



HU000027682T2

(19) **HU****MAGYARORSZÁG**
Szellemi Tulajdon Nemzeti Hivatala(11) Lajstromszám: **E 027 682**(13) **T2****EURÓPAI SZABADALOM**
SZÖVEGÉNEK FORDÍTÁSA(21) Magyar ügyszám: **E 11 749928**(22) A bejelentés napja: **2011. 07. 28.**

(96) Az európai bejelentés bejelentési száma:

EP 20110749928

(97) Az európai bejelentés közzétételi adatai:

EP 2599319 A1 **2012. 02. 02.**

(97) Az európai szabadalom megadásának meghirdetési adatai:

EP 2599319 B1 **2015. 12. 16.**(51) Int. Cl.: **H04N 13/00** (2006.01)

(86) A nemzetközi (PCT) bejelentési szám:

PCT/IB 11/053361

(87) A nemzetközi közzétételi szám:

WO 12014171

(30) Elsőbbségi adatok:

TO20100652 **2010. 07. 28.** **IT**

(72) Feltaláló(k):

PENNISI, Dario, I-00050 Loc. Aranova Fiumicino (rm)
(IT)**BALLOCCA, Giovanni, I-10131 Torino (IT)****D'AMATO, Paolo, I-00195 Roma (IT)**

(73) Jogosult(ak):

S.I.SV.EL. Societa' Italiana per lo Sviluppo
dell'Elettronica S.p.A., 10060 None (TO) (IT)

(74) Képviselő:

Danubia Szabadalmi és Jogi Iroda Kft.,
Budapest

(54)

Eljárás egy háromdimenziós tartalomra vonatkozó képek kombinálására

Az európai szabadalom ellen, megadásának az Európai Szabadalmi Közlönyben való meghirdetésétől számított kilenc hónapon belül, felszólalást lehet benyújtani az Európai Szabadalmi Hivatalnál. (Európai Szabadalmi Egyezmény 99. cikk(1))

A fordítást a szabadalmas az 1995. évi XXXIII. törvény 84/H. §-a szerint nyújtotta be. A fordítás tartalmi helyességét a Szellemi Tulajdon Nemzeti Hivatala nem vizsgálta.



(11) **EP 2 599 319 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
16.12.2015 Bulletin 2015/51

(21) Application number: **11749928.5**

(22) Date of filing: **28.07.2011**

(51) Int Cl.:
H04N 13/00 (2006.01)

(86) International application number:
PCT/IB2011/053361

(87) International publication number:
WO 2012/014171 (02.02.2012 Gazette 2012/05)

(54) **METHOD FOR COMBINING IMAGES RELATING TO A THREE-DIMENSIONAL CONTENT**

VERFAHREN ZUR KOMBINATION VON BILDERN IM ZUSAMMENHANG MIT EINEM
DREIDIMENSIONALEN INHALT

PROCÉDÉ POUR LA COMBINAISON D'IMAGES CONCERNANT UN CONTENU
TRIDIMENSIONNEL

(84) Designated Contracting States:
**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB
GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO
PL PT RO RS SE SI SK SM TR**

(30) Priority: **28.07.2010 IT TO20100652**

(43) Date of publication of application:
05.06.2013 Bulletin 2013/23

(73) Proprietor: **S.I.SV.EL.
Societa' Italiana per lo Sviluppo dell'Elettronica
S.p.A.
10060 None (TO) (IT)**

(72) Inventors:
• **PENNISI, Dario
I-00050 Loc. Aranova
Fiumicino (RM) (IT)**

• **BALLOCCA, Giovanni
I-10131 Torino (IT)
• D'AMATO, Paolo
I-00195 Roma (IT)**

(74) Representative: **Camolese, Marco et al
Metroconsult S.r.l.
Via Sestriere 100
10060 None (TO) (IT)**

(56) References cited:
**EP-A1- 1 705 929 EP-A1- 2 157 803
WO-A1-2010/064118 WO-A2-2008/038205
US-A1- 2010 103 168**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description**- TECHNICAL FIELD -**

[0001] The present invention relates to methods and devices for combining, within a stereoscopic display device, locally generated images superimposed on a three-dimensional content received by the device itself.

- PRIOR ART -

[0002] It is known that television apparatuses (television sets and decoders) can generate locally images containing text and graphics and superimpose them on the images being received; it is thus possible to supply useful information of various kinds to the user as the video is being played in the background.

[0003] These images can be generated by starting from information received along with the video signal (as is the case, for example, of subtitles and some electronic program guides, also known as EPGs), or they may provide information about the configuration and the settings of the decoder or television set (e.g. menus or the bar indicating the volume level and other parameters).

[0004] Nowadays the quantity of 3D contents available on the market has grown considerably; the fruition of these contents is no longer limited to cinemas, and users can watch 3D videos at home on their own television set.

[0005] Therefore, also for 3D videos there is a need to superimpose locally generated images on the television images being received.

[0006] Compared to a 2D video stream, image superimposition on a 3D video stream is more complex, since it is necessary to take into account the different depth arrangement of the objects included in the single images making up the video stream.

[0007] Patent application EP2157803A1 teaches how to position a text in such a position that it always remains in front of the television image. In particular, if the 3D content is broadcast as a two-dimensional image plus a depth matrix, the latter can also be used for defining the position of the text.

[0008] This solution has the drawback that it employs a very large depth map, since this is used for creating the pair of right and left images (which, when combined together, produce the 3D effect) by starting from a basic two-dimensional image. In addition to requiring a considerable computational effort when the map must be analyzed to define the text position, the size of this map also involves high bandwidth usage when the map is transmitted to a receiver.

[0009] Patent application WO2008/038205 describes a method for composing 3D images (to be displayed on a 3D display) composed by main 3D images and additional 2D or 3D images, as texts or graphics, not superimposed to the main ones. Every type of image is composed by a 2D image and a relating depth map. The depth map of every type of image is combined in recep-

tion with a corresponding 2D image for reconstructing the image to be displayed.

[0010] Patent application EP-1705929 describes a method for organizing the transmission of 3D images by combining in a frame one of the two 2D images composing the 3D image, with information relating to the depth, useful to rebuild the 3D image in reception.

- OBJECTS AND BRIEF DESCRIPTION OF THE INVENTION -

[0011] The object of the present invention is to provide a method and a system for combining images with a three-dimensional content transported by a stereoscopic video stream, which allows to overcome the drawbacks of the prior art.

[0012] In particular, it is one object of the present invention to provide a method for superimposing images on 3D content which requires a lower computational cost at 3D content player level.

[0013] It is another object of the present invention to provide a method for transmitting information necessary for superimposing images on those transported by a stereoscopic video stream, which does not require high bandwidth usage and which is robust to the stereoscopic video stream coding and decoding operations.

[0014] These and other objects of the present invention are achieved through a method and a system for superimposing images on those transported by a stereoscopic video stream incorporating the features set out in the appended claims, which are intended as an integral part of the present description.

[0015] The general idea at the basis of the present invention is to display an element superimposed on a stereoscopic stream by using in the playback phase a superimposition depth map coded into an image contained in a frame of a stereoscopic stream. The depth map used in the present invention is not aimed at coding the stereoscopic video stream, since it is transmitted with the only purpose of supplying the decoder or the television set with information useful for superimposing locally generated images on the stereoscopic image in an appropriate manner. To this end, the depth map has a lower resolution, and therefore a smaller number of pixels, than that of the stereoscopic pair, thus limiting bandwidth occupation. This is possible because the map is not used for generating the three-dimensional image, but only for properly positioning the overlays.

[0016] In a preferred embodiment, the frame carries a composite image that comprises a right image, a left image and the depth map, appropriately multiplexed.

[0017] In one embodiment, the right and left images are arranged in accordance with a traditional format, e.g. a side-by-side, top-bottom or checkerboard format, whereas the depth map is inserted into a free area of the composite frame and is not intended for display.

[0018] In an alternative embodiment, the right and left images are arranged in the composite frame in an inno-

vative manner. In this embodiment, the frame comprises a number of pixels which is higher than the sum of the pixels of the original format (i.e. prior to coding) of both the right and left images, which are thus entered without being subject to decimation.

[0019] In this embodiment, the pixels of the first image (e.g. the left image) are entered into the composite image without undergoing any changes, whereas the second image is subdivided into regions whose pixels are arranged in free areas of the composite image. This solution offers the advantage that one of the two images is left unchanged, which results in better quality of the reconstructed image.

[0020] Advantageously, the second image is then broken up into the smallest possible number of regions, so as to maximize the spatial correlation among the pixels and reduce the generation of artifacts during the compression step.

[0021] In an advantageous embodiment, the regions of the second image are entered into the composite image by means of translation or rototranslation operations only, thus leaving unchanged the ratio between horizontal and vertical resolution.

[0022] In a further embodiment, at least one of the regions into which the second image has been broken up undergoes a specular inversion step, i.e. it is overturned relative to one axis (in particular one side) and is arranged in the composite image in a manner such that one of its sides borders on one side of the other image having identical or similar pixels on the bordering side due to the strong correlation existing between homologous pixels of the two right and left images, i.e. pixels of the two images which are positioned in the same row and column.

[0023] This solution offers the advantage of reducing the generation of artifacts in the boundary area. More advantageously, the regions into which the second image is subdivided have a rectangular shape; compared to the solution that uses triangular regions arranged with boundary areas crossing the composite image in diagonal directions, this choice provides a reduction of the artifacts produced by a subsequent compression step, especially if the latter acts upon square blocks of pixels (e.g. 16x16 for the H.264 standard).

[0024] According to a particularly advantageous embodiment, the formation of artifacts is further reduced or even completely eliminated by introducing redundancy in the composite image, i.e. by copying some groups of pixels several times. In particular, this is attained by breaking up the basic image to be entered into the composite image into regions having such dimensions that the total number of pixels of these regions exceeds the number of pixels of the image to be broken up. In other words, the image is broken up into regions, of which at least two comprise an image portion in common. The common image portion is a boundary area between regions adjacent to each other in the disassembled image. The size of this common portion preferably depends on the type of compression to be subsequently applied to

the composite image, and may act as a buffer area which will be partially or completely removed when the disassembled image is reconstructed. Since compression may introduce artifacts in the boundary areas of said regions, by eliminating the buffer areas, or at least the outermost part thereof, it is possible to eliminate any artifacts and reconstruct an image which is faithful to the original one.

[0025] Further objects and advantages of the present invention will become more apparent from the following description of a few embodiments thereof, which are supplied by way of non-limiting example.

- BRIEF DESCRIPTION OF THE DRAWINGS -

[0026] Said embodiments will be described with reference to the annexed drawings, wherein:

Fig. 1 shows a block diagram of a device for multiplexing the right image and the left image into a composite image;

Fig. 2 is a flow chart of a method executed by the device of Fig. 1;

Fig. 3 shows a first form of disassembly of an image to be entered into a composite image;

Fig. 4 shows a first step of constructing a composite image according to one embodiment of the present invention;

Fig. 5 shows the complete composite image of Fig. 4;

Fig. 6 shows a second form of disassembly of an image to be entered into a composite image;

Fig. 7 shows a composite image that includes the image of Fig. 6;

Fig. 8 shows a third form of disassembly of an image to be entered into a composite image;

Fig. 9 shows a composite image that includes the image of Fig. 8;

Fig. 10 shows a block diagram of a receiver for receiving a composite image generated according to the method of the present invention;

Fig. 11 shows some steps of reconstructing the image disassembled according to the method of Fig. 8 and entered into the composite image received by the receiver of Fig. 10;

Fig. 12 is a flow chart of a method for reconstructing the right and left images multiplexed into a composite image of the type shown in Fig. 9;

Fig. 13 shows a composite image according to a fourth embodiment of the present invention;

Figs. 14a to 14f show a right image and a left image in different processing steps carried out for entering them into the composite image of Fig. 13.

[0027] Where appropriate, similar structures, components, materials and/or elements are designated by means of similar references in different figures.

- DETAILED DESCRIPTION OF THE INVENTION -

[0028] Fig. 1 shows a block diagram of a device 100 for generating a stereoscopic video stream 101 with a depth map for superimposing images on a video content transported by the video stream.

[0029] For the purposes of the present invention, a three-dimensional (or 3D) content is an image or a video which is perceived by the observer as having variable depth, where elements can protrude from the screen plane on which said image or video is being displayed or projected.

[0030] The expression "to superimpose two images" refers herein to any form of combination of two images, e.g. in transparency, half-transparency or complete opacity.

[0031] The present invention equally applies to any type of superimposition, whether static or dynamic, i.e. having fixed or time-variable graphic characteristics, which in turn may be either two-dimensional or three-dimensional.

[0032] The depth of a three-dimensional content relates to the dimension of the three-dimensional content which enters into the screen along an axis orthogonal to the screen on which the 3D content is being displayed. For the purposes of the present description, the screen corresponds to a zero depth point, while the "minimum depth" point is that point of the 3D content which is perceived by the user as closest to him/herself, i.e. farthest from the screen. Accordingly, the "maximum depth" point is that point which is perceived by the observer as deepest into the screen, i.e. farthest from him/herself, even beyond the screen plane.

[0033] In Fig. 1, the device 100 is inputted two image sequences 102 and 103, e.g. two video streams, respectively intended for the left eye (L) and the right eye (R), and a depth map sequence 106. Each depth map in the sequence 106 is associated with one pair of right and left images belonging to the sequences 102 and 103, respectively. In this embodiment, the depth map is generated by means of per se known algorithms which compare a right image with a left image and return a matrix (the depth map) having a size equal to the pixels of one of the two compared images, and whose elements have a value which is proportional to the depth of each pixel being displayed. Another depth map generation technique is based on measuring the distance of the object in the scene from the pair of video cameras that are shooting the scene: This distance can be easily measured by means of a laser. In the case of artificial video streams generated with the help of electronic computers, the video cameras are virtual ones, in that they consist of two points of view of a certain scene artificially created by a computer. In another embodiment, one depth map is associated with multiple pairs of right and left images; in this case, the value chosen for each element of the depth map is the minimum depth value of the pixel in the different frames. Preferably, in this embodiment the depth

map is entered once for each frame group it is associated with, so as to reduce the load on the device 100, which is also inputted a piece of information that allows associating one depth map with multiple pairs of right and left images.

[0034] As an alternative to the example of Fig. 1, the depth maps of the sequence 106 may be generated within the device 100. In this case, the device 100 comprises a suitable module which is inputted the images L and R of the sequences 102 and 103 and then generates corresponding depth maps.

[0035] The device 100 allows to implement a method for multiplexing two images of the two sequences 102 and 103 and the depth map of the sequence 106.

[0036] In order to implement the method for multiplexing the right and left images and the depth map, the device 100 comprises a disassembler module 104 for breaking up an input image (the right image in the example of Fig. 1) into a plurality of sub-images, each corresponding to one region of the received image, an under-sampling and filtering module 107 for processing the depth map, and an assembler module 105 capable of entering the pixels of received images, including the depth map, into a single composite image to be provided at its output. If no processing of the sequence 106 is necessary, the module 107 may be omitted. This may be the case, for example, when the depth map is laser-generated and has, right from the start, a lower resolution than that of the images L and R.

[0037] One example of a multiplexing method implemented by the device 100 will now be described with reference to Fig. 2.

[0038] The method starts in step 200. Subsequently (step 201), one of the two input images (right or left) is broken up into a plurality of regions, as shown in Fig. 3. In the example of Fig. 3, the disassembled image is a frame R of a video stream 720p, i.e. a progressive format with a resolution of 1280 x 720 pixels, 25/30 fps (frames per second).

[0039] The frame R of Fig. 3 comes from the video stream 103 which carries the images intended for the right eye, and is disassembled into three regions R1, R2 and R3.

[0040] The disassembly of the image R is obtained by dividing it into two portions of the same size and subsequently subdividing one of these portions into two portions of the same size.

[0041] The region R1 has a size of 640x720 pixels and is obtained by taking all the first 640 pixels of each row. The region R2 has a size of 640x360 pixels and is obtained by taking the pixels from 641 to 1280 of the first 360 rows. The region R3 has a size of 640x360 pixels and is obtained by taking the remaining pixels of the image R, i.e. the pixels from 641 to 1280 of the last 360 rows.

[0042] In the example of Fig. 1, the step of disassembling the image R is carried out by the module 104, which receives an input image R (in this case the frame R) and outputs three sub-images (i.e. three groups of pixels) cor-

responding to the three regions R1, R2 and R3.

[0043] Subsequently (steps 202, 203 and 205) the composite image C is constructed, which comprises the information pertaining to both the right and the left images and to the depth map received; in the example described herein, said composite image C is a frame of the output stereoscopic video stream, and therefore it is also referred to as container frame.

[0044] First of all (step 202), the input image received by the device 100 and not disassembled by the device 105 (the left image L in the example of Fig. 1) is entered unchanged into a container frame which is sized in a manner such as to include all the pixels of both input images. For example, if the input images have a size of 1280x720 pixels, then a container frame suitable for containing both will be a frame of 1920x1080 pixels, e.g. a frame of a video stream of the 1080p type (progressive format with 1920 x 1080 pixels, 25/30 frames per second).

[0045] In the example of Fig. 4, the left image L is entered into the container frame C and positioned in the upper left corner. This is obtained by copying the 1280x720 pixels of the image L into an area C1 consisting of the first 1280 pixels of the first 720 rows of the container frame C.

[0046] When in the following description reference is made to entering an image into a frame, or transferring or copying pixels from one frame to another, it is understood that this means to execute a procedure which generates (by using hardware and/or software means) a new frame comprising the same pixels as the source image.

[0047] The (software and/or hardware) techniques for reproducing a source image (or a group of pixels of a source image) into a target image are considered to be unimportant for the purposes of the present invention and will not be discussed herein any further, in that they are per se known to those skilled in the art.

[0048] In the next step 203, the image disassembled in step 201 by the module 104 is entered into the container frame. This is achieved by the module 105 by copying the pixels of the disassembled image into the container frame C in the areas thereof which were not occupied by the image L, i.e. areas being external to the area C1.

[0049] In order to attain the best possible compression and reduce the generation of artifacts when decompressing the video stream, the pixels of the subimages outputted by the module 104 are copied by preserving the respective spatial relations. In other words, the regions R1, R2 and R3 are copied into respective areas of the frame C without undergoing any deformation, exclusively by means of translation and/or rotation operations.

[0050] An example of the container frame C outputted by the module 105 is shown in Fig. 5. The region R1 is copied into the last 640 pixels of the first 720 rows (area C2), i.e. next to the previously copied image L.

[0051] The regions R2 and R3 are copied under the area C1, i.e. respectively in the areas C3 and C4, which

respectively comprise the first 640 pixels and the following 640 pixels of the last 360 rows.

[0052] As an alternative to the solution shown in Fig. 5, the regions R2 and R3 may be copied into the container frame C in disjointed areas (i.e. neither overlapping nor neighbouring) separated by a group of pixels, so as to reduce the boundary regions.

[0053] The operations for entering the images L and R into the container frame do not imply any alterations to the balance between horizontal and vertical resolution.

[0054] In the free pixels of the frame C, i.e. in the area C5, the module 105 enters, in the form of an image, the depth map (DM) pertaining to the stereoscopic pair L and R (step 205). Prior to step 205, the depth map DM may be undersampled, filtered or further processed by the module 107.

[0055] The depth map is preferably coded as a gray scale image, the information content of which can therefore be transported by the luminance signal alone, since chrominances are null; this allows to obtain an effective compression of the container frame C.

[0056] As shown in the example of Fig. 5, the depth map entered into the frame C is preferably an image superimposition depth map, and therefore its resolution is lower than that of the pair L and R, since it is transmitted with the only purpose of positioning the overlays in depth, not for generating the stereoscopic video stream. The chosen resolution of the depth map is a result of a compromise between the bit rate required for the transfer, which must be as low as possible, and the quality of the information necessary for properly positioning the overlays.

[0057] In a preferred embodiment, the superimposition depth map DM has a resolution of 640 x 360 pixels, corresponding to a 4-to-1 undersampling (or decimation) of the original depth map having a resolution of 1280 x 720 pixels, matching that of the images L and R. Each pixel of the undersampled map DM corresponds to a 2 x 2 pixel region of the original map. In particular, the 4-to-1 undersampling step can be executed by selecting one row out of two and one column out of two of the original map.

[0058] In another embodiment, after decimation the superimposition depth map DM undergoes a processing step wherein it is divided into 16x16-pixel macroblocks, and the pixels belonging to the same macroblock are assigned a single depth value. Preferably, this value equals the minimum depth within the macroblock, since this is the most significant value for properly positioning the overlays.

[0059] Alternatively, this value is equal to the mean depth value within the macroblock.

[0060] The choice of 16x16-pixel macroblocks is particularly advantageous when the compression standard in use is H.264, because such macroblocks coincide with those employed in the H.264 standard. With this solution, in fact, compression generates less artifacts and requires a lower bit rate.

[0061] The subdivision into blocks of 8 x 8 or 4 x 4 can also be considered to be advantageous in that, due to the particular characteristics of the H.264 compression algorithm, compression benefits are obtained if the pixels within these blocks are all equal. Alternatively, but giving up the subdivision into blocks or macroblocks within which the pixels are all equal, the 640x360 depth map may be filtered with a two-dimensional low-pass filter. Compression advantages are obtained in this case as well, because the highest spatial frequencies are eliminated or reduced.

[0062] Alternatively, the depth map may have a resolution of 160 x 90 pixels, resulting from a 64-to-1 under-sampling, wherein each pixel of the depth map DM corresponds to an 8x8 region of the original map.

[0063] In a further embodiment, the superimposition depth map DM entered into the container frame C may have an uneven resolution; in particular, the lower half or third of the superimposition depth map has a higher resolution than the upper part. This solution turns out to be particularly advantageous as concerns the positioning of subtitles or other information such as the audio volume, which are generally placed in the lower part of the image. The receiver can thus use more accurate information about the depth of the pixels in a region of interest, e.g. the lower third of the 3D image, and can therefore position the images (text or graphics) correctly in that region. At the very least, the superimposition depth map may even only contain information about the depth of the pixels (all or only a portion thereof) located in a region of interest, in particular in the lower half or in the lower third of the three-dimensional content.

[0064] In another embodiment, a region of the container frame which is not occupied by the right or left images, by portions thereof or by the superimposition depth map is intended for receiving a flag which is necessary for reconstructing the right and left images at demultiplexer level. For example, said flag may relate to how the composite image has been created. Preferably, the flag may contain information useful for properly using the depth map.

[0065] The pixels of this flag region are, for example, colored in two colors (e.g. black and white) so as to create a bar code of any kind, e.g. linear or two-dimensional, which carries the flag information.

[0066] Once the transfer of both images and of the superimposition depth map received (and possibly also of the flag) into the container frame has been completed, the method implemented by the device 100 ends, and the container frame can be compressed and transmitted on a communication channel and/or recorded onto a suitable medium (e.g. CD, DVD, Blu-ray, mass memory, etc.).

[0067] Since the multiplexing operations explained above do not alter the spatial relations among the pixels of one region or image, the video stream outputted by the device 100 can be compressed to a considerable extent while preserving good possibilities that the image

will be reconstructed very faithfully to the transmitted one without creating significant artifacts.

[0068] Before describing any further embodiments, it must be pointed out that, in the preferred embodiment, the division of the frame R into three regions R1, R2 and R3 corresponds to the division of the frame into the smallest possible number of regions, taking into account the space available in the composite image and the space occupied by the left image entered unchanged into the container frame.

[0069] Said smallest number is, in other words, the minimum number of regions necessary to occupy the space left available in the container frame C by the left image.

[0070] In general, therefore, the minimum number of regions into which the image must be disassembled is defined as a function of the format of the source images (right and left images) and of the target composite image (container frame C).

[0071] Preferably, the image to be entered into the frame is disassembled by taking into account the need for breaking up the image (e.g. R in the above example) into the smallest number of rectangular regions.

[0072] In a further embodiment, the right image R is disassembled as shown in Fig. 6.

[0073] The region R1' corresponds to the region R1 of Fig. 3, and therefore comprises the first 640 pixels of all 720 rows of the image.

[0074] The region R2' comprises the 320 columns of pixels adjacent to the region R1', whereas the region R3' comprises the last 320 columns of pixels.

[0075] The container frame C can thus be constructed as shown in Fig. 7, with the regions R2' and R3' turned by 90° and arranged in the areas C3' and C4' under the image L and the region R1'.

[0076] The regions R2' and R3' thus rotated occupy 720 pixels of 320 rows; therefore, the areas C3' and C4' are separated from the areas C1 and C2 that contain the pixels copied from the image L and from the region R1'.

[0077] Preferably, the areas C3' and C4' are separated from the other areas C1 and C2 by at least one safeguard line. In particular, it is advantageous and preferable to copy the pixels of the regions R2' and R3' into the last rows of the container frame C.

[0078] Since in this case the container frame is made up of 1080 rows, in the embodiment of Fig. 7 the rotated regions R2' and R3' are separated from the above image L and region R1' by a safeguard strip 40 pixels high.

[0079] In the example of Fig. 7, the regions R2' and R3' are separated from each other, so that they are surrounded by pixels of a predefined color (e.g. white or black) not coming from the right and left images. In this manner, the boundary areas between regions containing pixels coming from the right and left images are reduced, while also reducing any artifacts caused by image compression and maximizing the compression rate.

[0080] As an alternative to positioning R2' and R3' into the last rows of the container frame C (as described with

reference to Fig. 7), in a preferred embodiment R2' and R3' are positioned in a manner such that a safeguard strip 32 pixel rows high is left between the bottom edge of L and the upper edge of R2' and R3'. This provides a second safeguard strip 8 pixel rows high between the bottom edge of R2' and R3' and the bottom edge of C. By further exploiting the width of the container frame, it is possible to position R2' and R3' in a manner such that they get completely surrounded by pixels coming from neither the right image nor the left image.

[0081] Finally, in the area C5' in the bottom right corner of the frame C, the superimposition depth map (DM') is entered with a resolution of 160 x 90 pixels, obtained by undersampling the original depth map as previously described. In general, the superimposition depth map may have any resolution, as long as it is contained within a free space of the frame C. For better exploiting the available space, the superimposition depth map may undergo a rotation and/or disassembly step prior to being entered into the frame C.

[0082] In a further embodiment, which is described herein with reference to Figs. 8 and 9, the module 104 extracts three subimages R1", R2" and R3", whose total sum of pixels exceeds that of the disassembled image.

[0083] The region R1" corresponds to the region R1' of Fig. 6, whereas R2" and R3" include the area of the regions R2' and R3' plus an additional area (Ra2 and Ra3) which allows to minimize the creation of artifacts during the image compression step.

[0084] The segment R1" is thus a region having a size of 640 x 720 pixels and occupying the first columns of the frame R to be disassembled.

[0085] The segment R3" occupies the last columns of the frame R to be disassembled, and borders on the central region R2". R3" includes, on the left side (the one bordering on R2"), a buffer strip Ra3 containing pixels in common with the region R2". In other words, the last columns of R2" and the first ones of R3" (which constitute the buffer strip Ra3) coincide.

[0086] Preferably, the size of the buffer strip Ra3 is chosen as a function of the type of compression to be subsequently applied to the container frame C, and in general to the video stream containing it. In particular, said strip has a size which is twice that of the elementary processing unit used in the compression process. For example, the H.264 standard provides for disassembling the image into macroblocks of 16x16 pixels, each of which represents this standard's elementary processing unit. Based on this assumption, the strip Ra3 has a width of 32 pixels. The segment R3" therefore has a size of 352 (320+32)x720 pixels, and comprises the pixels of the last 352 columns of the image R.

[0087] The segment R2" occupies the central part of the image R to be disassembled and includes, on its left side, a buffer strip Ra2 having the same size as the strip Ra3. In the example taking into account the H.264 compression standard, the strip Ra2 is thus 32 pixels wide and comprises pixels in common with the region R1".

The segment R2" therefore has a size of 352x720 pixels and comprises the pixels of the columns from 608 (640 of R1"-32) to 978 of the frame R.

[0088] The three subimages pertaining to the regions R1", R2" and R3" outputted by the module 104 (visible in Fig. 8) are then entered into the container frame C as shown in Fig. 9. The regions R2" and R3" are turned by 90° and the pixels are copied into the last rows of the frame C (areas designated C3" and C4") by providing a certain number of safeguard pixels which separate the areas C3" and C4" from the areas C1 and C2 that include the pixels of the images L and R1". In the case shown in Fig. 9, this safeguard strip is 8 pixels wide.

[0089] In this embodiment as well, the superimposition depth map (DM') is entered into the area C5' in the bottom right corner of the frame C.

[0090] The frame C thus obtained is subsequently compressed and transmitted or saved to a storage medium (e.g. a DVD). For this purpose, compression means are provided which are adapted to compress an image or a video signal, along with means for recording and/or transmitting the compressed image or video signal.

[0091] Fig. 10 shows a block diagram of a receiver 1100 which decompresses the container frame (if compressed) received or read from a medium, reconstructs the two right and left images, and makes them available, along with the related superimposition depth map, to a display device (e.g. a television set) allowing fruition of 3D contents with images superimposed on a video content. The receiver 1100 may be a set-top-box or a receiver built in a television set. It should be noted that, when the receiver 1100 is a set-top-box not integrated into a television set, it must use the depth map internally to properly position the graphics it generates (e.g. subtitles, EPG and related menus). In addition, the receiver 1100 will have to send the depth map (e.g. via the HDMI interface) to the television set, in that the latter will need it to properly position its own graphics (e.g. its menus).

[0092] The same remarks made for the receiver 1100 are also applicable to a reader (e.g. a DVD reader) which reads a container frame (possibly compressed) and processes it in order to obtain one pair of frames corresponding to the right and left images entered into the container frame (possibly compressed) read by the reader.

[0093] Referring back to Fig. 10, the receiver receives (via cable or antenna) a compressed stereoscopic video stream 1101 and decompresses it by means of a decompression module 1102, thereby obtaining a video stream comprising a sequence of frames C' corresponding to the frames C. If there is an ideal channel or if container frames are being read from a mass memory or a data medium (Blu-ray, CD, DVD), the frames C' correspond to the container frames C carrying the information about the right and left images and the superimposition depth map, except for any artifacts introduced by the compression process.

[0094] These frames C' are then supplied to a reconstruction module 1103, which executes an image recon-

struction and depth map extraction method as described below with reference to Figs. 11 and 12.

[0095] It is apparent that, if the video stream is not compressed, the decompression module 1102 may be omitted and the video signal may be supplied directly to the reconstruction module 1103.

[0096] The reconstruction process starts in step 1300, when the decompressed container frame C' is received. The reconstruction module 1103 extracts (step 1301) the left image L by copying the first 720x1280 pixels of the decompressed frame into a new frame which is smaller than the container frame, e.g. a frame of a 720p stream. The image L thus reconstructed is outputted to the receiver 1100 (step 1302).

[0097] Subsequently, the method provides for extracting the right image R from the container frame C'.

[0098] The step of extracting the right image begins by copying (step 1303) a portion of the area R1" included in the frame C'. More in detail, the pixels of the first 624(640-16) columns of R1" are copied into the corresponding first 624 columns of the new frame representing the reconstructed image Rout, as shown in Fig. 11. As a matter of fact, this removes from the reconstruction step the 16 columns of R1" which are most subject to creation of artifacts, e.g. through the effect of the motion estimation procedure carried out by the H.264 compression standard.

[0099] Then a central portion of R2" is extracted (step 1304). From the decompressed frame C' (which, as aforesaid, corresponds to the frame C of Fig. 9), the pixels of the area C3" (corresponding to the source region R2") are selected and a 90° rotation inverse to the one executed in the multiplexer 100 is made, which brings them back to the original row/column condition, i.e. the one shown in Fig. 8. At this point, the first and last sixteen (16) columns of R2" are eliminated and the remaining 352-32=320 pixel columns are copied into the free columns adjacent to those just copied from R1".

[0100] By cutting the 16 outermost columns of the region R2", those columns are eliminated where formation of artifacts is most likely to occur. The width of the cut area (in this case 16 columns) depends on the type of compression used. Said area is preferably equal to the elementary processing unit used by the compression process; in the case described herein, the H.264 standard operates upon blocks of 16x16 pixels, and therefore 16 columns are to be cut.

[0101] As regards R3" (step 1305), the pixels of the region C4" are extracted from the frame C' and the subimage R3" is brought back to the original row/column format (see Fig. 8). Subsequently, the first 16 pixel columns are eliminated (corresponding to half the area Ra3) and the remaining 352-16=336 pixel columns are copied into the last free columns on the left of the reconstructed frame. Like R2", also in R3" the cut area is equal to the elementary processing unit used by the compression process.

[0102] Of course, for both regions R2" and R3" the rotation step may be carried out in a virtual manner, i.e. the

same result in terms of extraction of the pixels of interest may be obtained by copying into the reconstructed frame the pixels of a row of the area C3" (if R2", C4" if R3") in a column of the new frame Rout, except for the last 16 rows of the area C3" (if R2", C4" if R3") corresponding to the sixteen columns to be cut, shown in Fig. 8.

[0103] At this point, the right image Rout has been fully reconstructed and can be outputted (step 1306).

[0104] Finally, the reconstruction module 1103 extracts (step 1308) the superimposition depth map DM' by copying into a register the luminance values of the last 160 x 90 pixels of the decompressed container frame C', corresponding to the area C5'. The content of said register is outputted to the receiver 1100 (step 1309) and will be used for defining the position in depth of images (text or graphics) to be combined with the three-dimensional content transported by the stereoscopic video stream; in particular, it will be used for combining images to be superimposed on the three-dimensional content. As an alternative or in addition to outputting the content of the depth map and the images L and R extracted from the input frames, the receiver 1100 comprises a character generator and/or a graphic generator and combines other images with the images L and R, i.e. with the three-dimensional content. The images to be combined are selected from a memory area of the receiver and may be stored when manufacturing the receiver (e.g. the graphics of some menus or of the channel numbers) or may be extracted from the video stream (e.g. program guide information and subtitles).

[0105] These images are combined with the three-dimensional content in positions in depth that depend on the superimposition depth maps extracted from the video stream. In particular, for each stereoscopic image (produced by the pair of images L and R), the combined image is placed in the point of minimum depth of the stereoscopic image. After the images have been combined with the 3D content, in this embodiment the receiver 1100 outputs a pair of images L* and R* which, when reproduced, will be perceived by the user as a three-dimensional content corresponding to the original one (produced by the images L and R) with images superimposed thereon, e.g. subtitles, menus, graphics, etc.

[0106] The process for reconstructing the right and left images and the depth map contained in the container frame C' is thus completed (step 1307). Said process is repeated for each frame of the video stream received by the receiver 1100, so that the output will consist of two video streams 1104 and 1105 for the right image and for the left image, respectively, and one data signal deduced from the superimposition depth map.

[0107] The process for reconstructing the right and left images and the superimposition depth map described above with reference to Figs. 10, 11 and 12 is based upon the assumption that the demultiplexer 1100 knows how the container frame C was built and can thus extract the right and left images and the superimposition depth map.

[0108] Of course, this is possible if the multiplexing method is standardized.

[0109] In order to take into account the fact that the container frame may be generated according to any one of the above-described methods, or anyway according to any one of the methods that utilise the solution which is the subject of the appended claims, the demultiplexer uses the flag information contained in a predefined region of the composite image (e.g. a bar code, as previously described) in order to know how the contents of the composite image must be unpacked and how to reconstruct the right and left images and the superimposition depth map.

[0110] After decoding the flag, the demultiplexer will know the position of the unchanged image (e.g. the left image in the above-described examples), as well as the positions and any transformations (rotation, translation or the like) of the regions into which the other image was disassembled (e.g. the right image in the above-described examples) and the position of the superimposition depth map.

[0111] With this information, the demultiplexer can thus extract the unchanged image (e.g. the left image) and the depth map and reconstruct the disassembled image (e.g. the right image).

[0112] Although the present invention has been illustrated so far with reference to some preferred and advantageous embodiments, it is clear that it is not limited to said embodiments and that many changes may be made thereto by a man skilled in the art wanting to combine into a composite image two images relating to two different perspectives (right and left) of an object or a scene.

[0113] For example, the electronic modules that provide the above-described devices, in particular the device 100 and the receiver 1100, may be variously subdivided and distributed; furthermore, they may be provided in the form of hardware modules or as software algorithms implemented by a processor, in particular a video processor equipped with suitable memory areas for temporarily storing the input frames received. These modules may therefore execute in parallel or in series one or more of the video processing steps of the image multiplexing and demultiplexing methods according to the present invention.

[0114] It is also apparent that, although the preferred embodiments refer to multiplexing two 720p video streams into one 1080p video stream, other formats may be used as well, such as, for example, two 640x480 video streams into one 1280x720 video stream, or two 320x200 video streams into one 640x480 video stream.

[0115] Nor is the invention limited to a particular type of arrangement of the composite image, since different solutions for generating the composite image may offer specific advantages.

[0116] For example, the embodiments described above with reference to Figs. 1 to 12 offer the advantage that they only carry out translation or roto-translation op-

erations, thus only requiring little computational power.

[0117] Alternatively, it is conceivable that the images are also subjected to specular inversion steps, in addition to said rotation and/or translation operations, in order to obtain a composite image of the type shown in Fig. 13.

[0118] These additional operations are carried out for the purpose of maximizing the boundary perimeters between regions containing homologous pixels, thereby exploiting the strong correlation existing among them and minimizing the artifacts introduced by the subsequent compression step. In the example of Figs. 13 and 14 it has been assumed for clarity that the two right and left images are identical, even though they generally differ slightly.

[0119] In this figure, the left image L (shown in Fig. 14a) is positioned in the upper right corner of the container frame C, so as to occupy the last 1280 pixels of the first 720 rows. As in the examples previously described, the image L is thus copied unchanged into the container frame C.

[0120] Instead, the right image R is disassembled according to the example of Fig. 3; Fig. 14b shows the image R broken up into three regions R1, R2 and R3.

[0121] Subsequently, some regions (the regions R1 and R3 in the example of Fig. 14) undergo a specular inversion operation; the inversion may occur relative to a vertical axis (i.e. parallel to a column of the image) or to a horizontal axis (i.e. parallel to a row of the image).

[0122] In the case of inversion relative to a vertical axis, the pixels of the column N (where N is an integer between 1 and 1080, 1080 being the number of columns of the image) are copied into the column $1080+1-N$.

[0123] In the case of inversion relative to a horizontal axis, the pixels of the row M (where M is an integer between 1 and 720, 720 being the number of rows of the image) are copied into the row $720+1-M$.

[0124] Figs. 14c and 14d show the region R1 extracted from the image R and inverted ($R1_{rot}$) relative to a vertical axis, in particular relative to a vertical side.

[0125] The inverted region $R1_{inv}$ is entered into the first 640 pixels of the first 640 pixel rows.

[0126] As can be seen in the example of Fig. 13, when $R1_{inv}$ is entered rotated into the container frame C, the pixels of $R1_{inv}$ bordering on L are very similar to the pixels of L bordering on $R1_{inv}$. The spatial correlation among these pixels has the advantage of reducing the formation of artifacts.

[0127] Figs. 14e and 14f show the region R3 extracted from the image R of Fig. 14b and then inverted ($R3_{inv}$) relative to a horizontal axis, in particular relative to a horizontal side.

[0128] The region $R3_{inv}$ is entered into the last 640 pixels of the last 360 rows. This reduces the generation of artifacts, since the pixels of the boundary regions between $R3_{inv}$ and L are pixels having high spatial correlation. The pixels in this boundary region, in fact, reproduce similar or identical portions of the image.

[0129] The container frame C is then completed by en-

tering the region R2.

[0130] In this example R2 is not inverted and/or rotated because it would not be possible, in neither case, to match a boundary region of R2 with a boundary region made up of homologous pixels of another region of R or L.

[0131] Finally, it is also apparent that the invention also relates to any demultiplexing method which allows a right image and a left image to be extracted from a composite image by reversing one of the above-described multiplexing processes falling within the protection scope of the present invention.

[0132] The invention therefore also relates to a method for generating a pair of images starting from a composite image, which comprises the steps of:

- generating a first one (e.g. the left image) of said right and left images by copying one single group of contiguous pixels from a region of said composite image,
- generating a second image (e.g. the right image) by copying other groups of contiguous pixels from different regions of said composite image.

[0133] According to one embodiment, the information for generating said second image is extracted from an area of said composite image. Said information is preferably encoded according to a bar code.

[0134] In one embodiment of the method for generating the right and left images, the generation of the image which was disassembled in the composite image comprises at least one step of specular inversion of a group of pixels of one of said different regions. In one embodiment of the method for generating the right and left images, the generation of the image which was disassembled in the composite image comprises at least one step of removing pixels from one of the regions of the composite image that comprise the pixels of this image to be reconstructed. In particular, the pixels are removed from a boundary area of this region.

[0135] In one embodiment, the image which was disassembled into different regions of the composite image is reconstructed by subjecting the pixel regions that include the pixels of the image to be disassembled to translation and/or rotation operations only.

[0136] Although the above-described embodiment example refers to entering a superimposition depth map into a container frame in which either one of the two right and left images is disassembled into several parts, it is clear that the invention is not dependent on the manner in which the two right and left images are formatted within the container frame. For example, the two images may be undersampled and arranged side by side (side-by-side format) or one on top of the other (top-bottom format) in order to leave a free space in the frame wherein the superimposition depth map can be placed. Also, either one of the right and left images may be left unchanged, whereas the other one may be undersampled in order to free up space for the depth map.

[0137] Finally, it must be remarked that the embodiment examples described above with reference to the annexed drawings relate to a "whole" depth map, i.e. a depth map computed by decimating or filtering a depth map of the 3D content without however subdividing it into several parts, unlike one of the two images L and R, for example. Nevertheless, this is not a limitation of the present invention, and the superimposition depth map, once generated (or received), may be entered into the container frame by an encoder, which will break it up into multiple parts that will be arranged in different regions of the container frame. For example, as known, in order to code a stereoscopic content, an H.264 encoder has to enter eight additional rows which will be cut by the decoder; in one embodiment, the superimposition depth map can be entered into these eight additional rows by dividing it, for example, into 240 blocks of 8x8 in size, which when appropriately reassembled will form an image having dimensions proportional to the transported stereoscopic content. One example of block arrangement may be obtained by scanning the rows of a depth map decimated by 16, therefore with a 120x72 resolution, wherein strips of 120x8 pixels are lined up in order to obtain an 1080x8-pixel image. In another embodiment, the same decimated depth map may be subdivided into a greater number of strips 8 pixels high by using a 6-pixel offset instead of an 8-pixel one, so that the content becomes redundant and content protection is promoted at the boundary with the main image. This appears to be particularly advantageous whenever the stereoscopic content includes a pair of right and left images multiplexed into a top-bottom, side-by-side or checkerboard format, with such a resolution as to occupy all the potentially displayable pixels in the frame, e.g. the pixels of a 1920x1080 format.

[0138] Preferably, in the event that the frame includes a pair of asymmetrically decimated images (e.g. a side-by-side format wherein the columns are decimated more than the rows, or a top-bottom format wherein only the rows are decimated, not the columns), then the superimposition depth map is obtained by decimating a depth map with a row/column decimation ratio proportional to the one used for sampling the images placed in the same frame. By way of example, assuming that a side-by-side format is used for multiplexing the right and left images in the frame, the row/column decimation ratio will be 1:2, since all rows are kept and the columns are decimated by two. In this case, the superimposition depth map can be obtained by decimating a depth map with a 1:2 row/column decimation ratio.

[0139] It is also clear that different methods may be used for signaling the area occupied by the depth map to the receiver other than those described above, which provide for entering a flag into the image; in fact, such a flag may also be included in a data packet of the signal carrying the video stream.

Claims

1. A method for superimposing images on a three-dimensional content, wherein a video stream is received which comprises said three-dimensional content and a depth map (DM, DM') for superimposing images on said three-dimensional content, said depth map (DM, DM') containing information about the depth of said three-dimensional content, and being inserted as an image in a frame (C) of said video stream,
the method being **characterized by** using said depth map (DM, DM') in the playback phase only for the superimposition of images to said three-dimensional content, said superimposition being made in a position in depth depending on said superimposition depth map (DM, DM'),
and in that said depth map (DM, DM') has a smaller number of pixels than that of a two-dimensional image included in said frame (C) and associated with said three-dimensional content.
2. A method according to claim 1, wherein said superimposition depth map only contains information about the depth of pixels located in the lower half, preferably in the lower third, of said three-dimensional content.
3. A method according to claim 1 or 2, wherein the superimposition depth map has a non-uniform resolution, in particular the lower half or third of said depth map has a higher resolution than the upper part.
4. A method according to claim 1 or 2 or 3, wherein said superimposition depth map has a lower resolution than a two-dimensional image associated with said three-dimensional content.
5. A method according to claim 4, wherein said three-dimensional content is an image consisting of a plurality of pixels, and wherein said depth map is obtained by undersampling a depth map whose elements correspond to the depth of the pixels of said three-dimensional content.
6. A method according to claim 5, wherein, after undersampling said depth map, the undersampled map is divided into blocks and each pixel of the block is given a same value equal to the minimum depth of the pixels of said block or to the mean value of the depth of the pixels of the block.
7. A method according to claim 5, wherein, prior to undersampling said depth map, the depth map is divided into blocks and each pixel of the block is given a same value equal to the minimum depth of the pixels of said block or to the mean value of the depth of the pixels of the block.
8. A method according to claim 6 or 7, wherein said blocks have a size equal to a multiple of an elementary block of 2x2 pixels.
9. A method according to any one of the preceding claims, wherein said superimposition depth map is entered into a portion of said frame not intended for display.
10. A method according to any one of the preceding claims, wherein said depth map is broken up into blocks distributed in areas of said frame (C) which are not occupied by said three-dimensional content.
11. A method according to any one of the preceding claims, wherein said frame comprises a right image, a left image and said depth map, wherein said depth map is broken up into blocks distributed in regions of the frame (C) which are not occupied by said three-dimensional content, and wherein said frame (C) is coded according to the H.264 coding standard.
12. A method according to any one of claims 1 to 9, wherein said three-dimensional content comprises a two-dimensional image and information which allows to rebuild the other image of a stereoscopic pair, and wherein said superimposition depth map is entered into a portion of the two-dimensional image.
13. A method according to any one of claims 1 to 12, wherein said frame comprises a flag adapted to indicate to the receiver the position of said superimposition depth map within said frame.
14. A method according to any one of claims 1 to 13, wherein said video stream comprises a flag adapted to indicate to the receiver the position of said superimposition depth map within said frame, said flag being external to said frame.
15. A device for reproducing three-dimensional content, comprising means adapted to receive a video stream containing three-dimensional content,
means adapted to combine an image with said three-dimensional content,
characterized in that said means adapted to combine an image with said three-dimensional content are adapted to implement a method according to any one of claims 1 to 14, and comprise means for extracting said superimposition depth map and for combining said image to be superimposed on the three-dimensional content, based on a depth depending on said superimposition depth map.
16. A stereoscopic video stream (1101) comprising a plurality of frames and **characterized in that** said frames comprise at least one three-dimensional con-

tent and at least one superimposition depth map (DM, DM') inserted as an image within each frame of said plurality of frames, said superimposition depth map (DM, DM') being a depth map for superimposing images on said three-dimensional content, said superimposition being made in a position in depth depending on said superimposition depth map, and comprising a smaller number of pixels than that of a two-dimensional image associated with said three-dimensional content, said stereoscopic video stream being adapted to be used in a method according to any one of claims 1 to 14.

Patentansprüche

1. Verfahren zum Überlagern von Bildern auf einem dreidimensionalen Inhalt, wobei ein Videostrom empfangen wird, der den dreidimensionalen Inhalt und eine Tiefenkarte (DM, DM') zum Überlagern von Bildern auf dem dreidimensionalen Inhalt umfasst, wobei die Tiefenkarte (DM, DM') eine Information über die Tiefe des dreidimensionalen Inhalts beinhaltet und als ein Bild in einem Rahmen (C) des Videostroms eingefügt wird, wobei das Verfahren **gekennzeichnet ist durch** einen Nutzen der Tiefenkarte (DM, DM') in der Wiedergabephase nur für die Überlagerung von Bildern auf dem dreidimensionalen Inhalt, wobei die Überlagerung in einer Position in der Tiefe erfolgt, die von der Überlagerungstiefenkarte (DM, DM') abhängt, und wobei die Tiefenkarte (DM, DM') eine kleinere Anzahl von Pixeln aufweist als diejenige eines zweidimensionalen Bildes, das in dem Rahmen (C) enthalten ist und mit dem dreidimensionalen Inhalt assoziiert ist.
2. Verfahren nach Anspruch 1, wobei die Überlagerungstiefenkarte nur eine Information über die Tiefe von Pixeln beinhaltet, die in der unteren Hälfte, bevorzugt in dem unteren Drittel, des dreidimensionalen Inhalts gelegen sind.
3. Verfahren nach Anspruch 1 oder 2, wobei die Überlagerungstiefenkarte eine ungleichförmige Auflösung aufweist, wobei insbesondere die untere Hälfte oder das untere Drittel der Tiefenkarte eine höhere Auflösung aufweist als der obere Teil.
4. Verfahren nach Anspruch 1 oder 2 oder 3, wobei die Überlagerungstiefenkarte eine geringere Auflösung aufweist als ein zweidimensionales Bild, das mit dem dreidimensionalen Inhalt assoziiert ist.
5. Verfahren nach Anspruch 4, wobei der dreidimensionale Inhalt ein Bild ist, das aus einer Mehrzahl von Pixeln besteht, und wobei die Tiefenkarte durch ein Unterabtasten einer Tiefenkarte erhalten wird, deren

Elemente zu der Tiefe der Pixel des dreidimensionalen Inhalts korrespondieren.

6. Verfahren nach Anspruch 5, wobei, nach dem Unterabtasten der Tiefenkarte, die unterabgetastete Tiefenkarte in Blöcke aufgeteilt wird und jedem Pixel des Blocks ein selber Wert gegeben wird, der gleich der minimalen Tiefe der Pixel des Blocks oder dem mittleren Wert der Tiefe der Pixel des Blocks ist.
7. Verfahren nach Anspruch 5, wobei, vor dem Unterabtasten der Tiefenkarte, die Tiefenkarte in Blöcke unterteilt wird und jedem Pixel des Blocks ein selber Wert gegeben wird, der gleich der minimalen Tiefe der Pixel des Blocks oder dem mittleren Wert der Tiefe der Pixel des Blocks ist.
8. Verfahren nach Anspruch 6 oder 7, wobei die Blöcke eine Größe aufweisen, die gleich einem Mehrfachen eines elementaren Blocks von 2x2 Pixeln ist.
9. Verfahren nach einem der vorangegangenen Ansprüche, wobei die Überlagerungstiefenkarte in einen Teil des Rahmens eingefügt wird, der nicht zur Anzeige vorgesehen ist.
10. Verfahren nach einem der vorangegangenen Ansprüche, wobei die Tiefenkarte in Blöcke aufgeteilt wird, die über Bereiche des Rahmens (C) verteilt sind, welche nicht durch den dreidimensionalen Inhalt belegt sind.
11. Verfahren nach einem der vorangegangenen Ansprüche, wobei der Rahmen ein rechtes Bild, ein linkes Bild und eine Tiefenkarte umfasst, wobei die Tiefenkarte in Blöcke aufgeteilt wird, die in Regionen des Rahmens (C) verteilt sind, welche nicht durch den dreidimensionalen Inhalt belegt sind, und wobei der Rahmen (C) gemäß dem H.264-Kodierstandard kodiert wird.
12. Verfahren nach einem der Ansprüche 1 bis 9, wobei der dreidimensionale Inhalt ein zweidimensionales Bild und eine Information umfasst, die es erlaubt, das andere Bild eines stereoskopischen Paares wiederherzustellen, und wobei die Überlagerungstiefenkarte in einen Abschnitt des zweidimensionalen Bildes eingefügt wird.
13. Verfahren nach einem der Ansprüche 1 bis 12, wobei der Rahmen ein Flag umfasst, das angepasst ist, dem Empfänger die Position der Überlagerungstiefenkarte in dem Rahmen anzuzeigen.
14. Verfahren nach einem der Ansprüche 1 bis 13, wobei der Videostrom ein Flag umfasst, das angepasst ist, dem Empfänger die Position der Überlagerungstiefenkarte in dem Rahmen anzuzeigen, wobei das

Flag extern zu dem Rahmen ist.

15. Gerät zum Reproduzieren eines dreidimensionalen Inhalts, umfassend:

Mittel, die angepasst sind, einen Videostrom zu empfangen, der einen dreidimensionalen Inhalt beinhaltet,

Mittel, die angepasst sind, ein Bild mit dem dreidimensionalen Inhalt zu kombinieren,

dadurch gekennzeichnet, dass die Mittel, die angepasst sind, ein Bild mit dem dreidimensionalen Inhalt zu kombinieren, angepasst sind, ein Verfahren nach einem der Ansprüche 1 bis 14 zu implementieren, und Mittel zum Extrahieren der Überlagerungstiefenkarte und zum Kombinieren des auf dem dreidimensionalen Inhalt zu überlagernden Bildes basierend auf einer Tiefe, die von der Überlagerungstiefenkarte abhängt, umfasst.

16. Stereoskopischer Videostrom (1101), umfassend eine Mehrzahl von Rahmen und **dadurch gekennzeichnet, dass** die Rahmen zumindest einen dreidimensionalen Inhalt und zumindest eine Überlagerungstiefenkarte (DM, DM'), die als ein Bild in jeden Rahmen der Mehrzahl von Rahmen eingefügt ist, umfassen, wobei die Überlagerungstiefenkarte (DM, DM') eine Tiefenkarte zum Überlagern von Bildern auf dem dreidimensionalen Inhalt ist, wobei die Überlagerung in einer Position in der Tiefe erfolgt, die von der Überlagerungstiefenkarte abhängt, und eine kleinere Anzahl von Pixeln umfasst als diejenige eines zweidimensionalen Bildes, das mit dem dreidimensionalen Inhalt assoziiert ist, wobei der stereoskopische Videostrom angepasst ist, in einem Verfahren nach einem der Ansprüche 1 bis 14 genutzt zu werden.

Revendications

1. Procédé de superposition d'images sur un contenu tridimensionnel, dans lequel un flux vidéo est reçu, lequel comprend ledit contenu tridimensionnel et une carte de profondeur (DM, DM') permettant de superposer des images sur ledit contenu tridimensionnel, ladite carte de profondeur (DM, DM') contenant des informations à propos de la profondeur dudit contenu tridimensionnel, et étant insérée en tant qu'image dans une trame (C) dudit flux vidéo, le procédé étant **caractérisé par** l'utilisation de ladite carte de profondeur (DM, DM') dans la phase de lecture uniquement pour la superposition d'images audit contenu tridimensionnel, ladite superposition étant faite dans une position dans la profondeur en fonction de ladite carte de profondeur de superposition (DM, DM'),

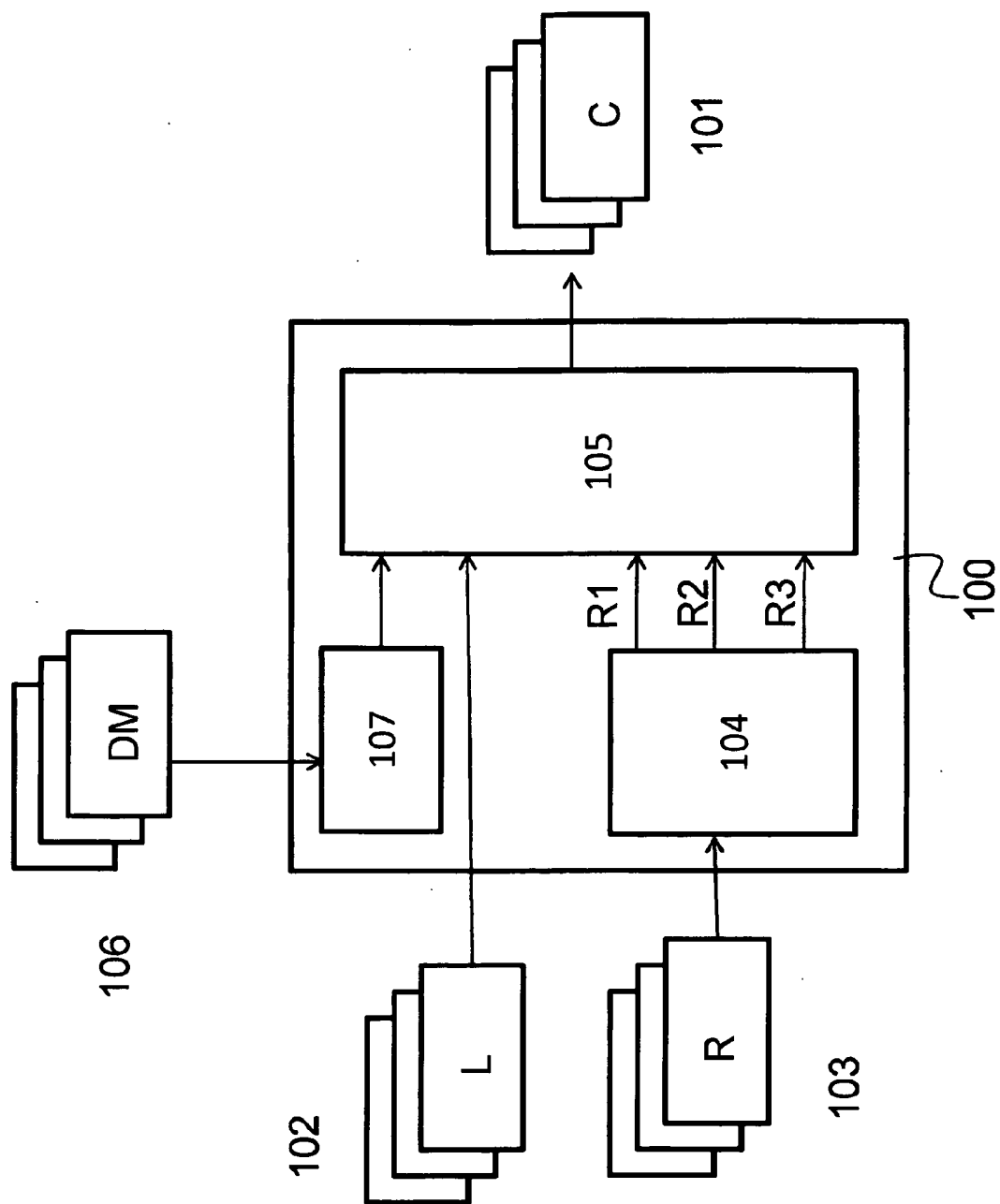
et en ce que ladite carte de profondeur (DM, DM') a un plus petit nombre de pixels que celui d'une image bidimensionnelle incluse dans ladite trame (C) et associée audit contenu tridimensionnel.

2. Procédé selon la revendication 1, dans lequel ladite carte de profondeur de superposition ne contient que des informations à propos de la profondeur des pixels situés dans la moitié inférieure, de préférence dans le tiers inférieur, dudit contenu tridimensionnel.
3. Procédé selon la revendication 1 ou 2, dans lequel ladite carte de profondeur de superposition a une résolution non uniforme, en particulier la moitié ou le tiers inférieur de ladite carte de profondeur a une résolution plus élevée que la partie supérieure.
4. Procédé selon la revendication 1, 2 ou 3, dans lequel ladite carte de profondeur de superposition a une résolution inférieure à une image bidimensionnelle associée audit contenu tridimensionnel.
5. Procédé selon la revendication 4, dans lequel ledit contenu tridimensionnel est une image constituée d'une pluralité de pixels, et dans lequel ladite carte de profondeur est obtenue en sous-échantillonnant une carte de profondeur dont les éléments correspondent à la profondeur des pixels dudit contenu tridimensionnel.
6. Procédé selon la revendication 5, dans lequel, après le sous-échantillonnage de ladite carte de profondeur, la carte sous-échantillonnée est divisée en blocs et chaque pixel du bloc se voit donner une même valeur égale à la profondeur minimale des pixels dudit bloc ou à la valeur moyenne de la profondeur des pixels du bloc.
7. Procédé selon la revendication 5, dans lequel, avant le sous-échantillonnage de ladite carte de profondeur, la carte de profondeur est divisée en blocs et chaque pixel du bloc se voit donner une même valeur égale à la profondeur minimale des pixels dudit bloc ou à la valeur moyenne de la profondeur des pixels du bloc.
8. Procédé selon la revendication 6 ou 7, dans lequel lesdits blocs ont une taille égale à un multiple d'un bloc élémentaire de 2 x 2 pixels.
9. Procédé selon l'une quelconque des revendications précédentes, dans lequel ladite carte de profondeur de superposition est entrée dans une portion de ladite trame non destinée à l'affichage.
10. Procédé selon l'une quelconque des revendications précédentes, dans lequel ladite carte de profondeur est scindée en blocs répartis dans des zones de la-

dite trame (C) qui ne sont pas occupées par ledit contenu tridimensionnel.

11. Procédé selon l'une quelconque des revendications précédentes, dans lequel ladite trame comprend une image droite, une image gauche et ladite carte de profondeur, dans lequel ladite carte de profondeur est scindée en blocs répartis dans des régions de la trame (C) qui ne sont pas occupées par ledit contenu tridimensionnel, et dans lequel ladite trame (C) est codée selon la norme de codage H.264. 5 10
12. Procédé selon l'une quelconque des revendications 1 à 9, dans lequel ledit contenu tridimensionnel comprend une image bidimensionnelle et des informations qui permettent de reconstruire l'autre image d'une paire stéréoscopique, et dans lequel ladite carte de profondeur de superposition est entrée dans une portion de l'image bidimensionnelle. 15 20
13. Procédé selon l'une quelconque des revendications 1 à 12, dans lequel ladite trame comprend un drapeau adapté pour indiquer au récepteur la position de ladite carte de profondeur de superposition au sein de ladite trame. 25
14. Procédé selon l'une quelconque des revendications 1 à 13, dans lequel ledit flux vidéo comprend un drapeau adapté pour indiquer au récepteur la position de ladite carte de profondeur de superposition au sein de ladite trame, ledit drapeau étant externe à ladite trame. 30
15. Dispositif de reproduction de contenu tridimensionnel, comprenant des moyens adaptés pour recevoir un flux vidéo contenant un contenu tridimensionnel, des moyens adaptés pour combiner une image avec ledit contenu tridimensionnel, **caractérisé en ce que** lesdits moyens adaptés pour combiner une image avec ledit contenu tridimensionnel sont adaptés pour implémenter un procédé selon l'une quelconque des revendications 1 à 14, et comprennent des moyens permettant d'extraire ladite carte de profondeur de superposition et de combiner ladite image à superposer sur le contenu tridimensionnel en se basant sur une profondeur en fonction de ladite carte de profondeur de superposition. 35 40 45
16. Flux vidéo stéréoscopique (1101) comprenant une pluralité de trames et **caractérisé en ce que** lesdites trames comprennent au moins un contenu tridimensionnel et au moins une carte de profondeur de superposition (DM, DM') insérée en tant qu'image au sein de chaque trame de ladite pluralité de trames, ladite carte de profondeur de superposition (DM, DM') étant une carte de profondeur permettant de superposer des images sur ledit contenu tridimensionnel, ladite superposition étant faite dans une po- 50 55

sition dans la profondeur en fonction de ladite carte de profondeur de superposition, et comprenant un plus petit nombre de pixels que celui d'une image bidimensionnelle associée audit contenu tridimensionnel, ledit flux vidéo stéréoscopique étant adapté pour être utilisé dans un procédé selon l'une quelconque des revendications 1 à 14.



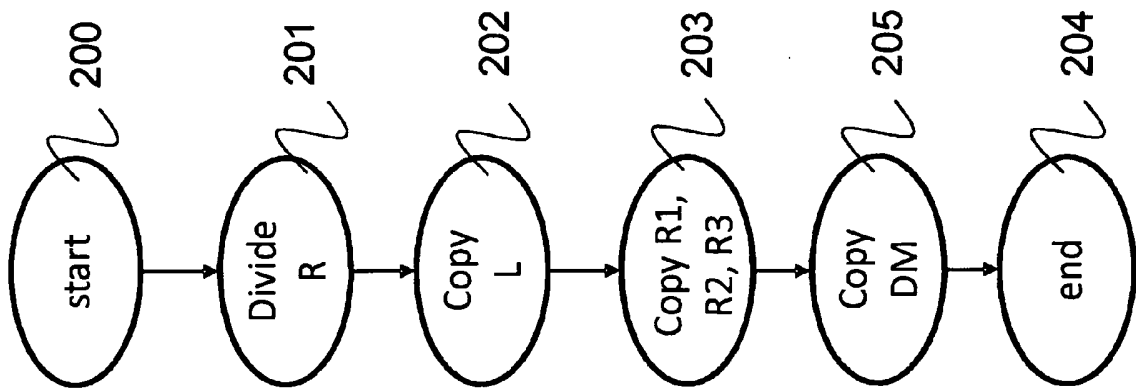
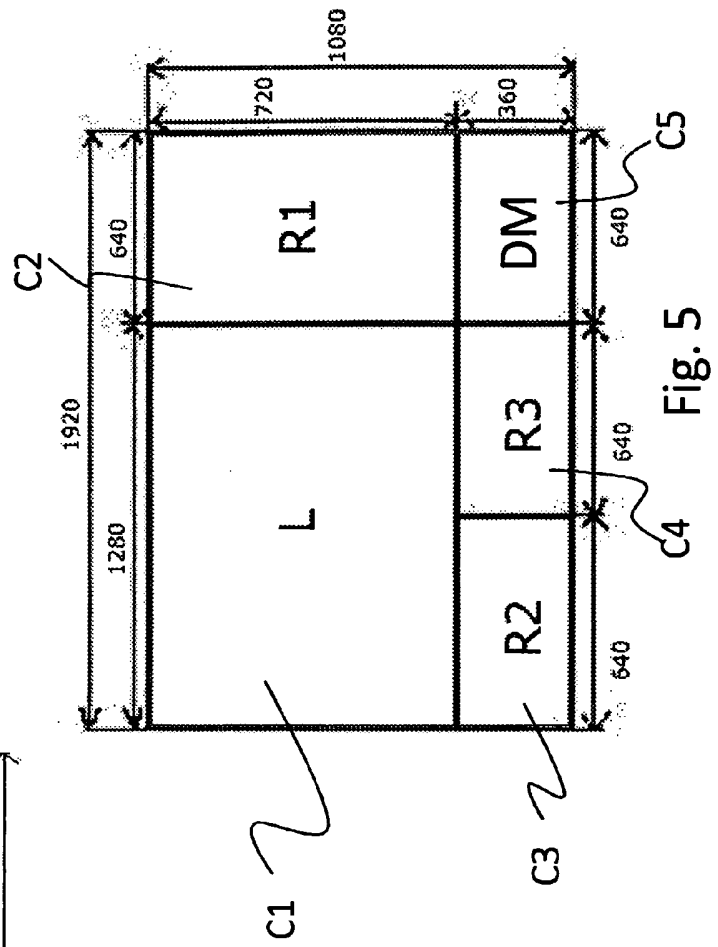
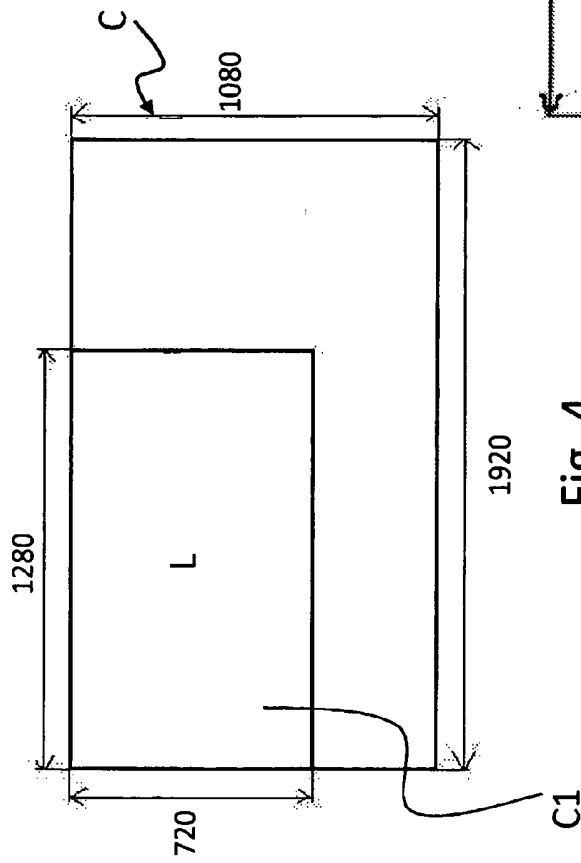


Fig. 2



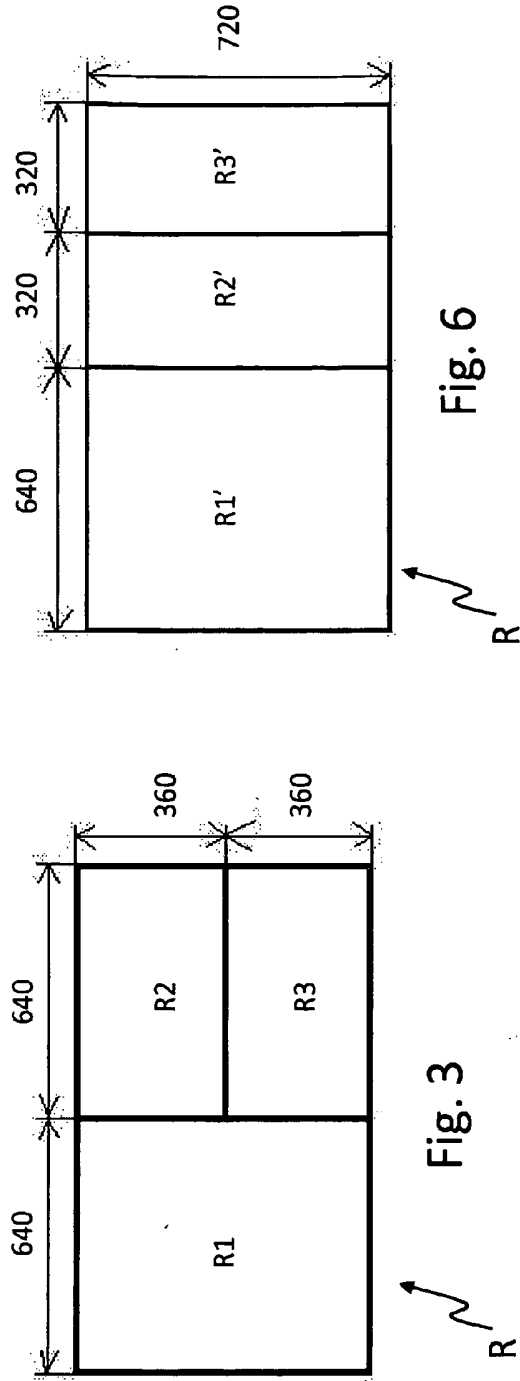


Fig. 6

Fig. 3

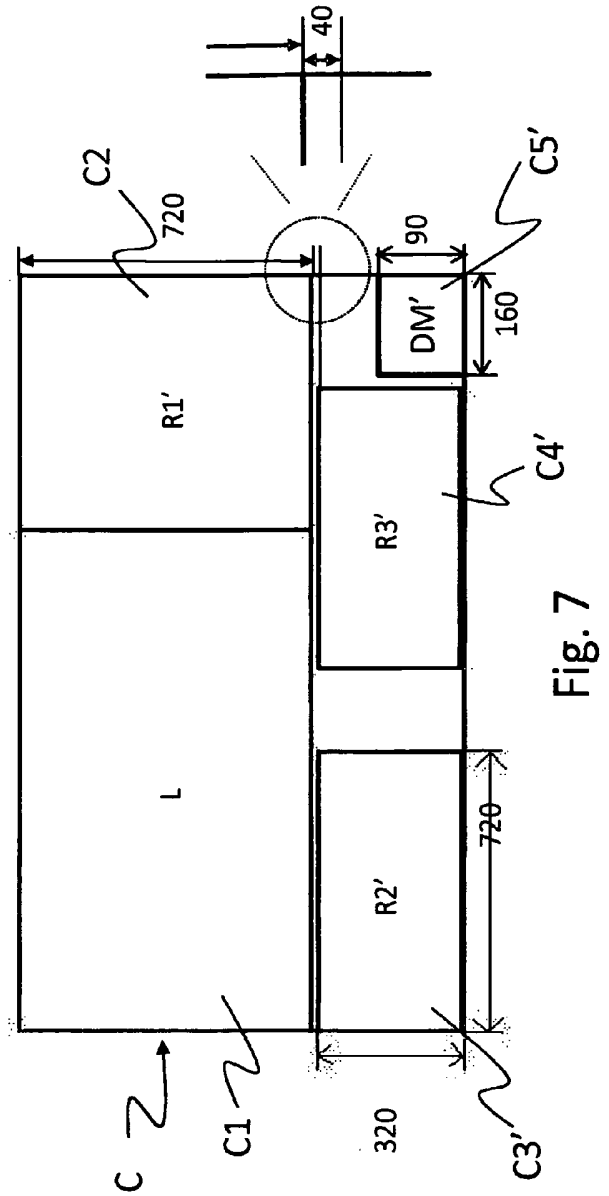


Fig. 7

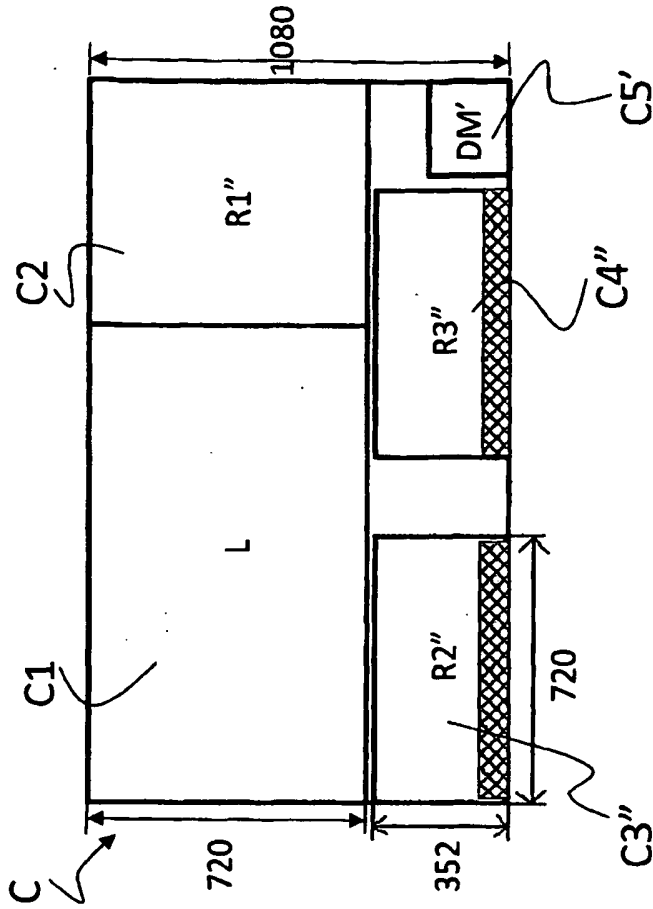


Fig. 9

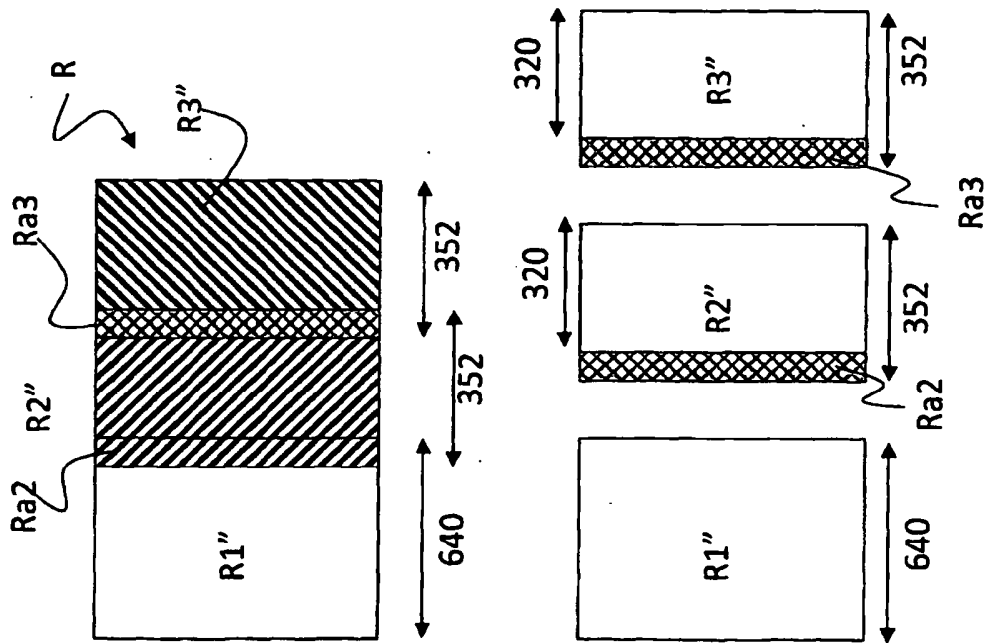


Fig. 8

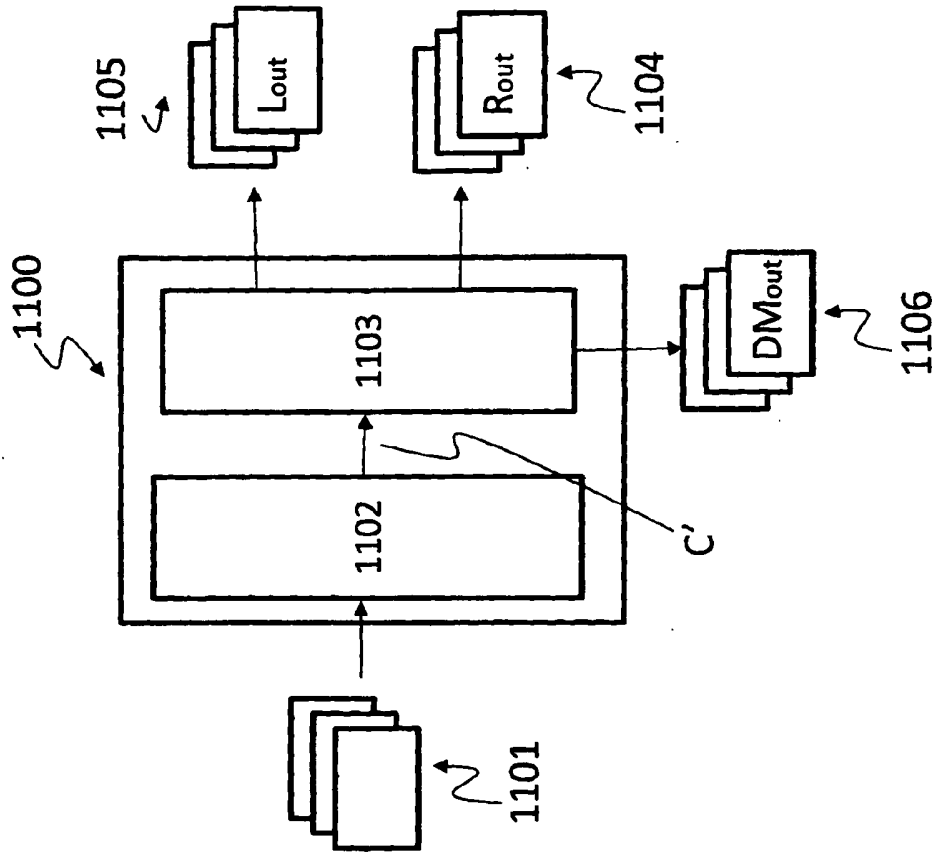


Fig. 10

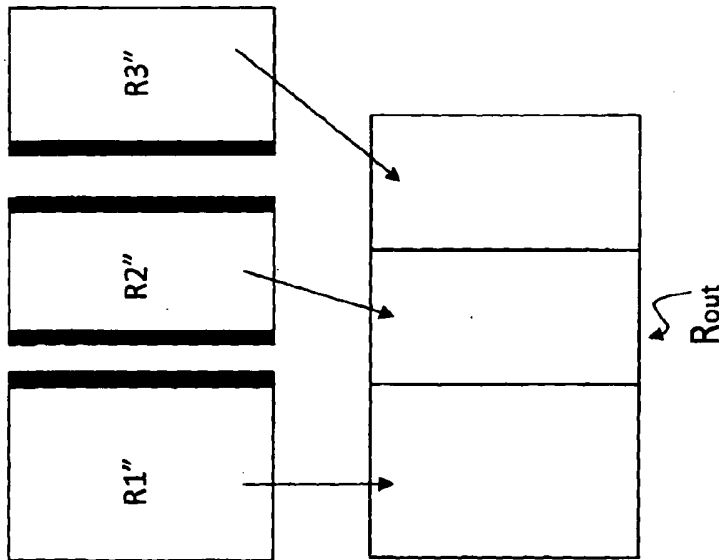


Fig. 11

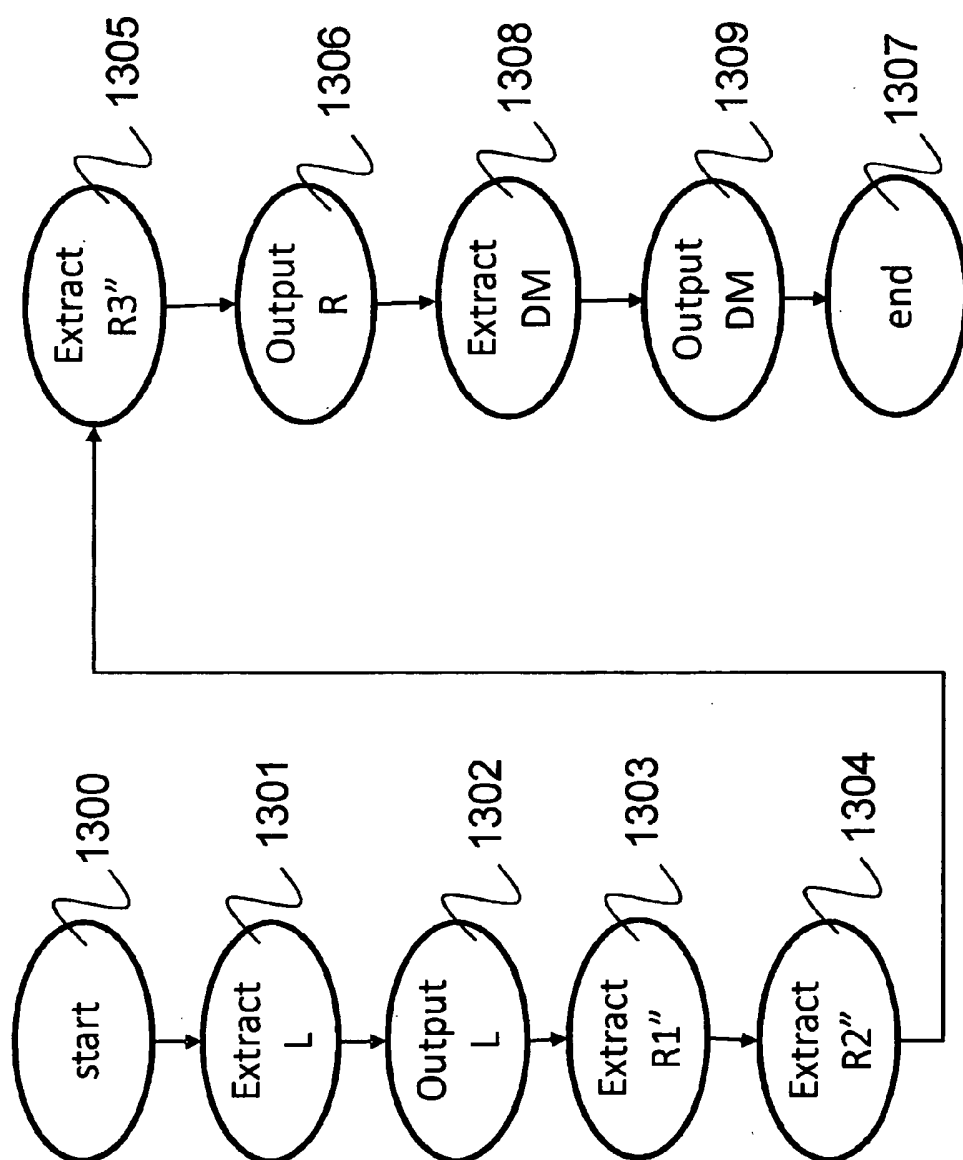


Fig. 12

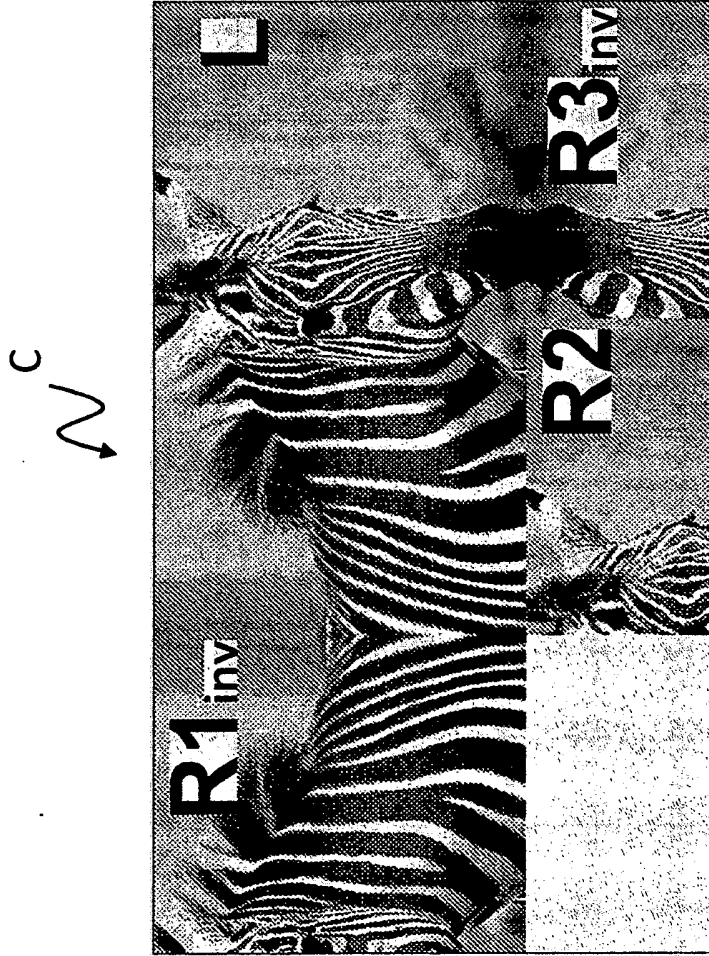


Fig. 13



Fig. 14a

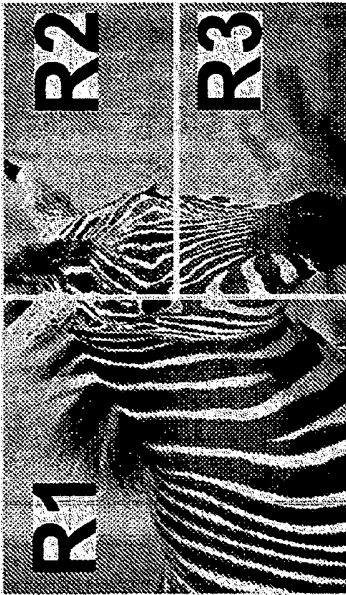


Fig. 14b



Fig. 14c



Fig. 14d

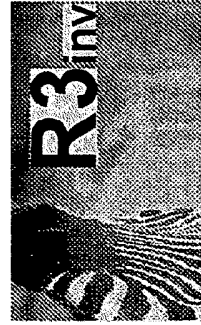


Fig. 14e



Fig. 14f

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- EP 2157803 A1 [0007]
- WO 2008038205 A [0009]
- EP 1705929 A [0010]

Eljárás egy háromdimenziós tartalomra vonatkozó képek kombinálására

Szabadalmi igénypontok

1. Eljárás képek egy háromdimenziós tartalomra illesztésére, ahol egy olyan videó folyamatot veszünk, amely magában foglalja az említett háromdimenziós tartalmat és a képek említett háromdimenziós tartalomra illesztéséhez egy mélység térképet (DM, DM'), az említett mélység térkép (DM, DM') az említett háromdimenziós tartalom mélységére vonatkozó információt tartalmaz, és képként van az említett videó folyam egy képkockájába (C) beleillesztve,

az eljárást az jellemzi, hogy az említett mélység térképet (DM, DM') a lejátszási fázisban csupán képek említett háromdimenziós tartalomra illesztéséhez használjuk, az említett ráillesztést az említett ráillesztési mélység térképtől (DM, DM') függő mélységi pozícióban végezzük,

és hogy az említett mélység térkép (DM, DM') kevesebb számú képpontot tartalmaz, mint az említett képkockában (C) lévő és az említett háromdimenziós tartalommal társított kétdimenziós kép képpontjainak száma.

2. Az 1. igénypont szerinti eljárás, ahol az említett ráillesztési mélység térkép csupán az említett háromdimenziós tartalom alsó felében, előnyösen alsó harmadában elhelyezkedő képpontok mélységére vonatkozó információt tartalmaz.

3. Az 1. vagy 2. igénypont szerinti eljárás, ahol a ráillesztési mélység térkép nem egyforma felbontású, különösen az említett mélység térkép alsó fele vagy harmada nagyobb felbontású, mint a felső rész.

4. Az 1. vagy 2. vagy 3. igénypont szerinti eljárás, ahol az említett ráillesztési mélység térkép kisebb felbontású, mint az említett háromdimenziós tartalommal társított kétdimenziós kép.

5. A 4. igénypont szerinti eljárás, ahol az említett háromdimenziós tartalom egy képpontokból álló kép, és ahol az említett mélység térképet egy olyan mélység térkép alulmintavételezése útján kapjuk, amelynek elemei az említett háromdimenziós tartalom képpontjainak mélységéhez tartoznak.

6. Az 5. igénypont szerinti eljárás, ahol az említett mélység térkép alulmintavételét követően az alulmintavételezett térképet blokkokra osztjuk, és egy blokk minden egyes képpontjának azonos értéket adunk, amely megegyezik az említett blokk képpontjainak minimális mélységével, vagy a blokk képpontjai mélységének a középértékével.

7. Az 5. igénypont szerinti eljárás, ahol az említett mélység térkép alulmintavételét megelőzően a mélység térképet blokkokra osztjuk, és egy blokk minden egyes képpontjának az említett blokk képpontjai legkisebb mélységével egyenlő, vagy a blokk képpontjai mélységének a közepes értékével egyező azonos értéket adunk.

8. A 6. vagy 7. igénypont szerinti eljárás, ahol az említett blokkok egy 2x2 képpont méretű elemi blokk többszörösével azonos méretűek.

9. Az előző igénypontok bármelyike szerinti eljárás, ahol az említett ráillesztési mélység térképet megjelenítésre nem szánt említett képkocka egy részébe adjuk be.

10. Az előző igénypontok bármelyike szerinti eljárás, ahol az említett mélység térképet az említett képkocka (C) olyan területein elosztott blokkokra törjük fel, amelyeket nem foglalt el az említett háromdimenziós tartalom.

11. Az előző igénypontok bármelyike szerinti eljárás, ahol az említett képkocka egy jobboldali képet, egy baloldali képet és az említett mélység térképet tartalmazza, ahol az említett mélység térképet a képkocka (C) azon régióiban elosztott blokkokra törjük fel, amelyeket nem foglalt el az említett háromdimenziós tartalom, és ahol az említett képkocka (C) H.264 kódolási szabvány szerint kódolt.

12. Az 1-9. igénypontok bármelyike szerinti eljárás, ahol az említett háromdimenziós tartalom egy kétdimenziós képet és olyan információt tartalmaz, amely lehetővé teszi egy sztereoszkópikus pár másik képének újjáépítését, és ahol az említett ráillesztési mélység térképet a kétdimenziós kép egy részébe adjuk be.

13. Az 1-12. igénypontok bármelyike szerinti eljárás, ahol az említett képkocka egy jelzöt tartalmaz, amely arra szolgál, hogy a vevő számára jelezze az említett ráillesztési mélység említett képkockán belüli térkép pozícióját.

14. Az 1-13. igénypontok bármelyike szerinti eljárás, ahol az említett videó folyam egy jelzöt tartalmaz, amely arra szolgál, hogy jelezze a vevőnek az említett ráillesztési mélység térkép pozícióját az említett képkockán belül, ahol az említett jelző az említett képkockán kívüli.

15. Berendezés háromdimenziós tartalom visszaadására, amely tartalmaz

olyan eszközöket, amelyek alkalmassá vannak téve háromdimenziós tartalmat magában foglaló videó folyam vételére,

továbbá olyan eszközöket, amelyek alkalmassá vannak téve egy kép említett háromdimenziós tartalommal történő kombinálására,

azzal jellemezve, hogy az említett eszközök, amelyek alkalmassá vannak téve egy kép említett háromdimenziós tartalommal történő kombinálására, alkalmassá vannak téve arra, hogy megvalósítsanak egy az 1-14. igénypontok bármelyike szerinti eljárást,

és eszközöket tartalmaznak az említett ráillesztési mélység térkép kivonására és az említett, a háromdimenziós tartalomra ráillesztendő kép kombinálására, az említett ráillesztési mélység térképtől függő mélység alapján.

16. Sztereoszkópikus videó folyam (1101), amely képkockákat tartalmaz, azzal jellemezve, hogy az említett képkockák legalább egy háromdimenziós tartalmat és az említett képkockák minden egyes képkockáján belül képként beillesztett legalább egy ráillesztési mélység térképet (DM, DM') tartalmaznak, az említett ráillesztési mélység térkép (DM, DM') egy mélység térkép képek az említett háromdimenziós tartalomra illesztésére szolgál, az említett ráillesztés olyan mélységi pozícióban történik, amely az említett ráillesztési mélység térképtől függ, és kisebb képpont számot tartalmaz, mint az említett háromdimenziós tartalommal társított kétdimenziós kép képpontjainak száma, ahol az említett sztereoszkópikus videó folyam alkalmassá van téve arra, hogy az 1-14. igénypontok bármelyike szerinti eljárásban használjuk.