



(51) International Patent Classification:

H04N 13/00 (2006.01) H04N 7/32 (2006.01)  
H04N 21/434 (2011.01)

(21) International Application Number:

PCT/SE2012/050865

(22) International Filing Date:

8 August 2012 (08.08.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant (for all designated States except US): TELEFONAKTIEBOLAGET L M ERICSSON (publ) [SE/SE]; Torshamnsgatan 21-23, S-164 83 Stockholm (SE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): GRAFUL-LA-GONZALÉZ, Beatriz [ES/SE]; Huvudstagan 1D, S-171 44 Solna (SE). GIRDZIJAUSKAS, Ivana [RS/SE]; Svanekegränd 2, S-164 46 Kista (SE). PETTERSSON, Martin [SE/SE]; Stenvallavägen 10, S-186 34 Vallentuna (SE).

(74) Agent: EGRELIUS, Fredrik; Ericsson AB, Patent Unit Kista Device, Service & Media, S-164 80 Stockholm (SE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: 2D/3D IMAGE FORMAT DETECTION

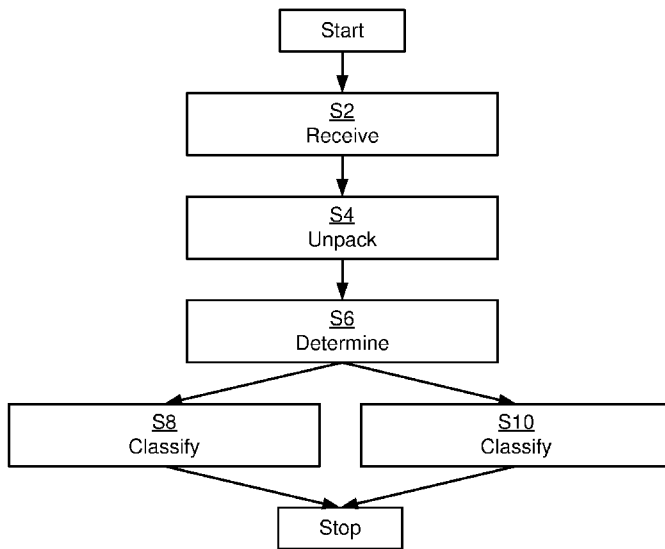


Fig. 13

(57) Abstract: The enclosed embodiments are related to classifying formats for a 3D video enabled device. A first image packed in a first image format is received. The first image is unpacked into at least one second image format, thereby generating at least one second image, wherein the at least one second image format is a 3D image format for a 3D video enabled device. Smoothness is determined for the first image and for the at least one second image. From the determined smoothness, the first image format is classified to be one 3D image format as a result of the at least one second image being smoother than the first image; or the first image format is classified to be a 2D image format as a result of the first image being equally smooth or smoother than the at least one second image.



**Declarations under Rule 4.17:**

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *of inventorship (Rule 4.17(iv))*

**Published:**

- *with international search report (Art. 21(3))*

## 2D/3D IMAGE FORMAT DETECTION

### TECHNICAL FIELD

Embodiments presented herein relate to image formats for a 3D video enabled device and more particularly to classifying image formats for a 3D video enabled device.

### BACKGROUND

The research in 3D imaging has gained a lot of momentum in recent years. A number of 3D movies are being produced every year, providing stereoscopic effects to the viewers. It is forecasted that soon there will be commercially available 3D-enabled mobile communications devices (such as mobile phones, so-called smart-phones, tablet computers, laptops, etc.).

Various 3D image and video formats coexist today. Some of the 3D image and video formats are based on a left image and a right image. The left image is to be seen by the left eye and the right image by the right eye. Various techniques exist to make sure one eye does not see both images, including the use of polarized screens and glasses, shutter based glasses and directional views. Many products require the knowledge of a 3D format in order to work properly. For example, a 3D video player typically needs to know how the left and the right images are arranged before it is able to play the content. A set top box typically needs to know the 3D format in order to properly place text or graphics on top of the video content etc.

Currently, there is no standard for signaling 3D formats over networks. Many standardization bodies are working on 3D-related standards, but it will take a couple of years until a consensus has been reached. What exists today is the HDMI 1.4a standard for transferring audio and video between devices and screens. HDMI 1.4a defines a number of 3D formats including interlaced, top-bottom, side-by-side and 2D + depth. One way of detecting the received 3D format is by manual detection. This process is cumbersome and should preferably be avoided.

In “3D image format identification by image difference,” by Tao Zhang in Proceedings of IEEE International Conference on Multimedia and Expo, July 2010, pp.1415-1420 there is proposed two different detection methods depending on if a format is assumed to be blending or non-blending. For the  
5 so-called blending formats, a pixel from one view is always surrounded by pixels from the other view. Examples of blending formats are interlaced and checkerboard. A format is non-blending if pixels from one view are surrounded by pixels from the same view except in places such as the view boundary. Side-by-side and top-bottom formats are examples of non-  
10 blending formats. The initial image is unpacked to the left and right view image given an assumption on its format. For non-blending formats, the author proposes computing an edge map on a difference between the left and right images. To simplify the algorithm, the initial edge map has some of the edges removed. The parameters used for edge removal as well as for  
15 subsequent measuring of remaining edge widths and format identification are based on the statistics of edge widths for these kinds of images. Non-blending formats, on the other hand, are tested by comparing histograms of left and right images. This algorithm may be considered as complex and sensitive to the order in which the formats are tested. Namely, to achieve  
20 higher accuracy, the authors propose to always test blending before non-blending formats. However, even with this suggestion, the success rate is not 100%.

Hence in view of the above there is a need for improved classification of different 3D formats.

## 25 **SUMMARY**

An object of the enclosed embodiments is to provide improved classification of different 3D formats.

As noted above, one disadvantage of the method proposed in “3D image format identification by image difference” is complexity. Edge detection in  
30 combination with morphological operations for insignificant edge removal is a computationally demanding process. Histogram matching further adds to

complexity. Finally, the algorithm is sensitive to the order in which the formats are being tested and does not in general guarantee the 100% success rate. One particular object is therefore to provide reliable low-complexity classification of different 3D formats.

- 5 According to a first aspect a method of classifying formats for a 3D video enabled device is provided. The method comprises receiving a first image packed in a first image format. The method further comprises unpacking the first image into at least one second image format, thereby generating at least one second image, wherein the at least one second image format is a 3D
- 10 image format for a 3D video enabled device. The method further comprises determining smoothness for the first image and for the at least one second image. The method further comprises classifying, from the determined smoothness, the first image format to be one 3D image format as a result of the at least one second image being smoother than the first image; or
- 15 classifying, from the determined smoothness, the first image format to be a 2D image format as a result of the first image being equally smooth or smoother than the at least one second image.

Advantageously the disclosed method enables automatic format detection based on the smoothness metric.

- 20 Advantageously the disclosed method is simple whilst providing a high format prediction accuracy and is not sensitive to the order in which the formats are tested.

- According to a second aspect a computer program of classifying formats for a 3D video enabled device is provided. The computer program comprises
- 25 computer program code which, when run on a processing unit, causes the processing unit to perform a method according to the first aspect.

According to a third aspect a computer program product comprising a computer program according to the second aspect and a computer readable means on which the computer program is stored are provided.

According to a fourth aspect a device for classifying formats for a 3D video enabled device is provided. The device comprises an image receiver arranged to receive a first image packed in a first image format. The device further comprises an image unpacker arranged to unpack the first image into at least one second image format, thereby generating at least one second image, wherein the at least one second image format is a 3D image format for a 3D video enabled device. The device further comprises a smoothness determiner arranged to determine smoothness for the first image and for the at least one second image. The device further comprises a classifier arranged to classify, from the determined smoothness, the first image format to be one 3D image format as a result of the at least one second image being smoother than the first image; or to classify, from the determined smoothness, the first image format to be a 2D image format as a result of the first image being equally smooth or smoother than the at least one second image. The functionalities of the image receiver, the image unpacker, the smoothness determiner and the classifier may be implemented by a processing unit.

It is to be noted that any feature of the first, second, third and fourth aspects may be applied to any other aspect, wherever appropriate. Likewise, any advantage of the first aspect may equally apply to the second, third and/or fourth aspects, respectively, and vice versa. Other objectives, features and advantages of the enclosed embodiments will be apparent from the following detailed disclosure, from the attached dependent claims as well as from the drawings.

Generally, all terms used in the claims are to be interpreted according to their ordinary meaning in the technical field, unless explicitly defined otherwise herein. All references to "a/an/the element, apparatus, component, means, step, etc." are to be interpreted openly as referring to at least one instance of the element, apparatus, component, means, step, etc., unless explicitly stated otherwise. The steps of any method disclosed herein do not have to be performed in the exact order disclosed, unless explicitly stated.

**BRIEF DESCRIPTION OF THE DRAWINGS**

The disclosed embodiments are now described, by way of example, with reference to the accompanying drawings, in which:

- Fig 1 is a schematic diagram illustrating image packing into different formats;
- 5 Fig 2 is a schematic diagram showing functional modules of a device;
- Fig 3 shows one example of a computer program product comprising computer readable means;
- Figs 4 and 5 are schematic diagrams illustrating image unpacking into different formats;
- 10 Figs 6-8 are photographs where different image formats have been used;
- Fig 9 illustrates a pixel surrounded by its neighbouring pixels;
- Fig 10 is a schematic diagram illustrating columns in an image;
- Fig 11 is a schematic diagram illustrating rows in an image;
- Fig 12 is a schematic diagram illustrating different areas in an image; and
- 15 Fig 13 is a flowchart of a method according to embodiments.

**DETAILED DESCRIPTION**

- The invention will now be described more fully hereinafter with reference to the accompanying drawings, in which certain embodiments of the invention are shown. This invention may, however, be embodied in many different
- 20 forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided by way of example so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art. Like numbers refer to like elements throughout the description.

Various 3D image and video formats coexist today. Some of the 3D image and video formats are based on a left image 20 and a right image 22 as illustrated in Fig 1 and as described above. Fig 1 also illustrates that the left image 20 and the right image 22 are down-sampled and then placed in one image  
5 frame. For illustrative purposes, in all formats illustrated in Fig 1 the left image part has been shaded and the right image part is white. In the so-called side-by-side format 24 and the so-called top-bottom format 26, left and right images are down-sampled in the horizontal or vertical direction respectively before being packed in a single frame. A frame in the so-called interlaced  
10 format 28 comprises left and right views down-sampled in vertical direction and packed such that odd lines correspond to one view and even lines to the other view. In the so-called checkerboard format 30, left and right views are sampled in such a way that each line has interlaced pixels from each view, yielding a mosaic format. The side-by-side format and the top-bottom  
15 formats are examples of non-blending formats. The interlaced and checkerboard formats are examples of blending formats. Side-by-side and top-bottom formats are typically used for transmission of 3D video, whereas interlaced and checkerboard formats are typically used as display input formats.

20 As noted above, an object of the enclosed embodiments is to provide improved classification of different 3D formats. A device 2 is therefore provided. Fig 2 schematically illustrates, in terms of a number of functional modules, the components of the device 2. A processing unit 4 is provided using any combination of one or more of a suitable central processing unit  
25 (CPU), multiprocessor, microcontroller, digital signal processor (DSP), application specific integrated circuit (ASIC) etc., capable of executing software instructions stored in a computer program product 14 (as in Fig 3). In the device 2 the computer program product 13 may be stored in the memory 6. Thus the processing unit 4 is thereby preferably arranged to  
30 execute methods as herein disclosed. The client device 2 further comprises an input/output (I/O) interface 8 having a transmitter 10 and a receiver 12, for communicating with other devices. The processing unit 4 may in turn



comprise an image receiver 4a, an image unpacker 4b, a smoothness determiner 4c and a classifier 4d, functionalities of which will be disclosed below. The functionalities of the image receiver 4a, the image unpacker 4b, the smoothness determiner 4c and the classifier 4d may thus be implemented  
5 by the processing unit 4. Other components, as well as the related functionality, of the client device 2 are omitted in order not to obscure the concepts presented herein.

Fig 13 is a flowchart illustrating embodiments of a method of classifying formats for a 3D video enabled device. The method is preferably performed in  
10 the device 2. The methods are advantageously provided as computer programs 16. Fig 3 shows one example of a computer program product 14 comprising computer readable means 18. On this computer readable means 18, a computer program 16 can be stored. This computer program 16 can cause the processing unit 4 of the device 2 and thereto operatively coupled  
15 entities and devices to execute methods according to embodiments described herein. In the example of Fig 3, the computer program product 14 is illustrated as an optical disc, such as a CD (compact disc) or a DVD (digital versatile disc) or a Blu-Ray disc. The computer program product could also be embodied as a memory (RAM, ROM, EPROM, EEPROM) and more  
20 particularly as a non-volatile storage medium of a device in an external memory such as a USB (Universal Serial Bus) memory. Thus, while the computer program 16 is here schematically shown as a track on the depicted optical disk, the computer program 16 can be stored in any way which is suitable for the computer program product 14.

25 In a step S2 a first image is received. The first image is received by the image receiver 4a. The first image is packed in a first image format. The first image format may be a format known by the device 2. In general, the device 2 is configured to pack and unpack images in a number of different formats. However, the device 2 does not beforehand know according to which one  
30 format of these different formats the first image has been packed into. Thus, the device 2 may not have any a priori knowledge regarding which of the known formats the first image has been packed into.

In order for the device 2 to determine what format the first image format is the received image is unpacked. Thus, in a step S4 the first image is unpacked into at least one second image format. The first image is unpacked by the image unpacker 4b. Thereby at least one second image is generated where  
5 each one of the at least one second image is in one of the at least one second image format. The at least one second image is generated so that properties, such as smoothness, can be compared between the first image and the at least one second image. The at least one second image format is a 3D image format for a 3D video enabled device. In general terms, a 3D video enabled device  
10 may be a 3D TV, a 3D TV set top box, an integrated DVD or Blue-ray player, a mobile 3D video device and any device which may be used for 3D video communication. Hence the first image may, by means of the at least one second format, be unpacked into a stereo pair. The unpacking follows the inverse steps of image packing, as explained above with reference to Fig 1.  
15 The process of unpacking is schematically illustrated by the arrows in Figs 4 and 5 for an image 28 packed in the interlaced format and an image 30 packed in a checkerboard format. The resulting unpacked left image 20', 20" and unpacked right image 22', 22" are then interpolated so that the dimensions of the original left image 20 and right image 22 (as illustrated in  
20 Fig 1) are obtained. For illustrative purposes, in all formats illustrated in Figs 4 and 5 the left image part has been shaded and the right image part is white.

The at least one second image format may or may not be equal to the first image format. A number of steps may be executed to determine, by  
25 classification, whether the at least one second image format is equal to the first image format or not, or alternatively whether the first image format is equal to another format (such as a third format) or not. The methods disclosed herein are hence based on evaluating a smoothness criterion and hence the first image format may be identified based on smoothness.

In general terms, a natural image is smoother than any image obtained by  
30 some of the packing arrangements, such as interlaced 28 or checkerboard 30 for example. To illustrate this, Fig 6 shows an example of a 2D image 32. Two frame-compatible 3D images (interlaced 34 as in Fig 7 and checkerboard 36

as in Fig 8) are provided for comparison. The upper part of each figure shows a zoomed-in part 32', 34', 36' of the entire image 32, 34, 36 illustrated in the lower part of each figure. Local variation of pixel values in the neighborhood of each pixel is smaller for the 2D image and this measure can be used as a reliable indicator of a format. The inventors of the enclosed embodiments have discovered that these findings can be used for format detection, namely that the correct image format is the one that produces the highest smoothness indicator.

Therefore smoothness is in a step S6 determined for the first image and for the at least one second image. The smoothness is determined by a smoothness determiner 4c. Hence, smoothness is determined both for the received first image and for the at least one second image which is an unpacked version of the first image. In general there may be different ways to determine smoothness. For example, the smoothness for one image may be determined from a pixel intensity variation between neighbouring pixels in said one image. Fig 9 and the following expression illustrate one example of a smoothness metric:

$$sm\_value(i, j) = \sum_{k=-l}^l \sum_{l=-1}^1 (p(i, j) - p(i+k, j+l))^2, \quad (1)$$

that reflects the pixel intensity variation as a smoothness value "sm\_value" for a pixel at position (i,j) and its immediate neighborhood determined by l number of pixels in vertical and horizontal direction, respectively. Fig 9 illustrates an image part 38 in which a center pixel (i,j) is related to its neighbouring pixels by means of arrows. The total smoothness value for the unpacked stereo pair may be obtained by summing the smoothness values determined according to expression (1) for all the pixels in both images (i.e. left and right). However, it should be clear to anyone skilled in the art that other metrics as well as pixel neighborhood can be used instead. Moreover, only parts of the images may be considered during the smoothness determination and/or comparison. This may simplify the smoothness

determination. It may also reduce the memory requirements involved for the smoothness determination.

From the determined smoothness, the first image format is, in a step S8, classified to be one 3D image format as a result of the at least one second  
5 image being smoother than the first image. Likewise, the first image format is, in a step S10, classified to be a 2D image format as a result of the first image being equally smooth or smoother than the at least one second image. The classifying is performed by the classifier 4d.

A straightforward way of detecting a format would be to unpack an image  
10 according to all the possible formats. However, as is clear from steps S8 and S10 the number of unpacking operations may be reduced by comparing smoothness for the first image and the at least one unpacked image.

Regular 2D images are in general terms smoother than images packed in 3D formats, such as the interlaced 28 or checkerboard 30 formats. Images  
15 packed according to the side-by-side 24 or top-bottom 26 formats may generally have similar smoothness as 2D images, with an exception of sharp transitions around a mid-horizontal line (in case of top-bottom) or a mid-vertical line (for side-by-side format). Therefore, additional criteria may be set to further distinguish a 2D format from side-by-side and top-bottom  
20 formats.

Further, the method of determining the first image format may be improved by considering that 2D formats generally have similar smoothness as images in the side-by-side and top-bottom formats (or any other non-blending format). Hence, in general terms, the at least one second image format may  
25 be chosen from a group of at least one blending image format and at least one non-blending image format. As noted above, the side-by-side format and the top-bottom formats are examples of non-blending formats whereas the interlaced and checkerboard formats are examples of blending formats. Thus, in case the at least one second image format is a blending format the second  
30 image format may be from a group of an interlaced image format and a

checkerboard image format. Likewise, in case the at least one second image format is a non-blending format the second image format may be from a group of a top-bottom image format and a side-by-side image format.

To further distinguish one blending format from another blending format the step S4 of unpacking may comprise unpacking the first image into an interlaced image format, thereby generating a first second image and/or unpacking the first image into a checkerboard image format, thereby generating a second second image. Hence, the first image may be unpacked into two second images; a first second image and a second second image, each being of a different format. Then the steps S8 and S10 of classifying may comprise classifying the first image format to be an interlaced image format as a result of the first second image being smoother than the first image and the second second image. Likewise, the steps S8 and S10 of classifying may comprise classifying the first image format to be a checkerboard image format as a result of the second second image being smoother than the first image and the first second image. If any of these pairs has smoother images than the unpacked image, the first image format is neither in a 2D format nor in a non-blending format. As a consequence thereof, if the smoothest image is the unpacked first image, the format is 2D, top-bottom or side-by-side.

To further determine the first image format, the presence of possible large pixel intensity variations around the mid-horizontal and mid-vertical lines may be investigated. Such variations may typically be indicators of top-bottom and side-by-side formats respectively. There may be different ways to investigate the presence of such variations. For example, the following expression may be used to determine the vertical pixel intensity variation “ver\_pixel\_variation”, following the notation from Fig 10 where  $v_1$ ,  $v_2$ ,  $v_3$  and  $v_4$  denote four columns of pixels in the image 40 and where  $H$  is the number of pixels in each column:

$$ver\_pixel\_variation = \sum_{i=1}^H |v_2(i) - v_3(i)| - \frac{1}{2} \left[ \sum_{i=1}^H |v_1(i) - v_2(i)| + \sum_{i=1}^H |v_3(i) - v_4(i)| \right]$$

30 (2a)

For example, the following expression may be used to determine the horizontal pixel intensity variation “hor\_pixel\_variation”, following the notation from Fig 11 where h1, h2, h3 and h4 denote four rows of pixels in the image 42 and where W is the number of pixels in each row:

$$5 \quad \text{hor\_pixel\_variation} = \sum_{j=1}^W |h_2(j) - h_3(j)| - \frac{1}{2} \left[ \sum_{j=1}^W |h_1(j) - h_2(j)| + \sum_{j=1}^W |h_3(j) - h_4(j)| \right] \quad (3a)$$

A ratio may then be determined between expressions (2a) and (3a) for at least a part of the first image. For example, the format is side-by-side if

$$\frac{\text{ver\_pixel\_variation}}{\text{hor\_pixel\_variation}} \geq \text{th\_sbs},$$

10 threshold. That is, if this ratio is larger than a first pre-determined threshold, “th\_sbs”, said first image format to be a side-by-side format. For example, if the ratio  $\frac{\text{hor\_pixel\_variation}}{\text{ver\_pixel\_variation}} \geq \text{th\_tb}$  the format is top-bottom. Thus, as a result of this ratio being larger than a second pre-determined threshold, “th\_tb” the first image format to be a top-bottom image format.

15 According to some embodiments, in addition to the above “ver\_pixel\_variation” a second vertical pixel variation, “ver\_pixel\_variation2”, is determined according to:

$$\text{ver\_pixel\_variation2} = \sum_{i=1}^H |v_4(i) - v_5(i)| - \frac{1}{2} \left[ \sum_{i=1}^H |v_3(i) - v_4(i)| + \sum_{i=1}^H |v_5(i) - v_6(i)| \right] \quad (2b)$$

20 where the calculation in comparison to the determination of “ver\_pixel\_variation” in expression (2a) has been shifted two pixels in horizontal direction, following the notation from Fig 10 where v4 and v5 denote two additional columns of pixels in the image 40.

A ratio may then be determined between expressions (2a) and (2b) for at least a part of the image. For example, the format may be detected as side-by-side if  $\frac{ver\_pixel\_variation}{ver\_pixel\_variation2} \geq th\_sbs$ .

Likewise, in addition to the above “hor\_pixel\_variation” a second horizontal pixel variation “hor\_pixel\_variation2” may according to embodiments be determined according to:

$$hor\_pixel\_variation2 = \sum_{i=1}^W |h_4(i) - h_5(i)| - \frac{1}{2} \left[ \sum_{i=1}^W |h_3(i) - h_4(i)| + \sum_{i=1}^W |h_5(i) - h_6(i)| \right]$$

(3b)

where the calculation in comparison to the determination of “hor\_pixel\_variation” in expression (3a) has been shifted two pixels in vertical direction”, following the notation from Fig 11 where h5 and h6 denote two additional rows of pixels in the image 42.

A ratio may then be determined between expressions (3a) and (3b) for at least a part of the image. For example, the format may be detected as top-bottom if  $\frac{hor\_pixel\_variation}{hor\_pixel\_variation2} \geq th\_tb$ .

Finally, if the ratios between the horizontal and vertical pixel variation do not exceed any of the thresholds, it is assumed that the format is 2D. Hence, a result of the ratio between horizontal and vertical pixel intensity variations being smaller than the first pre-determined threshold and the ratio between vertical and horizontal pixel intensity variations being smaller than the second pre-determined threshold, is that the first image format is a 2D image format.

As noted above,  $th\_sbs$  and  $th\_tb$  represent the thresholds indicating the side-by-side and top-bottom formats respectively. They can be hard-coded or obtained from the statistics of ratios between the pixel variations for different

types of images. Note that this is just one of the ways of distinguishing between the two formats.

In some cases each view in the side-by-side or top-bottom format could have a colored border (normally black) around it, resulting in that the above  
5 approach could have a difficulty to distinguish a side-by-side format or a top bottom format from a 2D format. However, such a border could easily be detected and removed by scanning for a solid color along and around the mid-horizontal or mid-vertical line.

The disclosed embodiments may also be used to detect if an overlay object,  
10 such as picture-in-picture, thumbnail or similar is present in a normal 2D video at an expected position. Instead of checking for possible large pixel intensity variations along a central line, large pixel intensity variations could additionally and/or alternatively be checked for around an expected object border. If the object exists large pixel intensity variation can be expected,  
15 otherwise not.

In an alternative embodiment of the current invention a slightly different approach may be used for distinguishing side-by-side-formats from top-bottom formats from 2D formats.

Instead of checking for possible large pixel intensity variations along a central  
20 line, a sufficiently large area could be selected in the first and second views. Each line of the two areas is then summed up and the global difference between the two areas is determined. To accommodate for horizontal shifts between the two views, the calculation is repeated with different shifts  $d$  between the first and the second view areas, as illustrated in Fig 12. The first  
25 image may therefore be unpacked into at least one third image format, thereby generating at least one third image, wherein the at least one third image format is a non-blending format for a 3D video enabled device. Each one of the at least one third image 44 includes a respective image pair comprising a left view and a right view. A first area 46 in a left view and a  
30 second area in the right view 48 are then selected. The first area and the



second area have the same dimensions. A pixel difference between the first area and the second area for at least one pixel shift between the first area and the second area is then determined. The pixel difference “global\_diff<sub>d</sub>” for a pixel shift d may be determined as

5

$$global\_diff_d = \sum_{y=0}^h \left( abs \left( \sum_{x=0}^w p_{Axy} - \sum_{x=d}^{w+d} p_{Bxy} \right) \right) \quad (4)$$

where w and h is the height and width of the selected areas of the views, p<sub>Axy</sub> and p<sub>Bxy</sub> are pixels in the left and right view and d denotes a shift as illustrated in Fig 12.

10 Then, a minimum difference may be determined as the minimum of the determined pixel difference over all shifts d. A ratio between the minimum difference and the number of pixels in the first area may then be determined. As a result of the ratio being larger than a third pre-determined threshold the first image format is determined to be the at least one third image format. Put  
15 in other words, to distinguish the unpacked format from a 2D format, the minimum global shift averaged per pixel,

$$\frac{global\_diff_{min}}{num\_pixels\_in\_area}, \quad (5)$$

can be compared to the third pre-determined threshold. If the averaged minimum global shift is below the threshold, an unpacked format of either  
20 side-by-side or top-bottom has been detected; otherwise it is detected to not be any of those formats.

Alternatively, to accommodate for the differences between different kinds of content, the global differences can be calculated also in the vertical direction, but using the same area and positions as optimized for the horizontal  
25 direction. Then, if the ratio

$$\frac{\textit{global\_diff\_horizontal}_{\min}}{\textit{global\_diff\_vertical}} \quad (6)$$

is close to 0 it is a strong indication that the unpacked format has been detected (as side-by-side or top-bottom). If the ratio is close to 1, this is a strong indication that the unpacked format is not the correct format.

- 5 The invention has mainly been described above with reference to a few embodiments. However, as is readily appreciated by a person skilled in the art, other embodiments than the ones disclosed above are equally possible within the scope of the invention, as defined by the appended patent claims. For example, although only side-by-side, top-bottom, interlaced and checker  
10 board 3D formats have been explicitly mentioned above it is to be understood that the disclosed embodiments can be extended to other 3D formats, including variations of the above mentioned formats. For example, although it has been assumed that the 3D video corresponds to stereo video, it should be clear to the skilled person that the disclosed embodiments equally apply  
15 for other 3D formats as well, including video plus depth (V+D), Multiview Video (MVV), Multiview Video plus Depth (MVD), Layered Depth Video (LDV) and Depth Enhanced Video (DES).

**CLAIMS**

1. A method of classifying formats for a 3D video enabled device, comprising the steps of:
  - receiving (S2) a first image packed in a first image format;
  - 5        unpacking (S4) the first image into at least one second image format, thereby generating at least one second image, wherein the at least one second image format is a 3D image format for a 3D video enabled device;
  - determining (S6) smoothness for the first image and for the at least one second image; and
  - 10        classifying (S8), from the determined smoothness, the first image format to be one 3D image format as a result of the at least one second image being smoother than the first image; or
  - classifying (S10), from the determined smoothness, the first image format to be a 2D image format as a result of the first image being equally
  - 15        smooth or smoother than the at least one second image.
  
2. The method according to claim 1, wherein the at least one second image format is from a group of at least one blending image format and at least one non-blending image format.
  
3. The method according to claim 2, wherein in case the at least one
- 20        second image format is a blending format the second image format is from a group of an interlaced image format and a checkerboard image format, and wherein in case the at least one second image format is a non-blending format the second image format is from a group of a top-bottom image format and a side-by-side image format.
  
- 25        4. The method according to claim 1 or 2 or 3, wherein the step of unpacking comprises unpacking the first image into an interlaced image format, thereby generating a first second image and unpacking the first image into a checkerboard image format, thereby generating a second second image.

5. The method according to claim 4, wherein the step of classifying comprises classifying the first image format to be an interlaced image format as a result of the first second image being smoother than the first image and the second second image.
- 5 6. The method according to claim 4, wherein the step of classifying comprises classifying the first image format to be a checkerboard image format as a result of the second second image being smoother than the first image and the first second image.
7. The method according to any one of the preceding claims, wherein the  
10 smoothness for one image is determined from a pixel intensity variation between neighbouring pixels in said one image.

8. The method according to claim 7, wherein the smoothness is determined as

$$sm\_value(i, j) = \sum_{k=-l}^l \sum_{l=-1}^1 (p(i, j) - p(i+k, j+l))^2 .$$

- 15 9. The method according to any one of the preceding claims, further comprising:  
determining, for at least a part of the first image, a ratio between horizontal and vertical pixel intensity variations.
10. The method according to any one of the preceding claims, further  
20 comprising:  
determining, for at least a part of the first image, a ratio between two horizontal pixel intensity variations, where the first horizontal pixel intensity variation is chosen along a horizontal line expected to divide the first image in a top part and a bottom part, respectively.
- 25 11. The method according to claim 9 or 10, further comprising:  
classifying, as a result of the ratio being larger than a first pre-determined threshold, said first image format to be a top-bottom image format for the 3D video enabled device.

12. The method according to any one of the preceding claims, further comprising:

determining, for at least a part of the first image, a ratio between vertical and horizontal pixel intensity variations.

5 13. The method according to any one of the preceding claims, further comprising:

determining, for at least a part of the first image, a ratio between two vertical pixel intensity variations, where the first vertical pixel intensity variation is chosen along a vertical line expected to divide the first image in a right part and a left part, respectively.

14. The method according to claim 12 or 13, further comprising:

classifying, as a result of the ratio being larger than a second pre-determined threshold, said first image format to be a side-by-side image format for the 3D video enabled device.

15 15. The method according to any one of claims 9 to 14, wherein the horizontal pixel intensity variation is determined as

$$ver\_pixel\_variation = \sum_{i=1}^H |v_2(i) - v_3(i)| - \frac{1}{2} \left[ \sum_{i=1}^H |v_1(i) - v_2(i)| + \sum_{i=1}^H |v_3(i) - v_4(i)| \right]; \text{ and}$$

wherein the vertical pixel intensity variation is determined as

$$hor\_pixel\_variation = \sum_{j=1}^W |h_2(j) - h_3(j)| - \frac{1}{2} \left[ \sum_{j=1}^W |h_1(j) - h_2(j)| + \sum_{j=1}^W |h_3(j) - h_4(j)| \right].$$

20 16. The method according to claim 11 and 14, further comprising:

classifying, as a result of the ratio between horizontal and vertical pixel intensity variations being smaller than the first pre-determined threshold and the ratio between vertical and horizontal pixel intensity variations being smaller than the second pre-determined threshold, said first image format to be a 2D image format.

17. The method according to any one of the preceding claims, further comprising:

unpacking the first image into at least one third image format, thereby generating at least one third image, wherein the at least one third image format is a non-blending format for the 3D video enabled device.

18. The method according to claim 14, wherein each one of the at least one  
5 third image includes a respective image pair comprising a left view and a right view, the method further comprising:

selecting a first area in a left view and selecting a second area in the right view, the first area and the second area having the same dimensions;

10 determining a pixel difference between the first area and the second area for at least one pixel shift between the first area and the second area.

19. The method according to claim 15, wherein the pixel difference for pixel shift  $d$  is determined as

$$global\_diff_d = \sum_h^{y=0} \left( abs \left( \sum_w^{x=0} P_{Axy} - \sum_{w+d}^{x=d} P_{Bxy} \right) \right).$$

20. The method according to claim 14 or 15, further comprising  
15 determining a minimum difference as the minimum of the determined pixel difference over all shifts  $d$ ;

determining a ratio between the minimum difference and the number of pixels in the first area; and

20 classifying, as a result of the ratio being larger than a third pre-determined threshold, the first image format to be the at least one third image format.

21. A computer program (18) of classifying formats for 3D video enabled devices, the computer program comprising computer program code which, when run on a processing unit (4), causes the processing unit to

25 receive (S2) a first image packed in a first image format;

unpack (S4) the first image into at least one second image format, thereby generating at least one second image, wherein the at least one second image format is a 3D image format for a 3D video enabled device;

determine (S6) smoothness for the first image and for the at least one

second image; and

classify (S8), from the determined smoothness, the first image format to be one 3D image format as a result of the at least one second image being smoother than the first image; or

- 5        classify (S10), from the determined smoothness, the first image format to be a 2D image format as a result of the first image being equally smooth or smoother than the at least one second image.

22. A computer program product (16) comprising a computer program (18) according to claim 21 and a computer readable means (20) on which the  
10 computer program is stored.

23. A device (2) for classifying formats for 3D video enabled devices, comprising

an image receiver arranged to receive a first image packed in a first image format;

- 15        an image unpacker arranged to unpack the first image into at least one second image format, thereby generating at least one second image, wherein the at least one second image format is a 3D image format for a 3D video enabled device;

20        a smoothness determiner arranged to determine smoothness for the first image and for the at least one second image; and

a classifier arranged to classify, from the determined smoothness, the first image format to be one 3D image format as a result of the at least one second image being smoother than the first image; or

- 25        to classify, from the determined smoothness, the first image format to be a 2D image format as a result of the first image being equally smooth or smoother than the at least one second image.

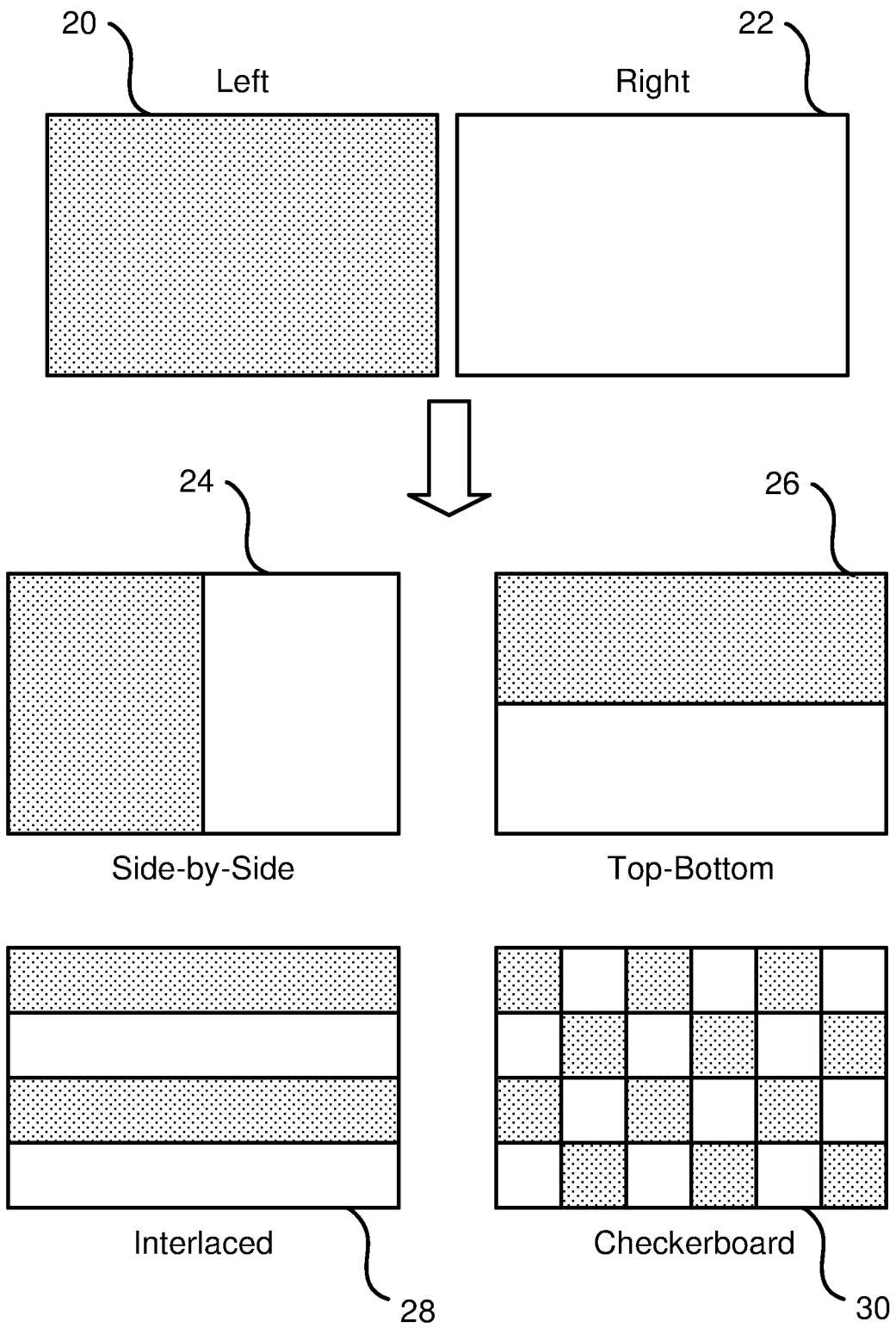


Fig. 1



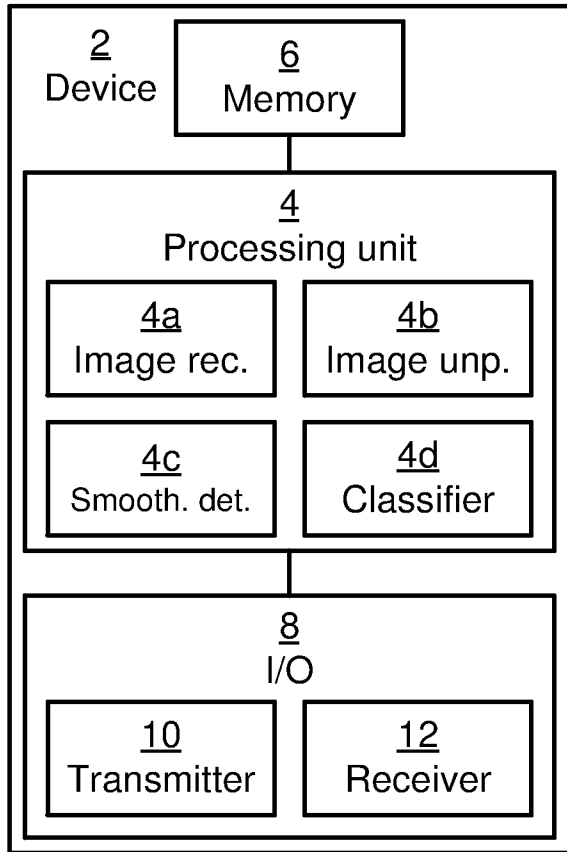
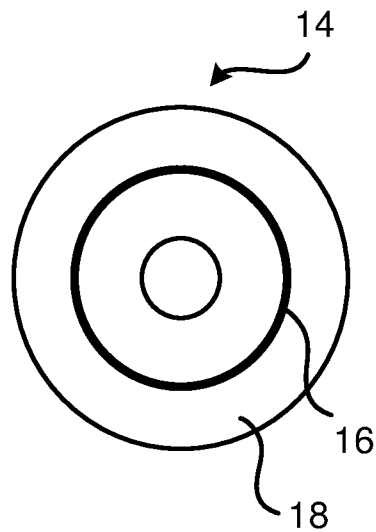


Fig. 2

Fig. 3



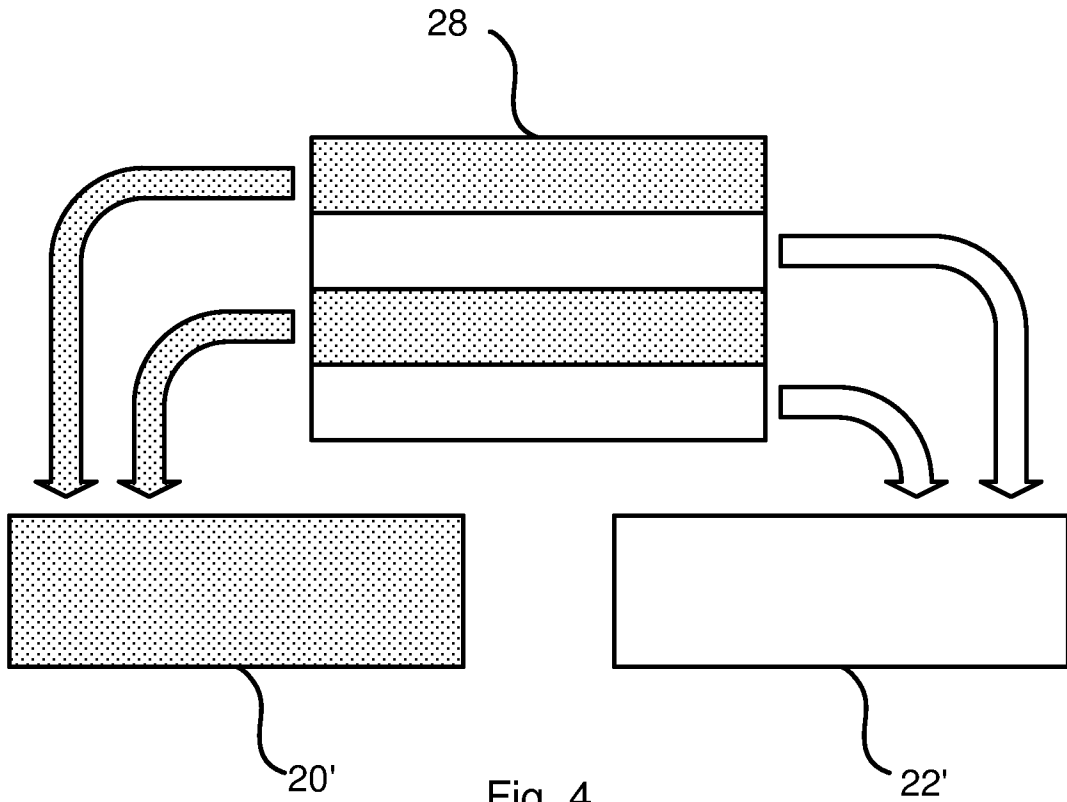


Fig. 4

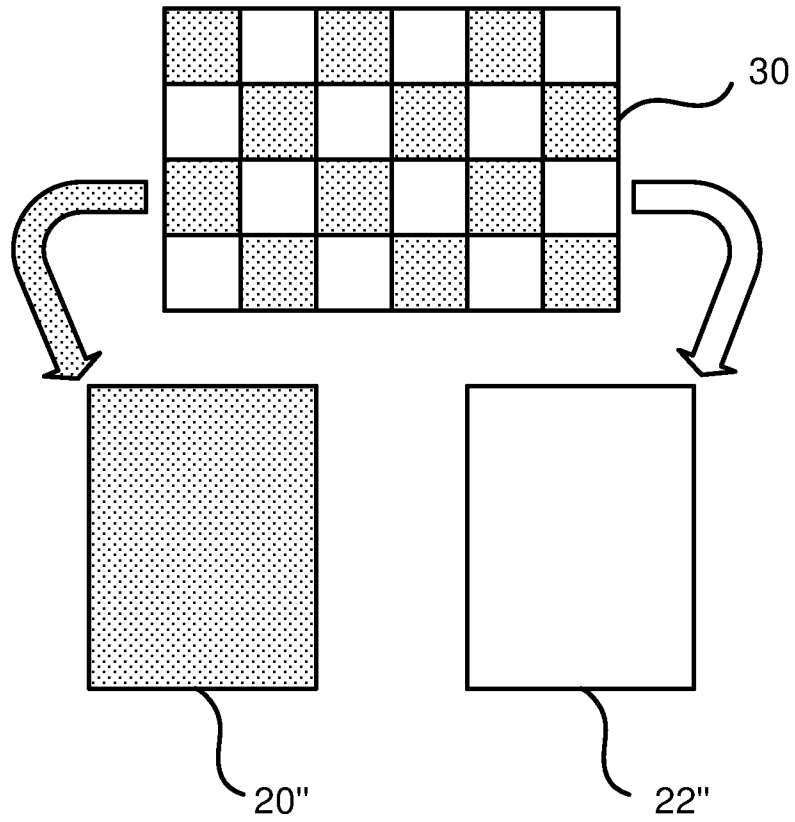


Fig. 5

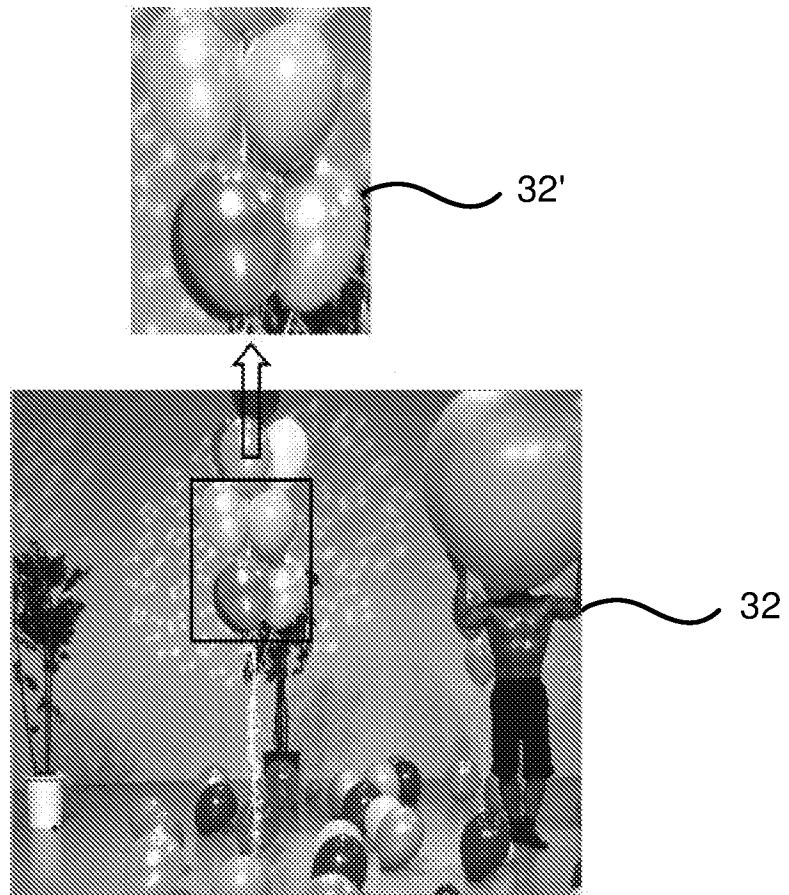
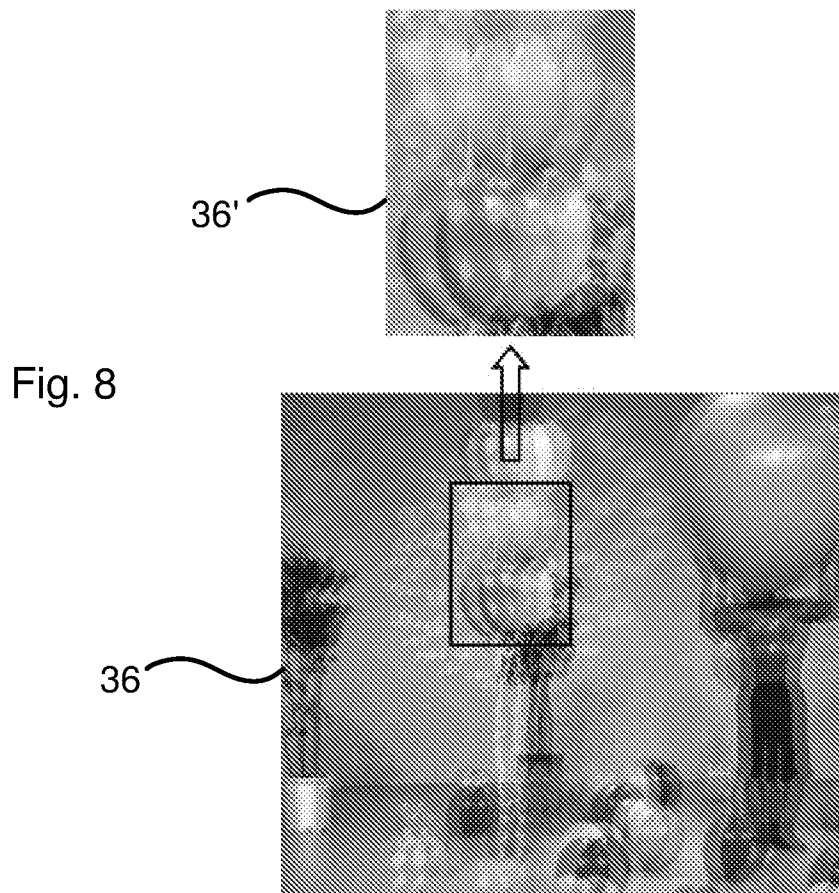
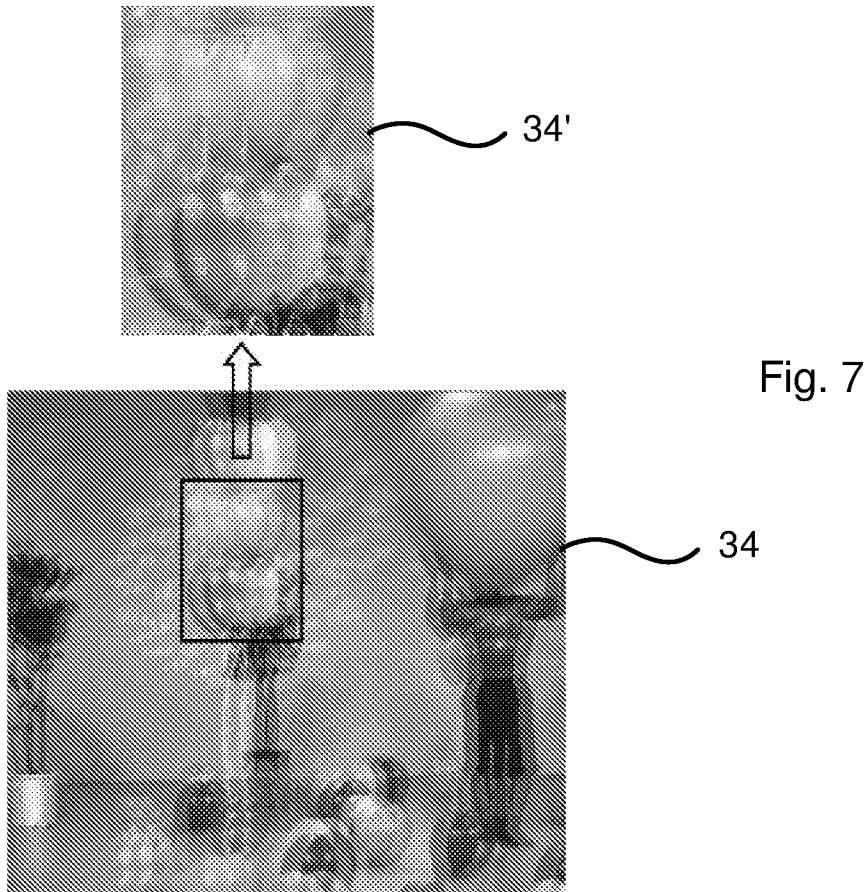


Fig. 6

5/7



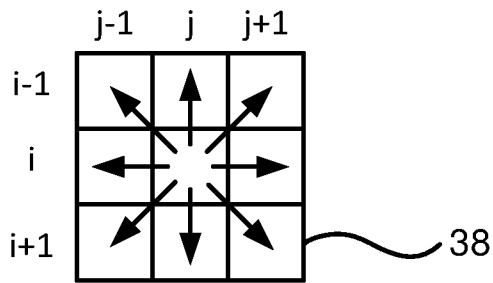


Fig. 9

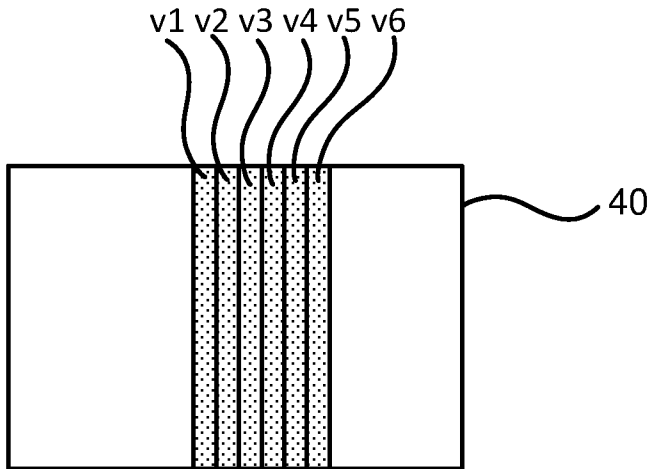


Fig. 10

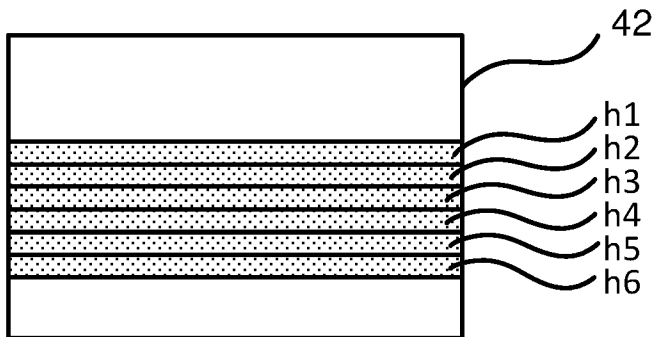


Fig. 11

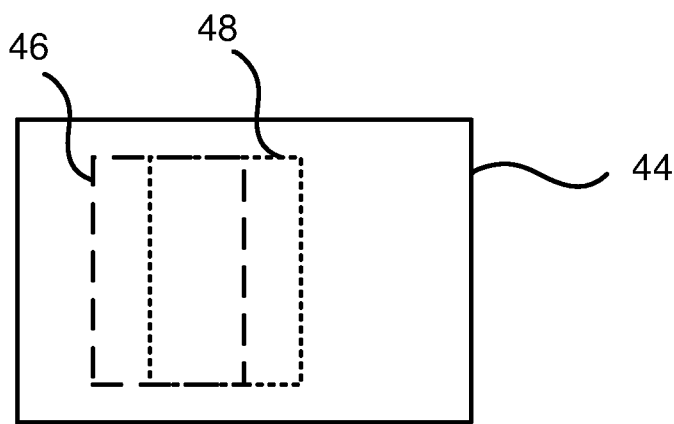


Fig. 12

←d→

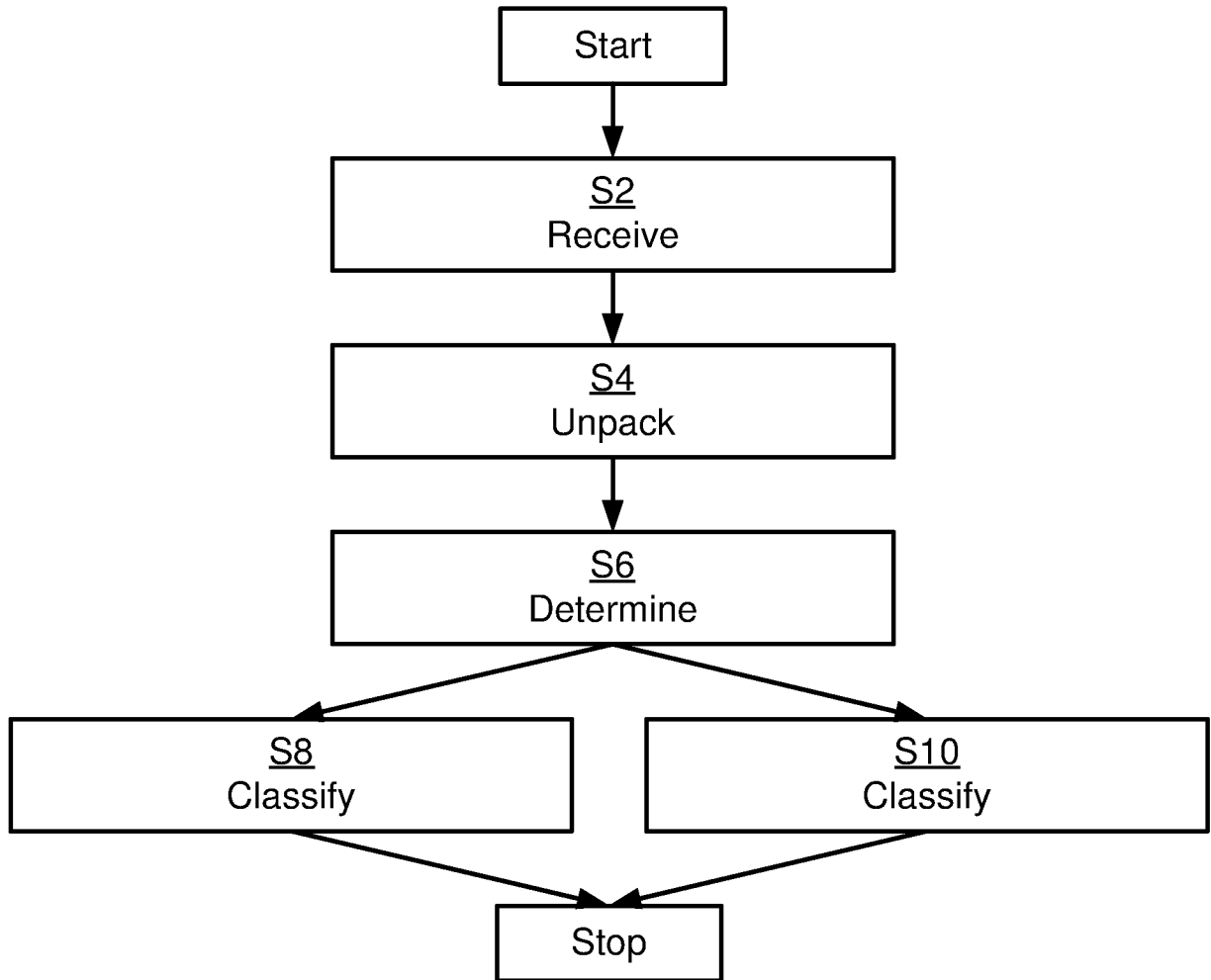


Fig. 13

**INTERNATIONAL SEARCH REPORT**

International application No  
PCT/SE2012/050865

**A. CLASSIFICATION OF SUBJECT MATTER**  
 INV. H04N13/00 H04N21/434 H04N7/32  
 ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
 EPO-Internal, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2011/164110 A1 (FORTIN ETIENNE [CA] ET AL) 7 July 2011 (2011-07-07)	1-6,11, 14-19, 21-23
Y	abstract paragraph [0006] paragraph [0013] figure 17 figure 10 paragraphs [0105] - [0108] paragraph [0046] paragraph [0151] claims 6-8  -----  -/--	7-10,12, 13,20

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search  18 April 2013	Date of mailing of the international search report  25/04/2013
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  Lauri, Lauro

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/SE2012/050865

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2010/321390 A1 (KIM YONG-TAE [KR] ET AL) 23 December 2010 (2010-12-23)	1,21,23
Y	abstract paragraph [0074] paragraph [0075] paragraph [0191]	7-10,12, 13,20
A	----- TAO ZHANG: "3D Image format identification by image difference", MULTIMEDIA AND EXPO (ICME), 2010 IEEE INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 19 July 2010 (2010-07-19), pages 1415-1420, XP031761509, ISBN: 978-1-4244-7491-2 the whole document	1-23
A	----- MIKE KNEE ET AL: "GETTING MACHINES TO WATCH 3D FOR YOU", INTERNATIONAL BROADCASTING CONFERENCE 2011; 8-9-2011 - 13-9-2011; AMSTERDAM,, 8 September 2011 (2011-09-08), XP030082329, the whole document	1-23
A	----- ANTHONY VETRO ET AL: "3D-TV Content Storage and Transmission", IEEE TRANSACTIONS ON BROADCASTING, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 57, no. 2, 1 June 2011 (2011-06-01), pages 384-394, XP011323527, ISSN: 0018-9316, DOI: 10.1109/TBC.2010.2102950 the whole document	1-23
A	----- GÖKTUG GURLER C ET AL: "Flexible Transport of 3-D Video Over Networks", PROCEEDINGS OF THE IEEE, IEEE. NEW YORK, US, vol. 99, no. 4, 1 April 2011 (2011-04-01), pages 694-707, XP011363631, ISSN: 0018-9219, DOI: 10.1109/JPROC.2010.2100010 the whole document	1-23
	----- -/--	



## INTERNATIONAL SEARCH REPORT

International application No  
PCT/SE2012/050865

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ANTHONY VETRO ET AL: "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard", PROCEEDINGS OF THE IEEE, IEEE. NEW YORK, US, vol. 99, no. 4, 1 April 2011 (2011-04-01), pages 626-642, XP011363626, ISSN: 0018-9219, DOI: 10.1109/JPROC.2010.2098830 the whole document	1-23
A	----- US 2012/113115 A1 (CUTTNER CRAIG DAVIS [US]) 10 May 2012 (2012-05-10) the whole document -----	1-23

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/SE2012/050865

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011164110	A1	07-07-2011	CN 102783158 A
			US 2011164110 A1
			WO 2011079376 A1
-----			
US 2010321390	A1	23-12-2010	CN 102461185 A
			EP 2422526 A2
			JP 2012531797 A
			KR 20100138806 A
			US 2010321390 A1
			WO 2010151049 A2
-----			
US 2012113115	A1	10-05-2012	US 2012113115 A1
			WO 2011008821 A1
-----			