(19)

(10) **FI 20205528 A1**

(12) **JULKISEKSI TULLUT PATENTTIHAKEMUS**
PATENTANSÖKAN SOM BLIVIT OFFENTLIG
PATENT APPLICATION MADE AVAILABLE TO THE
PUBLIC

(71) Hakija - Sökande - Applicant

1 • **Teknologian tutkimuskeskus VTT Oy,** Vuorimiehentie 3, 02150 ESPOO, SUOMI - FINLAND, (FI)

(72) Keksijä - Uppfinnare - Inventor

1 • **VALLI, Seppo,** VTT, SUOMI - FINLAND, (FI)

(74) Asiamies - Ombud - Agent

**Laine IP Oy,** Porkkalankatu 24, 00180 Helsinki

(54) Keksinnön nimitys - Uppfinningens benämning - Title of the invention
Menetelmä ja järjestelmä 3D-havainnon mahdollistamiseksi monokulaarikuvista

Förfarande och system för att möjliggöra 3D-detektering från monokulära bilder

Method and system for supporting 3D perception from monocular images

(57) Tiivistelmä - Sammandrag - Abstract

Menetelmä ja laite syvyysdatan johtamiseksi maisemalle monokulaarisista kuvista vastaanottamalla monokulaarinen kuvadata, joka sisältää vähintään kuvaparin: ensimmäisen kuvan maisemasta ensimmäisissä valaistusolosuhteissa ja toisen kuvan maisemasta toisissa valaistusolosuhteissa, jotka poikkeavat ensimmäisistä valaistusolosuhteista. Käsitellään ensimmäisen kuvan ja toisen kuvan erotusta erotuskuvan tuottamiseksi, ja määritetään sironneen ja/tai heijastuneen valon intensiteetti maiseman osista perustuen erotuskuvaan. Sitten johdetaan syvyyskuvaus, joka käsittää syvyystasoinformaation maiseman osille perustuen mainituista maiseman osista määritettyyn sironneen ja/tai heijastuneen valon intensiteettiin.

A method and apparatus for deriving depth data for a scene from monocular images by: receiving monocular image data containing at least a pair of images: a first image of a scene in a first illumination condition and a second image of the scene in a second illumination condition different from the first illumination condition. Processing a difference of the first image and second image to produce a difference image and determining an intensity of scattered and/or reflected light from portions of the scene based on the difference image. Then deriving a depth map comprising depth level information for the portions of the scene based on the determined intensity of scattered and/or reflected light from said portions of the scene.
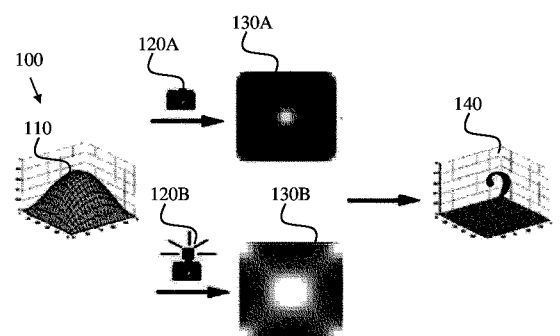
METHOD AND SYSTEM FOR SUPPORTING 3D PERCEPTION FROM MONOCULAR IMAGES

BACKGROUND

5   **[0001]**   Stereoscopic rendering based content and services are now available on television sets, in computer games, and via Augmented Reality (AR) or Virtual Reality (VR) glasses. Even modern mobile phones provide for certain types of stereoscopic rendering.

[0002]   While the applications employing stereoscopic rendering have increased, 10   there are still many hurdles involved in such offerings. They require complex stereoscopic scene capture, content production and delivery. Further, current stereoscopic rendering techniques may result in discomfort when viewing stereoscopic (S3D) displays.

[0003]   This discomfort when viewing S3D content is due at least in part to the fact that present displays force a viewer to accommodate or focus on one display surface, 15   although at the same time, the viewer's eyes are choosing image details with varying vergence. In other words the current displays force a user to assume varying levels of eye-cross, forcing a user to go cross-eyed. In contrast, within real-world viewing, focus and vergence are strongly related.

[0004]   Multiple focal plane (MFP) displays avoid this vergence-accommodation 20   conflict (VAC) by rendering each scene detail on a focal plane close to its real distance/depth. However, in addition to challenging optical arrangements for display, MFP display requires capturing depth corresponding to each image pixel. That is, MFP displays must know depth properties of a captured scene.

[0005]   Therefore, in order to avoid the problems associated with S3D displays and 25   employ MFP displays, a depth map must be formed. However, such depth maps are typically created using stereoscopic cameras or complicated methods involving neural networks. Stereoscopic cameras tend to make content production more complicated, and typically make assumptions on the orientation when viewing the content. Neural networks again increase complication, typically needing a large training set with known or measured 30   depth properties.

[0006] Further, capturing depth by stereovision is relatively inaccurate due to inaccuracies in deducting true disparity of stereo views. Red Green Blue - Depth (RGB-D) approaches are relatively expensive, consume high power, and require accurate and potentially complex calibration between cameras and depth sensors. Moreover, although RGB-D approaches may capture depth accurately, they tend to result with low spatial resolution. Depth of focus approaches are also rather complex, and for good accuracy, may require multiple scene captures with different focal lengths. On the other hand, motion parallax methods may be inaccurate, and in particular, unable to capture 3D data from a single or few image captures.

SUMMARY OF THE INVENTION

[0007] Instead of deriving depth data by capturing stereoscopic disparity from stereoscopic images, for example by applying motion estimation methods, the present invention allows for the derivation of depth data from monoscopic images using differences in shading from images captured under differing lighting. Embodiments of the invention allow for stereoscopic rendering without the need for stereoscopic image capture. Certain embodiments of the invention even provide for the derivation of depth data from images captured without the use of Red Green Blue - Depth (RGB-D) sensors. For example via sensors of a mobile phone, small consumer camera or other small consumer device.

[0008] Embodiments of the invention described herein provide for a much simpler method of deriving depth data and are able to derive scene depth for both stereoscopic 3D (S3D) and accommodative displays. Further, embodiments of the present invention are well suited for use in situations with an arbitrary stereo baseline, that is, head tilt when viewing 3D content.

[0009] The invention is defined by the features of the independent claims. Some specific embodiments are defined in the dependent claims.

[0010] According to a first aspect of the present invention, there is provided a method for deriving depth data for a scene from monocular images via shape-from-shading (SfS), the method comprising:

–	receiving monocular image data containing at least a pair of images comprising a first image of a scene in a first illumination condition and a second image of the scene in a second illumination condition different from the first illumination condition,

5	–	processing a difference of the first image and second image to produce a difference image,

–	determining an intensity of scattered and/or reflected light from portions of the scene based at least partly on the difference image,

–	deriving a depth map comprising depth level information for the portions of the scene based at least partly on the determined intensity of scattered and/or reflected

10	light from said portions of the scene.

[0011]	According to a second aspect of the present invention, there is provided an apparatus comprising at least one processing core, at least one memory including computer

15	program code, the at least one memory and the computer program code being configured to, with the at least one processing core, cause the apparatus at least to:

–	receive monocular image data containing at least a pair of images comprising a first image of a scene in a first illumination condition and a second image of the scene in a second illumination condition different from the first illumination condition,

20	–	process a difference of the first image and second image to produce a difference image,

–	determine an intensity of scattered and/or reflected light from portions of the scene based at least partly on the difference image,

–	deriving a depth map comprising depth level information for the portions of the

25	scene based at least partly on the determined intensity of scattered and/or reflected light from said portions of the scene.

BRIEF DESCRIPTION OF THE DRAWINGS

30	[0012]	FIGURES 1A and 1B illustrate processes employed by at least some embodiments of the present invention;

[0013]   FIGURE 2 shows a process in accordance with certain embodiments of the present invention;

[0014]   FIGURE 3 illustrates a process according to some embodiments of the present invention employing a serial stream of images;

5   **[0015]**   FIGURE 4 illustrates an example apparatus capable of supporting at least some embodiments of the present invention, and

[0016]   FIGURES 5A – 5F display images as processed by certain embodiments of the present invention.

10   EMBODIMENTS

[0017]   Embodiments of the present invention provide a novel approach for capturing depth using monocular images using shape-from-shading (SfS) techniques. The approach is much simpler than existing methods while still being capable of providing scene depth information for both stereoscopic 3D (S3D) and accommodative displays. Furthermore,
15   using monocular image capture of scenes as per certain embodiments of the present invention allows for the provision of arbitrary stereo baseline to accommodate head tilt, when viewing content created via such embodiments. At least some embodiments of the present invention may be realized with simple consumer devices, for example a mobile phone or consumer still or video camera.

20   **[0018]**   According to certain embodiments 3D information is captured from monocular images. In such embodiments, a scene is actively illuminated, and captures of the backscattered and/or reflected light are used to estimate distances of corresponding surfaces and pixels. For example, by processing a difference of images with and without flash, an estimation for the intensity of the scattered and reflected light is obtained. At least
25   some embodiments of the present invention serve to limit or eliminate the effect of ambient lighting when deriving 3D information.

[0019]   FIGURES 1A and 1B illustrate, broadly, processes employed by at least some embodiments of the present invention. Within Figure 1A, at least two images 130A and 130B are captured of a scene 100 comprising a three dimensional surface 110. Images
30   130A and 130B are captured in differing illumination conditions, in this instance by

employing a flash of the camera 120A and 120B. As illustrated, a single camera in the same position only in alternating states of flash can be employed to capture the two images 130A and 130B. From these two images 130A and 130B a reconstruction of the scene 140 may be performed via methods as will be outlined below. In essence, the process can be used to derive a surface which would result in the two images 130A and 130B when illuminated in a similar fashion to the illumination present during capture of the images 130A and 130B. FIGURE 1B, shows how incident light 150 is reflected from a surface 190 as both diffuse reflections 160 and a specular reflection 170.

[0020]   FIGURE 2 shows a process in accordance with certain embodiments of the present invention. As seen, a method for deriving depth data for a scene from monocular images via shape-from-shading (SfS) is outlined, the method comprising at least four steps. Within step 210 monocular image data is received containing at least a pair of images comprising a first image of a scene in a first illumination condition and a second image of the scene in a second illumination condition different from the first illumination condition. As will be outlined below, these illumination conditions can be actively varied or naturally variable. Within step 220, a difference of the first image and second image is processed to produce a difference image. The method continues within step 230 wherein an intensity of scattered and/or reflected light from portions of the scene are determined based at least partly on the difference image. Then within step 240 a depth map is derived comprising depth level information for the portions of the scene based at least partly on the determined intensity of scattered and/or reflected light from said portions of the scene.

[0021]   According to certain embodiments of the present invention the depth map comprises an image of the scene containing information relating to the distance of surfaces of scene from the capture point. In some embodiments the depth may be referred to or is analogous to a depth view, depth buffer or z-buffer. Within at least some embodiments the depth map contains values from 0 to 255. Such values may be encoded on a pixel by pixel basis or on an aggregate basis for example. Within embodiments employing video coding schemes a slightly limited range may be beneficial, leaving some headroom and footroom outside the values reserved for the digital signal. Certain depth maps use more than 8 bits per pixel, for example depth maps derived from RGB-D sensors. In such case, the applied video coding method may use correspondingly more bits per pixel and thus provide higher dynamics.

[0022] Within at least some embodiments depth level information comprises information on distances of portions of a scene from a fixed point, for example from the capture point of the image or some other reference point. As discussed this depth level information may be provided in the form of a depth map for a portion of a scene, for example for surfaces of the scene. Certain embodiments provide for compensation for motion between an instant of capture for the first image and an instant of capture for the second image. This may be accomplished, for example, by spatially compensating for a possible mismatch in captured images prior to processing the images to derive depth information.

[0023] The difference in illumination conditions as discussed above may be arrived at in a variety of ways. For example, within a first illumination condition there may be no active illumination, while in a second illumination condition there may be active illumination provided by, for example, a flash of the capturing device, such as an infrared flash. In certain embodiments this active illumination is provided within the visible light spectrum. In some embodiments active illumination is provided in the infrared spectrum or a combination of spectrums.

[0024] Some embodiments rely on active illumination provided at two separate intensity levels to provide for different illumination conditions. This varying of active illumination is less noticeable by a user as, in certain embodiments, the flash appears to be on consistently and the variance in active illumination can be adjusted such that it is less perceptible during image capture. For example, a constant fill-flash light may be employed.

[0025] Illumination may be provided without active participation of the capture device, for example if the natural illumination of the scene to be captured varies then no active illumination may be required. However, even in cases when active illumination is desired but there is natural illumination, the effect of natural or non-actively provided illumination is effectively neutralized as it should impact image pairs equally and as such would cancel out during the processes described herein.

[0026] Within certain embodiments of the present invention, the exact location of the illumination source is not necessary to derive depth data from the scene. For example, when illumination not actively provided by the capture source is utilized the exact location of other illumination sources, such as the sun, electric light, fire light or other, is not

needed to derive depth data. Even in embodiments wherein an illumination source of the capture device is utilized, the exact location of the illumination source relative to the capture point may not be necessary. Within certain embodiments the location of the illumination source is not provided but may be derived from captured image data.

[0027]   Within embodiments having a greater illumination in the second illumination condition than the first illumination condition processing a difference of the first image and second image may comprise subtracting the first image from the second image. Such a difference image may be formed based at least in part by the difference in reflections from points of the scene, for example diffuse and specular reflection.

[0028]   Certain embodiments of the present invention provide for the generation of at least one of a stereoscopic image pair and a Multiple Focal Plane (MFP) stack based at least in part on the derived depth map. Such MFP stacks may also be used to create stereoscopic image pairs. Some embodiments provide for production of an image pair or MFP stack which uses a viewer's head orientation or tilt and/or motion data to support stereoscopic disparity and motion parallax. Certain embodiments generate an MFP stack and then shift the MFP stack to obtain a stereoscopic MFP stack.

[0029]   Within some embodiments processing of the difference images comprises a square root operation. Certain embodiments further employ a step of leveling at least one of the captured, difference or processed images such that the images are adjusted to account for overly bright or dark areas. Some embodiments employ mapping operations, for example scaling operations, and/or filtering operations at some point during the process.

[0030]   Some embodiments of the present invention employ a transmitter and receiver topography whereby an image capture device captures a series of images under varying illumination conditions, potentially multiplexes said images and then transmits them to a receiver. Within the receiver the images are then processed according to methods disclosed herein to arrive at 3D data of the captured images. Such a transmitter and receiver combination could be, for example, a mobile phone running an application and a server configured to receive information from the application.

[0031]   FIGURE 3 illustrates a process according to certain embodiments of the present invention which employs a serial stream of images with alternating illumination

conditions, for example, as could be provided by a flash of a capture device which is activated every other frame of capture. Within some embodiments image data, for example a video, is captured while a flash is intermittently fired, for example, such that every other frame is illuminated.

[0032]   As in certain embodiments of the present invention, first a serial stream of images with alternating illumination conditions is received 310. Optionally, this serial stream of images is demultiplexed 320 and motion during capture is estimated and compensated for 330. Then a formation of differences between any two images in the stream is formed 340 and processed into a difference map 350, for example by taking the square root. Leveling, by adjusting bright levels backwards, may be performed 360 prior to filtering for smoothing of the depth map 370. Finally a depth map approximation is forward to a next stage of depth map generation 380, for example by taking into account another image pair from the stream.

[0033]   FIGURE 4 illustrates a device capable of supporting at least certain embodiments of the claimed invention, for example a portable device such as a mobile phone, tablet or camera. Device 400 may comprise memory 420. Memory 420 may comprise random-access memory and/or permanent memory. Memory 420 may comprise at least one RAM chip. Memory 420 may comprise solid-state, magnetic, optical and/or holographic memory, for example. Memory 420 may be at least in part accessible to processor 410. Memory 420 may be at least in part comprised in processor 410. Memory 420 may be means for storing information. Memory 420 may comprise computer instructions that processor 410 is configured to execute. When computer instructions configured to cause processor 410 to perform certain actions are stored in memory 420, and device 400 overall is configured to run under the direction of processor 410 using computer instructions from memory 420, processor 410 and/or its at least one processing core may be considered to be configured to perform said certain actions. Memory 420 may be at least in part comprised in processor 410. Memory 420 may be at least in part external to device 400 but accessible to device 400.

[0034]   Device 400 may comprise a transmitter 430. Device 400 may comprise a receiver 440. Transmitter 430 and receiver 440 may be configured to transmit and receive, respectively, information in accordance with at least one cellular or non-cellular standard. Transmitter 430 may comprise more than one transmitter. Receiver 440 may comprise

more than one receiver. Transmitter 430 and/or receiver 440 may be configured to operate in accordance with global system for mobile communication, GSM, wideband code division multiple access, WCDMA, 5G, long term evolution, LTE, IS-95, wireless local area network, WLAN, Ethernet and/or worldwide interoperability for microwave access, WiMAX, standards, for example.

[0035]   Device 400 may comprise user interface, UI, 460. UI 460 may comprise at least one of a display, a keyboard, a touchscreen, a vibrator arranged to signal to a user by causing device 400 to vibrate, a speaker and a microphone. A user may be able to operate device 400 via UI 460, for example to accept incoming telephone calls, to originate telephone calls or video calls, to browse the Internet, to manage digital files stored in memory 420 or on a cloud accessible via transmitter 430 and receiver 440, or via NFC transceiver 450, and/or to play games.

[0036]   Processor 410 may be furnished with a transmitter arranged to output information from processor 410, via electrical leads internal to device 400, to other devices comprised in device 400. Such a transmitter may comprise a serial bus transmitter arranged to, for example, output information via at least one electrical lead to memory 420 for storage therein. Alternatively to a serial bus, the transmitter may comprise a parallel bus transmitter. Likewise processor 410 may comprise a receiver arranged to receive information in processor 410, via electrical leads internal to device 400, from other devices comprised in device 400. Such a receiver may comprise a serial bus receiver arranged to; for example, receive information via at least one electrical lead from receiver 440 for processing in processor 410. Alternatively to a serial bus, the receiver may comprise a parallel bus receiver.

[0037]   Device 400 may comprise further devices not illustrated in FIGURE 4. For example, where device 400 comprises a smartphone, it may comprise at least one digital camera. Some devices 400 may comprise a back-facing camera and a front-facing camera, wherein the back-facing camera may be intended for digital photography and the front-facing camera for video telephony. Device 400 may comprise a fingerprint sensor arranged to authenticate, at least in part, a user of device 400. In some embodiments, device 400 lacks at least one device described above. For example, some devices 400 may lack a NFC transceiver 450 and/or user identity module 470.

[0038] Processor 410, memory 420, transmitter 430, receiver 440 and/or UI 460 may be interconnected by electrical leads internal to device 400 in a multitude of different ways. For example, each of the aforementioned devices may be separately connected to a master bus internal to device 400, to allow for the devices to exchange information. However, as the skilled person will appreciate, this is only one example and depending on the embodiment various ways of interconnecting at least two of the aforementioned devices may be selected without departing from the scope of the present invention.

[0039] As shown at least certain embodiments of the present invention comprise at least one processing core 410, at least one memory 420 including computer program code, the at least one memory 420 and the computer program code being configured to, with the at least one processing core 410, cause the apparatus at least to:

- receive monocular image data containing at least a pair of images comprising a first image of a scene in a first illumination condition and a second image of the scene in a second illumination condition different from the first illumination condition,
- process a difference of the first image and second image to produce a difference image,
- determine an intensity of scattered and/or reflected light from portions of the scene based at least partly on the difference image,
- deriving a depth map comprising depth level information for the portions of the scene based at least partly on the determined intensity of scattered and/or reflected light from said portions of the scene.

[0040] Certain embodiments further comprise an image sensor 450 and illuminator 470, such as the image sensor and flash of a mobile phone, the apparatus being further configured to capture the monocular image data while varying illumination conditions of the scene. In some embodiments the scene is actively illuminated in the second illumination condition by the illuminator 470 of the apparatus. 18. Certain apparatuses are configured such that within the first illumination condition, the scene is actively illuminated by a flash or the illuminator 470 at a lower intensity level than it is illuminated by a flash or illuminator 470 in the second illumination condition.

[0041] Apparatuses according to the current invention may be configured such that: there is greater illumination in the second illumination condition than the first illumination

condition and processing a difference of the first image and second image comprises subtracting the first image from the second image. They may be further configured to generate at least one of a stereoscopic image pair and a Multiple Focal Plane (MFP) stack based at least in part on the derived depth map. Still further, they may be configured to generate a Multiple Focal Plane (MFP) stack and then shift the MFP stack to obtain a stereoscopic MFP stack.

[0042]   FIGURES 5A – 5F display images as processed by certain embodiments of the present invention, including virtual viewpoints shown as printed cross-eyed stereograms. FIGURE 5A illustrates a scene captured under two different illumination conditions. One monoscopic camera is used to capture the images at slight different points in time to produce the image pair as shown. As discussed above, there are a variety of ways in which to achieve the differences in illumination. As the images are captured at different times there may have been motion of the capture means between image captures, as such, certain embodiments correct for this motion prior to further processing.

[0043]   Within FIGURE 5B the differences of the images of Figure 5A are processed to first form a difference image (a). In certain embodiments the difference image (a) is further processed by scaling, for example based on a histogram, and remapped to arrive at image (b). As the intensity of reflections decays inversely proportional to the squared distance, this scaling may be accomplished by a square root function. This image (b) may be further processed by filtering to reduce steps or coarseness to arrive at image (c).

[0044]   Surface properties generally bias reflected intensities so that dark surfaces corresponding to more diffuse and small reflections show up further away than light surfaces corresponding to less diffuse and stronger reflections. Correspondingly, a mechanism (here denoted as leveling) may be employed for preventing biasing of light and dark areas to different distances, while they are in fact at about the same depth. Figure 5C shows the effects of this leveling process as can be seen by comparing, for example, the shade differences at the doll's hair, sleeves, and chest within Figure 5C to those of image (c) of Figure 5B. In at least some embodiments this leveling is a recursive process whereby a depth map is used for the first leveling round, which in turn produces an improved depth map, used for the second leveling round, and so on, until the result converges or some threshold is met.

[0045]   FIGURE 5D shows a depth map of FIGURE 5B image (c) as broken down into three weight maps (a), (b) and (c), for example by using linear depth blending. The number of weight maps here is for illustrative purposes and corresponds to the number of focal planes desired; however this number may be increased for higher quality outputs. In certain embodiments the weigh maps sum to a value of one and therefore can be used to create focal planes which sum to the original image, as illustrated herein the original image being a normally lighted doll picture. FIGURE 5E shows the three focal planes (a), (b), (c) derived from the weight maps in Figure 5D. FIGURE 5F shows a cross-eyed stereo pair formed using the three focal planes shown in Figure 5E.

[0046]   At least some embodiments of the present invention capture orientation information during image capture, for example via a sensor within the image capture device which will sense the direction of gravitational pull. In such embodiments a "gravity stabilized" capture mode can be implemented whereby a user does not need to be concerned with regard to the orientation of the image capture device during capture. This orientation information can then be used during generation of an MFP stack or stereoscopic image pair. For example, when synthesizing disparity for an S3D stereo video or a stereoscopic MFP video, the orientation of the baseline may be decided and supported freely in the receiver. In certain instances this may be done based on the viewers head tilt. Further embodiments may support synthetic motion parallax in viewing. Certain embodiments of the present invention provide for a 3D video service which allows for shooting and viewing of stable S3D or accommodative MFP videos with less restrictive assumptions for both capture and viewing orientations.

[0047]   Within some embodiments the capture device captures orientation information, such as a gravity vector and includes that, potentially in multiplexed and/or encoded image data, during capture of images as per methods discussed above. That orientation information is then used to stabilize or adjust the captured image data such that they have a consistent orientation. This stabilization may occur at any point in the processing of the image as per the methods described herein. For example, the stabilization may occur within a capture device prior to processing or sending the captured information. In at least some embodiments this stabilization occurs after the formation of depth map information.

[0048] Certain embodiments of the present invention employ both orientation derived during image capture and orientation derived during display in order to match an orientation of a captured scene with a head tilt of a viewer of a displayed scene, potentially a scene derived from a stereoscopic image pair or MFP stack created as per methods described herein.

[0049] For purposes of display, embodiments of the present invention may employ conventional auto-stereoscopic display, a shutter glasses based S3D display, or a stereoscopic near-eye display, for example as can be implemented via a mobile phone or tablet.

[0050] Certain embodiments of the present invention provide the ability of a user to adjust the brightness of a rendered 3D image between those levels of illumination obtained during image capture.

[0051] At least some embodiments of the present invention employ a calibration pattern prior to capturing images. The known properties of the calibration pattern can then be used to improve the derivation of 3D information.

[0052] Certain embodiments of the present invention employ a step of displaying the generated at least one of a stereoscopic image pair and MFP stack. Within some embodiments data is received regarding orientation of the head of a viewer and said orientation data is used when generating at least one of a stereoscopic image pair and a Multiple Focal Plane (MFP) stack.

[0053] Within some embodiments the image data is received as multiplexed data containing alternating images under alternating illumination conditions.

[0054] At least some embodiments filter high frequency textures from the depth map. Certain embodiments employ a step of processing the difference image prior to the derivation of the depth map, for example by scaling, linearization and/or filtering of the difference image.

[0055] It is to be understood that the embodiments of the invention disclosed are not limited to the particular structures, process steps, or materials disclosed herein, but are extended to equivalents thereof as would be recognized by those ordinarily skilled in the

relevant arts. It should also be understood that terminology employed herein is used for the purpose of describing particular embodiments only and is not intended to be limiting.

[0056] Reference throughout this specification to "one embodiment" or "an embodiment" means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment.

[0057] As used herein, a plurality of items, structural elements, compositional elements, and/or materials may be presented in a common list for convenience. However, these lists should be construed as though each member of the list is individually identified as a separate and unique member. Thus, no individual member of such list should be construed as a de facto equivalent of any other member of the same list solely based on their presentation in a common group without indications to the contrary. In addition, various embodiments and example of the present invention may be referred to herein along with alternatives for the various components thereof. It is understood that such embodiments, examples, and alternatives are not to be construed as de facto equivalents of one another, but are to be considered as separate and autonomous representations of the present invention.

[0058] Furthermore, the described features, structures, or characteristics may be combined in any suitable manner in one or more embodiments. In the following description, numerous specific details are provided, such as examples of lengths, widths, shapes, etc., to provide a thorough understanding of embodiments of the invention. One skilled in the relevant art will recognize, however, that the invention can be practiced without one or more of the specific details, or with other methods, components, materials, etc. In other instances, well-known structures, materials, or operations are not shown or described in detail to avoid obscuring aspects of the invention.

[0059] While the forgoing examples are illustrative of the principles of the present invention in one or more particular applications, it will be apparent to those of ordinary skill in the art that numerous modifications in form, usage and details of implementation can be made without the exercise of inventive faculty, and without departing from the

principles and concepts of the invention. Accordingly, it is not intended that the invention be limited, except as by the claims set forth below.

[0060] The verbs "to comprise" and "to include" are used in this document as open limitations that neither exclude nor require the existence of also un-recited features. The features recited in depending claims are mutually freely combinable unless otherwise explicitly stated. Furthermore, it is to be understood that the use of "a" or "an", i.e. a singular form, throughout this document does not exclude a plurality.

CLAIMS:

1. A method for deriving depth data for a scene from monocular images via shape-from-shading (SfS), the method comprising:

5   – receiving monocular image data containing at least a pair of images comprising a first image of a scene in a first illumination condition and a second image of the scene in a second illumination condition different from the first illumination condition,

        – processing a difference of the first image and second image to produce a difference
10   image,

        – determining an intensity of scattered and/or reflected light from portions of the scene based at least partly on the difference image, and

        – deriving a depth map comprising depth level information for the portions of the scene based at least partly on the determined intensity of scattered and/or reflected
15   light from said portions of the scene.

2. The method of claim 1, wherein in the second illumination condition, the scene is actively illuminated, for example by a flash, such as an infrared flash.

20   3. The method of any preceding claim, wherein in the first illumination condition there is no active illumination.

4. The method of any preceding claim, wherein there is greater illumination in the second illumination condition than the first illumination condition and processing a difference of
25   the first image and second image comprises subtracting the first image from the second image.

5. The method of any preceding claim, further comprising the step of generating at least one of a stereoscopic image pair and a Multiple Focal Plane (MFP) stack based at least in
30   part on the derived depth map.

6. The method of any preceding claim, further comprising the step of generating a Multiple Focal Plane (MFP) stack and then shifting the MFP stack to obtain a stereoscopic MFP stack.

7. The method of any preceding claim, wherein the image data further comprises orientation information captured during image capture, such as orientation information relative to the gravity direction.

8. The method of claim 7, further comprising the step of displaying the stereoscopic image pair or MFP stack wherein orientation information is captured during said display, such as orientation information relative to the gravity direction, the orientation information being used during generation of the stereoscopic image pair or MFP stack.

9. The method of any preceding claim wherein the image data consists of image data captured while a flash is intermittently fired, for example, such that every other frame is illuminated.

10. The method of any preceding claim, further comprising the step of compensating for motion between an instant of capture for the first image and an instant of capture for the second image.

11. The method of any preceding claim, wherein in the first illumination condition, the scene is actively illuminated by a flash at a lower intensity level than it is illuminated by a flash in the second illumination condition.

12. An apparatus comprising at least one processing core, at least one memory including computer program code, the at least one memory and the computer program code being configured to, with the at least one processing core, cause the apparatus at least to:
  - receive monocular image data containing at least a pair of images comprising a first image of a scene in a first illumination condition and a second image of the scene in a second illumination condition different from the first illumination condition,
  - process a difference of the first image and second image to produce a difference image,

    – determine an intensity of scattered and/or reflected light from portions of the scene based on the difference image, and

    – deriving a depth map comprising depth level information for the portions of the scene based on the determined intensity of scattered and/or reflected light from said portions of the scene.

13. The apparatus of claim 12, further comprising an image sensor and illuminator, such as the image sensor and flash of a mobile phone, the apparatus being further configured to capture the monocular image data while varying illumination conditions of the scene.

14. The apparatus of claim 12 or 13 wherein in the second illumination condition, the scene is actively illuminated.

15. The apparatus of any of claims 12 – 14, wherein there is greater illumination in the second illumination condition than the first illumination condition and processing a difference of the first image and second image comprises subtracting the first image from the second image.

16. The apparatus of any of claims 12 – 15, wherein the apparatus is further configured to generate at least one of a stereoscopic image pair and a Multiple Focal Plane (MFP) stack based at least in part on the derived depth map.

17. The apparatus of any of claims 12 – 16, wherein the apparatus is further configured to generate a Multiple Focal Plane (MFP) stack and then shift the MFP stack to obtain a stereoscopic MFP stack.

18. The apparatus of any of claims 12 – 17, wherein the image data further comprises orientation information captured during image capture, such as orientation information relative to the gravity direction.

19. The apparatus of any of claim 18, further comprising the step of displaying the stereoscopic image pair or MFP stack wherein orientation information is captured during said display, such as orientation information relative to the gravity direction, the

orientation information being used during generation of the stereoscopic image pair or MFP stack.

20. The apparatus of any of claims 12 – 19, wherein in the first illumination condition, the scene is actively illuminated by a flash at a lower intensity level than it is illuminated by a flash in the second illumination condition.

21. A non-transitory computer readable medium having stored thereon a set of computer readable instructions that, when executed by at least one processor, cause an apparatus to at least perform a method in accordance with at least one of claims 1 – 11.

1. Menetelmä syvyysdatan johtamiseksi maisemalle monokulaarisista kuvista varjomuodon (shape-from-shading, SfS) kautta, missä menetelmä käsittää sen, että:

   – vastaanotetaan monokulaarinen kuvadata, joka sisältää vähintään kuvaparin, joka käsittää ensimmäisen kuvan maisemasta ensimmäisissä valaistusolosuhteissa sekä toisen kuvan maisemasta toisissa valaistusolosuhteissa, jotka poikkeavat ensimmäisistä valaistusolosuhteista,

   – käsitellään ensimmäisen kuvan ja toisen kuvan erotusta erotuskuvan tuottamiseksi,

   – määritetään sironneen ja/tai heijastuneen valon intensiteetti maiseman osista perustuen ainakin osittain erotuskuvaan, ja

   – johdetaan syvyyskuvaus, joka käsittää informaation maiseman osien etäisyyksistä kiinteästä pisteestä,

   tunnettu siitä, että syvyyskuvaus johdetaan perustuen ainakin osittain mainituista maiseman osista määritettyyn sironneen ja/tai heijastuneen valon intensiteettiin.

2. Patenttivaatimuksen 1 menetelmä, jossa toisissa valaistusolosuhteissa maisemaa valaistaan aktiivisesti, esimerkiksi salamavalolla, kuten infrapunasalamalla.

3. Jommankumman edeltävän patenttivaatimuksen menetelmä, jossa ensimmäisissä valaistusolosuhteissa ei ole aktiivista valaistusta.

4. Jonkin edeltävän patenttivaatimuksen menetelmä, jossa toisissa valaistusolosuhteissa on suurempi valaistus kuin ensimmäisissä valaistusolosuhteissa ja jossa ensimmäisen kuvan ja toisen kuvan erotuksen käsittely käsittää sen, että vähennetään ensimmäinen kuva toisesta kuvasta.

5. Jonkin edeltävän patenttivaatimuksen menetelmä käsittäen lisäksi vaiheen, jossa generoidaan ainakin jompikumpi stereoskooppisesta kuvaparista ja monen polttotason (Multiple Focal Plane, MFP) pinosta perustuen ainakin osittain johdettuun syvyyskuvaukseen.

6. Jonkin edeltävän patenttivaatimuksen menetelmä käsittäen lisäksi vaiheen, jossa generoidaan monen polttotason (MFP) pino ja sitten siirretään MFP-pinoa stereoskooppisen MFP-pinon saamiseksi.

5

7. Jonkin edeltävän patenttivaatimuksen menetelmä, jossa kuvadata käsittää lisäksi suuntainformaation, joka on kaapattu kuvan kaappauksen aikana, kuten esimerkiksi suuntainformaation suhteessa painovoiman suuntaan.

10   8. Patenttivaatimuksen 7 menetelmä käsittäen lisäksi vaiheen, jossa näytetään stereoskooppinen kuvapari tai MFP-pino, jossa kuvadata käsittää lisäksi suuntainformaation, joka kaapataan mainitun näyttämisen aikana, kuten esimerkiksi suuntainformaatio suhteessa painovoiman suuntaan, missä suuntainformaatiota käytetään stereoskooppisen kuvaparin tai MFP-pinon generoimisen aikana siten, että kaapatun

15   maiseman suuntaus täsmää näytetyn maiseman katsojan pään kallistumaan.

9. Jonkin edeltävän patenttivaatimuksen menetelmä, jossa kuvadata koostuu kuvadatasta, joka kaapataan samalla, kun salamavalo ajoittaisesti laukaistaan, esimerkiksi siten että joka toinen yksittäinen kuva valaistaan.

20

10. Jonkin edeltävän patenttivaatimuksen menetelmä käsittäen lisäksi vaiheen, jossa kompensoidaan liikettä ensimmäisen kuvan kaappaushetken ja toisen kuvan kaappaushetken välillä.

25   11. Jonkin edeltävän patenttivaatimuksen menetelmä, jossa ensimmäisissä valaistusolosuhteissa maisemaa valaistaan aktiivisesti salamavalolla matalammalla intensiteettitasolla, kuin sitä valaistaan salamavalolla toisissa valaistusolosuhteissa.

12. Laite käsittäen vähintään yhden prosessointiytimen ja vähintään yhden muistin sisältäen tietokoneohjelmakoodia, missä mainitut vähintään yksi muisti ja tietokoneohjelmakoodi konfiguroidaan mainitun vähintään yhden prosessointiytimen avulla saamaan laitteen vähintään:

- vastaanottamaan monokulaarista kuvadataa, joka sisältää vähintään kuvaparin, joka käsittää ensimmäisen kuvan maisemasta ensimmäisissä valaistusolosuhteissa sekä toisen kuvan maisemasta toisissa valaistusolosuhteissa, jotka poikkeavat ensimmäisistä valaistusolosuhteista,

- käsittelemään ensimmäisen kuvan ja toisen kuvan erotuksen erotuskuvan tuottamiseksi,

- määrittämään sironneen ja/tai heijastuneen valon intensiteetin maiseman osista perustuen erotuskuvaan, ja

- johtamaan syvyyskuvauksen, joka käsittää informaation maiseman osien etäisyyksistä kiinteästä pisteestä,

   tunnettu siitä, että syvyyskuvaus on johdettu perustuen ainakin osittain mainituista maiseman osista määritettyyn sironneen ja/tai heijastuneen valon intensiteettiin.

13. Patenttivaatimuksen 12 laite käsittäen lisäksi kuva-anturin ja valonlähteen, kuten matkapuhelimen kuva-anturin ja salamavalon, missä laite konfiguroidaan lisäksi kaappaamaan monokulaarisen kuvadatan samalla, kun se vaihtelee maiseman valaistusolosuhteita.

14. Patenttivaatimuksen 12 tai 13 laite, jossa toisissa valaistusolosuhteissa maisemaa valaistaan aktiivisesti.

15. Jonkin patenttivaatimuksista 12–14 laite, jossa toisissa valaistusolosuhteissa on suurempi valaistus kuin ensimmäisissä valaistusolosuhteissa ja jossa ensimmäisen kuvan ja toisen kuvan erotuksen käsittely käsittää sen, että vähennetään ensimmäinen kuva toisesta kuvasta.

16. Jonkin patenttivaatimuksista 12–15 laite, jossa laite on lisäksi konfiguroitu generoimaan vähintään yhden stereoskooppisen kuvaparin sekä monen polttotason (Multiple Focal Plane, MFP) pinon perustuen ainakin osittain johdettuun syvyyskuvaukseen.

17. Jonkin patenttivaatimuksista 12–16 laite, jossa laite on lisäksi konfiguroitu generoimaan monen polttotason (MFP) pinon ja sitten siirtämään MFP-pinoa stereoskooppisen MFP-pinon saamiseksi.

18. Jonkin patenttivaatimuksista 12–17 laite, jossa kuvadata käsittää lisäksi suuntainformaation, joka on kaapattu kuvan kaappauksen aikana, kuten esimerkiksi suuntainformaation suhteessa painovoiman suuntaan.

19. Patenttivaatimuksen 18 laite käsittäen lisäksi vaiheen, jossa näytetään stereoskooppinen kuvapari tai MFP-pino, jossa kuvadata käsittää lisäksi suuntainformaation, joka kaapataan mainitun näyttämisen aikana, kuten esimerkiksi suuntainformaatio suhteessa painovoiman suuntaan, missä suuntainformaatiota käytetään stereoskooppisen kuvaparin tai MFP-pinon generoimisen aikana siten, että kaapatun maiseman suuntaus täsmää näytetyn maiseman katsojan pään kallistumaan.

20. Jonkin patenttivaatimuksista 12–19 laite, jossa ensimmäisissä valaistusolosuhteissa maisemaa valaistaan aktiivisesti salamavalolla matalammalla intensiteetillä, kuin sitä valaistaan salamavalolla toisissa valaistusolosuhteissa.

21. Pysyvä tietokoneella luettava media, johon on tallennettu joukko tietokoneella luettavia komentoja, jotka suoritettaessa vähintään yhdessä prosessorissa saavat laitteen ainakin suorittamaan ainakin yhden patenttivaatimuksista 1–11 mukaisen menetelmän.
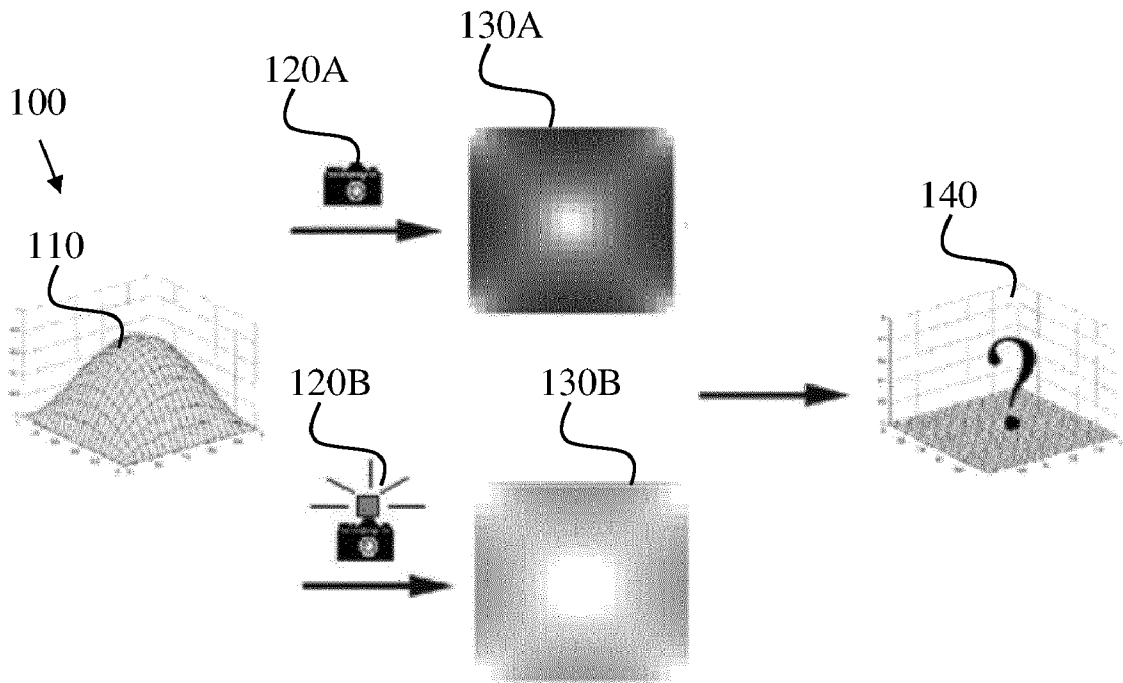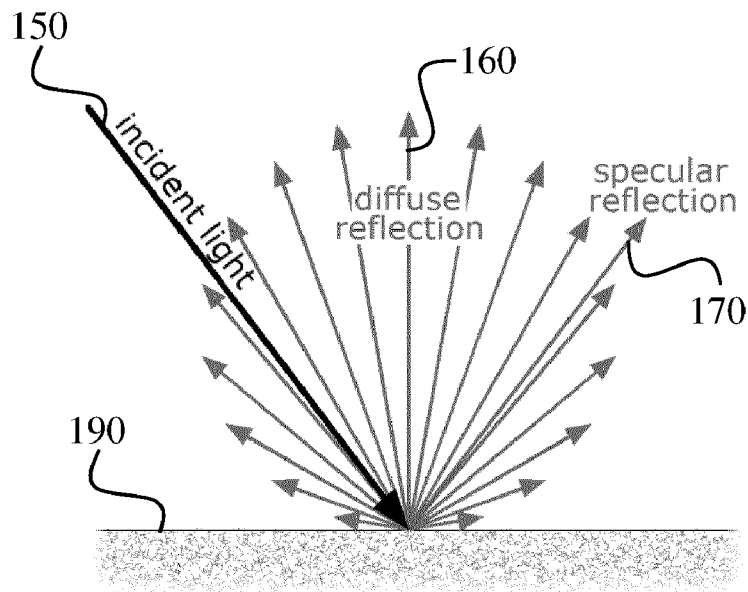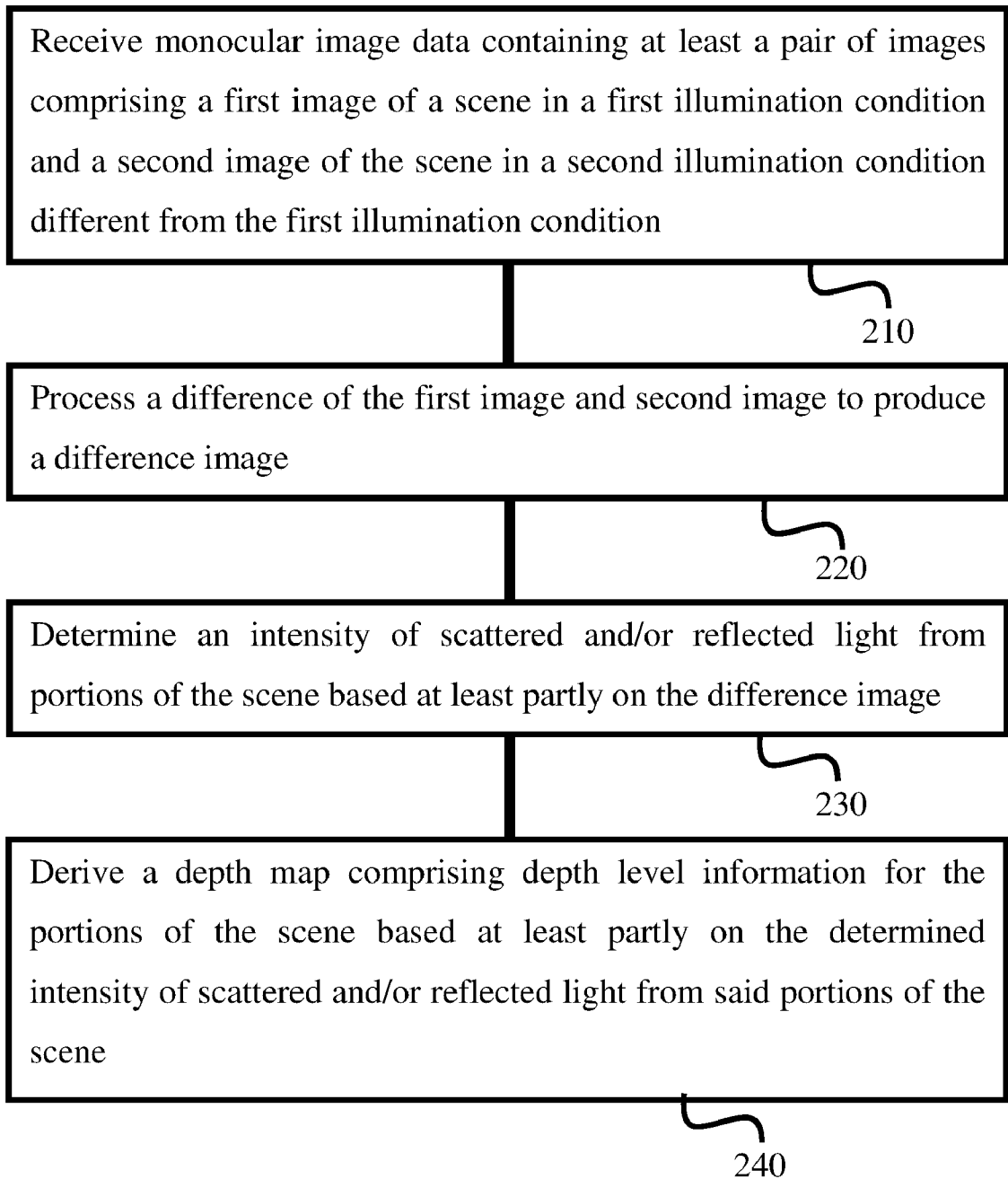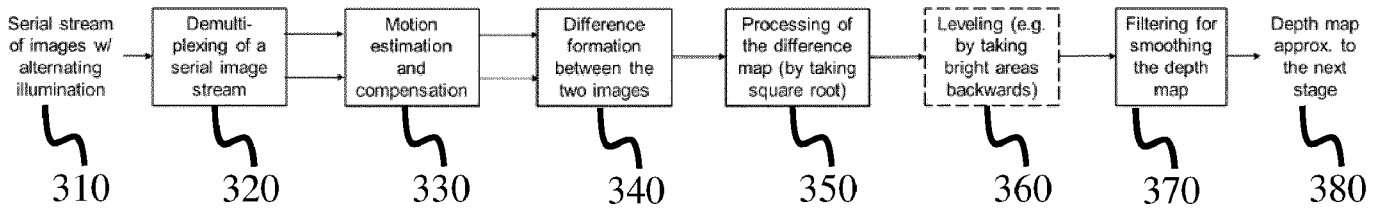
FIGURE 1A

FIGURE 1B

Receive monocular image data containing at least a pair of images comprising a first image of a scene in a first illumination condition and a second image of the scene in a second illumination condition different from the first illumination condition

210

Process a difference of the first image and second image to produce a difference image

220

Determine an intensity of scattered and/or reflected light from portions of the scene based at least partly on the difference image

230

Derive a depth map comprising depth level information for the portions of the scene based at least partly on the determined intensity of scattered and/or reflected light from said portions of the scene

240

FIGURE 2

Serial stream of images w/ alternating illumination	Demulti-plexing of a serial image stream	Motion estimation and compensation	Difference formation between the two images	Processing of the difference map (by taking square root)	Leveling (e.g. by taking bright areas backwards)	Filtering for smoothing the depth map	Depth map approx. to the next stage
310	320	330	340	350	360	370	380

FIGURE 3

400

__430__

__440__

__460__

__470__

__410__

__420__

__450__

FIGURE 4

(a)   (b)

FIGURE 5A

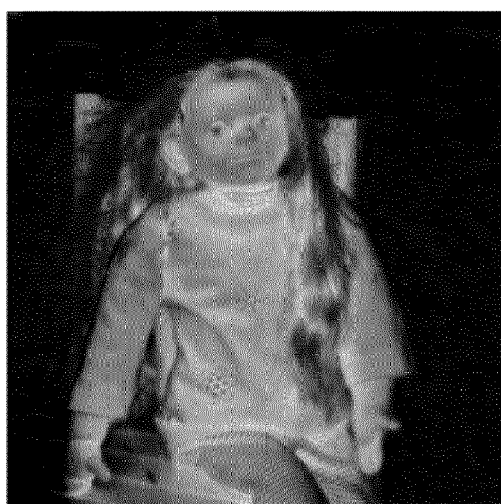(a)   (b)   (c)

FIGURE 5B
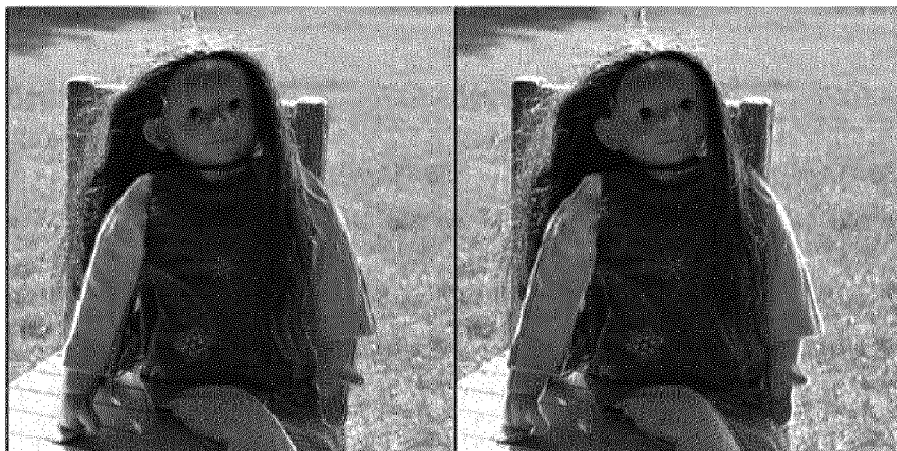

FIGURE 5C
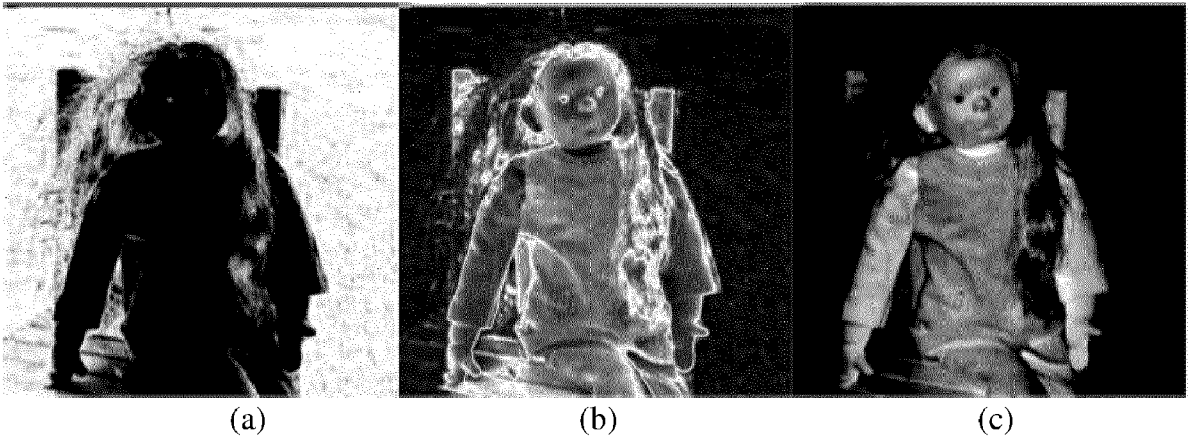
FIGURE 5D
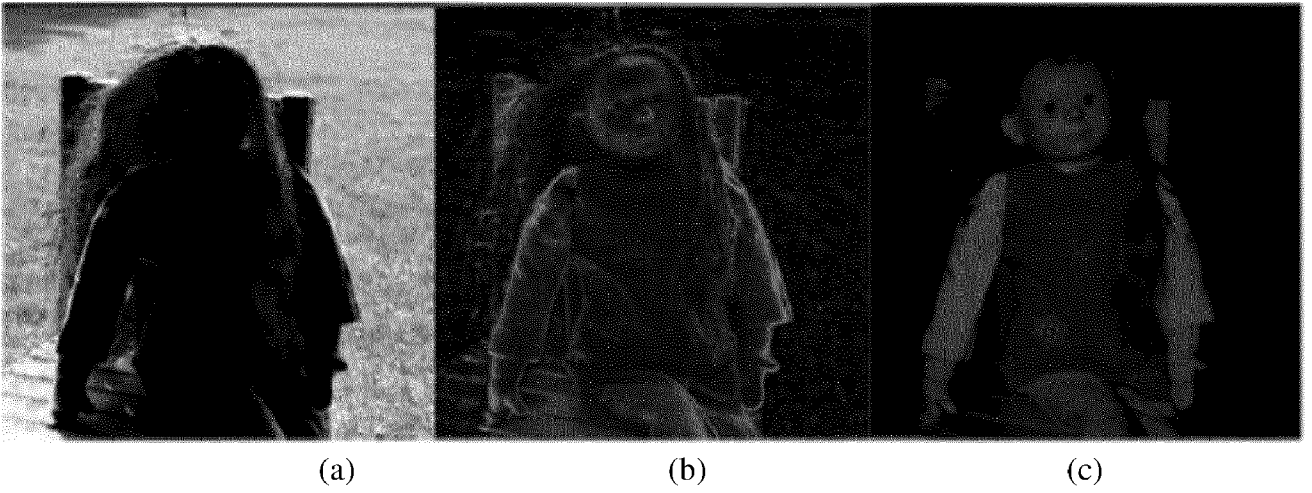
FIGURE 5E

FIGURE 5F

FINNISH PATENT AND REGISTRATION OFFICE

SEARCH REPORT

PATENT APPLICATION No.	CLASSIFICATION	
20205528	**IPC** ***G06T 7/586*** (2017.01) ***H04N 13/261*** (2018.01) ***H04N 13/207*** (2018.01) ***G02B 30/52*** (2020.01) ***G01C 11/06*** (2006.01) ***G01B 11/24*** (2006.01) ***H04N 13/264*** (2018.01)	**CPC** **G06T 7/586** **H04N 13/261** **H04N 13/207** **G02B 30/52** **G01C 11/06** **G01B 11/24** **H04N 13/264**

PATENT CLASSES SEARCHED (classification systems and classes)		
IPC: G06T, H04N, G02B, G01C, G01B		

DATABASES CONSULTED DURING THE SEARCH		
EPODOC, EPO-Internal full-text databases, Full-text translation databases from Asian languages, WPIAP, XP3GPP, XPAIP, XPESP, XPETSI, XPI3E, XPIEE, XPIETF, XPIOP, XPIPCOM, XPJPEG, XPMISC, XPOAC, XPRD, XPTK, COMPDX, INSPEC, TDB, Internet		

DOCUMENTS CONSIDERED TO BE RELEVANT

Category*)	Bibliographic data on the document and relevant passages	Relevant to claims
X	US 2004212725 A1 (RASKAR RAMESH [US]) 28 October 2004 (28.10.2004) paragraphs [0040], [0046], [0049], [0059], [0067]-[0069], [0074], [0080], [0218]-[0236], [0242]-[0243]; Fig. 3; claim 41	1-21
X	KIM, C. et al. 'Structured Light Based Depth Edge Detection for Object Shape Recovery.' Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEExplore IEEE [online], 2005-09-21, [retrieved on 2020-12-10]. Retrieved from <https://ieeexplore.ieee.org/abstract/document/1565416>, <DOI:10.1109/CVPR.2005.536> abstract, section 1.1; Fig. 1	1, 12, 21

Continued on the next sheet ☒

*) X Document indicating that the invention is not novel or does not involve an inventive step with respect to the state of the art.
Y Document indicating that the invention does not involve an inventive step with respect to the state of the art if combined with one or more other documents in the same category.
A Document representing the general state of the art.

O Document referring to disclosure through lecture, use or other non-written means.
P Document published prior to the filing date but not prior to the earliest priority date.
T Document published after the filing date or priority date and illustrating the principle or theory underlying the invention.
E Earlier patent or utility model application that either is Finnish or designates Finland published on or after the filing date (priority date).
D Document that is mentioned in the application.
L Document which may throw doubts on priority claim(s), is cited to establish the publication date of another citation or is referred to for some other reason.

& Document member of the same patent family.

This document has been electronically signed.   **Further information given in the annex** ☐

Date	Patent Examiner
17.12.2020	Kai Willner **Telephone** 029 509 5000

PATENT APPLICATION No.
20205528

DOCUMENTS CONSIDERED TO BE RELEVANT, CONTINUED		
Category*)	Bibliographic data on the document and relevant passages	Relevant to claims
A	WANG, J. et al. 'Stereoscopic Dark Flash for Low-light Photography.' 2019 IEEE International Conference on Computational Photography (ICCP). IEEExplore IEEE [online], 2019-05-15, [retrieved on 2020-12-10]. Retrieved from <https://ieeexplore.ieee.org/abstract/document/8747337>, <DOI:10.1109/ICCPHOT.2019.8747337> section 1	1-21
A	WO 2020010018 A1 (PCMS HOLDINGS INC [US]) 09 January 2020 (09.01.2020) paragraph [0243]	1-21