



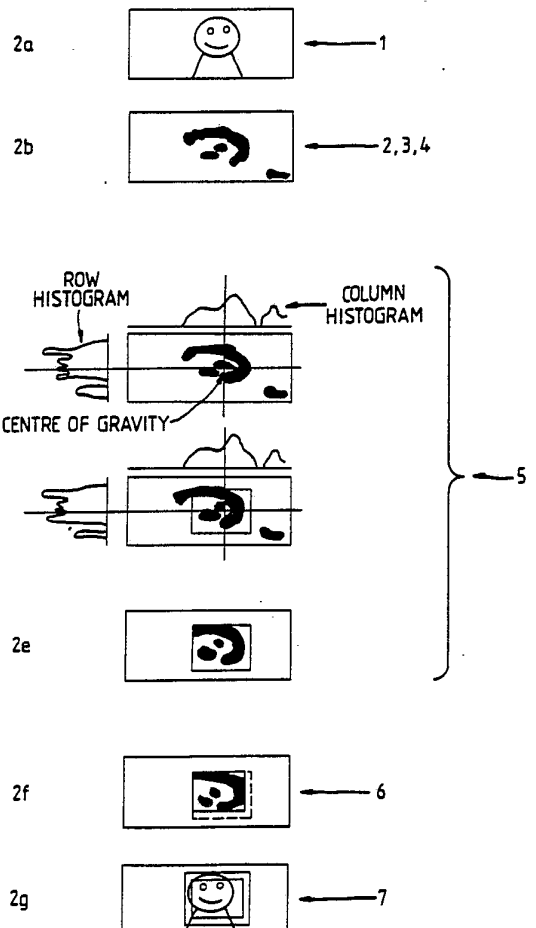
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁴ : H04N 7/13, 7/137</p>	<p>A1</p>	<p>(11) International Publication Number: WO 88/ 09101 (43) International Publication Date: 17 November 1988 (17.11.88)</p>
<p>(21) International Application Number: PCT/GB88/00357 (22) International Filing Date: 6 May 1988 (06.05.88) (31) Priority Application Number: 8710737 (32) Priority Date: 6 May 1987 (06.05.87) (33) Priority Country: GB</p> <p>(71) Applicant (for all designated States except US): BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).</p> <p>(72) Inventor; and (75) Inventor/Applicant (for US only) : SEXTON, Graham, Grainger [GB/GB]; 3 Nayland Road, Cavendish Park, Felixstowe, Suffolk IP11 8XJ (GB).</p> <p>(74) Agent: LLOYD, Barry, George, William; British Telecom, Intellectual Property Unit, 151 Gower Street, London WC1E 6BA (GB).</p>		<p>(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>

(54) Title: VIDEO IMAGE PROCESSING

(57) Abstract

A video processor identifies the head area from a head against background scene using Vector Quantisation operating on a composite vector codebook including "head" vectors and "background" vectors. The codebook is initially derived by storing several frames of an image training sequence (1); differencing adjacent frames (2); thresholding the difference data against a luminance threshold (3); median filtering (4); clustering the data sets (5); determining the minimum rectangle which will contain all the remaining non-zero pels (6); generating a border around the rectangles (7); generating a head codebook from the pels of the original frames of the image that fall within the borders (8); similarly generating a background codebook; and finally generating a composite codebook (9) in which "head" and "background" vectors are distinguished by flags. Then the head is tracked by Vector Quantising (10) the image with this codebook (9); and analysing the flags (11) (12) to obtain the head position (13).



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.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	LI	Liechtenstein	SN	Senegal
CH	Switzerland	LK	Sri Lanka	SU	Soviet Union
CM	Cameroon	LU	Luxembourg	TD	Chad
DE	Germany, Federal Republic of	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	US	United States of America
FI	Finland				

VIDEO IMAGE PROCESSING

This invention relates to analysis and processing of video images.

5 A video image (which will be understood to encompass frozen images such as facsimile images, in addition to moving images) will in general include at least one object which is of interest and a "background" of lesser interest (and hence of lesser importance).

10 To analyse the image, e.g. detect the presence/absence or position of a particular object of interest, is often desirable in a variety of applications.

15 In an image transmission system an improved picture quality might be achieved if data relating to important parts of the scene, i.e. objects of interest, is coded using relatively more bits than data relating to unimportant (i.e. background) parts. For example, in a videophone system a typical image comprises a head and shoulders against a background, and the face area of the head is visually the most important; it is thus desirable
20 to be able to identify the head area from the shoulders and background so as to be able to process the head at a higher refreshment rate than the rest, so that the impression of smooth head motion is conveyed. The ability to locate a head within a head and shoulders scene can
25 thus be used to modify the spatial allocation of video data, enabling a degree of visual importance to be attributed to blocks within the data.

30 Also, if the position of an object is accurately tracked with time it will be possible to predict its motion, thus allowing "motion compensated" DPCM.

One way of identifying different regions of an image is to utilise the method proposed by Nagao (M. Nagao - "Picture recognition and data structure", Graphic

Languages - ed Nake and Rossenfield, 1972). This method has been used in a videophone type system, on an image of a head and shoulders against a background. Some success was achieved in determining the sides of the head when the subject was clean shaven, but very little success was achieved in other cases; so this method is not considered reliable enough for the basis of an area identifying method.

Conventional coders, for instance hybrid discrete cosine transform coders, use no 'scene content' information to code the data within the scene, so each part of the scene is operated on as if it has the same visual importance as every other part.

Other image analysis applications are manifold (for example, in automated manufacturing systems).

It is also known to code video images for transmission using Vector Quantisation (VQ). In VQ coding, the image is represented initially by an array of digital data corresponding to the image frame. Blocks of array points ("sub-arrays") are compared with vectors from a codebook, and the best-matching vector selected using a "least squares" difference criterion. A code designating this vector is then transmitted to represent the sub-array. At the receiving end the indicated vector is selected from an identical codebook and displayed.

The underlying principle of the invention, however, is to use VQ as an identification (e.g. object location) method. The extent of the various aspects of the invention are defined in the claims appended hereto.

The different areas of a video image, when vector quantised (VQ), can be operated on differently provided each entry in the VQ codebook has an associated flag indicating which area that entry represents. So in the example of the videophone two different flag entries are

required, one for the head and the other for the remainder of the scene.

An embodiment of the invention will now be described by way of non-limitative example concerned with the
5 identification of a head in a head and shoulders against a background scene, with reference to the accompanying drawings in which:

10 figure 1 is a block diagram illustrating the initial stages of operation of parts of a coder embodying the invention;

figures 2a-g show schematically various stages in the training sequence used to derive the codebook;

figure 3 is a block diagram illustrating the operation of a coder embodying the invention;

15 figure 4a shows schematically a frame to be analysed;

figure 4b illustrates the sub-array blocks used in vector quantising figure 4a;

figure 4c shows the state of flags corresponding to the vector quantised image of figure 4b;

20 figure 4d shows schematically the result of analysing the frame of figure 4a according to the invention; and

figure 5 shows schematically a coder embodying the invention.

To enable the invention to operate, it is necessary
25 to have provided a composite codebook which includes vectors flagged as being "head". Preferably others are flagged as being "background". It is possible to derive a "standard" codebook for either an average or a given speaker, but to allow flexibility and greater accuracy of
30 identification, this codebook is derived at the start in an initial "training" sequence. A preferred way of implementing such a sequence will now be described.

To generate "head" and "background" parts of the codebook, it is necessary to unambiguously obtain some

"head only" data and "background only" data; a crude initial head detection algorithm is required.

Referring to figures 1 and 2, in order to detect the head, digital data representing several contiguous frames of the head and shoulders image are captured; for instance
5 in a store 1. One of these frames is depicted in figure 2a. This data does not need to be extremely accurate, but rather representative.

On the assumption that the prime moving areas within the data sequence are directly associated with the head
10 area, frame differencing 2 is applied to the data representing each adjacent pair of frames. This process typically yields a set of difference data for each adjacent pair representing moving areas together with
15 random noise across the whole image area.

For all picture elements (pels) represented by each set of difference data, each pel above a given threshold value of intensity is set to maximum intensity (255) and each pel below the threshold is set to minimum intensity
20 (0). This 'thresholding' 3 removes a large quantity of the random noise and some of the moving areas.

Median filtering 4 is next applied to each set of difference data, which very effectively removes most of the remaining random noise, but erodes only small amounts
25 of the moving areas.

The image represented by each set of data at this stage will rarely provide a clear outline of the head; unless the head to background contrast is very high and the movement of the head between adjacent frames is more
30 than one pel. Often only one side and the top of the head may be depicted as shown in figure 2b.

Generally, the moving areas will be clustered in regions around the head area, but some isolated clusters may arise due to motion in other areas of the image.

A clustering process 5 is used to remove some of the isolated clusters: two orthogonal histograms are generated, one representing the number of 'moving' pels in the columns of the image represented by the data and one representing the number of moving pels in the rows of the image represented by the data. The first order moments are calculated and the 'centre of gravity' of the image determined, as shown in figure 2c. A rectangle is then generated, centred on these co-ordinates, of such dimensions that a given percentage of moving area is included within it, see figure 2d. The pels remaining outside this rectangle are set to zero intensity, figure 2e. By a suitable choice of rectangle isolated clusters are removed by this process.

Constraints are imposed on the selection of the rectangles in order to reduce the occurrence of incorrect rectangles. Since a very small movement of the head between one frame and the next may produce a very small rectangle, the rate of change of size of the rectangle from one set of data to the next is restricted: either each of the boundary lines of the rectangle are constrained to lie within a small distance of the corresponding boundary in the immediately preceding set of data; or the maximum rate of change of the size of the rectangle is linked to the frame difference energy (eg. the square of the difference data), so if the difference energy is small the change is kept small, but if the difference energy is large the rate of change may be greater.

The rectangle - rectangles are used because they require very few bits of data to define - is then shrunk if necessary, at 6 in figure 1, and as shown in figure 2f, to become the smallest rectangle that can be placed around the data to enclose all the remaining non-zero pels. This

rectangle is assumed to represent an approximate model of the head.

5 A border is then created, at 7 in figure 1, around the final rectangle, as shown in figure 2g. This border defines an exclusion zone from where no data will later be taken. This ensures that when the border is applied to the respective frame of the original image the data outside the border will be exclusively head data and the data outside the border will be exclusively background data.

10 If five frames of data are initially captured in the store 1, then four adjacent pairs of frames are analysed and four sets of data result. After the four borders have been set 7 the head area data and the background area data are extracted from the first four frames of the original image respectively and the Linde-Buso-Grey algorithm is applied to generate a VQ codebook for each area 8, for example, a 9 bit background codebook and 10 bit head codebook (i.e. codebooks containing respectively 2^9 and 2^{10} entries). The two codebooks are then combined 9 to form one codebook, each entry of which has an associated flag indicating its origin.

25 Referring now to Figures 3 and 4a-d, after this training sequence is completed, the composite codebook is used to locate the head in successive image frames. The VQ coder operates just as it would in a prior art system using VQ as the transmission coding, but for each block of pels coded 10, the code generated will include a flag (for example, the first digit) indicating whether that block is "head" or "background" so that the head position is known for each frame.

30 It will of course be appreciated that when the codebook is derived at the coder as indicated above, VQ cannot be used as the transmission code (unless this

codebook is made known to the decoder first by transmitting an indication of the vectors).

5 Since the quantisation process is inherently approximate, it will be appreciated that occasionally blocks from the head part of the image may best match a vector from the "background" part of the codebook, or vice versa. The actual identification of the head will thus usually involve ignoring isolated "head" blocks using erosion and clustering 11, 12 (for example, as described above), or designating the area with the highest concentration of "head" blocks as the actual head.

10 Another method involves detecting isolated "head" blocks and then examining the error between the block and the "head" vector, and that between the block and the best-matching "background" vector, and if the two scores are similar (i.e. there is ambiguity about whether the block is "head" or "background"), reflagging the block to "background" instead.

15 If the head blocks are too scattered, it may be that the codebook is insufficient to characterise the head. In this case, a retraining sequence may be employed to regenerate the codebook.

20 This retraining sequence may either simply be a further sequence of the kind described above, or it may attempt to improve (rather than simply redefine) the codebook. For example, a count may be kept of the number of "incorrect" (i.e. scattered) as opposed to "correct" (i.e. concentrated in the head area) occurrences of each vector, and the scatter may this be reduced by rejecting from the codebook vectors which occur incorrectly too often.

25 Or, alternatively, the approximate head location derived by locating the greatest concentration of "head"

blocks may be used, in the same manner as described above, as an area for generating a new "head" codebook.

These latter approaches, in which VQ coder "learns" from each retraining sequence, are preferred on grounds of accuracy.

Figure 5 shows a block diagram of a video coding apparatus (eg for a video telephone) embodying the invention. Video signals are fed from an input 20 to a frame store 21 where individual picture element values are recorded in respective store locations so that desired sub-arrays of pels are accessible for further processing. The sub-array sizes may typically be 8 x 8. In an initial, training, phase of the apparatus a training control unit 22 - which may for example be a suitably programmed microprocessor system - carries out the codebook generation method described above, and enters the vectors (and flags) in a VQ codebook store 23. It will be understood that the VQ process involves matching 8 x 8 sub-array to the nearest one of the stored vectors, viz. a number of 8 x 8 patterns which are consistently fewer in number than the maximum possible number (2^{64}) of such patterns.

In the coding phase of the apparatus, the matching is carried out by VQ control logic 24 which receives successive sub-arrays from the frame store 21 and compares each of these with all the vectors in the codebook store. The simplest form of comparison would be to compute the mean square difference between the two; the vector giving the lowest result being deemed to be the best match. The output from the VQ control logic is the sequence of flags associated with the vectors thus identified.

The actual logic is carried out in this example by an inter frame differential coder 25, in which an inter-frame difference (in subtractor 26) is taken (in

conventional manner) between pels from the frame store 21 and a previous frame predictor 27. As is usual in such systems, a quantizer 28 and an output buffer 29 (to match the irregular rate of data generation to a transmission link operating at a constant rate) are shown. A receiver (not shown) uses the difference-information to update a reconstructed image in a frame store. The flag output from the VQ control logic 24 is connected (if required, via erode/cluster circuits 30) to the differential coder 25. When the flag indicates that "head" information is being processed, the coder operates normally. If however "background" is indicated, then the generation of difference information is carried out less frequently (eg on alternate frames only). This operation is illustrated by a switch 31 which, when the flag indicates "background", breaks the coding loop on alternate frames.

It will be apparent from the foregoing that any visually distinctive object or objects may be accurately detected, recognised or located using methods according to the invention.

CLAIMS

1. A method of identifying an object within an image array, comprising the repeated steps of:
 - 5 a) comparing a sub-array of the image with vectors from a set including members associated with the object, and
 - b) in the event of substantial similarity, labelling that sub-array as corresponding to the object.

- 10 2. A method according to claim 1, in which the set includes also members corresponding to the background, and each vector has an associated flag indicating which of the object or the background that vector is associated with, so that each sub-array may be labelled as corresponding to the object or to the background by the flag.

- 15 3. A method according to claim 1 or claim 2 in which the location of the object is identified by finding within the image the greatest concentration of sub-arrays labelled as corresponding to the object.

- 20 4. A method of detecting a plurality of different objects, according to any one of claim 1 to 3, in which the set includes members associated with each such object.

5. A method of detecting a human head within a video image, employing the method of any one of claims 1 to 4.

- 25 6. A method of encoding a video image signal comprising the steps of
 - a) identifying an area of the video image corresponding to an object of visual importance, using a method according to any one of claims 1 to 5, and

b) modifying the spatial allocation of coding data
in favour of that area,
whereby a degree of visual importance may be attributed to
that area of the image.

5 7. A method of encoding a video image signal according
to claim 6 wherein the video image signal is encoded so as
to update the area corresponding to the object at a higher
rate than other areas.

10 8. A method of generating a set of vectors for use in a
method of identifying a moving object according to any one
of claims 1 to 5, comprising the steps of

a) identifying an area of the image corresponding to
at least a part of the object, and

15 b) generating vectors from video-data representing
that area of the image, in which the areas of the
image corresponding to the object are identified
by analysing the difference between a pair of
temporally separated image frames, whereby the
object is detected by its movement.

20 9. An image analyser for identifying an object against
a background within an image, comprising vector
quantisation means arranged to compare sub-arrays of an
image array with vectors from a codebook and select
therefrom the most similar vector to each such sub-array,
25 the codebook comprising a subset of vectors associated
with the object and a subset of vectors associated with
the background, each such vector having an associated flag
indicating to which subset it belongs, whereby the
analyser may identify the object from the flags of the
30 vectors selected for the sub-arrays.

5 10. An image analyser according to claim 9, further comprising clustering means for determining the position of a boundary to enclose a given proportion of those sub-arrays flagged as corresponding to the object, whereby the image analyser may identify the position and spatial extent of the object.

11. An image analyser employing a method of identifying an object according to any one of claims 1 to 5.

10 12. A coder for encoding video image signals comprising:

a) an image analyser according to any one of claims 9 to 11 arranged to identify an object within an image, and

15 b) an encoder arranged to preferentially allocate video encoding data to the area of the image corresponding to the object.

13. A coder according to claim 12, in which the encoder is arranged to encode the unquantised image.

14. A coder according to claim 13, in which the coder is a discrete cosine transform encoder.

20 15. A coder according to claim 12, further comprising:

c) motion analysis means arranged to detect motion of the position of the identified object between temporally separated image frames, and to predict therefrom the position of the object in a subsequent image frame,

25 whereby the encoder may be a motion compensated DPCM encoder.

16. A coder arranged to employ a method of encoding according to claims 6 or claim 7.

5 17. A coder according to any one of claim 12 to 16, also arranged initially to generate vectors of the said subsets of the said codebook, further comprising identification means for identifying an area of the image corresponding to the object or to the background, whereby the respective vectors may be generated from data derived from the said area.

10 18. A coder according to claim 17 in which the identification means is arranged to analyse the difference between a pair of temporally separated image frames to identify areas of the image corresponding to the object.

15 19. A coder according to claim 18, in which the identification means comprises:

a) means for generating from a pair of frames of the image array elements within a notional two dimensional field, the value of each position within the field indicating whether the
20 difference between the luminance levels of the picture elements at corresponding positions in the two frames lies above or below a threshold;
and

25 b) clustering means for determining the centre of gravity within the said field of those array elements indicating a difference above the threshold and determining the position of a boundary about the centre of gravity which encloses a given proportion of those array
30 elements,

whereby all picture elements lying within a boundary so determined are identified as belonging to an area of the image corresponding to the object.

5 20. A coder according to claim 19 which means further comprises filtering means for median filtering the array elements within the notional two dimensional field prior to the determination of the centre of gravity.

10 21. A coder according to claim 19 or 20, wherein said boundary about the centre of gravity is of a finite number of elements in thickness.

22. A coder according to any one of claims 19, 20 or 21, wherein said boundary about the centre of gravity is of rectangular shape.

15 23. A coder according to claim 22, wherein the rectangular shaped boundary is centred upon the centre of gravity and each side of the rectangle is moved inward, if possible, until it abuts at least one of those array elements indicating a difference above the threshold.

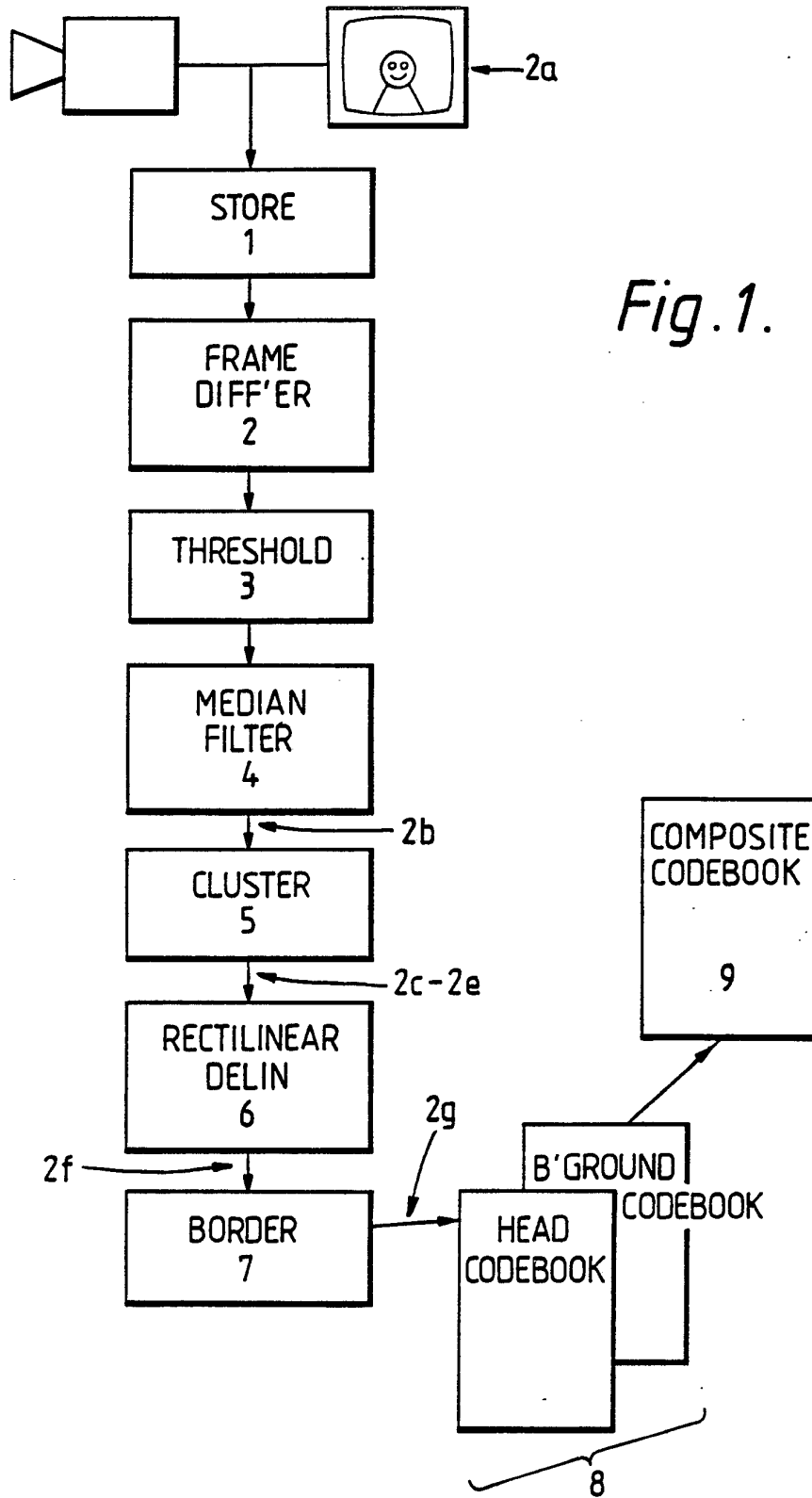


Fig. 2.

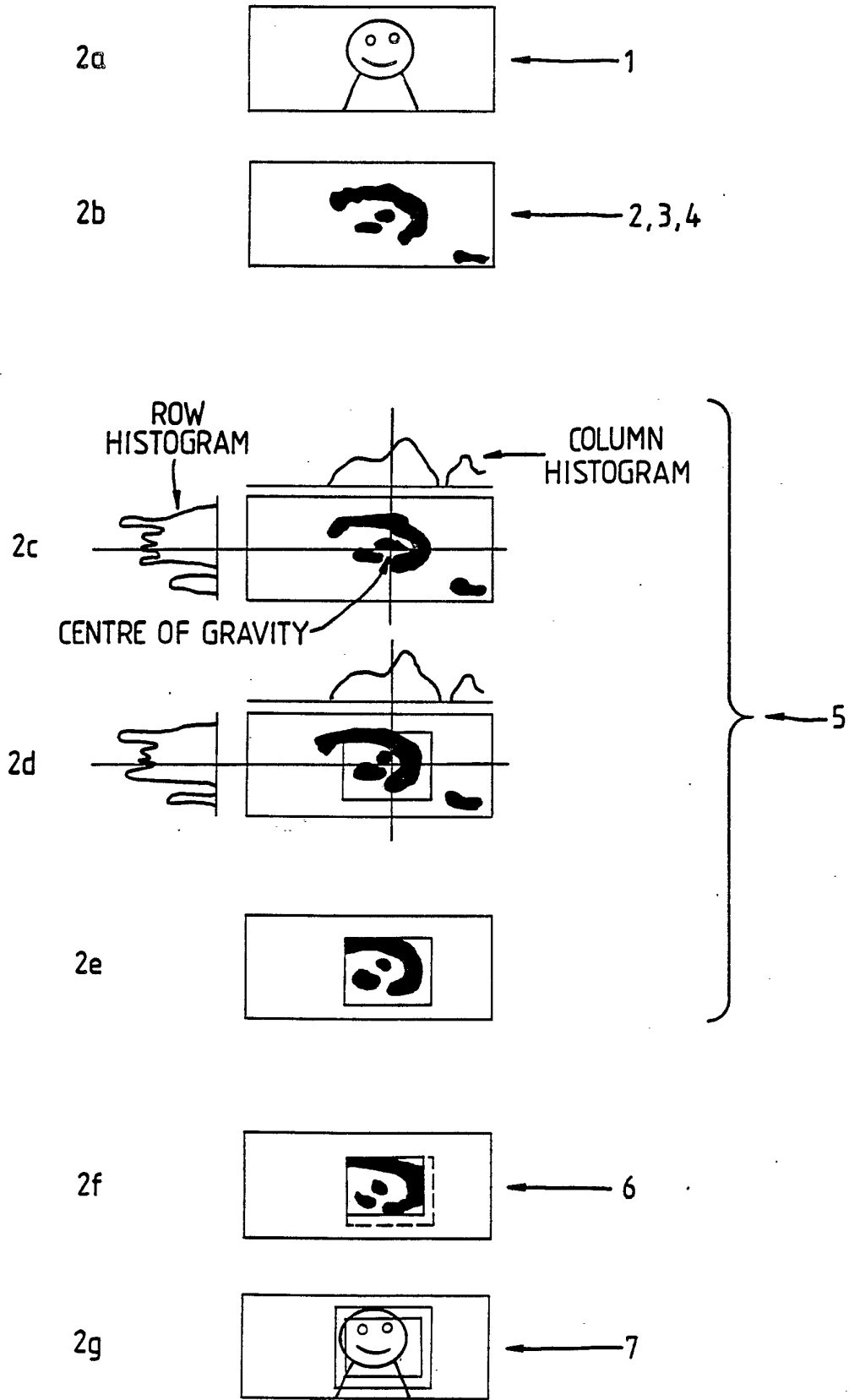


Fig. 3.

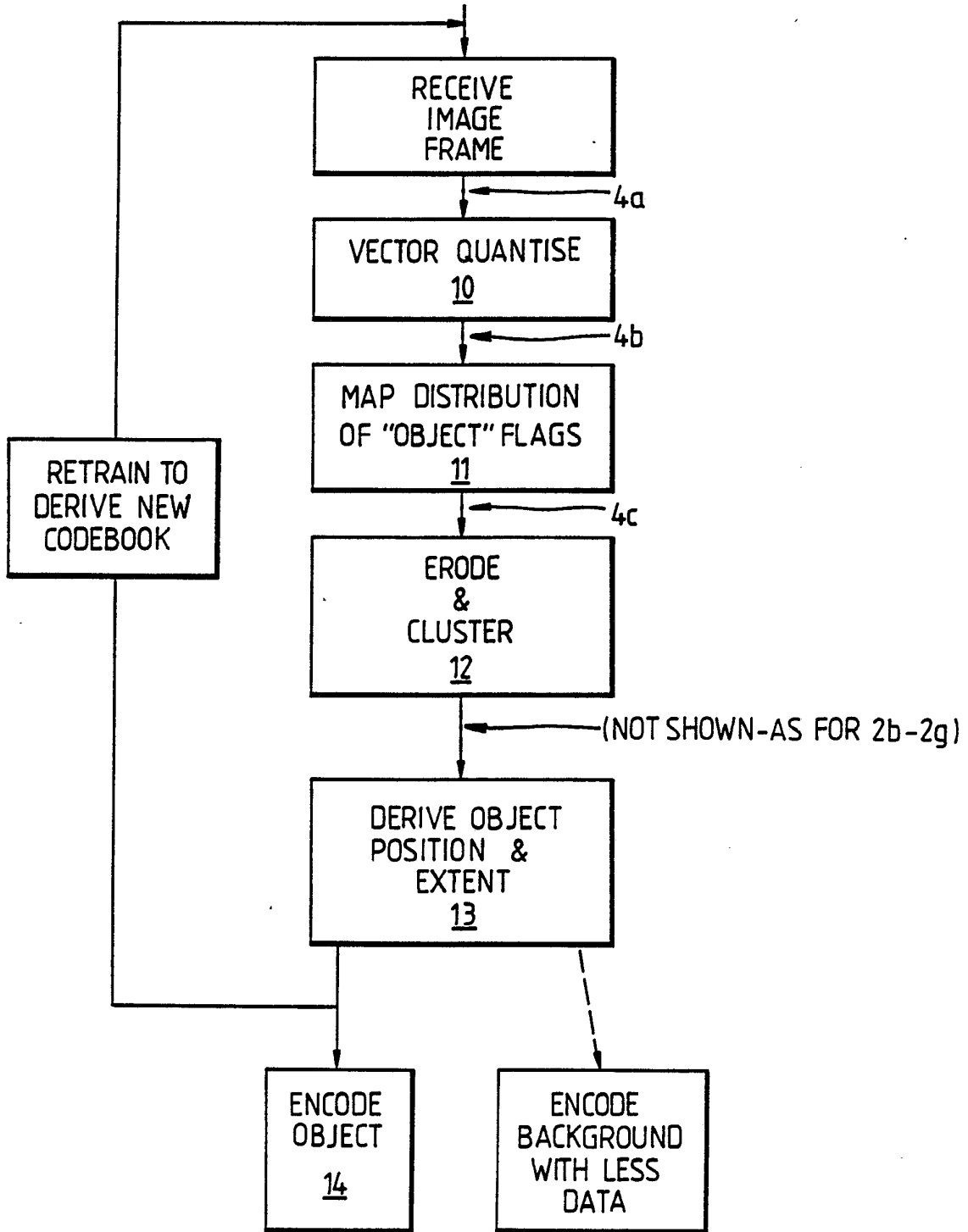


Fig. 4.

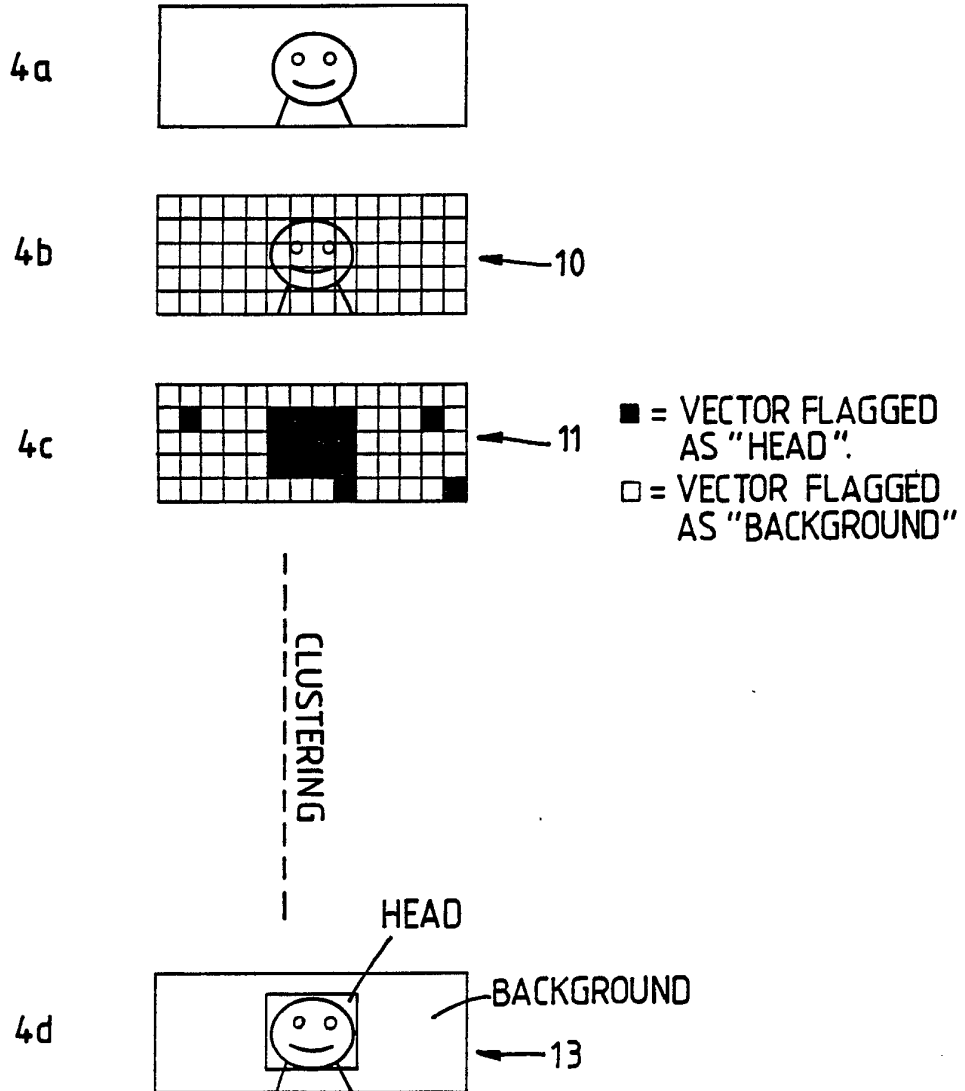
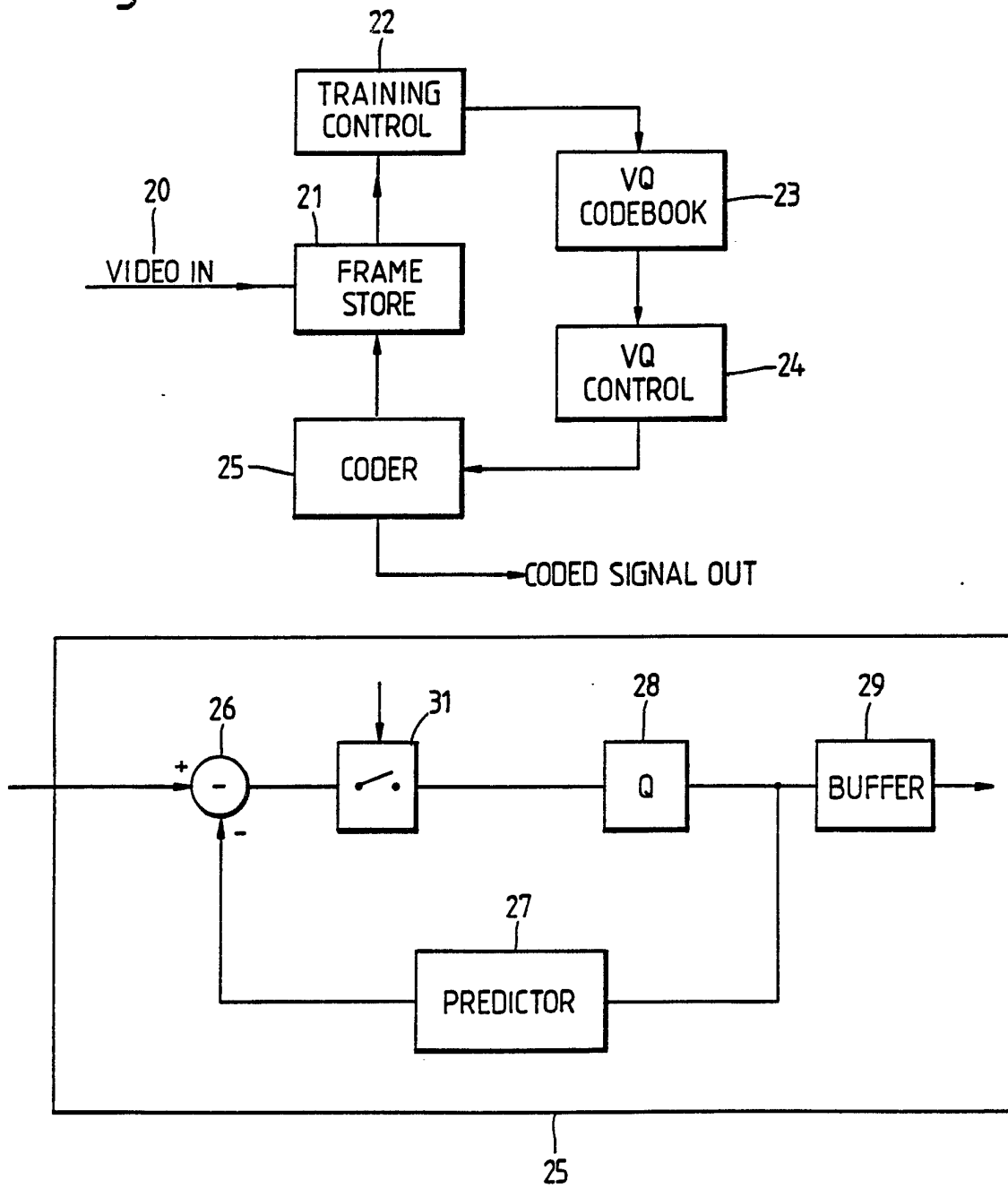


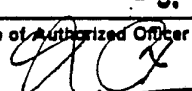
Fig. 5.



INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 88/00357

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) *		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC ⁴ : H 04 N 7/13; H 04 N 7/137		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
IPC ⁴	H 04 N 7	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are included in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹		
Category ⁹	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
X	IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 82, Proceedings, 3-5 May 1982, Paris, FR, volume 1 of 3, IEEE, (US), A. Gersho et al.: "Image coding using vector quantization", pages 428-431 see page 428, left-hand column, line 1 - right-hand column, line 35	1
Y		2,9
A		3,6,8,10-13,16,17
	--	
Y	IEEE International Conference on Communications, ICC '84, Links for the Future, Science, Systems & Services for Communications, 14-17 May 1984, Amsterdam, NL, volume 1, IEEE/Elsevier Science Publishers B.V. (North-Holland), (Amsterdam, NL), C.M. Lin et al.: "Motion compensated interframe color image coding", pages 516-520 ./.	2,9
<p>* Special categories of cited documents: ¹⁰</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"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</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"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.</p> <p>"&" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
25th July 1988	- 9. 09. 88	
International Searching Authority	Signature of Authorized Officer	
EUROPEAN PATENT OFFICE	 L. ROSSI	

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
	see page 516, left-hand column, line 26 - right-hand column, line 29	
A	--	1,7,11,16
A	IEEE International Conference on Communications, June 9182, volume 1, (Philadelphia, US), F. May et al.: "Picture coding with motion analysis for low rate trans- mission", pages 2G.7.1 - 2G.7.5 see page 2G.7.1, left-hand column, lines 1-10; page 2G.7.1, left-hand column, lines 25-33; page 2G.7.2, left-hand column, lines 1-40; paragraphs 3.2, 3.2.I, 3.3, 3,4; figures 4a,b,c	4,7,8,10, 13-16,18, 19,22
A	EP, A, 0143010 (THOMSON-CSF) 29 May 1985 see page 3, line 14 - page 6, line 4	8,18-20, 22
A	US, A, 3761613 (LIMB) 25 September 1973 see column 2, line 18 - column 3, line 19	7,8,16,18, 19,21
A	The Radio and Electronic Engineer, volume 48, no. 6, June 1978, Institution of Electronic and Radio Engineers, (London, GB), R.C. Nicol: "Interfacing intraframe d.p.c.m. with conditional replenishment coding for viewphone signals", pages 277-284 see page 279, right-hand column, lines 8-37; page 280, right-hand column, lines 5-44; figures 4,5	5,10,19, 23
A	Reconnaissance des Formes et Traitement des Images, congrès AFCET-IRIA, volume 2, February 1978, (Paris, FR), G.E. Lowitz: "Compression des données images par reconnaissance des formes et clustering", pages 699-714 see page 702, line 37 - page 703, line 8	19
X,P, L	EP, A, 0225729 (BRITISH TELECOMMUNICATIONS) 16 June 1987 see claims 1-7	1,3-6
A,P		2

**ANNEX TO THE INTERNATIONAL SEARCH REPORT
ON INTERNATIONAL PATENT APPLICATION NO.**

GB 8800357

SA 22065

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 02/09/88. The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0143010	29-05-85	FR-A, B 2551290	01-03-85
		JP-A- 60072390	24-04-85
		US-A- 4613894	23-09-86
US-A- 3761613	25-09-73	None	
EP-A- 0225729	16-06-87	JP-A- 62120179	01-06-87