



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁷ : H04N 13/00, 7/26</p>	<p>A1</p>	<p>(11) International Publication Number: WO 00/31981</p> <p>(43) International Publication Date: 2 June 2000 (02.06.00)</p>
<p>(21) International Application Number: PCT/EP99/08243</p> <p>(22) International Filing Date: 27 October 1999 (27.10.99)</p> <p>(30) Priority Data: 09/196,574 20 November 1998 (20.11.98) US</p> <p>(71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).</p> <p>(72) Inventors: CHALLAPALI, Kiran; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). CHEN, Richard, Y.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).</p> <p>(74) Agent: SCHMITZ, Herman, J., R.; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).</p>		<p>(81) Designated States: JP, KR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Published <i>With international search report.</i></p>
<p>(54) Title: EXTRACTION OF FOREGROUND INFORMATION FOR STEREOSCOPIC VIDEO CODING</p>		
<p>(57) Abstract</p> <p>An image processing device which improves the transmission of image data over a low bandwidth network by extracting foreground information and encoding it at a higher bit rate than background information.</p>		

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	Kazakstan	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						

EXTRACTION OF FOREGROUND INFORMATION FOR STEREOSCOPIC VIDEO CODING

BACKGROUND OF THE INVENTION

1. Field of the Invention

The invention relates in general to image processing and in particular to the extraction and variable bit rate encoding of foreground and background information from a stereo pair of images for video conferencing applications.

2. Description of the Prior Art

In all video conference applications, the bandwidth of communication between the participants is typically limited, about 64 kilo bits per second for a telephone line connection. Better compression standards have been developed over the years for efficiently compressing low-bitrate audio and video data, for example H.263 and MPEG-4. However, in typical video conference applications, a majority of the picture data in any given scene consists of irrelevant information, for example objects in the background. Compression algorithms cannot distinguish between relevant and irrelevant objects and if all of this information is transmitted on a low bandwidth channel, the result is a delayed jumpy looking video of a video conference participant.

Prior systems, as shown in German Patent DE 3608489 A1, use a stereo pair of cameras to image the video conference participant. A comparison is then made of the two images and using various displacement techniques the contour of the foreground information is located (as described in the above identified German patent and also in Birchfield and Tomasi, "Depth Discontinuities by Pixel-to-pixel Stereo," Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay India ["Birchfield"]). Once the contour of the foreground information is located, the background information is also known. A single static background image is then transmitted to a receiver to be stored in memory. The foreground images are encoded and transmitted along with address data which define where in the background image the foreground images should be placed.

The problems with such systems is that the background looks artificial since it lacks all motion and the contour of the video conference participant must be defined with a certain degree of accuracy. In addition the encoder which is typically optimized for a

rectangular image such as an 8 x 8 block of DCT coefficients must encode an oddly shaped image which follows the contour of the video conference participant. This "oddly" shaped information must also be transmitted separately which is a load on both bandwidth and computational resources at both the encoder and decoder sides.

5

SUMMARY OF THE INVENTION

Accordingly, it is an object of the invention to extract the foreground information of a video conference image and encode it at a first bit rate and encode the background information at a second lower bit rate. This object is achieved by the use of a pair of cameras arranged such that each camera has a slightly different view of the scene. Two images are produced and the difference in location of corresponding matching pixels in each image is computed and the disparity in location of these pixels is determined. A small disparity between the location of two identical pixels indicates the pixels constitute background information. A large disparity indicates the pixels constitute foreground information. The foreground pixels are then transmitted at the higher bit rate while the background pixels are transmitted at the lower bit rate.

It is a further object of the invention to avoid having to accurately represent the contour of the video conference participant. This object is achieved by using the 8 x 8 DCT blocks of coefficients to define the contour. Any block that includes a predefined number of foreground pixels is encoded at the higher bit rate, while those blocks that fall below this predefined number are encoded at the lower bit rate.

It is even a further object of the invention to encode the data using a standard encoder which encodes an 8 x 8 DCT block of coefficients. Again this object is achieved by defining foreground information based on a block of DCT data rather than the precise boundary of the video conference participant.

The invention accordingly comprises the methods and features of construction, combination of elements, and arrangement of parts which will be exemplified in the construction hereinafter set forth, and the scope of the invention will be indicated in the independent claims. The dependent claims define advantageous embodiments.

30

BRIEF DESCRIPTION OF THE DRAWINGS

For a better understanding of the invention reference is had to the following drawings:

Figure 1 shows a video conference scheme which uses a stereo pair of cameras;

Figures 2A and 2B show the images that result from the cameras in Fig. 1.

Figure 3A shows the identification of the foreground information;

Figure 3B shows the DCT blocks which are transmitted at the higher bit rate;

Figure 4 shows a block diagram of a video conference device in accordance
5 with the invention;

Figure 5 shows a PC configured for operating the instant invention; and

Figure 6 shows the internal structure of the PC in Figure 5.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT.

10 Fig. 1 shows a video conference set up in accordance with the invention. A video conference participant 30 sits at a desk 32 in front of two cameras 10 and 20 slightly spaced from one another. In the background there is a computer 40, a door 50 with people walking in and out, and a clock 60. The view of camera 10 is shown in Fig. 2A as follows: the video conference participant 30 is positioned to the right of the lens of camera 10, the
15 computer 40 since it is a distance from the cameras it remains basically in the center of the image. The door 50 is in the right hand portion of the image. The clock 60 is in the left hand corner of the image.

The view of camera 20 is shown in Fig. 2B as follows: The video conference participant 30 is off to the left in the image. The clock 60 is to the left of the video conference
20 participant 30. The computer 40 is to the right of the video conference participant 30 but still remains basically in the center of the image. The door 50 is in the upper right hand corner of the image.

The images received from the two cameras are compared to locate pixels of foreground information. (There are many algorithms that can be used to locate the foreground
25 information such as those described in DE 3608489 and Birchfield hereby incorporated by reference). In a preferred embodiment of the invention, the image from the left camera 10 (image A) is compared to the image from the right camera 20 (image B). The scan lines are lined up, e.g. scan line 19 of image A matches scan line 19 of image B. A pixel on scan line 19 of image A is then matched to its corresponding pixel in scan line 19 of image B. So for
30 example, if pixel 28 of scan line 19 of image A matches pixel 13 of scan line 19 of image B the disparity is calculated at $28-13=15$. Because of the closely located cameras, pixels of foreground information will have a larger disparity than pixels of background information. A disparity threshold is then chosen, e.g. 7, and any disparity above the threshold 7 indicates the pixel is foreground information while any disparity below 7 indicates the pixel is background

information. These calculations are all performed in the foreground detector 50 of Fig. 4. The output of the foreground detector is one of the images, e.g. image B, and another block of data which is of the same size as the image data and indicates which pixels are foreground pixels, e.g. '1', and which are background pixels, e.g. '0'. These two outputs are supplied to a DCT block classifier 52 which creates 8 x 8 DCT blocks of the image and also binary blocks which indicate which DCT blocks of the image are foreground and which are background information. Depending on the number of pixels in a particular DCT block that are foreground information, which can be a predefined threshold or vary as the bit rate capacity of the channel varies, the block will either be identified to the encoder 56 as a foreground block (triggering high bit-rate encoding 56A) or a background block (triggering low bit-rate encoding 56B).

Fig. 3A shows image B with the dashed lines representing the information that is encoded as foreground information in accordance with the invention. Assume each square represents an 8 x 8 DCT block. A foreground threshold is set such that if any pixel within an 8 x 8 block is foreground information then the entire block must be encoded as foreground information. The dashed lines in Fig. 3A indicate the DCT blocks identified as foreground information, these blocks will be encoded with a finer quantization level.

Fig. 3B shows a binary DCT disparity block which is the output of DCT block classifier 52. Encoder 56 receives both the image B and the binary DCT disparity blocks. Any DCT block which corresponds to a logic '1' DCT disparity block is encoded finely. Any DCT block which corresponds to a logic '0' DCT disparity block is encoded coarsely. The result is most of the bandwidth of the channel is dedicated to the foreground information and only a small portion allocated to background information. A decoder 58 (shown in Fig.4) receives the bitstream and decodes it according to the quantization levels provided in the bitstream.

This invention has applications wherever there is a transmission of moving images over a network such as the Internet, telephone lines, videomail, video phones, digital television receivers etc.

In a preferred embodiment of the invention, the invention is implemented on a digital television platform using a Trimedia processor for processing and the television monitor for display. The invention can also be implemented similarly on a personal computer.

Figure 5 shows a representative embodiment of a computer system 7 on which the present invention may be implemented. As shown in Figure 5, personal computer ("PC") 8 includes network connection 11 for interfacing to a network, such as a variable-bandwidth network or the Internet, and fax/modem connection 12 for interfacing with other remote sources such as a video camera (not shown). PC 8 also includes display screen 14 for

displaying information (including video data) to a user, keyboard 15 for inputting text and user commands, mouse 13 for positioning a cursor on display screen 14 and for inputting user commands, disk drive 16 for reading from and writing to floppy disks installed therein, and CD-ROM drive 17 for accessing information stored on CD-ROM. PC 8 may also have one or
5 more peripheral devices attached thereto, such as a pair of video conference cameras for inputting images, or the like, and printer 19 for outputting images, text, or the like.

Figure 6 shows the internal structure of PC 8. As shown in Figure 5, PC 8 includes memory 25, which comprises a computer-readable medium such as a computer hard disk. Memory 25 stores data 23, applications 25, print driver 24, and operating system 26. In
10 preferred embodiments of the invention, operating system 26 is a windowing operating system, such as Microsoft® Windows95; although the invention may be used with other operating systems as well. Among the applications stored in applications area 51 of the memory 25 are foreground information detector/DCT block classifier/video coder 21 ('video coder 21') and video decoder 22. Video coder 21 performs video data encoding in the
15 manner set forth in detail above, and video decoder 22 decodes video data which has been coded in the manner prescribed by video coder 21. The operation of these applications has been described in detail above.

Also included in PC 8 are display interface 29, keyboard interface 41, mouse interface 31, disk drive interface 42, CD-ROM drive interface 34, computer bus 36, RAM 37,
20 processor 38, and printer interface 43. Processor 38 preferably comprises a microprocessor or the like for executing applications, such those noted above, out of RAM 37. Such applications, including video coder 21 and video decoder 22, may be stored in memory 25 (as noted above) or, alternatively, on a floppy disk in disk drive 16 or a CD-ROM in CD-ROM drive 17. Processor 38 accesses applications (or other data) stored on a floppy disk via disk drive
25 interface 32 and accesses applications (or other data) stored on a CD-ROM via CD-ROM drive interface 34.

Application execution and other tasks of PC 8 may be initiated using keyboard 15 or mouse 13, commands from which are transmitted to processor 38 via keyboard interface 41 and mouse interface 31, respectively. Output results from applications running on PC 8
30 may be processed by display interface 29 and then displayed to a user on display 14 or, alternatively, output via network connection 11. For example, input video data which has been coded by video coder 21 is typically output via network connection 11. On the other hand, coded video data which has been received from, e.g., a variable bandwidth-network is decoded by video decoder 22 and then displayed on display 14. To this end, display interface 29

preferably comprises a display processor for forming video images based on decoded video data provided by processor 38 over computer bus 36, and for outputting those images to display 14. Output results from other applications, such as word processing programs, running on PC 8 may be provided to printer 19 via printer interface 43. Processor 38 executes print
5 driver 24 so as to perform appropriate formatting of such print jobs prior to their transmission to printer 19.

It will thus be seen that the objects set forth above, and those made apparent from the preceding description are efficiently obtained and, since certain changes may be made in the above construction without departing from the scope of the invention, it is
10 intended that all matter contained in the above description or shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

It is also to be understood that the following claims are intended to cover all the generic and specific features of the invention herein described, and all statements of the scope of the invention, which, as a matter of language, might be said to fall therebetween. In the
15 claims, any reference signs placed between parentheses shall not be construed as limiting claim. The word "comprising" does not exclude the presence of elements or steps other than those listed in a claim. The word "a" or "an" preceding an element does not exclude the presence of a plurality of such elements. The invention can be implemented by means of hardware comprising several distinct elements, and by means of a suitably programmed
20 computer. In the device claim enumerating several means, several of these means can be embodied by one and the same item of hardware.

CLAIMS:

1. An image processing device, comprising:
an input which receives a stereo pair of images;
a foreground extractor (50) which detects foreground pixel information from
the stereo pair of images; and
5 an encoder (56) coupled to the foreground extractor (50) which encodes the
foreground pixel information at a first high level of quantization and which encodes
background pixel information at a second lower level of quantization.
2. The image processing device as claimed in claim 1, wherein the foreground
10 extractor (50) computes the difference in location of like pixels in each image and selects the
foreground pixels as those pixels whose difference in location falls above a threshold distance.
3. The image processing device as claimed in claim 1, wherein the foreground
pixel information is defined in terms of entire blocks.
15
4. An image processing system, comprising:
a stereo pair of cameras (10,20) for taking a stereo pair of images;
a foreground extractor (50) which detects foreground pixel information from
the stereo pair of images; and
20 an encoder (56) coupled to the foreground extractor which encodes the
foreground pixel information at a first high level of quantization and which encodes
background pixel information at a second lower level of quantization.
5. A method of encoding a stereo pair of images, comprising:
25 receiving the stereo pair of images;
extracting foreground information from the stereo pair of images; and
encoding the foreground information at a first higher quantization level and
encoding background information of the stereo pair of images at a second lower quantization
level.

6. The method in accordance with claim 5, wherein the step of extracting includes the following steps:
- identifying the locations of like pixels in each of the stereo pair of images;
 - calculating the difference between the locations of like pixels; and
 - 5 determining for each set of like pixels whether the difference between locations falls above a threshold difference, and if so identifying those pixels as foreground information.
7. Computer-executable process steps to process image data from a stereo pair of images, the computer-executable process steps being stored on a computer-readable medium and comprising:
- 10 a foreground extracting step to detect foreground pixel information from the stereo pair of images; and
 - an encoding step for encoding foreground pixel information of at least one image at a first higher quantization level and for encoding background pixel information of the
 - 15 at least one image at a second lower quantization level.
8. The computer-executable process steps as claimed in claim 7, wherein the foreground extracting step determines which 8 x 8 DCT blocks contain at least a predetermined amount of foreground pixel information; and wherein the encoding step
- 20 encodes the entire 8 x 8 block of DCT coefficients at the first higher quantization level if the 8 x 8 block of DCT coefficients contains the predetermined amount of foreground pixel information.
9. An apparatus for processing a stereo pair of images, the apparatus comprising:
- 25 a memory (25) which stores process steps; and
 - a processor (38) which executes the process steps stored in the memory so as (i) to extract foreground information from the stereo pair of images and (ii) to encode the foreground information at a first high level of quantization and to encode background information at a second low level of quantization.

1/6

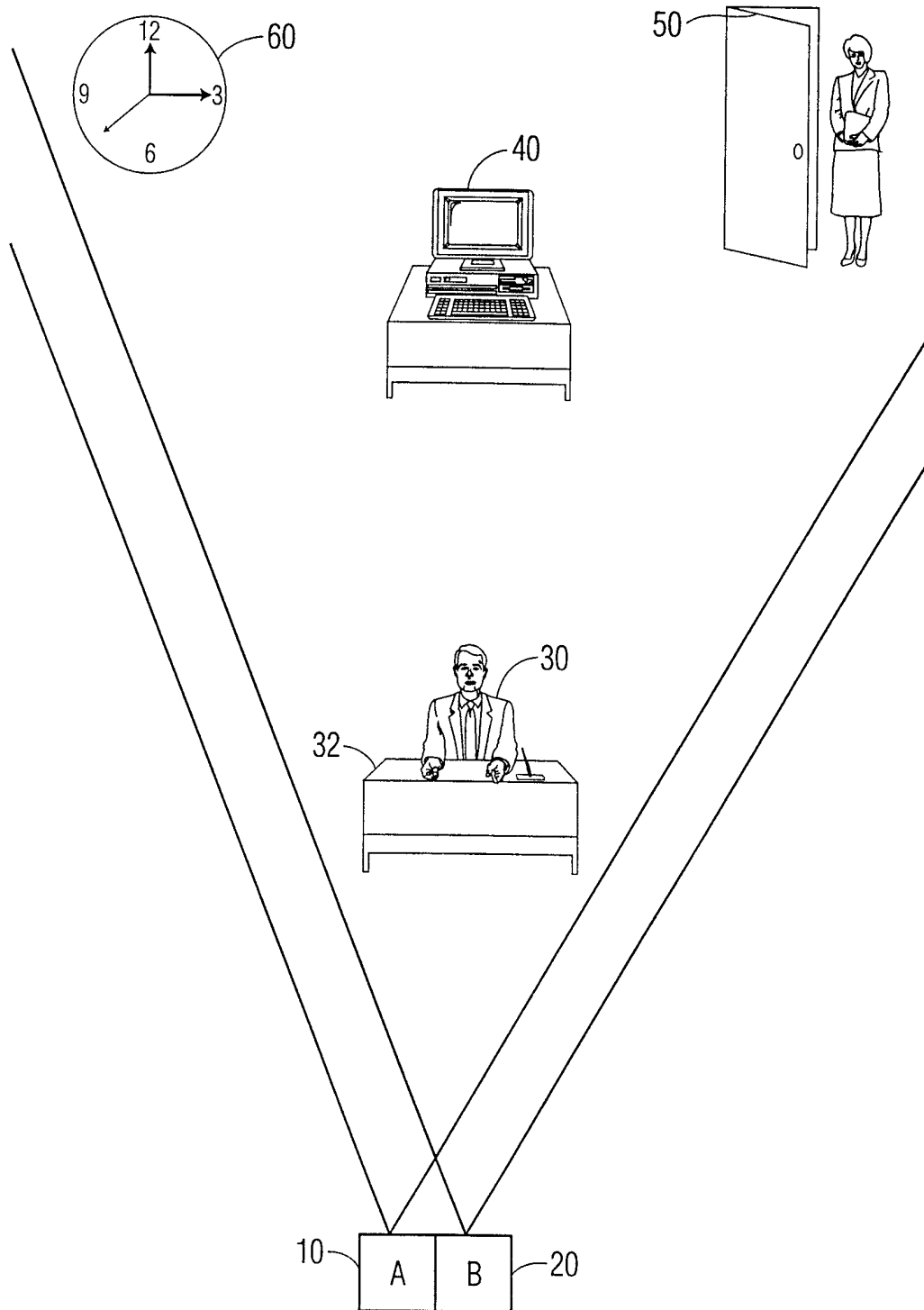


FIG. 1

2/6

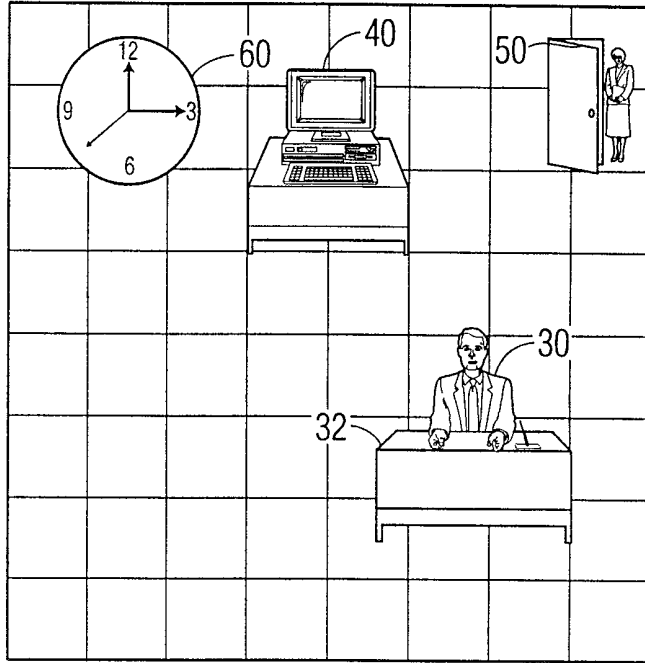


FIG. 2A

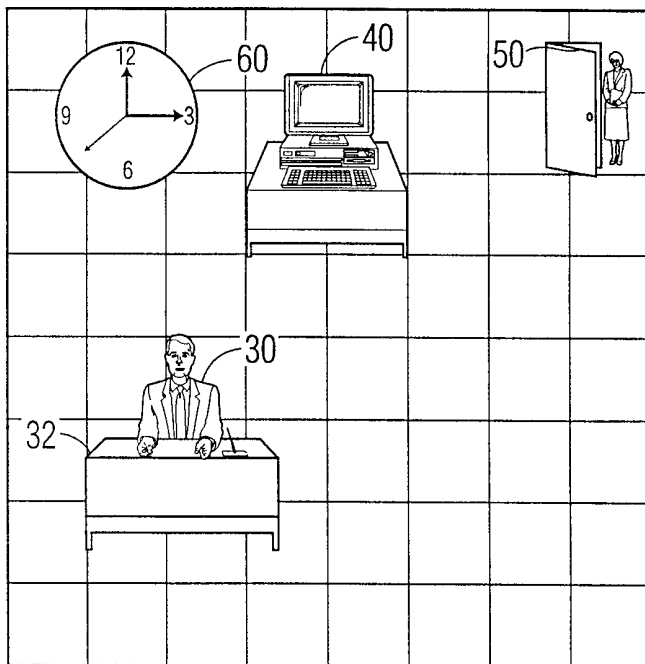


FIG. 2B

3/6

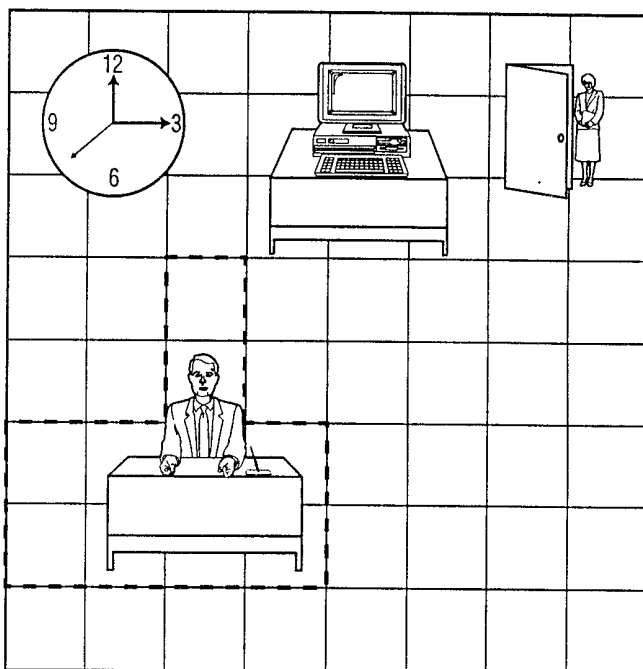


FIG. 3A

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	0	1	0	0	0	0	0
0	0	1	0	0	0	0	0
1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0
0	0	0	0	0	0	0	0

FIG. 3B

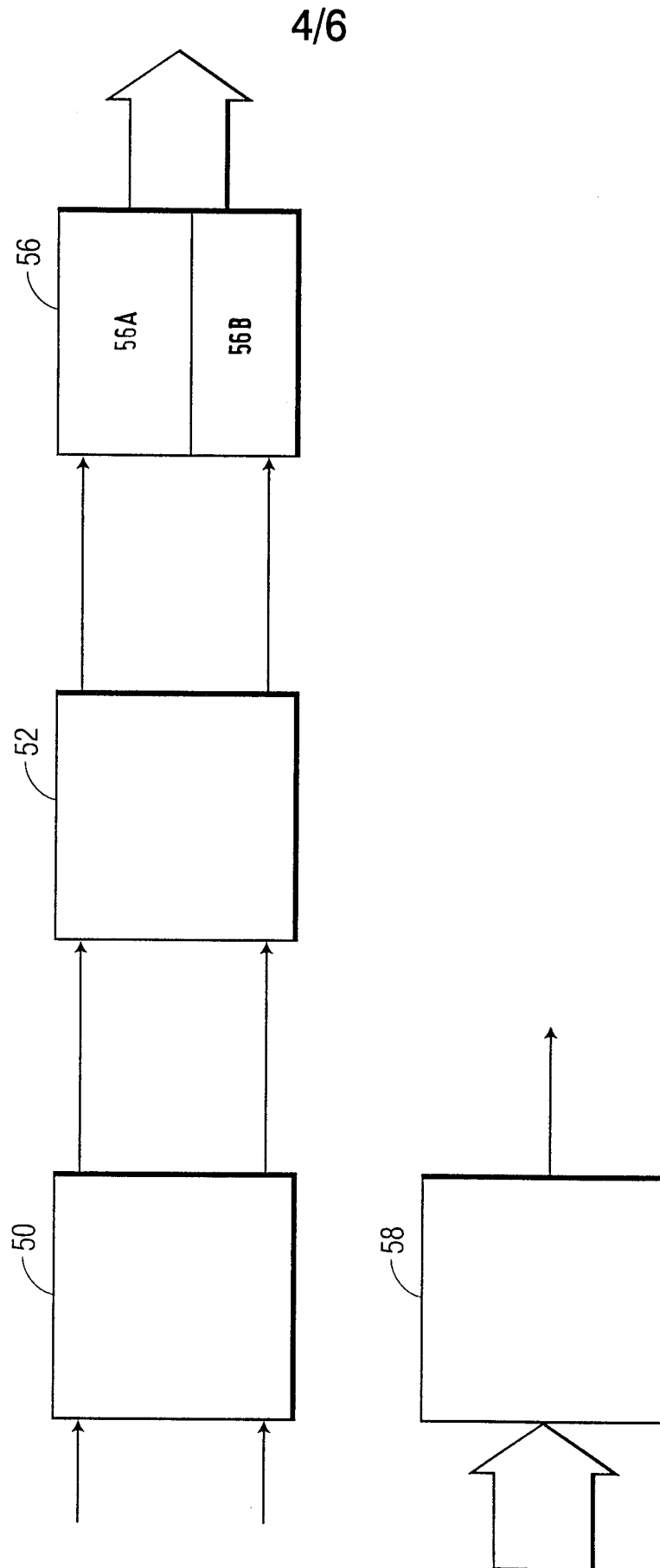


FIG. 4

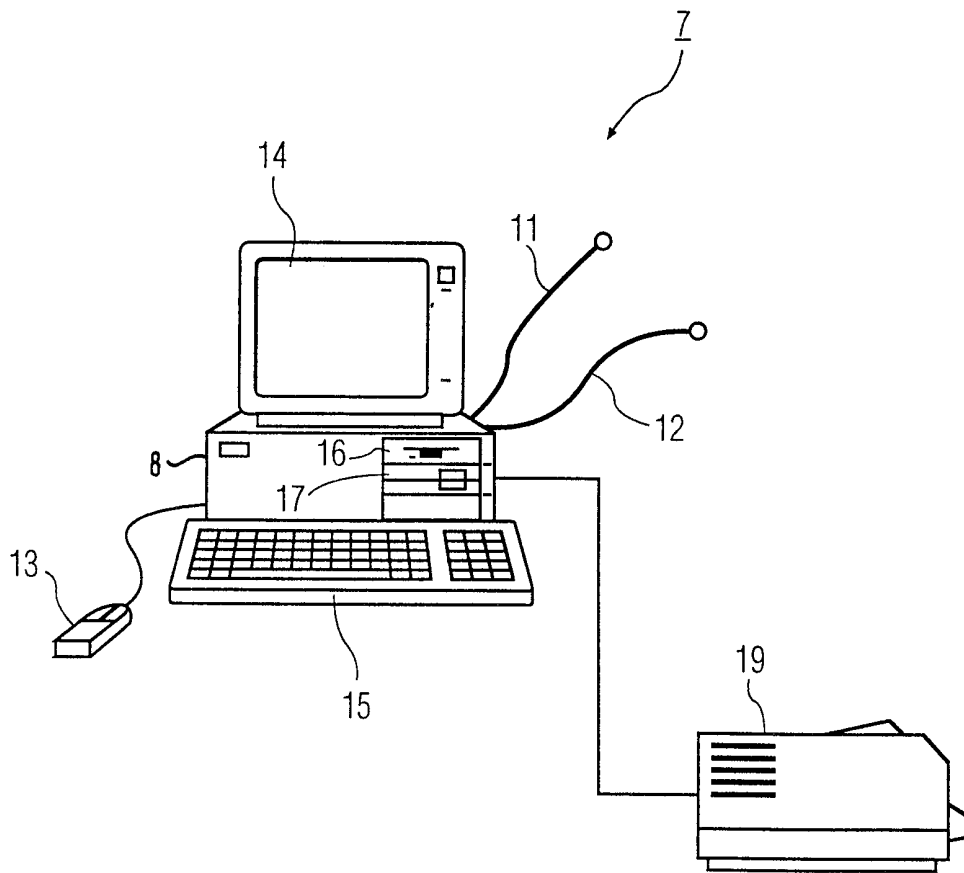


FIG. 5

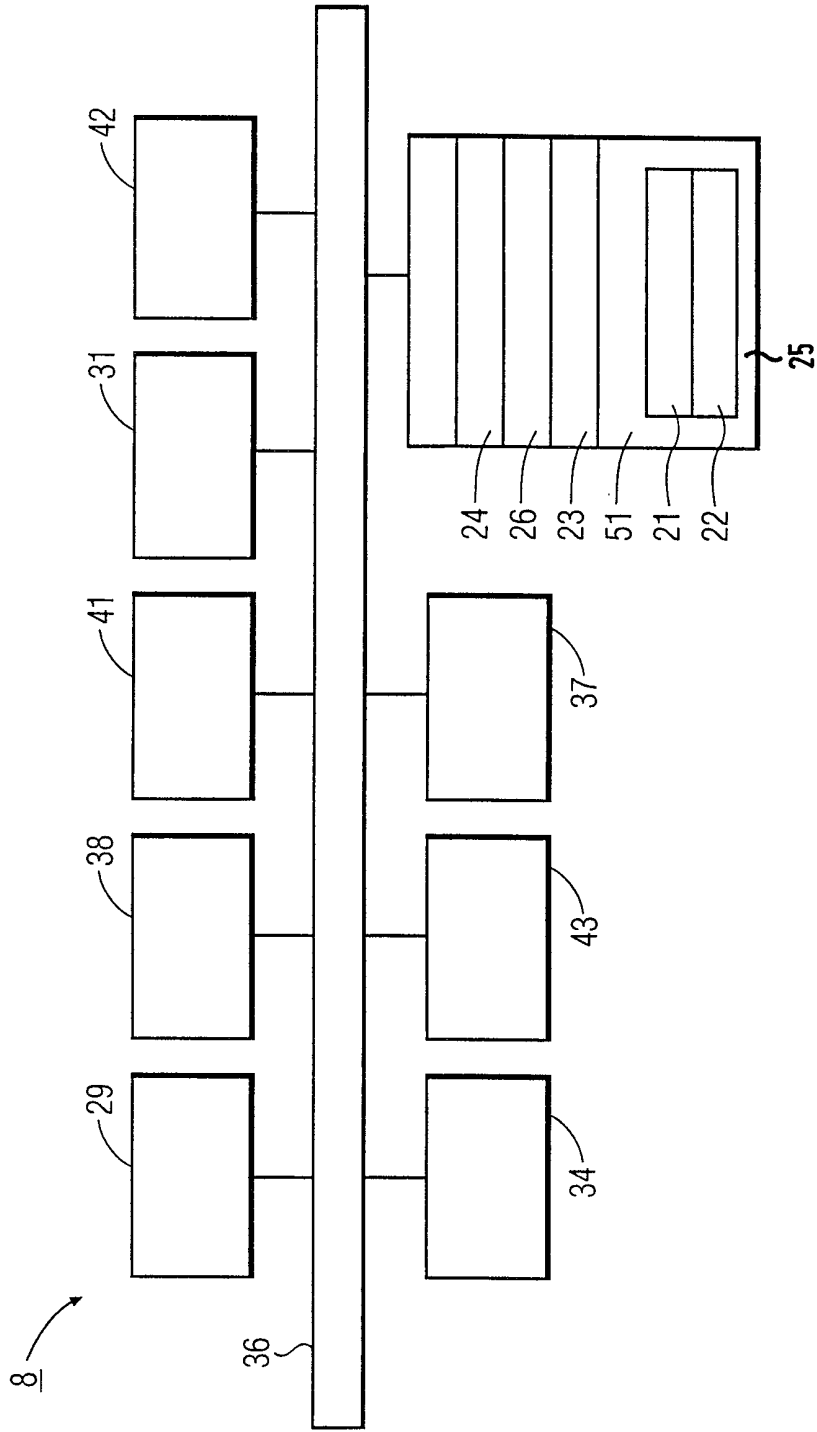


FIG. 6

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 99/08243

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N13/00 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 7 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	<p>WO 97 24000 A (RICHARD ANGUS DUNCAN ;XENOTECH RES PTY LTD (AU)) 3 July 1997 (1997-07-03) page 13, line 19 - line 29 page 17, line 30 -page 18, line 6; figure 15</p>	1-9
A	<p>EP 0 731 608 A (SHARP KK) 11 September 1996 (1996-09-11) page 4, column 5, line 51 - line 56 page 7, column 11, line 30 - line 51; figure 17</p>	1-9

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 21 February 2000	Date of mailing of the international search report 28/02/2000
--	---

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 De Paepe, W
--	--

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 99/08243

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>TZOVARAS D ET AL: "THREE-DIMENSIONAL CAMERA MOTION ESTIMATION AND FOREGROUND/BACKGROUND SEPARATION FOR STEREOSCOPIC IMAGE SEQUENCES" OPTICAL ENGINEERING, US, SOC. OF PHOTO-OPTICAL INSTRUMENTATION ENGINEERS. BELLINGHAM, vol. 36, no. 2, 1 February 1997 (1997-02-01), pages 574-579, XP000686883 ISSN: 0091-3286 the whole document</p>	1-9

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No PCT/EP 99/08243
--

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
WO 9724000	A	03-07-1997		AU 1088497 A	17-07-1997
				CA 2241078 A	03-07-1997
				CN 1209248 A	24-02-1999
				EP 0868818 A	07-10-1998
EP 0731608	A	11-09-1996		JP 8251583 A	27-09-1996
				US 5978515 A	02-11-1999
				US 5815601 A	29-09-1998