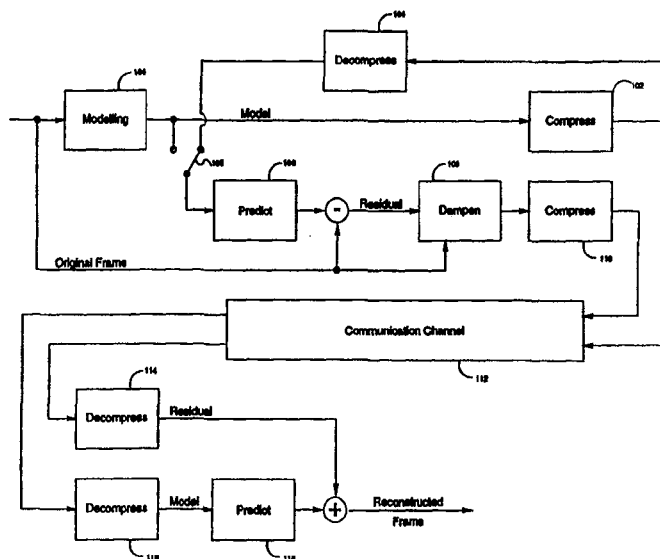




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : H04N 7/26	A1	(11) International Publication Number: WO 99/07154 (43) International Publication Date: 11 February 1999 (11.02.99)
(21) International Application Number: PCT/EP98/02949 (22) International Filing Date: 20 May 1998 (20.05.98) (30) Priority Data: 97112973.9 28 July 1997 (28.07.97) EP (34) Countries for which the regional or international application was filed: AT et al. (71) Applicant (for all designated States except US): IDT INTERNATIONAL DIGITAL TECHNOLOGIES DEUTSCHLAND GMBH [DE/DE]; Fraunhoferstrasse 9, D-85737 Ismaning (DE). (72) Inventors; and (75) Inventors/Applicants (for US only): MARTENS, Harald [NO/DK]; Skt. Peders Straede 25, DK-1453 Copenhagen (DK). MÜLLER, Klaus [DE/DE]; Oefelestrasse 3, D-81543 München (DE). REBERG, Jan, Otto [NO/NO]; Haugbro Terrasse 4, N-1405 Langhus (NO). RÖTTGERMANN, Clemens [DE/DE]; Thalkirchner Strasse 5, D-80337 München (DE). (74) Agent: BETTEN & RESCH; Reichenbachstrasse 19, D-80469 München (DE).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, 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: METHOD AND APPARATUS FOR COMPRESSION OF VIDEO IMAGES AND IMAGE RESIDUALS



(57) Abstract

A method of processing one or more video images to be compressed, each video image comprising pixels, and said method of processing comprising the step of: filtering one or more of said video images so as to improve compressibility, where said filtering is performed with the constraint such that the deviation between the filtered image and the non-filtered image lies within certain bounds, wherein said one or more video images to be filtered comprises one or more of original video images, residual images which are derived from one or more of said original images as the differences between said original images and predictions representing said original images using a model.

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	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHOD AND APPARATUS FOR COMPRESSION OF VIDEO IMAGES AND IMAGE RESIDUALS

RELATED APPLICATIONS

The application is related to the following applications assigned to the same applicant as the present invention and filed on even date herewith, the disclosures of which are hereby incorporated by reference:

Method and apparatus for compressing video sequences (Our file: IDT 018 WO)

Method and apparatus for Motion Estimation (Our file: IDT 020 WO).

FIELD OF INVENTION

This invention is related to compression of video frames, respectively video images, especially with residual character, so as to allow more efficient transmission and storage of the residuals.

Several video compression technologies are based on providing a prediction of a frame, based on some model involving a reference image. Examples include hybrid encoders with block based motion compensation, wire frame modelling and IDLE modelling, which is for example described in WO 95/08240. The final representation of the frame is then achieved by adding a residual or corrector image to the prediction.

OBJECTS OF THE INVENTION

It is an object of this invention to improve compression ratios for video images, especially residual images in video coding, under consideration of masking effects for the Human Visual System.

It is a further object of this invention to allow better image quality at a given compression ratio.

It is yet a further object of this invention to allow bit rate control for a compression system.

SUMMARY OF THE INVENTION

The invention is based on exploiting masking effects of the Human Visual System. Based on features of the original input frames or the reference image, bounds for the

residual are computed, so that a wanted image quality is assured as long as the residual stays within the bounds. The residual is then processed so as to minimise the number of bits necessary to transmit the residual, without violating the bounds. The processing includes several types of filtering. In one special case, the filtering is specialised for subsequent compression of the residuals using Discrete Cosine Transform, DCT.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 is a block diagram illustrating an apparatus for compressing video frames where the residual frames are dampened to achieve lower bitrates.

Fig. 2 shows an example for an upper and lower tolerance bound for one scanline of a video frame derived from masking effects of the Human Visual System.

Fig. 3a shows a plot as an example for upper and lower bounds derived from contrast masking.

Fig. 3b shows a plot as an example for upper and lower bounds derived from entropy masking.

Fig. 3c shows a plot as an example for upper and lower bounds derived from edge masking.

Fig. 3d shows a plot as an example for upper and lower bounds derived from temporal masking.

Fig. 4a is a block diagram illustrating the dampening of the residual frame by filtering within given bounds.

Fig. 4b is a block diagram illustrating the dampening of the residual frame by quantizing as coarse as possible within given bounds.

Fig. 5 is a block diagram illustrating the merging of compression and dampening of the residual frame by quantization.

Fig. 6a shows a plot as an example for an original frame and its prediction,

Fig. 6b shows a plot as an example for the corresponding residual frame with upper and lower bound and

Fig. 6c shows a plot as an example for the residual frame filtered in space.

Fig. 7a shows a plot as an example for an original frame and its prediction,

Fig. 7b shows a plot as an example for the corresponding residual frame with upper and lower bound and

Fig. 7c shows a plot as an example for the residual frame quantized in frequency.

Fig. 8 is a block diagram illustrating an apparatus for compressing video frames where the video frames are filtered to achieve higher compression ratios.

DETAILED DESCRIPTION OF THE INVENTION

The present invention will be described below with reference to the accompanying drawings.

First preferred embodiment

Fig. 1 shows the block diagram illustrating a first preferred embodiment of an apparatus for compressing video frames where the residual frames are dampened to achieve lower bitrates. A sequence of original frames is fed into a modelling module 100. The received model approximately describes the original sequence, e.g. based on prediction or motion compensation from earlier frames. The model is then input to a compress module 102 compressing the different components of the model. Because the model may not describe the original sequence sufficiently because of modelling artefacts or model compression artefacts, a residual frame for each original frame is calculated. Therefore, the model is again decompressed in a decompress module 104 and a prediction is made for each original frame in a predict module 106. Instead of the compressed and decompressed model, the original model may be fed to predict module 106 using switch 105. The residual frame is then derived by subtracting the prediction from the original frame. In many conventional encoding apparatus the residual frame would now be compressed in a compress module 110 without any pre-processing. Within the present invention the residual frame is pre-processed by a dampen module 108. The residual frame is influenced in such way that the output of the following compress module 110, based on a transform into frequency with following quantization of transform coefficients, requires less bits while retaining as much visual quality as possible. After transmitting the compressed model and the compressed

residual frame over any communication channel 112, both are fed to decompress modules 114 and 116. The reconstructed frame is derived by predicting the corresponding frame from the model with a predict module 118 and adding the decompressed residual frame.

The pre-processing of the residual frame is performed by changing the residual frame within a given tolerance bound derived from the original frame. Fig. 2 shows an example for such tolerance bound for one scanline of a video frame. The tolerance bound is derived from the corresponding original frame 200 considering masking effects of the Human Visual System (HVS). It consists of an upper bound 201 and a lower bound 202 giving for each pixel of the original frame the maximal allowed positive and negative change which is just not noticeable for the HVS. For good reconstruction quality, the reconstructed value for all pixels is confined to lie inside this bound as determined by the following expression:

$$I_{Org} + Chg_{Neg} < I_{Rec} < I_{Org} + Chg_{Pos} \quad (1)$$

where I_{Org} is the original frame, I_{Rec} is the reconstructed frame, Chg_{Neg} is the allowed negative intensity change and Chg_{Pos} is the allowed positive intensity change. Equivalently the bound can be given as an allowed range for the pre-processed residual as derived in the following. According to the present invention, the reconstructed frame is received by adding the pre-processed and transmitted residual frame $\tilde{I}_{Res}$ to the corresponding frame I_{Pred} predicted from the model:

$$I_{Rec} = I_{Pred} + \tilde{I}_{Res} \quad (2)$$

The original frame can be obtained by adding the original residual I_{Res} (before pre-processing and transmission) to the corresponding frame I_{Pred} predicted from the model:

$$I_{Org} = I_{Pred} + I_{Res} \quad (3)$$

Substituting I_{Rec} and I_{Org} in (1) by (2) and (3) leads to the following expression:

$$I_{Res} + Chg_{Neg} < \tilde{I}_{Res} < I_{Res} + Chg_{Pos} \quad (4)$$

This means, that for good reconstruction quality also the pre-processed residual $\tilde{I}_{Res}$ has to lie inside the given bound, where the bound can be derived from the original frame.

Any visual masking effect of the HVS or a combination of several effects may be considered to derive the tolerance bound. Well known masking effects are

- contrast masking ,
- entropy masking ,
- edge masking ,
- temporal masking and
- motion masking.

Figures 3a - 3d show a series of plots with examples for upper and lower bounds considering different masking effects. Fig. 3a illustrates bounds considering contrast masking. As in Fig. 2, one scanline of a video frame 301 and the corresponding upper bound 302 and lower bound 303 are drawn. The bound widens with increasing intensity due to Weber's law and at very low intensities due to the impact of ambient luminance. Several methods exist for deriving bounds from contrast masking effects of the HVS. One method, considering both Weber's law, ambient luminance and display characteristics, is described in Ruben Gonzalez, "Software Decodable Video for Multimedia based on a Computer Graphic Model", PhD Thesis, University of Technology, Sydney, 1994, which is hereby included by reference.

Fig. 3b illustrates bounds considering entropy masking. The distance between upper bound 312 and lower bound 313 increases in high entropy areas of the original video frame 311. Entropy of an video frame, sometimes also called activity or complexity of a video frame, may be estimated by calculating the standard deviation or the number of significant Fourier coefficients for a certain image area as described in Andrew B. Watson, Robert Borthwick and Mathias Taylor, "Image quality and entropy masking", in SPIE Proceedings, Human Vision, Visual Processing, and Digital Display VII, San Jose, volume 3016, 1997, which is hereby included by reference.

Fig. 3c illustrates bounds considering edge masking. The distance between upper bound 322 and lower bound 323 increases, possibly asymmetric, at intensity edges of the original video frame 321. One advantageous embodiment considers edge masking by deriving the bounds from the maximal difference between the original frame and a representation of the original frame shifted by a certain distance in any direction, e.g. 0.5 pixels. Alternatively, convolution with Sobel or Laplace masks or any other method for finding image discontinuities as described in R. Gonzalez and R. Woods, "Digital Image Processing", Addison-Wesley Publishing Company, 1992, pages 414-443, which is hereby included by reference, can be used.

Fig. 3d illustrates an example for temporal masking based on innovations. The intensity of one pixel of a video frame 332 and its corresponding upper bound 334 and lower bound 335 is plotted over time. The bound widens at time 331 when innovations appear and narrows with increasing temporal distance from the time of appearance. Examples for innovations are scene shifts or areas appearing after being covered by other objects in the sequence. The impact of temporal masking effects on visual quality perception is discussed in K. Boff, L. Kaufman and J. Thomas, editors, "Handbook of Perception and Human Performance", Chapter 6, Wiley New York, 1986, which is hereby included by reference.

The distance of the bounds may also depend on the complexity of the motion of the sequence. The bound may widen in areas with complex or fast motion or in areas with high accelerations.

Additionally, some information derived from the residual itself can be used to characterise the bounds. The residual contains data to compensate for artefacts due to modelling. Size and duration of an artefact can influence the distance of the bounds at the location of the artefact. The impact of the size of an artefact on its visibility is studied in H. R. Blackwell, "Contrast Thresholds of the Human Eye", Journal of the Optical Society of America, Vol. 36, No. 11, November 1946, pp. 624 - 643, which is hereby included by reference. As reported by Blackwell, the visibility of an artefact decreases with decreasing size. Therefore the bounds may widen in areas with only small artefacts. One advantageous embodiment considering the size of artefacts is to low pass filter the bounds derived according to equation (4).

The impact of the duration of an artefact on its visibility is similar to that of the duration of innovations. The bound narrows with increasing duration of an artefact. One advantageous embodiment derives the duration of an artefact by counting for each pixel the number of frames its residual value has exceeded the allowed positive or negative change. With increasing counter values the bound narrows correspondingly.

The technique described within the present invention may be incorporated into many existing types of video encoders. One possible encoding technique is MPEG-2 as described in ISO/IEC 13818, "Information Technology - Generic Coding of Moving Pictures and Associated Audio: Part 2 Video", which is hereby included by reference. Within the MPEG-2 encoding algorithm, B-frames may be pre-processed before compression, as described in the following. Referring again to Fig. 1, modelling 100 outputs two reference frames, I- or P-frames, two sets of motion vectors for each frame temporally located between the two reference frames and some mode information telling if a macroblock of a B-frame is not predicted (intra), forward predicted, backward predicted or interpolated. Compress module 102 and decompress module 104 are not used, because I- and P-frames were already compressed and decompressed within the modelling module 100 and motion vectors are compressed lossless, that means the output of the modelling module 100 is directly fed to predict module 106. For each frame located between the two reference frames a prediction is calculated from corresponding motion vectors, reference frames and mode information in predict module 106 using switch 105. The prediction is subtracted from original to derive the B-frame to be transmitted. Because a B-frame has residual character, it is then pre-processed in the dampen module 108 before performing DCT, quantization and run-length encoding within compress module 110.

Another possible encoding technique is the IDLE encoding technique as described in the patent application "Method and apparatus for compressing video sequences", already incorporated by reference. Referring to Fig. 1, the modelling module 100 outputs in this case two reference frames, two bilinear motion models and several Blend-fields possibly also represented as a bilinear model. The blend-fields indicate for each frame temporally located between the two reference frames, if a pixel or block of pixels is predicted from the first, from the second or from both reference frames of the model. Reference frames, motion models and blend-fields are compressed and again decompressed in modules 102 and 104 and a prediction is calculated for each frame

between the two reference frames in predict module 106. The prediction is subtracted from the original to derive the residual to be transmitted. The residual is then pre-processed in the dampen module 108 before performing DCT, quantization and run-length encoding within compress module 110.

Two versions of a technique for pre-processing the residual frame are described within the present invention. The aim of both versions is to reduce the amount of bits required for the compressed residual and to keep reconstruction quality at a given quality level. One preferred embodiment is the pre-processing by filtering of the residual frame in spatial domain. In another preferred embodiment, the filtering is performed by quantization of the residual frame in frequency domain which is even better tuned to a DCT compression scheme. Both versions are performed in such a way, that the reconstruction error resulting from pre-processing does not exceed a certain value defined by the tolerance bounds.

Second preferred embodiment

Fig. 4a shows the block diagram illustrating the pre-processing of the residual frame by filtering in spatial domain within given bounds. The original frame and residual frame are fed to a bound definition module 400 defining upper and lower bound of the tolerance bound, as previously described in equation (4). The residual frame is filtered in a filter module 401 and fed to the replace module 402. Within the replace module 402 it is checked if the pre-processed residual fits the tolerance bound. All values outside the tolerance bound are replaced by values which fit the tolerance bound. This may be the corresponding value from original residual frame or the corresponding nearest bound value.

The distance between upper bound and lower bound, which may vary from pixel to pixel depending on characteristics of original frame, leads to a certain reconstruction quality and to a certain bitrate after compression. A simple way to control the bitrate is to vary the distance of the bounds via a quality level parameter 405: increasing distance between bounds leads to a decreasing bitrate but also to a decreasing reconstruction quality.

Several filter types may be applied on the residual frame within filter module 401. In any case the filtering should lead to a reduced bitrate after compressing the filtered

residual frame within compress module 403. Possible filters in combination with compress module 403 based on a Discrete Cosine Transform (DCT) are

- fixed low pass filter by convolution with a fixed filter mask,
- low pass filter by convolution with an infinite filter range. This corresponds to an implementation where all pixel values within the tolerance bound are replaced by zero and all other pixel values remain untouched,
- adaptive low pass filter with a variable filter mask for each pixel value of the residual frame. The filter mask for each pixel is computed from the absolute value of the intensity gradient around the pixel. A higher absolute gradient leads to a smaller filter range. This ensures that intensity edges are not smoothed by filtering. One possible method for adaptive low pass filtering is described in the patent application "Method and apparatus for Motion Estimation", already incorporated by reference, where adaptive filtering is performed in a similar way on motion fields.

All methods reduce the amount of bits required for the compressed residual, because high frequency parts are removed from residual information which normally leads to a reduced number of significant transform coefficients after transformation into frequency domain. After quantizing transform coefficients, a higher number of zeros will be present which leads to higher compression ratios in the following runlength coding or entropy coding.

Filtering may be applied more than once. This is controlled by switch 404. In a first iteration, the original residual is filtered and values outside the tolerance bound are replaced. During following iteration steps, the filtered residual is fed back to filter module 401. The filter range may be changed with each iteration step. One possible embodiment of an iterative filtering is to start iteration with a wide filter range for all pixels and in following iteration steps to filter with a decreasing filter range only those pixels whose previously filtered values had violated the bound and were replaced by the original residual values. With such a mechanism each pixel is filtered as strongly as possible while retaining a certain quality level defined by the tolerance bound.

The pre-processing of the residual frame by filtering in spatial domain within given bounds is illustrated in Figures 6a - 6c. They show plots of one scanline of a video frame as examples for the preferred embodiment which is illustrated in Fig. 4a. Fig. 6a

shows a scan line of an original video frame 601 and a possible prediction 602. Compared to the original scanline 601, the predicted scanline 602 is blurred and contains a block artefact 603. Fig. 6b shows the residual 611 derived by subtracting the prediction 602 from the original 601 and the corresponding upper bound 612 and lower bound 613 defining the tolerance bound. Fig. 6c shows the low-pass filtered residual 621 which still stays within the tolerance bound defined by upper bound 622 and lower bound 623.

Third preferred embodiment

Fig. 4b depicts the invention in a third preferred embodiment. It shows a block diagram illustrating the filtering of the residual frame by quantization in frequency domain. The original frame and residual frame are fed to a bound definition module 410 defining upper and lower bound of the tolerance bound as previously described in equation (4). Within a DCT module 411 the residual frame is divided into blocks of pixels, e.g. 8 by 8 pixel blocks, and each block is transformed into frequency domain by performing DCT, producing a set of transform coefficients. In a quantize module 412 each set of transform coefficients is quantized with a constant quantizer. Then each quantized set of transform coefficients is transformed back into space within an inverse DCT module 413. The resulting quantized residual is fed to the replace module 414. Within the replace module 414 it is checked, if the quantized residual violates the tolerance bound. All values violating the tolerance bound are replaced by the corresponding value of the unquantized residual frame. The resulting replaced residual frame is finally fed to compress module 415, where the residual is compressed by performing DCT, quantizing transform coefficients and performing run length or entropy coding on quantized transform coefficients.

The quantizer value used in quantize module 412 should be coarser than that used finally in the compress module 415 to reduce amount of bits required for compressed residual. A coarser quantizer value of the transform coefficients leads to a higher number of zeros and a lower entropy of the non-zero coefficients. Therefore higher compression ratios are achieved when applying run-length coding or entropy coding.

One advantageous embodiment requires the ratio between the quantizer value used in quantize module 412 and the final quantizer value used in compress module 415 to be an integer value. Assuming that the quantization is based on the following

expression, quantization of each real valued transform coefficient c with quantizer value q_1 in quantize module 412 leads to the quantized transform coefficient c_{q_1} :

$$c_{q_1} = q_1 \cdot \text{round}(c/q_1). \quad (5)$$

If the corresponding block of pixels in space does not violate the tolerance bound, its transform coefficients after frequency transform within compress module 415 are equal to those derived after quantization in quantize module 412. Now each transform coefficient c_{q_1} is quantized with a second quantizer value q_2 which leads to

$$c_{q_2} = q_2 \cdot \text{round}(c_{q_1}/q_2) \text{ or} \quad (6)$$

$$c_{q_2} = q_2 \cdot \text{round}(q_1/q_2 \cdot \text{round}(c/q_1)). \quad (7)$$

If the ratio q_1/q_2 is an integer value, then the outer rounding operation can be discarded, and equation (7) can be written as

$$c_{q_2} = q_1 \cdot \text{round}(c/q_1). \quad (8)$$

that is, the transform coefficients after the second quantization are equal to those after first quantization. Therefore if a set of transform coefficients is first quantized using a coarser quantizer value where the quantization steps are an integer multiple of the quantization steps of the final quantization, then the final quantization will not change the values produced by the first quantization.

The distance between upper bound and lower bound, which may vary from pixel to pixel depending on characteristics of original frame, leads to a certain reconstruction quality and to a certain bitrate after compression. A simple way to control the bitrate is to vary the distance of the bounds via a quality level parameter 417: an increasing distance between bounds leads to a decreasing bitrate but also to a decreasing reconstruction quality.

Quantization may be applied more than once. This is controlled by switch 416. In a first iteration, each block of the original residual is quantized in frequency domain with a coarse quantizer. During following iteration steps, the blocks violating the tolerance

bound are fed back to DCT module 411 and quantized with a quantizer value decreasing with increasing iteration steps. This results finally in a residual where each block is quantized as coarsely as possible. Another possibility is to start iteration with a fine quantizer (e.g. the same quantizer value as used finally in compress module 415), feed back those blocks not violating the tolerance bound and quantize them with a quantizer value increasing with increasing iteration steps. Again, ratio between quantizer value used in quantize module 412 and final quantizer value used in compress module 415 may be an integer value.

In the case of filtering by quantization in frequency domain, pre-processing and compression of the residual frames can be combined. Referring again to Fig 1, the dampen module 108 and the compress module 110 can then be merged to one module. A possible structure of such a merged module is illustrated in Fig. 5. The original frame and residual frame are fed to a bound definition module 500 defining upper and lower bound of the tolerance bound as previously described in equation (4). The residual frame is divided into blocks of pixels, e.g. 8 by 8 pixel blocks, and each block is fed to a DCT module 501, which transforms it into frequency domain and produces a set of transform coefficients. In a quantize module 502 the set of transform coefficients is quantized with a coarse quantizer value. Then the quantized set of transform coefficients is transformed back into space within an inverse DCT module 503. The resulting quantized residual block is checked against the tolerance bound in a check module 504. As long as the block violates the bound, the unquantized set of transform coefficients is quantized with a finer quantizer value 505 until the corresponding block in spatial domain fits the tolerance bound. Then the quantized set of transform coefficients is coded in a code coefficients module 506 by performing run length or entropy coding. Instead of an iterative determination of the coarsest possible quantizer value also a parallel determination is possible. In this case, for each unquantized set of transform coefficients several alternative sets of quantized transform coefficients are produced by using a different quantizer value for each alternative set. After transforming the alternative sets of quantized transforming coefficients into space, all are checked against the tolerance bound. The alternative set of quantized transforming coefficients corresponding to the coarsest quantizer value while still satisfying the bounds is fed to code coefficients module 506.

The pre-processing of the residual frame by quantization within given bounds is illustrated in Figures 7a - 7c. They show plots of one scanline of a video frame as an examples for the preferred embodiment shown in Figures 4b and 5. Fig. 7a shows one scanline of an original video frame 701 and a possible prediction 702. Compared to the original scanline 701, the predicted scanline 702 is blurred and contains a block artefact 703. Fig. 7b shows the residual 711 derived by subtracting the prediction 702 from the original 701 and the corresponding upper bound 712 and lower bound 713 defining the tolerance bound. Fig. 7c shows the residual 721 quantized as coarsely as possible and still staying within the tolerance bound defined by upper bound 722 and lower bound 723. In the drawn example, the residual was divided into 8 by 8 pixel blocks before transformation into frequency. In Fig. 7c some block borders 724 to 725 are drawn. It can be seen that the quantized residual consists of consecutive independent patterns each composed of a set of cosine functions.

Fourth preferred embodiment

The previously described methods of pre-processing a residual can also be applied directly on video frames. Fig. 8 shows the block diagram illustrating a preferred embodiment for compressing video frames where the video frames are pre-processed to achieve higher compression ratios. A video frame is fed to a filter module 810 where it is filtered in spatial domain or frequency domain to reduce the number of bits produced by the following compress module 820. Filtering in spatial and frequency domain is performed similar to the methods described in the previous embodiments. Instead of a residual frame, the video frame is filtered. The bounds are determined only from the video frame according to the following equation:

$$I_{Video} + Chg_{Neg} < \tilde{I}_{Video} < I_{Video} + Chg_{Pos} \quad (9)$$

where I_{Video} is the video frame, $\tilde{I}_{Video}$ is the dampened and compressed video frame, Chg_{Neg} is the allowed negative intensity change and Chg_{Pos} is the allowed positive intensity change.

While the invention has been particularly shown and described with reference to the preferred embodiments thereof, it will be understood by those skilled in the art that

various changes in form and detail may be made therein without departing from the spirit and scope of the invention.

The invention as described herein can be implemented by a program which runs on a general purpose computer, it may also be implemented for example by a specially configured chip, such as an ASIC, or for example it may be implemented by means of a Digital Signal Processor DSP. It may also be implemented by a program stored on a computer readable data carrier or by means of a program which is transmitted to the user or to the computer on which it runs by any transmission link, like e.g. also via the internet.

CLAIMS

1. A method of processing one or more video images to be compressed, each video image comprising pixels, and said method of processing comprising the step of:

filtering one or more of said video images so as to improve compressibility, where said filtering is performed with the constraint such that the deviation between the filtered image and the non-filtered image lies within certain bounds, wherein said one or more video images to be filtered comprises one or more of

original video images,

residual images which are derived from one or more of said original images as the differences between said original images and predictions representing said original images using a model.

2. The method according to claim 1, wherein the bounds are one upper and one lower limit which are computed for each pixel or block of pixels of the video images, based on features of the video images.

3. The method according to any of the preceding claims, wherein said method further comprises the step of

compressing said filtered video images,

4. The method according claim 3, wherein

compressing the video image comprises using a transform into frequency.

5. The method according to one of the preceding claims,

wherein the bounds are calculated so as to exploit human visual system masking effects.

6. The method according to claim 5,

wherein the human visual system masking effect comprises one of or a combination of the following for each pixel or each block of pixels:

(1) localization of edges,

(2) smoothness of surrounding pixels or blocks of pixels,

- (3) estimated human visual sensitivity based on intensity,
- (4) degree of movement,
- (5) response time after scene shifts, or
- (6) response time after uncovered image areas.

7. The method according to claim 6,

wherein the lower and upper bounds depend on one of or any combination of the following:

the response of an edge filter for the corresponding surround region of the original image, the presence of an edge leading to wider bounds;

the estimated image entropy in the corresponding surrounding region of the original image, higher entropy leading to wider bounds, the intensity of the corresponding surrounding region of the original image;

lower estimated human sensitivity at the given intensity level leading to wider bounds;

the movement of the corresponding surrounding region of the original image, faster or more complex movement leading to wider bounds;

the time since scene shifts, a short time since scene shift leading to wider bounds;

the time since an image area was revealed, as judged by the model, newly uncovered areas having wider bounds.

8. The method according to one of the preceding claims,

wherein the filtering of a video image comprises convolving the video image with a low pass filter mask.

9. The method according to claim 8,

wherein the standard deviation of the low pass filter mask is made dependent on localization of edges in the original image,

high degree of presence of an edge leading to lower standard deviation of the low pass filter mask.

10. The method according to claim 9,
wherein the filtering comprises replacing the intensity with zeroes.
11. The method according to one of claims 1 to 10,
wherein first of all filtering is performed without considering bounds and then the constraint is enforced by replacing the filtered video image at those pixels where the filtered video image does not satisfy the constraint with either values of the unfiltered image or values fitting the exceeded bound.
12. The method according to any of the preceding claims,
wherein the filtering is performed more than once.
13. The method according to any of the preceding claims,
wherein the filtering so as to improve compressibility comprises the following steps:
- (1) dividing the video image into blocks of adjacent pixels,
 - (2) transforming each block into frequency, producing a set of transform coefficients,
 - (3) quantizing the transform coefficients using coarse quantization,
 - (4) transforming the quantized transform coefficients back into space, producing a reconstructed block, and
 - (5) for each block where the reconstructed block does not violate the bounds, replacing the original block of the video image with the reconstructed block.
14. The method according to claim 13,
wherein the transform into frequency is a Discrete Cosine Transform, (DCT).
15. The method according to one of claims 13 and 14,
wherein steps (2) to (5) are repeated with different coarse quantizations.
16. The method according to claim 15,
wherein the ratios between the quantizations steps for the coarse and the finer quantization are integers.

17. The method according to one of the preceding claims, wherein the filtering so as to improve compressibility comprises the following steps:

- (1) dividing the video image into blocks of adjacent pixels,
- (2) transforming each block into frequency, producing a set of transform coefficients,
- (3) for each of several different quantizers, quantizing the set of transform coefficients, producing alternative quantized sets of transform coefficients,
- (4) for each of the alternative quantized sets of transform coefficients, transform back into space, producing alternative reconstructed blocks,
- (5) for each of the alternative reconstructed blocks, checking against bounds, and
- (6) selecting the alternative quantized set of transform coefficients that corresponds to the coarsest quantizer, without the corresponding reconstructed block violating the bounds.

18. The method according to claim 17, further comprising:
transmitting the selected set of transform coefficients.

19. The method according to claim 18,
wherein the transmitting comprises run length coding or entropy coding.

20. The method according to one of claims 1 to 19,
wherein the model comprises I or P reference images and motion vectors, as used in MPEG 1 and MPEG 2 compression standards, and
the residual image is the difference between original image and motion compensated prediction.

21. The method according to one of claims 1 to 19,
wherein the model is an IDLE model comprising one or more reference images and one or more bilinear models of motion.

22. An apparatus for processing one or more video images to be compressed, each video image comprising pixels, and said apparatus comprising:

means for filtering one or more of said video images so as to improve compressibility, where said filtering is performed with the constraint such that the deviation between the filtered image and the non-filtered image lies within certain bounds, wherein said one or more video images to be filtered comprises one or more of original video images,

residual images which are derived from one or more of said original images as the differences between said original images and predictions representing said original images using a model.

23. An apparatus according to claim 22, further comprising:

means for performing a method according to one of claims 1 to 21.

24. An apparatus according to claim 23, said apparatus further comprising:

means for transmitting and/or receiving video images processed according to one of claims 1 to 21 over a communications link, and/or

means for reconstructing video images which have been processed according to one of claims 1 to 21.

25. Computer program product comprising:

a computer-usable medium having computer-readable program code means embodied therein for causing the processing of one or more video images to be compressed, the computer-readable program code means in said computer program product comprising:

computer-readable program code means for causing a computer to

filter one or more of said video images so as to improve compressibility, where said filtering is performed with the constraint such that the deviation between the filtered image and the non-filtered image lies within certain bounds, wherein said one or more video images to be filtered comprises one or more of

original video images,

residual images which are derived from one or more of said original images as the differences between said original images and predictions representing said original images using a model.

26. A video signal comprising video images, each image comprising pixels, said video signal being derived from an original video signal comprising original video images by a method according to one of claims 1 to 21.

Fig. 1

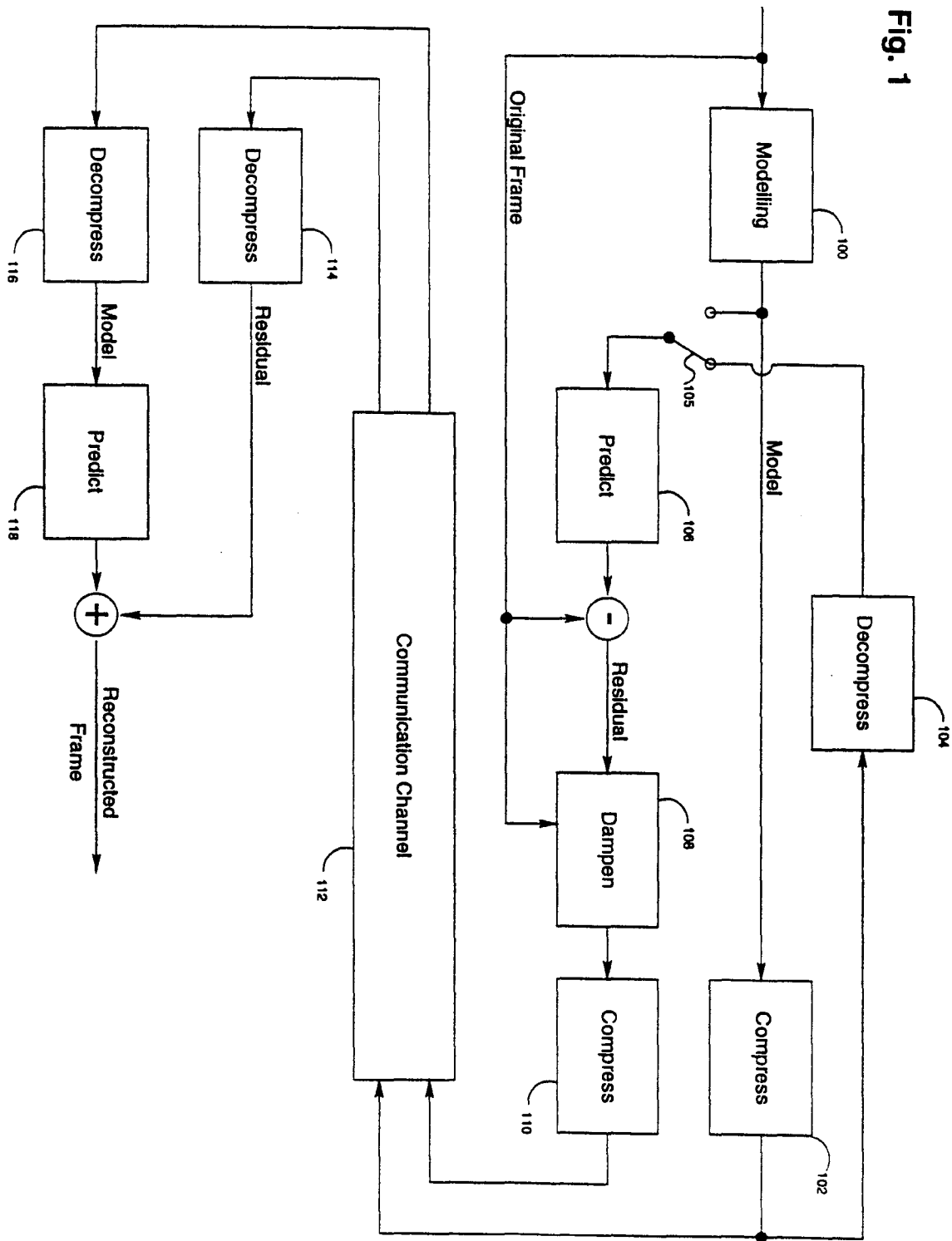


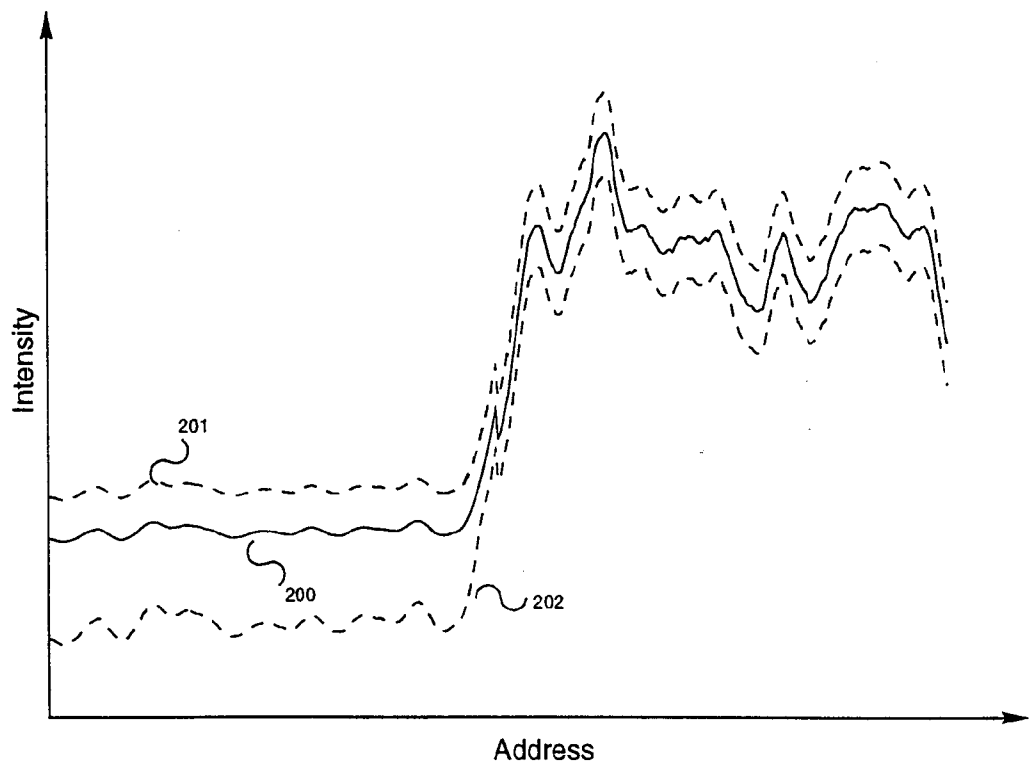
Fig. 2

Fig. 3a

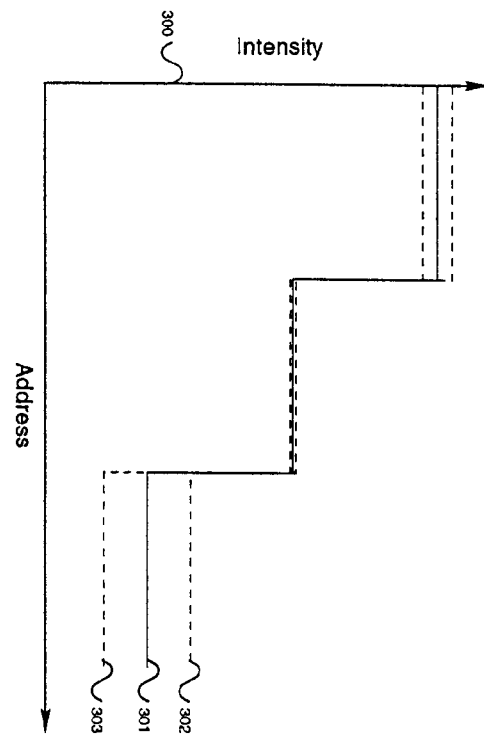


Fig. 3b

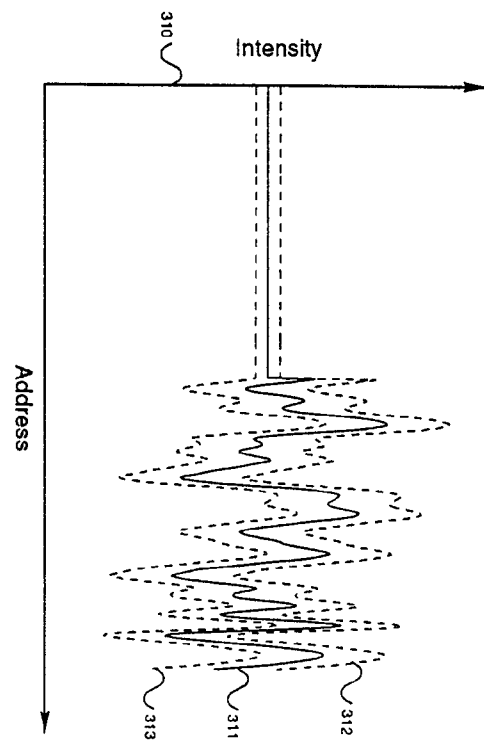


Fig. 3c

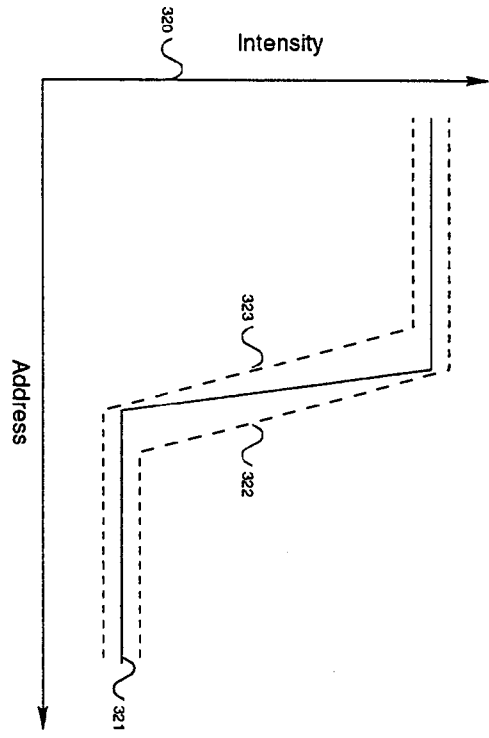
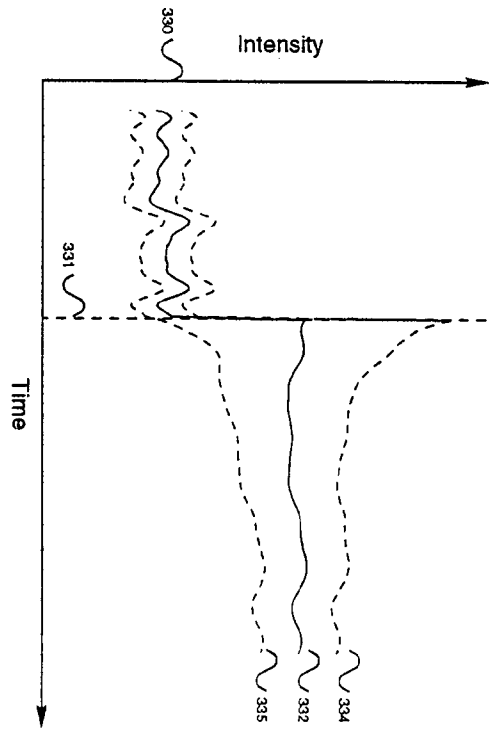


Fig. 3d



4/8

Fig. 4 a

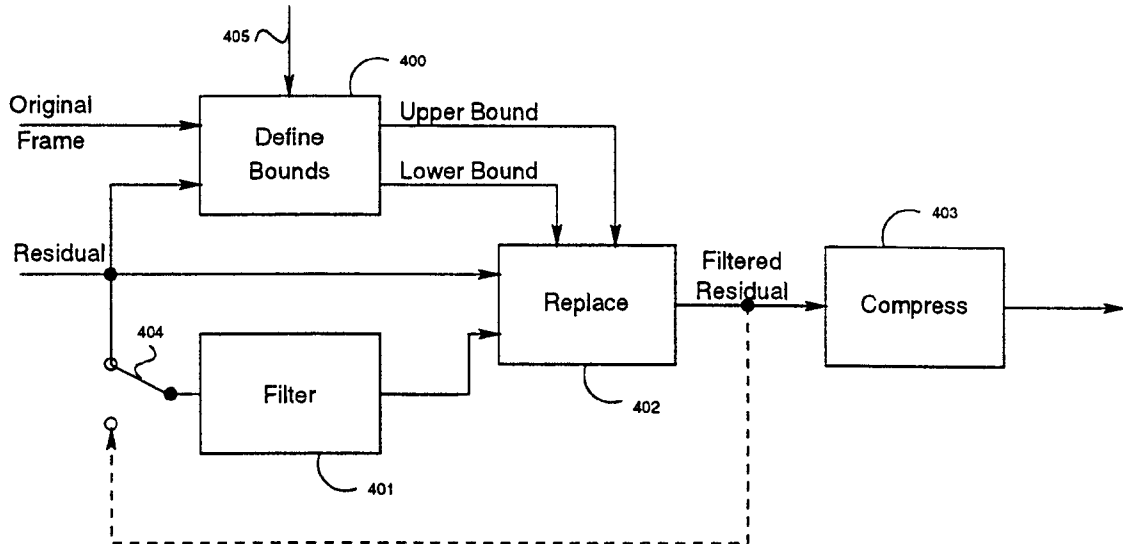


Fig. 4 b

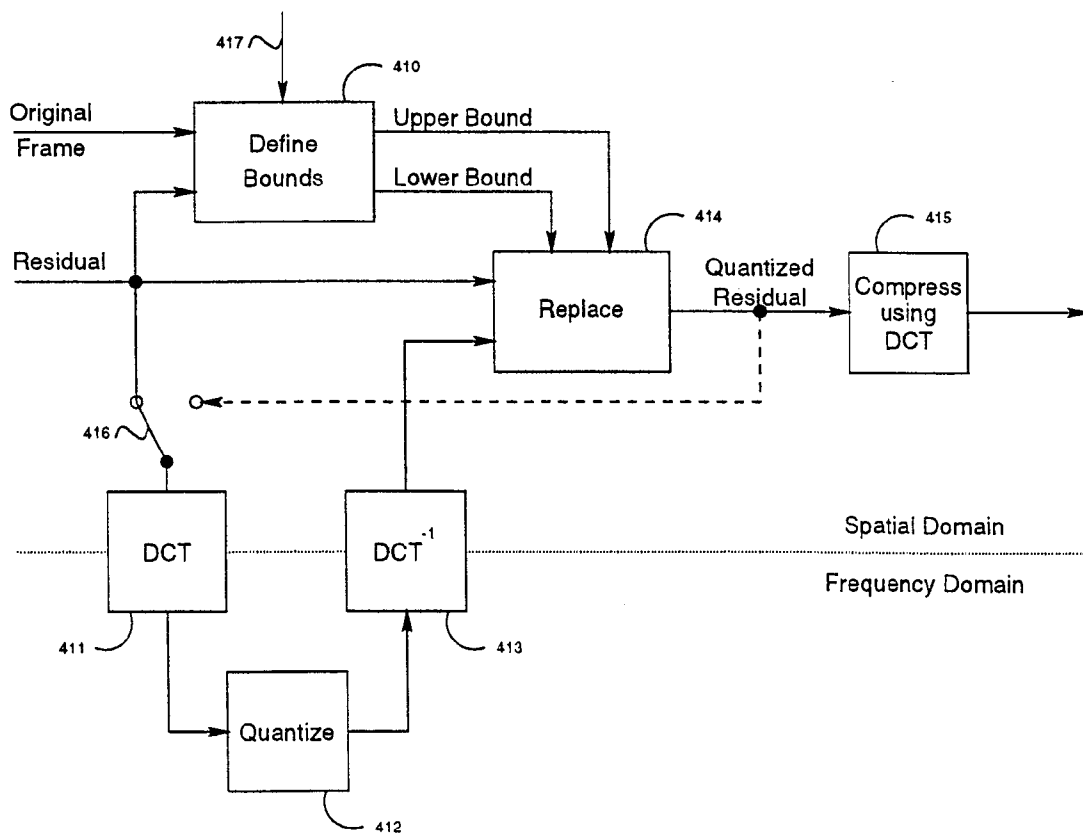


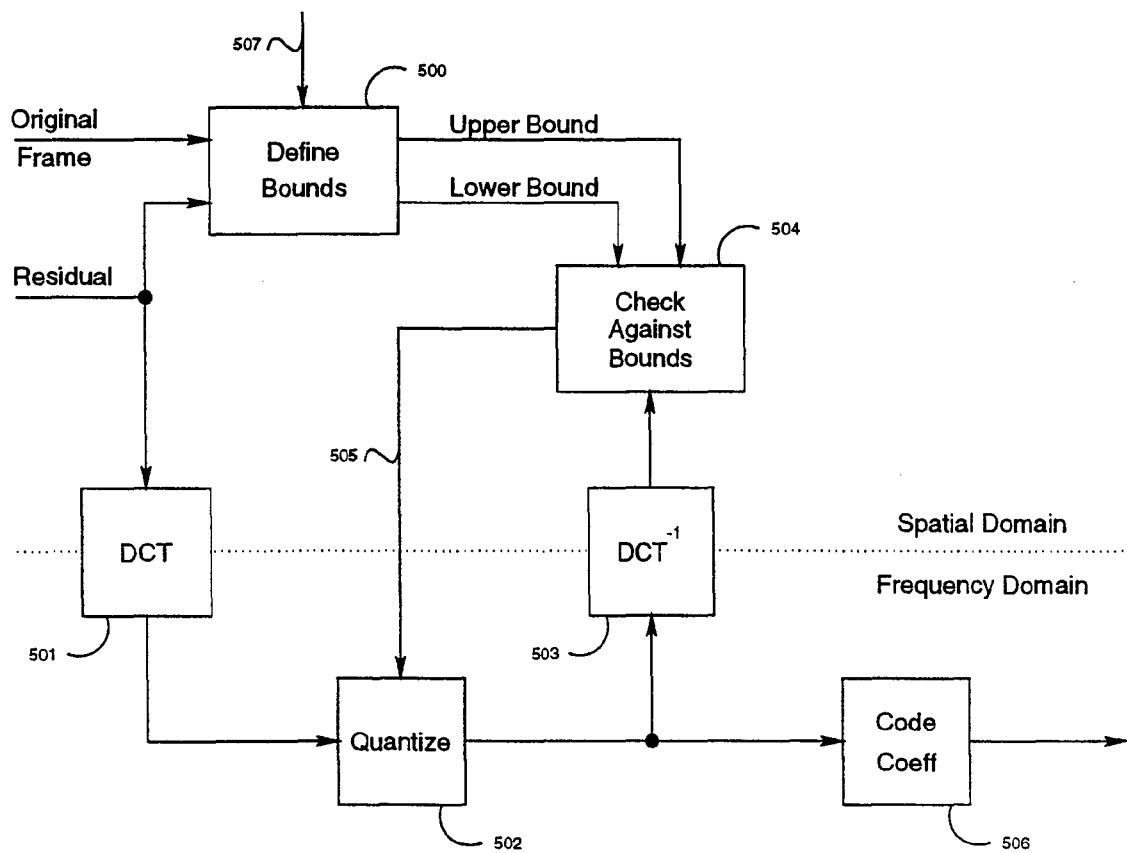
Fig. 5

Fig. 6a

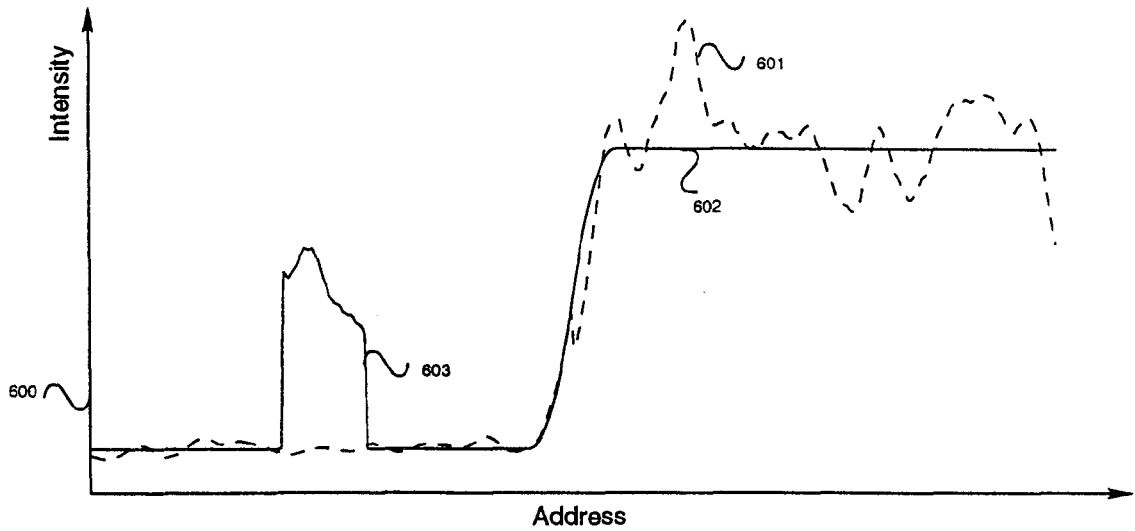


Fig. 6b

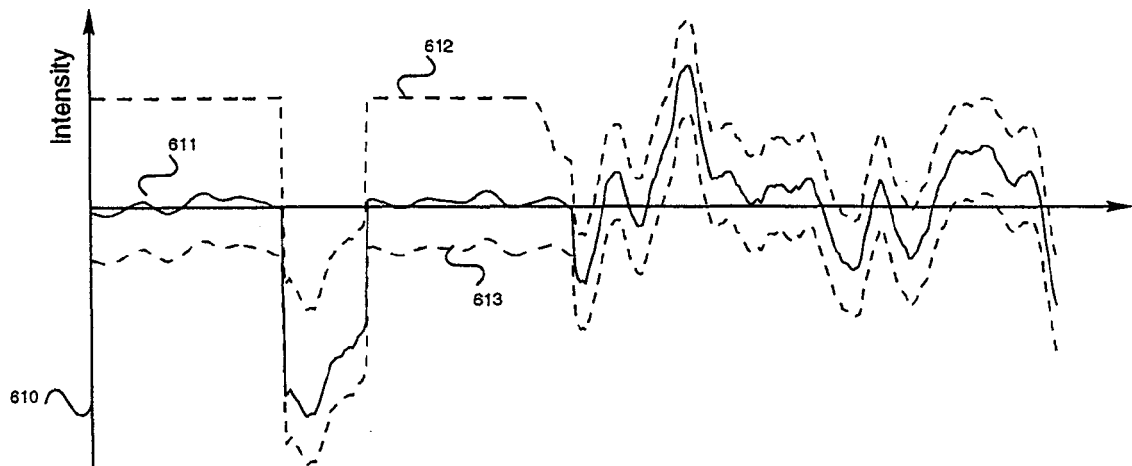
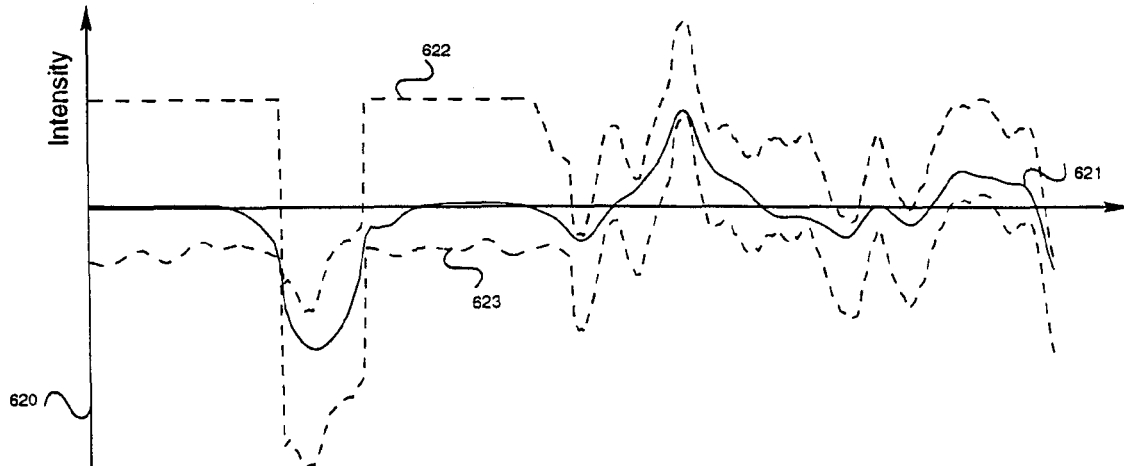


Fig. 6c



7/8

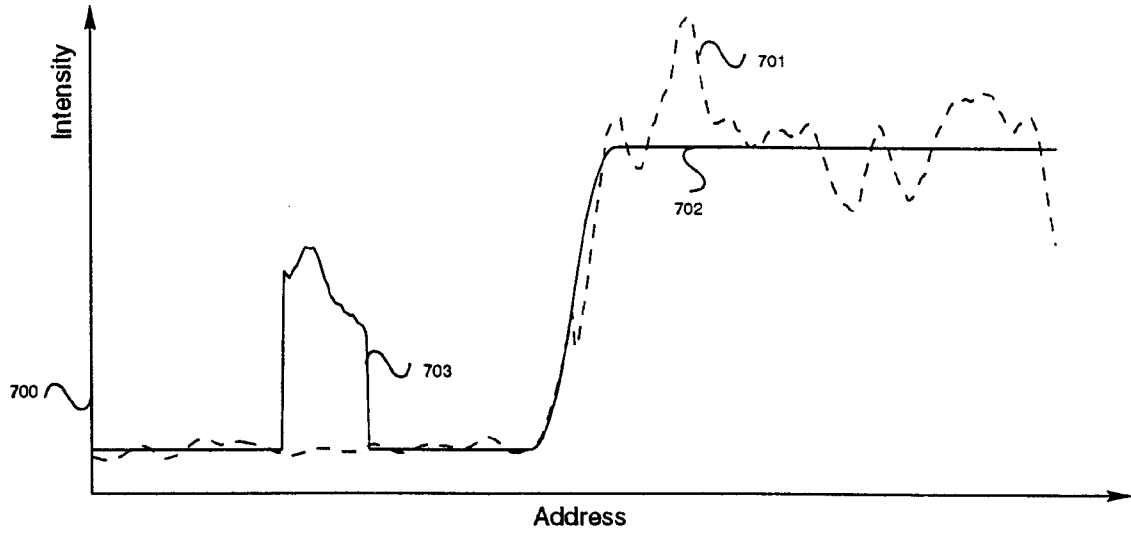
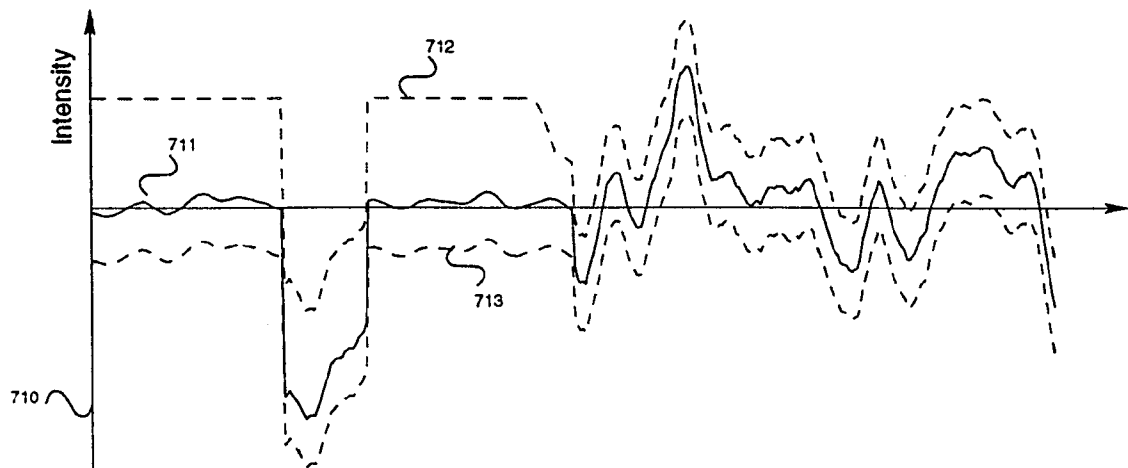
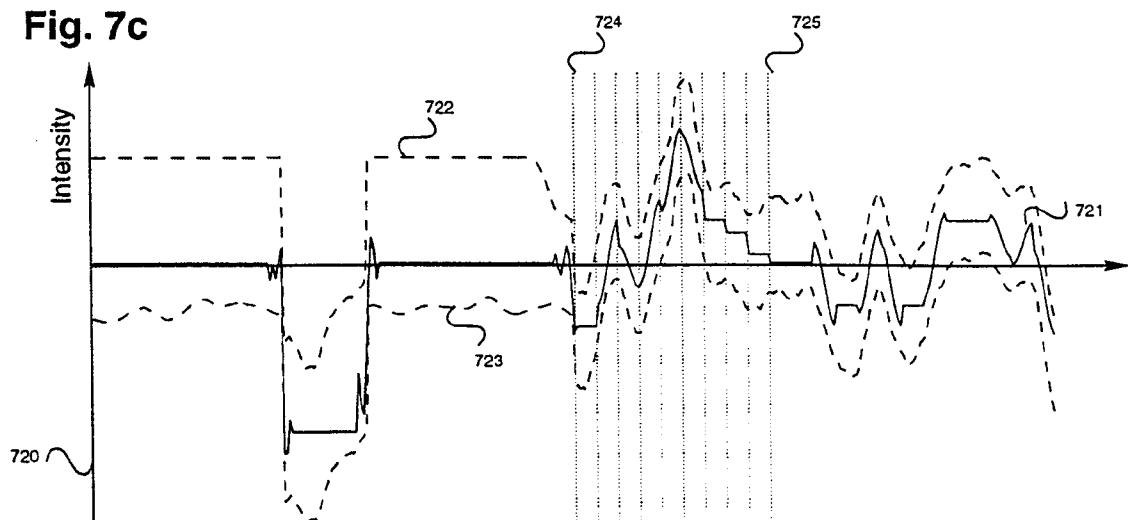
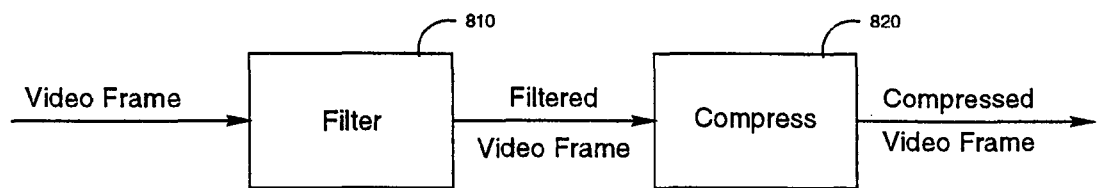
Fig. 7a**Fig. 7b****Fig. 7c**

Fig. 8

INTERNATIONAL SEARCH REPORT

Int: ional Application No
PCT/EP 98/02949

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04N7/26

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

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	WO 97 04589 A (MOTOROLA INC.) 6 February 1997	1, 3, 4, 8, 20, 22, 25
Y	see page 3, line 20 - page 4, line 5; figure 1 see page 6, line 20 - line 22 ---	13, 14
Y	EP 0 490 539 A (AMERICAN TELEPHONE & TELEGRAPH) 17 June 1992 see page 14 - page 17, line 15 ---	13, 14
A	JAYANT N ET AL: "SIGNAL COMPRESSION BASED ON MODELS OF HUMAN PERCEPTION" PROCEEDINGS OF THE IEEE, vol. 81, no. 10, October 1993, pages 1385-1421, XP000418793 see page 1404, column 1, paragraph 2 - page 1405, column 1, paragraph 1; figure 23 -----	1-26

☐ Further documents are listed in the continuation of box C.

☒ 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

10 September 1998

Date of mailing of the international search report

18/09/1998

Name and mailing address of the ISA

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

Authorized officer

Raeymaekers, P

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/EP 98/02949

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9704589 A	06-02-1997	US 5764307 A	09-06-1998
		AU 689048 B	19-03-1998
		AU 6275496 A	18-02-1997
		CA 2200730 A	06-02-1997
		CN 1159270 A	10-09-1997
		EP 0791270 A	27-08-1997
EP 0490539 A	17-06-1992	US 5136377 A	04-08-1992
		CA 2055046 A,C	12-06-1992
		JP 4341083 A	27-11-1992
		MX 9102424 A	01-06-1992