PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
H04N 7/36, 7/26, G06T 7/20

A1

(11) International Publication Number: **WO 97/40628**

(43) International Publication Date: 30 October 1997 (30.10.97)

(21) International Application Number: PCT/EP96/01644

(22) International Filing Date: 19 April 1996 (19.04.96)

(71) Applicant *(for all designated States except US)*: NOKIA MOBILE PHONES LTD. [FI/FI]; P.O. Box 68, FIN-24101 Salo (FI).

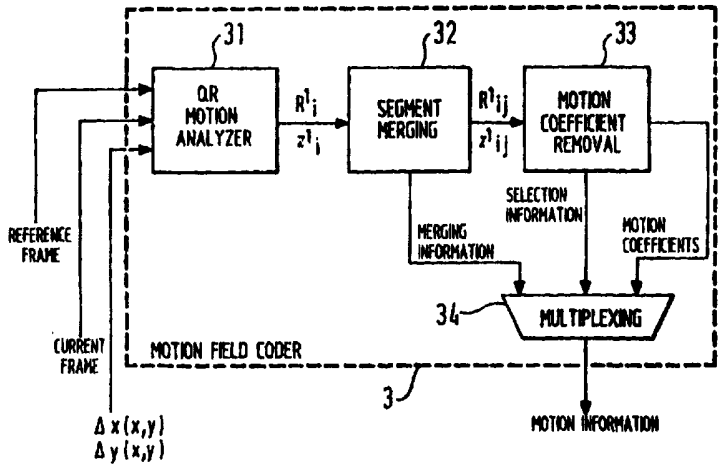
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: NIEWEGLOWSKI, Jacek [PL/FI]; Lemminkäisenkatu 8D76, FIN-33540 Tampere (FI). KARCZEWICZ, Marta [PL/FI]; Insinöörinkatu 60C146, FIN-33720 Tampere (FI).

(74) Agent: JOHANSSON, Folke; Nokia Mobile Phones Ltd., Patent Dept., P.O. Box 100, FIN-00045 Nokia Group (FI).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: VIDEO ENCODER AND DECODER USING MOTION-BASED SEGMENTATION AND MERGING



(57) Abstract

This invention relates to motion compensated (MC) coding of video and to an MC prediction scheme which allows fast and compact encoding of motion vector fields retaining at the same time very low prediction error. By reducing prediction error and number of bits needed for representation of motion vector field, substantial savings of bitrate are achieved. Reduction of bitrate needed to represent motion field is achieved by merging segments in video frames, by adaptation of motion field model and by utilization of motion field model based on orthogonal polynomials.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

VIDEO ENCODER AND DECODER USING MOTION-BASED SEGMENTATION AND MERGING

FIELD OF THE INVENTION

The present invention generally relates to video compression. More precisely, the invention relates to an encoder and method for performing motion compensated encoding of video data and of motion compensated encoding of video data. The invention also relates to a decoder for decoding video data thus encoded.

BACKGROUND OF THE INVENTION

The motion compensated prediction is a key element of majority of video coding schemes. Figure 1 is a schematic diagram of an encoder for compression of video sequences using motion compensation. Essential elements in the encoder are a motion compensated prediction block 1, a motion estimator 2 and a motion field coder 3. The operating principle of the motion compensating video encoder is to compress the prediction error $E_n(x,y)$, which is a difference between the incoming frame $I_n(x,y)$ to be coded, called the current frame, and a prediction frame $\hat{I}_n(x, y)$, wherein:

$$E_n(x, y) = I_n(x, y) - \hat{I}_n(x, y) \qquad (1)$$

The prediction frame $\hat{I}_n(x, y)$ is constructed by motion compensated prediction block 1 and is built using pixel values of the previous, or some other already coded frame denoted $\tilde{I}_{n-1}(x, y)$, called a reference frame, and motion vectors describing estimated movements of pixels between the current frame and the reference frame. Motion vectors are calculated by motion field estimators 2 and the resulting motion vector field is then coded in some way before applying it to the predictor block 1. The prediction frame is then:

$$\hat{I}_n(x, y) = \tilde{I}_{n-1}[x + \Delta x(x, y), y + \Delta y(x, y)] \qquad (2)$$

The pair of numbers $[\Delta x(x,y), \Delta y(x,y)]$ is called the motion vector of a pixel in location (x,y) in the current frame, whereas $\Delta x(x,y)$ and $\Delta y(x,y)$ are the values of horizontal and vertical displacement of this pixel. The set of motion vectors of all pixels in the current frame $I_n(x,y)$ is called motion vector field. The coded motion vector field is also transmitted as motion information to the decoder.

In the decoder shown in Fig. 2, pixels of the current frame $I_n(x,y)$ are reconstructed by finding the pixels' predictions $\hat{I}_n(x, y)$ in the reference frame $\tilde{I}_{n-1}(x, y)$. Motion compensated prediction block 21 generates the prediction frame using the received motion information and the reference frame $\tilde{I}_{n-1}(x, y)$ (in this picture the reference frame is the same as the current frame). In the prediction error decoder 22 a decoded prediction error frame $E_n(x,y)$ is then added with the

prediction frame, the result being the original current frame I_n.

The general object of the motion compensated (MC) prediction encoder is to minimize the amount of information which needs to be transmitted to the decoder. It should minimize the amount of prediction error measured, e.g. as the energy of $E_n(x,y)$, and minimize the amount of information needed to represent the motion vector field.

The document N. Nguen, E. Dubois, "Representation of motion information for image coding". Proc. Picture Coding Symposium '90, Cambridge, Massachusetts, March 26-18, 1990, pages 841-845, gives a review of motion field coding techniques.

As a rule of the thumb, a reduction of prediction error requires a more refined sophisticated motion field, i.e. more bits must be spent on its encoding. Therefore the overall goal of the video encoding is to encode as compactly as possible the motion vector field keeping at the same time the measure of prediction error as low as possible.

Due to very large number of pixels in the frame it is not efficient to transmit a separate motion vector for each pixel. Instead, in most of the video coding schemes the current frame is divided into image segments as shown for example in figure 3 so that all motion vectors of the segment can be described by few parameters. Image segments can be square blocks. For example 16 x 16 pixels blocks are used in codecs in accordance with international standard ISO/IEC MPEG-1 or ITU-T H.261, or they can comprise completely arbitrarily shaped regions obtained for instance by a

segmentation algorithm. In practice segments include at least few tens of pixels.

The motion field estimation block 1 of fig. 1 calculates motion vectors of all the pixels of a given segment which minimize some measure of prediction error in this segment, for example square prediction error. Motion field estimation techniques differ both in the model of the motion field and in the algorithm for minimisation of the chosen measure of prediction error.

In order to compactly represent the motion vectors of the pixels in the segments it is desirable that their values are described by a function of few parameters. Such function is called motion vector field model. A known group of models are linear motion models, in which motion vectors are approximated by linear combinations of motion field basis functions. In such models the motion vectors of image segments are described by a general formula:

$$\Delta x(x, y) = \sum_{i=1}^{N} c_i f_i(x, y)$$
$$\Delta y(x, y) = \sum_{i=N+1}^{N+M} c_i f_i(x, y) \qquad (3)$$

where parameters c_i are called motion coefficients and are transmitted to the decoder. Functions f_i (x,y) are called motion field basis functions which have a fixed form known both to encoder and decoder.

The problem when using the linear motion model having the above described formula is how to minimize in a computationally simple manner the number of motion

coefficients c_i which are sent to the decoder, keeping at the same time some measure of distortion, e.g. a chosen measure of prediction error $E_n(x,y)$, as low as possible.

The total number of motion coefficients which needs to be sent to the decoder depends both on the number of segments in the image and the number of motion coefficients per segment. Therefore, there exist at least two ways to reduce the total number of motion coefficients.

The first way is to reduce the number of segments by combining (merging) together those segments which can be predicted with a common motion vector field model without causing a large increase of prediction error. The number of segments in the frame can be reduced because very often the adjacent, i.e. neighbouring, segments can be predicted well with the same set of motion coefficients. The process of combining such segments is called motion assisted merging. Figure 3 shows a frame divided into segments. The prior art techniques for motion coefficient coding include several techniques for motion assisted merging. After motion vectors of all the segments have been estimated, the motion assisted merging is performed. It is done by considering every pair of adjacent segments S_i and S_j with their respective motion coefficients c_i and c_j. The area of combined segments S_i and S_j is denoted S_{ij}. If the area S_{ij} can be predicted with one set of motion coefficients c_{ij} without causing excessive increase of prediction error over the error resulting from separate predictions of S_i and S_j, then S_i and S_j are merged. The methods for motion assisted merging differ essentially in the way of finding a single set of motion coefficients c_{ij} which allow a good prediction of segments combined together.

One method is known as merging by exhaustive motion estimation. This method estimates "from scratch" a new set of motion parameters c_{ij} for every pair of adjacent segments S_i and S_j. If the prediction error for S_{ij} is not excessively increased then the segments S_i and S_j are merged. Although this method can very well select the segments which can be merged it is not feasible for implementation because it would increase the complexity of the encoder typically by several orders of magnitude.

Another method is known as merging by motion field extension. This method tests whether area of S_{ij} can be predicted using either motion parameters c_i or c_j without an excessive increase of the prediction error. This method is characterized by very low computational complexity because it does not require any new motion estimation. However, it fails to merge segments very often because motion compensation with coefficients calculated for one segment very rarely predicts well also the adjacent segments.

Still another method is known as merging by motion field fitting. In this method the motion coefficients c_{ij} are calculated by the method of approximation. This is done by evaluating few motion vectors in each of the segments. Some motion vectors in segments S_i and S_j are depicted in figure 3. The motion field for the segment S_{ij} is done by fitting a common motion vector field through these vectors using some known fitting method. The disadvantage of the method is that motion field obtained by fitting is not precise enough and often leads to an unacceptable increase of prediction error.

Methods for performing motion estimation with different models and selecting the most suitable one are proposed in

H. Nicolas and C. Labit, "Region-based motion estimation using deterministic relaxation schemes for image sequence coding," Proc. 1994 International Conference on Acoustics, Speech and Signal Processing, pp. III265-268 and P. Cicconi and H. Nicolas, "Efficient region-based motion estimation and symmetry oriented segmentation for image sequence coding," IEEE Tran. on Circuits and Systems for Video Technology, Vol. 4, No. 3, June 1994, pp. 357-364. The methods try to adapt the motion model depending on complexity of the motion by performing motion estimation with different models and selecting the most suitable one. The main disadvantage of these methods is their high computational complexity and low amount of different motion field models which can be tested in practice.

The second way to minimize the number of motion coefficients is to select for each segment a motion model which allows achieving satisfactorily low prediction error with as few coefficients as possible. Since the amount and the complexity of the motion varies between frames and between segments it is not efficient to use always all N+M motion coefficients per segment. It is necessary to find out for every segment what is the minimum number of motion coefficients which allows to achieve satisfactorily low prediction error. The process of such adaptive selection of coefficients is called motion coefficient removal.

None of the afore-described methods alone solves the problem how to minimize the number of motion coefficients c_i which are sent to the decoder, preserving at the same time said measure of prediction error as low as possible.

SUMMARY OF THE INVENTION

An object of the present invention is to create a motion compensated video encoder and a method of motion compensated encoding of video data and a video decoder for decoding motion compensation encoded video data, which allow reducing the amount of motion vector field data produced by some known motion estimator by a large factor without causing an unacceptable distortion of the motion vector field. The complexity of the motion field encoder should preferably be low for allowing practical implementation on available signal processors or general purpose microprocessors.

This invention is solved as defined in claims 1, 10, 20, 23, 26 and 39, respectively. Embodiments of the invention are defined in the dependent claims. In accordance with a first aspect of the invention the encoder can perform merging of segments and find for a merged segment in a computationally efficient way a set of motion coefficients minimizing a measure of distortion. This aspect of the invention also allows a simple and efficient evaluation of distortion due to segment merging, if desired. Advantageously, the encoder according to this aspect performs adaptive merging of adjacent segments of a video frame by means of calculating an additional distortion according to a predefined measure and merging the segments if the additional distortion is tolerable, e.g. below a given threshold or tolerable with regard to an achieved bit rate reduction. A chosen measure for calculating this distortion can be but is not limited to some measure of prediction error, for instance the energy or squared predition error in the segment. Another measure for distortion can be e.g. the squared difference between an original frame and the restored original frame after encoding

and decoding. For this purpose the motion field encoder included in the video encoder comprises two main blocks.

The first main block may be called a QR motion analyzer. Its task is to find a new representation of the inputted motion vector field produced by the motion field estimator. This new representation is applied to the second main block. Operations in this first main block include a plurality of steps comprising matrix operations: In the first step the prediction frame is linearized using some known approximation method so that the prediction frame becomes a linear function of the motion vectors. In the second step a matrix E_i and a matrix y_i are constructed for minimisation of an appropriate measure of prediction error, e.g. of the square prediction error. For this purpose matrix E_i is decomposed into a product of two matrices Q_i and R_i. In addition, an auxiliary vector z_i is calculated from the factor matrix Q_i and the matrix y_i. Part of the matrix R_i and the auxiliary vector z_i are applied to the second main block.

The second main block called a segment merging block performs merging operations for pairs of segments S_i, S_j. Advantageously, this block checks whether their combined area can be predicted using a common motion field model for the combined area. In the merging operations a matrix equation is firstly formed based on said factor matrices, thereafter the factor matrices are processed by using known matrix computation methods. The result is a matrix equation which allows calculating motion coefficients common for the pair of segments under consideration, in a simple and efficient way. Using these coefficients the chosen measure of distortion can be calculated in the area of the merged segments. If square prediction error is used as the measure of distortion, it can

be easily calculated on the basis of the terms included in one of the resulting matrices. If the change of said measure of prediction error is acceptable according to a chosen criterion, the segments are merged.

After all pairs of segments are considered the output of the segment merging block is a new segmentation of the image with a reduced number of segments. Moreover, for each new segment the block outputs a matrix R^1ij and a vector z^1_{ij}, which allow calculating all motion coefficients in a simple and efficient way. Also, the encoder provides the decoder with information enabling the reconstruction of the resulting new segments in the frame.

The segment merging block according to this aspect of the invention allows a computationally simple judgement whether segments can be merged into one. Merging of as many segments as possible can be achieved by judging for each pair of adjacent segments whether merging is possible, and repeating this process for the resulting segmentation of the frame until no pair of adjacent segments suitable for merging remains in the frame.

According to an embodiment of the invention the amount of distortion can be calculated based on a linear approximation of additional distortion due to approximating the motion vector field by a linear combination of basis functions.

An encoder according to a second aspect of the present invention allows removing coefficients from the set of coefficients representing the motion vector field of a segment and finding in a computationally efficient way the optimum remaining coefficients having regard to a measure of

distortion. Also, if desired, the invention according to this aspect allows of checking whether the omission of a particular coefficient of this set causes a significant increase of distortion in the motion field or not. This check is performed in a computationally efficient way such that for each segment S_i each coefficient of the set may be subjected to such checking. Only those coefficients need to be transmitted to the decoder that are found to increase the distortion significantly if omitted.

An encoder according to a second aspect comprises a first main block operating similar to the first main block of the first aspect of the invention. It furthermore comprises a second block which for every segment S_i receives a matrix R_i and a vector z_i produced by the first block and, if desired, which determines based on matrix R_i and vector z_i for each of the segments whether it is possible to simplify the motion field model by removing basis functions from the model without intolerable increase of distortion.

The operations in this second block are matrix operations, in which the matrix equation is firstly modified by removing one column and row of the matrix equation $R_i c = z_i$, c being a vector comprising the coefficients c_i of the model. Removal of one column and row corresponds to removal of one basis function from the motion model. Then the matrix equation is triangularized. Motion coefficients corresponding to the reduced set of basis functions can be calculated by solving the resulting linear equation. The equation can be solved using back substitution or some other well known algorithm. If the prediction error is used as a measure of distortion, its change for the segment caused by removal of a basis

function is a simple predetermined function of one term in the resulting equation.

More coefficients for the segment can be reduced by further repeating these matrix operations. Using such an approach one can easily find the distortion when using different reduced sets of basis functions. The set which yields the desired distortion is selected to represent the motion of the segment.

This block outputs for every processed segment selection information which tells which basis functions where removed from the motion field model. Additionally it outputs new motion coefficients corresponding to the remaining basis functions. Both selection information and motion coefficients are transmitted to the decoder.

The removal of a basis function from a set of basis functions is fully equivalent to setting the value of a coefficient corresponding to the removed basis function to zero. Thus, in an alternative implementation the encoder may instead of outputting the selection information followed by motion coefficients, output all motion coefficients with coefficients corresponding to removed basis functions having values equal to zero.

Preferably, a motion compensated video encoder takes advantage both of adaptively merging segments and of removing motion coefficients which are not significant for the overall distortion. Such a preferred encoder includes a QR motion analyser and a segment merging block according to the first aspect of the invention and also a coefficient removal block according to the second aspect of the invention. The

coefficient removal block receives a matrix R_{ij} and a vector z_{ij} from the segment merging block. Only after merging of adjacent segments and after manipulating matrix R and vector z for coefficient removal, the coefficients c_i of each segment are calculated for transmission, resulting in a substantial reduction in the amount of motion data output by the video encoder.

Preferably, a video encoder and decoder according to the present invention is implemented in hardware, e.g. as one or more integrated circuits, adapted to perform encoding and compressing of received video frames, and decoding encoded video data, respectively, according to the present invention.

It is common in the video coding art that different areas of the video frame are coded using different coding modes. This variability of coding modes includes also the motion compensated prediction method used in the codex. Several modes are used in all modern video codecs such as the ITU H.261 and H.263 as well as the ISO MPEG-1 and MPEG-2 video coding standards.

For example, some of the video frame areas are coded without using any temporal prediction at all (so called intra-blocks). No motion coefficients are transmitted for such image areas and the areas are coded without reference to any prior images. Practical embodiments of the present invention would also combine the invention with such intra-coding.

Furthermore, in typical video image sequences large areas of the video frame remain stationary for the duration of several frames (i.e., there is no motion). It is computationally much easier to detect that an area has remained stationary than to

estimate the motion of the video frame area. Therefore, practical video codecs often check the area to detect whether there is any motion at all and include a stationary prediction mode where no motion parameters need to be transmitted. Practical embodiments of the present invention would also combine the invention with such a stationary mode.

The MPEG video coding standards also include coding modes where the motion estimation is performed with respect to two reference frames (bi-directional prediction). This results in two different predictions for the area to be coded. The encoder may decide to use the better of these predictions, or it may decide to combine the two predictions (e.g., by averaging). The decision regarding the mode needs to be communication to the decoder. The ITU H.263 standard also includes a temporal prediction mode using two reference frames. It is clear that the present invention can benefit from similar techniques of using multiple reference frames.

It is thus clear for those skilled in the art that the present invention can be the basis for one or mode coding modes in a video codec where it is used together with prior art coding modes (such as intra coding, stationary modes, or multiple reference frames).

A preferred embodiment of the present invention will now be described with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS.

Fig. 1 is a schematic diagram of a known encoder;
Fig. 2 is a schematic diagram of a known decoder;
Fig. 3 depicts adjacent segments for merging;

Fig. 4 illustrates merging by motion field approximation;

Fig. 5 is the motion field coder according to the invention;

Fig. 6 is a schematic diagram of QR motion analyser.

Fig. 7 is a schematic diagram of an embodiment of a motion compensated prediction block according to the present invention.

DESCRIPTION OF PREFERRED EMBODIMENT

The output of the video encoder shown in fig. 1 is the compressed frame divided into segments S_i and each of the segments S_i is accompanied by information regarding motion vectors $[\Delta x(x,y), \Delta y(x,y)]$ for each pixel (x,y) of the respective segment. Then for a segment S_i which consists of P pixels with coordinates (x_i,y_i), i=1,2, ..., P, the task of the motion field encoder 3 is to find from the motion vector field $(\Delta x(x_i,y_i),\Delta y(x_i,y_i))$ output by motion field estimation block 2 the motion coefficients $c=(c_1, c_2, \ldots, c_{N+M})$ of a motion vector field $[\tilde{\Delta}x(\cdot),\tilde{\Delta}y(\cdot)]$ approximating $[\Delta x(x,y), \Delta y(x,y)]$ as precisely as necessary by a linear motion model of the form:

$$\tilde{\Delta}x(x,y) = \sum_{i=1}^{N} c_i f_i(x,y) \qquad (4a)$$

$$\tilde{\Delta}y(x,y) = \sum_{i=N+1}^{N+M} c_i f_i(x,y) \qquad (4b)$$

such that it minimizes the square prediction error SPE:

$$SPE = \sum_{i=1}^{P} \left(I_n(x_i, y_i) - I_{n-1}\left(x_i + \tilde{\Delta}x(x_i, y_i), y_i + \tilde{\Delta}y(x_i, y_i)\right)\right)^2 \quad (5)$$

Fig. 5 illustrates an embodiment of a motion field encoder in a video encoder according to the invention. It corresponds to the block 3 in Fig. 1 but it has as inputs also the reference frame and the current frame. The third input to this block is the motion vector field $[\Delta x(\cdot), \Delta y(\cdot)]$ produced by motion field estimation block 2, Fig. 1.

To fulfil said task, the motion field encoder 3 consists of three main building blocks which are the QR motion analyser block 31, the segment merging block 32 and the motion coefficient removal block 33. The segment merging block 32 and the motion coefficient removal block 33 reduce the amount of motion information which may result in a less accurate prediction and hence in an increase of the square prediction error.

The objective of the QR motion analyzer is to find a new representation of the motion field that is suitable for judging efficiently how large the impact of segment merging and coefficient removal on the prediction error is. This new representation is used later in the other two blocks for a fast and flexible finding of motion coefficients for merged segments and for coefficient removal. Fig. 6 shows an embodiment of the QR motion analyser according to this invention. This block comprises a gradient filter 41 receiving a previous video frame input I_{n-1}, or more preferably, receiving a previous video frame $\tilde{I}_{n-1}$ similar to the frame $\tilde{I}_{n-1}$ decoded by the video decoder shown in fig. 2. For this purpose also the video encoder of fig. 1 comprises a

prediction error decoder block 5 and means 7 for calculating video frames $\tilde{I}$ corresponding to the video frames decoded by the video decoder.

The outputs G_x, G_y of the gradient filter are input into a block 42 for building a matrix E and into a block 45 for building a vector y. Matrix building block 42 performs a linearization of the prediction frame $\tilde{I}$ such that the approximated prediction frame is a linear function of $\tilde{\Delta}x$ and $\tilde{\Delta}y$ and calculates on the basis of this linearization a matrix E the multiplication of which with a vector c of coefficients c_i in equations 4a, 4b above may be interpreted as prediction error resulting if Δx, Δy are replaced by a linear combination of basis functions $f_i(x,y)$ of a linear motion model.

Block 45 for building vector y receives the current frame I_n, predicted frame $\tilde{I}_{n-1}$, the outputs G_x, G_y of the gradient filter 41 and the motion vectors $[\Delta x(x,y), \Delta y(x,y)]$ estimated by block 2 in fig. 1 and calculates from these input right values said vector y which may be interpreted as the prediction error due to the motion in the scene, i.e. inherent prediction error.

Matrix E and vector y are received by a QR factorizer block 43 and a matrix multiplier block 46, resepctively. The function of these blocks can be regarded as a coordinate transformation of matrix E and vector y in order to prepare for finding coefficients c_i such that for all pixels of a given segment the prediction error resulting from the representation of Δx, Δy as a linear combination of basis functions f_i is as close as possible to the inherent

prediction error. This will be explained in further detail below.

Block 43 outputs a matrix R which results from representing matrix E in the coordinates of a matrix Q also output by block 43. Block 46 receives not only said vector y but also said matrix Q from block 43 and finds a vector z representing y in the coordinates of matrix Q. Preferably matrix Q is orthonormal. As will be shown in further detail below this representation of E and y as R and z, respectively, is very advantageous for judging whether adjacent segments can be merged with tolerable increase of prediction error, and also for finding the minimum number of coefficients necessary for representing the motion vector field of a merged or non-merged segment, i.e. for removing non-significant coefficients from the set of coefficients c_i in equations 4a, 4b.

Blocks 44 and 47 receive matrix R and vector z, respectively and select rows from these which are required for judging the effect of segment merging and/or motion coefficient removal. These operations are performed based on R and z without the need of calculating said cofficients c_i. Furthermore, all row manipulations refer both to the rows of R and to the corresponding rows of z such that R and z can be regarded for the purpose of segment merging and/or motion coefficient removal as a single column extended matrix A comprising the columns of R and comprising as an additional column, vector z. Accordingly, blocks 44 and 47 can be regarded and implemented as one block for manipulating matrix A by selecting appropriate rows of A and for outputting a modified matrix A' comprising the selected rows of A. A' comprises the

selected rows denoted R^1 of R and the selected rows denoted z^1 of z.

Segment merging block 32 receives R^1 and z^1, i.e. matrix A', for each segment and judges whether merging of two segments S_i, S_j by means of representing the motion vector fields of both segments with the same set of coefficients, results in a tolerable increase of prediction error. This is done by means of generating a row extended matrix B comprising all rows of matrix A'_i of segment S_i and of matrix A'_j of segment S_j. Segments S_i, S_j can be but don't have to be adjacent. Matrix B is subjected to a further coordinate transformation e.g. by means of triangularization of matrix B, resulting in a modified matrix B'. Block 32 in fig. 5 judges whether segment merging is possible, from selected elements in matrix B' in that column which results from vectors z'_i and z'_j and in rows which have zeros in the columns of B' resulting from matrices R^1_i and R^1_j. Preferably, said further coordinate transformation is orthonormal. Then the additional prediction error resulting from merging is the sum of the square of said selected elements.

Motion coefficient removal block 33 in fig. 5 receives for each segment after frame resegmentation said matrix A' if the corresponding segment remained unmerged, or matrix B' for merged segments and judges whether a removal of coefficients c_i is possible with a tolerable increase of prediction error. This is performed by block 33 by means of extracting a row from matrix A' and B', respectively, which row corresponds to coefficient c_i, and performing a preferably orthonormal coordinate transformation of the matrix resulting from matrix A' and B', respectively, after extraction of said row. The additional prediction error due to removing coefficient c_i

can then be calculated from a selected element of said transformed matrix, said selected element being located in the column resulting from z^1 of said transformed matrix and in the row of this matrix which has zeros in all columns resulting from R^1. Multiplexer 34 in fig. 5 receives merging information from block 32, information on which coefficients c_i are selected for transmission, and the selected coefficients c_i which are finally calculated based on said transformed matrix by block 33. This information output by multiplexer 34 is then output to the video decoder (not shown).

In more detail the operation of QR Motion Analyzer consists of following steps:

Step 1 is linearization of prediction error. In this step the prediction frame $\tilde{I}_{n-1}(\cdot)$ in equation (7) is approximated using some known approximation method so that it becomes linear with respect to $[\tilde{\Delta}x(\cdot), \tilde{\Delta}y(\cdot)]$. Then elements under sum in the formula (5) become linear combinations of coefficients c_i

$$SPE = \sum_{j=1}^{P} \left(e_{j,1}c_1 + e_{j,2}c_2 + \cdots + e_{j,N+M}c_{N+M} - y_j\right)^2 \qquad (6)$$

In the preferred implementation a quadratic polynomial motion vector field model with 12 coefficients is used:

$$\tilde{\Delta}x(x, y) = c_1 + c_2x + c_3y + c_4xy + c_5x^2 + c_6y^2 \qquad (4c)$$

$$\tilde{\Delta}y(x, y) = c_7 + c_8x + c_9y + c_{10}xy + c_{11}x^2 + c_{12}y^2 \qquad (4d)$$

This model in practice can handle very well even very complex motion in video sequences and yields good prediction results.

In the QR motion analyzer block, linearization in the step 1 is done by using Taylor expansion of $\tilde{I}_{n-1}(\cdot)$ at every pixel (x_i, y_i) where i=1,2,...P, around points:

$$x'_i = x_i + \tilde{\Delta}x(x_i, y_i)$$
$$y'_i = y_i + \tilde{\Delta}y(x_i, y_i)$$

Using property that $\sum a^2 = \sum (-a)^2$, the square prediction error SPE is then

$$SPE = \sum_{i=1}^{P} \left(\tilde{I}_{n-1}(x'_i, y'_i) + \left(\tilde{\Delta}x(x_i, y_i) - \Delta x(x_i, y_i)\right) G_x(x'_i, y'_i) \right.$$
$$\left. + \left(\tilde{\Delta}y(x_i, y_i) - \Delta y(x_i, y_i)\right) G_y(x'_i, y'_i) - I_n(x_i, y_i) \right)^2$$

Auxiliary values $g_j(x,y)$ are calculated using formula:

$$g_j(x_i, y_i) = \begin{cases} f_j(x_i, y_i) G_x(x_i, y_i) & \text{when } j = 1, 2, \ldots, N \\ f_j(x_i, y_i) G_y(x_i, y_i) & \text{when } j = N+1, N+2, \ldots, N+M \end{cases}$$

where function $f_j(x_i, y_i)$ is a predefined basis function according to the motion field model as defined in the equations (4a) and (4b) and more specifically, in equations (4c) and (4d).

Step 2 is construction of matrices. It is based on the fact that minimisation of SPE according to formula (6) is fully

equivalent to minimisation of the matrix expression $(Ec-y)^T(Ec-y)$, where E and y are:

$$E = \begin{bmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,N+M} \\ e_{2,1} & e_{2,2} & \cdots & e_{2,N+M} \\ \vdots & \vdots & \vdots & \vdots \\ e_{P,1} & e_{P,2} & \cdots & e_{P,N+M} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_P \end{bmatrix} \quad (7)$$

Matrix E and vector y in equation (9) are built using formulas:

$$E = \begin{bmatrix} g_1(x_1, y_1) & g_2(x_1, y_1) & \cdots & g_{N+M}(x_1, y_1) \\ g_1(x_2, y_2) & g_2(x_2, y_2) & \cdots & g_{N+M}(x_2, y_2) \\ \vdots & \vdots & \vdots & \vdots \\ g_1(x_P, y_P) & g_2(x_P, y_P) & \cdots & g_{N+M}(x_P, y_P) \end{bmatrix}$$

$$y = \begin{bmatrix} I_n(x_1, y_1) - \tilde{I}_{n-1}(x'_1, y'_1) + G_x(x'_1, y'_1)\Delta x(x_1, y_1) + G_y(x'_1, y'_1)\Delta y(x_1, y_1) \\ I_n(x_2, y_2) - \tilde{I}_{n-1}(x'_2, y'_2) + G_x(x'_2, y'_2)\Delta x(x_2, y_2) + G_y(x'_2, y'_2)\Delta y(x_2, y_2) \\ \vdots \\ I_n(x_P, y_P) - \tilde{I}_{n-1}(x'_P, y'_P) + G_x(x'_P, y'_P)\Delta x(x_P, y_P) + G_y(x'_P, y'_P)\Delta y(x_P, y_P) \end{bmatrix}$$

$G_x(x,y)$ and $G_y(x,y)$ are values of the horizontal and vertical gradient of the reference frame $\tilde{I}_{n-1}(x, y)$ calculated using following formula:

$$G_x(x, y) = \tilde{I}_{n-1}(x + 1, y) - \tilde{I}_{n-1}(x - 1, y)$$
$$G_y(x, y) = \tilde{I}_{n-1}(x, y + 1) - \tilde{I}_{n-1}(x, y - 1)$$

The pixel values of $\tilde{I}_{n-1}(x, y)$, $G_x(x,y)$ and $G_y(x,y)$ are defined only for integer coordinates x and y. When x or y are noninteger, the pixel value is calculated e.g. using a bilinear interpolation of closest pixels with integer co-ordinates.

Step 3 is QR Factorization. QR factorization of a matrix is as such well known and a suitable algorithm is described in D.H. Golub and C. van Loan, "Matrix computation" 2'nd edition, The Johns Hopkins University Press, 1989. This algorithm can be used to decompose E into a product of two matrices

$$E=Q\ R \qquad (8)$$

In other words, R is a representation of E in coordinates of Q. Q preferably is orthonormal and such that R is upper triangular, i.e. rows N+M+1 to P of R are all zero.

In this step is also calculated an auxiliary vector z where

$$z = Q^T y \qquad (9)$$

In step 4 the output of the QR motion analyzer block is calculated. The output comprises of a matrix R^1 consisting of N+M first rows of matrix R and a vector z^1 consisting of first N+M elements of z.

In the segment merging block merging operation is performed for pairs of adjacent segments S_i and S_j, see Fig. 4, by judging whether for combined segment S_{ij} the motion vector

field can be represented using a common motion field described by motion coefficients c_{ij}. Merging operation consists of the following steps:

Step 1 comprises matrix calculation. This invention utilizes a previously unknown property that motion coefficients c_{ij} minimizing the prediction error in the merged segment Sij also minimize the scalar value

$$\left(\begin{bmatrix} R^1{}_i \\ R^1{}_j \end{bmatrix} C_{ij} - \begin{bmatrix} z^1{}_i \\ z^1{}_j \end{bmatrix}\right)^T \left(\begin{bmatrix} R^1{}_i \\ R^1{}_j \end{bmatrix} C_{ij} - \begin{bmatrix} z^1{}_i \\ z^1{}_j \end{bmatrix}\right) \qquad (10)$$

where $R^1{}_i, z^1{}_i$ and $R^1{}_j$, $z^1{}_j$ are already produced by the QR analyzer block for segments S_i and S_j, respectively, as described above. This minimization of (10) is equivalent to solving in the least square sense the overdetermined system of equations

$$\begin{bmatrix} R^1{}_i \\ R^1{}_j \end{bmatrix} C_{ij} = \begin{bmatrix} z^1{}_i \\ z^1{}_j \end{bmatrix} \qquad (10a)$$

Step 2 comprises triangularization of the matrices obtained in the step 1. If QR factorization of E for segment S_i, that is E_i and of E for segment S_j, that is E_j according to the teaching of the afore-mentioned document is applied, matrices $R^1{}_i$, $R^1{}_j$ are upper triangular and the system

$$\begin{bmatrix} R^1{}_i \\ R^1{}_j \end{bmatrix} C_{ij} = \begin{bmatrix} z^1{}_i \\ z^1{}_j \end{bmatrix}$$

in (10) has the form:

$$\begin{bmatrix} X & X & X & \cdots & X \\ & X & X & \cdots & X \\ & & X & \cdots & X \\ & & & \ddots & \vdots \\ & & & & X \\ X & X & X & \cdots & X \\ & X & X & \cdots & X \\ & & X & \cdots & X \\ & & & \ddots & \vdots \\ & & & & X \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_{N+M} \end{bmatrix} = \begin{bmatrix} z^i{}_1 \\ z^i{}_2 \\ z^i{}_3 \\ \vdots \\ z^i{}_{N+M} \\ z^j{}_1 \\ z^j{}_2 \\ z^j{}_3 \\ \vdots \\ z^j{}_{N+M} \end{bmatrix} \qquad (11)$$

where symbol x denotes a nonzero element and $z^i{}_k$ denotes the kth element of vector $z^1{}_i$, $z^j{}_k$ denotes the kth element of vector $z^1{}_j$.

The system of equation (11) is triangularized by applying a series of multiplications of rows by scalars followed by additions of the rows; i.e. it is converted to the form:

$$\begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,N+M} \\ 0 & r_{2,2} & r_{2,3} & \cdots & r_{2,N+M} \\ 0 & 0 & r_{3,3} & \cdots & r_{3,N+M} \\ 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & r_{N+M,N+M} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_{N+M} \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{N+M} \\ q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_{N+M} \end{bmatrix} \quad (12)$$

For this triangularization again QR factorization according to said document mentioned above may be used.

In step 3 the merging error is evaluated. The change of the square prediction error ΔE_{ij} in the segment S_{ij} caused by merging of segments S_i and S_j is calculated according to

$$\Delta E_{ij} = \sum_{k=1}^{N+M} q^2_k \quad (13)$$

In this preferred embodiment, QR factorization of (11) results in Q being orthonormal, such that equation 13 is very simple. However, depending on the properties of Q in this factorization, ΔE_{ij} is in general a function of q^2_k, k=1,..., N+M, if as a measure for prediction error the square prediction error is used; of course, other measures

for prediction error are feasible and accordingly, other functional relations between q_k and ΔE_{ij} may be adapted.

Finally, in step 4 the segments are merged if the change of square prediction error in formula (13) is acceptable according to a chosen criterion. The segment merging block uses the following strategy for segment merging:

a. a threshold T is selected which corresponds to allowed increase of square prediction error in whole frame.
b. for all pairs of adjacent segments ΔE_{ij} is calculated using equation (13).
c. pair of segments with smallest ΔE_{ij} is merged.
d. points b-c are repeated until sum of ΔE_{ij} corresponding to all merged pairs of segments is greater than T.

For triangularization of the system in equation (11) a sequence of Givens rotations can be used.

For the resulting new segment S_{ij} matrix $R^1{}_{ij}$ and vector $z^1{}_{ij}$ are built by taking first N+M rows of the system in equation (12), i.e. are given by formula:

$$R^1{}_{ij} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,N+M} \\ & r_{2,2} & r_{2,3} & \cdots & r_{2,N+M} \\ & & r_{3,3} & \cdots & r_{3,N+M} \\ & & & \ddots & \vdots \\ & & & & r_{N+M,N+M} \end{bmatrix}, z^1{}_{ij} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{N+M} \end{bmatrix} \qquad (14)$$

After all pairs of segments of the frame are considered the output of the segment merging block is obtained. The output comprises of three kinds of information. Firstly, it gives a

new division of the image with reduced number of segments. Secondly, for each new segment the block outputs matrix R^1_{ij} and vector z^1_{ij}. Thirdly, it gives merging information which is sent to the decoder and helps the decoder to identify segments which were merged.

The motion coefficients $c_{ij}=(c_1,c_2,...c_{N+M})$ for the segment S_{ij} could be now calculated by solving the system of equations $R^1_{ij}c_{ij}=z^1_{ij}$ but their calculation is not yet necessary if the coefficient removal block 33 is used. Also, as will be described further below, at this stage and prior to performing coefficient removal it may be advantageous to orthogonalize the motion field model with respect to the segments obtained from segment merging. Such orthogonalized motion field model is less sensitive to quantization errors resulting from quantizing errors resulting from quantizing the motion coefficients. However, in order to keep the description of the preferred embodiment as clear as possible, the optimal motion field model orthogonalization will be omitted here.

Coefficient removal block 33 receives as the input said new division of the current frame into segments and for every segment S_k it receives said matrices R^1_k, z^1_k shown in (14), produced previously by the segment merging block. Motion vectors of every segment are represented by N+M motion coefficients, N and M being determined by the motion field model for Δx and Δy.

The motion coefficient removal block 33 determines for a given segment S_k whether it is possible to simplify the motion field model, without excessively increasing the prediction error. A simplified motion field model is obtained

when some basis functions are removed from the model in equations (3) described in the background art of this application and less coefficients are required to describe such a simplified motion field model.

The following procedure is performed by block 33 for all segments in order to find out whether the i'th basis function (and i'th coefficient) can be removed from the motion field model:

Step 1 includes a matrix modification, where the system of linear equations

$$R^1{}_k c_k = z^1{}_k \qquad (15)$$

is modified by removing i'th column from $R^1{}_k$ and i'th element from c_k.

Step 2 includes a matrix triangularization, preferably using said QR factorization algorithm described in the above-mentioned document, or using a sequence of Givens rotations. That is, the system in equation (15) is triangularized in a manner known as such, by applying a series of multiplications of rows by scalars followed by additions of the rows, i.e. it is converted to the form:

$$\begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,N+M-1} \\ 0 & r_{2,2} & r_{2,3} & \cdots & r_{2,N+M-1} \\ 0 & 0 & r_{3,3} & \cdots & r_{3,N+M-1} \\ 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & r_{N+M-1,N+M-1} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_{N+M-1} \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{N+M-1} \\ q_i \end{bmatrix} \quad (16)$$

Step 3 includes error evaluation. The change of the square prediction error for the segment caused by removal of i'th coefficient is simply equal to the term $q^2{}_i$ in equation (16). Again, this is valid based on Q in said QR factorization being orthonormal. In general depending on the properties of Q and the measure for prediction error, the change of the square prediction error is a function of q_i.

Step 4 includes removal of coefficients. If the change of the prediction error is acceptable according to a chosen criterion than the coefficient c_i is removed from the coefficient set. The new number of coefficients is now N+M-1. Matrix $R^1{}_k$ and vector $z^1{}_k$ are modified e.g. by means of QR factorization to form:

$$R^1{}_k = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,N+M-1} \\ & r_{2,2} & r_{2,3} & \cdots & r_{2,N+M-1} \\ & & r_{3,3} & \cdots & r_{3,N+M-1} \\ & & & \ddots & \vdots \\ & & & & r_{N+M-1,N+M-1} \end{bmatrix}, \quad z^1{}_k = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{N+M-1} \end{bmatrix} \quad (17)$$

The number of coefficients for the segment can be reduced further by repeating the steps 1-4 based on eqation (17).

In the motion coefficient removal block the following strategy for coefficient removal is used:

a. a threshold T is selected which corresponds to an allowed increase of square prediction error in the whole frame.

b. for all segments and their basis functions q_i^2 is calculated using equation (16).

c. a basis function of a segment with smallest q_i^2 is removed.

d. points b-c are repeated until the sum of all q_i^2's corresponding to all removed basis functions in different segments is greater than T.

Finally, step 5 includes coefficient calculation. After removal of suitable coefficients in this step the final motion coefficients for a segment S_k are calculated by solving the system of linear equations:

$$R^1{}_k C_k = z^1{}_k \qquad (18)$$

where matrix $R^1{}_k$ and vector $z^1{}_k$ are the result of the previous steps 1-4. The equation can be solved using one of well known algorithms, e.g. backsubstitution.

This preferred embodiment does not calculate coefficients c_i based on R^1_{ij} and z^1_{ij} but includes a further block not shown in the figures, for orthogonalizing the motion field model with respect to a rectangle circumscribing the given segment. This block receives as the input the partitioning of the current frame into segments and for every segment S_k matrices R^1_k and Z^1_k shown in (14) from the Segment Merging block. In this orthogonalization block the polynomial basis functions $f_i(\cdot)$ are replaced by orthogonal polynomials $\tilde{f}_i(x,y)$. Then the motion vector field of this segment can be represented as:

$$\Delta x(x,y) = \sum_{i=1}^{N} \tilde{c}_i \tilde{f}_i(x,y) \qquad (23)$$

$$\Delta y(x,y) = \sum_{i=N+1}^{N+M} \tilde{c}_i \tilde{f}_i(x,y)$$

Although the motion vector field in Eq 4 is fully equivalent to the one in Eq 23, the latter one is used because coefficients $\tilde{c}_i$ are less sensitive to quantization than c_i, hence can be represented with fewer bits.

Computation of the orthogonal polynomial basis functions is performed as follows both in the video encoder and in the video decoder, based on the shape of each segment and on the predefined basis functions f_i of the motion model.

Let the size of rectangle circumscribing the given segment be N_1 x N_2 pixels. At first two sequences of one dimensional polynomials, e.g Legendre polynomials, are computed:

1. $g_k(x)$, k=0,1,... orthogonal on the interval $[1,N_1]$,
2. $h_l(y)$, l=0,1,... orthogonal on the interval $[1,N_2]$,

The two dimensional (2-D) orthogonal polynomial basis functions $\tilde{f}_i(x,y)$, i=1,...,N+M are built as a tensor product of 1-D orthogonal polynomials:

$$\tilde{f}_i(x,y)=g_k(x)h_l(y) \qquad (24)$$

Details on the choice of polynomials can be taken from A. Akansu and R. Haddad, „Multiresolution Signal Decomposition", Academic Press Inc., USA, 1992, pages 55 to 56.

Orthogonalization with respect to the rectangle circumscribing the given segment performed in this way is computationally much less complex than other methods of orthgonalization, which are nevertheless applicable as alternatives, for example orthogonalization with respect to the shape of the segment.

In general, however, well known orthogonalization algorithms, e.g. the Gramm-Schmidt algorithm, can be used to convert ordinary polynomials to polynomials orthogonal in an arbitrary shaped segment area.

Orthogonal polynomial basis function $\tilde{f}_i(\cdot)$ are chosen so that they can be represented as a linear combination of polynomial basis functions $f_k(\cdot)$, k=1,2,...,i, i.e.,

$$\tilde{f}_i(x, y) = \sum_{k=1}^{i} t_{k,i} f_k(x, y), \; i = 1, \; 2, \ldots, N + M \quad (25)$$

This assumption guarantees that conversion from nonorthogonal to orthogonal basis functions can be implemented with a low computational complexity by simple matrix multiplications.

Matrices R_k^1 and Z_k^1 describing the motion vector field of the segment need to be recomputed to reflect the change of the basis functions from $f_i(\cdot)$ to their orthogonal version $\tilde{f}_i(\cdot)$. New matrices $\tilde{R}_k^1$ and $\tilde{Z}_k^1$ corresponding to orthogonal polynomial basis functions $\tilde{f}_i(\cdot)$ which satisfy equation (25) can be computed using matrix R_k^1 and vector Z_k^1 according to following formula:

$$\tilde{R}^1{}_k = R^1{}_k T \quad (26)$$

$$\tilde{Z}^1{}_k = Z^1{}_k \quad (27)$$

Matrix T is given by

$$T = \begin{bmatrix} t_{1,1} & t_{1,2} & t_{1,3} & \cdots & t_{1,N+M} \\ 0 & t_{2,2} & t_{2,3} & \cdots & t_{2,N+M} \\ 0 & 0 & t_{3,3} & \cdots & t_{3,N+M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & t_{N+M,N+M} \end{bmatrix} \quad (28)$$

where elements $t_{k,i}$ are taken from the equation (25). The new motion vector coefficients $\tilde{c}_k=(\tilde{c}_1,\tilde{c}_2,\ldots,\tilde{c}_{N+M})$ for the segment S_k corresponding to orthogonal polynomial basis functions $\tilde{f}_i(\cdot)$, i=1,...,N+M, can be calculated using coefficients $c_k=(c_1,\ldots,c_{N+M})$ corresponding to polynomial basis functions $f_i(\cdot)$, i=1,...,N+M,

$$\tilde{C}_k = T^{-1}C_k \qquad (29)$$

or by solving the system of equations:

$$\tilde{R}^1{}_k\tilde{C}_k = \tilde{z}^1{}_k \qquad (30)$$

The coefficient removal block 33 receives as the input new division of the current frame into segments and for every segment S_k it receives matrices $\tilde{R}_k^1$ and $\tilde{z}_k^1$ produced by the Orthogonalization block. The motion vector field of every segment is represented by N+M motion field model basis functions.

Motion Coefficient Removal block 33 determines for a given segment S_k if it is possible to simplify the motion field model, as described above.

Figure 7 shows an embodiment of motion compensated prediction block 1 of Fig. 1. This block receives motion information output by motion field coding block 3 and furthermore receives a reference frame $\tilde{I}_{n-1}(x,y)$. Based on this information, block 1 outputs a predicted frame $\hat{I}_n(x,y)$. As shown in Figure 7, motion compensated prediction bock 1 comprises a demultiplexer 11 receiving multiplexed motion

information from motion field encoding block 3 and outputting demultiplexed motion information components, i.e. image partitioning information, coefficient selection information and, finally, the value of the transmitted motion coefficients. Reference numeral 12 denotes an image partitioning block receiving said image partitioning information and said reference frame $\tilde{I}_{n-1}$ and outputting segments of the frame resulting from partitioning the image according to the image partitioning information. Reference numeral 13 denotes a basis functions building block. This block selects from a predefined set of basis functions the particular basis functions indicated in the selection information generated by the motion coefficient removal block 33 in motion field encoding block 3. Reference numeral 14 denotes a segment prediction block which receives for each segment of said reference frame $\tilde{I}_{n-1}$ the associated selection of basis functions and the associated motion coefficients, calculates the motion vectors $[\tilde{\Delta}x, \tilde{\Delta}y]$ and based on these motion vectors, calculates the predicted frame $\hat{I}_n(x,y)$ for each pixel (x,y) of each segment. Motion compensated prediction block 1 corresponds in its structure and function to motion compensated prediction block 21 of the video decoder depicted in figure 2. Both motion compensated prediction blocks base the prediction on the motion information output by motion field coding block 3 of the video encoder shown in figure 1.

As a result of all of the steps in all of the blocks, the motion field encoder according to the invention produces merging information for informing the decoder which segments are merged, selection information informing the decoder which basis functions are removed and motion coefficient information.

The main advantage of the present invention over prior art solutions is its ability to reduce the amount of motion information by a large factor without causing a large increase of prediction error. Additionally the complexity of the overall system is low which allows practical implementation on available signal processors or general purpose microprocessors.

The segment merging block has a unique ability of finding motion vectors of combined segments from given motion vectors estimated for separate segments. It can be proven that the motion vectors it produces are in fact optimal in terms of keeping minimal square error for the combined segment. This explains the ability of this block to dramatically reduce the number of segments with only very modest increase of square prediction error.

The motion coefficient removal block is a very powerful tool for instantaneous adaptation of the motion model to the actual amount and type of motion in the video scene. This block can easily test the result of prediction (value of square prediction error for a segment) with a very large number of models, e.g., with all possible combinations of motion field basis functions. A strong advantage of this scheme is that it does not need to repeat the process of motion estimation and hence is computationally simple.

By using motion estimation followed by QR motion analyzation the motion field coder can find new motion coefficients for any desired combination of image segments or any desired model of the motion field in the segment by solving very simple systems of linear equations.

The system can be implemented in a variety of ways without departing from the spirit and the scope of the invention. For instance, different linear motion models can be used in equation (3). Different methods can be used to linearize the term in the formula (5). Further, different criteria may be used to decide whether to merge of not to merge two segments. The strategy for deciding whether a given basis function should be removed from the model may vary. Triangularization of matrices in equations (10) and (15) can be performed using various algorithms and calculation of final coefficients by solving equation (18) can be done using a number of known algorithms for solving systems of linear equations. Finally, different interpolation methods can be used for values of $\tilde{I}_{n-1}(x,y)$, $G_x(x,y)$ and $G_y(x,y)$ in noninteger coordinates.

Claims

1. Encoder for performing motion compensated encoding of video data, comprising:

- means (2) for estimating a motion vector field (ΔX, ΔY) describing scene motion displacements of video frame pixels;

- means (3) for obtaining motion coefficients (c_i) representing said motion vector field;

- means (1) for predicting a video data frame based on a first video data frame and said motion vector field;

- means (6) for calculating a prediction error frame based on said predicted video data frame and on a second video data frame input;

- means (4, 8) for transmitting said motion coefficients and said prediction error frame to a decoder;

- said means (3) for obtaining motion coefficients comprising

- means for partitioning each frame into segments (S_i) ;

- means (41, 42, 45), for calculating and storing for each segment a distortion matrix E and a distortion vector such that a predefined measure ΔE for distortion in each segment due to approximating said motion vector field as coefficients (c_i) of a set of basis functions f_i is a function of (E c - y), c being a vector of said motion coefficients c_i;

- means (31) for decomposing distortion matrix E into a first matrix Q and a second matrix R such that

$$\det Q \neq 0 \text{ and}$$

$$Q\,R = E$$

a subset of the set of all columns of Q being a basis of a vector space defined by all possible linear combinations of all column vectors of E, columns of Q being orthogonal to each other;

- means (46) for calculating an auxiliary vector z according to $z = Q^{-1}\,y$, Q^{-1} being the inverse matrix of Q;

- means (44, 47) for generating for each segment a column extended matrix A comprising the columns of R and z as an additional column, and for selecting all rows of matrix A which have elements unequal to zero in all columns due to R;

- means (32) for generating a row extended matrix B consisting of said selected rows of matrix A of one segment (S_i) and of said selected rows of matrix A of another segment (S_j);

- means (32) for performing a series of multiplications of rows of matrix B with scalars unequal zero and additions of rows of matrix B in order to obtain a modified matrix B' having in the columns due to R as many rows as possible filled with zeros.

2. Encoder according to claim 1, characterized by

- means (32) for calculating an error value as a predetermined function of elements in matrix B'; and

- means (32) for instructing said means for partitioning the frame into segments (S_i), to merge said one segment (S_i) with said another segment (S_j) based on said error value.

3. Encoder according to claim 2, characterized by

- said means for calculating an error value being adapted to calculate a merging error as a function of elements of the column due to z of said modified matrix B' of all rows filled with zeros in the columns due to R.

4. Encoder according to any of claims 1 to 3, characterized by

- means (33) for calculating and outputting for each segment said motion coefficients c_i by solving the system of linear equations D c = d;

- D being a matrix obtainable by removing all non-selected rows of matrix A and then removing the column due to z, and d being the column of A due to z after removal of said non-selected rows, if the segment is a non merged segment;

- D being a matrix obtainable by removing from matrix B' all rows having zeros in the columns due to R and then removing said column due to z and d being the column of B' due to z after removal of said rows having zeros in the columns due to R, if said segment is a merged segment.

5. Encoder according to any of the preceding claims, characterized in that

- said means (31) for decomposing E is adapted to decompose E such that said matrix Q is an orthonormal matrix the column vectors q of which fulfill $q_i\ q_k = 1$ for i=k and 0 for i≠k; and

- said means (46) for calculating said auxiliary vector z is adapted to calculate Q^{-1} by exchanging column vectors and row vectors of matrix Q.

6. Encoder according to claim 5, characterized in that

   - said means (31, 43) for decomposing E is adapted to generate matrix R as a triangular matrix.

7. Encoder according to one of the preceding claims, characterized in that

   - said means (32) for calculating an error value is adapted to calculate said error value as the sum of squared elements of the column due to z of said modified matrix B' of all rows filled with zero in the columns due to R.

8. Encoder according to one of the preceding claims, characterized in that

   - said means (42) for calculating and storing said distortion matrix E comprises a gradient filter (41) and is adapted to calculate the elements e_{jk} of the j-th row and k-th column of E based on linear components of a Taylor series expansion of the prediction frame and based on said basis functions f_i.

9. Encoder according to claim 4, characterized by

   - means for orthogonalizing said motion coefficients by performing a coordinate transformation on said motion coefficients corresponding to a transformation of said set of basis functions f_i

into a set of basis functions being orthogonal to each other with respect to an area determined by the shape of the segment; and

- means for quantizing and coding the obtained orthogonalized motion coefficients.

10. Encoder for performing motion compensated encoding of video data, comprising:

- means (2) for estimating a motion vector field (ΔX, ΔY) describing scene motion displacements of video frame pixels;

- means (3) for obtaining motion coefficients (c_i) representing said motion vector field;

- means (1) for predicting a video data frame based on a first video data frame and said motion vector field;

- means (6) for calculating a prediction error frame based on said predicted video data frame and on a second video data frame input;

- means (4, 8) for transmitting said motion coefficients and said prediction error frame to a decoder;

- said means (3) for obtaining motion coefficients comprising

- means (31) for partitioning each frame into segments;

- means (43) for calculating and storing for each segment a distortion matrix E and a distortion vector y such that a predefined measure ΔE for distortion in each segment due to representing said motion vector field as coefficients (c_i) of a set of basis functions f_i is a function of (E c - y), c being a column vector of said motion coefficients c_i ;

- means for calculating from distortion matrix E a first matrix Q and a second matrix R such that

$$\det Q \neq 0 \text{ and}$$

$$Q R = E$$

a subset of the set of all columns of Q being a basis of a vector space defined by all possible linear combinations of all column vectors of E, all other columns of Q being orthogonal to each other;

- means (46) for calculating an auxiliary vector z according to $z = Q^{-1} y$, Q^{-1} being the inverse matrix of Q;

- means (44, 47) for generating for each segment a column extended matrix A comprising the columns of R and z as an additional column, and for obtaining a matrix B comprising all rows of matrix A which have elements unequal to zero in all columns due to R;

- means (33) for removing from matrix B the j-th column due to R corresponding to the j-th coefficient of said coefficients (c_i) and performing a series of multiplications of rows of the resulting matrix with scalars unequal to zero and additions of rows of said resulting matrix, to obtain a matrix C having a row filled with zeros in the columns due to R.

11. Encoder according to claim 10, characterized by

- means (33) for calculating an error value as a predetermined function of at least one element in matrix C; and

- means (33) for determining based on said error value whether said j-th coefficient is to be transmitted to said decoder.

12. Encoder according to claim 11, characterized by

- said means (33) for calculating an error value being adapted to calculate said error value as a predetermined function of the element in matrix located in said row filled with zeros in the columns due to R and located in the column due to z.

13. Encoder according to any of claims 10 to 12, characterized in that

- said means (33) for obtaining matrix C is adapted to perform said multiplications and additions such that matrix C is triangular.

14. Encoder according to any of claims 10 to 13, characterized by

- means for calculating and outputting for each segment said motion coefficients c_i to be transmitted to said decoder by solving the system of linear equations D c = d;

- D being a matrix obtainable by removing from matrix C said row filled with zeros in the columns due to R and then removing the column due to z and d being the column of C due to z after removal of said row.

15. Encoder according to any of claims 10 to 14, characterized in that

- said means (43) for decomposing E is adapted to decompose E such that said matrix Q is an orthonormal matrix the column vectors q of which fulfill $q_i\ q_k = 1$ for i=k and 0 for $i \neq k$; and

- said means (46) for calculating said auxiliary vector z is adapted to calculate Q^{-1} by exchanging column vectors and row vectors of matrix Q.

16. Encoder according to claim 15, characterized in that

- said means (43) for decomposing E is adapted to generate matrix R as a triangular matrix.

17. Encoder according to any of claims 10 to 16, characterized in that

- said means (31) for calculating and storing said prediction error matrix E comprises a gradient filter (41) and is adapted to calculate the elements e_{jk} of the j-th row and k-th column of E based on linear components of a Taylor series expansion of the prediction frame and based on said basis functions f_i.

18. Encoder according to one of the preceding claims, characterized by

- a prediction error decoder (5) for decoding said prediction error frame (E); and

- means (7) for calculating said first video data frame based on said predicted video data frame and said decoded prediction error frame.

19. System for transmitting or storing video data, comprising an encoder according to any of claims 1 to 18 and comprising a decoder for performing decoding of motion compensation encoded video data, said decoder comprising:

- means for storing a video data frame;

- means for predicting a video data frame based on said video data frame and on said motion coefficients received from the encoder;

- means for decoding prediction error data received from said encoder and obtaining a prediction error frame; and

- means for calculating and outputting an updated video data frame based on said predicted video data frame and said decoded prediction error frame, and storing the updated video data frame in said storing means.

20. Method of motion compensated encoding of video data, comprising the steps:

- estimating a motion vector field describing scene motion displacements of video frame pixels;

- determining motion coefficients (c_i) representing said motion vector field;

- predicting a video data frame based on a first video data frame and said motion vector field;

- calculating a prediction error frame based on said predicted video data frame and on a second video data frame input;

- transmitting said motion coefficients and said prediction error frame to a decoder;

- said step of calculating motion coefficients comprising the steps

- partitioning each frame into segments;

- calculating and storing for each segment a distortion matrix E and a distortion vector y such that a predefined measure ΔE for distortion in each segment due to approximating said motion vector field as coefficients (c_i) of a set of basis functions f_i is a function of (E c - y), c being a column vector of said motion coefficients c_i ;

- decomposing distortion matrix E into a first matrix Q and a second matrix R such that

$$\det Q \neq 0 \text{ and}$$

$$Q\,R = E$$

a subset of the set of all columns of Q being a basis of a vector space defined by all possible linear combinations of all column vectors of E, all other columns of Q being orthogonal to said vector space;

- calculating an auxiliary vector z according to $z = Q^{-1} y$, Q^{-1} being the inverse matrix of Q;

- generating for each segment a column extended matrix A comprising the columns of R and z as an additional column, and for selecting all rows of

matrix A which have elements unequal to zero in all columns due to R;

- generating a row extended matrix B consisting of said selected rows of matrix A of one segment and of said selected rows of matrix A of an adjacent segment; and

- performing a series of multiplications of rows of matrix B with scalars unequal zero and additions of rows of matrix B in order to obtain a modified matrix B' having in the columns due to R as many rows as possible filled with zeros.

21. Method according to claim 20, characterized by

- calculating an error value as a predetermined function of elements in matrix B'; and

- instructing said means for partitioning the frame into segments, whether to merge said one segment with said adjacent segment based on said error value.

22. Method according to claim 21, characterized by

- calculating as said error value a merging error as a function of elements of the column due to z of said modified matrix B' of all rows filled with zero in the columns due to R.

23. Method of performing motion compensated encoding of video data, comprising:

- estimating a motion vector field describing scene motion displacements of video frame pixels;

- determining motion coefficients (c_i) representing said motion vector field;

- predicting a video data frame based on a first video data frame and said motion vector field;

- calculating a prediction error frame based on said predicted video data frame (P) and on a second video data frame input;

- transmitting said motion coefficients and said prediction error frame to a decoder;

- said step of calculating motion coefficients comprising

- partitioning each frame into segments;

- calculating and storing for each segment a prediction matrix E and a distortion vector y such that a predefined measure ΔE for distortion in each segment due to representing said motion vector field as coefficients (c_i) of a set of basis functions f_i is a function of (E c - y), c being a column vector of said motion coefficients c_i ;

- calculating from distortion matrix E a first matrix Q and a second matrix R such that

$$\det Q \neq 0 \text{ and}$$

$$Q\,R = E$$

a subset of the set of all columns of Q being a basis of a vector space defined by all possible linear combinations of all column vectors of E, all other columns of Q being orthogonal to said vector space;

- calculating an auxiliary vector z according to $z = Q^{-1}\,y$, Q^{-1} being the inverse matrix of Q;

- generating for each segment a column extended matrix A comprising the columns of R and z as an additional column, and for obtaining a matrix B comprising all rows of matrix A which have elements unequal to zero in all columns due to R;

- removing from matrix B the j-th column due to R corresponding to the j-th coefficient of said coefficients (c_j) and performing a series of multiplications of rows of the resulting matrix with scalars unequal to zero and additions of rows of said resulting matrix, to obtain a matrix C having a row filled with zeros in the columns due to R.

24. Method according to claim 23, characterized by

- calculating an error value as a predetermined function of at least one element in matrix C; and

- determining based on said error value whether said j-th coefficient is to be transmitted to said decoder.

25. Method according to claim 24, characterized by

- calculating said error value as a predetermined function of the element in matrix C located in said row filled with zeros in the columns due to R and located in the column due to z.

26. A video encoder comprising:
a motion field encoder (3) responsive to a current frame (I_n), a reference frame ($\tilde{I}_{n-1}$) and a motion vector field for producing a coded motion vector field and for minimizing the number of the motion coefficients, the motion field encoder (3) comprising:

- a first main block (31) including linearization means (41,42,45) for linearizing a squared error function between a segment (S_i) of the current video frame (I_n) and a segment of a prediction frame, and matrix processing means (43,44,46,47) for forming a first matrix R_i^1 and a first vector z^1_i, said matrix and vector being a new

representation of the motion vector field of the segment;

- a second main block (32) connected to the first main block (31), the second main block (32) including merging means for merging pairs of segments (S_i, S_j) if the combined area can be predicted using a common motion field, wherein, for each merged segment, the block (32) outputs a first merged matrix R^1_{ij} and a second merged vector z^1_{ij};

- a third main block (33) connected to the second main block (32), the third main block including coefficient removal means for removing motion coefficients (c_i) of the segment until the change of the squared prediction error is not acceptable;

- wherein in the outputted coded motion vector field each motion vector is based on a linear motion model comprising basis functions and motion coefficients.

27. A video encoder as in claim 26, wherein linearization means transform the square prediction error to a form according to the formula:

$$\sum_{j=1}^{P} \left(e_{j,1}c_1 + e_{j,2}c_2 + \ldots + e_{j,N+M}c_{N+M} - y_j \right)^2$$

where c_i is a motion coefficient and N+M is the number of available motion field model basis functions.

28. A video encoder as in claim 26,
wherein the matrix processing means include means for constructing following matrices:

$$E_i = \begin{bmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,N+M} \\ e_{2,1} & e_{2,2} & \cdots & e_{2,N+M} \\ \vdots & \vdots & \vdots & \vdots \\ e_{P,1} & e_{P,2} & \cdots & e_{P,N+M} \end{bmatrix}, y_i = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_P \end{bmatrix}$$

N+M being the number of available basis functions of the motion field model, P being the number of pixels of segment i.

29. A video encoder as in claim 28,
wherein the matrix processing means further include QR factorization means for decomposing the matrix E_i into a product of a matrix Q_i and a matrix R_i and for forming an auxiliary vector z_i of the form $z_i=Q^T{}_i y_i$ and for outputting a first matrix $R_i{}^1$ consisting N+M first rows of matrix R_i and the first vector $z^1{}_i$ consists of first N+M elements of z_i.

30. A video encoder as in claim 26,
wherein the merging means includes means for triangulization the linear equation

$$\begin{bmatrix} R^1{}_i \\ R^1{}_j \end{bmatrix} C_{ij} = \begin{bmatrix} z^1{}_i \\ z^1{}_j \end{bmatrix}$$

wherein R^1_i, z^1_i and R^1_j, z^1_j are produced by the first main block segments S_i and S_j respectively, to the form:

$$\begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,N+M} \\ 0 & r_{2,2} & r_{2,3} & \cdots & r_{2,N+M} \\ 0 & 0 & r_{3,3} & \cdots & r_{3,N+M} \\ 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & r_{N+M,N+M} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_{N+M} \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{N+M} \\ q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_{N+M} \end{bmatrix}$$

31. A video encoder as in claim 30,
wherein the merging means further includes means for calculating the change of the square prediction error ΔE_{ij} in the area S_{ij} caused by merging of segments S_i and S_j in accordance with the following formula:

$$\Delta E_{ij} = \sum_{k=1}^{N+M} q^2{}_k$$

wherein the segments are merged if the change of the square prediction error is acceptable according to a given criterion.

32. A video encoder as in claim 30,

- wherein the merging means further includes means for building for the resulting merged segment S_{ij} a matrix R^1_{ij} and a vector z^1_{ij} by taking first N+M rows of the triangularized matrix equation, wherein the matrix R^1_{ij} and vector are described by the following formula:

$$R^1_{ij} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,N+M} \\ & r_{2,2} & r_{2,3} & \cdots & r_{2,N+M} \\ & & r_{3,3} & \cdots & r_{3,N+M} \\ & & & \ddots & \vdots \\ & & & & r_{N+M,N+M} \end{bmatrix}, z^1_{ij} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{N+M} \end{bmatrix}$$

said matrix and vector being outputted from the second main block.

33. A video encoder as in claim 26,
comprising means for sending information enabling a reconstruction of merged partitions of the frame to a decoder.

34. A video encoder as in claim 26,
wherein the coefficient removal means includes means for reducing the matrix equation $R^1_k c_k = z^1_k$ by removing i'th

column from a matrix R^1_k and i'th element from the vector c_k.

35. A video encoder as in claim 34,
wherein the coefficient removal means further includes means for triangulizing the reduced matrix equation to the form:

$$\begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & \cdots & r_{1,N+M-1} \\ 0 & r_{2,2} & r_{2,3} & \cdots & r_{2,N+M-1} \\ 0 & 0 & r_{3,3} & \cdots & r_{3,N+M-1} \\ 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & r_{N+M-1,N+M-1} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_{N+M-1} \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{N+M-1} \\ q_i \end{bmatrix}$$

and means for calculating the change of the square prediction error caused by removal of the i'th column and i'th element as a function of q^2_i.

36. A video encoder in claim 35,
wherein the coefficient removal means further includes means for comparing the change of the prediction error to a chosen criterion.

37. A video encoder as in claim 36,
wherein the coefficient removal means further includes means for removing the motion coefficient c_i from the coefficient matrix if said change of prediction error is acceptable, and means for forming a modified matrix R^1_k and a modified vector z^1_k by removing the bottom row from the triangularized matrix, said modified matrix and

modified vector being applied to the coefficient removal means.

38. A video encoder as in claim 36, wherein the coefficient removal means further includes means for calculating of final motion coefficients for the segment by solving the system of linear equations $R^1{}_k c_k = z^1{}_k$, if the further change of prediction error is not acceptable.

39. Decoder for decoding of motion compensation encoded video data, comprising:

- means for storing a video data frame;
- means for predicting a video data frame based on said stored video data frame and on received motion information;
- means for decoding received prediction error data and obtaining a prediction error frame; and
- means for calculating and outputting an updated video data frame based on said predicted video data frame and said decoded prediction error frame, and storing the updated video data frame in said storing means;
- said means for predicting a video data frame comprising
- means (11) for demultiplexing received motion data into at least two of the following: data concerning

the partitioning of said updated video data frame into segments (S_i), data concerning a selection of basis functions from a set of motion field model basis functions, and data concerning coefficients of selected basis functions;

- means (13) for reconstructing said motion vector field in each segment (S_i) from a linear combination of said selected basis functions and coefficients; and

- means for calculating said prediction frame based on said reconstructed motion vector field and based on said stored video data frame.

40. Decoder according to claim 39, characterized in that

- said means (13) for reconstructing said motion vector field is adapted to receive data regarding the structure of each member of said set of basis functions for each segment (S_i) of said updated video data frame.

41. Decoder according to claim 39 or 40, characterized in that

- said means (13) for reconstructing said motion vector field is adapted to calculate said set of motion field model basis functions for each segment (S_i) as being orthogonal to each other with respect to an area determined by the shape of the segment (S_i).

42. System for transmitting video data, comprising an encoder according to any of claims 1 to 18 and 26 to 38 and comprising a decoder according to any of claims 39 to 41.

43. System for storing video data, comprising an encoder according to any of claims 1 to 18 and 26 to 38 and comprising a decoder according to any of claims 39 to 41.

Fig.1

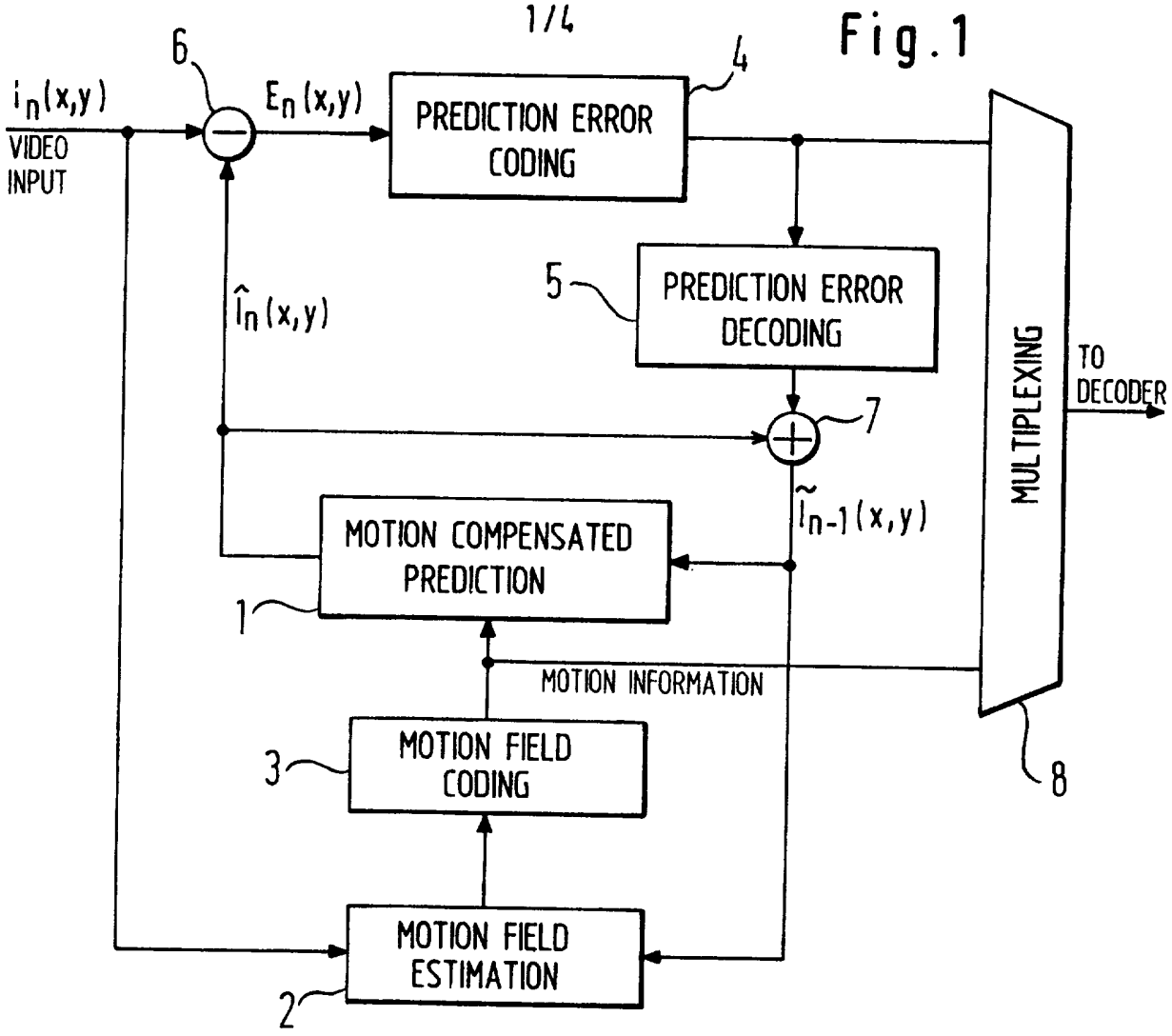

6
$i_n(x,y)$
VIDEO INPUT
$E_n(x,y)$
PREDICTION ERROR CODING
4
$\hat{I}_n(x,y)$
5
PREDICTION ERROR DECODING
MULTIPLEXING
TO DECODER
7
$\tilde{I}_{n-1}(x,y)$
MOTION COMPENSATED PREDICTION
1
MOTION INFORMATION
3
MOTION FIELD CODING
8
MOTION FIELD ESTIMATION
2

Fig.2

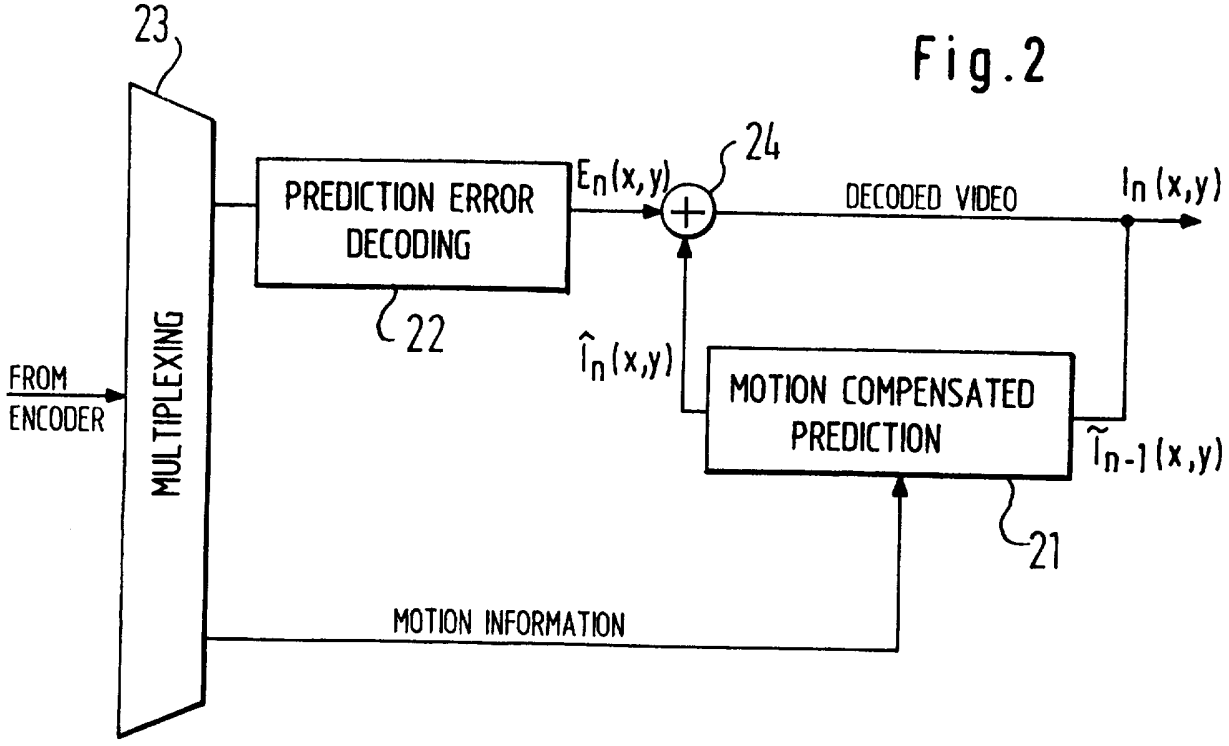

23
FROM ENCODER
MULTIPLEXING
PREDICTION ERROR DECODING
22
$E_n(x,y)$
24
DECODED VIDEO
$I_n(x,y)$
$\hat{I}_n(x,y)$
MOTION COMPENSATED PREDICTION
$\tilde{I}_{n-1}(x,y)$
21
MOTION INFORMATION

Fig.3

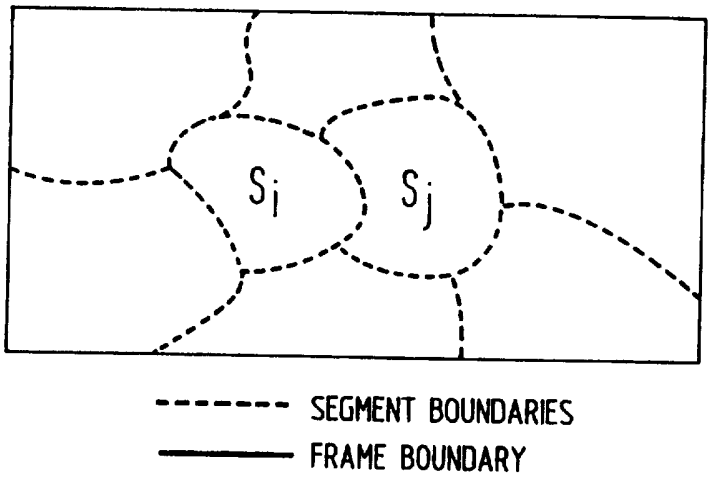

S_i
S_j
SEGMENT BOUNDARIES
FRAME BOUNDARY

Fig.4

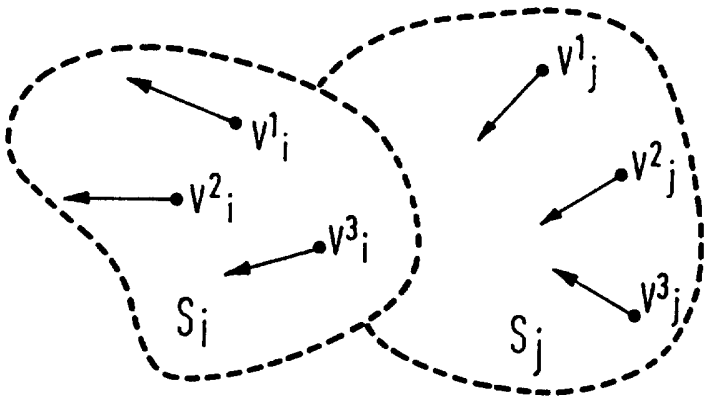

V^1_i
V^2_i
V^3_i
S_i
V^1_j
V^2_j
V^3_j
S_j

Fig.5

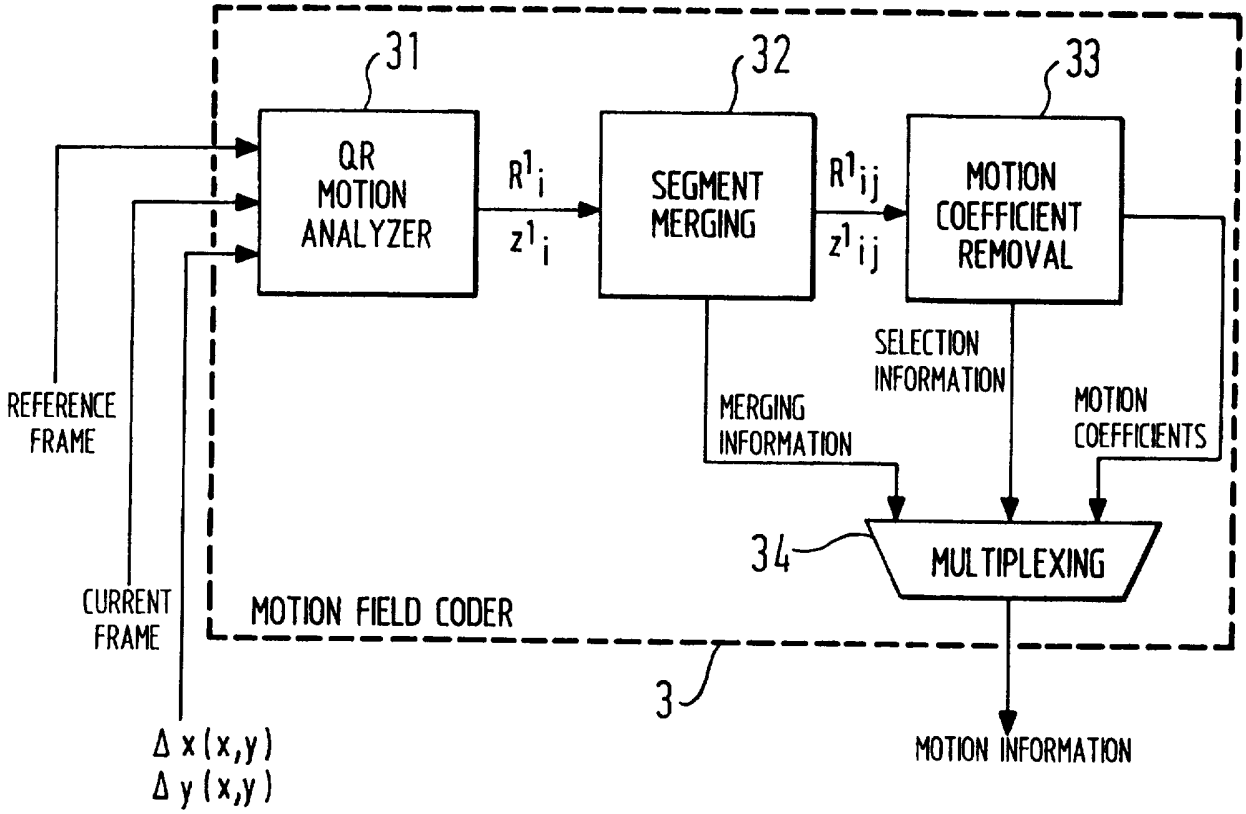

31
32
33
QR MOTION ANALYZER
SEGMENT MERGING
MOTION COEFFICIENT REMOVAL
R^1_i
z^1_i
R^1_{ij}
z^1_{ij}
REFERENCE FRAME
CURRENT FRAME
$\Delta x(x,y)$
$\Delta y(x,y)$
MERGING INFORMATION
SELECTION INFORMATION
MOTION COEFFICIENTS
34
MULTIPLEXING
MOTION FIELD CODER
3
MOTION INFORMATION

Fig.6

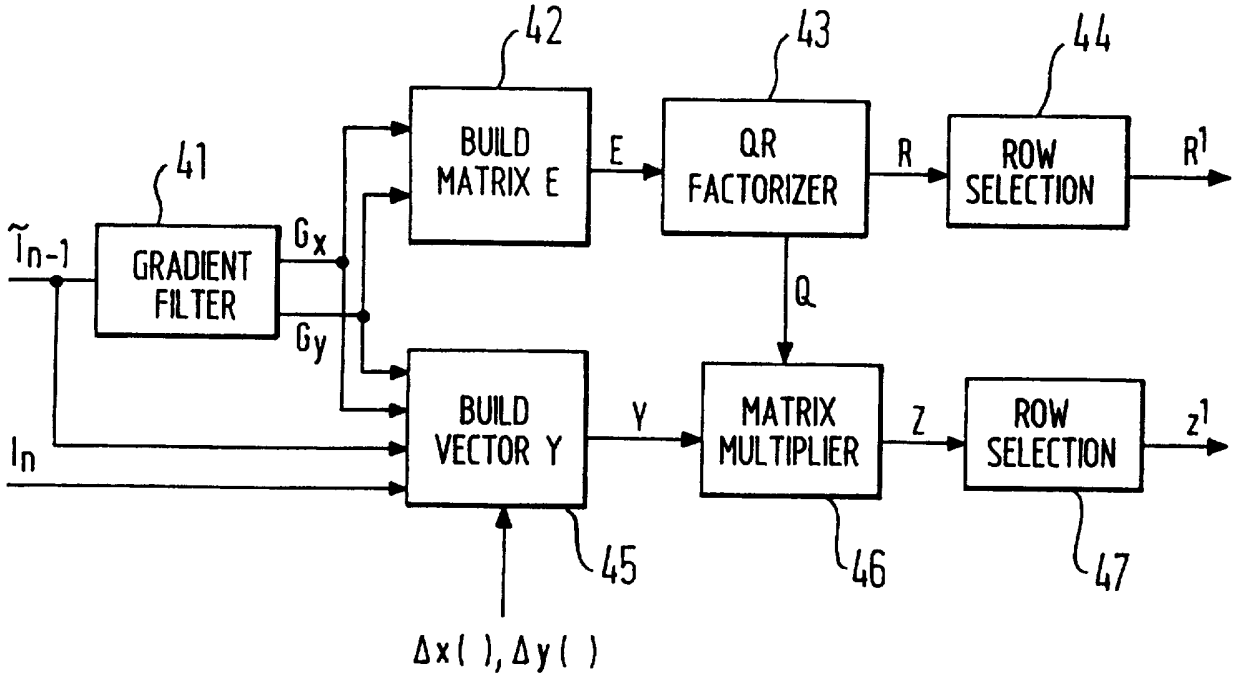
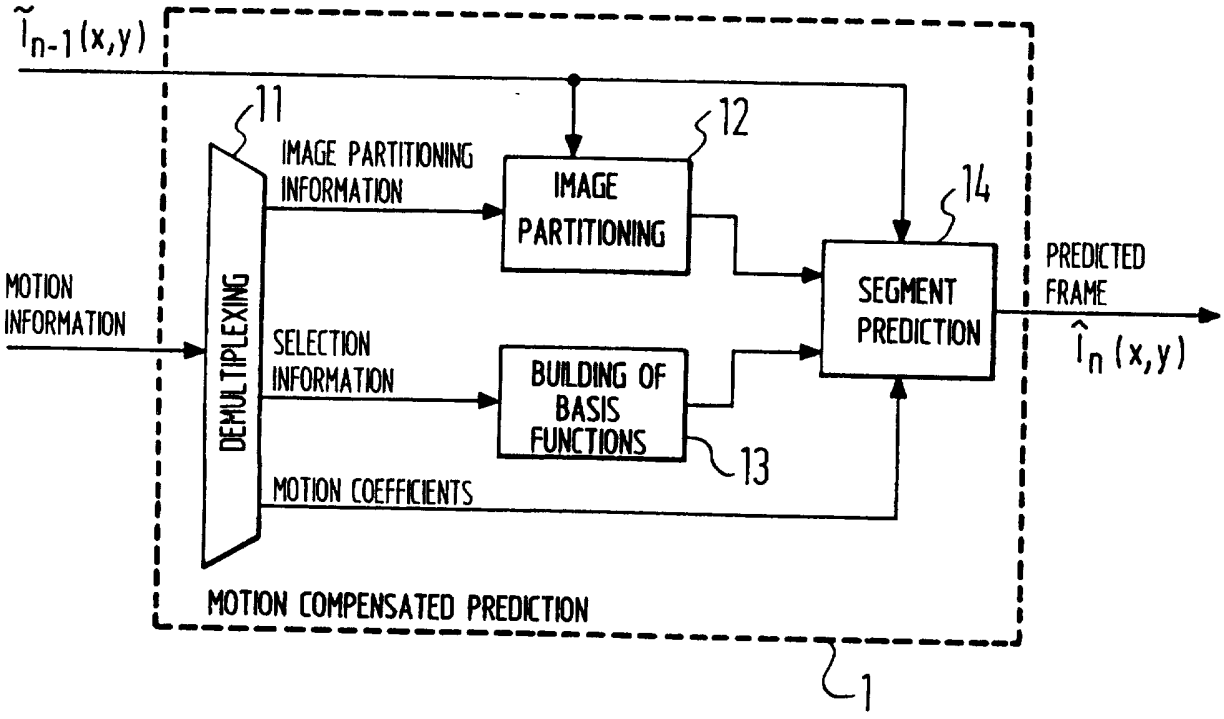

42
43
44
41
BUILD MATRIX E
E
QR FACTORIZER
R
ROW SELECTION
R^1
$\tilde{I}_{n-1}$
GRADIENT FILTER
G_x
G_y
Q
BUILD VECTOR Y
Y
MATRIX MULTIPLIER
Z
ROW SELECTION
z^1
I_n
45
46
47
$\Delta x(\ ), \Delta y(\ )$

Fig. 7


$\tilde{I}_{n-1}(x,y)$
11
IMAGE PARTITIONING INFORMATION
12
IMAGE PARTITIONING
14
MOTION INFORMATION
DEMULTIPLEXING
SELECTION INFORMATION
BUILDING OF BASIS FUNCTIONS
13
SEGMENT PREDICTION
PREDICTED FRAME
$\hat{I}_n(x,y)$
MOTION COEFFICIENTS
MOTION COMPENSATED PREDICTION
1

INTERNATIONAL SEARCH REPORT

Int onal Application No
PCT/EP 96/01644

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04N7/36 H04N7/26 G06T7/20

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 6 H04N G06T

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ICIP '95, vol. 2, 23 October 1995, LOS ALAMITOS, CA, USA, pages 197-200, XP000623945 KARCZEWICZ ET AL.: "Motion Estimation and Representation for Arbitrarily Shaped Image Regions" see the whole document ---	1-43
A	EP 0 625 853 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 23 November 1994 see the whole document --- -/--	1-43

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 March 1997	27.03.97

Name and mailing address of the ISA	Authorized officer
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	Foglia, P

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int. ional Application No
PCT/EP 96/01644

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DIGITAL VIDEO COMPRESSION: ALGORITHMS AND TECHNOLOGIES 1996, vol. 2668, 31 January 1996, SAN JOSE, CA, USA, pages 345-354, XP000612400 BAKER ET AL.: "Segmentation-Based Coding of Motion Fields for Video Compression" see the whole document	1-43
A	IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS, vol. E77-D, no. 6, June 1994, TOKYO JP, pages 694-704, XP000468524 CHEN ET AL.: "Segmentation Based on Accumulative Observation of Apparent Motion in Long Image Sequences" see the whole document	1-43
A	ICASSP-92, vol. 3, 23 March 1992, SAN FRANCISCO, CA, USA, pages 525-528, XP000378984 LIU ET AL.: "Segmentation-Based Coding of Motion Difference and Motion Field Images for Low Bit-Rate Video Compression" see the whole document	1-43
A	US 5 295 201 A (YOKOHAMA) 15 March 1994	

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. ional Application No
PCT/EP 96/01644

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 625853 A	23-11-94	JP 7046595 A	14-02-95
		US 5598215 A	28-01-97
US 5295201 A	15-03-94	JP 5199510 A	06-08-93
		CA 2087802 A	22-07-93