(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 October 2015 (22.10.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/158570 A1

(51) International Patent Classification:
G06T 7/00 (2006.01) *H04N 13/00* (2006.01)
G06T 5/00 (2006.01)

(21) International Application Number:
PCT/EP2015/057534

(22) International Filing Date:
8 April 2015 (08.04.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14165051.5 17 April 2014 (17.04.2014) EP

(71) Applicant: **KONINKLIJKE PHILIPS N.V.** [NL/NL]; High Tech Campus 5, NL-5656 AE Eindhoven (NL).

(72) Inventors: **BRULS, Wilhelmus Hendrikus Alfonsus**; c/o High Tech Campus 5, NL-5656 AE Eindhoven (NL). **WILDEBOER, Meindert Onno**; c/o High Tech Campus 5, NL-5656 AE Eindhoven (NL).

(74) Agents: **NOWACKA, Ewa Barbara** et al.; Philips IP&S, High Tech Campus 5, NL-5656 AE Eindhoven (NL).

(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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, 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, SA, 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, ST, 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,

[Continued on next page]

WO 2015/158570 A1

(54) Title: SYSTEM, METHOD FOR COMPUTING DEPTH FROM VIDEO

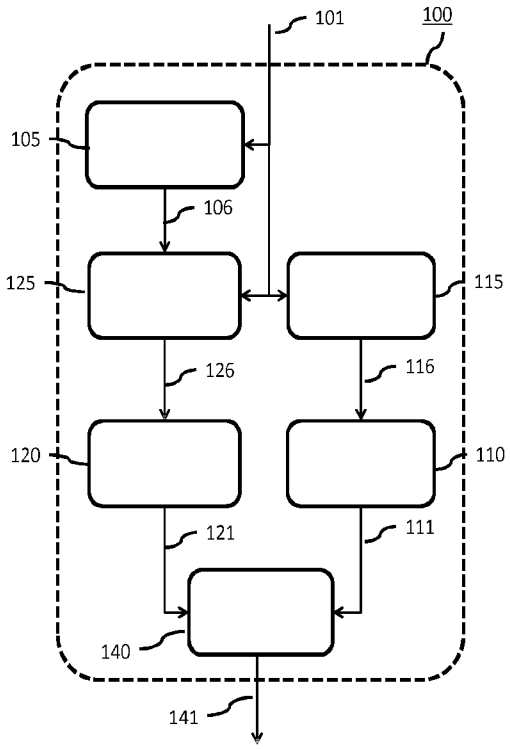


Fig. 1

(57) Abstract: A system (100) is arranged for computing a depth value (141) from a video sequence (101). The depth value represents depth at a current pixel location in a current image in the video sequence. The video sequence comprises a sequence of images. The system determines a current local motion vector (116) representing motion at the current pixel location, and determines a current dynamic depth value (111) based on the current local motion vector. The system determines in a previous image, a previous pixel location (106) comprising image content matching image content at the current pixel location. The system determines a previous local motion vector (126) representing motion at the previous pixel location, and determines a previous dynamic depth value (121) based on the previous local motion vector. Finally, the depth value is based on the current dynamic depth value and the previous dynamic depth value.

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

System, method for computing depth from video

FIELD OF THE INVENTION

The invention relates to computing depth values from a video sequence. The video sequence comprises video frames, each video frame being an image having two dimensions: a horizontal and a vertical dimension. Depth values represent a third dimension in addition to said horizontal and vertical dimension. The video sequence and corresponding depth values may be converted to a three-dimensional (3D) format for being displayed on a 3D display.

BACKGROUND OF THE INVENTION

EP 2629531 describes a method for computing a depth map from a video sequence by using motion-based depth cues. The method determines motion vectors for an image of the video sequence, wherein each of the motion vectors correspond to an image pixel of the image. A depth value corresponding to an image pixel is then computed as being proportional to the length of a motion vector corresponding to that image pixel. The depth values combined for the entire image then forms a depth map.

A drawback of said method is that meaningful depth values can only be obtained for an image of the video sequence when motion is present at said image in the video sequence. In contrast, when the image presents a static scene, a depth map cannot be computed.

SUMMARY OF THE INVENTION

It is an object of the invention to provide an improved system and method for computing a depth value based on motion in a video sequence.

An aspect of the invention is a system for computing a depth value from a video sequence, the video sequence comprising a sequence of images, each of the images comprising a two-dimensional array of image pixels, each of the image pixels having a respective pixel location in the two-dimensional array, the depth value representing depth of image content at a current pixel location, the current pixel location being a pixel location in a current image in the video sequence, the system comprising: a current motion unit arranged

for determining a current local motion vector representing motion of image content at the current pixel location; a current depth unit arranged for determining a current dynamic depth value based on the current local motion vector; a previous pixel location unit arranged for determining a previous pixel location being a pixel location in a previous image of the video sequence, the previous pixel location comprising image content corresponding to the image content at the current pixel location; previous motion unit arranged for determining a previous local motion vector representing motion of the image content at the previous pixel location; a previous depth unit arranged for determining a previous dynamic depth value based on the previous local motion vector; and a depth unit arranged for determining the depth value based on the current dynamic depth value and the previous dynamic depth value.

The video sequence comprises a sequence of images (video frames), each image comprising a two-dimensional array of image pixels. Output of the system is a depth value for a current image, being one of the images of the video sequence. The depth value represents depth of image content at the current image pixel, which is a pixel location in the current image. For example, the depth value may be one of the depth values of a depth map, whereas the depth map holds all depth values corresponding to the respective image pixels of the current image.

In the depth unit, the depth value is based on two dynamic depth values from two respective images, being (a) the current dynamic depth value from the current image, and (b) the previous dynamic depth value from the previous image. The current depth value represents depth of image content at said current pixel location, whereas the previous depth value represents depth of image content at the previous pixel location. The previous pixel location is a pixel location in the previous image, and image content at that previous pixel location corresponds to image content at the current pixel location.

In the current motion unit, the current local motion vector is determined. The current local motion vector represents motion of image content at the current pixel location. The current local motion vector may have a horizontal and vertical component indicating a two-dimensional displacement of the image content at the current pixel location. For example, the local motion vector may refer to a displacement of said image content between the current image and an image adjacent to the current image in the video sequence. For example, determining the current local motion vector may be done using a motion estimation algorithm that estimates motion of the image content at the current pixel location, or may be selected from local motion vectors already provided with the video sequence.

The current dynamic depth value is a motion-based depth value from the current image, and is based on the current local motion vector. The length of the current local motion vector may be used as an indicator of the current dynamic depth value.

In the previous location unit, the previous pixel location is determined as a pixel location in the previous image where the image content matches the image content at the current pixel location (which is in the current image). For example, the image content at the current pixel location and the image content at the previous pixel location are related. The image content may portray a portion of a face of an actor, for example. Yet, image content at the current and previous pixel location correspond to respective different moments in time, namely to the respective moments in time corresponding to the current and previous image, respectively. Determining the previous pixel location may be done, for example, by applying a motion estimation algorithm (or other image matching algorithm) to locate a pixel location where the image content matches the image content at the current pixel location, and to determine the previous pixel location as that located pixel location. (Such a process for locating the previous pixel location by means of matching image content is well known to a person skilled in the art of motion estimation in a video sequence.)

It should be noted that according to above definitions, the 'previous pixel location' consistently refers to (a pixel location in) the previous image, and that the 'current pixel location' consistently refers to (a pixel location in) the current image.

In the previous motion unit, the previous local motion vector is determined. The previous local motion vector represents motion of image content at the previous pixel location, analogous to the current local motion vector at the current pixel location. For example, the previous local motion vector may refer to a displacement of said image content between the previous image and an image adjacent to the previous image in the video sequence.

In the previous depth unit, the previous dynamic depth value is based on the previous local motion vector, which represents motion of image content at the previous pixel location. Determining the previous local motion vector for the previous pixel location is done in an analogous manner as determining the current local motion for the current pixel location. The previous dynamic depth value is thus a motion-based depth value from the previous image.

In the depth unit, the depth value is determined based on the current dynamic depth value and the previous dynamic depth value. For example, the depth value may be

computed based on a. linear combination of the current depth value and previous dynamic depth value.

An effect of the invention is therefore that the depth value is not only dependent on motion in the current image, but may also benefit from motion in the previous image. When the current image in general and the current pixel in particular contain little motion, the depth value may be improved by using a motion-based depth value from the previous image, in addition to a motion-based depth value from the current dynamic depth value.

Optionally, the depth unit is arranged for determining the depth value by computing the depth value as a combination of the current dynamic depth value and the previous dynamic depth value, the relative contributions of the current dynamic depth value and the previous dynamic depth value in the combination being defined by a current weight factor and a previous weight factor, respectively. The current weight factor and the previous weight factor may be used to make the depth value more reliant on the current image or on the previous image. In such a way, the weight factors may effectively enable a tuning of the depth value towards (a) using motion image from the current image or (b) using motion from the previous image. For example, the depth value may be computed as a linear combination of the current dynamic depth value and the previous dynamic depth value, thus as a sum of (a) the current dynamic depth value multiplied by the current weight factor and (b) the previous dynamic depth value multiplied by the previous weight factor.

Optionally, the depth unit is further arranged for defining the current weight factor by determining a local motion indicator indicating an amount of motion at the current pixel location, and determining at least one of the current dynamic weight factor and the previous dynamic weight factor based on the local motion indicator. For example, the determined amount of motion may indicate a large amount of motion at the current pixel location and, in response, the current weight factor may be determined to be 0.9 and the dynamic weight factor may be determined to be 0.1. The depth value then depends strongly on the current dynamic depth value than on the previous dynamic depth value, because the current weight factor is large (i.e. close to 1.0).

Optionally, the depth unit is arranged for determining the local motion indicator by computing the local motion parameter based on one of: a length of the current local motion vector, an absolute value of a horizontal component of the current local motion vector, and an absolute value of a vertical component of the current local motion vector. The local motion indicator is an indicator of the amount of motion present at the current pixel

location. The local motion indicator may be computed as the length of the current local motion vector or as the absolute value of one of its components.

Optionally, the depth unit is arranged for computing the depth value by determining the current weight factor based on a difference of amount of motion at the current pixel location and an amount of motion at the previous local motion vector. For example, the depth unit may determine that the amount of motion at the current pixel location is smaller than the amount of motion at the previous pixel location. The image content at the current pixel location then effectively 'decelerates'. By consequently decreasing the current weight factor, the depth value relies less on motion at the current pixel location and consequently more on motion at the previous pixel location. The depth value thus relies more on depth from the previous image in this case.

Optionally, the current depth unit is arranged for determining the current dynamic depth value by computing the current dynamic depth value based on one of: a length of the current local motion vector, an absolute value of a horizontal component of the current local motion vector, and an absolute value of a vertical component of the current local motion vector. For example, the length of the current local motion vector is used to compute the current dynamic depth value.

Optionally the current motion unit is arranged for determining the current local motion vector by: (a) determining a local motion vector indicating motion of the image content at the current pixel location, (b) determining a global motion vector indicating global motion of image content of at least a region of the current image, and (c) computing the current local motion vector as a relative motion vector based on the local motion vector relative to the global motion vector. A current dynamic depth value based on the relative motion vector compensates for movement of the background in a video scene, because a static foreground in front of a moving background and a moving foreground in front of a static background will both correspond to a large relative motion of the foreground. In both cases, the current dynamic depth value will therefore be larger if the current pixel location is part of the foreground than if it is part of the background.

Optionally, the depth unit is arranged for determining the depth value based on a static depth value being a non-motion-based depth value based on only the current image. The static depth value may be determined even when motion is small or absent in the current image and the current dynamic depth value may therefore become unreliable. For example, the static depth value may be based on a so-called 'slant' being a depth map that defines depth as a vertical gradient, i.e. having large depth values (close to the viewer) at the bottom

of the current image and having smaller depth values (farther away from the viewer) toward the top or middle of the current image.

Optionally, the depth unit is further arranged for determining the depth value by (a) determining a combined dynamic depth value by combining the current dynamic depth value and the previous dynamic depth value, and (b) determining the depth value by combining the combined dynamic depth value with the static depth value into the depth value, relative contributions of the dynamic depth value and the static depth value in the combining being dependent on an amount of global motion present in the current image. Based on the amount of global motion present in the current image, the depth value may rely more on the static depth value or more on the combined dynamic depth value. First, the combined dynamic depth value may be determined by the depth unit of the system above, thus based on the current dynamic depth value and previous dynamic depth value. Second, the depth value may then be determined as a linear combination of the static depth value and the combined dynamic depth value. The relative contributions of the static depth value and the combined dynamic depth value in said linear combination may be represented by respective weight factors. For example, if the amount of global motion is small, this may indicate that said combined dynamic depth value is unreliable. In such a case, it may be desirable to make the depth value more dependent on the static depth value. The latter may be achieved by means of a high relative contribution of the static depth value and a low relative contribution of the combined dynamic depth value.

Optionally, the previous pixel unit is arranged for determining the previous pixel location by determining the previous pixel location according a non-motion compensated manner by determining the previous pixel location having a same coordinate in the two-dimensional array of the previous image as a coordinate of the current pixel location in the two-dimensional array of current image. According to the non-motion-compensated manner, the previous pixel location has the same (X,Y) coordinate in the 2D array of the previous image as the current pixel location has in the 2D array of the current image. Accordingly, the previous pixel location is effectively determined by straightforward copying said (X,Y) coordinate from the current pixel location in the current image to the previous pixel in the previous image. The non-motion-compensated is straightforward as no motion estimation is needed to determine the previous pixel location.

Optionally, the system is further arranged for determining the depth value by using a predetermined non-linear function for limiting the depth value to a predetermined depth value range. Large values of the current local motion vector may result in an excessive

value of the depth current dynamic depth value that lies outside the predetermined depth value range. For example, the depth value may be used to convert the image pixel at the current pixel location to a three-dimensional format, in order to be displayed at a 3D-display with a limited output depth range. For example, limiting the depth value may be achieved by applying e.g. a hard- or soft-clipping function to the depth value.

Optionally, the previous pixel location unit is arranged to determine the previous pixel location in the previous image that corresponds to a later moment in time than the current image. For example, a depth map may be computed for each image in the video sequence, processing the images one-by-one, starting with the first image and ending with the last image of the video sequence (i.e. in a regular temporal order), or the other way around, thus starting with the last image and ending with the first image of the video sequence (i.e. in a reverse temporal order). In this context, the regular temporal order implies that the video sequence is intended to be played starting with the first image (corresponding to an early time instance) and ending with the last image (corresponding to a later time instance).

A further aspect of the invention is a method for computing a depth value from a video sequence, the video sequence comprising a sequence of images, the depth value representing depth of image content at a current pixel location, the current pixel location being a pixel location in a current image in the video sequence, the video sequence comprising a sequence of images, each of the images comprising a two-dimensional array of image pixels, each of the image pixels having a respective pixel location in the two-dimensional array, the method comprising the steps of: determining a current local motion vector representing motion of image content at the current pixel location, determining a current dynamic depth value based on the current local motion vector, determining a previous pixel location being a pixel location in a previous image of the video sequence, the previous pixel location comprising image content corresponding to the image content at the current pixel location, determining a previous local motion vector representing motion of the image content at the previous pixel location, determining a previous dynamic depth value based on the previous local motion vector, and determining the depth value based on the current dynamic depth value and the previous dynamic depth value.

A further aspect of the invention is a three-dimensional display system comprising: a display unit comprising a display arranged for displaying a three-dimensional image; an input unit arranged for receiving a video sequence comprising a sequence of images, the video sequence comprising a sequence of images, each of the images comprising a two-dimensional array of image pixels, each of the image pixels having a respective pixel

location in the two-dimensional array; a conversion unit arranged for converting an image of the video sequence to the three-dimensional image using a depth map comprising depth values of the image; and a processing unit comprising the system above, the processing unit being arranged for computing depth values of the depth map from the video sequence using the system above.

A further aspect of the invention is a computer program product comprising computer program code means adapted to perform all the steps of the method above when said computer program code is run on a computer.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other aspects of the invention are apparent from and will be elucidated with reference to the embodiments described hereinafter.

In the drawings,

FIG.1 illustrates a system for computing a depth value from a video sequence,

FIG.2 illustrates the current motion unit being arranged to compute the current local motion vector as a relative local motion vector

FIG.3 illustrates the depth unit being arranged to compute the depth value,

FIG.4 illustrates a method for computing the depth value from the video sequence, and

FIG.5 illustrates a 3D display system for showing a 3D image derived from a current image of the video sequence.

It should be noted that items that have the same reference numbers in different figures, have the same structural features and the same functions. Where the function and/or structure of such an item have been explained, there is no necessity for repeated explanation thereof in the detailed description.

DETAILED DESCRIPTION OF THE INVENTION

FIG.1 illustrates a system 100 for computing a depth value 141 from a video sequence 101. The video sequence 101 comprises a sequence of images (video frames) which are two-dimensional, each image being a two-dimensional array of image pixels. The depth value 141 corresponds to an image pixel, being an image pixel at a current pixel location in a current image of the video sequence. For example, the depth value 141 may be one of the depth values of a depth map, whereas the depth map holds all depth values corresponding to the respective image pixels of the current image.

The depth is defined as the third dimension in addition to the horizontal and vertical dimension of an image of the video sequence 101. If the image is converted to a 3D-format and shown on a (3D) display, then the depth dimension would be substantially perpendicular to the plane of the display. For a viewer standing in front of the display, image content (e.g. a foreground object) having large (high) depth values would be perceived as standing more out of the display towards the viewer, than if said image content would have smaller (lower) depth value. In other words, a large depth value corresponds to a smaller perceived distance to the viewer, whereas a smaller depth value corresponds to a larger perceived distance to the viewer.

In summary, the system 100 works as follows, when in operation. A current motion unit 115 determines, from the video sequence 101, a current local motion vector 116 representing motion of image content at the current pixel location. A current depth unit 110 determines a current dynamic depth value 111 from the current local motion vector 116, which serves as an indicator of depth at the current pixel location. A previous pixel location unit 105 determines, from a previous image in the video sequence, a previous pixel location 106 having image content that matches image content at the current pixel location, thus in the current image. A previous motion unit 125 determines, from the video sequence 101, a previous local motion vector 126 representing motion of the image content at the previous pixel location 106. A previous depth unit 120 then determines a previous dynamic depth value 121 from the previous local motion vector 126, which is used as an indicator of depth at the previous pixel location 106. Finally, a depth unit 140 determines the depth value 141 based on the current dynamic depth value 111 and the previous dynamic depth value 121.

It should be noted that 'image content at the current pixel location' refers to a small group of image pixels near the current pixel location. For example, the group of image pixels may be a small patch of 8x8 image pixels. The image content at the current pixel location then refers a portion of the current image portrayed by the small patch of 8x8 image pixels.

It should be noted that that the 'current pixel location' in this document always refers to a pixel location in the current image, and the 'previous pixel location' always refers to a pixel location in the previous image. 'The current pixel location' thus implies 'the current pixel location in the current image', and 'the previous pixel location' thus implies 'the previous pixel location in the previous image'.

In the current motion unit 115, a current local motion vector 116 is thus determined for the current pixel location. The current local motion vector 116 represents

motion of the image content at the current pixel location. The current local motion vector may have a horizontal and vertical component indicating a two-dimensional displacement of said image content. For example, the current local motion vector 116 (dX, dY) represents how said image content moves between (a) the current image and (b) an image directly succeeding the current image in the video sequence. Said image content moves by a number of pixels dX in the horizontal (X) direction and by a number of pixels dY in the vertical direction (Y).

For example, the video sequence is a movie with a frame rate of 50 Hz, i.e. 50 images per second. The current image may correspond to a moment in time t1. The image directly succeeding the current image then corresponds to a moment in time t2=t1+0.02 sec. A current local motion vector 116 may be (dX, dY) = (+4,-2), representing that said image content moves horizontally by 4 pixels to the right and moves vertically by 2 pixels downward, within the time period of 0.02 sec between t1 and t2.

Determining the current local motion vector 116 may be done by estimating local motion vectors for all image pixels of the current image, and then selecting, from the local motion vectors, the current local motion vector 116 corresponding to the current pixel location. For example, local motion vectors may be determined by searching for matching (similar) image content in both the current image and an adjacent image in the video sequence. An example of such an algorithm is the so-called '3D Recursive Search' (see also 'True estimation with 3-D recursive search block matching', IEEE Transactions on Circuits and Systems for Video Technology, Vol.3 No.5, Oct 1993). Motion may be determined on a pixel-basis, implying that a local motion vector is computed (estimated) for each image pixel in the current image. Local motion vectors may initially also be determined on a block-basis, which implies that a single local motion vector is determined for each block of pixels in the current image, e.g. a block of 8x8 image pixels. In that case, the single local motion vector represents motion of every image pixel in the block of image pixels. Optionally, a refinement algorithm then may be applied to refine the block-specific local motion vector to a pixel-specific local motion vector.

Alternatively, the current local motion vector 116 may be determined by selecting it from local motion vectors that are pre-computed and provided with the video sequence 101. For example, the video sequence 101 may be obtained by decoding an encoded video stream that also contains said the pre-computed local motion vectors. For example, an MPEG-encoded video stream typically contains both the video sequence 101

and said local motion vectors. In such a case, motion estimation from the video sequence 101 is not needed, as the local motion vectors are readily available.

Rather than computing the current dynamic depth value 111 based on motion (e.g. according to the method of EP 2629531), it may be instead based on relative motion. The current local motion vector 116 can be determined as a relative motion vector. A relative local motion vector may be defined as a local motion vector relative to a global motion vector. The global motion vector may represent the overall motion in the current image, and may be computed e.g. as an average of all local motion vectors in the current image. The current local motion vector may thus be calculated as the relative motion vector by subtracting the global motion vector from the local motion vector at the current pixel location. To emphasize the difference with a relative local motion vector, a local motion vector may also be referred to, in what follows, as an absolute local motion vector.

The previous motion unit 125 may be arranged to determine the previous local motion vector as a relative local motion vector. The previous motion unit 125 is arranged in an analogous manner to the current motion unit 115, which is arranged to determine the current local motion vector as a relative local motion vector.

A benefit of computing the current local motion vector 116 as a relative local motion vector is that the current dynamic depth value 111 (which is based on the current local motion vector 116) is largely invariant to background movement in a scene of the video sequence. A scene is defined here as multiple consecutive images of the video sequence that portray related image content (for example an actor) at respective consecutive time instances. Invariance to said background movement is explained based on the following two cases.

In a first case, a scene comprising a fast moving foreground object before a static background is considered. The foreground object moves in that scene, while the background is static. Using (absolute) local motion vectors as direct indicators of depth, depth values are assigned correctly to pixels: large depth values are assigned to pixels of the foreground object and small depth values are assigned to pixels of the background. Using relative local motion vectors as direct indicators of depth has the same result in this case, because the background is static, so that the relative local motion vectors are effectively absolute motion vectors.

In a second case, a scene comprising a static foreground object before a fast moving background is considered. The foreground object is static in that scene, while the background moves. Using (absolute) local motion vectors as direct indicators of depth, depth values are incorrectly assigned: large depth values are assigned to pixels of the background

and small depth values are assigned to pixels of the foreground object. Using relative local motion vectors as direct indicators of depth has a different result in this case, because relative local motion vectors of the foreground object are large, whereas relative local motion vectors of the background are (by definition) small. Consequently, depth values are then assigned correctly to pixels: large depth values are assigned to pixels of the foreground object and small depth values are assigned to pixels of the background.

These two cases illustrate the invariance of the current dynamic depth value 111 for background movement, when the current dynamic depth value is based on the current local motion vector 115 being a relative local motion vector. Computing the current dynamic depth value 111 based on the relative local motion vector may therefore be preferred.

The depth unit 110 is arranged to determine the current dynamic depth value 111 from the current local motion vector 116. This can be done in two steps. The first step is to determine an intermediate depth value. The second step is to map the intermediate depth value to the current dynamic depth value.

The intermediate depth value can be determined as the length of the current local motion vector 116. For example, if the length of the current local motion vector is 5 pixels, then the intermediate depth value is 5. Alternatively, the intermediate depth value can be determined as the absolute value of either the horizontal component or the vertical component of the current local motion vector 116.

To map the intermediate depth value to the current dynamic depth value 111, a mapping function of the linear type $y=ax+b$ may be applied, having a gain a, an offset b, an input x for the intermediate depth value, and an output y for the current dynamic depth value 111. The mapping function maps the intermediate depth value to a depth range. For example, if the intermediate value lies typically in a range of [0..10] and needs to be mapped to a depth range of [-10..10], then the mapping function may be y=2x-10. For example, using this mapping function, the intermediate depth values 0, 5, and 10 are then mapped a current dynamic depth value 115 being -10, 0, and 10, respectively.

In alternative embodiments, non-linear mapping functions y=f(x) may also be used to map the intermediate depth value to the current dynamic depth value 111. The mapping function may alternatively include a suppression of large depth values. Such suppression may be accomplished by a function that has a low derivative for large values of x, thus for large values of the intermediate depth value. For example, the mapping function may be defined as y being proportional to the square root of x. Or as another example, the mapping function may be defined as a soft-clipping function wherein y, representing the

current dynamic depth, gradually tends to a predetermined maximum value for large values of x, representing the intermediate depth value. Without such suppression, the conversion of the current image to a 3D image to be viewed on a 3D display may result in extreme depth values causing a discomfort for a viewer. Such extreme depth values may also cause a problem in portraying the resulting 3D image on a 3D display that can only display a 3D image having limited depth output range. An auto-stereoscopic 3D display is an example of such a 3D display having a limited depth output range.

The unit 105 is arranged to determine the previous pixel location 106. For example, a motion estimation algorithm may be used to locate the previous pixel location. The previous pixel location 106 may be determined as a pixel location in the previous image where image content matches image content at the current pixel location. The previous pixel location 106 is then determined in a motion-compensated manner. As another example, the previous pixel location 106 can be determined as a copy of the current pixel location: the coordinates of the previous pixel location 106 within the previous image are the same as the coordinates of current pixel location within the current image. The previous pixel location 106 is then determined in a non-motion-compensated manner.

An advantage of the non-motion compensated manner may be that it is straightforward, because it requires no motion estimation. In the non-motion-compensated manner image content at the previous pixel location may still match image content at the current pixel location, when the image content moves only little between the previous and current image.

The previous motion unit 125 is arranged to determine the previous local motion vector 126 at the previous pixel location 106, in an analogous manner as 115 determines the current local motion vector 116 at the current pixel location. The previous depth unit 120 determines the previous dynamic depth value 121 based on the previous local motion vector 126, in an analogous manner as 110 determines the current dynamic depth value 111 based on the current local motion vector 116.

The depth unit 140 is arranged to compute the depth value 141 by combining the current dynamic depth value 111 and the previous dynamic depth value 121. Further below, where FIG.3 is discussed, the depth unit 140 will be explained in further detail.

FIG.2 illustrates the current motion unit 115 being arranged to compute the current local motion vector 116 as a relative local motion vector. A sub-unit 230 is arranged to determine a local motion vector 231 at the current pixel location. The local motion vector 231 may result directly from applying a motion estimation algorithm that determines the

motion of image content at the current pixel location. The current local motion vector 116 represents motion between the current image and an image adjacent to the current image in the video sequence 101. A sub-unit 220 is arranged to determine a global motion vector 211, for example based on local motion vectors in the current image. The sub-unit 220 may receive the local motion vectors from the sub-unit 230 or, alternatively, they may be provided with the video sequence. A global motion vector 221 may be determined by computing the average of the local motion vectors of the current image. The sub-unit 210 is arranged to compute the current local motion vector 116 as a relative motion vector, by subtracting the global motion vector 221 from the local motion vector 231.

Alternatively, the horizontal component (X) of the global motion vector 221 may be determined by computing a trimmed mean of the horizontal components of the respective local motion vectors. The trimmed mean is an average of said horizontal components, wherein the largest 10% and smallest 10% of the horizontal components are excluded from that average. Similarly, the vertical component (Y) of the global motion vector 221 is determined by computing a trimmed mean of the vertical components of the respective local motion vectors.

Alternatively, the global motion vector is determined by computing the horizontal component of the global motion vector as a median of the horizontal components of the local motion vectors, and by computing the vertical component of the global motion vector as the median of the vertical components of the local motion vectors.

Alternatively, the global motion vector is determined by a so-called projection method, which does not require use of the local motion vectors. In a projection method, the horizontal component of the global motion vector is computed as follows. A first profile is computed by vertically averaging all pixel lines of the current image. The first profile is thus a pixel line itself, comprising the vertically averaged pixel lines of the current image. In a similar way, a second profile is computed by vertically averaging all pixel lines of the previous image. The horizontal component is then determined as the amount of horizontal pixels that the first profile needs to be shifted to best match the second profile. A best match may be determined as the shift for which the difference between the (shifted) first profile and the second profile is at a minimum. The vertical component of the global motion vector is determined in an analogous way, thus based a first and second profile from horizontally averaging pixel columns of the current and previous image. For further details on said projection method, one is referred to US-patent US20090153742.

Alternatively, the global motion vector may be computed from a region in the current image, rather than from the entire image. For example, the region may consist of the current image excluding the center and the bottom of the current image. Since a foreground is expected to be at the center and bottom of the current image, the global motion vector is then computed from local motion vectors predominantly corresponding to the background. Consequently, subtracting that global motion vector from the local motion vector results in a relative motion vector that represents local motion relative to background motion. This may be considered as a more accurate and appropriate way of computing the relative motion vector.

FIG.3 illustrates the depth unit 140 being arranged to compute the depth value 141. The depth unit 140 may be arranged to combine the current dynamic depth value 111 and the previous dynamic depth value 121 using respective weight factors. The depth unit 140 includes a sub-unit 310 and a sub-unit 320.

Sub-unit 320 is arranged to combine the current dynamic depth value 111 and the previous depth value 121 into the depth value 141, using a current weight factor 311. The units 110 and 120 provide the current dynamic depth value 111 and the previous depth value 121, respectively. The current weight factor 311 is provided by the sub-unit 310 and determines to what extent the depth value 141 depends on the current dynamic depth value 111 and on the previous dynamic depth value 121. For example, the depth value 141 is determined as a weighted average:

$$D = W_1 D_1 + W_2 D_2 \qquad \text{[Eq. 1]}$$

wherein W_1 and W_2 represent the current weight factor and the previous weight factor, respectively, while D, D_1, and D_2 represent the depth value 141, the current dynamic depth value 111 and the previous dynamic depth value 121, respectively. The weight factors W_1 and W_2 are typically numbers between 0.0 and 1.0. The weight factors, W_1 and W_2 may be related according to the following equation:

$$W_2 = 1 - W_1. \qquad \text{[Eq. 1b]}$$

In such a case the sub-unit 310 only needs to provide W_1 to sub-unit 320, as sub-unit 320 derives W_2 from W_1.

In an embodiment, the depth value 141 is based on more than two images, i.e. more than the current image and the previous image. The depth value 141 may be determined based on multiple depth values from respective multiple images. For example, the depth value 141 may be based on a first, second, third, and fourth dynamic depth value from a respective first, second, third and fourth image. For sake of clarity, the terms ‘first’ and

'second' are used as substitutes for the terms 'current' and 'previous'. Each of the first, second, third and fourth depth values may be based on a first, second, third and fourth local motion vector. Similarly to Eq. 1, the depth value 141 may be computed as

$$D = W_1 D_1 + W_2 D_2 + W_3 D_3 + W_4 D_4 \quad \text{[Eq. 2]}$$

wherein:

- D_3 and D_4 represent the third and fourth dynamic depth value, respectively;
- W_3 and W_4 represent the third and fourth weight factor, respectively; and
- Variables W_1, D_1, W_2, and D_2 are defined as in Eq.1.

Eq. 2 describes a finite impulse response (FIR) filter that has an output D, input values D_1, D_2, D_3, and D_4, and respective filter coefficients W_1, W_2, W_3, and W_4.

In another embodiment, computation of the depth value D 141 is implemented as an infinite impulse response (IIR) filter as follows:

$$D = W D_1 + (1-W) C_2 \quad \text{[Eq. 3]}$$

$$C_2 = W D_2 + (1-W) C_3 \quad \text{[Eq. 4]}$$

$$C_3 = W D_3 + (1-W) C_4 \quad \text{[Eq. 5]}$$

wherein:

- W is the current weight factor 311;
- D_1, D_2, D_3, D_4 are as in Eq.2; and
- C_2, C_3, C_4 represent cumulative depth values of the second, third, fourth image, respectively. Accordingly, D represents a 'temporal update' of the cumulative depth value C_2, and is calculated as a weighted average of D_1 and the cumulative depth value C_2. In its turn, the cumulative depth value C_2 is computed as a 'temporal update', of the cumulative depth value C_3, thus as a weighted average of D_2 and the cumulative depth value C_3. In its turn, C_3 is computed as a 'temporal update' of a cumulative depth value C_4 of a fourth image, etcetera. The depth value 141 is thus computed using a temporal recursive filter, (a) because of the recursion described by of Eqs.3-5 and (b) because of the depth value is derived based on images in the video sequence 101 associated with a different time instances. Said recursion requires an initial value at some preceding image. For example, said recursion may be initialized at the fourth image by $C_4 = D_4$.

The effect of varying the current weight factor W 311 is as follows. If W is close to 1.0 then D is very similar to D_1 and depends weakly (or to a minor extent) on images preceding the current image. This means that when the video sequence 101 is shown in real time on a 3D-display, the depth value 141 is less dependent on depths from preceding images. If W is close to 0.0 then D is very similar to D_2 and largely depends on images

preceding the current image. In other words, the depth value 141 then includes only a relatively small contribution of the current depth value 121. When the video sequence 101 is shown in real time on a 3D-display, in such a case, the depth value 141 is strongly dependent on depths from preceding earlier images.

When W is close to 1.0, the depth value D 141 thus depends weakly on depths from preceding images. Consequently, D depends strongly on depth from the current image. D is therefore based on 'up-to-date image content, because D represents depth at the current pixel location which is in the current image, while D is strongly based on depth from the same, current image. A benefit of W being close to 1.0 may thus be that the depth value D 141 is based on up-to-date image content.

Another consequence of W being close to 1.0 is that D is computed to only a minor extent from motion in preceding images of the video sequence, when the current image itself contains little or no motion to compute meaningful depth. A drawback of W being close to 1.0 may thus be that the depth value D 141 may not represent a meaningful value when the current image itself contains little or no motion.

When W is close to 0.0, the depth value D 141 thus depends strongly on depths from preceding images. Consequently, D depends weakly on depth from the current image. D is therefore not based on 'up-to-date image content, because D represents depth at the current pixel location which is in the current image, while D is only weakly based on depth from the same, current image. A drawback of W being close to 0.0 may thus be that the depth value D 141 is not based on up-to-date image content.

Another consequence of W being close to 0.0 is that D can be computed from motion in preceding images of the video sequence, even when the current image itself contains little or no motion to compute meaningful depth from. A benefit of W being close to 0.0 may thus be that the depth value D 141 may be computed from motion in the video sequence, even when the current image itself contains little or no motion.

By adapting W to motion in the current image and to motion in the previous image, said benefits for computing D may be combined in an advantageous way. In what follows, it will be explained how W may be determined by such an adaptation in a first, second and third embodiment.

In said first embodiment, D may be computed primarily from motion in the current image when the current image contains more motion than the previous image. Motion in the current image may then be considered as more reliable for computing D than motion in the previous image. This may work as follows. An amount of motion at the current pixel

location is computed as the length of the current local motion vector. An amount of motion at the previous pixel location is computed as the length of the previous local motion vector. When the amount of motion at the current pixel location is larger than the amount of motion at the previous pixel location, then W may be determined as being close to 1.0, e.g. W=0.8 or W=1.0. Consequently, D then depends strongly on D_1 and is thus based on up-to-date image content.

In said second embodiment, D may be computed primarily from motion in the previous image when:

- the current image contains less motion than the previous image, and
- the current image contains little or no motion.

This may work as follows. When the amount of motion at the current pixel location is:

- smaller than the amount of motion at the previous pixel location, and
- is also smaller than a predetermined amount of motion at the current pixel location,

W may be determined as being close to 0.0, e.g. W=0.2 or W=0.0. Consequently, the effect is that D then depends strongly on D_2 and is thus computed from motion in the previous image, while the amount of motion at the current pixel location itself is small.

For example, said predetermined amount of motion at the current pixel location may be two, for example. Then, if the length, $\text{sqrt}(dX^2+dY^2)$, of the current local motion vector 116 (X,Y) is smaller than two, the image content at the current pixel location may be defined as containing 'little or no motion'.

In said third embodiment, D may be computed from both motion in the current image and motion in the previous image, when:

- the current image contains less motion than the previous image and
- the current image contains a moderate or large amount of motion.

This may work as follows. When the amount of motion at the current pixel location is:

- smaller than the amount of motion at the previous pixel location, and
- is larger than the predefined amount of motion at the current pixel location,

W may be determined as an intermediate value, e.g. W=0.5. Consequently, the effect is that D depends on both D1 and D2 to a similar extent.

As a refinement of said third embodiment, W may have an intermediate value that further depends on the amount of motion at the current pixel location. This may work as

follows. For example, when the amount of motion at the current pixel location is:

- smaller than the amount of motion at the previous pixel location, and
- is smaller than or equal to 2 then W=0.0, or
- is larger than 2 then W=0.2, or
- is larger than 4 then W=0.4, or
- is larger than 6 then W=0.6, or
- is larger than 8 then W=0.8.

D therefore depends more on the D_1 as the amount of motion at the current pixel location is larger.

The following pseudo-code comprises a combination of abovementioned first, second and third embodiment.:

```
if(currmot >= prevmot){
   W=1.0;} /* first embodiment */
else { /* second/third embodiment */
   if (currmot <=2) { W=0.0;}  /* second/third embodiment */
   if (currmot >2)  { W=0.2;}  /* third embodiment */
   if (currmot >4)  { W=0.4;}  /* third embodiment */
   if (currmot >6)  { W=0.6;}  /* third embodiment */
   if (currmot >8)  { W=0.8;}  /* third embodiment */ }
depth=W*currdepth+(1-W)*prevdepth; /* computing depth value 141 */
```

wherein:

- `currmot` represents the length of the current local motion vector 116,
- `prevmot` represents the length of the previous local motion vector 126,
- `currdepth` represents the current dynamic depth value D_1 111,
- `prevdepth` represents the previous dynamic depth value D_2 121,
- `depth` represents the depth value D 141,
- `W` represents the current weight factor W 311.

In previous embodiments determining W is based on two features:

- the amount of motion at the current pixel location and
- the amount of motion at the previous pixel location.

Additional embodiments are obtained by replacing said two features by:

- the amount of global motion in the current image and
- the amount of global motion in the previous image, respectively.

The amount of global motion in the current image may be determined as:

- a length of the global motion vector of the current image or
- an average length of local motion vectors of the current image.

The amount of global motion in the previous image is determined in a similar manner as the amount of global motion in the current image.

The depth value 141 may be part of a depth map that contains depth values of all respective image pixels of the current image. Such a depth map depends on motion, and therefore may be referred to as a 'dynamic' depth map. In addition, a so-called 'static' depth map may supplement the dynamic depth map. The static depth map is based on depth cues from within the current image and thus does not depend on motion. For example, the static depth map may be a so-called 'slant' that describes the depth as a gradient of the vertical dimension of the current image, gradually changing from nearby (i.e. large depth values) at the bottom of the current image too far away (i.e. small depth values) at the middle or top of the current image. The static depth map may be based on other non-motion dependent depth cues from the current image. For example, using a face detector, large depth values may be assigned to areas in the current image containing faces, so that the faces become part of the foreground of the current image. As another example, using a sky detector may, low depth values may be assigned to areas of the current image detected as sky, so that said areas become part of the background in the current image.

A static depth value (from a static depth map) may be added to the depth value 141. In this way, the depth value 141 contains a meaningful depth value, even when not enough motion is present in the current image for a reliable computation of the dynamic depth value. Adding the static depth value to the depth value 141 is particularly advantageous for the case (a) when stationary periods in the video sequence (i.e. with little or no motion) last too many consecutive images for the recursive temporal filter to allow reliable derivation of the depth from previous images, or (b) when the video sequence starts with a stationary scene having little motion so that no previous images are available to reliably compute a motion-based depth value.

Adding the static depth value to the depth value 141 may be rephrased as follows. First, a 'dynamic depth value' is defined as the depth value 141 before adding the static depth value. The dynamic depth value is thus based on only the current dynamic depth value 111 and the previous dynamic depth value 121. Next, the depth value 141 is determined by combining the dynamic depth value and the static depth value. For example,

as described above, the depth value 141 is a result from adding the static depth value and the dynamic depth value.

Combining the static depth value and the dynamic depth value may depend on the amount of global motion in the current image. For example, when the amount of global motion is low, the depth value 141 relies more on the static depth value than on the dynamic depth value, and, conversely, when the amount of global motion is high, the depth value 141 relies less on the static depth value than on the dynamic. Such a combination of the static depth value and the dynamic depth value may be implemented as a weighted average, wherein the respective relative contributions (i.e. weight factors) of the static and dynamic depth value in the weighted average gradually vary with the amount of global motion. In a variant of this embodiment, the depth value 141 simply equals the largest of the static depth value and the dynamic depth value, corresponding to the relative contributions being 0 or 1.

FIG.4 illustrates a method 400 for computing the depth value 141 from the video sequence 101. The method 400 comprises the following steps. Step 410 comprises determining the current local motion vector 116 representing motion of image content at the current pixel location. Step 420 comprises determining the current dynamic depth value 111 based on the current local motion vector. Step 430 comprises determining the previous pixel location 106 being a pixel location in the previous image of the video sequence; the previous pixel location comprises image content corresponding to the image content at the current pixel location. Step 440 comprises determining the previous local motion vector 126 representing motion of the image content at the previous pixel location 106. Step 450 comprises determining the previous dynamic depth value 121 based on the previous local motion vector 126. Step 460 comprises determining the depth value 141 based on the current dynamic depth value 111 and the previous dynamic depth value 121.

Operations performed by the steps 410-460 of method 400 are consistent with operations performed by units 115, 110, 105, 125, 120, and 140 of system 100, respectively.

The method 400 described above may also be implemented as computer program code means. The computer program code means may be adapted to perform the steps of the method 400 when said computer program code is run on a computer. The computer program code may be provided via a data carrier, such as a DVD or solid-state disk, for example.

FIG.5 illustrates a 3D display system 500 for showing a 3D image 531. The 3D image 531 is derived from a current image of the video sequence 101. A display unit 540 comprises a 3D display arranged for showing the 3D image 531. For example, the display

may be a stereoscopic display that requires viewer to wear stereo glasses, or an auto-stereoscopic multiview display. An input unit 510 may receive the video sequence 101, for example for reading the video sequence 101 from a storage medium or receiving it from a network. A processing unit 520 comprises the system 100 for computing a depth map 521 which comprises depth values of respective image pixels of the current image. A conversion unit 530 is for converting the current image to the 3D image 531 based on the depth map 521 and the current image.

For example, the 3D display system may be a 3D-TV. The processing unit 520 and the conversion unit 530 may be part of a video processing board that renders a 3D-video sequence in real-time for being displayed on the 3D display. The processing unit 520 computes a depth map for each image in the video sequence, whereas the conversion unit 530 converts each image and the corresponding depth map into a 3D-image. As another example, the 3D display system 500 device may implemented as a smart phone having a 3D-display capability or as studio equipment for offline conversion of the video sequence 101 into a 3D video sequence.

It should be noted that the above-mentioned embodiments illustrate rather than limit the invention, and that those skilled in the art will be able to design many alternative embodiments without departing from the scope of the appended claims. It will be appreciated by those skilled in the art that two or more of the above-mentioned embodiments, implementations, and/or aspects of the invention may be combined in any way deemed useful.

Modifications and variations of the method or the computer program product, which correspond to the described modifications and variations of the monitoring subsystem, can be carried out by a person skilled in the art on the basis of the present description.

In the claims, any reference signs placed between parentheses shall not be construed as limiting the claim. Use of the verb "comprise" and its conjugations does not exclude the presence of elements or steps other than those stated in a claim. The article "a" or "an" preceding an element does not exclude the presence of a plurality of such elements. The invention may be implemented by means of hardware comprising several distinct elements, and by means of a suitably programmed computer. In the device (or system) claim enumerating several means, several of these means may be embodied by one and the same item of hardware. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage.

The invention is defined in the independent claims. Advantageous yet optional embodiments are defined in the dependent claims.

CLAIMS:

1. A system (100) for computing a depth value (141) from a video sequence (101), the video sequence comprising a sequence of images, each of the images comprising a two-dimensional array of image pixels, each of the image pixels having a respective pixel location in the two-dimensional array, the depth value representing depth of image content at a current pixel location, the current pixel location being a pixel location in a current image in the video sequence, the system (100) comprising:

- a current motion unit (115) arranged for determining a current local motion vector (116) representing motion of image content at the current pixel location,
- a current depth unit (110) arranged for determining a current dynamic depth value (111) based on the current local motion vector
- a previous pixel location unit (105) arranged for determining a previous pixel location (106) being a pixel location in a previous image of the video sequence, the previous pixel location comprising image content corresponding to the image content at the current pixel location
- a previous motion unit (125) arranged for determining a previous local motion vector (126) representing motion of the image content at the previous pixel location
- a previous depth unit (120) arranged for determining a previous dynamic depth value (121) based on the previous local motion vector, and
- a depth unit (140) arranged for determining the depth value (141) based on the current dynamic depth value (111) and the previous dynamic depth value (121),

characterized by the current motion unit (115) being arranged for determining the current local motion vector (116) by

- determining a local motion vector indicating motion of image content at the current pixel location,
- determining a global motion vector representing global motion of image content in at least a region of the current image, wherein the global motion vector is an average of all local motion vectors in said region of the current image, and
- computing the current local motion vector as a relative motion vector based on the local motion vector relative to the global motion vector.

2. A system (100) according to claim 1, wherein the depth unit (140) is arranged for determining the depth value (141) by computing the depth value as a combination of the current dynamic depth value and the previous dynamic depth value, the relative contributions of the current dynamic depth value and the previous dynamic depth value in the combination being defined by a current weight factor (311) and a previous weight factor, respectively.

3. A system (100) of claim 2, wherein the depth unit (140) is further arranged for defining the current weight factor (311) by
- determining a local motion indicator indicating an amount of motion at the current pixel location, and
- determining at least one of the current dynamic weight factor and the previous dynamic weight factor based on the local motion indicator.

4. A system (100) of claim 3, wherein the depth unit (140) is arranged for determining the local motion indicator by computing the local motion parameter based on one of
- a length of the current local motion vector,
- an absolute value of a horizontal component of the current local motion vector, and
- an absolute value of a vertical component of the current local motion vector.

5. A system (100) of claim 2, wherein the depth unit (140) is arranged for computing the depth value (141) by determining the current weight factor based on a difference of amount of motion at the current pixel location and an amount of motion at the previous local motion vector.

6. A system (100) according to any of the previous claims, wherein the current depth unit (110) is arranged for determining the current dynamic depth value (111) by computing the current dynamic depth value based on one of:
- a length of the current local motion vector,
- an absolute value of a horizontal component of the current local motion vector,

and

- an absolute value of a vertical component of the current local motion vector.

7. A system (100) according to any of the previous claims, wherein the depth unit (140) is arranged for determining the depth value (141) based on a static depth value being a non-motion-based depth value based on only the current image.

8. A system (100) of claim 7, wherein the depth unit (140) is further arranged fordetermining the depth value (141) by

- determining a dynamic depth value based on the current dynamic depth value and the previous dynamic depth value, and
- combining the dynamic depth value with the static depth value into the depth value (141), relative contributions of the dynamic depth value and the static depth value in the combining

being dependent on the global motion vector in the current image.

9. A system (100) according to any of the previous claims, wherein the previous pixel unit (105) is arranged for determining the previous pixel location (106) according to a non-motion compensated manner by determining the previous pixel location having a same coordinate in the two-dimensional array of the previous image as a coordinate of the current pixel location in the two-dimensional array of current image.

10. A system (100) according to any of the previous claims, wherein the system is further arranged for determining the depth value (141) by using a predetermined non-linear function for limiting the depth value to a predetermined depth value range.

11. A system (100) according to any of the previous claims, wherein the previous pixel location unit (105) is arranged to determine the previous pixel location (106) in the previous image, the previous image corresponding to a later moment in time than the current image.

12. A method (400) for computing a depth value (141) from a video sequence (101), the video sequence comprising a sequence of images, the depth value representing depth of image content at a current pixel location, the current pixel location being a pixel

location in a current image in the video sequence, the video sequence comprising a sequence of images, each of the images comprising a two-dimensional array of image pixels, each of the image pixels having a respective pixel location in the two-dimensional array, the method comprising the steps of:

- determining a current local motion vector (116) representing motion of image content at the current pixel location,
- determining a current dynamic depth value (111) based on the current local motion vector,
- determining a previous pixel location (106) being a pixel location in a previous image of the video sequence, the previous pixel location comprising image content corresponding to the image content at the current pixel location,
- determining a previous local motion vector (126) representing motion of the image content at the previous pixel location,
- determining a previous dynamic depth value (121) based on the previous local motion vector, and
- determining the depth value (141) based on the current dynamic depth value (111) and the previous dynamic depth value (121),

said method characterized by the step of determining the current local motion vector being arranged for determining the current local motion vector (116) by

- determining a local motion vector indicating motion of image content at the current pixel location,
- determining a global motion vector representing global motion of image content in at least a region of the current image, wherein the global motion vector is an average of all local motion vectors in the current image, and
- computing the current local motion vector as a relative motion vector based on the local motion vector relative to the global motion vector.

13. A three-dimensional display system (500) comprising:

- a display unit (540) comprising a display arranged for displaying a three-dimensional image (531),
- an input unit (510) arranged for receiving a video sequence (101) comprising a sequence of images, the video sequence comprising a sequence of images, each of the images comprising a two-dimensional array of image pixels, each of the image pixels having a respective pixel location in the two-dimensional array,

- a conversion unit (530) arranged for converting an image (512) of the video sequence to the three-dimensional image using a depth map (521) comprising depth values of said image (512) and
- a processing unit (520) comprising the system (100) of claim 1 for computing depth values of the depth map.

14. A computer program product a computer program product comprising computer program code means adapted to perform all the steps of the method (400) according to claim 12 when said computer program code is run on a computer.

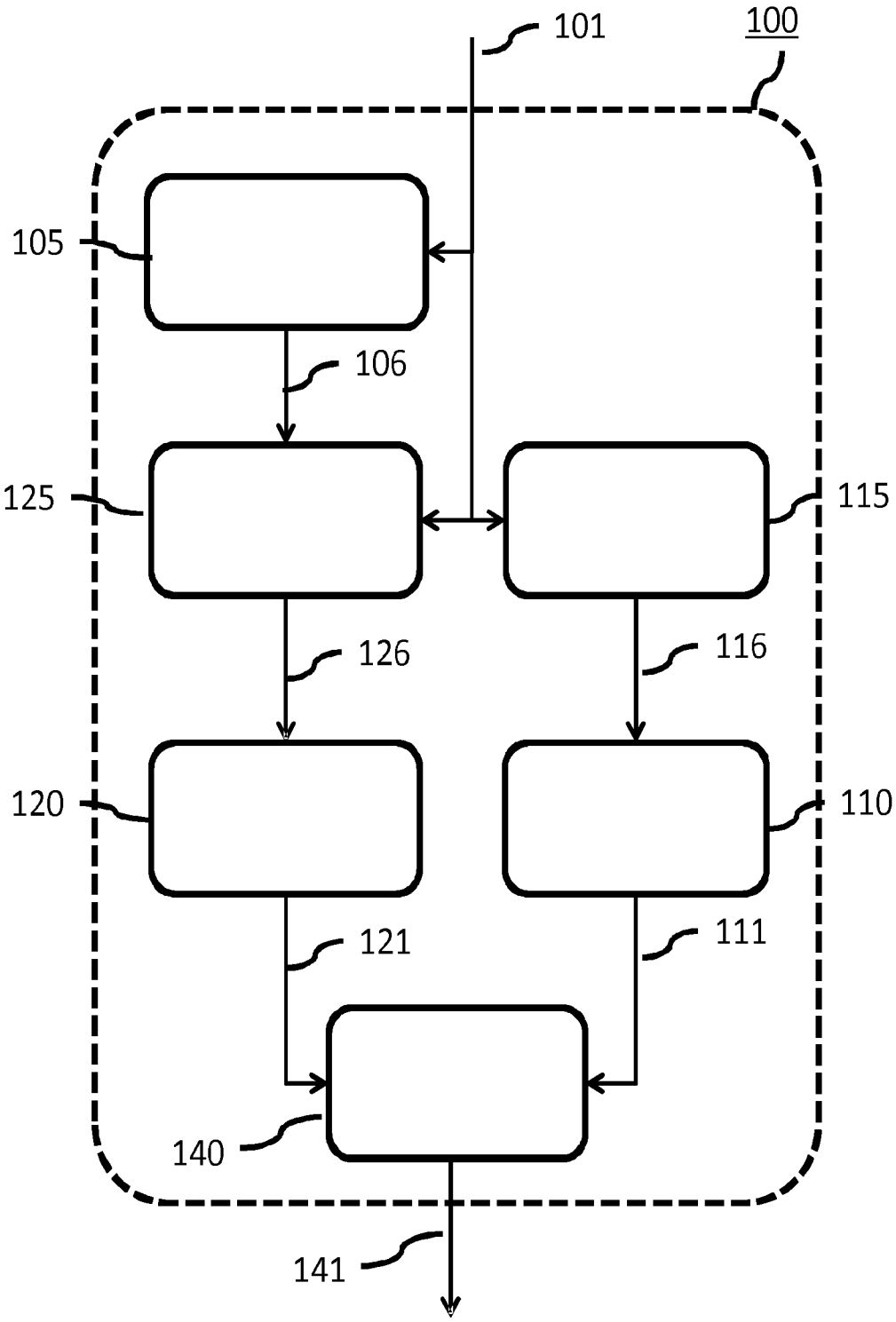


Fig. 1

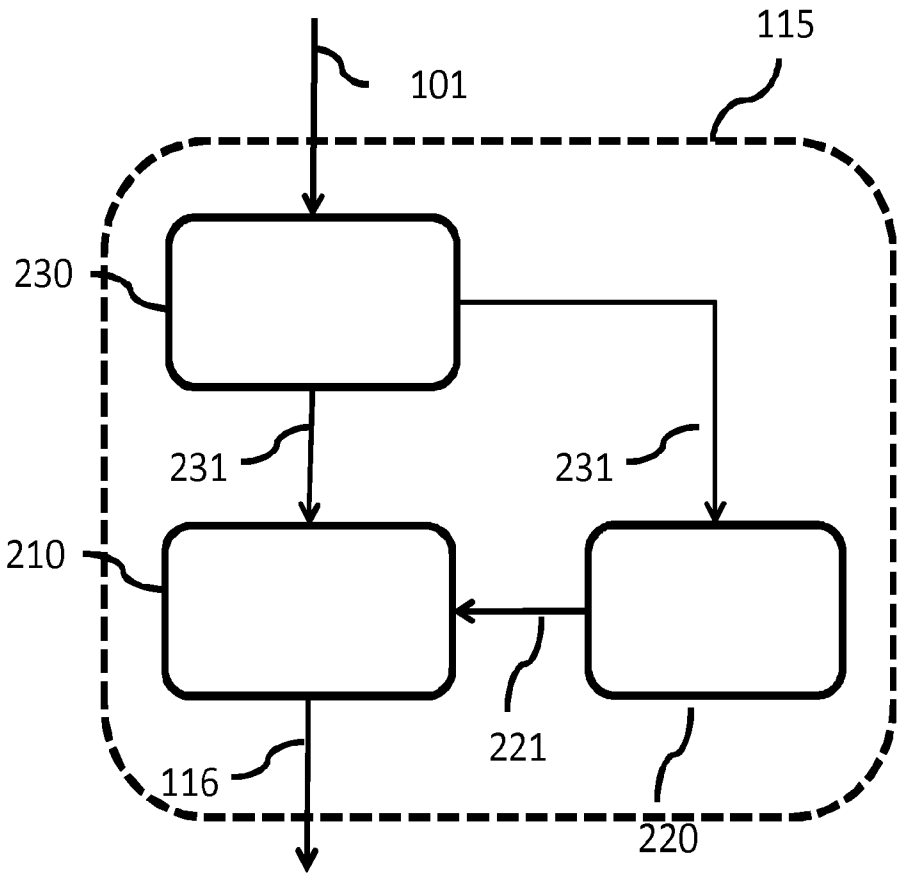


Fig. 2

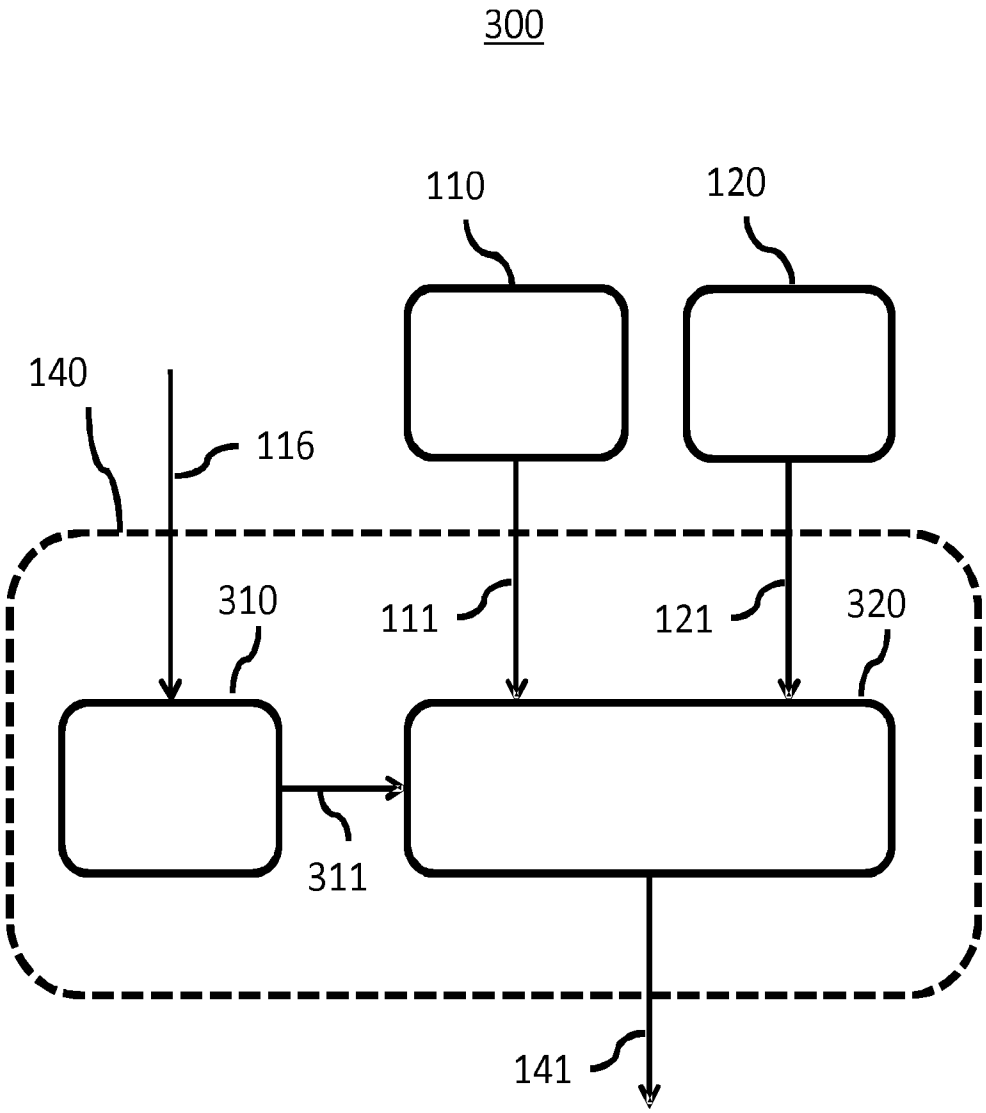


Fig. 3

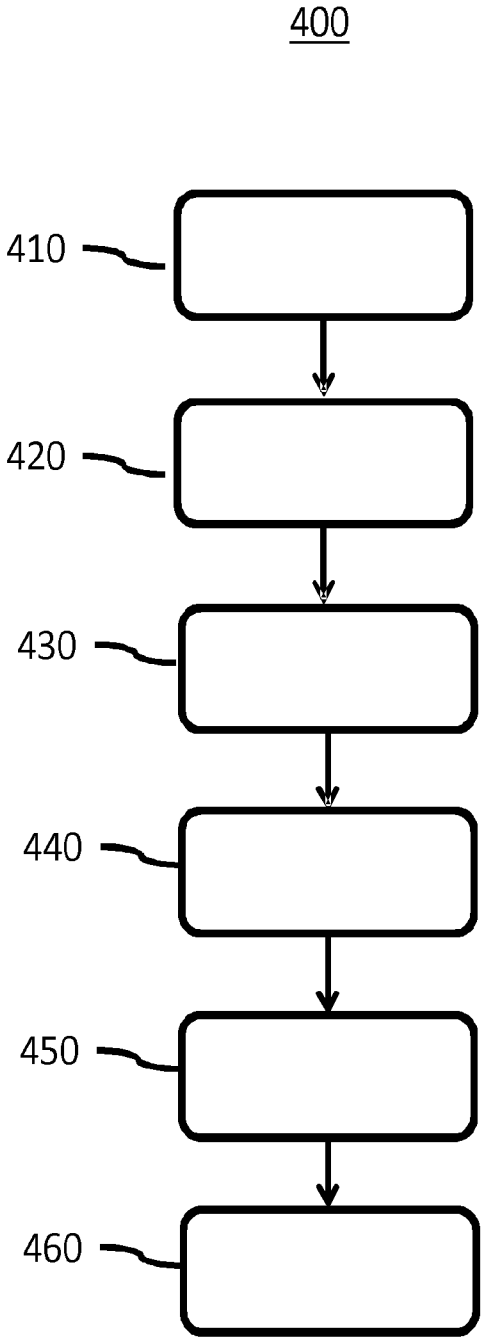


Fig. 4

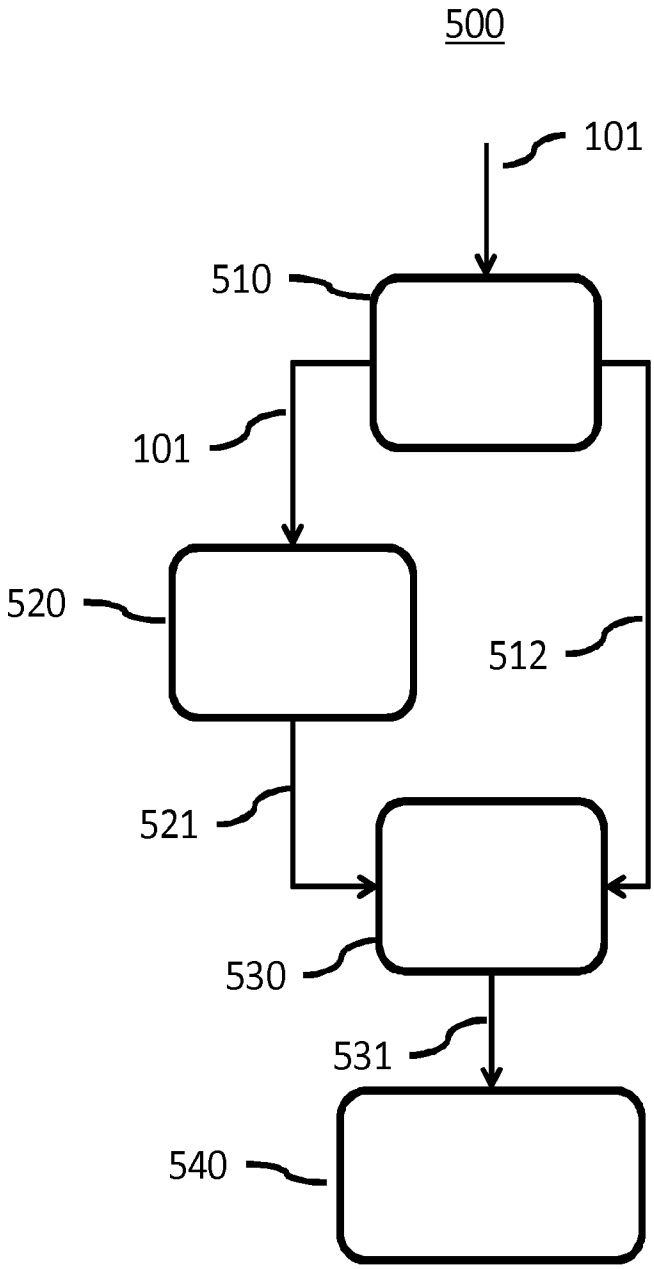


Fig. 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/057534

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06T7/00 G06T5/00 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)
G06T 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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2011/096832 A1 (ZHANG RONG [US] ET AL) 28 April 2011 (2011-04-28) paragraphs [0002], [0005] - paragraph [0008] paragraphs [0105], [0113], [0115], [0125] -----	1-14
Y	US 2013/336577 A1 (LU PEI-YING [TW] ET AL) 19 December 2013 (2013-12-19) paragraph [0050] -----	1,2,7-9, 12-14

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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	Date of mailing of the international search report
24 June 2015	02/07/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Winkler, Gregor

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MAHSA T. POURAZAD ET AL: "Generating the Depth Map from the Motion Information of H.264-Encoded 2D Video Sequence", EURASIP JOURNAL ON IMAGE AND VIDEO PROCESSING, vol. 2010, 1 January 2010 (2010-01-01), pages 1-13, XP055033766, ISSN: 1687-5176, DOI: 10.1155/2010/108584 page 5, left-hand column, line 28 - right-hand column, line 2	1,2,7-9, 12-14
A	US 2013/050427 A1 (CHOU HONG-LONG [TW] ET AL) 28 February 2013 (2013-02-28) paragraph [0045]	1,12,14
A	BING HAN ET AL: "Motion-segmentation-based change detection", PROCEEDINGS OF SPIE, 27 April 2007 (2007-04-27), XP055195537, ISSN: 0277-786X page 8, last paragraph	1,12,14
Y	WO 2010/087955 A1 (THOMSON LICENSING [FR]; LAI POLIN [US]; TIAN DONG [US]) 5 August 2010 (2010-08-05) figure 14 claims 7,8 page 9, line 9 - page 10, line 20	1,3-6, 11-14
Y	REDERT A: "Visualization of arbitrary-shaped 3D scenes on depth-limited 3D displays", 3D DATA PROCESSING, VISUALIZATION AND TRANSMISSION, 2004. 3DPVT 2004. PROCEEDINGS. 2ND INTERNATIONAL SYMPOSIUM ON THESSALONIKI, GREECE 6-9 SEPT. 2004, PISCATAWAY, NJ, USA,IEEE, 6 September 2004 (2004-09-06), pages 938-942, XP010725305, DOI: 10.1109/TDPVT.2004.1335416 ISBN: 978-0-7695-2223-4 page 940, left-hand column, line 3 - line 8	10

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2015/057534

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011096832 A1	28-04-2011	CN 102598674 A	18-07-2012
		EP 2491722 A2	29-08-2012
		JP 5575908 B2	20-08-2014
		JP 2013509104 A	07-03-2013
		KR 20120071407 A	02-07-2012
		TW 201130306 A	01-09-2011
		US 2011096832 A1	28-04-2011
		WO 2011050304 A2	28-04-2011
US 2013336577 A1	19-12-2013	NONE	
US 2013050427 A1	28-02-2013	TW 201310970 A	01-03-2013
		US 2013050427 A1	28-02-2013
WO 2010087955 A1	05-08-2010	CN 102356637 A	15-02-2012
		EP 2384584 A1	09-11-2011
		JP 2012516637 A	19-07-2012
		KR 20110122817 A	11-11-2011
		US 2011273529 A1	10-11-2011
		WO 2010087955 A1	05-08-2010