PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G09B 5/00	A1	(11) International Publication Number: WO 99/35631
		(43) International Publication Date: 15 July 1999 (15.07.99)

(21) International Application Number: PCT/US99/00407

(22) International Filing Date: 7 January 1999 (07.01.99)

(30) Priority Data:
60/070,626 7 January 1998 (07.01.98) US

(71) Applicant: ELECTRIC PLANET, INC. [US/US]; 3200 Ash Street, Palo Alto, CA 94304 (US).

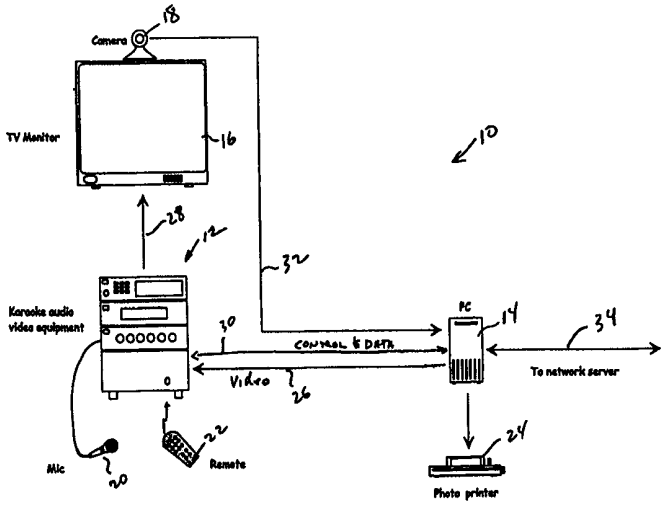
(72) Inventors: KUMAR, David; 810 Ladis Court, Sunnyvale, CA 94086 (US). AHMAD, Subutai; 946 Colorado Street, Palo Alto, CA 94303 (US).

(74) Agent: COLEMAN, Brian, R.; Hickman Stephens & Coleman, LLP, P.O. Box 52037, Palo Alto, CA 94303-0746 (US).

(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, HR, 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, 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, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: METHOD AND APPARATUS FOR PROVIDING INTERACTIVE KARAOKE ENTERTAINMENT



(57) Abstract

An interactive karaoke system includes a microphone (20) developing an audio input from at least one karaoke performer; a camera (18) producing a series of video frames including the at least one karaoke performer; and a karaoke processor system including a video environment and a related audio environment for the karaoke performer. The karaoke processor system is coupled to the camera to create extracted images of the at least one karaoke performer from the series of video frames and to composite the extracted images with a background derived from the viedo environment. The video environment is affected by at least one of a position and a movement of the at least one karaoke performer. A karaoke network includes a local area network, a local karaoke server coupled to the local area network and storing local karaoke content; and a number of karaoke systems coupled to the local area network, each of which can request karaoke content from the local karaoke server.

FOR THE PURPOSES OF INFORMATION ONLY

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

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

Method and Apparatus for Providing Interactive Karaoke Entertainment

Description

Technical Field

This invention relates generally to multimedia entertainment systems, and more particularly to karaoke systems.

Background Art

Karaoke is a form of entertainment, originating in Japan, that features a live singer with pre-recorded accompaniment. Karaoke is a Japanese abbreviated compound word, where "kara" comes from "karappo" meaning empty, and "oke" is the abbreviation of "okesutura," or orchestra. Therefore, karaoke literally means “empty orchestra.” While originating in Japan, the karaoke boom has spread abroad, and is popular in Korea, China and other parts of Southeast Asia, as well as in the U.S. and Europe.

Karaoke music was originally recorded on audio tape, but quickly evolved with the advent of the compact disk, which not only allows rapid, non-serial access new songs, but which also can include multimedia effects such as video and lyrics. Therefore, the advent of the compact disk made it possible to enhance the karaoke experience with video scenes synchronized with the music and the accompanying lyrics.

Using technological innovations such as the video disk, laser disk, and CD graphics, karaoke has grown to be a major entertainment industry. Family-use karaoke sets are also available. However, there is an obstacle to this end of the business: since most Japanese houses stand close each other and are still built of wood, with poor soundproofing, it would be very annoying of the neighbors to sing into an amplified karaoke system at night.

Reacting to the opportunity created by this problem, entrepreneurs created the “karaoke box”, a roadside facility containing closed-door insulated rooms for singing. They are advertised as a place where you can “sing to your heart's content.” The first karaoke box appeared in 1984 in a rice field in the countryside of Okayama Prefecture, just west of the Kansai area. It was built from a converted freight car. Since then, karaoke boxes have been built on unoccupied grounds all over Japan, and in urban areas, karaoke rooms, which consist of compartments made by partitioning and soundproofing rooms in a building, were introduced and set up one after another.

Karaoke is a common form of entertainment for Japanese business people. It is not at all uncommon for workers to drop into a bar with colleagues after work, have a drink, and enjoy singing popular songs to the accompaniment of karaoke. Karaoke has been entertaining people ever since its invention 20 years ago, and has become firmly established in Japanese society.

Today, karaoke is available in a wide variety of formats, suitable for any venue, from a soloist rehearsing up to large crowds at community gatherings. However, a typical karaoke show includes one or two singers, and a possibly a karaoke operator to operate the karaoke equipment. Couples will often enjoy a karaoke session together. The equipment typically includes a player, an amplifier, and a television monitor for the music video. There may be an additional television monitor facing the singers to display the lyrics, or the lyrics can be displayed on the television monitory that is displaying the music video.

While karaoke is very popular, it may be reaching a saturation point, at least in Japan. This is because there are many thousands of karaoke boxes and bars having karaoke systems and, as such, the novelty is beginning to wear off.

One attempt to increase the interest in karaoke is the use of "blue screen" technology which allows a video camera to capture the image of one or more persons standing in front of a blue screen, and inserting the images of those persons into the music video. However, this technology is somewhat cumbersome in that it requires a specialized stage including the blue screen, and in that the karaoke customers are merely superimposed upon a background image of the music video without any interactivity with that background scene.

What would therefore be desirable is a karaoke system which allows new, enhanced, and interactive participation of karaoke customers with their karaoke experience.

Disclosure of the Invention

In one embodiment of the interactive karaoke system of the present invention, a personal computer (PC) is paired with a karaoke audio/video system and a video camera to provide interactivity between the karaoke customers (i.e. the karaoke performers) and the karaoke system. In one aspect of the present invention, images of the karaoke customers are captured with a video camera, processed in the personal computer, and composited into the musical video presentation. However, unlike prior art "blue screen" technologies, no special blue screen is required, and the user can interact with the karaoke content as portrayed on the TV monitor. For example, the karaoke customer may make gestures which to cause the images on the TV monitor to change.

A process for providing interactive karaoke entertainment includes the acts of determining if there is a user initiation and, if so, whether the request of content is local. If not, the content is retrieved. Next, a "frame" of video information is received by the video camera, and background subtraction is performed. Then, there is a tracking analysis, with the results being put into a tracking buffer. A gesture analysis is then performed. Next, the image is "composited" based upon the tracking and gesture analysis and the request of content. The resulting multimedia content is then outputted and, preferably, recorded. The next frame is then retrieved from the video camera and the process is repeated.

The interactive karaoke entertainment system is designed so that it can form a part of a larger network of karaoke entertainment systems. More particularly, a number of interactive karaoke entertainment systems are adapted to coupled to a local area network (LAN) which is served by a local PC server. The local PC server can communicate with an Internet based content server to download content that is not locally available and to upload accounting information.

The local PC server includes the acts of determining whether it has been polled by a content server and, if so, accounting information is transferred to the contents server and other information or software or content can be uploaded or downloaded with the content server. If there has been no polling, the local PC server then determines whether there is a request from a local PC that is coupled to the local area network. If there is, it is determined whether the content is locally available and, if not, the local PC server communicates with remote content server to obtain the desired content. The content is then downloaded to the requesting PC over the local area network and an accounting entry is created at the local PC server reflecting the karaoke customers use of that content.

It will therefore be appreciated that the interactive karaoke system of the present invention will add a new dimension of enjoyment to the karaoke experience. The interactive nature allows the karaoke to transcend a simple performance and take on aspects of an interactive game. This increases the enjoyment and therefore the use of the interactive karaoke systems of the present invention.

Brief Description of the Drawings

Fig. 1 is a representation of interactive karaoke entertainment system in accordance with the present invention;

Fig. 2 is a block diagram of a portion of the system of Fig. 1;

Fig. 3 is a pictorial representation of the personal computer (PC) portion of the system of Fig. 1;

Fig. 4 is a flow diagram illustrating the computer implement operations performed by the personal computer of Fig. 3;

Fig. 4A is a illustration of the compositing act of Fig. 4;

Fig. 4B is an illustration of the compositing act of Fig. 4;

Fig. 5 is a representation of a networked karaoke entertainment system of the present invention;

Fig. 6 is a flow diagram illustrating computer implemented acts performed by the local PC server of Fig. 5;

Fig. 7 is a pictorial representation illustrating one implementation of the interactive karaoke entertainment system of the present invention;

Figs. 8A and 8B illustrate illustrating another, more integrated, implementation of the interactive karaoke entertainment system of the present invention;

Fig. 9 is a more detailed view of the karaoke module used in the DVD and VCD player of Figs. 8B;

Fig. 10 is a block diagram of the vision processor of the karaoke module illustrated in Fig. 9;

Figs. 10A and 10B illustrate a preferred integrated circuit package arrangement for the vision processor of Fig. 10;

Fig. 11 illustrates a typical set-up of an interactive karaoke entertainment system of the present invention;

Fig. 12 illustrates an embodiment of the present invention that utilizes a digital television system;

Figures 13A and 13B are flowcharts showing a preferred embodiment of a method for model-based compositing of the present invention;

Figure 14 is a flowchart showing a process for capturing a frame of an average (background) image;

Figure 15 is a flowchart showing a process for updating the background model;

Figure 16 is a flowchart showing a process for updating the minimum and maximum values for pixels in the average image;

Figure 17A is a replica of a sample background model or average image;

Figure 17B is a replica of a sample input image consisting of the background image including the object being composited;

Figure 18A is a flowchart showing a process for subtracting a background to isolate the object being composited;

Figure 18B shows an initial alpha image of an object being composited after the background subtraction procedure described with respect to Figure 8A is done;

Figures 19A and 19B are flowcharts a process for showing a preferred embodiment of the shadow reduction process;

Figures 20A through 20D are flowcharts showing a process for matching the object to a model of the object made up of object part templates;

Figure 21 is a flowchart showing the process for fitting parts of the object with the templates;

Figure 22A is a flowchart showing a process for eliminating background artifacts and clutter close to the boundary of the object so that such items are not unintentionally composited with the object onto the destination image;

Figure 22B shows an alpha image of the object after the shadow reduction, hole filling, and background clutter procedures have been performed; and

Figure 23 is a flowchart showing a process for blending the object from the input image onto the destination image using the alpha image as a blending coefficient.

Best Modes for Carrying out the Invention

In Fig. 1, an interactive karaoke entertainment system 10 in accordance with the present invention includes karaoke audio and video equipment 12, a personal computer (PC) 14, a TV monitor 16, and a video camera 18. Associated with the karaoke audio and video equipment 12

is an input microphone 20 and a remote control 22. An optional photo-printer 24 can be coupled to the PC 12.

The karaoke audio and video equipment can be provided by any number of vendors. In this embodiment of the entertainment system 10, only the audio portion of the karaoke equipment 12 is used. In other words, as a karaoke customer sings into the microphone 20, the karaoke equipment 12 will amplify and process the sound and play it from speakers (not shown) and/or the TV monitor 16. However, the image for the TV monitor 16, in the present embodiment, is provided by the PC 14 via a video input line 26 to the karaoke equipment 12. Karaoke equipment, such as karaoke equipment 12, typically have an external video input to receive external video information. The combined video and audio is then provided by the karaoke equipment 12 to the TV monitor 16 as illustrated by arrow 28.

In addition, the karaoke equipment 12 typically includes a control and data port (often a serial port) which is coupled to the PC by a bus 30. The output of the video camera 18 is coupled to the PC 14 by a cable 32 and, in alternate embodiments of the invention, may be coupled to the PC by a control cable to allow specialized software and utilities to be loaded into the camera 18 from the PC 14. The photo-printer 24 allows the capture of images that are displayed on the TV monitor that can be printed as photographs, photographic buttons, rubber stamps, etc. There are several vendors for such photo-printers. Preferably, the PC 14 is coupled to a local network server by a local area network (LAN) cable 34.

In Fig. 2, the PC 14 and some peripheral components connected thereto are illustrated in block diagram form. The PC 14 is preferably a standard microcomputer available from a variety of sources including a microprocessor 36 that is coupled to dynamic random access memory (DRAM) 38 and to read only memory (ROM) 40. The microprocessor 36 is also coupled to one or more I/O buses 42 to which peripherals, such as peripheral 44 is coupled. For example, peripheral 44 can be a CD-ROM drive, a DVD drive, a hard disk drive, or any number of input/output (I/O) interfaces. The voice input from the microphone 20 is coupled to the karaoke audio video equipment 12 via a cable 46 and, optionally, to the I/O bus 42 by an audio input card 48. The image input from the video camera 18 is input to a video input card 50 which, also, is coupled to I/O bus 42. The LAN 34 is coupled to the I/O bus 42 by a network card 52. A video output card 54 is coupled to the I/O bus and produces NTSC (and possibly stereo) output for the karaoke audio visual system 12 on the line 26. A parallel card 56 is coupled to the I/O bus 42 and produces photo-printer output signals for the photo-printer 24. An audio card 58 produces an audio output for a power amplifier (not shown) that may be hooked up to loudspeakers (also not shown). A control card 60 an be provided for purposes such as lighting control.

In Fig. 3, a preferred physical implementation of the PC 14 as illustrated. In the present embodiment, the PC 14 is of a "tower" design which provides a multiplicity of I/O slots for he various cards of the present invention. More particularly, a memory expansion board 62, a video card 54, the audio card 58, the camera interface card 50, the network interface card 52, the control card 60, and the parallel card 58 are preferably plugged into I/O slots within the PC tower 14. A keyboard 64 and a mouse 66 are coupled to the PC tower 14 in a conventional manner. Likewise, the PC tower 14 is preferably provided with a CD-ROM drive, a floppy drive, and a pair of hard disks in a conventional fashion. It is preferred to have two hard disks operating in parallel (i.e. "mirroring" each other) for redundancy, since this is the most common area of failure in the PC. By having redundant hard disks drives, the karaoke operator can be virtually assured that the karaoke entertainment system will be continuously operable.

In Fig. 4, the computer implemented process running on the PC 14 is illustrated in flow-diagram. More particularly, the process 68 begins at 70 and, in a decision operation, it is determined whether a user (i.e. a "karaoke customer") is initiating the use of the karaoke entertainment system. This is typically accomplished by using the remote control 22 to activate the selection of a karaoke song. If there is no user initiation, the operation 72 cycles until an initiation is detected. Once an initiation is detected, the process 68 determines whether the requested content is local. By "content" it is meant the requested music video, along with any accompanying multi-media affects and software required for the interactivity with the karaoke entertainment system. If the content is not local, an operation 76 retrieves the content.

Next, in an operation 78, a "frame" of video data is retrieved from the video camera 18. Once the frame has been retrieved and buffered in the memory of the personal computer 14, a background subtraction is performed. Next, a tracking analysis operation 82 is performed and the results are placed in a tracking buffer of the PC 14 and an operation 84. Next, a gesture analysis operation 86 is performed. Subsequently, the images composited based upon the tracking and gesture analysis of operations 82 and 86, respectively, and by the content requested by the karaoke customer. Finally, in operation 90 the resulting composited multi-media content is outputted and, preferably, recorded in a suitable recording device such as a video cassette recorder, recordable CD-ROM, recordable DVD disk, etc. It is the determined in operation 92 if the karaoke customer is done with their particular karaoke session. If so, process control is returned to operation 72 and if not, process control is returned to operation 78 to retrieve a new frame from the video camera.

In Fig. 4A, the operation of the process 68 is illustrated. More particularly, a "frame" 94 of video derived from the camera 18 is loaded into the memory 62 of the PC 14. Those skilled in the art of digital video are well acquainted with the concept of frames. The frame 94 includes the

"true" background image 96 and the images of two karaoke customers or "players" or "performers" 96 and 98. The frame is retrieved by operation 78 and a background subtraction is performed by operation 80 to remove all but the karaoke customers 96 and 98. It should be noted that this background subtraction is accomplished without the use of the awkward blue screen apparatus of the prior art. With the background subtracted, the operation 82 performs the tracking analysis operation 82 to provide a tracked image 100. The compositing operation 88 then composites the karaoke customers 96 and 98 into an interactive environment 102.

The aforementioned technologies permit the karaoke customers 96 and 98 to interact with the environment 102. For example, when karaoke customer 96 raises her hands above her head, animated sparks 104 can be caused to fly from her fingertips. As another example, the grasping of the hand of the karaoke customer 96 by the karaoke customer 98 can be used a gesture which produces the images of hearts 106 in the interactive environment 102. Other gestures or body positions can also interact with various objects 108 in the interactive environment, or change the scene of the interactive environment. Therefore, with the technology of the present invention, karaoke becomes a truly interactive activity, somewhat akin to a game, wherein the multi-media, enhanced reality, and virtual reality effects are possible. It should also be noted that this is a true multi-media experience for the karaoke customers. In addition to video and audio outputs, there are the lyrics 110 of the song, animation effects, etc.

In Fig. 4B, some of the activities of the compositing operation 88 are illustrated in a conceptual form. The operation 88 includes a "media merging" engine 112 which has input, lyrics, audio (e.g. such as from the microphone), sound effects, graphics, animation, camera images, alpha images, tracking information, and gestures. The output is a video stream which provides the video signals for a television monitor, and an audio stream which provides the audio signals for the television monitor and/or separate loudspeakers.

In Fig. 5, a network configuration for the interactive karaoke entertainment system 10 is illustrated. More particularly, a karaoke entertainment system 10 is shown in the lower left hand corner of the page, while a number of other similar systems 10A, 10B, 10C, etc. are also illustrated. Each of the interactive karaoke entertainment systems 10 are coupled to a local area network (LAN) backbone or hub 114 to communicate with a local PC server 116. Preferably, the local PC server 116 is simply a powerful personal computer system.

Also preferably, the local PC server 116 and the interactive karaoke entertainment systems are in fairly close proximity, e.g. within the same building. For example, each of the interactive karaoke entertainment systems 10 can be located in its own, soundproofed room, while the local PC server can be provided in a server or operator room in the same building. The implementation of local area networks are well known to skilled in the art. Preferably, the local

PC server is coupled to a content server 118 by a telephone line 120. The content server 118 includes karaoke "content", which is defined as musical video accompanied by lyrics and any data or software programs required for the interactive use of the "content." The telephone line connecting the local PC server to the content server can be a standard analog telephone line (with the use of appropriate modems at both the local PC 16 and the content server 118), or can be a digital line such as an ISDN line, T1 line, etc. digital line. The advantage of the digital lines are, of course, a significantly higher data transfer rate, with the disadvantage of higher cost. Other data transmission medium are also well known to those skilled in the art.

In the present example, the content server 118 is a "mirror site" that is coupled to a remote content server 122 by, for example, the Internet 124. As is well known to those skilled in the art, a "mirror site" is a site which is updated on a periodic basis, to reflect or "mirror" the contents of another or "master" site, such as content server 122. The purpose of the mirror site 118 is to prevent unnecessary communication delays, especially when transferring large amounts of data, over a relatively slow transmission media such as the Internet 124. For example, one or more content servers can be provided in various cities in Japan while a single content server can be provided in Palo Alto, California. A number of content development systems 124 can then be used to load new content on content server 22 which, as explained previously, creates a mirror image of itself at the content server mirror site 118 via the Internet 124 on a periodic basis.

In Fig. 6, a computer implemented process 126 running on the local PC server 116 begins at 128 and, in an operation 130, it determines whether it has been polled by the content server mirror site 118. It should be noted here that the mirror sites 118 are not required, as the local PC server could communicate directly with the content server 122 via the Internet 124. However, for purposes of efficiency, it is often more desirable to access a local mirror site 118.

If the local PC server 116 determines that it has been polled, it connects with the appropriate content server and transfers accounting information in an operation 132. This accounting information can include the number of times a particular karaoke video has been played and what the appropriate charge for the karaoke operator should be. In addition, an operation 134 can be used to upload and download other information, content, software, etc. Process control is then returned to operation 30.

If operation 130 does not detect a polling from a content server, an operation 136 determines whether here is a request from a local PC, i.e. one of the interactive karaoke entertainment systems 10. If not, process control is returned to operation 130. If there is a request from a local PC, an operation 138 determines whether the requested content is locally available. If not, the content is retrieved from the content server in an operation 140. It should be noted that the local PC server 116 can be connected to the content server mirror site 118 either

on a continuous basis (such as with a ISDN line) or on an "on demand" basis, such as with dial-up modem access. Next, an operation 142 downloads the requested content to the requesting local PC, and in operation 144 creates an accounting entry at the local PC server 116. This accounting entry, along with other data, is what is transferred to the content server in the operation 132.

In Fig. 7, an alternative interactive karaoke entertainment system 10' includes a DVD and VCD player 146, a karaoke adapter 148 of the present invention, a recorder 150, a binocular camera 18', and a television monitor 16. Preferably, the player 146 and adapter 148 are controlled by a remote control 152. In this embodiment, a microphone 20 is coupled to the player 146, and a number of DVD and/or VCD disks 154 are inserted to the player 146. The output of the player 146 goes into the adapter 148, as does the output of the camera 18'.

In this embodiment of the present invention, the adapter 48 performs the functionality described previously with regards to the PC 14 running the computer implemented process 68 of Fig. 4. However, the advantage of this systems is that a separate, dedicated personal computer 14 is not required, since that functionality has been integrated into the adapter 148. The output of the adapter 148 is input into the television monitor and/or loudspeakers (not shown). In addition, a VCR, recordable CD-ROM or recordable DVD recorder 150 can be used to record the output of the adapter 148.

In Figs. 8A and 8B, yet another alternate embodiment of the present invention integrates the functionality of the player 146 with the adapter 158 of Fig. 7. More particularly, a combination DVD/VCD karaoke player 156 is shown in a front elevational view in Fig. 8A and a top plan view with the top lid removed in Fig. 8B.

With primary reference to Fig. 8B, the combined unit 156 includes a VCD and DVD logic module 158, a disk loader 160, a VCD and DVD drive 162, and a karaoke module 164. A power supply 166 is coupled to a source of AC power by a cord and plug 68. The unit 156 has, as inputs, an input 170 from the server, and an input 172 from the camera. The unit 156 has, as outputs, an output 174 to the television monitor 16 and an output 176 to recorder 150. The advantaged of integrating the karaoke module 164 into a DVD and VCD player includes both size and cost reductions.

In Fig. 9, the interactive video karaoke module 164 is shown in a conceptual form. It includes, as inputs, an input 174 for receiving video input from the disk player, and an input 180 for receiving input from the camera 18'. In addition, the module 164 includes an output 182 to the television monitor 16 and an optional output 184 to the camera 18'. It is therefore contemplated that the camera 18' being used with the interactive karaoke entertainment system

10' may be a "smart" camera which can receive programs, data, and commands from the karaoke module 164. The karaoke module 164 includes a vision processor 186 and an ASIC 188 to handle data communications between the karaoke module 164 and the rest of the unit 156.

In Fig. 10, a block diagram of the major components of the vision processor 186 is illustrated. More particularly, the vision processor 186 includes a digital video interface 190, the color processing unit 192, a microcontroller 194, a vision algorithm core 196, a compression unit 198, an ASIC 200 to handle various glue logic functions, memory 202, a Universal Serial Bus (USB) module 204, a memory controller 206, a field programmable gate array (FPGA) controller 208, and a PAL/NTSC module 210. J-Tag circuitry can be included to provide boundary scan capabilities. The input signals (at the digital video interface 190) are processed by the vision processor 186 under microcontroller 194 control. A first output 212 is provided by the USB, and a second output, either for European (PAL) or U.S. (NTSC) video formats is provided at an output 214. External DRAM 216 is coupled to the memory controller 206, and an external FTGA 218 is coupled to the FPGA controller 208. A top plan view of a preferred packaging for the vision processor 186 is shown in Fig. 10A, with a side elevational view taken along line 10B-10B is shown in Fig. 10B.

In Fig. 11, an exemplary use of an interactive karaoke entertainment system 10 is illustrated. The camera 18 of the unit is aimed toward a play area 220 where the karaoke customers may sing and otherwise perform. It is preferred that the customer stay within the play area 220 so as to remain within the "field of sight" 222 of the camera 18. A wired or wireless microphone 20 can be used by the karaoke customers as they sing, and a remote control can be used to activate the system and to select the karaoke music video they wish to accompany. As the karaoke customers moves about in the play area 220 and make pre-determined gestures and poses, they can interact with the video and other content displayed on the television monitor 16.

While the present invention has been described primarily with reference to standard television (analog) monitors, an embodiment of the present invention utilizes the new digital television standards. More particularly, in Fig. 12 a digital television 224 is used as the display unit an interactive karaoke entertainment system 10''. The real time video interaction and vision technologies 226, as disclosed herein provide an interaction between the digital television and a number of peripheral sources. More particularly, the real time video interaction vision technologies provide an interaction with a computer 228, a digital camera 230, a DVD player 232, a VCD player 234, a game console 236, a digital broadcast receiver 238, a video telephone 240, a "set top" box 242, a satellite receiver 244, or a camcorder 246. It will be appreciated by those skilled in the art, the functionality of the interactive karaoke entertainment systems 10 as

described with reference to the analog television monitor are quite transportable to the digital television system as well.

Figures 13A and 13B are flowcharts showing a preferred embodiment of a method for model-based compositing of the present invention. At 300 the system is initialized by setting the variable N to zero. N is the number of iterations the system will perform to create a background model. A background model ("average image") is created or built by averaging several frames of a background image. An average image is essentially an image of a backdrop that does not contain the object that is being composited. For example, an average image could be a simple image of the karaoke room. The background model is essentially a model of the generally static (i.e., non-moving or unchanging) background in which the object being composited, such as a person, will enter. Before the object enters, however, this background model must be created. At 302 the system captures one frame of the background image. With every image frame captured, the system updates the background model as shown block 304. The process of updating the background model involves updating the average image and maintaining a minimum and maximum pixel value chart for each pixel in the average image examined as well as updating the average image. These charts are maintained by comparing the value of the pixel to the minimum and maximum value of that pixel based on previous images. This process is described in greater detail in Figure 16. Blocks 302 and 304 are described in greater detail in Figures 14 and 15 respectively.

At 306 the system determines whether the background model update is complete by checking if the number of iterations (i.e., number of captured frames of the average image) has reached N. As mentioned above, N is the number of iterations the user wants the system to perform in order to create the background model. For example, the user may want the system to go through 30 or a 100 iterations to build the background model depending on how much time the user wants to spend building the model and how accurate the user wants it to be. If the number of iterations has not reached N, the system returns to block 302 where another frame of the average image is retrieved. If the number of iterations has reached N and no more frames are needed to build the background model, the system proceeds to block 308.

At 308 the system retrieves the minimum and maximum values for each pixel in the average image from the minimum and maximum pixel value charts discussed above. At 310 the system computes the tolerance for each pixel value to be the difference between the maximum pixel value and the minimum pixel value for that pixel. For many of the pixels that are stationary, this tolerance will be close to zero. The tolerance for non-static pixels will likely be greater than zero. At 312 the system checks whether there are any more pixels. If

there are more pixels, the process returns to block 308 where the minimum and maximum values for the next pixel in the average image are retrieved. If there are no more pixels in the average image, the process continues with block 314.

At 314 the system captures a frame of an input image. The input image is essentially a background image containing the object being composited. For example, the object could be a human being (e.g., a child) and the background image could be a living room or bedroom. At 316 the system begins creating a new image called the alpha image which contains a representation of the object being composited by first isolating the object. This is first done by subtracting the background from the input image. The background subtraction block is described in greater detail in Figure 18A.

In a preferred embodiment, the system performs a procedure for improving the alpha image referred to generally as shadow reduction at 318. This procedure reduces the effect of shadows cast by the object on other background objects in creating the alpha image. It is described in greater detail in Figures 19A and 19B.

In another preferred embodiment, the system performs another procedure for improving the alpha image called model fitting as shown at 320. In this block the system creates a configuration of templates where each template fits entirely within a particular part of the object. For example, if the object is a person, one template could be for the torso or head. The configuration of templates make up the model which fits the object. The model fitting allows the system to fill up holes in the object while the object is being created in the alpha image. This process is described in greater detail in Figures 20A to 20D. The block following the creation of the templates is simply that of matching each template to its appropriate object part and setting the alpha pixels within the templates to one. This object fill process is shown at 322 and is described in greater detail in Figure 11. Figure 18 is the alpha image, which now contains less holes than the previous alpha image, after the object fill block.

At 324 the system eliminates as much background clutter and artifacts as possible without affecting the object itself. In order to do this it assumes that artifacts greater than a predetermined distance from the closest template (created in block 320 above) is clutter or some type of extraneous artifact not part of the object being composited and ensures that it is not composited onto the destination image.

At 326 the system uses the alpha image to blend the object from the input image onto the destination image. This procedure, known in the art as an alpha blend, uses the value of the pixels in the alpha image to determine which pixels from the input image should be

blended onto the destination image. It is described in greater detail in Figure 23. Once the alpha blend is complete, the system checks whether there are any other images to be composited at 328. If there are, the system returns to block 314 where it captures a frame of the next input image that contains the new object to be composited.

Figure 14 is a flowchart showing a process for capturing a frame of an average (background) image. At 400 the system computes the sum of all differences between the current pixel value of an image just captured and the previous pixel value from an image captured immediately before the current image. In a preferred embodiment, this is done through the formula: $\text{sum} = \sum_i \sum_j P(i,j) - |P_0(i,j)|$, where i and j are coordinates for each pixel. The current pixel is from the image that was captured in the present iteration. Thus, at 400 the system looks at the current value and the previous value for each pixel in the average image frame. At 402 the system prepares for the next iteration by setting the value of the previous pixel value to the current pixel value (P_0=P).

At 404 the system determines whether the sum of the differences computed block 400 is greater than a predetermined threshold value. If not, the system proceeds to 408 where the number of iterations is incremented by one. If the number of iterations reaches N, the process of capturing frames of an average image is complete. If the sum of differences is greater than the threshold value, then there has been too much activity in the background image thereby preventing a background model from being built. This can occur, for example, if a large object passes through the image or if an object in the image is moved. The threshold value is set such that some non-static activity, such as a television screen displaying images or a window showing a road with occasional passing objects, is acceptable and will not prevent a background model from being built. However, significant activity will cause the system to re-initialize itself (setting N to zero) and re-starting the process from block 300 of Figure 13A as shown block 406.

Figure 15 is a flowchart showing a process for updating the background model. The background model is updated, if necessary, with each new background image frame captured as described in Figure 14. Once the number of frames captured equals N, the updating process is complete and the background model has been created. Like all images discussed herein, the average (background) image is comprised of pixels. At 500 the system retrieves a pixel from the average image. At 502 the system updates the average image color pixel value. Each pixel in the average image has an average color value. The average color value for the pixels is determined in a preferred embodiment according to the RGB color scheme, well-known in the art. Other color schemes such as YUB can also be used in another preferred embodiment. A low pixel color value indicates a dark pixel. Thus, a color pixel

value of zero would essentially be a black pixel. Similarly, the brightest pixel will have the maximum color value for a pixel. By way of example, the pixel from the average image corresponding to the pixel retrieved in block 500 can have a color pixel value of .4 and the pixel in the current frame can have a color pixel value of 0.3. If an averaging coefficient of 0.5 is used, the system would just update the average color pixel value for that particular pixel from .4 to .35, *i.e.*, (.5)(.4)+(.5)(.3)=0.35.

In building the background model, the system also maintains a minimum image and a maximum image. The minimum color image and the maximum color image are used to provide a tolerance or variance for each pixel in the background model. A pixel that is part of a stationary object, for example a piece of furniture in the living room, will have little variance or none at all. Any variance for such a pixel would most likely result from camera noise. On the other hand, a pixel that is part of a background image that is dynamic, such as a television screen or the view through a window, will have a greater tolerance. Such pixels are not stationary and the brightness of such pixels can vary while the background model is being updated. For these pixels, the system needs to have a variance or tolerance level. At 504 the system updates the minimum and maximum values for each pixel if needed. The minimum and maximum values for each pixel provides the tolerance for each pixel. Thus, if the new color pixel value is less than the previous minimum color value for that pixel, the minimum value is updated. Similarly, if the color pixel value is greater than the maximum value for that pixel the maximum value is updated. Block 504 is described in greater detail in Figure 6. At 506 the system checks to see whether there are any more pixels in the average image that need to be checked. If there are, the process returns to block 500 where the next pixel from the average image is retrieved. If not, the system returns to the background model update process as shown in block 306 of Figure 13A.

Figure 16 is a flowchart showing a process for updating the minimum and maximum values for pixels in the average image. At 600 the system determines whether the color pixel value of the pixel just retrieved is greater than the maximum value of the corresponding pixel from previous frames. If the current color pixel value is greater, the system sets the maximum color pixel value to the current color pixel value in block 602. Once this is done, the maximum color value for the pixel in that location is set to a new high value. If the current color pixel value is not greater than the maximum value, the system proceeds to block 604. At 604, the same process as in blocks 600 and 602 takes place except the minimum color pixel value is compared to the color pixel value of the pixel just retrieved. If the current color pixel value is less than the minimum value, the system sets the new minimum color pixel value to the current color pixel value in block 606. Once the system determines

whether the minimum or maximum pixel values need to be updated, the system continues the process of updating the background model.

Figure 17A is a replica of a sample background model or average image. It shows a typical karaoke area without the object to be composited. Figure 17B is a replica of a sample input image (discussed below) that consists of the average image including the object being composited, in this example, a figure of a person.

Figure 18A is a flowchart showing a process for subtracting a background to isolate the object being composited. Background subtraction is basically the first process in creating an alpha image of the object that is being composited to a destination image. Each frame of an alpha image is digitally composed such that each pixel is either a 0 or 1 based on whether that pixel is either part of the object. If a pixel has a value of one, that pixel is within the object being composited. Where the value of the alpha pixel is zero, the pixel is not part of the object (i.e., it may be part of the background) and is not composited onto the destination image. As will be described in greater detail below, the alpha image is used in an alpha blend, a technique well-known in the art, to blend the object in the input image with the destination image.

At 800 of Figure 18A the system retrieves a pixel in the input image frame. As mentioned above, the input image contains the background and the object being composited. The system also determines its value and sets it to be the current pixel value. At 802 the system determines whether the absolute value of the difference between the current pixel value and the value of its corresponding pixel from the average image is greater than the tolerance of the current pixel plus a constant. As described block 310 of Figure 13A, each pixel in the average image has a tolerance which is essentially the difference between the maximum and minimum pixel values. If the absolute value of the difference between the current pixel value and the average image pixel value is greater than the tolerance of the current pixel, the system proceeds to block 804 where the system sets the alpha pixel value to one. This indicates that the pixel retrieved from the input image is part of the object because that pixel's color value has changed greater than a "tolerable" amount. A color value change this significant means that there is a new pixel in that position, and that new pixel could be part of the object since the object is the main change in the background model. If the absolute value of the difference is not greater than the tolerance of the current pixel value, the system proceeds to block 806 and simply checks whether there are any more pixels in the input image frame. If there are pixels remaining in the frame the system returns to block 800 and repeats the process. Otherwise, the background subtraction process is complete. It should be noted that in a preferred embodiment, the alpha image is initially set to all zeros and the value for each pixel in the alpha image is changed to one only if the corresponding pixel in the input image

frame is determined to be part of the object that is being composited. Otherwise the value of the alpha pixel is unaffected.

The system provides a means for recognizing gestures, positions, and movements made by one or more subjects (karaoke singers) within a sequence of images and performing an operation based on the semantic meaning of the gesture. In a preferred embodiment, a subject, such as a human being, enters the viewing field of a camera connected to a computer and performs a gesture. The gesture is then examined by the system one image frame at a time. Positional data is derived from the input frame and compared to previously derived data representing gestures known to the system. The comparisons are done in real time and the system can be trained to better recognize known gestures or to recognize new gestures.

In a preferred embodiment, a computer-implemented gesture recognition system is described. A background image model is created by examining frames of an average background image before the subject that will perform the gesture enters the image. A frame of the input image containing the subject, such as a human being, is obtained after the background image model has been created. The frame captures the person in the action of performing the gesture at one moment in time. The input frame is used to derive a frame data set that contains particular coordinates of the subject at that given moment. These sequence of frame data sets taken over a period of time is compared to sequences of positional data making up one or more recognizable gestures *i.e.*, gestures already known to the system. If the gesture performed by the subject is recognizable to the system, an operation based on the semantic meaning of the gesture may be performed by the system.

The gesture recognition procedure may include a routine setting its confidence level according to the degree of mismatch between the input gesture data and the patterns of positional data making up the system's recognizable gestures. If the confidence passes a threshold, a material is considered found.

The gesture recognition procedure may further include a partial completion query routine that updates a status report which provides information on how many of the requirements of the known gestures have been met by the input gesture. This allows queries of how much or what percentage of a known gesture is completed by probing the status report. This is done by determining how many key points of a recognizable gesture have been met.

The gesture recognition procedure preferably includes a routine for training the system to recognize new gestures or to recognize certain gestures performed by an individual more efficiently. Several samples of the subject, *i.e.*, individual, performing the new gesture are used by the system to extract the number of key points, the dimensions, and other relevant

characteristics of the gesture. A probability distribution for each key point indicating the likelihood of producing a particular observable output at that key point is also derived. Once a characteristic data pattern is obtained for the new gesture, it can be compared to patterns of previously stored known gestures to produce a confusion matrix. The confusion matrix describes possible similarities between the new gesture and known gestures as well as the likelihood that the system will confuse these similar gestures.

It will therefore be appreciated that an interactive karaoke system of the present invention includes a microphone developing an audio input from at least one karaoke performer, a camera producing a series of video frames including the at least one karaoke performer, and a karaoke processor system including a video environment and a related audio environment for the karaoke performer. The karaoke processor system is coupled to the camera to create extracted images of the at least one karaoke performer from the series of video frames and to composite the extracted images with a background derived from the video environment, where the video environment is affected by at least one of a position and a movement of the at least one karaoke performer as detected, for example, by a gesture recognizer. The karoake performer image can be recognized for position, movement, and/or semantic content either before or after image extraction from the background.

It will also be appreciated that the present invention includes a karaoke network including a local area network, a local karaoke server coupled to the local area network and storing local karaoke content; and a plurality of karaoke systems coupled to the local area network, each of which can request karaoke content from the local karaoke server. Preferably, the karaoke network also includes a distal content server system coupled to the local karaoke server.

CLAIMS

1. An interactive karaoke system comprising:

a microphone developing an audio input from at least one karaoke performer;

a camera producing a series of video frames including said at least one karaoke performer; and

a karaoke processor system including a video environment and a related audio environment for said karaoke performer, said karaoke processor system being coupled to said camera to create extracted images of said at least one karaoke performer from said series of video frames and to composite said extracted images with a background derived from said video environment, where said video environment is affected by at least one of a position and a movement of said at least one karaoke performer.

2. An interactive karaoke system as recited in claim 1 wherein said related audio environment is affected by at least one of a position and a movement of said at least one karaoke performer.

3. An interactive karaoke system as recited in claim 1 wherein there are multiple karaoke performers, and wherein said video environment is affected by at least one of the positions and movements of said multiple karaoke performers.

4. An interactive karaoke system as recited in claim 3 wherein said related audio environment is affected by at least one of the positions and movements of said multiple karaoke performers.

5. An interactive karaoke system as recited in claim 1 wherein said karaoke processor system includes a karaoke unit having a microphone input, a control and data input, a video input, an audio output, and a video output, and wherein said karaoke processor system

further includes a digital computer system having a camera input coupled to said camera, a control and data output coupled to said control and data input of said karaoke unit, and a video output coupled to said video input of said karaoke unit.

6. An interactive karaoke system as recited in claim 5 wherein said digital computer system is coupled to a network.

7. An interactive karaoke system as recited in claim 6 further comprising a karaoke server coupled to said network for two-way communication with said digital computer system.

8. An interactive karaoke system as recited in claim 5 further comprising a video display unit coupled to said video output of said karaoke processor system.

9. A method for providing interactive karaoke entertainment comprising:

receiving a plurality of video frames which include images of at least one karaoke performer;

subtracting background images from said video frames to create extracted images of said at least one karaoke performer;

performing an analysis of at least one of the position and motion of said extracted images to provide a visual performer input; and

providing background images with accompanying sound, at least one of which is affected, by said visual performer input.

10. A method for providing interactive karaoke entertainment as recited in claim 9 wherein said background images and accompanying sound are based upon stored content.

11. A method for providing interactive karaoke entertainment as recited in claim 10 further comprising retrieving said stored content from a server.

12. A method for providing interactive karaoke entertainment as recited in claim 9 wherein said analysis includes a tracking analysis and a gesture analysis.

13. A method for providing interactive karaoke entertainment as recited in claim 12 further comprising compositing said extracted images with said background images to create composited images.

14. A method for providing interactive karaoke entertainment as recited in claim 13 further comprising recording said composited images and accompanying sound.

15. A karaoke network comprising:

a local area network;

a local karaoke server coupled to said local area network and storing local karaoke content; and

a plurality of karaoke systems coupled to said local area network, each of which can request karaoke content from said local karaoke server.

16. A karaoke network as recited in claim 15 further comprising a distal content server system coupled to said local karaoke server.

17. A karaoke network as recited in claim 16 wherein said distal content server system includes a connection over a world-wide network system.

18. A karaoke network as recited in claim 16 wherein said distal content server system is coupled to said local karaoke server, at least in part, by a local telephone exchange.

19. A karaoke network as recited in claim 16 wherein said distal content server system includes a mirror site content server coupled to a master site content server.

20. A karaoke network as recited in claim 19 wherein said mirror site content server is coupled to said master site content server by a TCP/IP network.

21. A karaoke network as recited in claim 20 wherein said mirror site content server is coupled to said local karaoke server, at least in part, by a local telephone exchange.

22. A karaoke network as recited in claim 15 wherein said local karaoke server further stores accounting information concerning requests for karaoke content from said plurality of karaoke systems.

23. A karaoke network as recited in claim 15 wherein at least one of said plurality of karaoke systems include:

a microphone developing an audio input from at least one karaoke performer;

a camera producing a series of video frames including said at least one karaoke performer; and

a karaoke processor system including a video environment and a related audio environment for said karaoke performer, said karaoke processor system being coupled to said camera to create extracted images of said at least one karaoke performer from said series of video frames and to composite said extracted images with a background derived from said video environment, where said video environment is affected by at least one of a position and a movement of said at least one karaoke performer.

24. A method for providing networked karaoke entertainment comprising:

determining whether an accounting polling event has occurred and, if so, uploading accounting information to a remote server;

providing a requested karaoke content to a local karaoke unit; and

creating an accounting entry concerning the provision of said karaoke content to said local karaoke unit.

25. A method for providing networked karaoke entertainment as recited in claim 24 further comprising determining whether said requested karaoke content is available locally and, if not, obtaining said requested karaoke content from said remote server.

26. A method for providing networked karaoke entertainment as recited in claim 24 wherein at least one local karaoke unit:

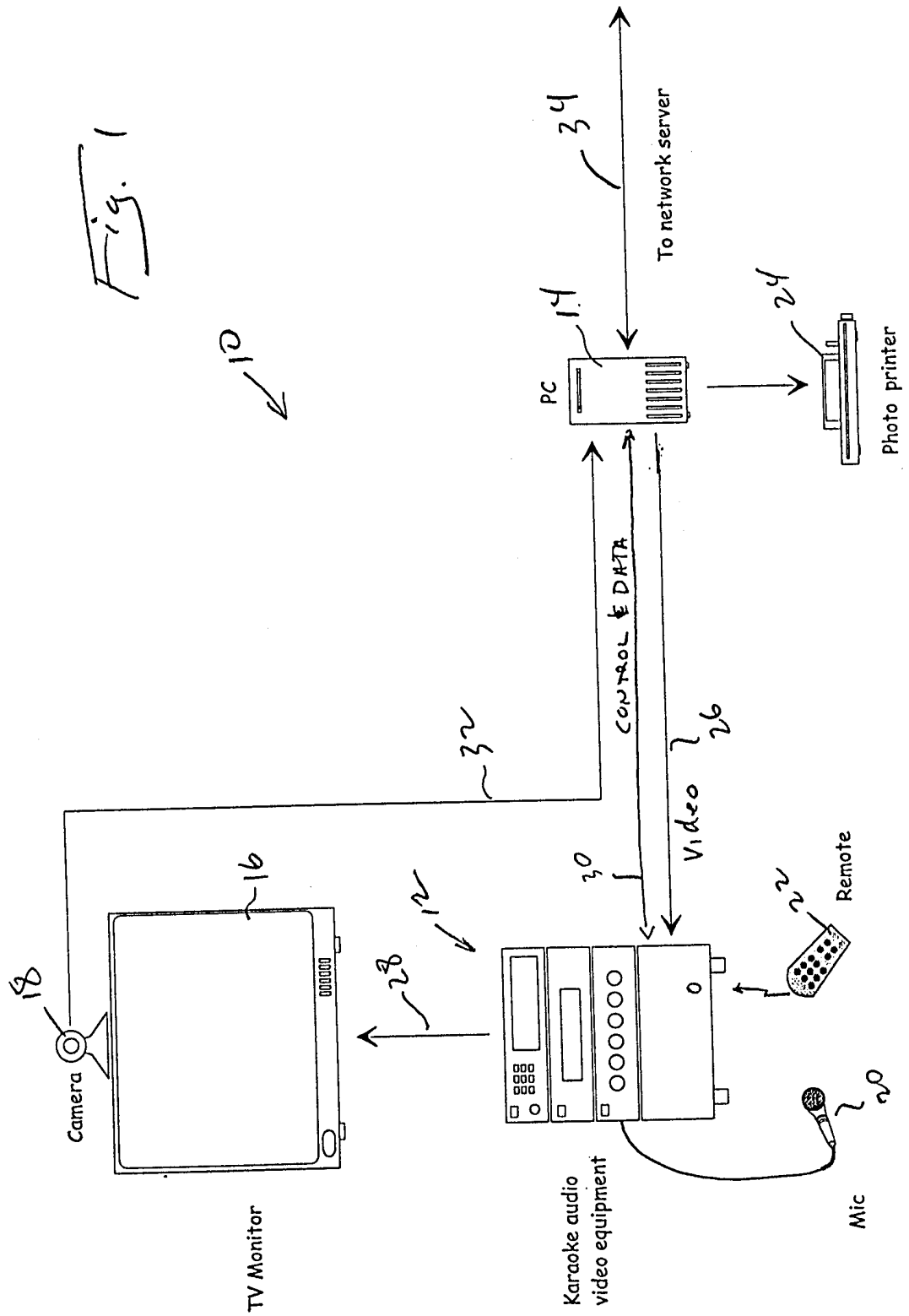
receives a plurality of video frames which include images of at least one karaoke performer;

subtracts background images from said video frames to create extracted images of said at least one karaoke performer;

performs an analysis of at least one of the position and motion of said extracted images to provide a visual performer input; and

provides background images with accompanying sound, at least one of which is affected, by said visual performer input.

Fig. 1


10
34
To network server
14
PC
24
Photo printer
32
CONTROL & DATA
26
Video
30
12
22
Remote
28
16
18
Camera
TV Monitor
Karaoke audio
video equipment
20
Mic

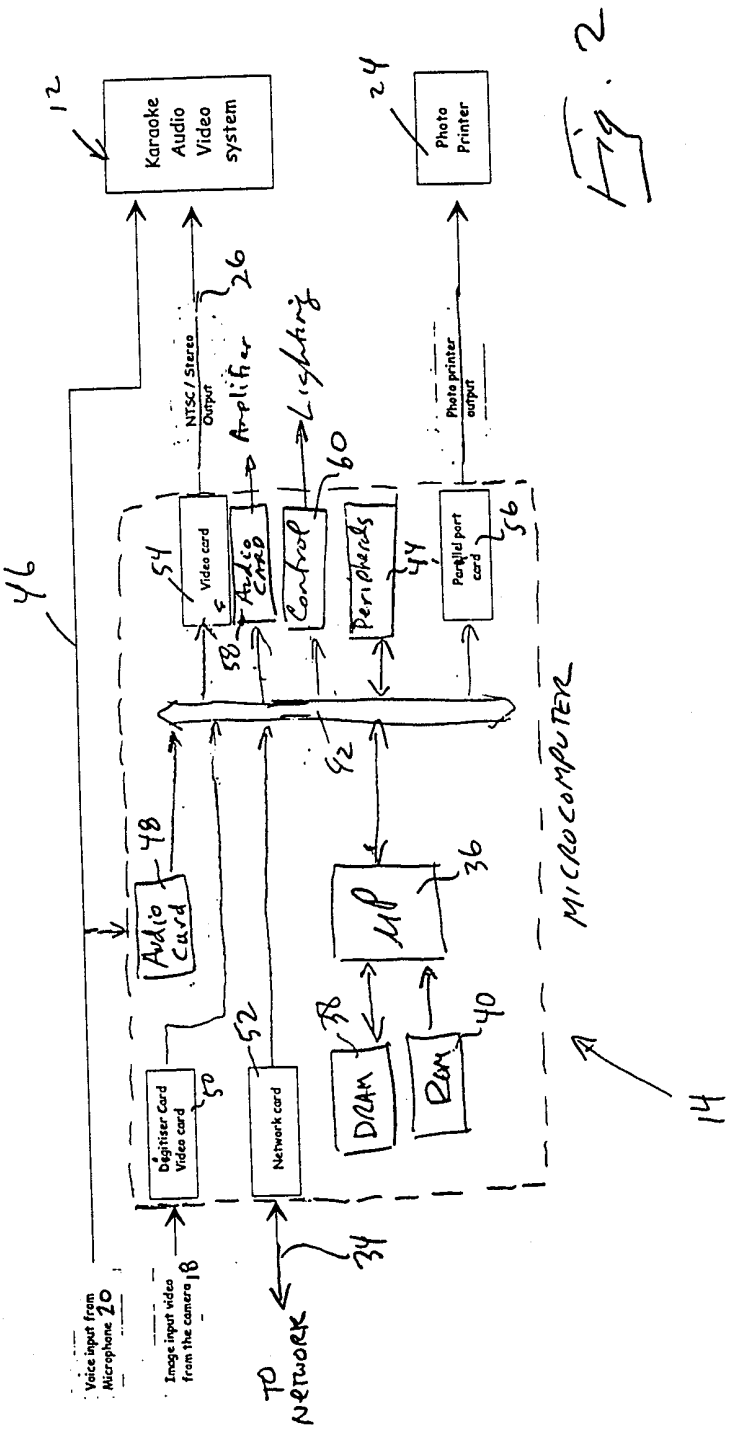

Karaoke Audio Video system
12
Photo Printer
24
NTSC / Stereo Output
Amplifier
26
Lighting
60
Photo printer output
Video card
54
Audio card
58
Control
Peripherals
44
Parallel port card
56
46
92
Audio Card
48
MP
36
DRAM
38
ROM
40
Digitiser Card Video card
50
Network card
52
34
Voice input from Microphone 20
Image input video from the camera 18
TO NETWORK
MICROCOMPUTER
14

Fig. 2

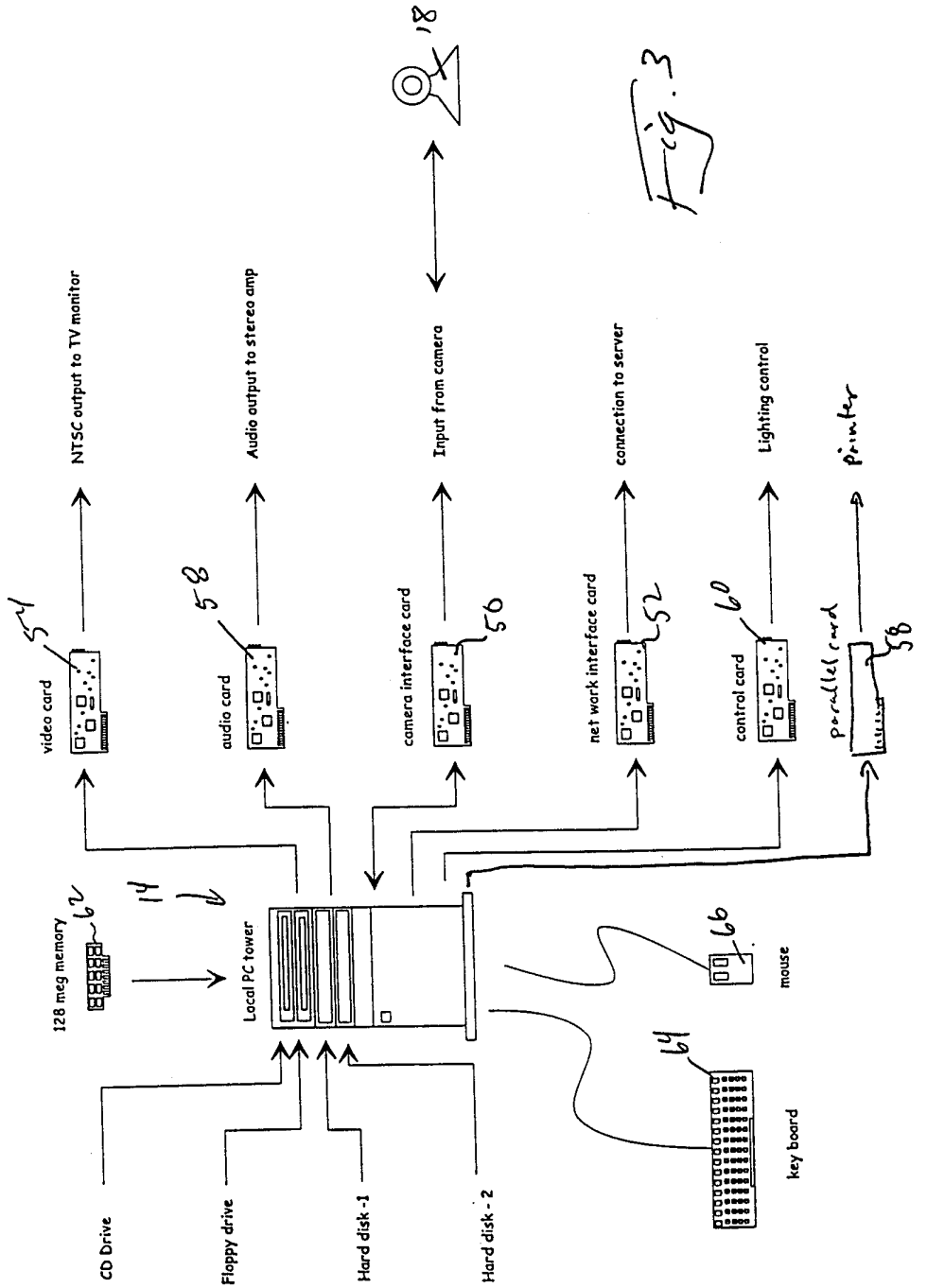

NTSC output to TV monitor
Audio output to stereo amp
Input from camera
connection to server
Lighting control
Printer
video card
audio card
camera interface card
net work interface card
control card
parallel card
128 meg memory
Local PC tower
mouse
key board
CD Drive
Floppy drive
Hard disk -1
Hard disk - 2

Fig. 3

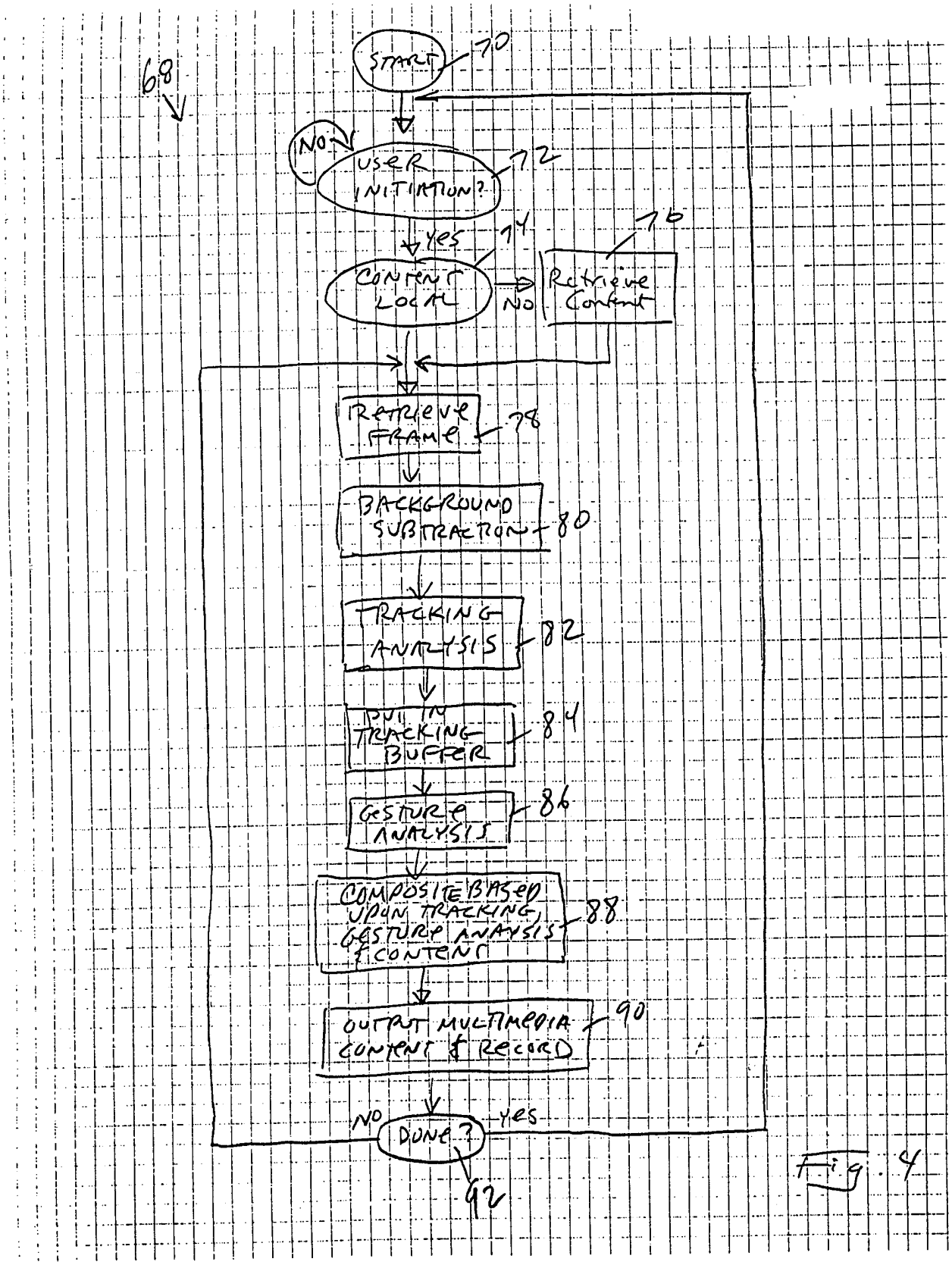

68
START
70
NO
USER INITIATION?
72
YES
74
CONTENT LOCAL
NO
76
Retrieve Content
RETRIEVE FRAME
78
BACKGROUND SUBTRACTION
80
TRACKING ANALYSIS
82
INPUT TO TRACKING BUFFER
84
GESTURE ANALYSIS
86
COMPOSITE BASED UPON TRACKING, GESTURE ANALYSIS & CONTENT
88
OUTPUT MULTIMEDIA CONTENT & RECORD
90
NO
DONE?
YES
92

Fig. 4

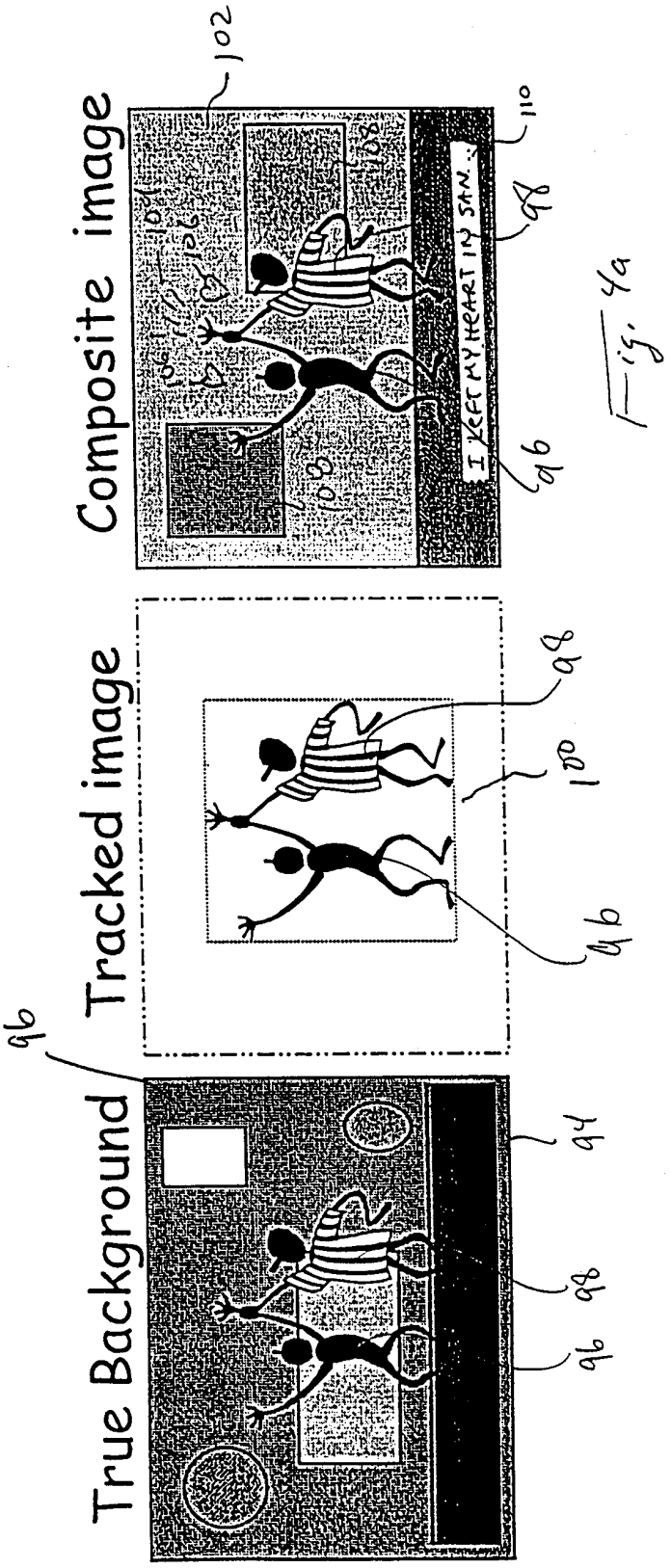

Composite image
102
110
98
96
I KEEP MY HEART IN SAN...
Tracked image
98
100
96
True Background
96
94
98
96

Fig. 4a

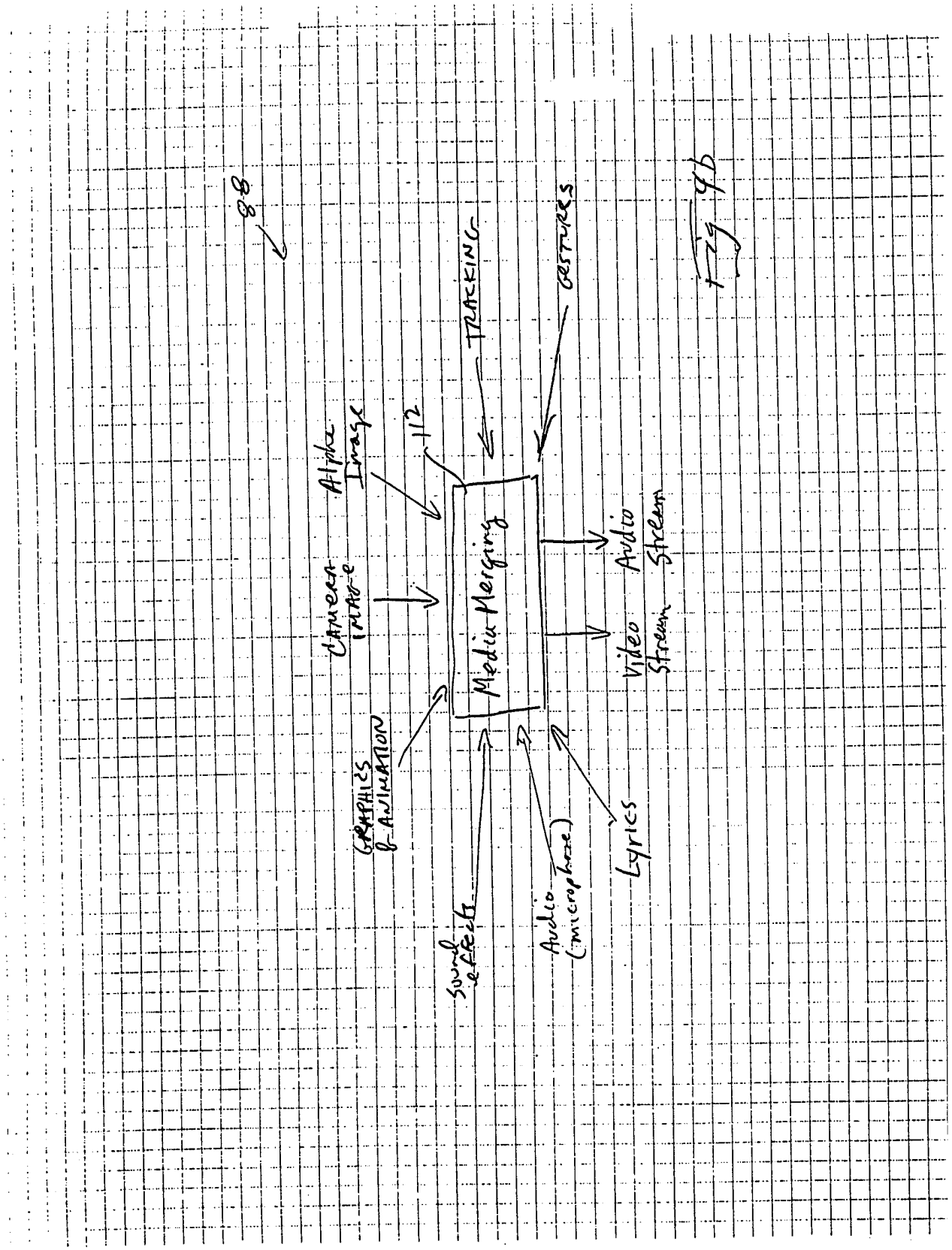

88
Camera Interface
Alpha Image
112
Graphics & Animation
Sound effects
Audio (microphone)
Media Merging
Tracking
Gestures
Lyrics
Video Stream
Audio Stream

Fig. 4b

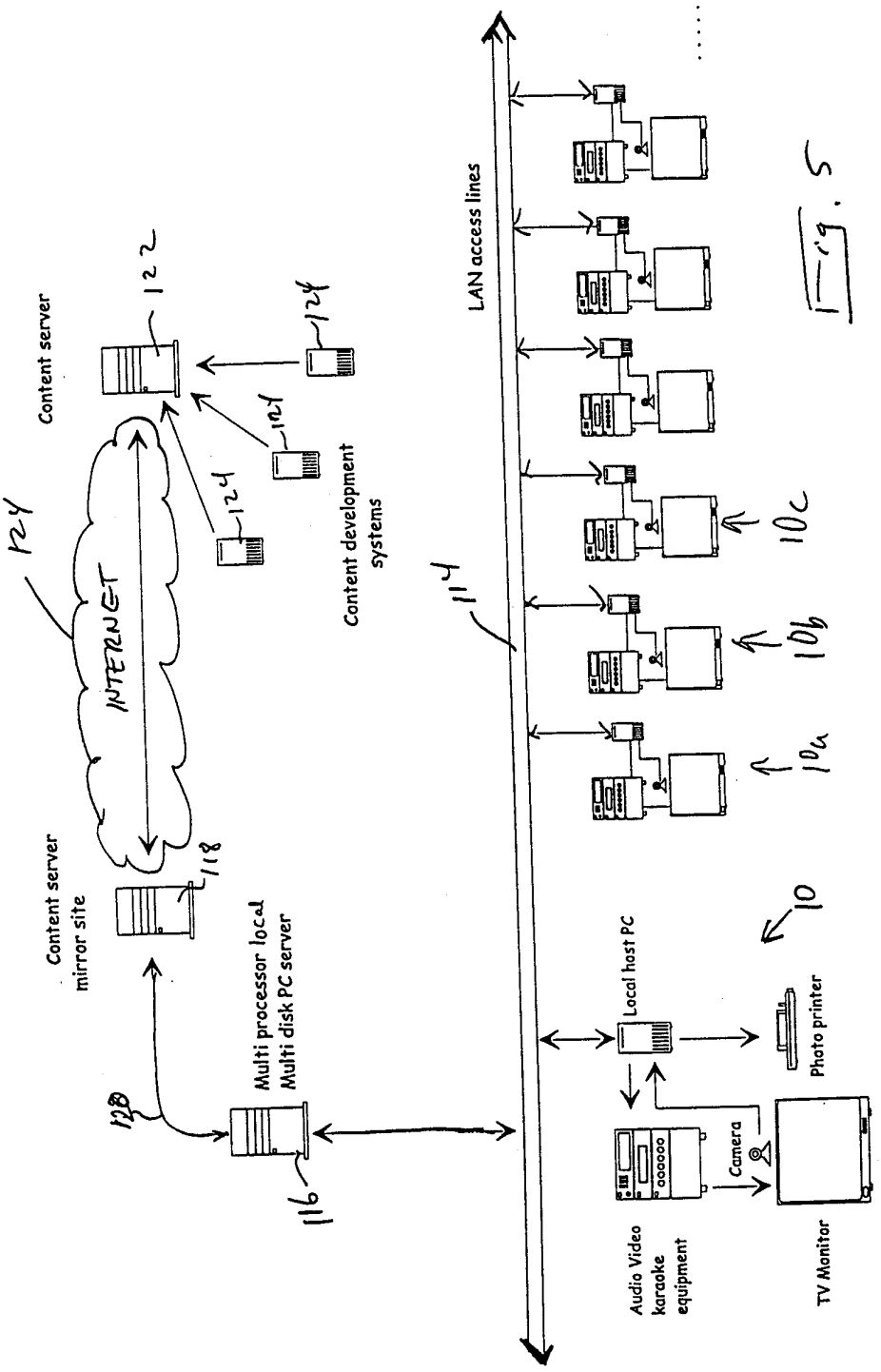

Content server
122
124
124
124
Content development systems
INTERNET
124
Content server mirror site
118
Multi processor local Multi disk PC server
120
116
LAN access lines
114
10a
10b
10c
10
Local host PC
Photo printer
Camera
Audio Video karaoke equipment
TV Monitor

Fig. 5

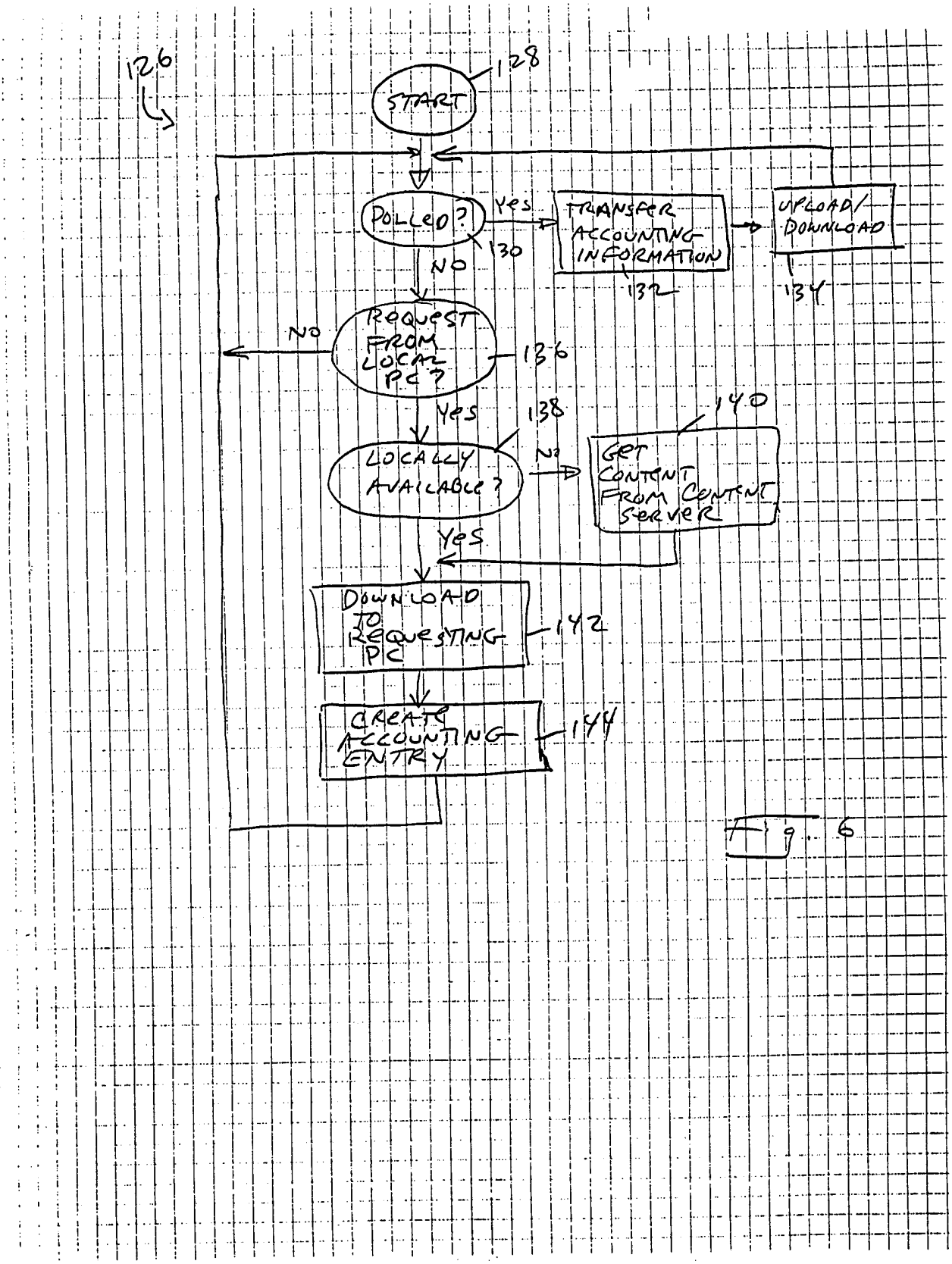

126
START
128
POLLED?
Yes
130
TRANSFER ACCOUNTING INFORMATION
132
UPLOAD/ DOWNLOAD
134
NO
REQUEST FROM LOCAL PC?
NO
136
Yes
LOCALLY AVAILABLE?
138
NO
GET CONTENT FROM CONTENT SERVER
140
Yes
DOWNLOAD TO REQUESTING PC
142
CREATE ACCOUNTING ENTRY
144

Fig. 6

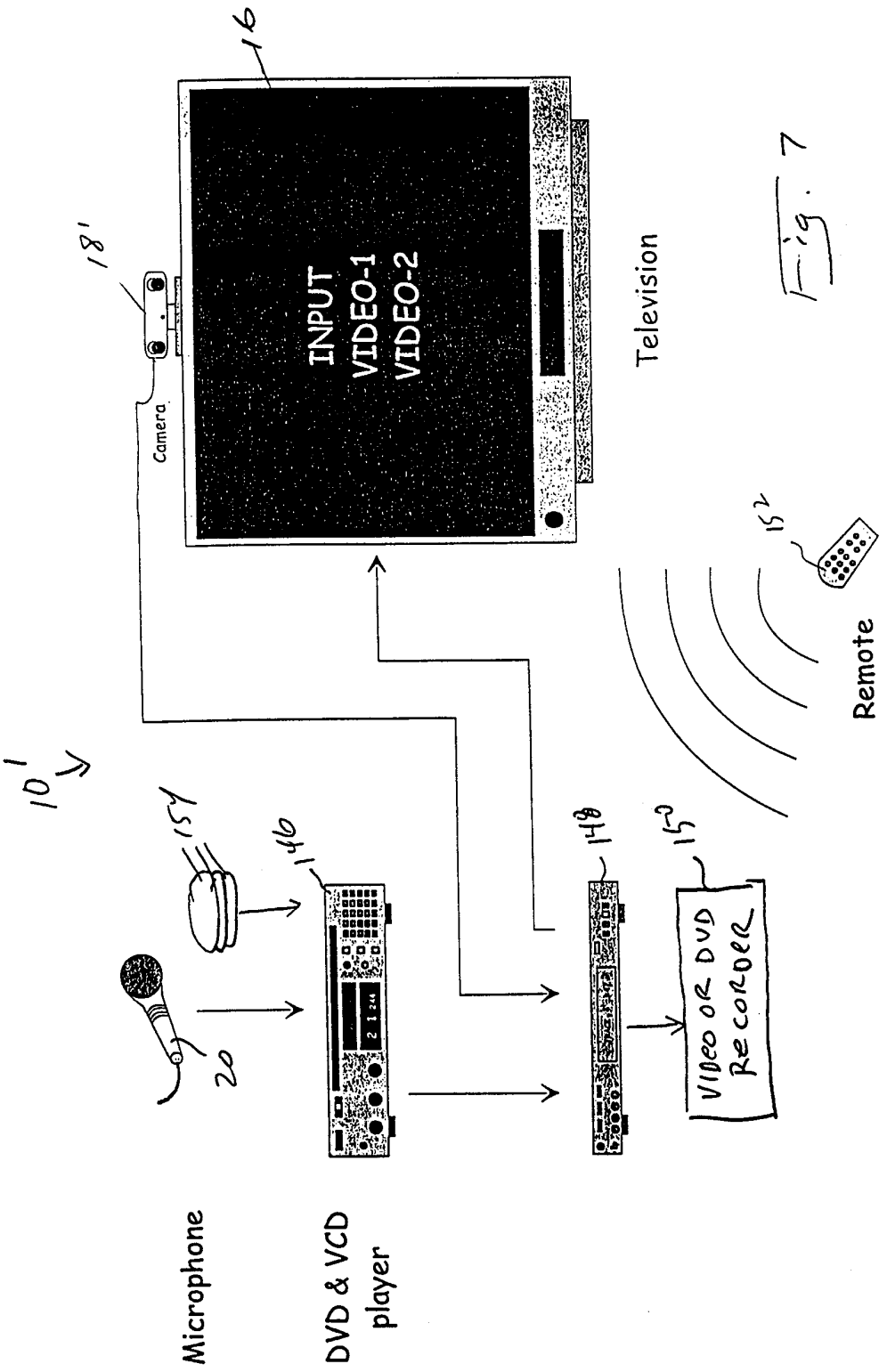

16
18'
Camera
INPUT
VIDEO-1
VIDEO-2
Television
152
Remote
10'
154
146
148
150
VIDEO OR DVD RECORDER
20
Microphone
DVD & VCD player

Fig. 7

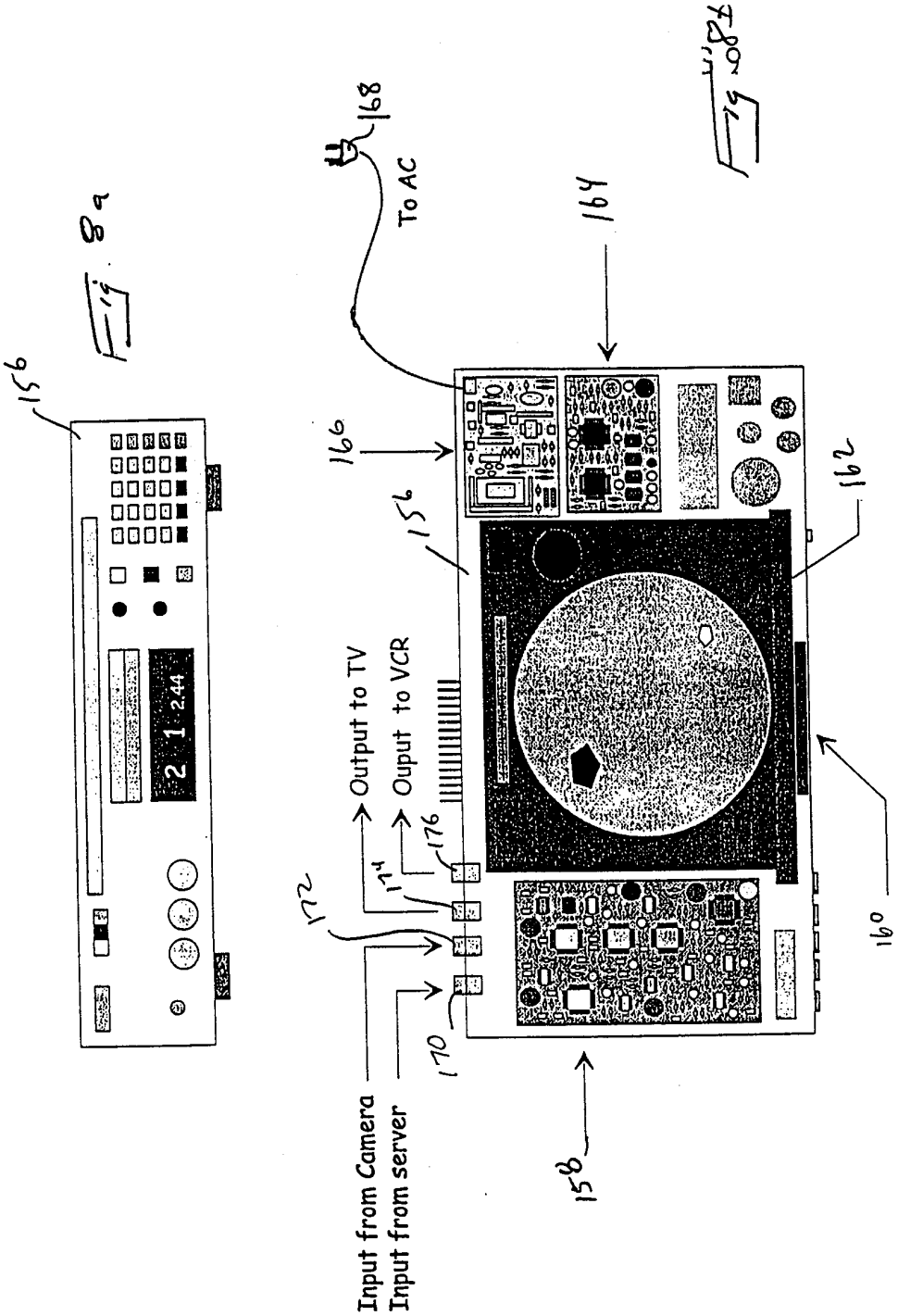


Fig. 8a
Fig. 8b
156
168
To AC
164
166
156
162
Output to TV
Ouput to VCR
176
174
172
170
160
Input from Camera
Input from server
158

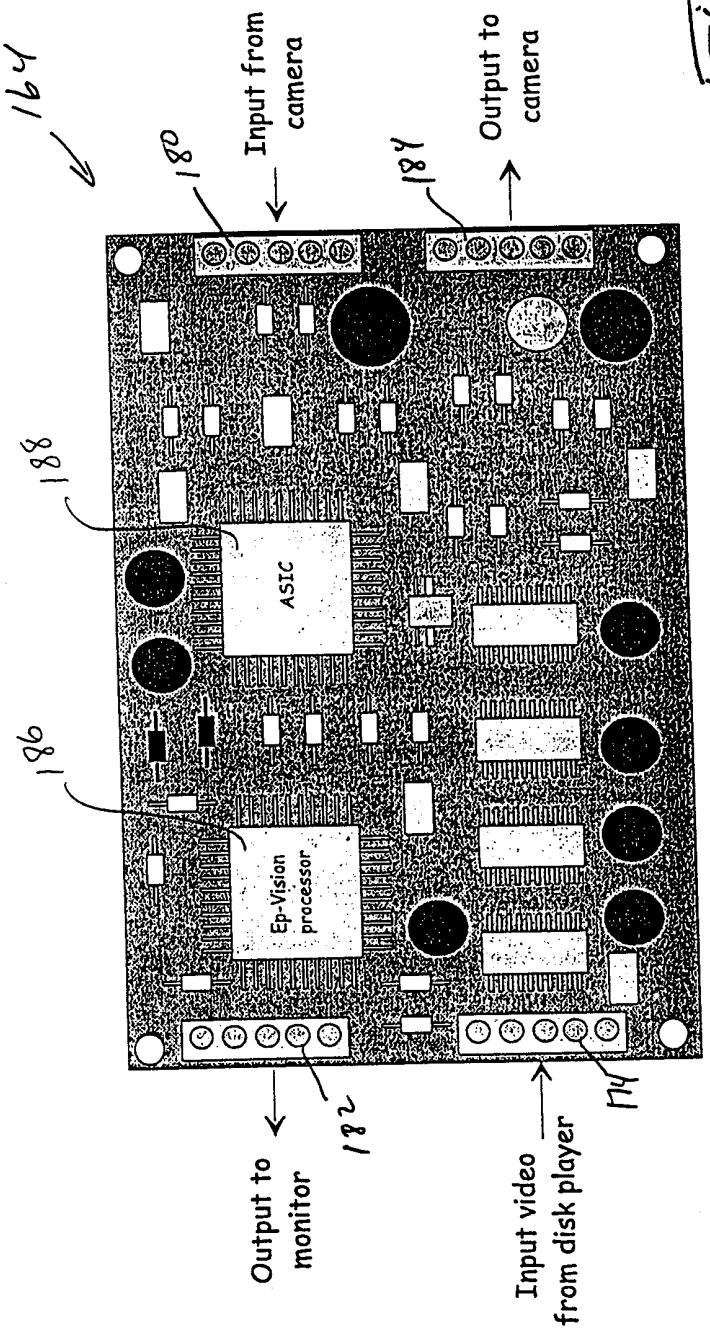
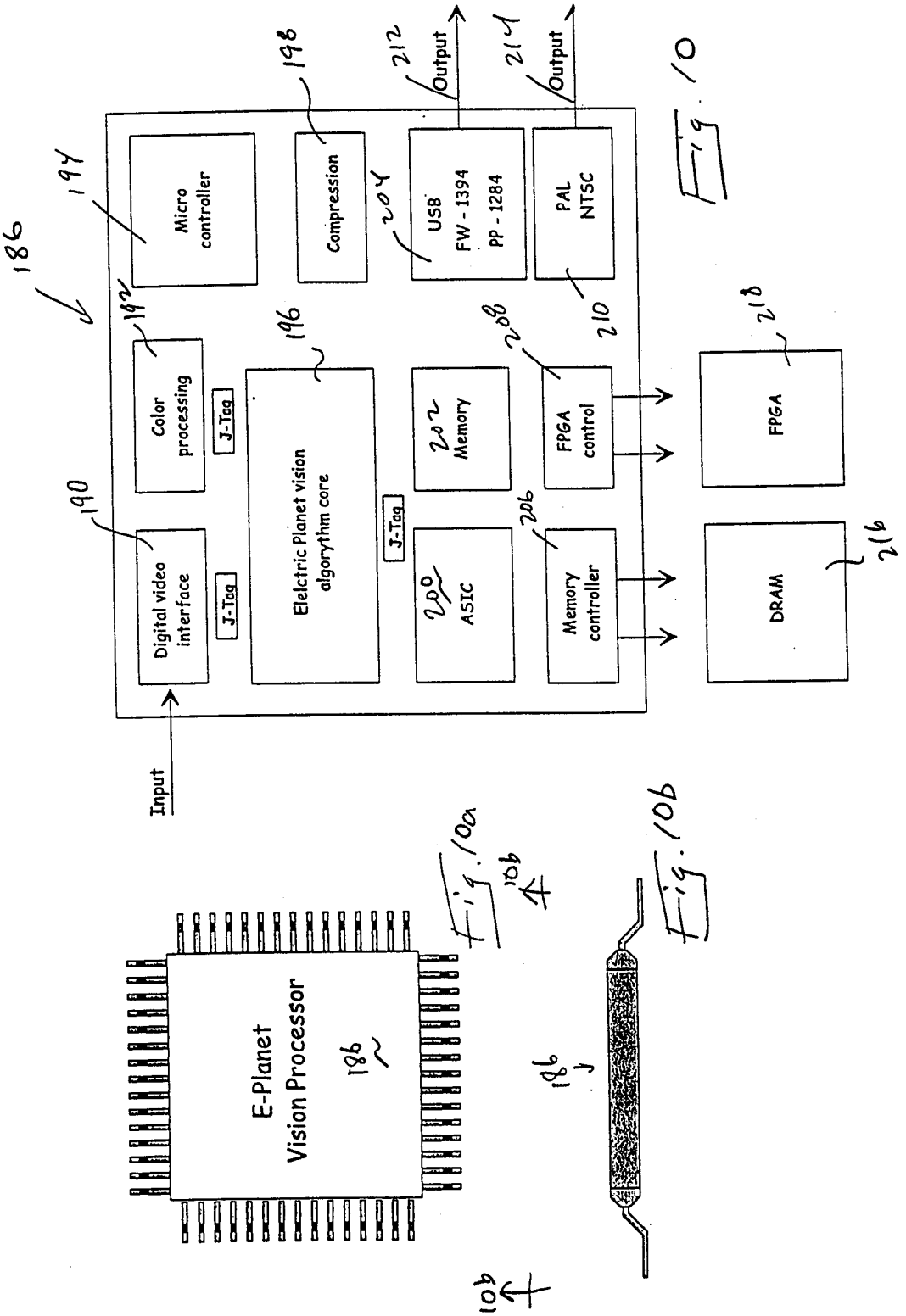

164
180
Input from camera
184
Output to camera
188
ASIC
186
Ep-Vision processor
182
Output to monitor
174
Input video from disk player

Fig. 5


186
194
Micro controller
198
Compression
212
Output
214
Output
204
USB FW - 1394 PP - 1284
PAL NTSC
210
192
Color processing
J-Tag
196
Elelctric Planet vision algorythm core
J-Tag
202 Memory
208
FPGA control
218
FPGA
190
Digital video interface
J-Tag
200 ASIC
206
Memory controller
216
DRAM
Input
Fig. 10
E-Planet Vision Processor
186
401
Fig. 10a
196
401
Fig. 10b

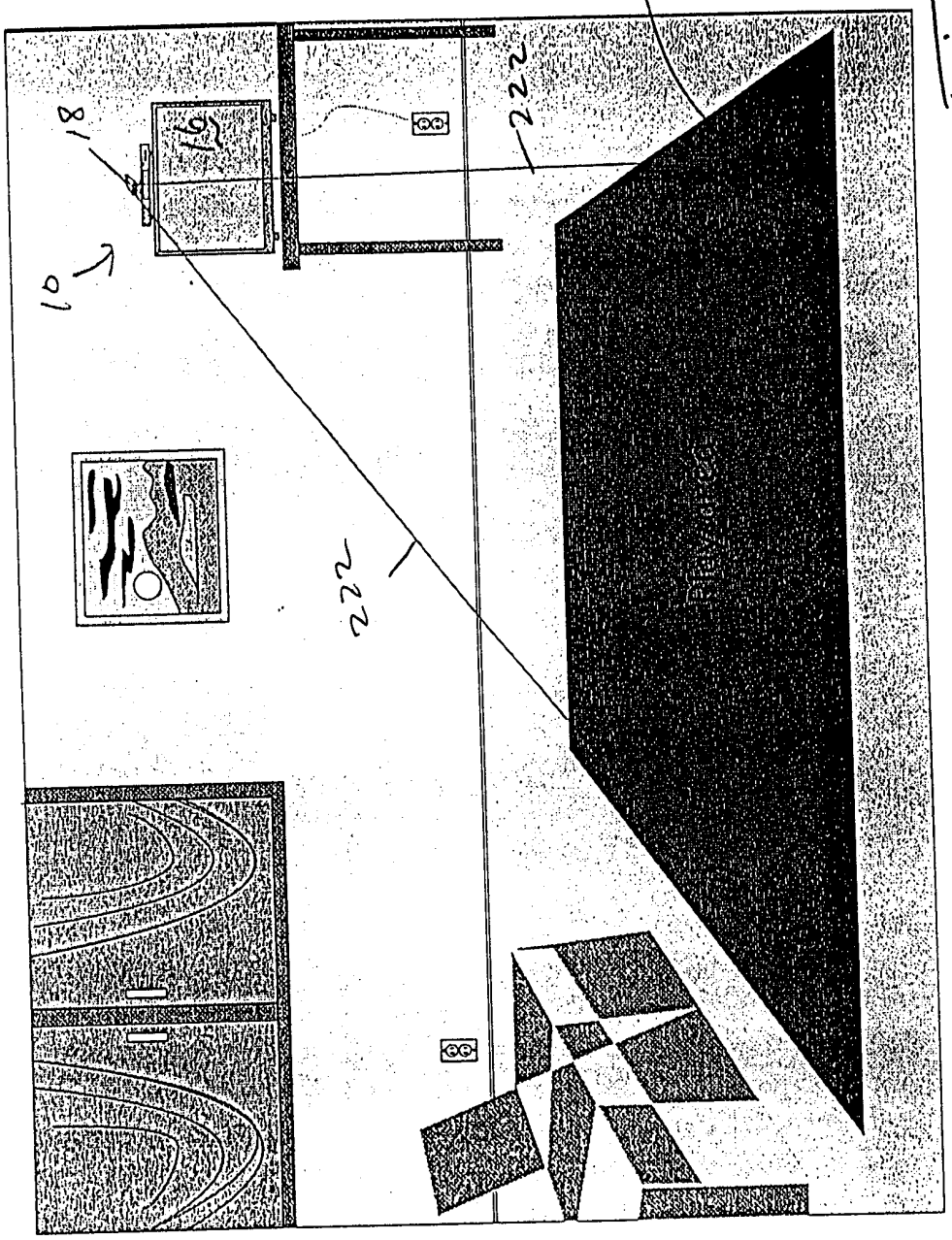


Fig. 11

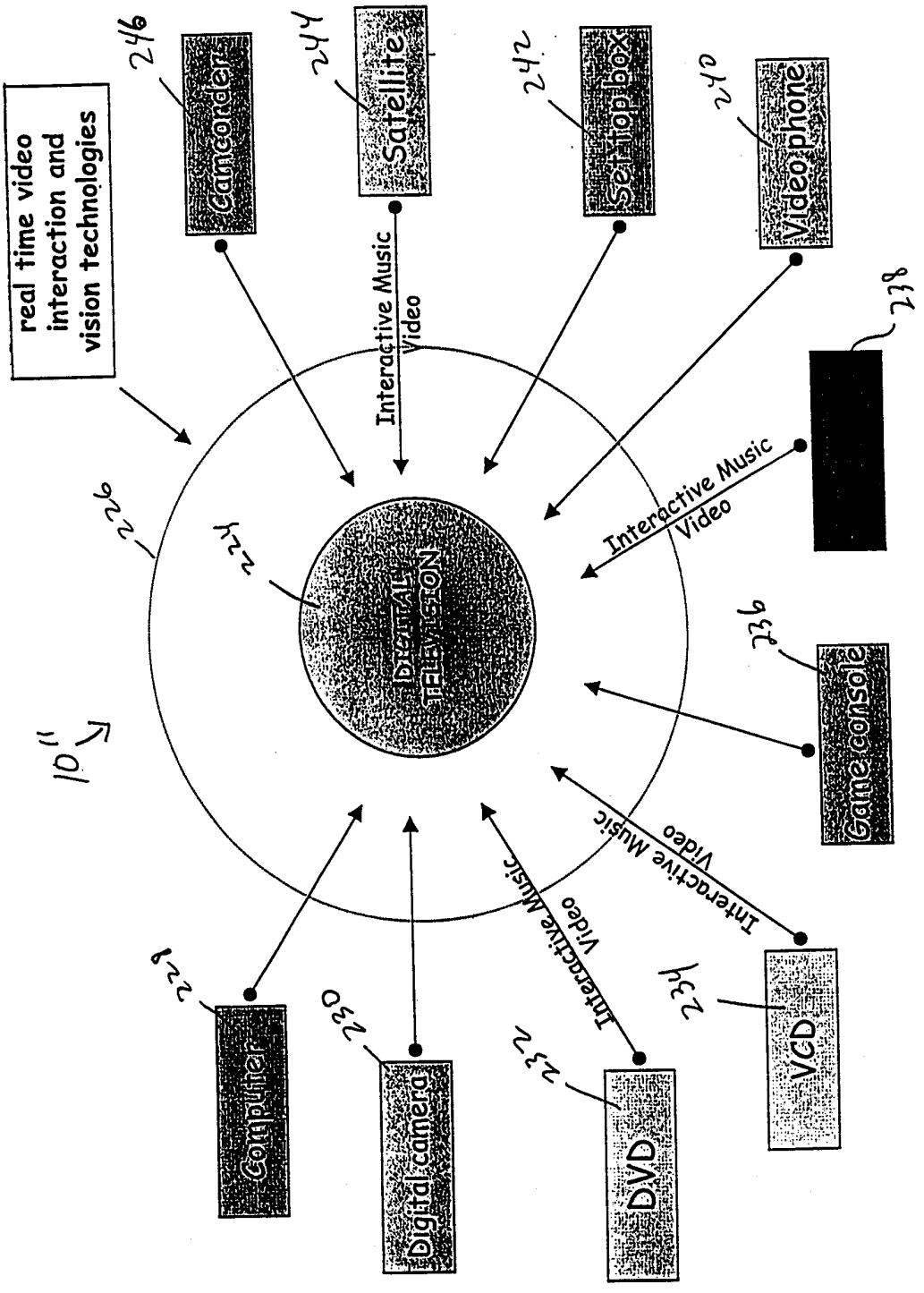

real time video interaction and vision technologies
Camcorder
246
Satellite
244
Set-top box
242
Videophone
240
238
Interactive Music Video
Interactive Music Video
DIGITAL TELEVISION
224
222
Game console
236
Interactive Music Video
Interactive Video
Interactive Music
VCD
234
DVD
232
Digital camera
230
Computer
228
10

Fig. 12

FIG. 3A

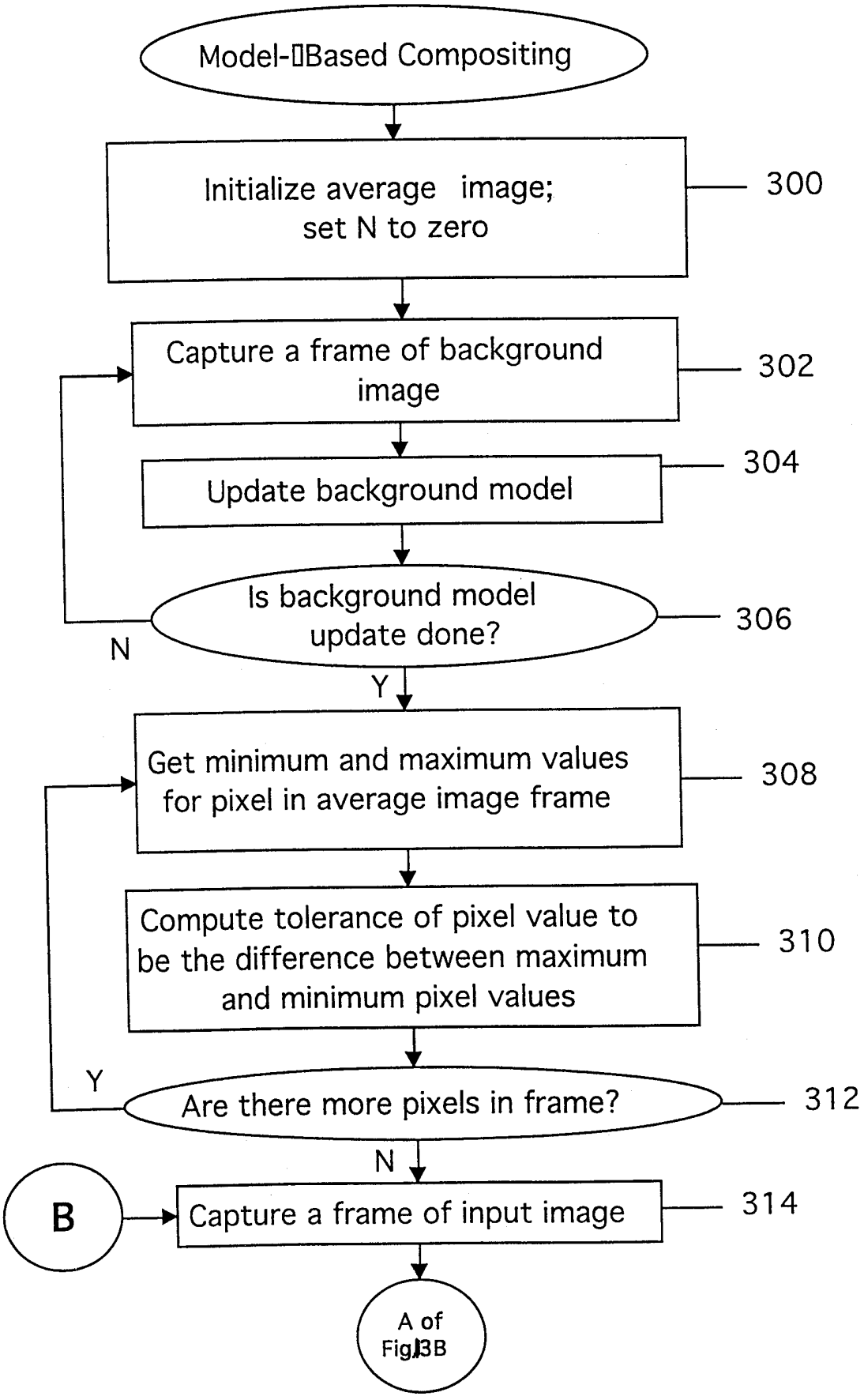

Model-Based Compositing
Initialize average image; set N to zero
300
Capture a frame of background image
302
Update background model
304
Is background model update done?
306
N
Y
Get minimum and maximum values for pixel in average image frame
308
Compute tolerance of pixel value to be the difference between maximum and minimum pixel values
310
Y
Are there more pixels in frame?
312
N
B
Capture a frame of input image
314
A of Fig. 3B

Fig. 3B

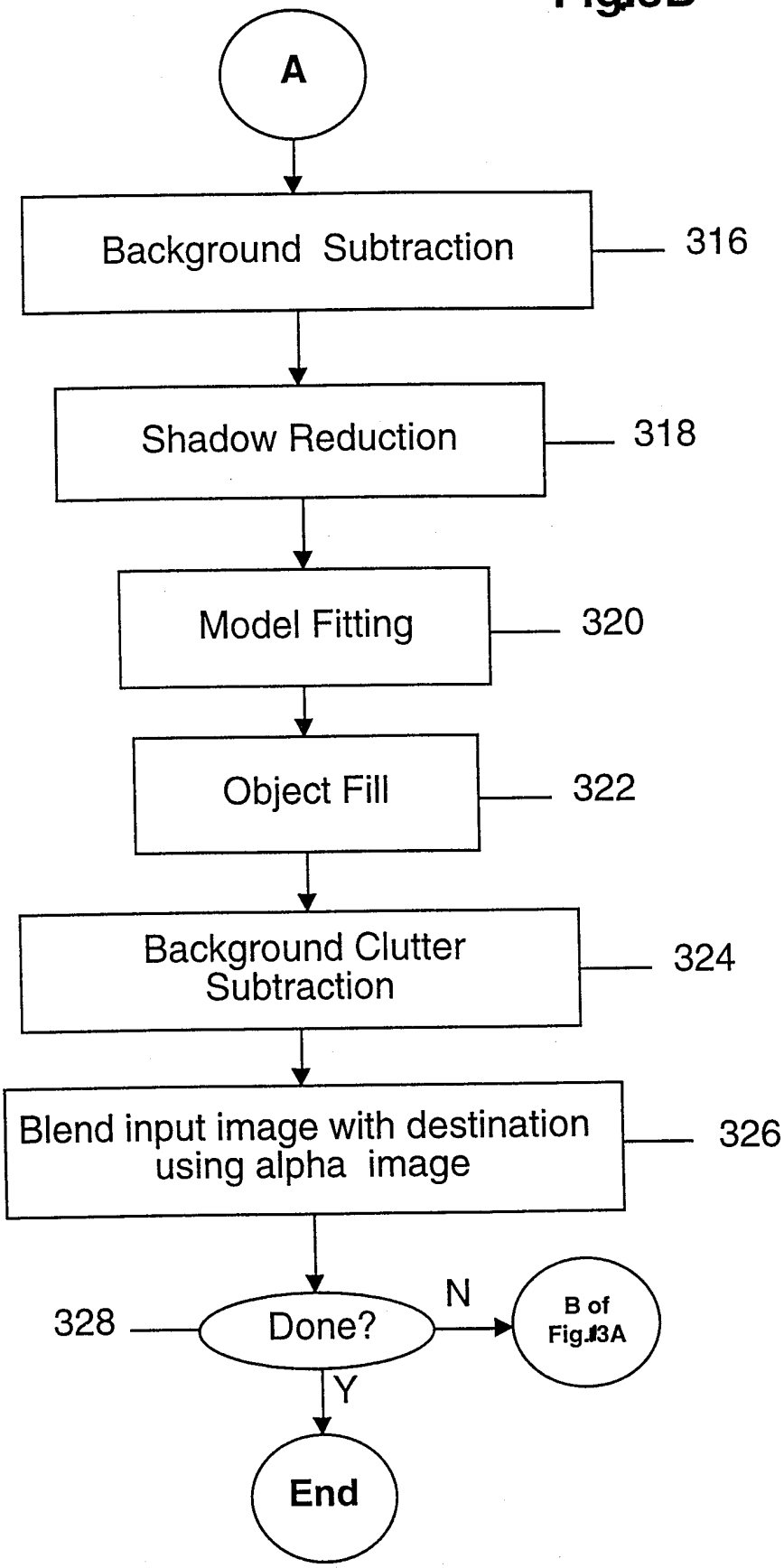

A
Background Subtraction
316
Shadow Reduction
318
Model Fitting
320
Object Fill
322
Background Clutter Subtraction
324
Blend input image with destination using alpha image
326
328
Done?
N
B of Fig. 3A
Y
End

Fig.4

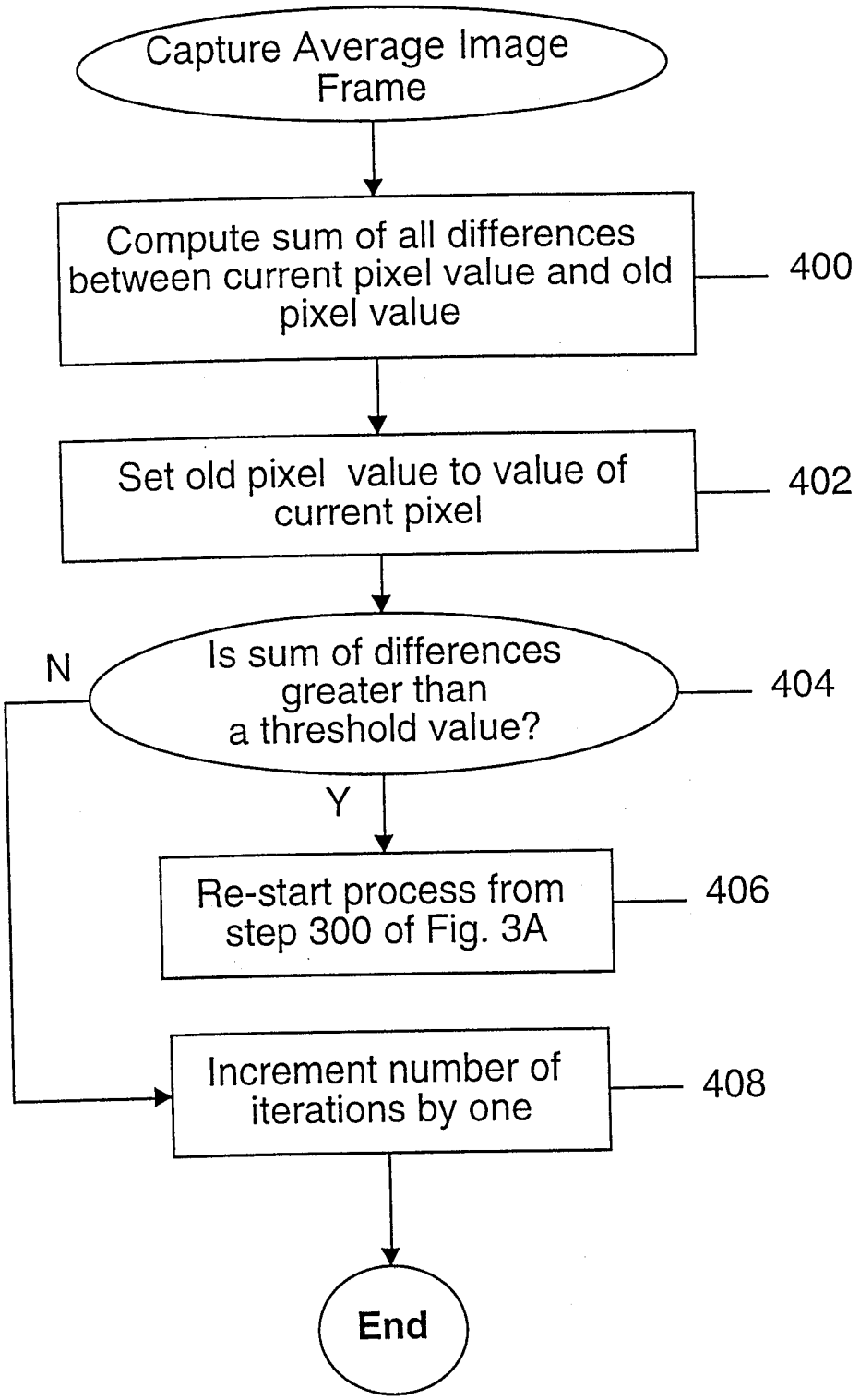

Capture Average Image Frame
Compute sum of all differences between current pixel value and old pixel value
400
Set old pixel value to value of current pixel
402
Is sum of differences greater than a threshold value?
404
N
Y
Re-start process from step 300 of Fig. 3A
406
Increment number of iterations by one
408
End

FIG. 15

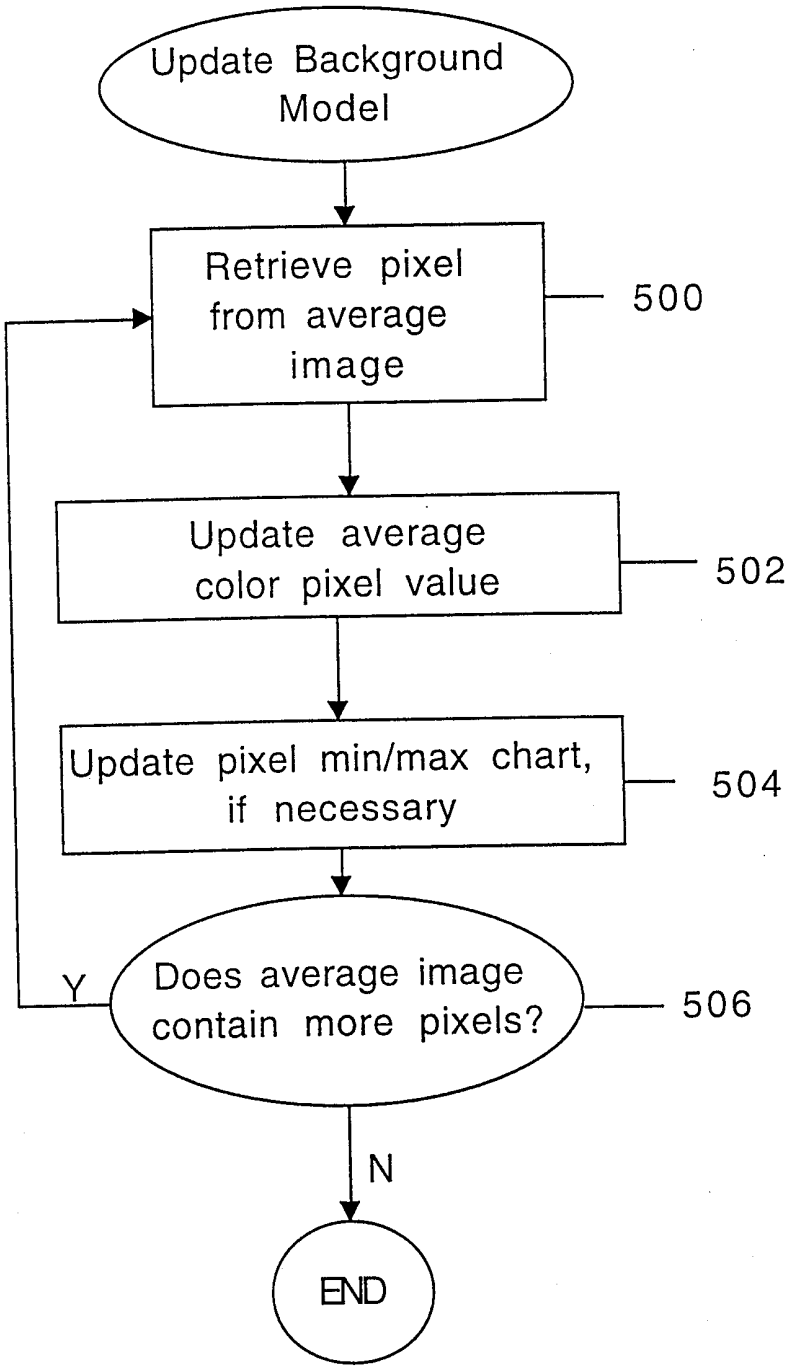
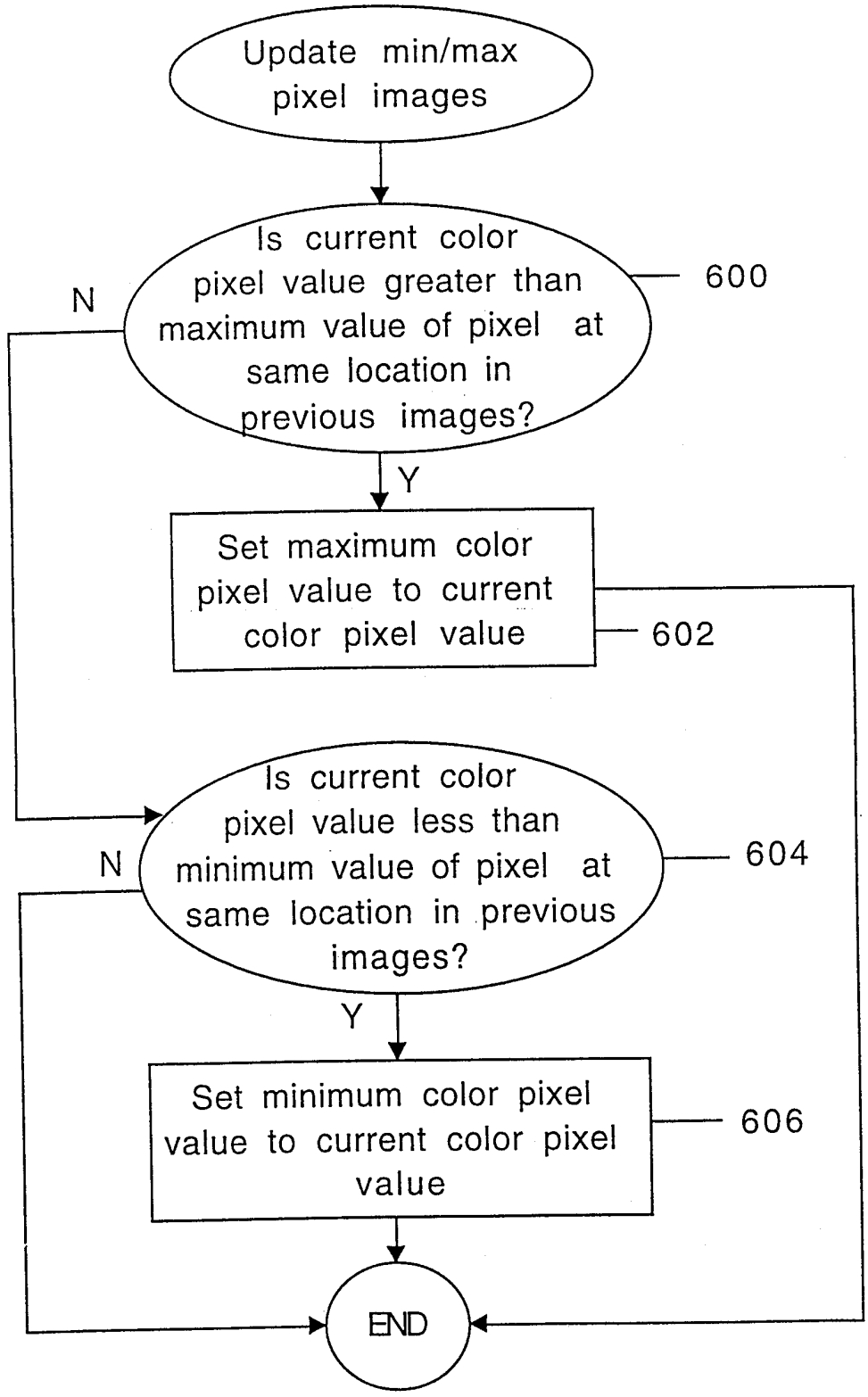

Update Background Model
Retrieve pixel from average image
500
Update average color pixel value
502
Update pixel min/max chart, if necessary
504
Does average image contain more pixels?
506
Y
N
END

FIG. 16


Update min/max pixel images
Is current color pixel value greater than maximum value of pixel at same location in previous images?
600
N
Y
Set maximum color pixel value to current color pixel value
602
Is current color pixel value less than minimum value of pixel at same location in previous images?
604
N
Y
Set minimum color pixel value to current color pixel value
606
END

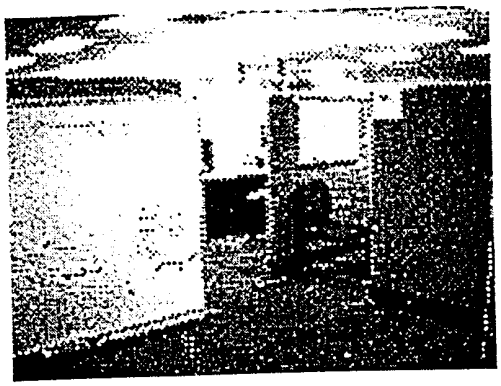


Figure 17A

Figure 17B

Fig18A

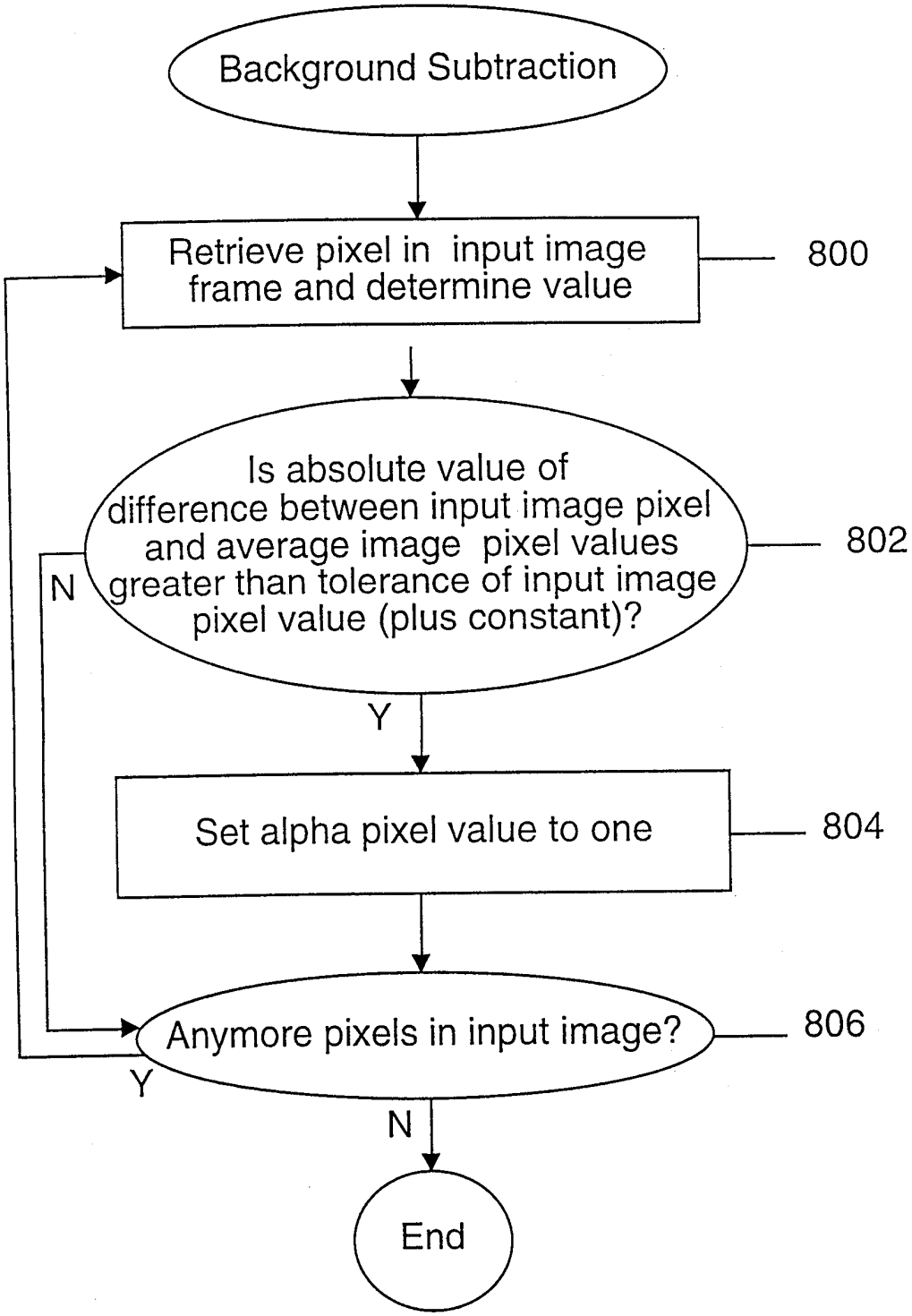

Background Subtraction
Retrieve pixel in input image frame and determine value
800
Is absolute value of difference between input image pixel and average image pixel values greater than tolerance of input image pixel value (plus constant)?
802
N
Y
Set alpha pixel value to one
804
Anymore pixels in input image?
806
Y
N
End

Figure 18B

Fig19A

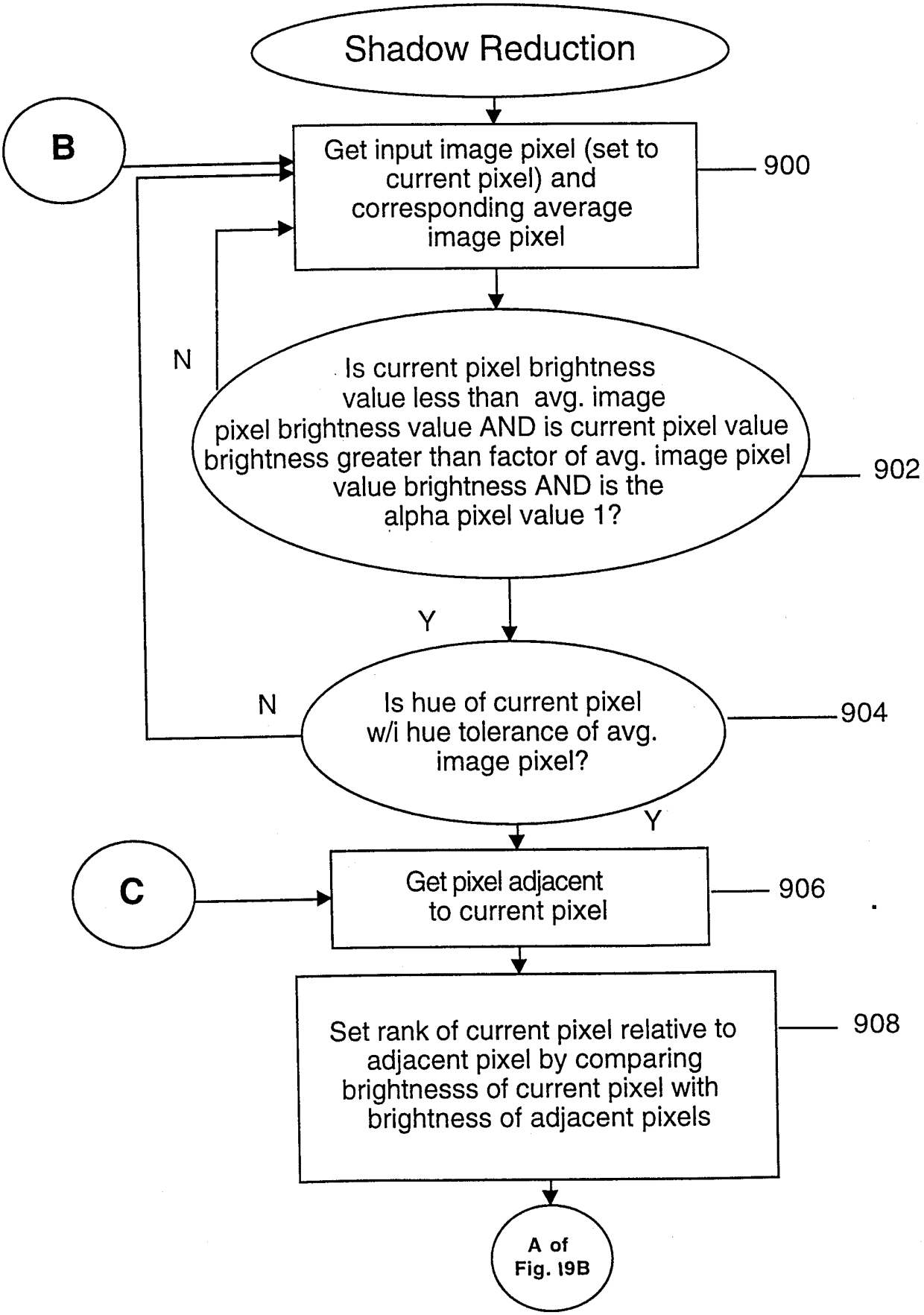

Shadow Reduction
B
Get input image pixel (set to current pixel) and corresponding average image pixel
900
N
Is current pixel brightness value less than avg. image pixel brightness value AND is current pixel value brightness greater than factor of avg. image pixel value brightness AND is the alpha pixel value 1?
902
Y
N
Is hue of current pixel w/i hue tolerance of avg. image pixel?
904
Y
C
Get pixel adjacent to current pixel
906
Set rank of current pixel relative to adjacent pixel by comparing brightnesss of current pixel with brightness of adjacent pixels
908
A of Fig. 19B

Fig.19B

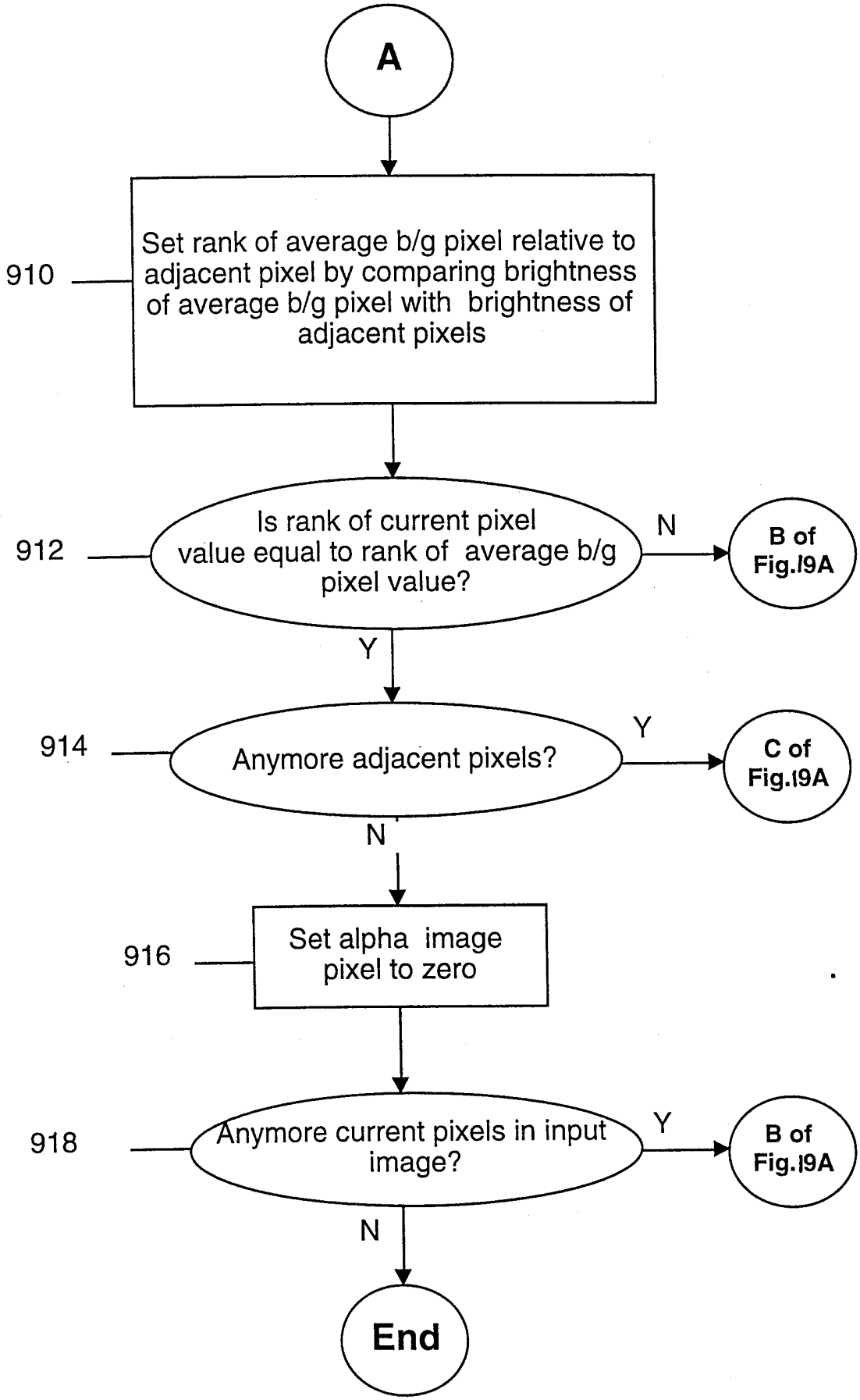

A
910
Set rank of average b/g pixel relative to adjacent pixel by comparing brightness of average b/g pixel with brightness of adjacent pixels
912
Is rank of current pixel value equal to rank of average b/g pixel value?
N
B of Fig.19A
Y
914
Anymore adjacent pixels?
Y
C of Fig.19A
N
916
Set alpha image pixel to zero
918
Anymore current pixels in input image?
Y
B of Fig.19A
N
End

Fig.10A

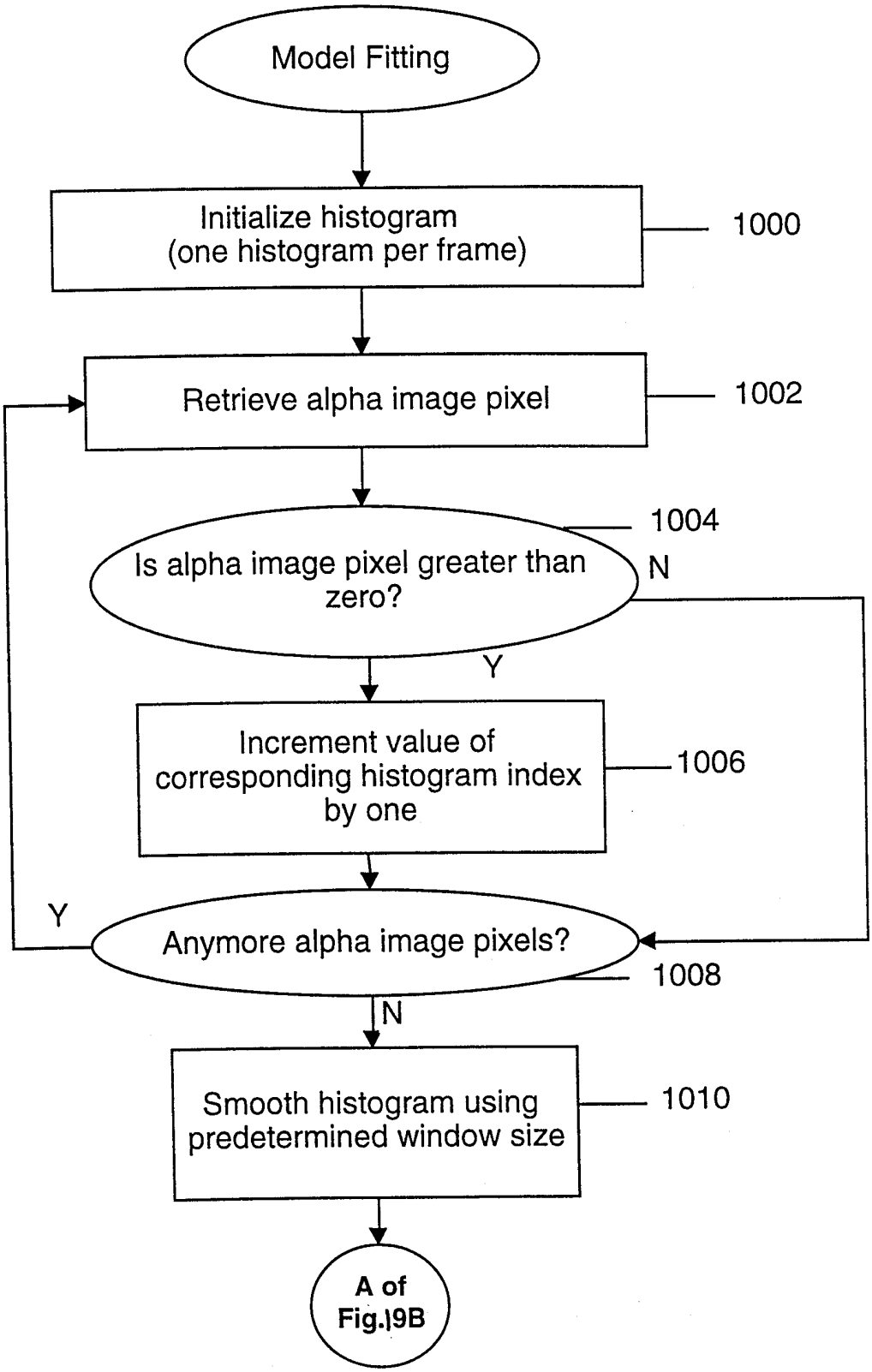

Model Fitting
Initialize histogram (one histogram per frame)
1000
Retrieve alpha image pixel
1002
1004
Is alpha image pixel greater than zero?
N
Y
Increment value of corresponding histogram index by one
1006
Y
Anymore alpha image pixels?
1008
N
Smooth histogram using predetermined window size
1010
A of Fig.19B

Fig.10B

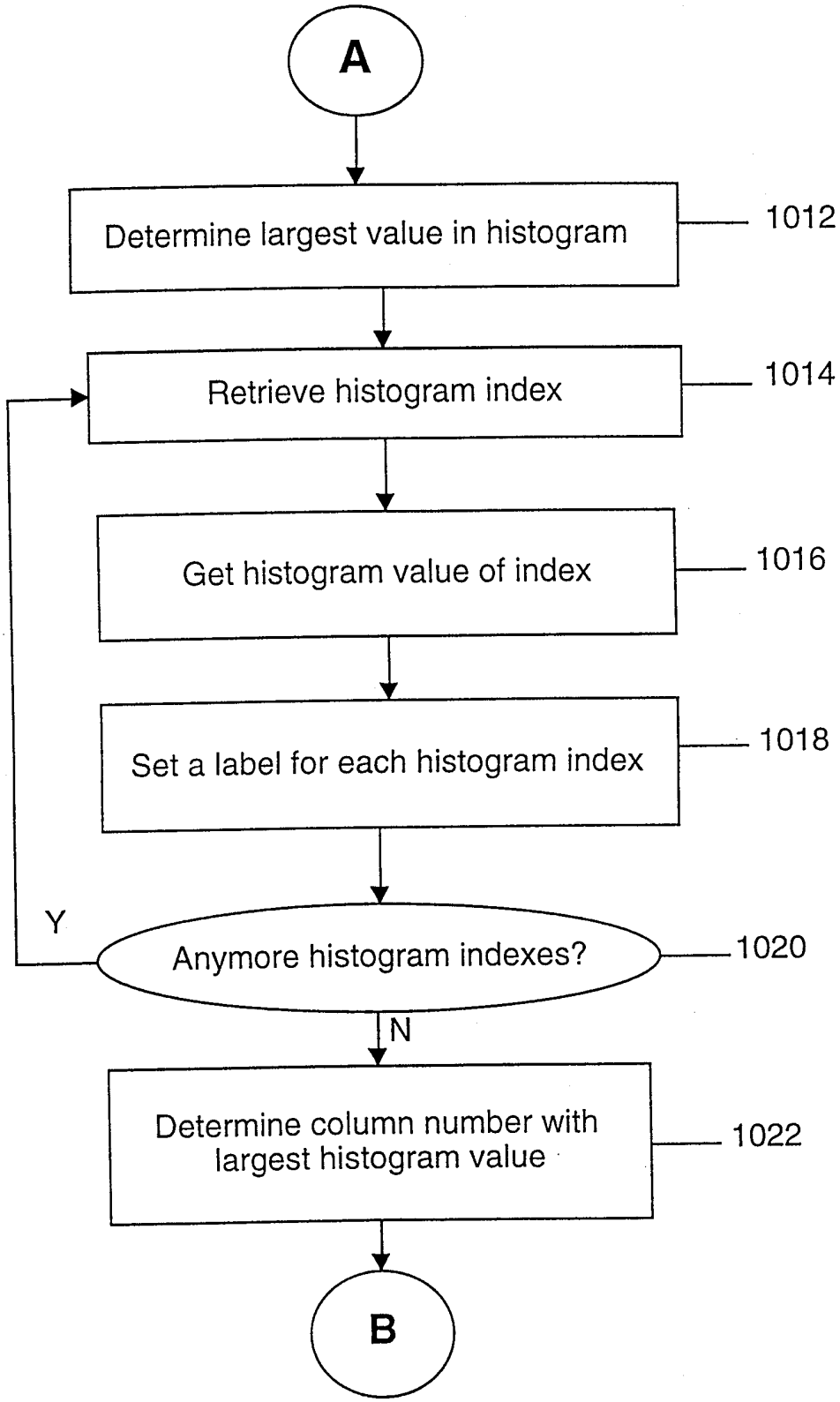

A
Determine largest value in histogram
1012
Retrieve histogram index
1014
Get histogram value of index
1016
Set a label for each histogram index
1018
Y
Anymore histogram indexes?
1020
N
Determine column number with largest histogram value
1022
B

Fig.10C

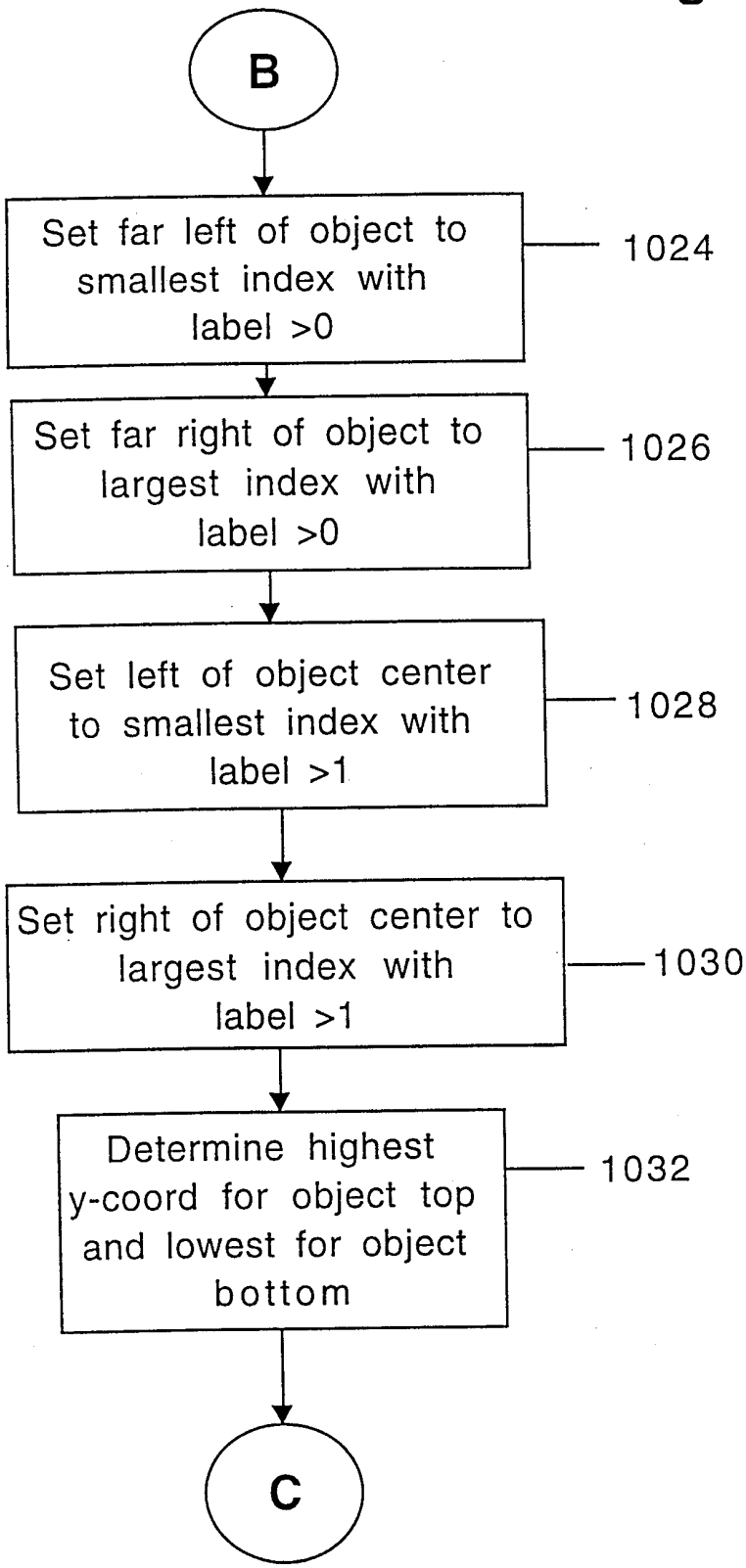

B
Set far left of object to smallest index with label >0
1024
Set far right of object to largest index with label >0
1026
Set left of object center to smallest index with label >1
1028
Set right of object center to largest index with label >1
1030
Determine highest y-coord for object top and lowest for object bottom
1032
C

Fig.10D

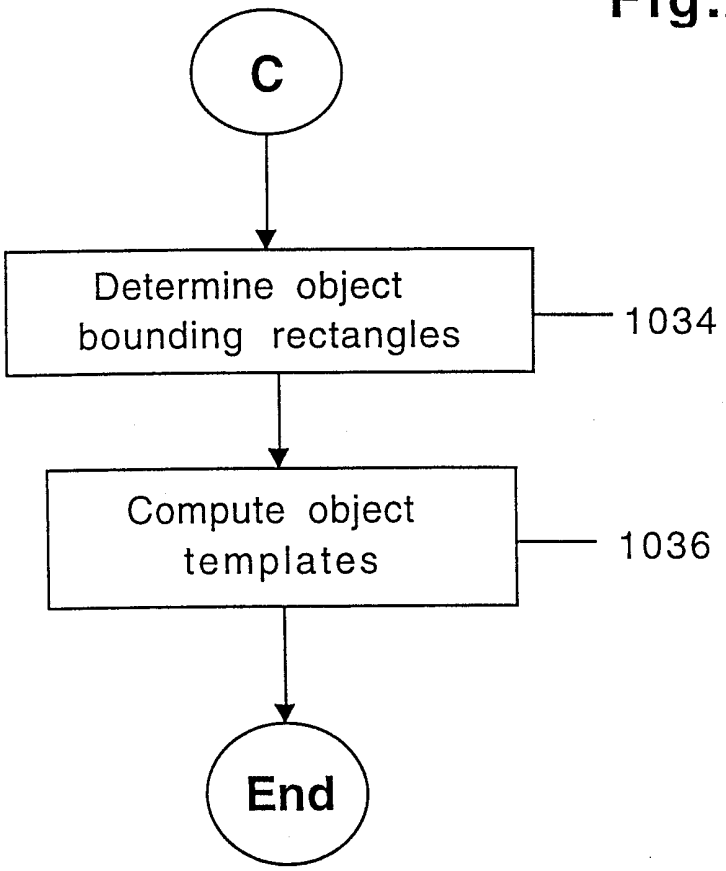

C
Determine object bounding rectangles
1034
Compute object templates
1036
End

Fig.11

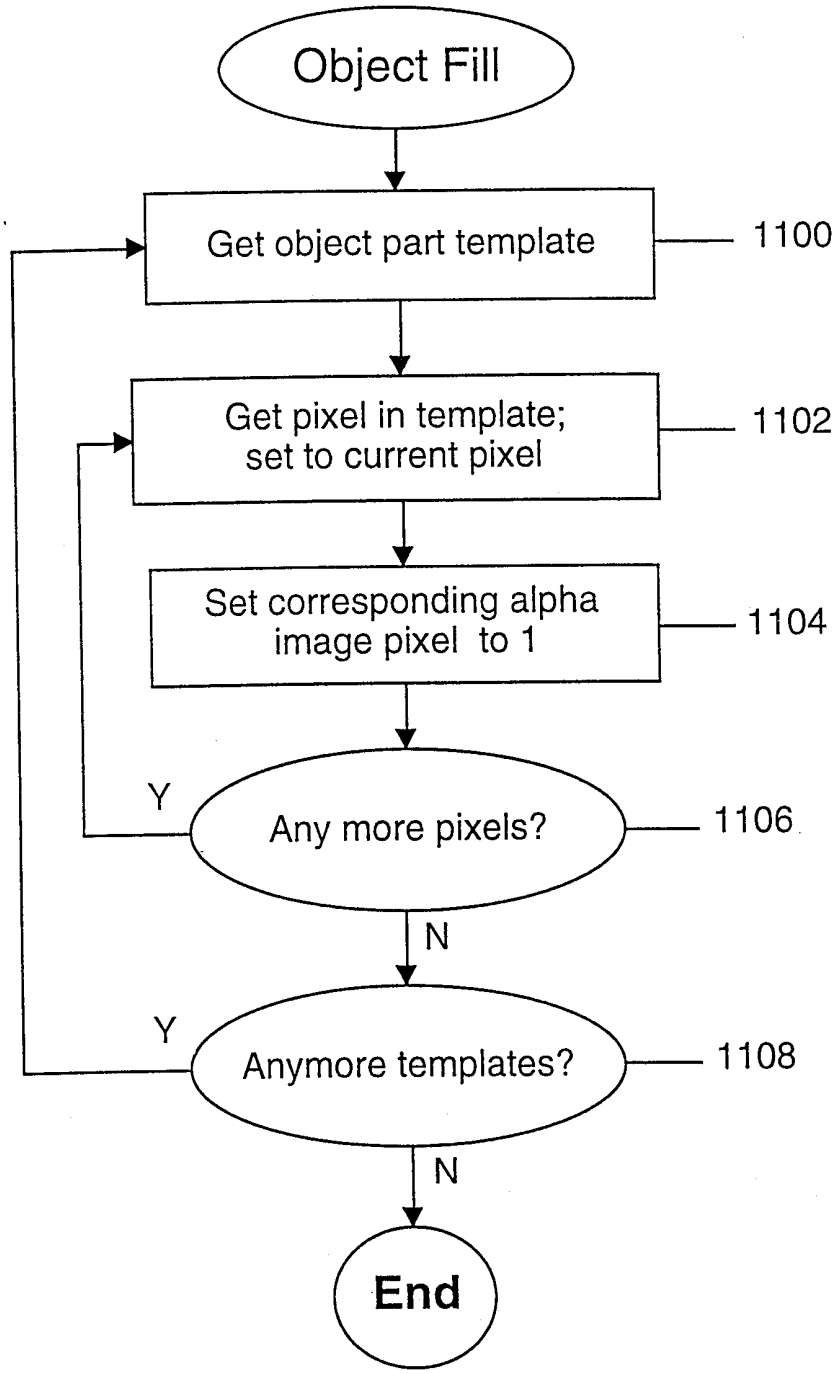

Object Fill
Get object part template
1100
Get pixel in template; set to current pixel
1102
Set corresponding alpha image pixel to 1
1104
Any more pixels?
1106
Y
N
Anymore templates?
1108
Y
N
End

Fig.12A

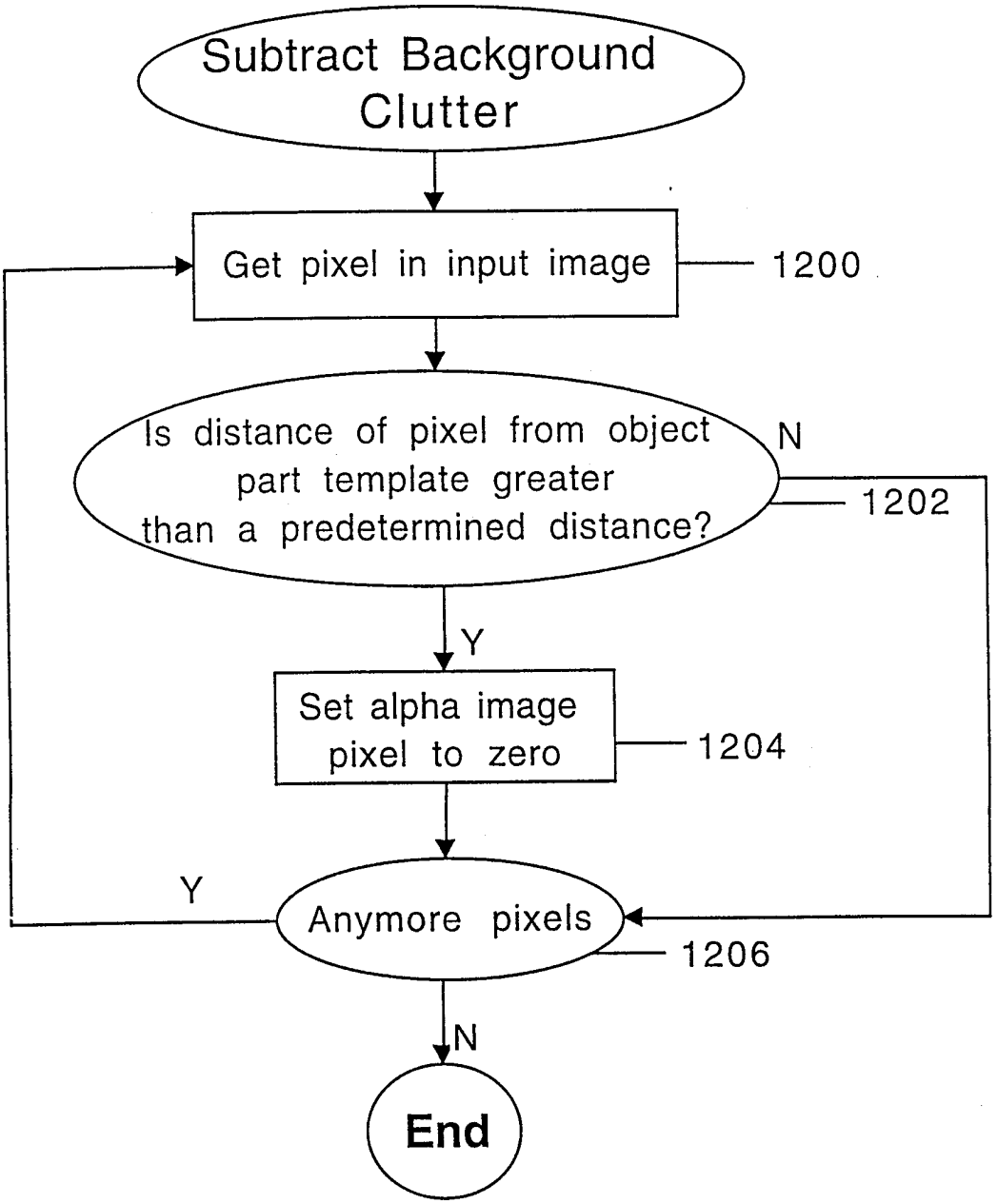
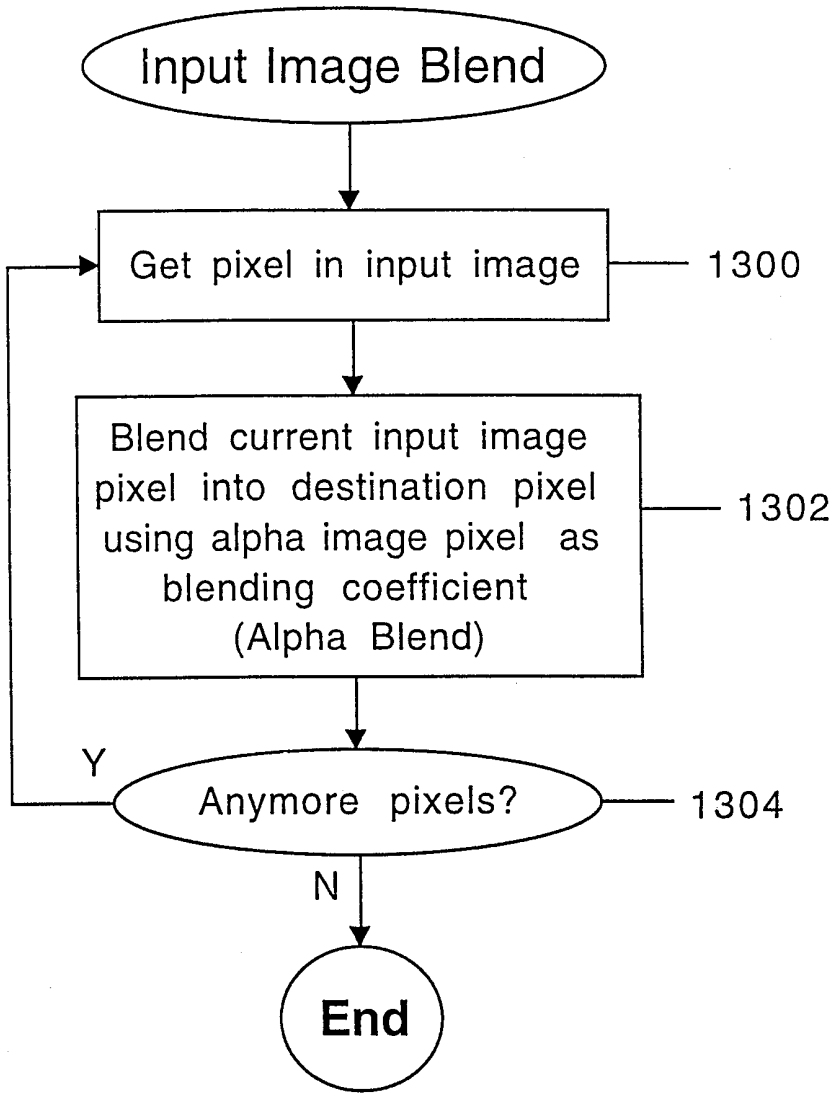

Subtract Background Clutter
Get pixel in input image
1200
Is distance of pixel from object part template greater than a predetermined distance?
N
1202
Y
Set alpha image pixel to zero
1204
Y
Anymore pixels
1206
N
End

Figure 12B

Fig.13


Input Image Blend
Get pixel in input image
1300
Blend current input image pixel into destination pixel using alpha image pixel as blending coefficient (Alpha Blend)
1302
Y
Anymore pixels?
1304
N
End

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/00407

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G09B 5/00
US CL :434/307 A

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)

U.S. : 434/307 A, 307 R

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 practicable, search terms used)

APS
search terms: karaoke, camera, composite

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,144,454 A (CURY) 01 SEPTEMBER 1992	NONE
A	US 5,151,793 A (ITO ET AL) 29 SEPTEMBER 1992	NONE
X	US 5,689,081 A (TSURUMI) 18 NOVEMBER 1997, SEE WHOLE DOCUMENT	15-22, 24-25
X	US 5,691,494 A (SAI ET AL) 25 NOVEMBER 1997, SEE WHOLE DOCUMENT	15-22, 24-25
X, E	US 5,725,383 A (FUNAHASHI ET AL) 10 MARCH 1998, SEE WHOLE DOCUMENT	15-22, 24-25
X, E	US 5,810,603 A (KATO ET AL) 22 SEPTEMBER 1998, SEE WHOLE DOCUMENT	15-22, 24-25
A, E	US 5,803,747 A (SONE ET AL) 08 SEPTEMBER 1998	NONE

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"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
"A"	document defining the general state of the art which is not considered to be of particular relevance	"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
"E"	earlier document published on or after the international filing date	"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
"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)	"&"	document member of the same patent family
"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		

Date of the actual completion of the international search
08 APRIL 1999

Date of mailing of the international search report
22 APR 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
John Rovnak
Telephone No. (703) 308-3087

Sheila Veney
Paralegal Specialist
Technology Center 3700

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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/00407

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A, E	US 5,827,990 A (FUJITA) 27 OCTOBER 1998	NONE