Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2380105 A1 2003/10/09

(21) 2 380 105

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(22) Date de dépôt/Filing Date: 2002/04/09

(41) Mise à la disp. pub./Open to Public Insp.: 2003/10/09

(51) Cl.Int.^{7}/Int.Cl.^{7} H04N 13/00

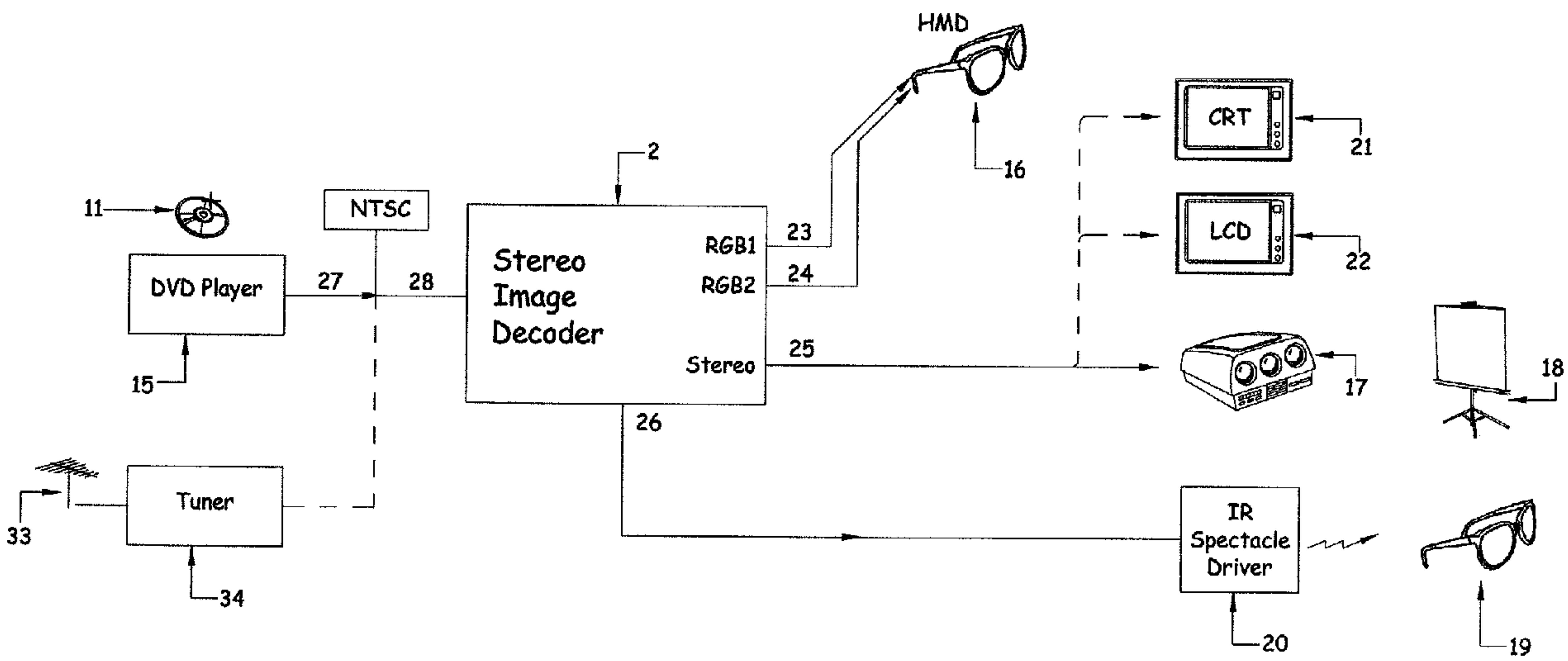
(71) Demandeurs/Applicants:
ROUTHIER, NICHOLAS, CA;
BELZILE, JEAN, CA;
...

(72) Inventeurs/Inventors:
ROUTHIER, NICHOLAS, CA;
BELZILE, JEAN, CA;
THIBEAULT, CLAUDE, CA;
MALOUIN, DANIEL, CA;
CARPENTIER, PIERRE-PAUL, CA;
DALLAIRE, MARTIN, CA

(74) Agent: ROBIC

(54) Titre : PROCESSUS ET SYSTEME D'ENREGISTREMENT ET DE LECTURE DE SEQUENCES VIDEO STEREOSCOPIQUES

(54) Title: PROCESS AND SYSTEM FOR ENCODING AND PLAYBACK OF STEREOSCOPIC VIDEO SEQUENCES



(21) 2 380 105

(13) A1

(71) **Demandeurs(suite)/Applicants(continued)**: THIBEAULT, CLAUDE, CA; MALOUIN, DANIEL, CA; CARPENTIER, PIERRE-PAUL, CA; DALLAIRE, MARTIN, CA

PROCESS AND SYSTEM FOR ENCODING AND PLAYBACK OF STEREOSCOPIC VIDEO SEQUENCES

BACKGROUND OF THE INVENTION

1. Field of the invention:

The present invention relates generally to a process and system for encoding and decoding a dual program image sequence, and, more particularly, to a process and system for compressing two image sequence signals on a single video signal and decoding said single video signal to reproduce two image sequence programs or a three-dimensional stereoscopic program in multiple viewing formats. Although the invention will be described hereinafter by reference to processing of three-dimensional stereoscopic programs such as movies, it should be deemed to be within the scope of the present invention to apply to the processing of any pair of video sequences, regardless of any differences in the respective video content of each sequence.

2. Brief description of the prior art:

Since the invention of the stereoscope in 1947, several systems have been developed to enable a viewer to view three-dimensional (3D) programs through the reproduction of a first image sequence intended for viewing by the viewer's left eye and a second sequence of images of the same scene and at the same time but with a parallax with respect to the first image sequence, intended to be viewed exclusively by the viewer's right eye, thereby replicating the principles of natural three-dimensional vision. Since the 1950's, many films have been made using dual camera head systems to pick up stereo pairs of images in time-synchronism and with a parallax to enable a viewer at reproduction to perceive the effect of depth, so to provide a more complete and exciting viewing experience.

At present, home theatre systems are rapidly penetrating the household market and very sophisticated and high quality systems are gaining in popularity, responding to a need for high quality cinematographic experience at home. Nevertheless, existing stereoscopic reproduction systems are still far from fulfilling the expectations of viewers and are still not integrated into even the most advanced home theatre systems available. The reason mostly lies on the relatively pour image quality (fade colours and/or stair-stepping diagonals) and the fatigue and discomfort caused by the usual flicking and lack of spatial realism. Indeed, since two different programs are being presented with equipment intended for single video program presentation, such as a television set, sharing of the technical resources between two video signals leads to loss of image spatial resolution and flicking due to the reduction by half of the frame presentation rate for each eye and contrast between image fields and a black background.

A typical existing stereoscopic reproduction technology consists in encoding the first image sequence information in the even line field of an interlaced video signal and the information of the second image sequence in the odd line field of the signal. At playback, shutter spectacles are used to block one of the viewer's eyes during presentation of the even lines and the other eye during presentation of the odd lines. As normal images comprising even and odd lines are typically presented in two successive scan periods of 1/60s, each eye sees the stereoscopic program as a sequence of 1/60s images followed by 1/60s blackout periods, to enable each eye to view 30 frames per second (fps). Moreover, each reproduced image is constituted by alternating image lines and black lines. Obviously, the stereoscopic images so reproduced loose half of their topological information and the 50% duty cycles (both in space and in time) induce loss of brightness and flicking, as confirmed by experience.

A solution to such limitations, shortcomings and drawbacks, is to present complete stereoscopic images at a rate of at least 60 fps (30 full frames per second per eye) which would normally require at least twice the signal bandwidth required by a non-stereo (planar) program. Elimination of flicking in a room presenting relatively high contrast between the displayed pictures and ambient lighting, further requires a vertical scan (and shutter spectacle) frequency of up to 120 Hz, to enable presentation of up to 60 full definition images per second to each eye. While such a frequency is not currently available, flickerless presentation of stereoscopic program can be set up by using two digital progressive scan video projectors of current manufacture, receiving respectively a first and a second image sequence of the stereoscopic program at a continuous rate of 30 fps each. The output of each projector is optically filtered to produce a vertically and an horizontally polarized output projecting images in register and in perfect time synchronism on a special silver coated screen. Eyewear comprising differently polarized glasses can be worn by a viewer to reveal the tri-dimensional effects. Such a solution is obviously very expensive and does not meet market expectations for a home theatre system.

However, very fast and relatively affordable projectors using the DLP (Digital Light Processing) technology are now available that could provide a presentation rate up to 120 fps, so that a single projector could alternatively present images of stereo pair sequences at a sufficiently high rate to substantially eliminate flicking even in a high contrast environment. Also, high-end CRT projectors and computer monitors could provide such a compatible definition and refresh rate.

Nevertheless, a major limitation of such systems remains that most current standards for storage and broadcast (transport) of video program information are limiting the flow of full frame images to 30 fps, which is approximately half of the capacity required to store and present a high quality stereoscopic program originally comprised of two 24 (American Motion Picture), 25 (European Video) or 30 fps programs. Furthermore, since motion picture movies are always captured and recorded at a rate of 24 frames per second, the dual problem of compressing two 24 fps programs into a single 30 fps signal and thereafter expanding such a signal to present the two programs at a rate of 30 to 60 fps each must be addressed. Therefore, the future of the 3D home theatre lies on the capacity to encode and decode a stereoscopic video signal to comply with standard recorders, players and broadcast equipment of present manufacture treating a 30 fps signal compressed and decompressed using a protocol such as MPEG-1 or MPEG-2 (Moving Picture Image Coding Group) compression/decompression protocol of the MAIN profile (vs MVP), so that negligible loss of information or distortion is induced throughout the process.

A few technologies of the prior art have been teaching solutions to overcome one or more of the above mentioned shortcomings and limitations. Firstly, the 3:2 pull-down compression method can be used to create a 30 fps stereoscopic interlaced signal from a 24 fps interlaced picture sequence. With this method the original image sequence is timely expanded by creating and inserting one new picture after every four pictures of the original sequence. The new picture comprises the even lines of the preceding picture in one field and the odd lines of the next picture in its second field. Obviously, each picture of the original program may be comprised of a first field comprising a portion of a left view image and a second field comprising a portion of a right view image of a stereoscopic program. A 30 fps stereoscopic program can thereby be obtained from a 24 fps left eye sequence and a 24 fps right eye sequence. With such a technique however, the resulting 30 fps program presents anachronism and topological distortion due to the combination in certain pictures of lines belonging to images captured at different times. This yields a poor result, lacking in realism and causing eye fatigue and discomfort to the viewer. When used to present a stereoscopic program, this technique further suffers from the same limitations and drawbacks as discussed hereinabove about the interlaced signal compression technique.

Furthermore, many stereoscopic display devices have been developed using different input signals incompatible with one another and requiring different transport (storage or distribution) formats (column interleaved, row interleaved, simultaneous dual presentation, page flipping, anaglyphic, etc.). A solution to bring a stereoscopic video program to different systems at the same time while allowing for 2D viewing would be to simultaneously broadcast or store on several physical media in all the existing formats. Obviously, that would neither be practical nor economical. Therefore, the future of stereoscopic video at home requires a stereoscopic video signal and a video processing apparatus that have the ability to generate multiple/universal stereoscopic output formats compatible with current and future stereoscopic display devices while allowing for normal 2D viewing.

Many patents are also teaching compression techniques to reduce two 30 fps signals to be carried through a single channel with a 30 fps capacity, some of them being designed to be transparent for the MPEG compression/decompression process. However, these techniques do not feature temporal interpolation as needed to create the missing frames to convert for instance a 24 fps sequence to 30 fps, or to convert a 30 fps sequence to a 48, 60, 72, 96 or 120 fps sequence, while preserving image quality and comfortable viewing experience. Furthermore, they do not have the ability to generate multiple stereoscopic output formats from the same video signal and video processing apparatus.

For instance, US patent No 5,626,582 granted to Muramoto et al on May 6, 1997, teaches a time-base compression method in which two 30 fps video signals are digitized and stored in DRAM memory at a given clock frequency. Subsequently, the memory is read at twice that write frequency so that two samples of an original period of 1/30 can be concatenated in a 1/30 interval. However, depending on the selected sampling frequency, the final signal will either lack definition because the information of two adjacent pixels will be averaged in a single digital data (low sampling frequency and normal playback frequency), or exceed the capacity of a data storage medium such as a DVD or a broadcast channel. This invention also lacks the ability to generate multiple output formats from a given source format, and requires two parallel circuits for reconstruction of the original sequences.

Further, in International application No WO 97/43863, by Briede, laid open on November 20 1997, images from a first and a second sequence of images are decimated by removing in half of the picture elements and then concatenated into side by side fields of a combined stereo image sequence to be transmitted through a channel. At the receiving end, the juxtaposed fields are demultiplexed from the stereo image sequence and are sent to two parallel expanding circuits that simultaneously recreate the missing picture elements of their respective stereoscopic video sequence (right and left). The thereby reconstructed original first and second images sequences are then outputted to two displays for reproduction.

While that technology provides an interesting method for spatially compressing/decompressing full frames, for storage or distribution using a limited capacity channel (transport medium), it does not address the problem of converting a two 24 or 25 fps image sequences into a 30 fps stereo sequence or boosting the playback rate to prevent flicking. Furthermore, it does not allow playback in other stereoscopic formats, including the page flipping mode using a single display monitor or projector through time sequencing of the rebuilt first and second image sequences. Also, as for the previous example, two parallel circuits are again required to carry out the reconstruction process on both image sequences.

Although the above examples show that different methods and systems are known for the encoding of two video signals or images sequences into a single signal and for decoding such a composite signal to substantially retrieve the original signals or sequences, these methods and systems of the prior art are nevertheless lacking important features to provide a functional system which enables high fidelity recording, broadcast and playback of two 24 fps motion picture movies as well as 30 fps stereoscopic video programs, using a single channel and conventional recording, playback and display equipment of present manufacture, as required for instance to meet the expectations of the home theatre market for 3D movies reproduction.

There is thus a need for a novel stereoscopic program encoding and playback method and system which can be readily used with existing home theatre equipment to provide a high quality stereoscopic reproduction, still at an affordable cost, while enabling playback of a specific stereoscopic video transport signal in a plurality of output formats.

SUMMARY OF THE INVENTION

More specifically, in accordance with the invention as broadly claimed, there is provided a stereoscopic program encoding and playback process and a system implementing the process, comprising the steps of:

1- Topologically decimating first and second moving image sequences having a first display rate, into reduced mosaics to form adjacent fields of a third image sequence having a second display rate which is less than twice said first display rate;
2- Encoding and transporting said third image sequence through a data transport medium;
3- Retrieving said third image sequence from said data transport medium, and,

4- Topologically interpolating reduced image mosaics from fields of said third image sequence to substantially rebuild images of said first and second image sequences.

The process preferably comprises an additional step 5 wherein said first and second rebuilt image sequences are time sequenced to form an output image sequence comprised of successively alternating images from each of said first and second rebuilt image sequences (page flipping mode). According to alternate real time reading modes of the first and second rebuilt image sequences, output sequence format can be arranged to comply with a line interleaved, a column interleaved, an anaglyphic or a two-dimensional presentation mode.

There is further disclosed another embodiment of the process according to the present invention wherein step 4 further comprises:

Creating pairs of new images by temporal interpolation of successive images of the rebuilt first and second image sequences, and inserting one image from each of said temporally interpolated images pairs into each of said first and second rebuilt image sequences to thereby increase their display rate.

There is further disclosed another embodiment of the process according to the present invention wherein step 4 further comprises:

repeating image pairs into said first and second rebuilt image sequences to thereby increase their display rate.

There is further disclosed another embodiment of process according to the present invention wherein, first and second moving image sequences have their display rate increased by inserting pairs of new temporally interpolated (created) or repeated (read twice) images before carrying out step 1 of the above process.

According to another embodiment of the present invention, said output image sequence has a normal display rate of R images per second and is comprised of 12 images per second from the first and second image sequences, 36 images per second spatially interpolated from reduced mosaics of the first and second image sequences, and R - 48 inserted images. Inserted images can be repeated images or temporally interpolated images. Should R be equal to 96 images per second, said output image sequence comprises 12 images per second from the first and second image sequences, 36 images per second spatially interpolated from reduced mosaics of the first and second image sequences, each image being repeated twice.

According to a further embodiment of the present invention, said output image sequence has a normal display rate of R images per second and is comprised of 60 images per second spatially interpolated from reduced mosaics of the first and second image sequences, and R - 60 inserted images. Inserted images can be repeated images or temporally interpolated images. Should R be equal to 120 images per second, said output image sequence comprises 60 spatially interpolated images repeated twice.

There is further disclosed a system according to the present invention, comprising a digital video disk player, a decoder and a video display device, wherein said decoder:

inputs a signal from the video disk player representative of a sequence of images having a left field and a right field respectively representing reduced image mosaics,

topologically interpolates reduced image mosaics from fields of said image sequence to form a first and second sequences of rebuilt images,

creates new images by temporal interpolation of successive images of the rebuilt first and second image sequences,

time sequences rebuilt images and created images to form an output image sequence comprised of successively alternating images from

each of said first and second rebuilt image sequences, and,
outputs signals representative of said output image sequence to said video display device.

All embodiments of the system can be further provided with synchronized sound capability, characterized in that at playback, the recorded sound track(s) is(are) time delayed by the decoder so to be outputted by a sound reproduction system in phase with the visual program.

As it will become more apparent from reading of the following detailed description, the present invention overcomes the limitations and drawbacks of the above mentioned solutions of the prior art, and amongst other advantageous features the following can be enlighten:

- The present invention provides an encoding method and a system implementing the method which enable the compression of two 24 or 30 fps image sequences in a format suitable for storage into a conventional digital video disk (DVD) or broadcasting using conventional equipment, and without substantially perceivable loss of spatial and temporal information.

- The present invention provides a playback method and system which enable high visual quality reproduction of stereoscopic programs by reconstruction of the original sequences and rate augmentation for presentation at rates from 24 to 60 full resolution images per second per eye, in progressive or interlaced mode.

- The present invention provides an encoding and playback method and system which features full compatibility with commercially available data storage medium playback equipment and display equipment, and more specifically with MPEG main view profile compression/decompression protocols and commercial circuits.

- The present invention provides a stereoscopic program playback method and system which provide universal output signals which can be directly used or converted to enable reproduction with any existing technology such as head mounted displays (HMD), LCD, DLP and CRT rear and front projection TV'S, direct view TV'S and computer monitors operating under any universal standard (NTSC, PAL, SDTV, HDTV, etc.), with shutter spectacles, polarized eyewear, or anaglyphic glasses.

- The present invention provides an encoding and playback method and system which provide elimination or substantial reduction of flicking and fatigue usually encountered in viewing stereoscopic movies with the prior art methods and apparatus.

- The present invention further provides a method and system which enable encoding and decoding of two independent image sequences potentially representing different and unrelated scenes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1a is a schematic representation of a system according to the present invention for the compression encoding of two planar image sequences into a stereoscopic image sequence to be recorded onto a data storage medium or broadcast on a single channel.

Figure 1b is a schematic representation of a system according to the present invention, for expansion decoding and playback of a stereoscopic image sequence previously encoded with a system such as represented in Figure 1a.

Figure 2a is a schematic representation of a portion of a digitized image 60 topologically separated into two complementary mosaics of picture elements, forming fields A and B of a merged image 60.

Figure 2b is a schematic representation of a portion of two digitized images 50 and 50', topologically decimated into reduced mosaics respectively forming field A and field B of a merged image 60.

Figure 2c is a schematic representation of a process for the spatial interpolation of a pair of decimated images comprised in a merged image 60, to rebuild two full-definition images 72 and 72'.

Figure 2d is a schematic representation of a time-interpolation process for the creation of a new image 52 from two images 50.4 and 50.5 with a time delay.

Figure 3a is a schematic representation of a first embodiment of a compression process according to the present invention, for compression encoding two planar image sequences into a stereoscopic combined image sequence.

Figure 3b is a schematic representation of a first embodiment of a decompression process according to the present invention, for reconstruction and temporal expansion of a stereoscopic image sequence previously encoded according to a process such as represented in Figure 3a.

Figure 4a is a schematic representation of a second embodiment of a compression process according to the present invention, for compression encoding two planar image sequences into a stereoscopic image sequence.

Figure 4b is a schematic representation of a second embodiment of a compression process according to the present invention, for reconstruction and temporal expansion of a stereoscopic image sequence previously encoded according to a process such as represented in Figure 4a.

Similar reference numerals refer to similar parts throughout the various Figures.

DETAILED DESCRIPTION OF THE DRAWINGS

Preferred embodiments of the method and associated systems for encoding and playback of stereoscopic video sequences according to the present invention will now be described in detail referring to the appended drawings.

Referring to Figure 1, there is illustrated a typical system set-up according to the present invention, for the compression encoding of two planar image sequences into a stereoscopic image sequence. A first and a second source of image sequences represented by cameras **3** and **6** are stored into common or respective digital data storage media **4** and **7**. Alternatively, image sequences may be provided from digitised movie films or any other source of digital picture files stored in a digital data storage medium or inputted in real time as a digital video signal suitable for reading by a microprocessor based system. Cameras **3** and **6** are shown in a position wherein their respective captured image sequences represent different views with a parallax of a scene **100**, simulating the perception of a left eye and a right eye of a viewer, according to the concept of stereoscopy. Therefore, appropriate reproduction of said first and second captured image sequences would enable a viewer to perceive a three-dimensional view of scene **100**.

Stored digital image sequences, typically available in a 24 fps digital Y U V format such as Betacam 4:2:2 (motion pictures), are then converted to an RGB format by processors such as **5** and **8** and fed to inputs **29** and **30** of moving image mixer unit **1**, representing the main element of the encoding system. It should be noted however that he two image sequences can alternatively be converted on a time-sharing basis by a common processor, to deserve a need for cost reduction. Mixer **1** compresses the two planar RGB input signals into a 30 fps stereo RGB signal delivered at output **31** and then converted by processor **9** into a betacam 4:2:2 format at output **32** and in turn compressed into a standard MPEG2 bit stream format by a typical circuit **10**. The resulting MPEG2 coded stereoscopic program can than be recorded on a conventional medium such as a Digital Video Disk (DVD) **11** or broadcasted on a single standard channel through, for example, transmitter **13** and antenna **14**. Alternative program transport media could be for instance a cable distribution network or internet.

Turning now to Figure 1b, there is illustrated a typical system according to the present invention for the decoding and playback of the stereoscopic program recorded or broadcasted using the system of Figure 1. The stereo DVD **11** (3DVD) comprising the compressed information from said first and second images sequences, is played by a conventional player **15** of current manufacture, delivering a NTSC serial analog signal to the input **28** of the stereo image decoder **2**, the main element of the decode/playback system. Alternatively, any ATSC DTV signal in its analogue or digital format can be accepted.

Decoder **2** produces a synchronized pair of RGB signals at outputs **23** and **24**, representative of said first an second image sequences, to drive a dual input stereoscopic progressive display device such as a head mounted display (HMD) **16**. Further, decoder **2** produces a time-sequenced stereo RGB signal at output **25**, to supply a single input progressive display device such as projector **17**, a LCD display **22**, a CRT monitor or a SDTV or HDTV **21**, whereby images from said first and second image sequences are presented in an alternating page flipping mode. Alternatively, the stereo RGB signal from output **25** may be converted into an interlaced NTSC signal to be reproduced by an analog CRT television set or in other stereoscopic formats (ex: column interleaved for autostereoscopic lenticular displays). Also, decoder 2 may be so internally configured to output the stereo RGB signal at one of RGB outputs **23** or **24**, thus eliminating output **25**.

Decoder **2** further produces a sync-timing signal at output **26** to drive an infrared shutter spectacle driver **20**, driving spectacles **19**. Shutter spectacles **19** can be worn by a viewer to view a three-dimensional program projected for instance on screen **18** by projector **17** fed by stereo output **25**, by enabling the viewer to alternately see an image from the first image sequence with one eye and an image from the second image sequence with his second eye.

As stated in the foregoing description, the two original image sequences contain too much information to enable direct storage onto a conventional DVD or broadcast through a conventional channel using the MPEG2 or equivalent multiplexing protocol handling information at a rate of 30 fps. Therefore mixer **1** carries out a decimation process to reduce each picture's information by half. The spatial decimation carried out by mixer **1** will now be described by reference to Figures 2a and 2b.

Figure 2a illustrates a portion of an image **50** as defined by a RGB video signal processed by mixer **1** and decoder **2**. In the RGB format, each pixel is defined by a vector of 3 digital numbers respectively indicative of the red, green and blue intensity. In the system of the present invention, each image is defined as being a superposition of two complementary mosaics: a type A mosaic and a type B mosaic. By definition, mosaic A comprises the series of pixels (indicated by a solid dot) starting with the first pixel of the image (pixels one of line one), followed by the third one and so forth throughout the image, from the left to the right of each row and from the upper line to the last one. The remaining pixels (indicated by the empty dots) are defined as mosaic B. In the example of Figure 2a, the two complementary mosaics of image 50 are shown as being respectively stored in field A and field B of a common merged image 60, which rather provides a separation than a decimation process per se.

As better illustrated in Figure 2b, basically, images are spatially compressed by 50% by keeping only mosaic A of images of the first sequence (ex left eye sequence) such as **50**, and field B of the images of the second sequence (ex. right eye sequence) such as **50'**. Keeping mosaics of different types for each sequence promotes higher fidelity at playback when first and second sequences represent different views of a same scene. Alternatively, spatial compression could be carried out by saving mosaic A for even numbered images and mosaic B for odd numbered images, for both input sequences, so that two successive images of the same eye would be rebuilt from mosaics of different types and potentially stored in the same compressed frame.

The above operation is accomplished by inputting the data of one pixel at a time in a three-pixel input buffer **55** as shown in Figure 2b. Pixel information is then transferred into the appropriate memory location of one or more frame buffer(s), each serving to build a different merged image. Mosaics from different input images are concatenated side by side by pair to form two adjacent fields (left field and right field) of a new series of merged frames of the original size such as **60**. In the example illustrated at Figure 2b, image **50'** is currently being processed, while processing of image **50** is completed, yielding a complete type A mosaic stored in the left field (A) of merged image **60.** It should be pointed out however that the merged frames do not necessarily comprise an image from the first sequence and an image from the second sequence, or images captured at the same time, as will be apparent from the detailed description of preferred embodiments of the compressing/encoding (mixing) method. As a matter of fact, in the example of Figure 2a, field A and field B of the merged image **60** are respectively filled with mosaic A and mosaic B from a same image **50**. While that situation has been chosen to simplify the illustration and corresponds to an actual situation according to one of the embodiments of the invention contemplated herein, it shall be deemed that merged images such as **60** could comprise mosaics originating from any of the inputted images. That side-by-side compressed transport format is totally transparent and unaffected by the compression/decompression processing characterizing the MPEG2 main view protocol downstream in the process.

Upon decoding of the merged images, reconstruction of the complete images is carried out by spatially interpolating missing pixels from the compressed half-size images (mosaics) located in the fields of the merged images such as **60**. As illustrated in Figure 2c, this is accomplished in real time when each pixel of an input merged frame **60** decoded in decoder 2 is being transferred to memory. Data of one pixel at a time is stored into a three-pixel input buffer **65**. As shown, the three pixels of the shadowed portion of input image **60** have been stored in input buffer **65**, two adjacent pixels from the same mosaic being identified as P_i and p_{i+1}. Data of a third pixel P_j is then calculated as being the arithmetic mean of each of the 3 components of the RGB vectors of adjacent pixels (P_i and P_{i+1}). For example, if pixel P_i has an intensity vector of (10,0,30) and pixel P_{i+1} has an intensity vector of (20,0,60), then, pixel P_j will be calculated as being (15,0,45). Therefore, the mean of two identical pixels is another identical pixel. That calculated (topologically interpolated) pixel will replace the missing pixel decimated upon creation of the mosaics from original image sequences such as **50**.

Pixels P_i, P_{i+1} and P_j are next stored in appropriate memory locations of a frame buffer where the corresponding image is to be reconstructed (image **72** in the present example). Passed the centre of each line of the merged frame **60** (entering the right field), data is stored into a second frame buffer **72'**, to rebuild the image from the mosaic stored in the right hand field of the stereo image. The process is followed line by line from left to right, until the two images are spatially reconstructed in their respective buffer. In order to assure flickerless viewing, the decoding method further comprises temporal expansion of image sequences as will be described in detail in the following description. When frame buffers are completely filled to provide complete rebuilt or temporally interpolated images (no more than three frame buffers required in any embodiment of the reconstruction process and system), they may be read according to different modes to provide different types of desired output signals.

A first embodiment of the mixing method carried out by mixer **1** according to the present invention is schematically represented in Figure 3a of the appended drawings and will now be described in detail.

A first sequence of images in RGB 24 fps format **50**, identified as L1 to L4, is first time expanded by 25% to form a 30 fps sequence of images such as **51**, by the creation and insertion of a new image **52** after every fourth image of the original sequence **50**. New image **52** is time-interpolated from the topological information of the immediately preceding and following images (#4 and #5 of original sequence **50**). Each pixel of the new image **52** is calculated as the arithmetic mean of the corresponding pixel in the precedent and following image, in a manner similar to the spatial interpolation technique explained in the foregoing description. Figure 2d provides a specific illustration of the time-interpolation process where a new image **52** is created from two time-successive images **50.4** and **50.5** of input image sequence **50**. Creation of new images in the present invention is generally accomplished according to that technique to provide improved fluidity at playback, as compared for instance to simple repetition of frames, which would require less processing power. Alternatively, any known method, such as movement anticipation based methods, could be used for performing time-interpolation of images.

Images of the time-expanded sequence **51** are then spatially compressed according to the technique illustrated in Figure 2b and described in detail in the foregoing description, to form mosaics represented by a new sequence **53**. Similarly, the second input image sequence **50'** is time expanded into the 30 fps sequence **51'** and spatially compressed into mosaics represented by sequence **53'**. In this specific embodiment, pairs of compressed images (mosaics) from the first sequence **53** and the second sequence **53'** respectively and representing scene **100** at a same time, are then concatenated to form a left field and a right field of merged images of a 30 fps RGB sequence **60**.

Since time-interpolated images may be inserted in the sequence (when input sequences comprise 24 fps), compressed sequence 60 becomes irregular. Therefore, the encoding (mixing) process according to the present embodiment of the invention further includes insertion of information in the compressed sequence **60** to enable identification of frame numbers as needed by the reconstruction process to identify image content and rebuild sequences with the proper sequential order (timing) and insert additional interpolated images at appropriate locations in the sequence. Such information may be stored in blank lines of merged images for instance. The usefulness of that procedure will become more apparent upon reading of the following description of the decoding process. This completes the mixing procedure per se carried out by mixer **1**. Further in the process, as described in reference to Figure 1a, RGB merged image sequence **60** (ex. AVI file) can be converted to a digital Y U V format prior to being multiplexed into an MPEG2 bit stream format or be directly converted into an MPEG2 format identified by numeral **62** in Figure 3a.

It is worth mentioning that in spite of the schematic diagram of Figure 3a, the processing of the image sequences is preferably not carried out in parallel and with long sequences of images. Actually, only three images are being buffered at a given time to enable temporal interpolation, and images are alternatively imported from the first and the second (left and right) sequences and processed on a pixel by pixel basis more like the process steps represented in Figures 2b and 2d.

The decoding and reconstruction carried out by decoder **2** according to the first embodiment of the present invention will now be described by referring to Figures 3b, 2c and 2d.

In the example shown in Figure 3b, stereo MPEG signal **62** is read from a DVD and is converted by the DVD player **15** (Fig. 1b) into an analog NTSC signal **70** inputted by the decoder **2**. NTSC signal **70** is first converted into an RGB format to recuperate merged images such as in sequence **60**. Reconstruction of first and second original sequences **50**, **50'** can then be started by spatially interpolating and separating mosaics from the left and right fields of the merged images from sequence **60** on a pixel by pixel basis, as previously described with reference to Figure 2c, to form 30 fps decompressed buffered images such as **72** and **72'**. Therefore, spatial interpolation and separation are actually carried out simultaneously. Sequences of RGB images **72** and **72'** could be directly outputted and displayed on a dual input device to reproduce the original programs or stereoscopic program signals at a 60 fps (30 per eye) presentation rate. Further processing could also be performed to present the image sequences in an interlaced mode or in sequence (page flipping mode), anaglyphic, column interleaved, conventional 2D mode, etc. on a plurality of existing single input display devices.

However, in order to enable comfortable and fatigue free viewing, decoder **2** significantly reduces flicking by providing output signals at a typical rate of 36 full definition frames per eye per second, while satisfying results may be obtained at 30 fps per eye with high definition frames to match refresh rates of SDTV or HDTV for instance. On the other hand, output signals up to 120 fps (60 images per second per eye) can be provided by decoder **2** for a very high fidelity reproduction, such an output being compatible however with display devices such as DLP projectors and a limited number of high end devices. By experience, a playback rate of 72 fps provides very good results, provided image quality is preserved throughout the coding/decoding process as contemplated herein, such a frequency being a standard for most display devices currently encountered in home theatre systems.

Therefore, the playback process carried out by decoder **2** preferably includes a further step to increase the presentation rate of sequences **72** and **72'**. Additional images are inserted at regular intervals in the image sequence, using the temporal interpolation technique already explained in the foregoing description of the mixing process referring to Figures 3a and 2c. The position of insertion can be carefully controlled through the frame number information stored in blank lines of input sequence **60** at mixing. Alternatively, images from sequences **72** and **72'** can be repeated (read twice) to increase the rate at presentation. For instance every image of the sequences could be read twice to double the rate of presentation.

In the example illustrated in Figure 3b, one new intermediate image pair **73, 73',** is time-interpolated using information from images #2 and #3 of sequences **72** and **72'** respectively and inserted between images #2 and #3 to thereby increase the rate of the resulting sequences **74** and **74'** to 36 fps (total of 72 fps for the stereoscopic program). The process is partly illustrated in Figure 2c, where images #2 and #3 of sequence **72** are identified as **72.2** and **72.3**. Alternatively, further images could be time-interpolated and inserted to provide a rate of say 48 frames per second per sequence for a total of 96 fps. A rate of 60 fps per sequence (eye) (total 120 fps for the stereoscopic program) is also an interesting case where no interpolation is required. All of the images of sequences **72** and **72'** are merely duplicated to double the number of images. At playback, shutter spectacles are driven at a rate of 120 Hz and all the images of a given sequence are presented twice to the corresponding eye in 1/30 s. An unsurpassed clarity is thereby provided, but only a very limited range of display devices can handle such a high refresh rate.

It should be noted that the foregoing description has been based on the fact that input sequences are supplied at a rate of 24 fps, which is common for motion picture movies. However, one can easily appreciate that the mixing process can be easily adapted to the case whereby two 30 fps sequences (ex. TV programs) would be supplied, by merely skipping the preliminary step of temporal interpolation represented by time-expanded sequences **51** and **51'** of Figure 3a. Obviously, since the decoding process always operates on a 30 fps input sequence, no substantial adaptation is required to that part of the process.

A second embodiment of the mixing method carried out by mixer **1** according to the present invention will now be described in detail, by reference to Figure 4a of the appended drawings. This second embodiment is particularly advantageous for addressing the problem of converting two image sequences available in a 24 fps format to produce a 30 fps MPEG2 (main view profile) fully compatible sequence.

Full definition images from the two 24 fps sequences **50** and **50'** comprising mosaics A and B by definition are identified as L_iAB and R_iAB respectively (supposing two sequences of a stereoscopic program), index "i" representing the sequential number of a given image at time t_i. Dashed lines in Figures 4a and 4b indicate frame sequence. In a similar manner as for the first embodiment previously described, eight input images are spatially compressed and time-expanded to form five new merged images in a new 30 fps sequence **80**. It should be noted that in this embodiment of the present invention, 25% more of the original image information is preserved to be recorded or broadcasted. Indeed, two out of the eight original images (images L_1 and L_2 in the example shown) have both of their mosaics A and B saved in fields of the compressed sequence **80** instead of one according to the first embodiment.

These fully saved images are nevertheless encoded in the form of two complementary mosaics stored in side-by-side merged images fields to ascertain homogeneity of the encoded sequence and compatibility with the MPEG2 compression/decompression protocol, by providing a certain temporal redundancy between successive images. Better definition and fidelity is thus generally obtained at playback with respect to the previously described embodiment, but at the expense of increased processing power requirement and system hardware cost. As for the above-described first embodiment, the encoding (mixing) process according to the present embodiment of the invention also further includes insertion of information in the compressed sequence **80** to enable identification of frame numbers as needed by the reconstruction process to identify image content and rebuild sequences with the proper sequential order and insert interpolated images at appropriate locations in the sequence. Again, such information may be stored in blank lines of merged images for instance.

The corresponding decoding process carried out by decoder **2** according to the present invention is schematically represented in Figure 4b and operates as follows.

The five merged frames **81** to **85** representative of 30 fps RGB input sequence **80** are expanded to twelve images (six per channel) providing playback sequences **90** and **100** at 36 fps total (72 total, 36 per eye in the case of a tree-dimensional stereoscopic program). In total, each group of twelve successive images of playback sequences **90** and **100**, presented in a page flipping mode according to the frame sequence indicated by dashed lines **110,** comprises two integral original images, six spatially interpolated images and four temporally interpolated images. Alternatively, sequences **90** and **100** could be outputted separately in parallel on two separate channels, as required by some display devices such as a head mounted or auto-stereoscopic devices. In the illustrated example:

1. Image **91** (L_1AB) is totally rebuilt from mosaic L_1A stored in the left field of frame **81** of sequence **80**, and mosaic L_1B stored in the right field thereof;
2. Image **101** (R_1AX) is spatially interpolated from mosaic R_1A, taken from the left field of frame **82** of sequence **80**;
3. Image **103** (R_2BX) is spatially interpolated from mosaic R_2B, taken from the right field of frame **82** of sequence **80**;
4. Image **102** is temporally interpolated from image **101** and image **103**;
5. Image **93** (L_2AB) is totally rebuilt from mosaic image L_2A, stored in the left field of frame **83** of sequence **80**, and mosaic L_2B stored in the right field thereof;
6. Image **92** is temporally interpolated from image **91** (L_1AB) and image **93** (L_2AB);
7. Image **94** (L_3AX) is spatially interpolated from mosaic L_3A, stored in the left field of frame **84** of sequence **80**;
8. Image **96** (L_4BX) is spatially interpolated from mosaic L4B, stored in the right field of frame **84** of sequence **80**;
9. Image **95** is temporally interpolated from images **94** and **96**;
10. Image **104** (R_3AX) is spatially interpolated from mosaic R_3A, stored in the left field of frame **85** of sequence **80**;
11. Image **106** (R_4BX) is spatially interpolated from mosaic R_4B, stored in the right field of frame **85** of sequence **80**; and

12. Image **105** is temporally interpolated from image **104** and image **106**.

Obviously, one may easily understand that such a reconstruction process requires proper identification of frame order in the 5 frame sequences constituting input sequence **80**. Therefore, a frame recognition circuit is provided in decoder **2** to interpret frame number information stored by mixer **1** in merged image sequence **80**.

It can be observed that in this latter embodiment as well as in the first one disclosed in the foregoing description, the first and second image sequences are being encoded and decoded totally independently, without any inference between each other, enabling processing of original video sequences referring to independent scenes.

The above described example of the second embodiment, processing sources at 24 fps to yield a presentation rate of 72 fps, is only illustrative of a more general process applicable to 24 or 30 fps sources to produce a stereo output at presentation rates such as 60, 72, 96 or 120 fps. The chart below provides additional exemplary arrangements for 24 or 30 fps sources and 60, 72, 96 or 120 fps presentation rates:

Source (fps)	Output (fps)	Original images	Spatially-interpolated images	Temporally-interpolated images	Repeated images
24+24	60	12	36	12 or 0	0 or 12
24+24	72	12	36	24	0
24+24	96	12	36	0	48
24+24	120	12	36	12	60
30+30	60	0	60	0	0
30+30	72	0	60	12 or 0	0 or 12

30+30	96	0	60	36	0
30+30	120	0	60	0	60

As stated above, RGB sequences **90** and **100** obtained through the above described processing could be directly outputted and displayed on a dual input device to reproduce the original programs or stereoscopic program signals at a 72 fps (36 per eye) presentation rate. Further processing is however carried out by decoder **2** to provide a combined stereoscopic RGB output signal (not shown) comprising images of sequences **90** and **100** in a time sequenced arrangement as indicated by dashed arrows such as **110**. Still referring to the example of Figure 4b, images would be time sequenced by alternating left eye and right eye images in the following order: **91**, **101**, **92**, **102**, **92**, **103**, **94**, **104**, **95**, **105**, **96**, **106**. This is accomplished through an appropriate read sequence of the complete images stored in memory buffers.

Presentation of the time sequenced combined signal with a standard projector or another progressive scan display device is thus enabled to display the stereoscopic program in a page-flipping mode. Decoder **2** provides the necessary timing signals to a driver of shutter spectacles which can be worn by a viewer to view the displayed stereoscopic program in a three-dimensional mode, with high fidelity, negligible flicking and high comfort. As stated above, presentation rate can be increased up to 120 fps by inserting additional temporally interpolated image pairs or by repeating certain image pairs in the decoding process. It is also contemplated in the present invention that the RGB combined stereo output signal could be converted to another known standard presentation format such as an interlaced format or a conventional 2D format.

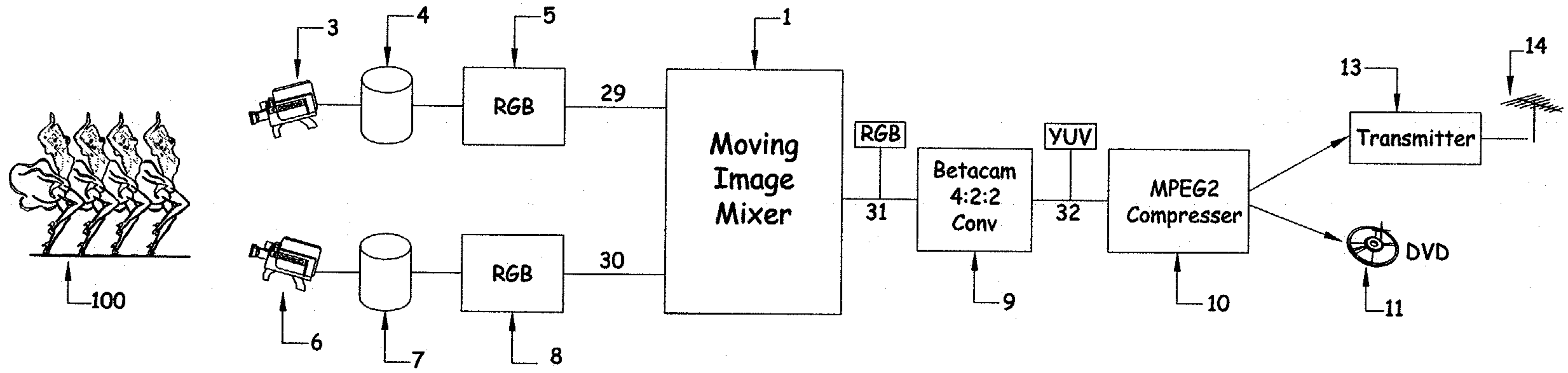
It is further contemplated in the present invention that sound tracks normally accompanying planar visual programs could similarly be associated with the stereo encoded programs on the transport medium. At playback, sound tracks would normally be directed to a sound reproduction system and outputted as they are inputted from said transport medium. However, since processing of image sequences by playback decoder **2** causes a given time delay between its input and output, said decoder also delays the audio tracks by the same time delay to ensure that sound tracks and the visual program are played with proper temporal synchronization. Since the time delay introduced by a given image processing mode is a known invariable information, proper sound delaying can accordingly be automatically selected and performed by the decoder **2**.

Therefore, one can easily appreciate that the above described embodiments of the present invention provide effective and practical solutions for the recording of two motion picture sequences on a conventional data storage medium, and playback with conventional videodisk player or broadcast source and display device, to enable viewing of stereoscopic 3D movies at home with unmatched performance and comfort, still at an affordable cost, in a plurality of output modes to match input signal requirement of a broad range of display devices. For example a universal set top box fed with a single input signal format as defined in the foregoing description, can be provided with selectable modes such as: page flipping, row interleaved, column interleaved, simultaneous dual presentation, anaglyphic, etc. The encoding/playback method and system of the present invention can thus be advantageously used in miscellaneous applications, including the processing of video sequences representing independent scenes, with numerous advantages over the solutions of the prior art.

Although the present invention has been described by means of preferred embodiments thereof, it is contemplated that various modifications may be made thereto without departing from the spirit and scope of the present invention. Accordingly, it is intended that the embodiment described be considered only as illustrative of the present invention and that the scope thereof should not be limited thereto but be determined by reference to the claims hereinafter provided and their equivalents.

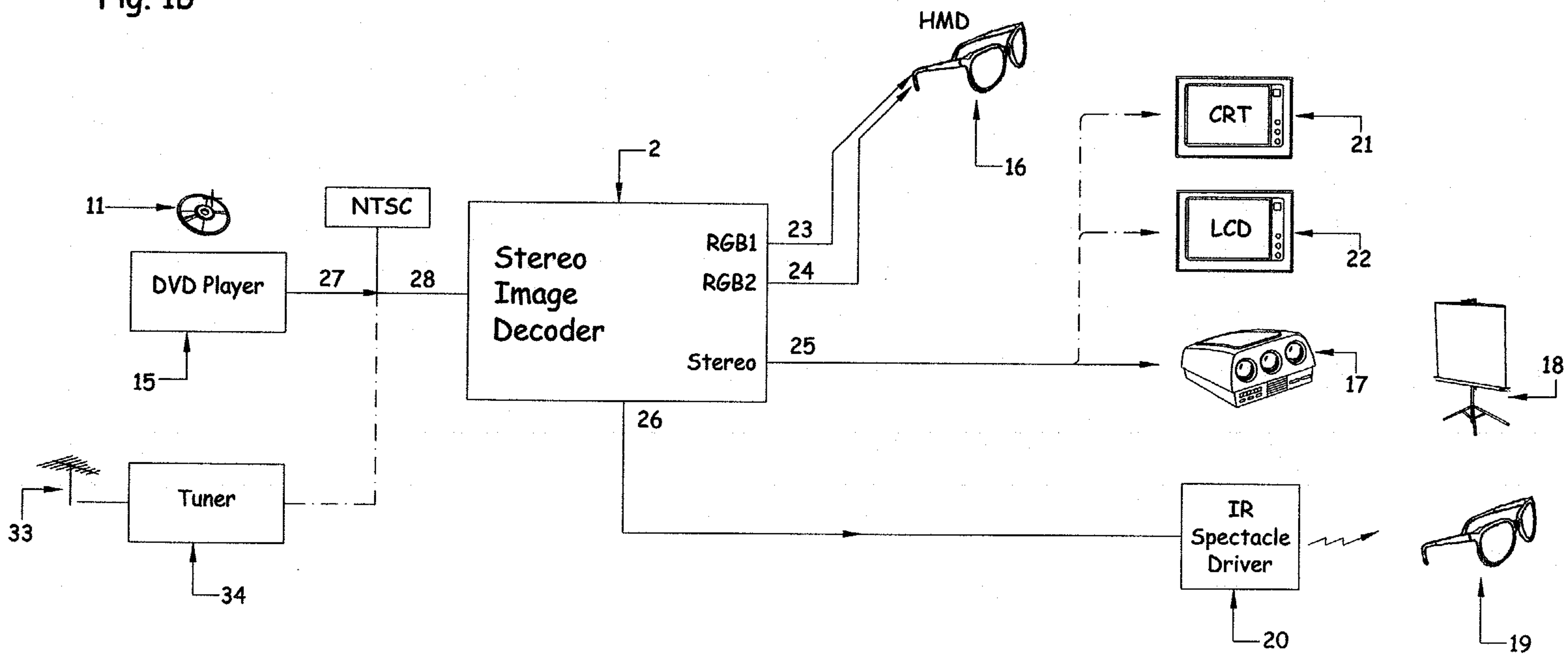
Inventors

Fig. 1a



Moving Image Mixer
RGB
RGB
RGB
YUV
Betacam 4:2:2 Conv
MPEG2 Compresser
Transmitter
DVD
1
3
4
5
6
7
8
9
10
11
13
14
29
30
31
32
100

Fig. 1b



DVD Player
Tuner
NTSC
Stereo Image Decoder
RGB1
RGB2
Stereo
HMD
CRT
LCD
IR Spectacle Driver
2
11
15
16
17
18
19
20
21
22
23
24
25
26
27
28
33
34

Fig. 2a Separation

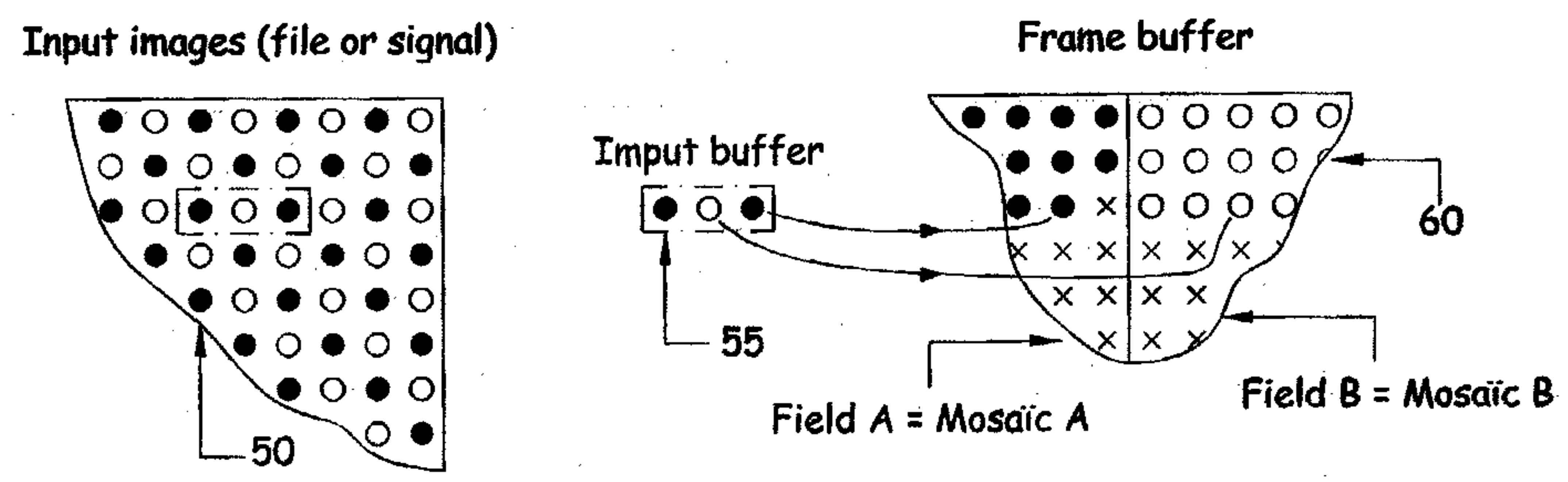


Fig. 2b Decimation

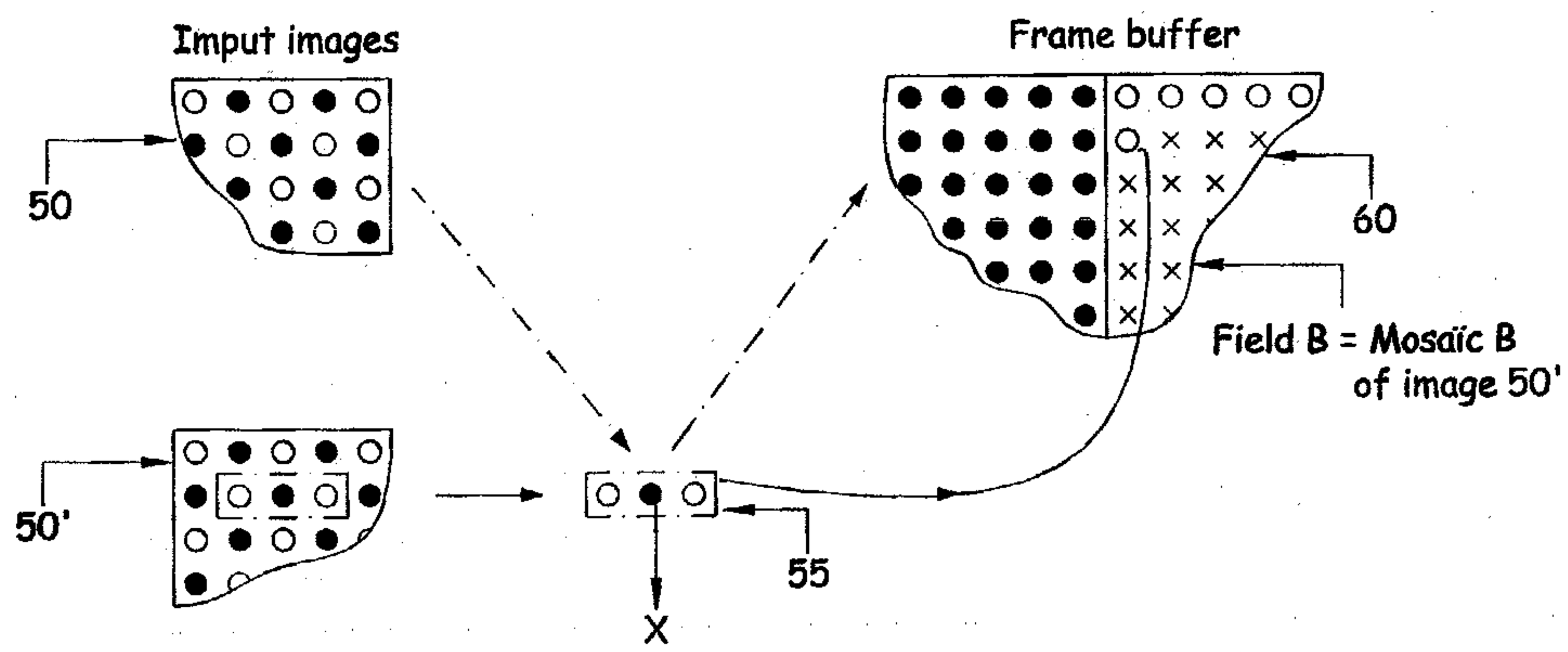


Fig. 2c

Spatial interpolation

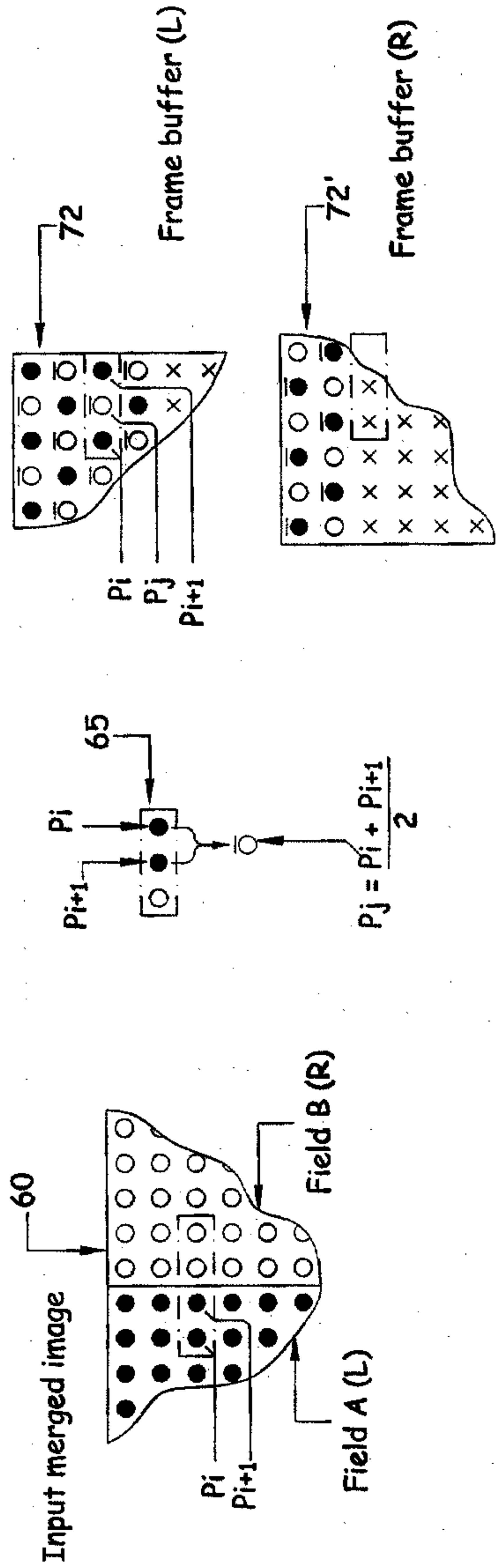


Fig. 2d

Time interpolation

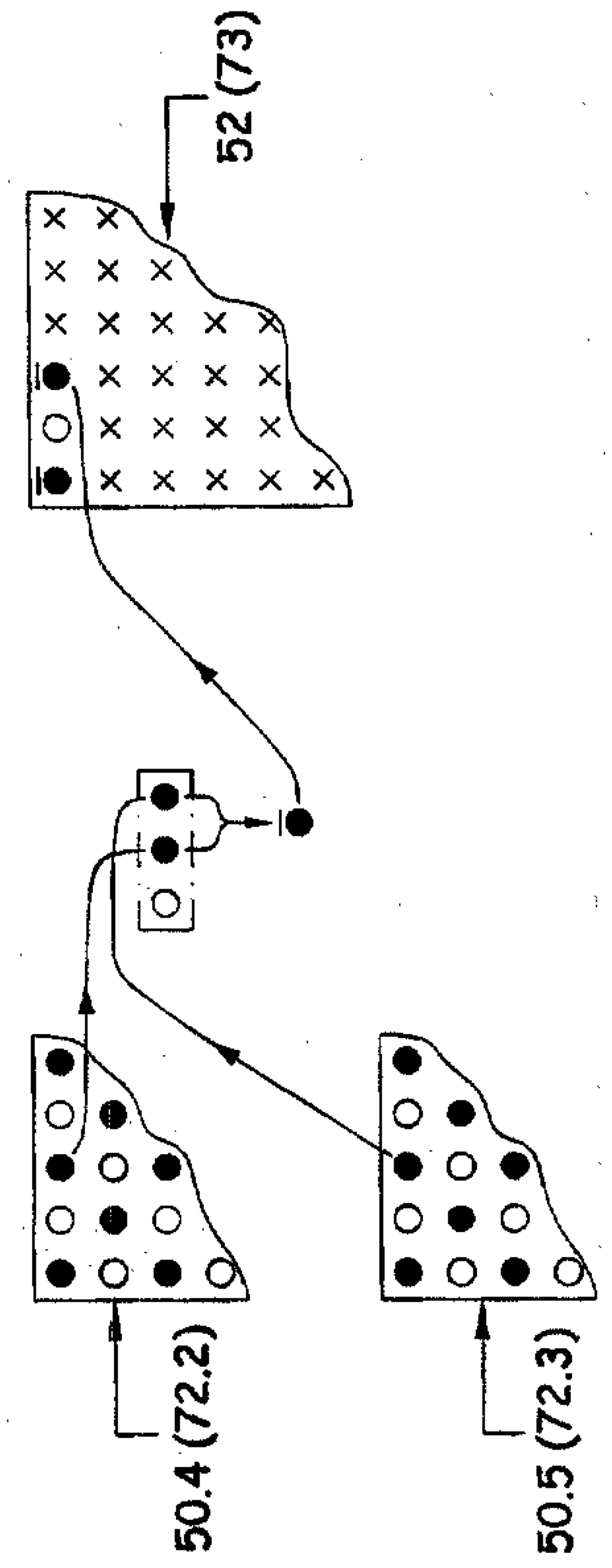
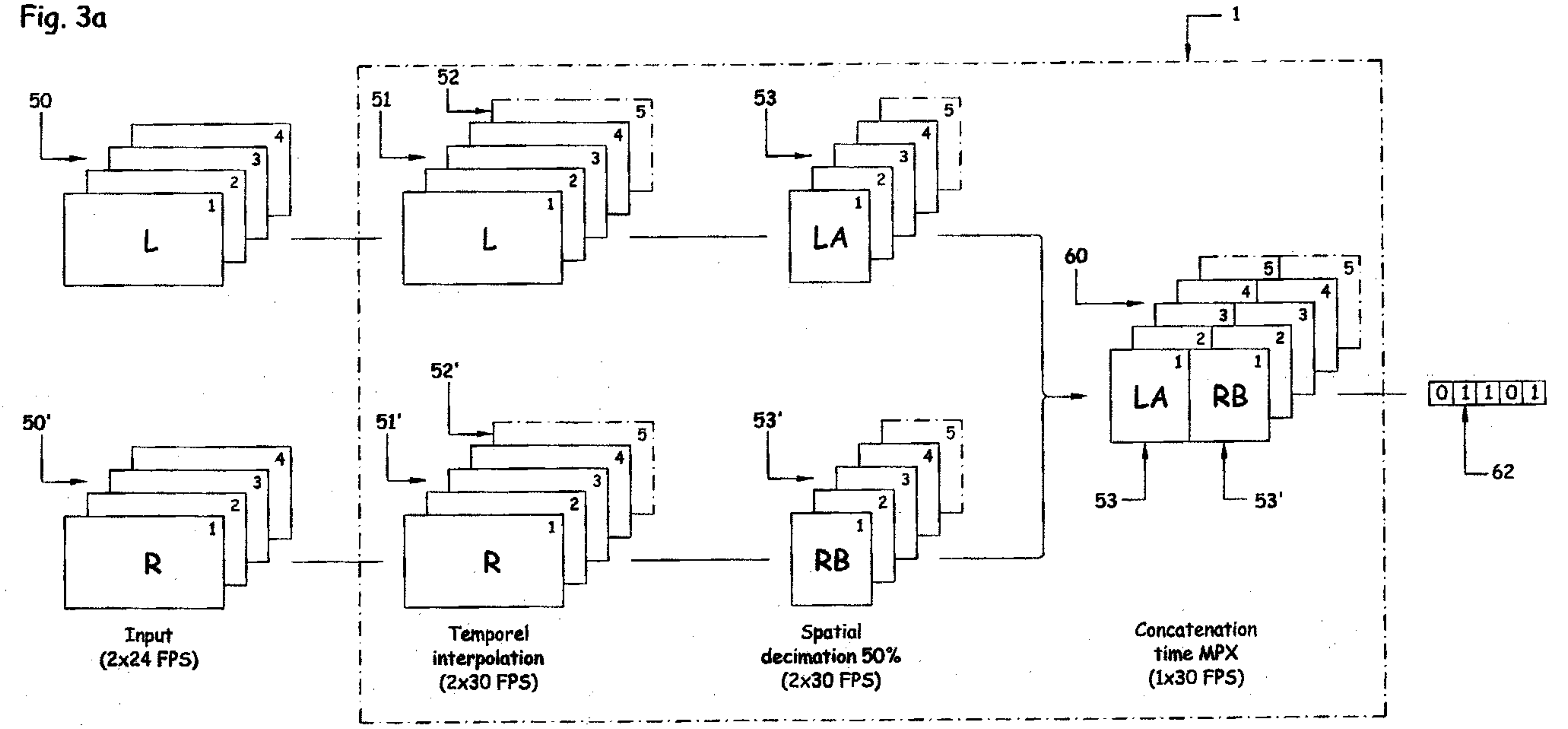
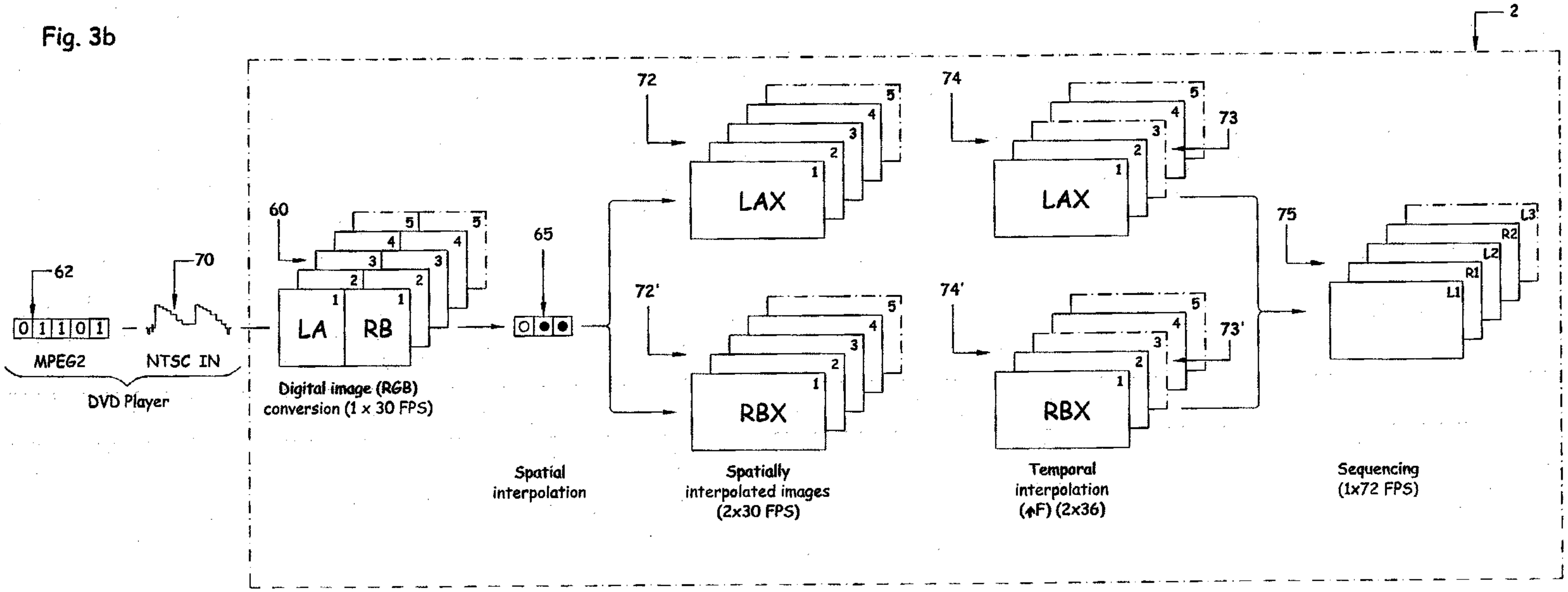


Fig. 3a



1
50
50'
L
R
Input
(2x24 FPS)
51
52
51'
52'
Temporel
interpolation
(2x30 FPS)
53
53'
LA
RB
Spatial
decimation 50%
(2x30 FPS)
60
Concatenation
time MPX
(1x30 FPS)
0 1 1 0 1
62

Fig. 3b



2
62
0 1 1 0 1
MPEG2
70
NTSC IN
DVD Player
60
LA
RB
Digital image (RGB)
conversion (1 x 30 FPS)
65
Spatial
interpolation
72
72'
LAX
RBX
Spatially
interpolated images
(2x30 FPS)
74
74'
73
73'
Temporal
interpolation
(↑F) (2x36)
75
L1
R1
L2
R2
L3
Sequencing
(1x72 FPS)

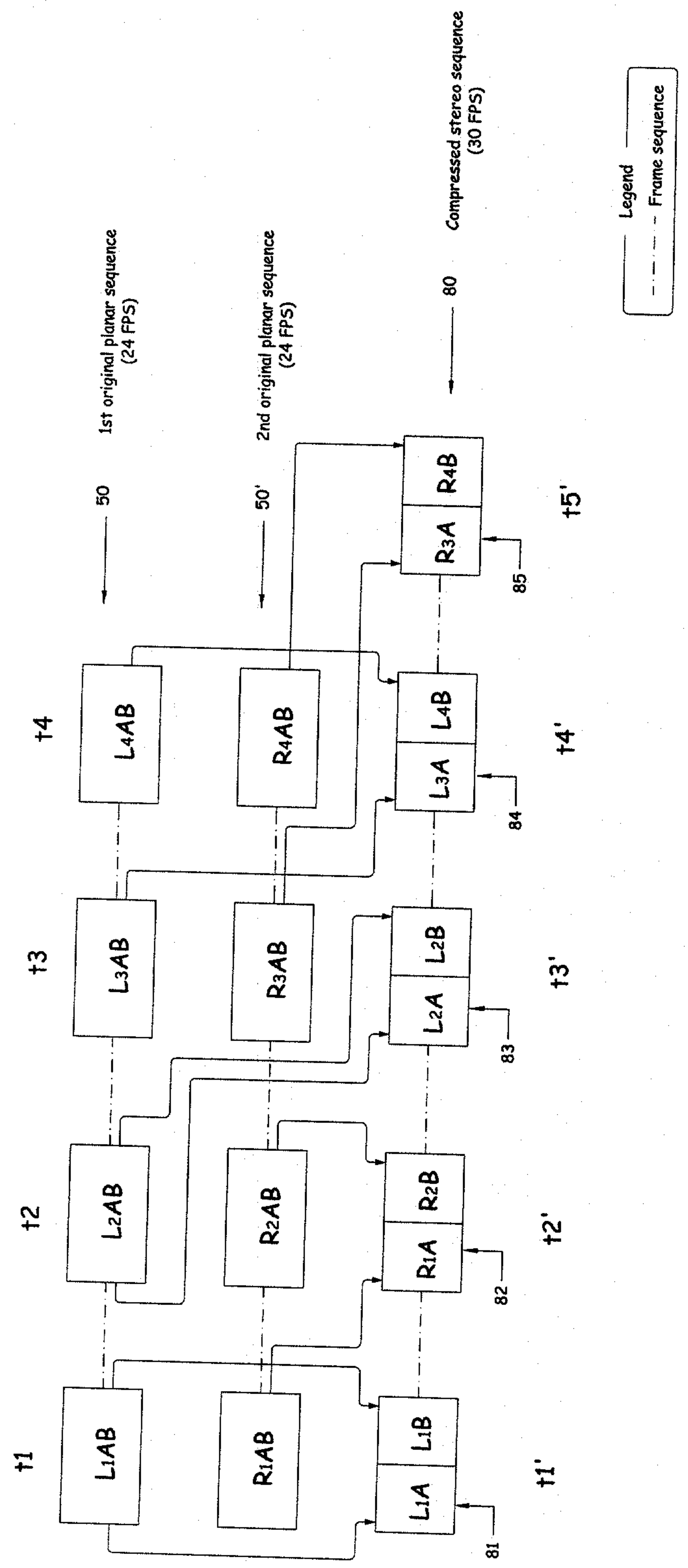

t1
t2
t3
t4
L1AB
L2AB
L3AB
L4AB
R1AB
R2AB
R3AB
R4AB
50
1st original planar sequence
(24 FPS)
50'
2nd original planar sequence
(24 FPS)
80
Compressed stereo sequence
(30 FPS)
L1A
L1B
R1A
R2B
L2A
L2B
L3A
L4B
R3A
R4B
81
82
83
84
85
t1'
t2'
t3'
t4'
t5'
Legend
Frame sequence

Fig. 4a

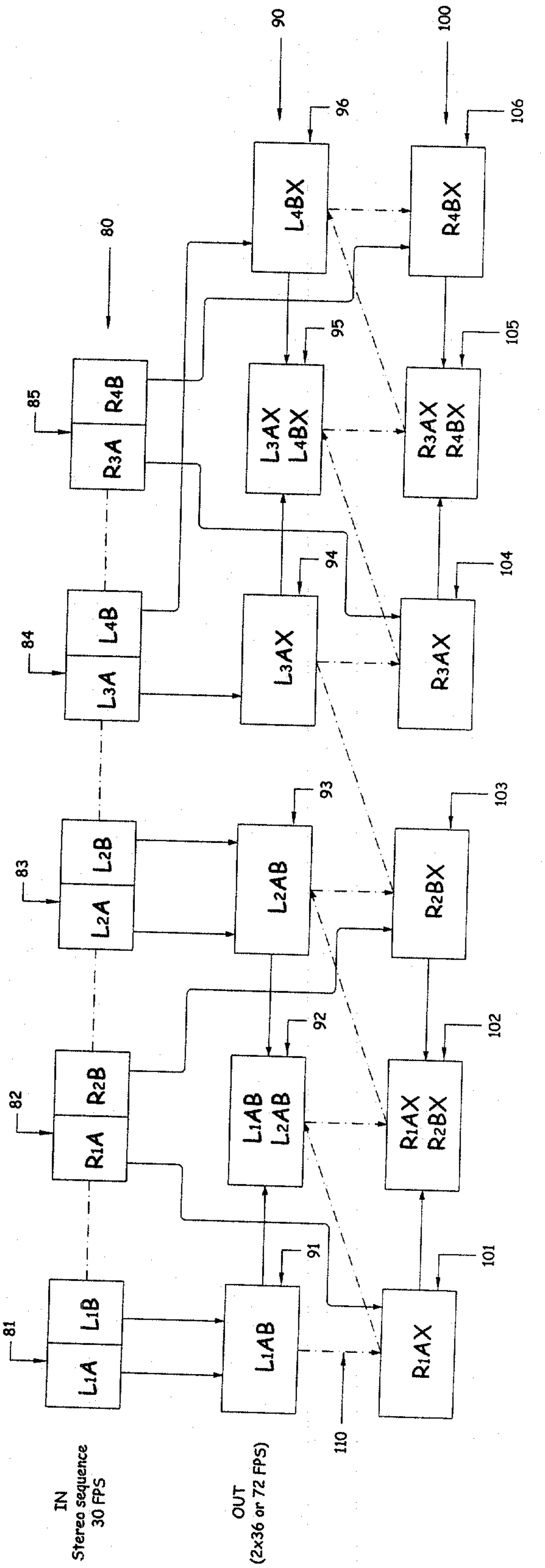

IN
Stereo sequence
30 FPS
OUT
(2x36 or 72 FPS)
80
81
82
83
84
85
L1A
L1B
R1A
R2B
L2A
L2B
L3A
L4B
R3A
R4B
L1AB
L1AB
L2AB
L2AB
L3AX
L3AX
L4BX
L4BX
R1AX
R1AX
R2BX
R2BX
R3AX
R3AX
R4BX
R4BX
90
91
92
93
94
95
96
100
101
102
103
104
105
106
110

Fig. 4b


HMD
CRT
LCD
2
11
NTSC
Stereo
Image
Decoder
RGB1
RGB2
Stereo
23
24
25
26
27
28
DVD Player
15
16
21
22
17
18
Tuner
33
34
IR
Spectacle
Driver
20
19