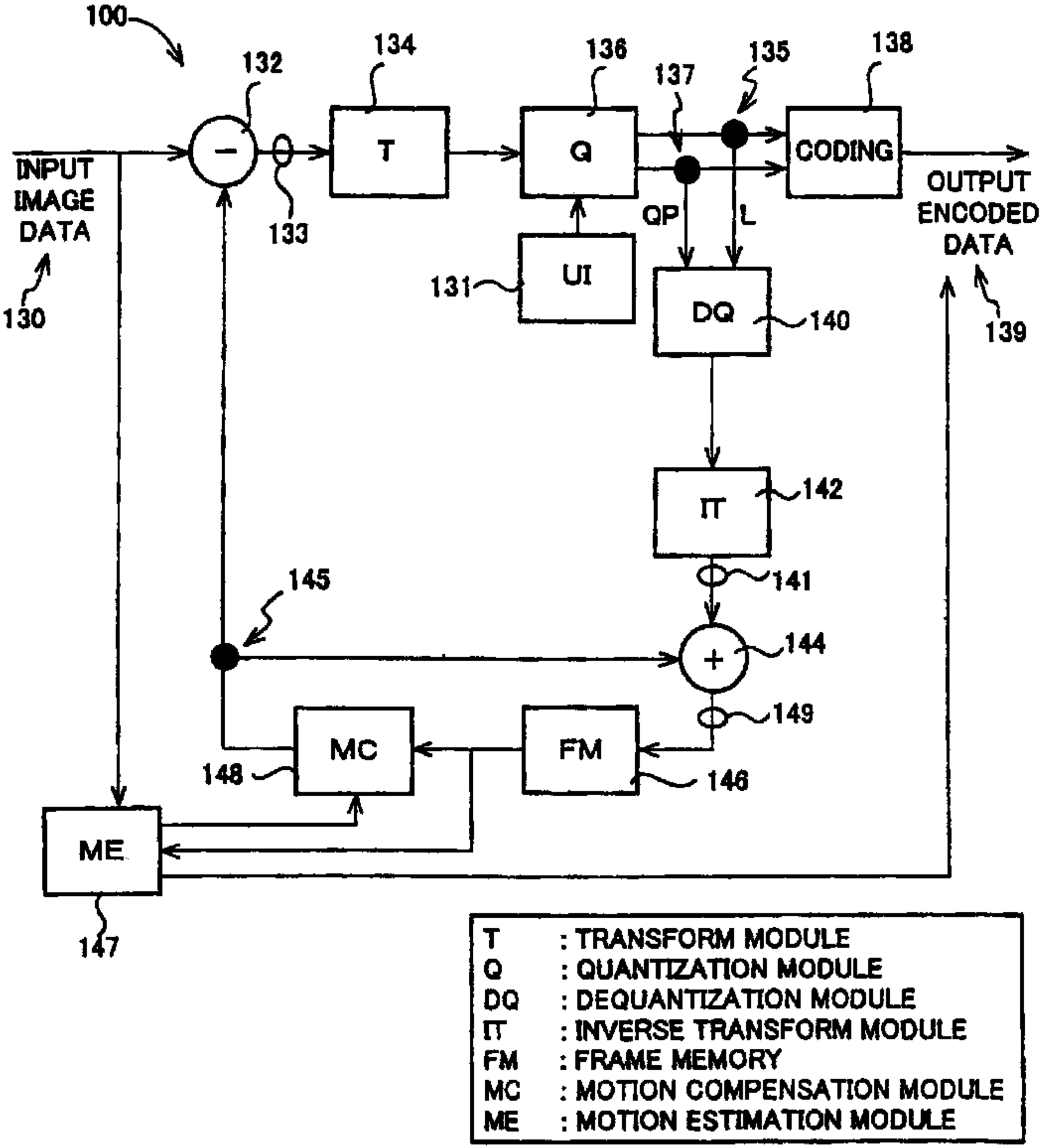

190
198
L
2nd MANTISSA PORTION PROCESSING MEANS
2nd EXPONENTIAL PORTION PROCESSING MEANS
X1
192
194
195
196
QP

FIG.9


200
208
L
2nd MANTISSA PORTION PROCESSING MEANS
2nd SHIFTING PORTION PROCESSING MEANS
X1
202
204
205
206
QP


100
132
134
136
135
138
137
INPUT IMAGE DATA
130
133
T
Q
CODING
OUTPUT ENCODED DATA
139
131
UI
QP
L
DQ
140
IT
142
141
145
144
149
MC
148
FM
146
ME
147
T : TRANSFORM MODULE
Q : QUANTIZATION MODULE
DQ : DEQUANTIZATION MODULE
IT : INVERSE TRANSFORM MODULE
FM : FRAME MEMORY
MC : MOTION COMPENSATION MODULE
ME : MOTION ESTIMATION MODULE