



(51) International Patent Classification:
H04N 13/00 (2006.01)

(21) International Application Number:
PCT/SE2013/050728

(22) International Filing Date:
19 June 2013 (19.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: TELEFONAKTIEBOLAGET L M ERIC-
SSON (PUBL) [SE/SE]; S-164 83 Stockholm (SE).

(72) Inventors: GRAFULLA-GONZÁLEZ, Beatriz; Huvud-
stagatan 1D, S-171 44 Solna (SE). DADASH POUR, Me-

hdi; c/o Abdollahpour, Mjölnerbacken 46-2tr., S-174 48
Sundbyberg (SE).

(74) Agent: EGRELIUS, Fredrik; Ericsson AB, Patent Unit
Kista DSM, 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, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC,

[Continued on next page]

(54) Title: 3D VIDEO SWITCHING WITH GRADUAL DEPTH TRANSITION

(57) Abstract: There is provided 3D video sequence depth parameter determina-
tion. A first depth perception parameter value of a first 3D video sequence is ac-
quired. A second depth perception parameter value of a second 3D video se-
quence is acquired. An intermediate depth perception parameter value is deter-
mined based on the first depth perception parameter value and the second depth per-
ception parameter value. The intermediate depth perception parameter value is to
be used during a switch between the first 3D video sequence and the second 3D
video sequence.

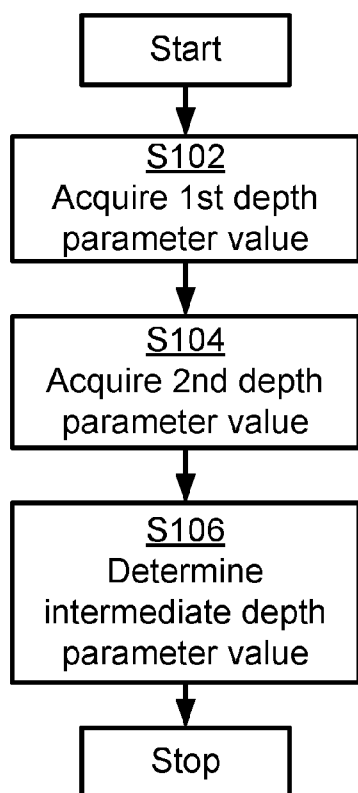


Fig. 8





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, KM, ML, MR, NE, SN, TD, TG).

Published:

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

3D VIDEO SWITCHING WITH GRADUAL DEPTH TRANSITION

TECHNICAL FIELD

Embodiments presented herein relate to video communication in general and particularly to a method, a device, a computer program, and a computer
5 program product for 3D video sequence depth parameter determination.

BACKGROUND

For video communication services, there is a challenge to obtain good performance and capacity for a given communications protocol, its parameters and the physical environment in which the video communication
10 service is deployed.

In recent years, video conferencing has become an important tool of daily life. In the business environment, it enables a more effective collaboration between remote locations as well as the reduction of travelling costs. In the private environment, video conferencing makes possible a closer, more
15 personal communication between related people. In general, although 2D video conferencing systems provide a basic feeling of closeness between participants, the user experience could still be improved by supplying a more realistic/immersive feeling to the conferees. Technically, this could be achieved, among others, with the deployment of 3D video, which adds depth
20 perception to the user visual experience and also provides a better understanding of the scene proportions.

3D video conferencing may be enabled in many different forms. To this effect, 3D equipment such as stereo cameras and 3D displays have been deployed. 3D video or 3D experience commonly refers to the possibility of, for a viewer,
25 getting the feeling of depth in the scene or, in other words, to get a feeling for the viewer to be in the scene. In technical terms, this may generally be achieved both by the type of capture equipment (i.e. the cameras) and by the type of rendering equipment (i.e. the display) that are deployed in the system.

In general terms, in multi-party 3D video conferencing the current main speaker of the 3D video conference is displayed in a larger format than the remaining participants of the 3D video conference. Once a new main speaker is identified a switch is performed so as to display the new current main
5 speaker in the larger format. The current strategy for video switching in 2D video conferencing aims at selecting the loudest client as the main speaker. In other words, the client with the highest audio level is displayed in full screen, whereas the other clients participating in the call are rendered in the thumbnails at the bottom part of the screen. The switch is performed
10 abruptly, i.e. without transition between the previous and the new main speakers. In current 3D video conferencing, the main speaker is also selected depending on the highest audio level among the connected clients. However, each transmitted 3D video sequence has different 3D properties. The user requires hence time to adapt to the new 3D rendering, which may generate
15 user discomfort, fatigue and/or eye strain.

Consequently, there is a need for an improved transition between 3D video sequences.

SUMMARY

An object of embodiments herein is to provide improved transition between
20 3D video sequences.

The inventors of the enclosed embodiments have realized that the different 3D properties of each transmitted 3D video sequence relate, for example, to depth bracket or depth perception. The inventors of the enclosed
embodiments have therefore further realized that it may occur therefore that
25 the depth bracket or the depth perception shown to the user may change abruptly and greatly when switching from one main speaker to another. The inventors of the enclosed embodiments have further realized that the 3D properties of each transmitted 3D video sequence may be matched to each other by adapting provisionally the 3D video sequences transmitted.

A particular object is therefore to provide improved transition between 3D video sequences based on adaption of the transmitted 3D video sequences.

According to a first aspect there is presented a method for 3D video sequence depth parameter determination. The method is performed by an electronic
5 device. The method comprises acquiring a first depth perception parameter value of a first 3D video sequence. The method comprises acquiring a second depth perception parameter value of a second 3D video sequence. The method comprises determining an intermediate depth perception parameter value based on the first depth perception parameter value and the second
10 depth perception parameter value. The intermediate depth perception parameter value is to be used during a switch between the first 3D video sequence and the second 3D video sequence.

Advantageously this provides improved transition between 3D video sequences. This in turn leads to a good user experience. Further
15 advantageously this enables improved 3D video switching in 3D video conference multi-party calls, thus not causing user discomfort. Fatigue and eye strain may thereby be avoided.

According to a second aspect there is presented an electronic device for 3D video sequence depth parameter determination. The electronic device
20 comprises a processing unit. The processing unit is arranged to acquire a first depth perception parameter value of a first 3D video sequence. The processing unit is arranged to acquire a second depth perception parameter value of a second 3D video sequence. The processing unit is arranged to determine an intermediate depth perception parameter value based on the
25 first depth perception parameter value and the second depth perception parameter value. The intermediate depth perception parameter value is to be used during a switch between the first 3D video sequence and the second 3D video sequence.

According to a third aspect there is presented a 3D video conference system
30 comprising at least three electronic devices according to the second aspect.

According to a fourth aspect there is presented a computer program for 3D video sequence depth parameter determination. The computer program comprises computer program code which, when run on an electronic device, causes the electronic device to perform a method according to the first aspect.

- 5 According to a fifth aspect there is presented a computer program product comprising a computer program according to the fourth aspect and a computer readable means on which the computer program is stored. The computer readable means may be non-volatile computer readable means.

- 10 It is to be noted that any feature of the first, second, third, fourth and fifth aspects may be applied to any other aspect, wherever appropriate. Likewise, any advantage of the first aspect may equally apply to the second, third, fourth, and/or fifth aspect, 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
15 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
20 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

- Embodiments of the inventive concept will now be described, by way of non-limiting examples, references being made to the accompanying drawings, in
25 which:

Fig 1 is a schematic diagram illustrating a video communications system according to an embodiment;

Fig 2a is a schematic diagram showing functional modules of an electronic device representing a video conferencing client device according to an embodiment;

Fig 2b is a schematic diagram showing functional modules of an electronic device representing a central controller according to an embodiment;

Fig 3a is a schematic diagram showing functional units of a memory according to an embodiment;

Fig 3b is a schematic diagram showing functional units of a memory according to an embodiment;

Fig 4 shows one example of a computer program product comprising computer readable means according to an embodiment;

Fig 5 is a schematic diagram illustrating a view as rendered by a 3D video sequence rendering unit according to an embodiment;

Fig 6 is a schematic diagram illustrating a parallel sensor-shifted setup according to an embodiment;

Fig 7 is a schematic diagram illustrating stereo display setup according to an embodiment; and

Figs 8 and 9 are flowcharts of methods according to embodiments.

DETAILED DESCRIPTION

The inventive concept will now be described more fully hereinafter with reference to the accompanying drawings, in which certain embodiments of the inventive concept are shown. This inventive concept may, however, be embodied in many different 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 inventive concept to those skilled in the art. Like numbers refer to like elements throughout the description.

Fig 1 is a schematic diagram illustrating a video communications system 1a where embodiments presented herein can be applied. The video communications system 1a comprises a number of electronic devices 2a, 2b, 2c representing video conferencing client devices. The electronic devices 2a, 2b, 2c are operatively connected via a communications network 8. The communications network 8 may comprise an electronic device 9 representing a central controller. The central controller may be arranged to control the communications between the video conferencing client devices.

Each electronic device 2a, 2b, 2c representing a video conferencing client device comprises, or is operatively connected to, a 3D video sequence capturing unit 6 (i.e. one or more cameras) and/or a 3D video sequence rendering unit 7 (i.e. a unit, such as a display, for rendering received video sequences) that require different video formats and codecs. As the skilled person understands this are just one example of a video communications system where the disclosed embodiments can be applied. Thus, although only three electronic devices 2a, 2b, 2c representing video conferencing client devices are illustrated in Fig 1, there may in practical situations be a large combination of electronic devices 2a, 2b, 2c representing video conferencing client devices with different 2D/3D equipment.

Fig 5 schematically illustrates a view 51 as rendered by the 3D video sequence rendering unit 7. In the center of the view the current main speaker of the 3D video conference is displayed at 52. Along a bottom border of the view the remaining participants of the 3D video conference are displayed as thumbnails at 53 and 54. Each participant of the 3D video conference corresponds to one video conferencing client device. In the upper right-hand corner a thumbnail 55 of the user of the 3D video sequence rendering unit 7 is shown. Along a left border settings for the video conferencing client device associated with the 3D video sequence rendering unit 7 are displayed at 56. The settings may be related to capturing parameters, microphone level, speaker level, screen adjustments, video recording mode, video quality mode, etc. As the skilled person understands this is just one illustrative example of a view; it is anticipated and within the scope of the herein presented

embodiments that the view may have another appearance as long as the principles of the herein disclosed embodiments apply. As noted above, once detected that a new participant is to be selected as the new main speaker, the view is changed such that the current main speaker is replaced by the new
5 main speaker.

The central controller may be arranged to only route/switch received video sequences. In this case, the video conferencing client devices transmit multiple video sequences with different resolutions, e.g. a high-quality video sequence for the main speaker case and low-quality video sequences for the
10 thumbnails cases. The central controller then decides which video sequence is sent to which video conferencing client device, depending on the main speaker and the video conferencing client device itself. The central controller may alternatively be arranged to transcode and/or re-scales received video sequences. In this case, the video conferencing client devices only transmit
15 one high-quality video sequence which is processed by the central controller depending on whether the video sequence represents the main speaker or a thumbnail. Then, the central controller transmits the correct video sequence resolution to the each video conferencing client device. The central controller may yet alternatively be arranged to mix the video sequences. In this case, the
20 central controller decodes the received video sequences and composes the rendering scene depending on the main speaker and thumbnails. This implies that video sequences are transcoded and/or re-scaled. Then, the central controller transmits the composed video sequences to the video conferencing client devices, which only have to render the received video
25 sequence.

The process of changing the view from one current main speaker to a new main speaker is referred to a switching. As further noted above, each 3D video sequence rendering unit 7 may be associated with its own 3D properties, thus resulting in each transmitted 3D video sequence also to be
30 associated with its own 3D properties. Examples of such 3D properties are the depth bracket and the depth perception shown to the user. One object of the embodiments disclosed herein is to provide mechanisms for 3D video

switching in multi-party calls that do not generate user discomfort by enabling a smooth transition from 3D properties of the current main speaker to 3D properties of the new main speaker.

More particularly, the embodiments disclosed herein relate to 3D video
5 sequence depth parameter determination. In order to obtain 3D video sequence depth parameter determination there is provided an electronic device, a method performed by the electronic device, a computer program comprising code, for example in the form of a computer program product, that when run on an electronic device, causes the electronic device to perform
10 the method.

Fig 2a schematically illustrates, in terms of functional modules, an electronic device 2 representing a video conferencing client device. The electronic device 2 may be part of a stationary computer, a laptop computer, a tablet computer, or a mobile phone. A processing unit 3 is provided using any
15 combination of one or more of a suitable central processing unit (CPU), multiprocessor, microcontroller, digital signal processor (DSP), application specific integrated circuit (ASIC) etc., capable of executing software instructions stored in a computer program product 13 (as in Fig 4). Thus the processing unit 3 is thereby preferably arranged to execute methods as herein
20 disclosed. The electronic device 2 further comprises an input/output (I/O) interface in the form of a transmitter (TX) 4 and a receiver (RX) 5, for communicating with other electronic devices over the communications network 8, with a capturing unit 6 and a display unit 7. Other components, as well as the related functionality, of the electronic device 2 are omitted in
25 order not to obscure the concepts presented herein.

Fig 2b schematically illustrates, in terms of functional modules, an electronic device 9 representing a central controller. The electronic device 9 is preferably part of a network server functioning as a media resource function processor (MRFP), but may also be part of a stationary computer, a laptop
30 computer, a tablet computer, or a mobile phone acting as a host for a 3D video communication service. A processing unit 10 is provided using any

combination of one or more of a suitable central processing unit (CPU), multiprocessor, microcontroller, digital signal processor (DSP), application specific integrated circuit (ASIC) etc., capable of executing software instructions stored in a computer program product 13 (as in Fig 4), Thus the
5 processing unit 10 is thereby preferably arranged to execute methods as herein disclosed. The central device 9 further comprises an input/output (I/O) interface in the form of a transmitter (TX) 11 and a receiver (RX) 12, for communicating with electronic devices 2a, 2b, 2c representing video conferencing client devices over the communications network 8. Other
10 components, as well as the related functionality, of the electronic device 9 are omitted in order not to obscure the concepts presented herein.

Fig 3a schematically illustrates functional units of the memory 4 of the electronic device 2; an acquiring unit 4a, a determining unit 4b, a switching unit 4c, a generating unit 4d, an adjusting unit 4e, an initiating unit 4f, and a
15 checking unit 4g. The functionality of each functional unit 4a-f will be further disclosed. In general terms, each functional unit 4a-f may be implemented in hardware or in software. The processing unit 3 may thus be arranged to from the memory 4 fetch instructions as provided by a functional unit 4a-g and to execute these instructions.

20 Fig 3b schematically illustrates functional units of the memory 11 of the electronic device 9; an acquiring unit 11a, a determining unit 11b, a switching unit 11c, a generating unit 11d, an adjusting unit 11e, an initiating unit 11f, and a checking unit 11g. The functionality of each functional unit 11a-f will be further disclosed. In general terms, each functional unit 11a-f may be
25 implemented in hardware or in software. The processing unit 10 may thus be arranged to from the memory 11 fetch instructions as provided by a functional unit 11a-g and to execute these instructions.

Figs 8 and 9 are flowcharts illustrating embodiments of methods for 3D video sequence depth parameter determination. The methods are performed by an
30 electronic device 2, 9 representing a video conferencing client device (as in Fig 2) or a central controller (as in Fig 3). The methods are advantageously

provided as computer programs 14. Fig 4 shows one example of a computer program product 13 comprising computer readable means 15. On this computer readable means 15, a computer program 14 can be stored. This computer program 14 can cause the processing unit 3 of the electronic device 2 and thereto operatively coupled entities and devices to execute methods according to embodiments described herein. The computer program 14 can alternatively or additionally cause the processing unit 10 of the electronic device 9 and thereto operatively coupled entities and devices to execute methods according to embodiments described herein. The computer program 14 and/or computer program product 13 thus provides means for performing any steps as herein disclosed.

In the example of Fig 4, the computer program product 13 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 13 could also be embodied as a memory (RAM, ROM, EPROM, EEPROM) and more 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 14 is here schematically shown as a track on the depicted optical disk, the computer program 14 can be stored in any way which is suitable for the computer program product 13.

For the capturing side, it is assumed that the capturing units 6 are configured with the so-called parallel sensor-shifted setup, as illustrated in Fig 6. Other configurations, such as the so-called toed-in setup, are possible too, although extra processing would be required to align left and right views, in general terms yielding a worse stereoscopic quality.

In Fig 6, f denotes the capturing unit's camera focal length, t_c is the baseline distance (or the distance between the camera optical centers, and Z_c is the distance to the convergence plane or the convergence distance. In the parallel sensor-shifted setup, the convergence of cameras is established by a small shift ($h/2$) of the sensor targets. Suppose the captured object is on the distance (i.e. depth) Z from the cameras. The distance between the image

points in the left and the right images that refer to the same captured point is called the disparity d .

The parameters mentioned above are mathematically related and it is not difficult for a skilled person to derive the following expression that connects
5 them:

$$d = h - \frac{t_c f}{Z} = t_c f \left(\frac{1}{Z_c} - \frac{1}{Z} \right) \quad (1)$$

Objects captured at $Z = Z_c$ have zero disparity, which further yields:

$$h = \frac{t_c f}{Z_c} \quad (2)$$

In a similar way, objects captured at $Z < Z_c$ have negative disparity, and
10 objects captured at $Z > Z_c$ have a positive disparity.

There is hence not a single disparity for the whole stereo pair (i.e. the corresponding left and right views), but rather a set of disparities. Indeed, disparity is the distance between the image points in the left and right images that refer to the same captured point. Hence there will be as many disparities
15 as matched points between the views.

3D displays (as part of the rendering unit) create the feeling of depth by showing simultaneously two slightly different images for the left and the right eye. One parameter that controls the depth perception is the so-called screen parallax P , which reflects the spatial distance between the points in the left
20 and right views on the screen. In general terms, the depth perception, among other parameters, depends on the amount and type of parallax. The so-called positive parallax means that the point in the right-eye view lays more right than the corresponding point in the left-eye view. Zero parallax means that the points lay at the same position, while negative parallax means that the
25 point in the right-eye view lays more left than the corresponding point in the left-eye view. With positive parallax the objects are perceived in the so-called

screen space, whereas with zero and negative parallax they are perceived on and in front of the screen space (viewer space) respectively.

- Suppose, without limitation, that the distance between the viewer's eyes is t_e (the so-called inter-ocular distance) and that the viewer sits at a distance Z_D from the screen, as schematically illustrated in Fig 7. A simple geometric study yields the following expression for the perceived depth:

$$Z_p = \frac{Z_D \cdot t_e}{t_e - P} \quad (3)$$

- From this equation it follows that objects with a positive parallax are perceived to be in the screen space ($Z_p > Z_D$), objects with zero parallax exactly on the screen surface ($Z_p = Z_D$), and objects with negative parallax in the viewer space ($Z_p < Z_D$).

A 3D display is characterized with a parallax range $[P_{DB \min}, P_{DB \max}]$ for which 3D viewing is comfortable for a user and which indeed defines the depth budget.

- The maximum value of parallax that human eyes can handle without diverging is equal to the inter-ocular distance, i.e. $P_{DB \max} = t_e$. This is, however, a border case, which usually does not hold in real stereo setups, where the furthest objects are usually placed at some distance comfortable for the viewers.
- The minimum value of the parallax can be approximated by:

- $P_{DB \min} = t_e - Z_D \cdot \Delta\alpha_{total}$, where $\Delta\alpha_{total}$ is the total convergence angle that itself is the sum of the two convergence ranges – one for the viewer space in front of the display and one for the screen space behind the display. An established rule of thumb is to set $\Delta\alpha_{total}$ to 0.02 rad. Although conservative from the current knowledge point of view, this bound yields a safe estimate. A screen may have other recommended values for $P_{DB \min}$. Indeed, another

recommendation could be to limit the depth budget to $1/30$ of the display width to avoid stereoscopic problems.

Screen parallax values that are outside the recommended parallax range may be tolerated for short periods of time, but they are not recommended for
5 extended viewings as they would lead to discomfort and fatigue.

When rendering 3D video, there is not one single parallax for a rendered stereo pair, but rather a set of parallaxes (as for the disparities in the production side). The value range of achieved parallaxes defines the depth bracket.

10 A method for 3D video sequence depth parameter determination is based on determining an intermediate depth perception parameter value between a first 3D video sequence and a second 3D video sequence. The first 3D video sequence may have been captured by a first video conferencing client device and the second 3D video sequence may have been captured by a second video
15 conferencing client device. The method is performed by an electronic device 2, 2a, 2b, 2c, 9. The processing unit 3, 10 of the electronic device 2, 2a, 2b, 2c, 9 is arranged to, in a step S102, acquire a first depth perception parameter value of a first 3D video sequence. These instructions may be provided by the acquiring units 4a, 11a. Hence the acquiring units 4a, 11a may be configured
20 to acquire the first depth perception parameter value. The computer program 14 and/or computer program product 13 may thus comprise means for performing instructions according to step S102.. The processing unit 3, 10 of the electronic device 2, 2a, 2b, 2c, 9 is arranged to, in a step S104, acquire a second depth perception parameter value of a second 3D video sequence.
25 These instructions may be provided by the acquiring units 4a, 11a. Hence the acquiring units 4a, 11a may be configured to acquire the second depth perception parameter value. The computer program 14 and/or computer program product 13 may thus comprise means for performing instructions according to step S104. Based on the first depth perception parameter value
30 and the second depth perception parameter value the processing unit 3, 10 of the electronic device 2, 2a, 2b, 2c, 9 is arranged to, in a step S106, determine

an intermediate depth perception parameter value. These instructions may be provided by the determining units 4b, 11b. Hence the determining units 4b, 11b may be configured to determine the intermediate depth perception parameter value. The computer program 14 and/or computer program
5 product 13 may thus comprise means for performing instructions according to step S106. The intermediate depth perception parameter value is to be used during a switch between the first 3D video sequence and the second 3D video sequence.

According to one embodiment the steps as herein disclosed are performed in
10 real-time. Hence, the herein disclosed mechanisms for 3D video sequence depth parameter determination are readily applicable in 3D video conferencing systems.

By means of the determined intermediate depth perception parameter value a smooth transition between the first 3D video sequence and the second 3D
15 video sequence is thereby enabled. Hence, the switching from the first 3D video sequence to the second 3D video sequence may thereby be performed without causing user discomfort. Fatigue and eye strain may thereby be avoided. According to one embodiment the processing unit 3, 10 of the electronic device 2, 2a, 2b, 2c, 9 is further arranged to, in an optional step
20 S112, switch between the first 3D video sequence and the second 3D video sequence. Hence the electronic device 2, 2a, 2b, 2c, 9 may be arranged to perform the actual switch. These instructions may be provided by the switching units 4c, 11c. Hence the switching units 4c, 11c may be configured to switch between the first 3D video sequence and the second 3D video
25 sequence. The computer program 14 and/or computer program product 13 may thus comprise means for performing instructions according to step S112.

As noted above, the switch may be performed once a new main speaker has been determined. In general terms the electronic device may be arranged to, in an optional step S108, receive an indication for switching between the first
30 3D video sequence and the second 3D video sequence. The switching may then be performed based on the indication. The electronic device may

therefore be arranged to, in an optional step S110, initiate switching between the first 3D video sequence and the second 3D video sequence based on the indication. These instructions may be provided by the initiating units 4f, 11f. Hence the initiating units 4f, 11f may be configured to initiate the switching.

- 5 The computer program 14 and/or computer program product 13 may thus comprise means for performing instructions according to step S108.

According to one embodiment the electronic device comprises at least one of a 3D video sequence capturing unit 6 arranged to capture the first and/or second 3D image video sequence, and a 3D video sequence rendering unit 7
10 arranged to render the first and/or second 3D image video sequence. The electronic device may further comprise a communications interface 12 arranged to receive the first and/or second 3D image video sequence from a 3D video sequence capturing unit device 6, and to transmit the first and/or second 3D image video sequence to a 3D video sequence rendering unit
15 device 7. As noted above, the electronic device 2 may represent a video conferencing client device. The electronic device may thus either be located at the capturing side or the rendering side. As also noted above, the electronic device 9 may alternatively represent a central controller. According to one embodiment the electronic device is thus located in the communications
20 network 8. Thus, depending on its role the electronic device 2, 9 may be arranged to, in an optional step S114, transmit a rendition of the first 3D video sequence, the intermediate 3D video sequence, and/or the second 3D video sequence to at least one of a rendering device 7 and a control device. In a case the electronic device 2 represents a video conferencing client device
25 the renditions of the 3D video sequences may be transmitted to the control device, such as the central controller. In a case the electronic device 9 represents a central controller the renditions of the 3D video sequences may be transmitted to a video conferencing client device.

There may be different ways to perform the switching in step S112. According
30 to one embodiment the processing unit 3, 10 of the electronic device 2, 2a, 2b, 2c, 9 is further arranged to, in an optional step S112a, use the intermediate depth perception parameter value to generate an intermediate 3D video

sequence. These instructions may be provided by the generating units 4d 11d. Hence the generating units 4d 11d may be configured to generate the intermediate 3D video sequence. The computer program 14 and/or computer program product 13 may thus comprise means for performing instructions
5 according to step S112a. The intermediate 3D video sequence may be rendered between the first 3D video sequence and the second 3D video sequence. The intermediate 3D video sequence may thus be regarded as a means for adapting the first 3D video sequence to the second 3D video sequence. Particularly, according to the present embodiment the processing
10 unit 3, 10 of the electronic device 2, 2a, 2b, 2c, 9 is further arranged to, in an optional step S112b, switch from the first 3D video sequence to the intermediate 3D video sequence, and, in an optional step S112d, switch from the intermediate 3D video sequence to the second 3D video sequence. These instructions may be provided by the switching units 4c, 11c. Hence the
15 switching units 4c, 11c may be configured to switch from the first 3D video sequence to the intermediate 3D video sequence, and to switch from the intermediate 3D video sequence to the second 3D video sequence. The computer program 14 and/or computer program product 13 may thus comprise means for performing instructions according to step S112d.

20 Further details of the switching may relate to how the intermediate 3D video sequence is related to the first 3D video sequence and the second 3D video sequence, respectively, and vice versa. Further details of the switching may further relate to how the intermediate depth perception parameter can be used to relate the first 3D video sequence to the second 3D video sequence.
25 According to one embodiment the processing unit 3, 10 of the electronic device 2, 2a, 2b, 2c, 9 is further arranged to, in an optional step S112c, adjust the depth perception parameter value of the first 3D video sequence from the first depth perception parameter value to the intermediate depth perception parameter value while displaying the first 3D video sequence. These
30 instructions may be provided by the adjusting units 4e, 11e. Hence the adjusting units 4e, 11e may be configured to adjust the depth perception parameter value of the first 3D video sequence in this way. The computer

program 14 and/or computer program product 13 may thus comprise means for performing instructions according to step S112c. Hence, this may enable a gradual change of the depth perception parameter value of the displayed first 3D video sequence from the depth perception parameter value of the first 3D video sequence to the intermediate depth perception parameter value. The processing unit 3, 10 of the electronic device 2, 2a, 2b, 2c, 9 is then further arranged to, in an optional step S112e, adjust the depth perception parameter value of the second 3D video sequence from the intermediate depth perception parameter value to the second depth perception parameter value while displaying the second 3D video sequence. These instructions may be provided by the adjusting units 4e, 11e. Hence the adjusting units 4e, 11e may be configured to adjust the depth perception parameter value of the second 3D video sequence in this way. The computer program 14 and/or computer program product 13 may thus comprise means for performing instructions according to step S112e. Hence, this may enable a gradual change of the depth perception parameter value of the displayed second 3D video sequence from the intermediate depth perception parameter value to the depth perception parameter value of the second 3D video sequence. This may enable a particularly smooth transition between the first 3D video sequence and the second 3D video sequence since the intermediate depth perception parameter is used for displaying both the first 3D video sequence and the second 3D video sequence.

As the skilled person understands, steps S112a, S112b, S112c, S112d, and S112e may be combined into one embodiment, where the steps are performed in the thus formed sequence. According to this combined embodiment the intermediate 3D video sequence thus has two parts; one based on the first 3D video sequence and one based on the second 3D video sequence.

Two overall embodiments related to the depth perception parameter value will now be disclosed.

In a first overall embodiment, the switch is performed according to the depth bracket range for each 3D video sequence. Thus, according to the first overall

embodiment the depth perception parameter value is based on a depth bracket range or a produced disparity range. That is, according to the first overall embodiment the depth perception parameter value may be based on parameters relating to the capturing side of the 3D video sequence. In brief, the first overall embodiment is based on the first 3D video sequence of the current main speaker having its depth bracket range reduced until a certain value (as given by the intermediate depth perception parameter value). Then the video switch is performed (for example as disclosed above) to the next main speaker stream (i.e., the second 3D video sequence) which starts at the same depth bracket range value. Finally, the second 3D video sequence has its depth bracket range adapted to its normal value (as given by the second depth perception parameter value).

In a second overall embodiment, the switch is performed based on the perceived depth. Thus, according to the second overall embodiment the depth perception parameter value is based on a perceived depth or parallax. That is, according to the first overall embodiment the depth perception parameter value may be based on parameters relating to the rendering side of the 3D video sequence. In brief, the second overall embodiment is based on the 3D video sequence for the current main speaker having its perceived depth reduced up to a certain value (as given by the intermediate depth perception parameter value). Then the video switch is performed (for example as disclosed above) to the next main speaker stream (i.e., the second 3D video sequence) which starts at the same perceived depth value. Finally, the second 3D video sequence has its perceived depth adapted to its normal value (as given by the second depth perception parameter value).

Particular details of the first overall embodiment and the second overall embodiment will now be disclosed in further detail.

First overall embodiment

The first overall embodiment is based on modification of the depth bracket range during the 3D video switching. In general terms, the depth bracket is

defined as the set of achieved parallaxes at the rendering side. Therefore, to modify the depth bracket range, the produced parallaxes at the rendering side needs to be changed. The shift between the left and the right views does not modify the depth bracket range, which is kept constant, but just shifts it.

- 5 In general terms, 3D video capturing and rendering sides are linked by a magnification factor S_M :

$$S_M = \frac{P}{d} = \frac{W_D}{W_S}, \quad (4)$$

- 10 where d is the disparity the capturing side, P is the screen parallax on the rendering side, where W_D is the rendering screen width, and where W_S is the capturing sensor width. By introducing S_M , the capturing side and rendering side geometries can be combined. In this sense the depth bracket may be translated into the capturing side, where the produced disparity range is defined.

- 15 In this sense, by adjusting disparities at the capturing side, parallaxes at the rendering side may be adjusted, and hence the depth bracket range may be adjusted. From Equation (1) follows that disparities may be modified by varying the sensor shift h , the baseline t_c or the focal length f . Since the object range or depth Z depends on the scene, it is not consider it as a parameter that can be used to adjust the depth bracket range. It is easy for a skilled person to derive that the effect of the sensor shift h is the same as for the parallax shift. In other words, disparities are changed (or shifted), but the depth bracket range is kept constant. Conversely, the baseline t_c and the focal length f have the ability to change the depth bracket range and therefore can be used to adjust the above disclosed 3D video switch.

- 25 Consider Z in the range $[Z_{\min}; Z_{\max}]$, where $Z_{\min} \geq 0$ and $Z_{\max} \rightarrow +\infty$ (although in practice $Z_{\max}$ will have a finite value). The depth bracket range DB is defined as:

$$DB = d_{\max} - d_{\min} , \quad (5)$$

where $d_{\min}$ indicates the minimum value of the disparity set for the given Z range and $d_{\max}$ indicates the maximum value of the disparity set for the given Z range. It is easy for a skilled person to derive that:

$$5 \quad d_{\min} = h - \frac{t_c f}{Z_{\min}} \quad \text{and} \quad d_{\max} = h - \frac{t_c f}{Z_{\max}} , \quad (6)$$

which yields that:

$$DB = h - \frac{t_c f}{Z_{\max}} - h + \frac{t_c f}{Z_{\min}} = t_c f \left(\frac{1}{Z_{\min}} - \frac{1}{Z_{\max}} \right) \quad (7)$$

It is also easy for a skilled person to deduce that the depth bracket range will be increased when either the baseline t_c and/or the focal length f are increased; whereas the depth bracket range will be decreased when either the
 10 baseline t_c and/or the focal length f are decreased.

Therefore, when in this disclosure referring to an increase or a decrease of the depth bracket range, it is referred to the modification of one of these parameters at the capturing side.

15 Although both the baseline t_c and the focal length f may be deployed to change the depth bracket range value, the baseline may typically be preferred. This is mainly because the modification of the focal length may also change substantially the captured field of view. Although the baseline may also change the field of view, the effect is less strong than with the focal
 20 length.

In general terms, there is at least one different way of changing the depth bracket range at the rendering side (although perhaps not preferred since the resulting video quality is not as good, as discovered by the inventors of the enclosed embodiments). Instead of requiring from the capturing side a new

view at a different baseline t_c , one can synthesize this new view at this particular baseline in the rendering side, provided that enough information is available (e.g. a depth map).

The first overall embodiment may be divided into four main parts: an initial
5 phase where the perceived depth ranges are calculated and stored for subsequent switches; a 3D video switch; on-call modifications; and disconnection.

Note that these four parts may be performed sequentially (as they are going to be explained), but they can also occur in parallel. For example, if three
10 video conferencing client devices are connected, there may be already 3D video switching being performed between the three video conferencing client devices before a fourth video conferencing client device joins the multi-party 3D video conference. Likewise, depth bracket range modifications may be determined before other video conferencing client devices are connected. The
15 division in these three main parts is hence given purely for descriptive purposes.

Initial phase

According to one embodiment within the first overall embodiment the initial phase comprises the following.

- 20 The first video conferencing client device requests a connection to a multi-party video conference. The connection request is sent by the communications interface 5 of the electronic devices 2a representing the first video conferencing client device. The connection request may be sent to an electronic device 9 representing a central controller, such as an MRFP.
- 25 The first client and the central controller negotiate connection properties, such as audio and video codecs, for example through SIP/SDP (Session Initiation Protocol / Session Description Protocol) negotiation or according to other protocols, such as H.323. During the negotiation, the first client also signals its capturing capabilities, namely its depth bracket range or produced

disparity range (it does not matter which parameter is signaled, provided that all the remaining video conferencing client device provide the same parameter). If the first client is arranged only to capture 2D video through its capturing device 6, the depth bracket range will be zero.

- 5 The central controller also assigns the first client with a unique ID, e.g. "client 1". The central controller stores the client ID in the memory 11 and the value of the depth bracket range into a local or remote data base to which it will access the data when required.

- The first client starts transmitting the captured video to the central
- 10 controller. Thus, a first 3D video sequence is captured by the capturing unit 6 of the electronic devices 2a representing the first video conferencing client device 2a and the transmitted by its communications interface 5. The first 3D video sequence is then received through the communications interface 12 of the electronic devices 9 representing the central controller. At this point it is,
15 however, assumed that no other clients are connected to the 3D video conference. The central controller may thus disregard the received first 3D video sequence and not transmit any 3D video sequences to the clients. For example, the processing unit 3 may be arranged to on the rendering unit 7 display a message indicating the lack of other connected clients.
 - 20 Then another client (hereinafter a second client, as represented by a second video conferencing client device 2b) requests a connection to the central controller.

- The same SIP/SDP negotiation (or any other negotiation protocol) takes place between the central controller and the second client. That is, the second
- 25 client also signals its capturing capabilities, namely its depth bracket range or produced disparity range.

The central controller also assigns the second client with a unique ID, e.g. "client 2". The central controller updates the data base with the new client ID and the value of the depth bracket range.

3D video communication between the first client and the second client may then start. At this point only two clients are connected to the call, implying that no video switching may yet be required.

Then another client (hereinafter a third client, as represented by a third video
5 conferencing client device 2c) requests a connection to the central controller.

The same SIP/SDP negotiation (or any other negotiation protocol) takes place between the central controller and the third client. That is, the third client also signals its capturing capabilities, namely its depth bracket range or produced disparity range.

10 The central controller also assigns the third client with a unique ID, e.g. "client 3". The central controller updates the data base with the new client ID and the value of the depth bracket range.

As the skilled person understands, also at least one further client may join the 3D video conference according to the steps as outlined above.

15 At this point, three or more clients are connected to the 3D video conference, and therefore a 3D video switching may be required, e.g., depending on the audio level of each client. Further details of the switching are provided next.

3D video switching

In general terms, the central controller may be arranged to handle the main
20 speaker switching between the clients based, e.g., on the audio level. When performing the switching in 3D video conferencing, a number of different scenarios may be considered depending on the video sequences transmitted:

2D-to-2D video switch: This first scenario corresponds to current video conferencing systems and an explanation thereof is therefore omitted.

25 *2D-to-3D video switch:* This second scenario deals with switching from a 2D video sequence to a 3D video sequence. In other words, the user experience corresponds to transition from a scene without depth perception (or flat) to a

scene with depth perception. In order to make the switch smooth, the 3D video sequence could be adapted so that its depth bracket range is zero at the switching moment. This implies that, when the central controller determines that the main speaker video sequence needs to be changed, the central
5 controller requests the client device transmitting the 3D video sequence to adapt its 3D video sequence so that its depth bracket range is zero. Thus, in this second scenario the intermediate depth perception parameter value corresponds to a depth bracket range of zero. Then, once the switch is carried out, the client device transmitting the 3D video sequence increases its depth
10 bracket range progressively up to its optimized value (i.e. the value stored at the central controller data base).

3D-to-2D video switch: This third scenario deals with switching from a 3D video sequence to a 2D video sequence. In other words, the user experience corresponds to transition from a scene with depth perception to a scene
15 without depth perception (or flat). When the central controller determines that the main speaker video sequence needs to be changed, the central controller requests the client transmitting the 3D video to progressively adapt its 3D video sequence so that its depth bracket range becomes zero. Thus, in this third scenario the intermediate depth perception parameter value
20 corresponds to a depth bracket range of zero. Once the depth bracket range is zero (i.e. the video sequence is now 2D), the switch to the 2D video is performed.

3D-to-3D video switch: This fourth scenario deals with switching from a first 3D video sequence to a second 3D video sequence. In other words, the user
25 experience corresponds to transition of a scene with a first depth perception to a scene with a second, different, depth perception. For this scenario, there are different possibilities on how the transition may be performed. Particular embodiments (each within the first overall embodiment) will be disclosed next. According to one embodiment the intermediate depth perception
30 parameter value takes a predetermined value. For example, the predetermined value may correspond to a more narrow depth bracket than

the depth bracket of the first 3D video sequence and the depth bracket of the second 3D video sequence.

According to a first embodiment, both 3D video sequences meet at a depth bracket range (as determined by the intermediate depth perception parameter value) that is comfortable for both clients. For example, consider switching from a first client that transmits a first 3D video sequence with a depth bracket range of 60 to a second client that transmits a second 3D video sequence with a depth bracket range of 30. The first client is requested to progressively reduce its depth bracket range to e.g. 20 (as determined by the intermediate depth perception parameter value). Then, the switch is carried out between the first 3D video sequence and the second 3D video sequence at the depth bracket of 20 (as the second client has also been requested to change its depth bracket range to 20). Finally, the second client is requested to increase its depth bracket range progressively to its optimized value of 30 (i.e. the one stored at the central controller data base). Hence, this first embodiment follows the steps S112 and S112a-d as outlined above.

According to a second embodiment, both 3D video sequences meet at an intermediate depth bracket range, which may or may not be comfortable for both clients. For example, consider switching from a first client that transmits a first 3D video sequence with a depth bracket range of 60 to a second client that transmits a second 3D video sequence with a depth bracket range of 30. The first client is requested to progressively reduce its depth bracket range to e.g. 45 (as determined by the intermediate depth perception parameter value). Thus, according to one embodiment the intermediate depth perception parameter value is a mean value of the first depth perception parameter value and the second depth perception parameter value. Then, the switch is carried out between the first 3D video sequence and the second 3D video sequence at the depth bracket of 45 (as the second client has also been requested to change its depth bracket range to 45). Finally, the second client is requested to decrease its depth bracket range progressively to its optimized value of 30 (i.e. the one stored at the central controller data

base). Hence, this second embodiment follows the steps S112 and S112a-d as outlined above.

According to a third embodiment, both 3D video sequences meet at the 2D case, i.e. the clients are requested to decrease and increase respectively their depth bracket range values to and from zero. In other respect the third
5 embodiment follows the steps as presented for the first embodiment. That is, the intermediate depth perception parameter value corresponds to a depth bracket range of zero. Hence, this third embodiment follows the steps S112 and S112a-d as outlined above.

10 According to a fourth embodiment, both 3D video sequences meet at a so-called black frame. According to this embodiment, the depth bracket range is thus not modified. The first 3D video sequence is shaded progressively until only a black frame is displayed. Then the switch is then carried out and the second 3D video sequence is displayed with progressing intensity from the
15 black screen until it is fully seen on the display of the rendering unit 7.

On-call modifications

During the call, the scene of one of the clients may change, e.g. if a new object is introduced to the scene captured by the capturing unit 6. Depending on the location of such a new object (i.e. depending on its depth, Z_{object}), the depth
20 bracket range for this client may also change, either because the new object is too close or too far from the capturing unit 6 (i.e. $Z_{object} < Z_{min}$ or $Z_{object} > Z_{max}$ respectively).

In order to detect these cases, a periodical check of the depth bracket range value may be carried out at each client during the call. According to one
25 embodiment the processing unit 3, 10 is thus further arranged to, in an optional step S116, periodically check for a change of at least one of the first depth perception parameter value and the second depth perception parameter value. These instructions may be provided by the checking units 4g, 11g. Hence the checking units 4g, 11g may be configured to periodically

check for such a change. The computer program 14 and/or computer program product 13 may thus comprise means for performing instructions according to step S116.

5 If the depth bracket range is the same as the previous check (and therefore as the one stored in the central controller data base), then nothing happens. However, if a change in the depth bracket range is detected, then the client communicates e.g. through RTCP (RTP control protocol; where RTP is a real-time transport protocol) messages this change to the central controller, which updates its data base with the new value.

10 *Disconnection*

Finally, when a client is disconnected from the multi-party call, in addition to performing a common disconnect procedure, the processing unit 10 of the central controller erases the data of the thus disconnected client in the data base.

15 Second overall embodiment

The main difference between the first overall embodiment and the second overall embodiment is the criterion used for the 3D video switching. In the previously disclosed first overall embodiment the depth bracket range is considered as a representative of the depth perception parameter value
20 whereas in the present second overall embodiment the perceived depth range is considered as a representative of the depth perception parameter value. Processing, such as initial phase processing, switching processing, on-call modifications and disconnection, is according to the second overall embodiment handled in a similar manner as the first overall embodiment
25 and is repeated here for completeness.

Modification of perceived depth range

The second overall embodiment is based on modification of the perceived depth range, or parallax, during the 3D video switching. As described above the perceived depth is the distance between the viewer's eyes and the location

of the object in the virtual space (since the object is actually rendered at the screen). From Equation (3) one can observe that the perceived depth is dependent on the inter-ocular distance t_e , the viewing distance Z_D and the parallax P . Since the inter-ocular distance t_e and the viewing distance Z_D are constant values, the perceived depth may only be modified with parallax P .

As has been described above, parallax is defined as the spatial distance between the points in the left and right views on the screen. This implies that the parallax may be changed by shifting the left and right views to make the parallax larger or smaller, provided that the resulting values are kept within the limits of the depth budget.

In order to illustrate how the perceived depth range may be modified, consider parallaxes P in the range $[P_{\min}; P_{\max}]$, where $P_{\min} \geq P_{DB \min}$ and $P_{\max} \leq P_{DB \max}$ (or in other words, where the limits of the parallaxes are contained within the limits of the depth budget). The perceived depth range is therefore defined as:

$$\Delta Z_p = Z_{p \max} - Z_{p \min} \quad , \quad (8)$$

where $Z_{p \max}$ indicates the maximum perceived depth within the parallaxes range and $Z_{p \min}$ indicates the minimum perceived depth within the parallaxes range. It is easy for a skilled person to derive:

$$Z_{p \max} = \frac{Z_D t_e}{t_e - P_{\max}} \quad \text{and} \quad Z_{p \min} = \frac{Z_D t_e}{t_e - P_{\min}} \quad , \quad (9)$$

which yields that:

$$\Delta Z_p = Z_D t_e \left(\frac{1}{t_e - P_{\max}} - \frac{1}{t_e - P_{\min}} \right) \quad (10)$$

It is also easy for a skilled person to deduce that when the parallaxes are shifted so that they become larger, the perceived depth range is increased; and when the parallaxes are shifted so that they become smaller, the perceived depth range is reduced. Therefore, references is made to an
5 increase or a decrease of the perceived depth range, relate to modification (i.e. shift) of parallaxes at the rendering side.

The second overall embodiment may be divided into four main parts: an initial phase where the perceived depth ranges are calculated and stored for subsequent switches; a 3D video switch; on-call modifications; and
10 disconnection.

Note that these four parts may be performed sequentially (as they are going to be explained), but they can also occur in parallel. For example, if three video conferencing client devices are connected, there may be already 3D video switching being performed between the three video conferencing client
15 devices before a fourth video conferencing client device joins the multi-party 3D video conference. Likewise, depth bracket range modifications may be determined before other video conferencing client devices are connected. The division in these three main parts is hence given purely for descriptive purposes.

20 *Initial phase*

According to one embodiment within the second overall embodiment the initial phase comprises the following.

The first video conferencing client device requests a connection to a multi-party video conference. The connection request is sent by the
25 communications interface 5 of the electronic devices 2a representing the first video conferencing client device. The connection request may be sent to an electronic device 9 representing a central controller, such as an MRFP.

The first client and the central controller negotiate connection properties, such as audio and video codecs, for example through SIP/SDP (Session

Initiation Protocol / Session Description Protocol) negotiation or according to other protocols, such as H.323. During the negotiation, the first client also signals its rendering capabilities, namely its perceived depth range, or parallax (it does not matter which parameter is signaled, provided that all the remaining video conferencing client device provide the same parameter). If the first client is arranged only to render 2D video through its rendering device 7, the perceived depth range will be zero.

The central controller also assigns the first client with a unique ID, e.g. "client 1". The central controller stores the client ID in the memory 11 and the value of the perceived depth range into a local or remote data base to which it will access the data when required.

The first client starts transmitting the captured video to the central controller. Thus, a first 3D video sequence is captured by the capturing unit 6 of the electronic devices 2a representing the first video conferencing client device 2a and the transmitted by its communications interface 5. The first 3D video sequence is then received through the communications interface 12 of the electronic devices 9 representing the central controller. At this point it is, however, assumed that no other clients are connected to the 3D video conference. The central controller may thus disregard the received first 3D video sequence and not transmit any 3D video sequences to the clients. For example, the processing unit 3 may be arranged to on the rendering unit 7 display a message indicating the lack of other connected clients.

Then another client (hereinafter a second client, as represented by a second video conferencing client device 2b) requests a connection to the central controller.

The same SIP/SDP negotiation (or any other negotiation protocol) takes place between the central controller and the second client. That is, the second client also signals its rendering capabilities, namely its perceived depth range or parallax.

The central controller also assigns the second client with a unique ID, e.g. "client 2". The central controller updates the data base with the new client ID and the value of the perceived depth range.

3D video communication between the first client and the second client may then start. At this point only two clients are connected to the call, implying that no video switching may yet be required.

Then another client (hereinafter a third client, as represented by a third video conferencing client device 2c) requests a connection to the central controller.

The same SIP/SDP negotiation (or any other negotiation protocol) takes place between the central controller and the third client. That is, the third client also signals its rendering capabilities, namely its perceived depth range or parallax.

The central controller also assigns the third client with a unique ID, e.g. "client 3". The central controller updates the data base with the new client ID and the value of the perceived depth range.

As the skilled person understands, also at least one further client may join the 3D video conference according to the steps as outlined above.

At this point, three or more clients are connected to the 3D video conference, and therefore a 3D video switching may be required, e.g., depending on the audio level of each client. Further details of the switching are provided next.

3D video switching

In general terms, the central controller may be arranged to handle the main speaker switching between the clients based, e.g., on the audio level. When performing the switching in 3D video conferencing, a number of different scenarios may be considered depending on the video sequences transmitted:

2D-to-2D video switch: This first scenario corresponds to current video conferencing systems and an explanation thereof is therefore omitted.

2D-to-3D video switch: This second scenario deals with switching from a 2D video sequence to a 3D video sequence. In other words, the user experience corresponds to transition from a scene without depth perception (or flat) to a scene with depth perception. In order to make the switch smooth, the 3D video sequence could be adapted so that its perceived depth range is zero at the switching moment. This implies that, when the central controller determines that the main speaker video sequence needs to be changed, the central controller requests the client device transmitting the 3D video sequence to adapt its 3D video sequence so that its perceived depth range is zero. Thus, in this second scenario the intermediate depth perception parameter value corresponds to a perceived depth range of zero. Then, once the switch is carried out, the client device transmitting the 3D video sequence increases its perceived depth range progressively up to its optimized value (i.e. the value stored at the central controller data base).

3D-to-2D video switch: This third scenario deals with switching from a 3D video sequence to a 2D video sequence. In other words, the user experience corresponds to transition from a scene with depth perception to a scene without depth perception (or flat). When the central controller determines that the main speaker video sequence needs to be changed, the central controller requests the client transmitting the 3D video to progressively adapt its 3D video sequence so that its perceived depth range becomes zero. Thus, in this third scenario the intermediate depth perception parameter value corresponds to a perceived depth range of zero. Once the perceived depth range is zero (i.e. the video sequence is now 2D), the switch to the 2D video is performed.

3D-to-3D video switch: This fourth scenario deals with switching from a first 3D video sequence to a second 3D video sequence. In other words, the user experience corresponds to transition of a scene with a first depth perception to a scene with a second, different, depth perception. For this scenario, there are different possibilities on how the transition may be performed. Particular embodiments (each within the second overall embodiment) will be disclosed next. According to one embodiment the intermediate depth perception

parameter value takes a predetermined value. For example, the predetermined value may correspond to a flatter perceived depth than the perceived depth of the first 3D video sequence and the perceived depth of the second 3D video sequence.

- 5 According to a first embodiment, both 3D video sequences meet at a perceived depth range (as determined by the intermediate depth perception parameter value) that is comfortable for both clients. For example, consider switching from a first client that transmits a first 3D video sequence with a perceived depth range of 150 cm to a second client that transmits a second
- 10 3D video sequence with a perceived depth range of 75 cm. The first client is requested to progressively reduce its perceived depth range to e.g. 60 cm (as determined by the intermediate depth perception parameter value). Then, the switch is carried out between the first 3D video sequence and the second 3D video sequence at the perceived depth range of 60 cm (as the second
- 15 client has also been requested to change its perceived depth range to 60 cm). Finally, the second client is requested to increase its perceived depth range progressively to its optimized value of 75 cm (i.e. the one stored at the central controller data base). Hence, this first embodiment follows the steps S112 and S112a-d as outlined above.
- 20 According to a second embodiment, both 3D video sequences meet at an intermediate perceived depth range, which may or may not be comfortable for both clients. For example, consider switching from a first client that transmits a first 3D video sequence with a depth bracket range of 125 cm to a second client that transmits a second 3D video sequence with a depth bracket
- 25 range of 75 cm. The first client is requested to progressively reduce its perceived depth range to e.g. 100 cm (as determined by the intermediate depth perception parameter value). Thus, according to one embodiment the intermediate depth perception parameter value is a mean value of the first depth perception parameter value and the second depth perception
- 30 parameter value. Then, the switch is carried out between the first 3D video sequence and the second 3D video sequence at the perceived depth range of 100 cm (as the second client has also been requested to change its perceived

depth range to 100 cm). Finally, the second client is requested to decrease its perceived depth range progressively to its optimized value of 75 cm (i.e. the one stored at the central controller data base). Hence, this second embodiment follows the steps S112 and S112a-d as outlined above.

- 5 According to a third embodiment, both 3D video sequences meet at the 2D case, i.e. the clients are requested to decrease and increase respectively their perceived depth range values to and from zero. In other respect the third embodiment follows the steps as presented for the first embodiment. That is, the intermediate depth perception parameter value corresponds to a
- 10 perceived depth range of zero. Hence, this third embodiment follows the steps S112 and S112a-d as outlined above.

- According to a fourth embodiment, both 3D video sequences meet at a so-called black frame. According to this embodiment, the perceived depth range is thus not modified. The first 3D video sequence is shaded progressively
- 15 until only a black frame is displayed. Then the switch is then carried out and the second 3D video sequence is displayed with progressing intensity from the black screen until it is fully seen on the display of the rendering unit 7.

On-call modifications

- During the call, the scene of one of the clients may change, e.g. if a new object is introduced to the scene captured by the capturing unit 6. Depending on the
- 20 location of such a new object (i.e. depending on its parallax, P_{object} , and therefore of its perceived depth $Z_{p\ object}$), the perceived depth range for this client may also change, either because the new element is too close or too far from the capturing unit 6 (i.e. $Z_{p\ object} < Z_{p\ min}$ or $Z_{p\ object} > Z_{p\ max}$ respectively).

- 25 In order to detect these cases, a periodical check of the perceived depth range value may be carried out at each client during the call. According to one embodiment the processing unit 3, 10 is thus further arranged to, in an optional step S116, periodically check for a change of at least one of the first depth perception parameter value and the second depth perception

parameter value. These instructions may be provided by the checking units 4g, 11g. Hence the checking units 4g, 11g may be configured to periodically check for such a change. The computer program 14 and/or computer program product 13 may thus comprise means for performing instructions
5 according to step S116.

If the perceived depth range is the same as the previous check (and therefore as the one stored in the central controller data base), then nothing happens. However, if a change in the perceived depth range is detected, then the client communicates e.g. through RTCP (RTP control protocol; where RTP is a real-
10 time transport protocol) messages this change to the central controller, which updates its data base with the new value.

Disconnection

Finally, when a client is disconnected from the multi-party call, in addition to performing a common disconnect procedure, the processing unit 10 of the
15 central controller erases the data of the thus disconnected client in the data base.

A 3D video conference system 1 may comprise at least three electronic devices according to any one of the herein disclosed embodiments.

The inventive concept has mainly been described above with reference to a
20 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 inventive concept, as defined by the appended patent claims.

CLAIMS

1. A method for 3D video sequence depth parameter determination, the method being performed by an electronic device, comprising the steps of:
acquiring (S102) a first depth perception parameter value of a first 3D
5 video sequence;
acquiring (S104) a second depth perception parameter value of a second
3D video sequence; and
determining (S106) an intermediate depth perception parameter value
based on the first depth perception parameter value and the second depth
10 perception parameter value, wherein the intermediate depth perception
parameter value is to be used during a switch between the first 3D video
sequence and the second 3D video sequence.
2. The method according to claim 1, further comprising:
switching (S112) between the first 3D video sequence and the second 3D
15 video sequence.
3. The method according to claim 2, further comprising:
transmitting (S114) a rendition of the first 3D video sequence, the
intermediate 3D video sequence, and the second 3D video sequence to at
least one of a rendering device and a control device.
- 20 4. The method according to claim 2 or 3, wherein the switching (S112)
further comprises:
generating (S112a) an intermediate 3D video sequence, wherein the
intermediate depth perception parameter value is used for generating the
intermediate 3D video sequence;
25 switching (S112b) from the first 3D video sequence to the intermediate
3D video sequence; and
switching (S112d) from the intermediate 3D video sequence to the
second 3D video sequence.
5. The method according to claim 4, wherein the switching further
30 comprises:

adjusting (S112c) the depth perception parameter value of the first 3D video sequence from the first depth perception parameter value to the intermediate depth perception parameter value while displaying the first 3D video sequence; and

- 5 adjusting (S112e) the depth perception parameter value of the second 3D video sequence from the intermediate depth perception parameter value to the second depth perception parameter value while displaying the second 3D video sequence.

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

receiving (S108) an indication for switching between the first 3D video sequence and the second 3D video sequence; and

initiating (S110) switching between the first 3D video sequence and the second 3D video sequence based on the indication.

- 15 7. The method according to any one of the preceding claims, further comprising:
periodically checking (S116) for a change of at least one of the first depth perception parameter value and the second depth perception parameter value.

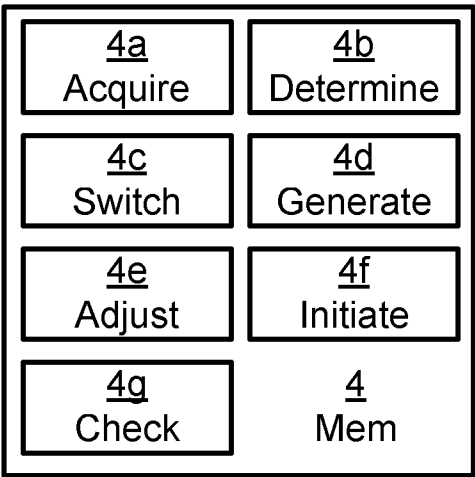
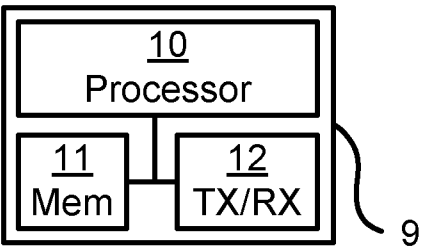
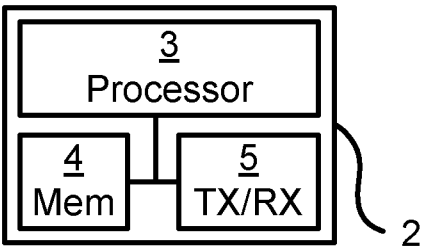
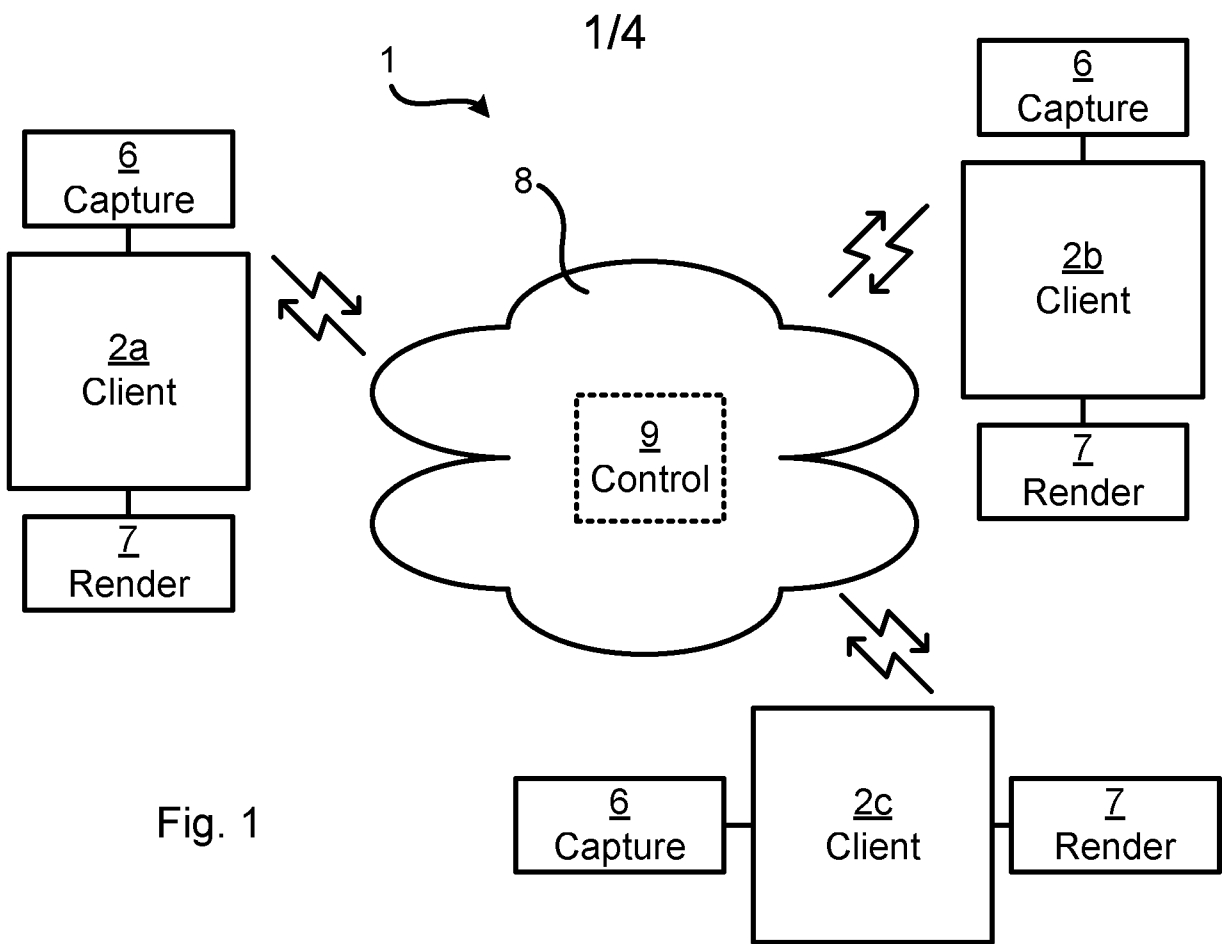
- 20 8. The method according to any one of the preceding claims, wherein the intermediate depth perception parameter value is a mean value of the first depth perception parameter value and the second depth perception parameter value.

9. The method according to any one of the preceding claims, wherein the
25 intermediate depth perception parameter value takes a predetermined value.

10. The method according to claim 9, wherein the predetermined value corresponds to a more narrow depth bracket than the depth bracket of the first 3D video sequence and the depth bracket of the second 3D video sequence.

11. The method according to claim 9, wherein the predetermined value corresponds to a flatter perceived depth than the perceived depth of the first 3D video sequence and the perceived depth of the second 3D video sequence.
12. The method according to any one of the preceding claims, wherein the
5 depth perception parameter value is based on a depth bracket range or a produced disparity range.
13. The method according to any one of the preceding claims, wherein the depth perception parameter value is based on a perceived depth or parallax.
14. The method according to any one of the preceding claims, wherein the
10 steps are performed in real-time.
15. An electronic device (2, 9) for 3D video sequence depth parameter determination, the electronic device comprising:
a processing unit (3, 10) arranged to acquire a first depth perception parameter value of a first 3D video sequence;
15 the processing unit further being arranged to acquire a second depth perception parameter value of a second 3D video sequence; and
the processing unit further being arranged to determine an intermediate depth perception parameter value based on the first depth perception parameter value and the second depth perception parameter value, wherein
20 the intermediate depth perception parameter value is to be used during a switch between the first 3D video sequence and the second 3D video sequence.
16. The electronic device according to claim 15, further comprising at least one of:
25 a 3D video sequence capturing unit (6) arranged to capture one of the first 3D image video sequence and the second 3D video sequence; and
a 3D video sequence rendering unit (7) arranged to render at least one of the first 3D image video sequence and the second 3D image video sequence.

17. The electronic device according to claim 15 or 16, further comprising:
a communications interface (12) arranged to receive at least one of the
first 3D image video sequence and the second 3D image video sequence from
a 3D video sequence capturing unit device and to transmit at least one of the
5 first 3D image video sequence and the second 3D image video sequence to a
3D video sequence rendering unit device.
18. A 3D video conference system (1) comprising at least three electronic
devices according to any one of claims 15, 16, or 17.
19. A computer program (14) for 3D video sequence depth parameter
10 determination, the computer program comprising computer program code
which, when run on an electronic device (2, 9), causes the electronic device
to:
acquire (S102) a first depth perception parameter value of a first 3D
video sequence;
15 acquire (S104) a second depth perception parameter value of a second
3D video sequence; and
determine (S106) an intermediate depth perception parameter value
based on the first depth perception parameter value and the second depth
perception parameter value, wherein the intermediate depth perception
20 parameter value is to be used during a switch between the first 3D video
sequence and the second 3D video sequence.
20. A computer program product (13) comprising a computer program (14)
according to claim 19, and computer readable means (15) on which the
computer program is stored.



2/4

Fig. 3b

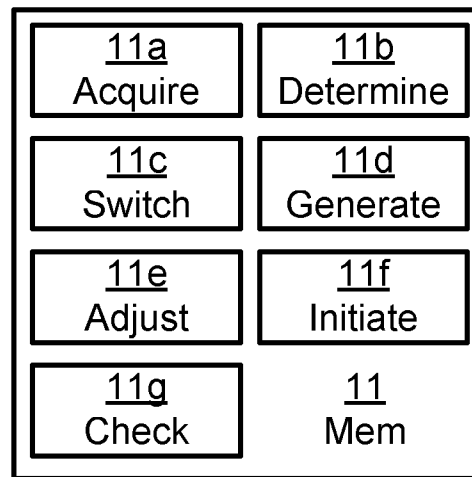


Fig. 4

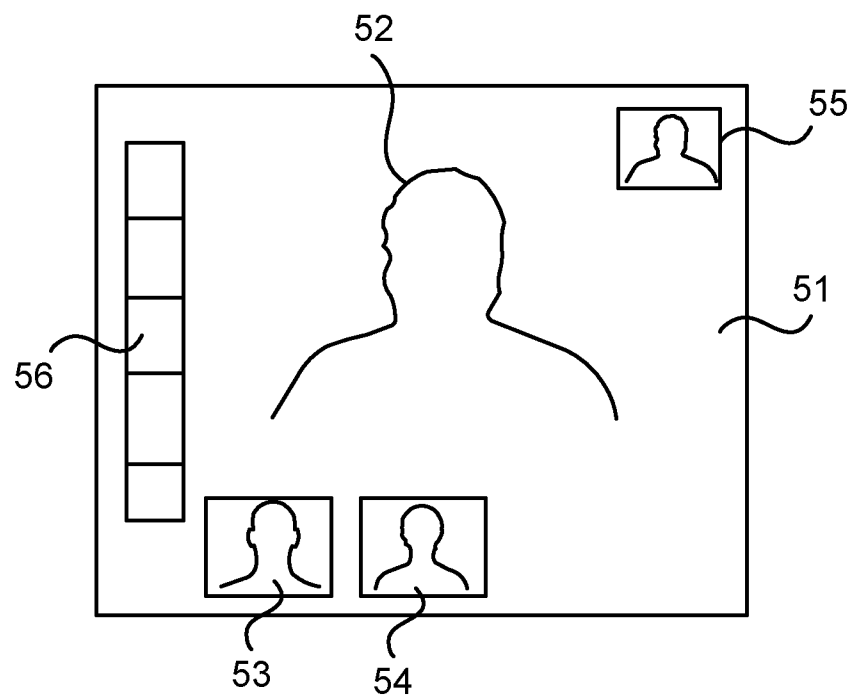
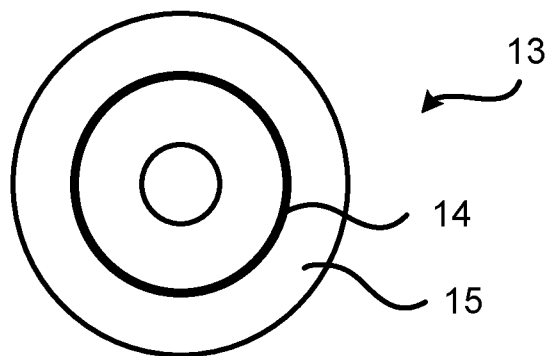


Fig. 5

3/4

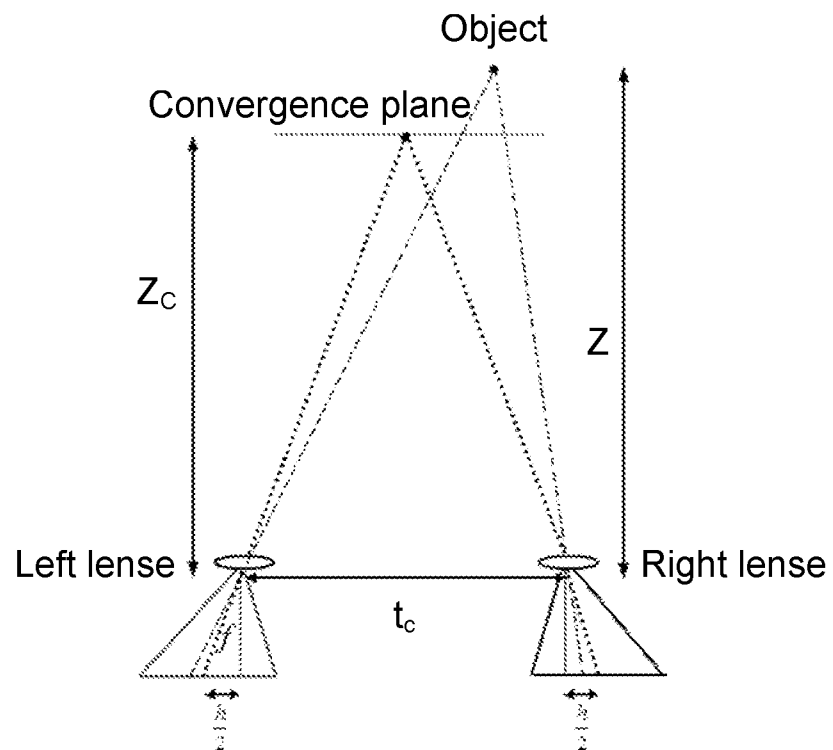


Fig. 6

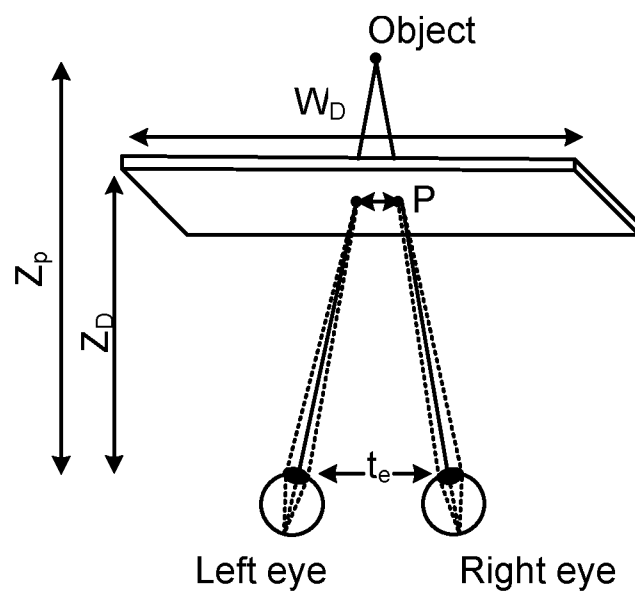


Fig. 7

4/4

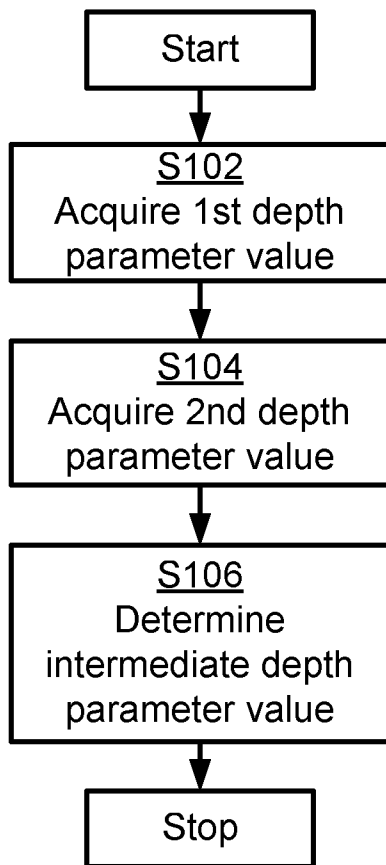


Fig. 8

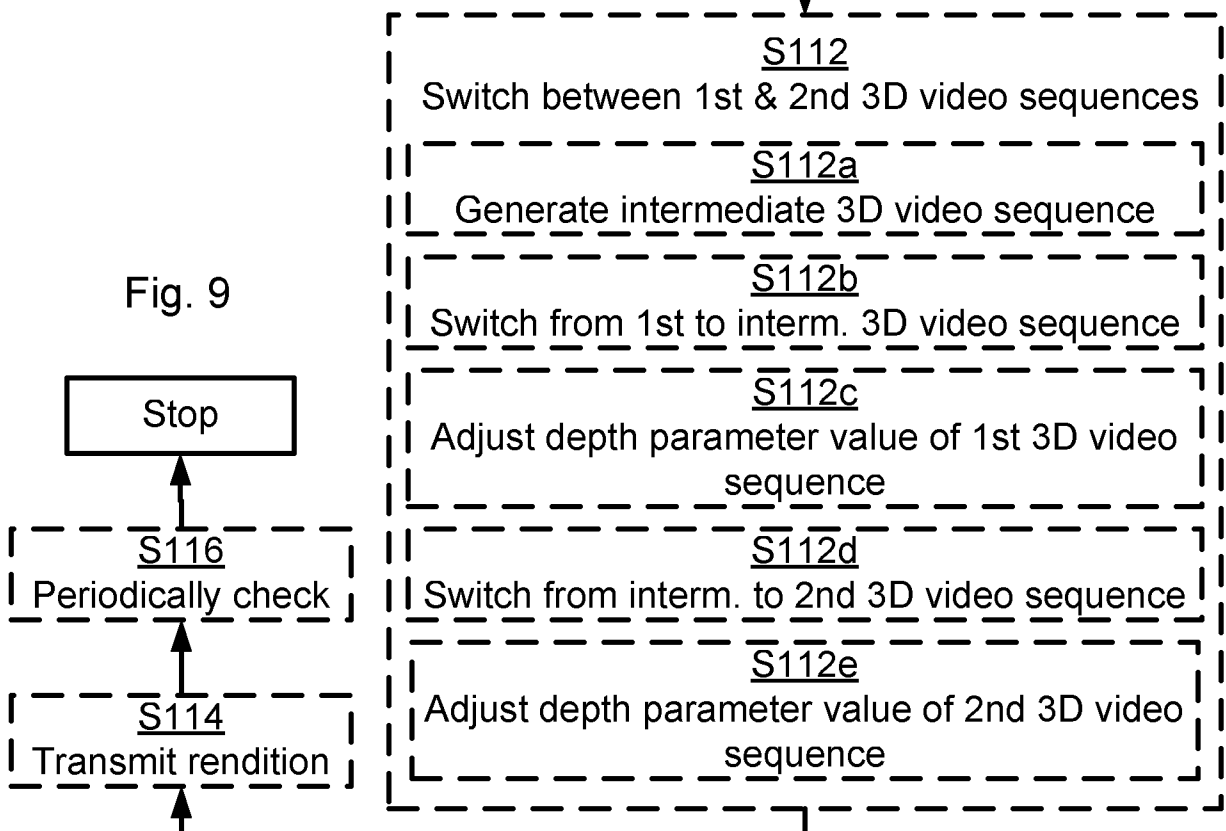
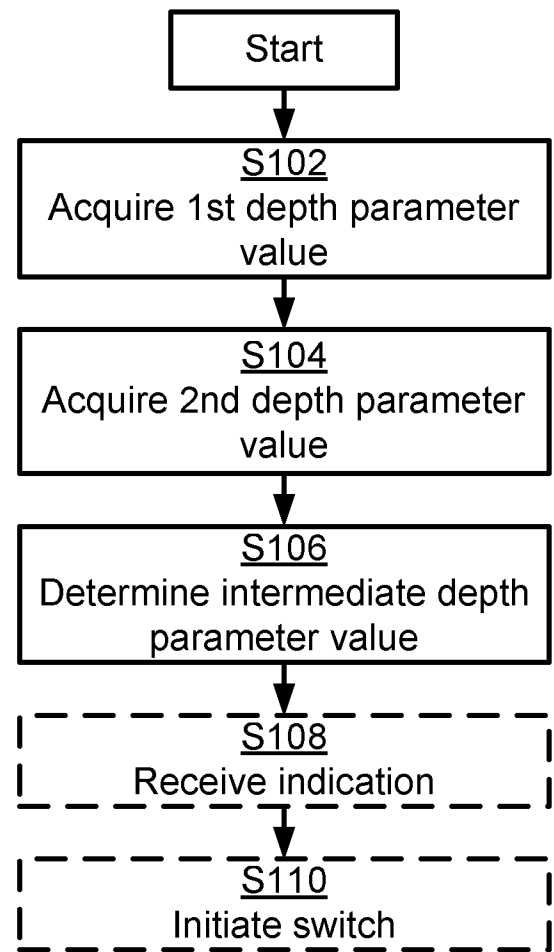


Fig. 9

INTERNATIONAL SEARCH REPORT

International application No
PCT/SE2013/050728

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N13/00
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)

EP0-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 2011/109731 A1 (KOO JAE-PHIL [KR] ET AL) 12 May 2011 (2011-05-12) paragraphs [0018] - [0040]; figures 1-7 -----	1-8, 13-20
X	US 2011/261160 A1 (TADOKORO EIJI [JP] ET AL) 27 October 2011 (2011-10-27) the whole document -----	1-20
X	US 2011/310982 A1 (YANG JEONG HYU [KR] ET AL) 22 December 2011 (2011-12-22) paragraphs [0091] - [0098]; figures 7-9 -----	1-8, 12-20
X	US 2009/244268 A1 (MASUDA TOMONORI [JP] ET AL) 1 October 2009 (2009-10-01) figures 1-23 ----- -/--	1-11, 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

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

31 July 2013

Date of mailing of the international search report

08/08/2013

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

Mao, Pauline

INTERNATIONAL SEARCH REPORT

International application No

PCT/SE2013/050728

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2012/037075 A1 (THOMSON LICENSING [FR]; ZHANG TAO [US]; HE SHAN [US]; ZOU DEKUN [US]) 22 March 2012 (2012-03-22) abstract; claims 1-17; figures 4-5 -----	1-9, 12-20

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/SE2013/050728

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011109731 A1	12-05-2011	NONE	

US 2011261160 A1	27-10-2011	CN 102067616 A	18-05-2011
		EP 2293584 A1	09-03-2011
		JP 2010258723 A	11-11-2010
		KR 20120034255 A	10-04-2012
		RU 2010151791 A	27-06-2012
		TW 201043003 A	01-12-2010
		US 2011261160 A1	27-10-2011
		WO 2010123053 A1	28-10-2010

US 2011310982 A1	22-12-2011	EP 2387243 A2	16-11-2011
		KR 20110116166 A	25-10-2011
		US 2011310982 A1	22-12-2011
		WO 2010079921 A2	15-07-2010

US 2009244268 A1	01-10-2009	EP 2106150 A2	30-09-2009
		JP 4695664 B2	08-06-2011
		JP 2009239388 A	15-10-2009
		KR 20090102705 A	30-09-2009
		US 2009244268 A1	01-10-2009

WO 2012037075 A1	22-03-2012	US 2013162641 A1	27-06-2013
		WO 2012037075 A1	22-03-2012
