

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
3 February 2011 (03.02.2011)

(10) International Publication Number  
**WO 2011/014419 A1**

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

(21) International Application Number:  
PCT/US20 10/043022

(22) International Filing Date:  
23 July 2010 (23.07.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
61/230,131 31 July 2009 (31.07.2009) US

(71) Applicant (for all designated States except US): **3DMEDIA CORPORATION** [US/US]; 1005 Slater Road, Suite 300, Durham, NC 27517 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MCNAMER, Michael** [US/US]; 21 Deerfield Road, Apex, NC 27523 (US). **ROBERS, Marshall** [US/US]; 3401 Meadowrun Drive, Durham, NC 27707 (US). **MARKAS, Tassos** [US/

US]; 10399 Holt, Chapel Hill, NC 275 17 (US). **HURST, Jason, Paul** [US/US]; 205 Drakeford Drive, Cary, NC 275 13 (US).

(74) Agent: **OLIVE, Bentley, J.**; Olive Law Group, PLLC, 130 Edinburgh South Drive, Suite 107, Cary, NC 275 11 (US).

(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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,

[Continued on next page]

(54) Title: METHODS, SYSTEMS, AND COMPUTER-READABLE STORAGE MEDIA FOR CREATING THREE-DIMENSIONAL (3D) IMAGES OF A SCENE

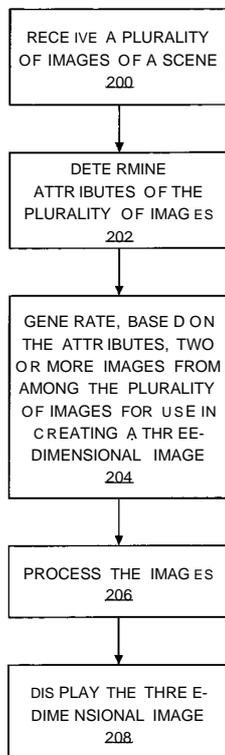


Figure 2

(57) Abstract: Methods, systems, and computer program products for generating three-dimensional images of a scene are disclosed herein. According to one aspect, a method includes receiving a plurality of images of a scene. The method also includes determining attributes of the plurality of images. Further, the method includes determining, based on the attributes, a pair of images from among the plurality of images for use in creating a three-dimensional image. A user can, by use of the subject matter disclosed herein, use an image capture device for capturing a plurality of different images of the same scene and for converting the images into a three-dimensional, or stereoscopic, image of the scene. The subject matter disclosed herein includes a three-dimensional conversion process. The conversion process can include identification of suitable pairs of images, registration, rectification, color correction, transformation, depth adjustment, and motion detection and removal.

WO 2011/014419 A1

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, **Published:**  
EE, ES, **FI**, FR, GB, GR, HR, HU, IE, IS, **IT**, **LT**, LU,  
LV, MC, MK, MT, NL, NO, PL, **PT**, RO, SE, SI, SK, ~ *m<sup>th</sup> international search report (Art. 21(3))*  
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
GW, ML, MR, NE, SN, TD, TG).

## DESCRIPTION

METHODS, SYSTEMS, AND COMPUTER-READABLE STORAGE MEDIA FOR  
CREATING THREE-DIMENSIONAL (3D) IMAGES OF A SCENE

## CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. provisional patent application number 61/230,131, filed July 31, 2009, the disclosure of which is incorporated herein by reference in its entirety. The disclosures of the following U.S. provisional patent applications, commonly owned and simultaneously filed July 31, 2009, are all incorporated by reference in their entirety: U.S. provisional patent application number 61/230,133; and U.S. provisional patent application number 61/230,138.

## TECHNICAL FIELD

[0002] The subject matter disclosed herein relates to generating an image of a scene. In particular, the subject matter disclosed herein relates to methods, systems, and computer-readable storage media for creating three-dimensional images of a scene.

## BACKGROUND

[0003] Stereoscopic, or three-dimensional, imagery is based on the principle of human vision. Two separate detectors detect the same object or objects in a scene from slightly different positions and/or angles and project them onto two planes. The resulting images are transferred to a processor which combines them and gives the perception of the third dimension, i.e. depth, to a scene.

[0004] Many techniques of viewing stereoscopic images have been developed and include the use of colored or polarizing filters to separate the two images, temporal selection by successive transmission of images using a shutter arrangement, or physical separation of the images in the viewer and projecting them separately to each eye. In addition, display devices have been developed recently that are well-suited for displaying stereoscopic images. For example, such

display devices include digital still cameras, personal computers, digital picture frames, set-top boxes, high-definition televisions (HDTVs), and the like.

[0005] The use of digital image capture devices, such as digital still cameras, digital camcorders (or video cameras), and phones with built-in cameras, for use in capturing digital images has become widespread and popular. Because images captured using these devices are in a digital format, the images can be easily distributed and edited. For example, the digital images can be easily distributed over networks, such as the Internet. In addition, the digital images can be edited by use of suitable software on the image capture device or a personal computer.

[0006] Digital images captured using conventional image capture devices are two-dimensional. It is desirable to provide methods and systems for using conventional devices for creating three-dimensional images.

#### SUMMARY

[0007] Methods, systems, and computer program products for creating three-dimensional images of a scene are disclosed herein. According to one aspect, a method includes receiving a plurality of images of a scene. The method also includes determining attributes of the plurality of images. Further, the method includes determining, based on the attributes of all captured images, a pair of images from among the plurality of images for use in creating a three-dimensional image.

[0008] According to another aspect, a user can, by use of the subject matter disclosed herein, use an image capture device for capturing a plurality of different images of the same scene and for creating a three-dimensional, or stereoscopic, image of the scene. The subject matter disclosed herein includes a process for creating three-dimensional images. The generation process can include identification of suitable pairs of images, registration, rectification, color correction, transformation, depth adjustment, and motion detection and removal. The functions of the subject matter disclosed herein can be implemented in hardware and/or software that can be executed on an image capture device or a suitable display device. For example, the functions can be implemented using a digital still camera, a personal computer, a digital picture frame, a set-top box, an HDTV, and the like.

[0009] According to another aspect, a system for creating a three-dimensional image of a scene is disclosed. The system may include a memory having stored therein computer-executable instructions. The system may also include a computer processor that executes the computer-

executable instructions. Further, the system may include an image generator configured to receive a plurality of images of a scene. The image generator may also be configured to determine attributes of the plurality of images. The image generator may also be configured to determine, based on the attributes, a pair of images from among the plurality of images for use in generating a three-dimensional image.

[0010] According to another aspect, the pair of images comprises different perspective views of the scene.

[0011] According to another aspect, the image generator may be configured to compare color attributes of the images; determine images in which the color attributes are within a predetermined color threshold; and modify the images within the predetermined color threshold for creating the three-dimensional image.

[0012] According to another aspect, the image generator is configured to: identify similar objects in the plurality of images; and compare images including the similar objects for occlusion.

[0013] According to another aspect, the image generator is configured to: identify similar objects in the plurality of images; and compare vertical displacement of the similar objects with respect to a predetermined threshold level.

[0014] According to another aspect, the image generator is configured to compare color differences between images.

[0015] According to another aspect, the image generator is configured to apply horizontal and vertical edge line analysis to images.

[0016] According to another aspect, the image generator is configured to: apply a Hough transform to identify lines within the images; and determine perspective view changes between the images based on the identified lines.

[0017] According to another aspect, the three-dimensional image includes a stereoscopic pair representing a left view image and a right view image.

[0018] According to another aspect, the image generator is configured to automatically determine the left view image and the right view image for creating the stereoscopic pair.

[0019] According to another aspect, the image generator is configured to: receive the left view image and the right view image from different sources; and automatically create the three-dimensional image using the left view image and the right view image.

[0020] According to another aspect, the image generator is configured to determine interest points within the left view image and the right view image for one or more of rectification and registration.

[0021] According to another aspect, the image generator is configured to: perform horizontal and vertical edge detection; filter for strong edges of a minimum length; and identify crossing points of the strong edges.

[0022] According to another aspect, the image generator is configured to: determine a parallax disparity between the left view image and the right view image; determine whether the parallax disparity meets a predetermined criteria; and, if the parallax disparity does not meet the predetermined criteria, adjust an attribute of at least one pixel in one of the left view image and the right view image such that the parallax disparity meets the predetermined criteria.

[0023] According to another aspect, the image generator is configured to crop one of the left view image and the right view image.

[0024] According to another aspect, the image generator is configured to: determine a parallax disparity between at least a portion of an object in the left view image and the right view image; determine whether the parallax disparity is greater than a predetermined threshold level; and, if the parallax disparity is greater than the predetermined threshold level, remove the object from the left view image and the right view image.

[0025] According to another aspect, the system further includes an image capture device for capturing the plurality of images of the scene.

[0026] According to another aspect, the image capture device is one of a digital still camera, a video camera, a mobile phone, and a smart phone for capturing the plurality of images of the scene.

[0027] According to another aspect, the image generator is configured to generate a three-dimensional image of the scene using the pair of images.

[0028] According to another aspect, the image generator is configured to implement: one or more of registration, rectification, color correction, matching edges of the pair of images, transformation, depth adjustment, motion detection, and removal of moving objects.

[0029] According to another aspect, the image generator is configured to display the three-dimensional image on a suitable three-dimensional image display.

[0030] According to another aspect, the image generator is configured to display the three-dimensional image on a device such as a digital still camera, a computer, a video camera, a digital picture frame, a set-top box, phone, or a high-definition television.

[0031] According to an aspect, a system for generating a three-dimensional video of a scene is disclosed. The system may include a memory having stored therein computer-executable instructions. The system may also include a computer processor that executes the computer-executable instructions. Further, the system may include an image generator configured to: receive a video sequence comprising a plurality of still images of a scene; determine attributes of the plurality of still images; and determine, based on the attributes, a plurality of stereoscopic image pairs from among the plurality of still images for use in generating a three-dimensional video.

[0032] According to another aspect, the computer processor and memory are configured to adjust at least one of a parallax and perspective of each image pair such that at least one of a predetermined orientation and depth results.

[0033] According to another aspect, the computer processor and memory are configured to sequence stereoscopic image pairs in the three-dimensional video according to an order of the video sequence of the still images.

[0034] According to another aspect, the computer processor and memory are configured to compress the image pairs of the three-dimensional video.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0035] The foregoing summary, as well as the following detailed description of various embodiments, is better understood when read in conjunction with the appended drawings. For the purposes of illustration, there is shown in the drawings exemplary embodiments; however, the invention is not limited to the specific methods and instrumentalities disclosed. In the drawings:

[0036] Figure 1 is a block diagram of an exemplary device for creating three-dimensional images of a scene according to embodiments of the present invention;

[0037] Figure 2 is a flow chart of an exemplary method for creating a three-dimensional image of a scene using the device shown in Figure 1 or any other suitable device described herein in accordance with embodiments of the present invention;

[0038] Figures 3A - 3D are a flow chart of an exemplary method for creating a three-dimensional image of a scene in accordance with embodiments of the present invention;

[0039] Figure 4A is a front view of a user moving between positions for capturing different images using a camera in accordance with embodiments of the present invention;

[0040] Figure 4B is a front view of a user moving between positions for capturing images using a camera in accordance with embodiments of the present invention;

[0041] Figure 5 is a flow chart of an exemplary method for a preliminary, quick analysis to pre-screen whether an image pair may be a valid stereoscopic pair in accordance with embodiments of the present invention;

[0042] Figures 6A - 6C are a flow chart of an exemplary method for edge-based analytics and matching for image correspondence, determination of right/left image, and camera toe-in/parallel plane configuration according to embodiments of the present invention;

[0043] Figure 7 is a graphical depiction of an edge detection example in accordance with embodiments of the present invention;

[0044] Figures 8A and 8B are a flow chart of an exemplary method for determining whether an image pair is a valid stereoscopic pair and which image is left and right according to embodiments of the present invention;

[0045] Figure 9 is a diagram depicting a technique for identifying corresponding points in left and right view images in accordance with embodiments of the present invention;

[0046] Figure 10 is a flow chart of an exemplary method for determining pixel disparities according to embodiments of the present invention;

[0047] Figure 11 is a flow chart of an exemplary method for adjusting parallax of segmented, moving objects according to embodiments of the present invention;

[0048] Figure 12 is an exemplary diagram of a method for adjusting parallax of moving, segmented objects according to embodiments of the present invention;

[0049] Figures 13A, 13B, and 13C illustrate an exemplary process for disparity interpolation according to embodiments of the present invention

[0050] Figure 14 is a flow chart of an exemplary method for adding/removing objects from a single image according to embodiments of the present invention;

[0051] Figure 15 is an exemplary diagram of a process for adding/removing objects from a single image according to embodiments of the present invention;

[0052] Figure 16 illustrates an exemplary process for creating three-dimensional still images from a standard two-dimensional video sequence by identifying stereoscopic pairs in accordance with embodiments of the present invention;

[0053] Figure 17 illustrates an exemplary process for creating three-dimensional video from a standard two-dimensional video sequence according to embodiments of the present invention;

[0054] Figure 18 illustrates an exemplary process of creating three-dimensional video with changing parallax and no translational motion from a standard two-dimensional video sequence in accordance with an embodiment of the subject matter disclosed herein; and

[0055] Figure 19 illustrates an exemplary environment for implementing various aspects of the subject matter disclosed herein.

### **DETAILED DESCRIPTION**

[0056] The subject matter of the present invention is described with specificity to meet statutory requirements. However, the description itself is not intended to limit the scope of this patent. Rather, the inventors have contemplated that the claimed subject matter might also be embodied in other ways, to include different steps or elements similar to the ones described in this document, in conjunction with other present or future technologies. Moreover, although the term "step" may be used herein to connote different aspects of methods employed, the term should not be interpreted as implying any particular order among or between various steps herein disclosed unless and except when the order of individual steps is explicitly described.

[0057] Methods, systems, and computer program products for creating one or more three-dimensional images of a scene are disclosed herein. The three-dimensional images can be viewed or displayed on a stereoscopic display. The three-dimensional images may also be viewed or displayed on any other display capable of presenting three-dimensional images to a person using other suitable equipment, such as, but not limited to, three-dimensional glasses. In addition, the functions and methods described herein may be implemented on a device capable of capturing still images, displaying three-dimensional images, and executing computer executable instructions on a processor. The device may be, for example, a digital still camera, a video camera (or camcorder), a personal computer, a digital picture frame, a set-top box, an HDTV, a phone, or the like. Such devices may be capable of presenting three-dimensional images to a person without additional equipment, or if used in combination with other suitable equipment such as three-dimensional glasses. The functions of the device may include methods for rectifying and registering at least two images, matching the color and edges of the images, identifying moving objects, removing or adding moving objects from or to the images to equalize them, altering a perceived depth of objects, and

any final display-specific transformation to create a single, high-quality three-dimensional image. The techniques described herein may be applied to still-captured images and video images, which can be thought of as a series of images; hence for the purpose of generalization the majority of the description herein is limited to still-captured image processing.

[0058] It should be noted that any of the processes and steps described herein may be implemented in an automated fashion. For example, any of the methods and techniques described herein may be automatically implemented without user input after the capture of a plurality of images.

[0059] Figure 1 illustrates a block diagram of an exemplary device 100 for creating three-dimensional images of a scene according to embodiments of the present invention. In this example, device 100 is a digital camera capable of capturing several consecutive, still digital images of a scene. In another example, the device 100 may be a video camera capable of capturing a video sequence including multiple still images of a scene. A user of the device 100 may position the camera in different positions for capturing images of different perspective views of a scene. The captured images may be suitably stored, analyzed and processed for creating three-dimensional images as described herein. For example, subsequent to capturing the images of the different perspective views of the scene, the device 100, alone or in combination with a computer, may use the images for creating a three-dimensional image of the scene and for displaying the three-dimensional image to the user.

[0060] Referring to Figure 1, the device 100 may include a sensor array 102 of charge coupled device (CCD) or CMOS sensors which may be exposed to a scene through a lens and exposure control mechanism as understood by those of skill in the art. The device 100 may also include analog and digital circuitry such as, but not limited to, a memory 104 for storing program instruction sequences that control the device 100, together with a CPU 106, in accordance with embodiments of the present invention. The CPU 106 executes the program instruction sequences so as to cause the device 100 to expose the sensor array 102 to a scene and derive a digital image corresponding to the scene. The digital image may be stored in the memory 104. All or a portion of the memory 104 may be removable, so as to facilitate transfer of the digital image to other devices such as a computer 108. Further, the device 100 may be provided with an input/output (I/O) interface 110 so as to facilitate transfer of digital image even if the memory 104 is not removable. The device 100 may also include a display 112 controllable by the CPU 106 and operable to display the images for viewing by a user.

[0061] The memory 104 and the CPU 106 may be operable together to implement an image generator function 114 for creating three-dimensional images in accordance with embodiments of the present invention. The image generator function 114 may generate a three-dimensional image of a scene using two or more images of the scene captured by the device 100. Figure 2 illustrates a flow chart of an exemplary method for creating a three-dimensional image of a scene using the device 100, alone or together with any other suitable device, in accordance with embodiments of the present invention. Referring to Figure 2, the method includes receiving 200 a plurality of images of a scene. For example, a user of the device 100 may use the input features of the device and move the device to different positions for capturing a plurality of images of a scene to which the sensor array 102 is exposed. The different images can include images of different perspective views of the scene. The CPU 106 may then implement instructions stored in the memory 104 for storing the captured images in the memory 104.

[0062] The method of Figure 2 includes determining 202 attributes of the plurality of images. For example, attributes of an image captured by an image capture device may include, but are not limited to, analysis of color(s), including mean, standard deviation, histogram correlation, cross correlation, edges, junctions, identified objects, size, orientation, and timestamps of images. For each captured image, the image generator function 114 can determine one or more attributes. Additional exemplary details of determining attributes of images are provided herein.

[0063] The method of Figure 2 also includes generating 204, based on the attributes, two or more images from among the plurality of images for use in creating a three-dimensional image. For example, the image generator function 114 may compare the measured value of an attribute of one image to the measured value of an attribute of another image for determining a difference of the measured values. The image generator function 114 may then determine whether the difference meets a threshold value level. If the threshold criterion is met, the image generator function 114 determines that the images may be selected for use in creating a three-dimensional image. This process may be used for preliminarily and efficiently determining whether images are candidates for pairing as a three-dimensional image as described in further detail herein.

[0064] The generated two or more images may also be suitably processed 206. For example, the images may be corrected and adjusted for display as described herein.

[0065] The method of Figure 2 includes displaying 206 the three-dimensional image. For example, the three-dimensional image may be displayed on the display 112. In another example, the

three-dimensional image may be communicated to and displayed on another device such as, but not limited to, a computer, video camera, digital picture frame, a set-top box, and a high-definition television.

[0066] Although the above examples are described for use with a device capable of capturing images, embodiments of the present invention described herein are not so limited. Particularly, the methods described herein for creating a three-dimensional image of a scene may for example be implemented in any suitable system including a memory and computer processor. The memory may have stored therein computer-executable instructions. The computer processor may execute the computer-executable instructions. The memory and computer processor may be configured for implementing methods in accordance with embodiments of the present invention described herein.

[0067] Figures 3A - 3D illustrate a flow chart of an exemplary method for creating a three-dimensional image of a scene in accordance with embodiments of the present invention. The method can convert a plurality of images to a three-dimensional image that can be viewed on a stereoscopic display. Referring to Figures 3A - 3D, the method can begin with receiving 300 a plurality of images of a scene. For example, the images can be captured by a standard digital video or still camera, or a plurality of different cameras of the same type or different type. A camera user may capture an initial image. Next, the camera user may capture subsequent image(s) at positions to the left or right of the position at which the initial image was captured. These images may be captured as still images or as a video sequence of images. The images may be captured using a device such as the device 100 shown in Figure 1. The images may be stored in a memory such as the memory 104 shown in Figure 1. In another example, the images may be received at a device after they have been captured by a different device.

[0068] Images suitable for use as a three-dimensional image may be captured by a user using any suitable technique. For example, Figure 4A illustrates a front view of a user 400 moving between positions for capturing different images using a camera 402 in accordance with embodiments of the present invention. Referring to Figure 4A, the user 400 is shown in solid lines in one position for capturing an image using the camera 402. The user 400 is shown in broken lines in another position for capturing another image using the camera 402. The camera 402 is also at different positions for capturing images offering different perspective views of a scene. In this example, the user 400 stands with his or her feet separated by a desired binocular distance, then captures the first image while aligning the camera over his or her right foot (the position of the user

400 shown in solid lines). Then the user captures the second image, and possibly other images in between, while aligning the camera 402 over his or her left foot (the position of the user 400 shown in broken lines). The captured images may be used for creating a three-dimensional image in accordance with embodiments of the present invention.

[0069] In another example, Figure 4B illustrates a front view of a user 410 moving between positions for capturing different images of a scene using a camera 412 in accordance with embodiments of the present invention. Referring to Figure 4B, the user 410 stands with his or her feet together and uses the camera 412 to capture the first image while maintaining a centered pose (the position of the user 410 shown in solid lines). Then the user moves one of his or her feet away from the other by twice the desired binocular distance while maintaining a centered pose and uses the camera 412 to capture the second image, and possibly other images in between (the position of the user 410 shown in broken lines). The captured images may be used for creating a three-dimensional image in accordance with embodiments of the present invention.

[0070] The distance between positions at which images are captured (the stereo baseline) for creating a three-dimensional image can affect the quality of the three-dimensional image. The optimal stereo baseline between the camera positions can vary anywhere between 3 centimeters (cm) and several feet, dependent upon a variety of factors, including the distance of the closest objects in frame, the lens focal length or other optics properties of the camera, the camera crop factor (dependent on sensor size), the size and resolution of the display on which the images will be viewed, and the distance from the display at which viewers will view the images. A general recommendation is that the stereo baseline should not exceed the distance defined by the following equation:

$$E = \frac{12D}{30FC/50}$$

where B is the stereo baseline separation in inches, D is the distance in feet to the nearest object in frame, F is the focal length of the lens in millimeters (mm), and C is the camera crop factor relative to a full frame (36 x 24 square mm.) digital sensor (which approximates the capture of a 35 mm analog camera). In the examples provided herein, it is assumed that at least two images have been captured, at least two of which can be interpreted as a stereoscopic pair.

[0071] The identification of stereo pairs in 302 is bypassed in the cases where the user has manually selected the image pair for 3D image registration. This bypass can also be triggered if a 3D-enabled capture device is used that identifies the paired images prior to the registration process.

Returning to Figures 3A - 3D, the method includes selecting 302 two images among the plurality of captured images for use as a stereoscopic pair. For example, the image generator function 114 shown in Figure 1 may be used for selecting captured images for use as a stereoscopic pair. One or more metrics can be defined for measuring one or more attributes of the plurality of images for selecting a stereoscopic pair. For example, a buffer of M consecutive images may be maintained, or stored in the memory 104. The attributes of image with index m are compared with the corresponding attributes of image m + 1. If there is no match between those two images, image m + 1 is compared with image m + 2. If images are determined to be sufficiently matched so as to be stereoscopic, and after those images have been processed as described below to generate a three-dimensional image, the m and m + 2 images are compared to also identify a possible stereoscopic pair. The process may continue for all or a portion of the images in the buffer.

[0072] A preliminary, quick analysis may be utilized for determining whether images among the plurality of captured images are similar enough to warrant a more detailed analysis. This analysis may be performed by, for example, the image generator function 114 shown in Figure 1. Figure 5 illustrates a flow chart of an exemplary method for a preliminary, quick analysis to pre-screen whether an image pair may be a valid stereoscopic pair in accordance with embodiments of the present invention. Referring now to Figure 5, the method includes defining 500 a candidate stereoscopic pair. For example, the image generator function 114 may define the image with index m and the image m + 1 as a candidate stereoscopic pair.

[0073] The method of Figure 5 includes performing 502 a series of measurements of the candidate stereoscopic image pair. The measurements may be of attributes of the image pair. For example, for each color, the image generator function 114 may measure or calculate the following values:

- Average image value  $AV = \left( \frac{1}{image\_size} \right) \sum_{i=1}^{image\_size} I_i$
- Segmented average image value: Divide image in k segments and take the average of those segments
- Minimum pixel value for each color of the image (MIN)
- Maximum pixel value for each color of the image (MAX)

[0074] The method of Figure 5 includes applying 504 criteria to the measurements. For example, the image function generator 114 shown in Figure 1 may apply several criteria for

determining if the images are a possible stereoscopic pair. Exemplary equations defining the application of these criteria to the image  $m$  and image  $m + 1$  follow:

$$\text{Image pair is not stereoscopic} = \text{ABS}(A V_m - A V_{m+1}) > \text{ThresholdAV}$$

OR

$$\text{For all } k, \text{ABS}(SA V_{k,n} - SA V_{k,m+1}) > \text{ThresholdSAV}$$

OR

$$\text{ABS}(\text{MAX}_m - \text{MAX}_{m+1}) > \text{ThresholdMAX}$$

OR

$$\text{ABS}(\text{MIN}_n - \text{MIN}_{n+1}) > \text{ThresholdMIN}$$

ThresholdAV, ThresholdSAV, ThresholdMAX, and ThresholdMIN are threshold value levels for the average, segmented average, maximum and minimum, respectively. These equations can be applied to all or at least some of the colors.

[0075] The method of Figure 5 includes determining 506 whether any of the criteria are met. The image generator function 114 may determine whether any of the criteria are met. If the differences between the values for each image are less than a defined threshold, analysis can continue using more complex techniques for determining whether the images are a suitable stereoscopic pair. For example, the method of Figures 6A - 6C, described below, can be applied for determining whether the images are a suitable stereoscopic pair 508. Otherwise, if all the differences are greater than the defined threshold, the images are rejected as a stereoscopic pair 510.

[0076] Referring again to Figures 3A - 3D, after images are determined to be a potential stereoscopic pair, the method includes applying 304 rudimentary color adjustment to the images. For example, the image generator function 114 shown in Figure 1 may apply color adjustment to the images. This optional color adjustment can be a normalized adjustment or DC-correction applied to a single image to allow luminance-based techniques to work better. In addition, several additional criteria may typically be applied to the luminance planes (or optionally to all color planes), including, but not limited to, a Hough transform analysis 306, segmentation 308, edge detection 310, and the like. For example, segmented objects or blocks with high information content can be compared between the two image views using motion estimation techniques, based on differential error measures, such as, but not limited to, sum of absolute difference (SAD) or sum of squared errors (SSE), or correlation based measures, such as phase correlation or cross correlation. Rotational changes between the two images may be considered and identified during this procedure.

Segmented objects that are in one view only are indicative of occlusion, and having a significant number of occluded regions is indicative of a poor image pair for stereoscopy. Regions of occlusion identified during this process are recorded for use in later parts of the conversion process. Similarly, motion vector displacement between matching objects may be recorded or stored for further use.

[0077] Using the results of the motion estimation process used for object similarity evaluation, vertical displacement can be assessed. Vertical motion vector components are indicative of vertical parallax between the images, which when large can indicate a poor image pair. Vertical parallax must be corrected via rectification and registration to allow for comfortable viewing, and this correction will reduce the size of the overlapping region of the image in proportion to the original amount of vertical parallax.

[0078] Using the motion vectors from the similarity of objects check, color data may be compared to search for large changes between images. Such large changes can represent a color difference between the images regardless of similar luminance.

[0079] The method of Figures 3A - 3D includes performing 312 edge-based analytics and matching for determining whether camera planes are parallel 313. For example, Figures 6A - 6C illustrate a flow chart of an exemplary method for edge-based analytics and matching for image correspondence, determination of right/left image, and camera toe-in/parallel plane configuration according to embodiments of the present invention. This method may be implemented by the image generator function 114 shown in Figure 1. Referring to Figures 6A - 6C, two images are provided 600. Edge detection, when applied to both images, can be the foundation for image analysis and correspondence. In an example, the image generator function 114 shown in Figure 1 may apply 602 numerous edge operators for this analysis. The edge operators may include, but are not limited to, zero-cross or gradient-based operations. Following the application of edge operators, the image generator function 114 may apply 604 a slice-based horizontal and vertical edge locator function for extracting edges from the binary images.

[0080] Referring to Figures 6A - 6C, the image generator function 114 may also determine whether enough edges have been found 606. This stage (606) involves a comparison of the edges found from the two pictures to make sure that they meet a predefined minimum edge count. This stage also confirms the similarity of the two pictures by comparing the edge count of the two images to assure they are within a predefined percentage of each other. If enough edges have not been found, techniques alternative to edge-based processing techniques may be used 608. These techniques include motion-estimation-based image matching and quadrant-based or full-image

based cross correlation of the input images. Edge extensions from slice boundaries can be generated 610 if enough edges have been found. This scheme simplifies the overall amount of calculation to determine the edges within the images by first considering edge segments that span a slice and then growing these edge segments to determine the exact edge size and endpoints. Next, the method of Figures 6A - 6C includes comparing and correlating 612 the lengths, slope, curvature, midpoints/offset position, boundaries, and primary points of the resulting edges within the two images as an additional check of potential correspondence. Block 614 tests each edge to see if it intersects with a single image boundary (i.e., top, bottom, left or right boundary of image). In the event of an edge intersecting a single image boundary, block 616 classifies its primary point as the endpoint which is not on the boundary. Otherwise, block 618 classifies its primary point as the midpoint of the edge. Block 620 involves solving a minimization problem via vertical edge matching in order to determine the optimal selection for the vertical shift between the two images. In an example for block 620, the following equation may be used:

$$\min_{\delta_y = -\epsilon_y}^{\epsilon_y} \sum_{i=1}^M \min_{j=1}^N (\min (\|P_i(x, y + \delta_y) - Q_j(x, y)\|, \epsilon_x + \epsilon_y))$$

For each vertical edge in one image, determine the closest edge in the other image, subject to meeting criteria for length, slope and curvature. For distance, use the distance between the primary points. If this distance is larger than  $\epsilon$ , it is deemed that no edge matches, and this edge contributes  $\epsilon$  to the cost function. The end result of the optimization is the determination of  $\delta$ , the optimal shift between the two images based on this vertical edge matching. In box 622, the same optimization process from box 620 is repeated; this time, however, is for horizontal edge matching, and utilizes the vertical  $\delta$  already determined from box 620.

[0081] In an example for block 622, the following equation may be used:

$$\min_{\delta_x = -\epsilon_x}^{\epsilon_x} \sum_{i=1}^M \min_{j=1}^N \min (\|P_i(x + \delta_x, y + \delta_y) - Q_j(x, y)\|, \epsilon_x + \epsilon_y)$$

Block 624 then uses the calculated horizontal and vertical  $\delta$ 's to match each edge with its closest edge that meets the length, slope and curvature criteria. In an example for block 624, the following equation may be used:

$$C_{IJ} = \begin{cases} 1 & \text{if } P_i \text{ matches } Q_j \\ 0 & \text{otherwise} \end{cases}$$

The output of this stage is the matrix C, which has 1 in location i,j if edge i and j are matching edges and otherwise 0. This matrix is then pruned in Box 626 so that no edge is matched with multiple

other edges. In the event of multiple matches, the edge match with minimal distance is used. Finally, in Box 628, the edge matches are broken down into regions of the image. The set of matching edges within each region are then characterized by the mean shift, and this mean shift is then the characteristic shift of the region. By examining the direction of the shifts of each subregion, it is thus possible to determine which picture is left and which is right. It is also possible to determine whether the second captured picture was captured with a focal axis parallel to the first picture. If not, there is some amount of toe-in or toe-out which can be characterized by the directional shifts of the subregions.

[0082] Referring to Figures 6A - 6C, the extracted edge sets from the two input images can be compared as part of a minimal optimization, in order to solve for the optimal delta translation between images. This  $\delta$  value allows for determination of which image is left and right, as well as whether the cameras were in parallel configuration. When the cameras focal axes are parallel (or near parallel), the algorithm (from Figures 3A - 3D) can proceed to the image registration without performing image rectification.

[0083] Figure 7 is a graphical depiction of an edge detection example in accordance with embodiments of the present invention. The example of Figure 7 shows the edge detection map when edge detection is applied to a captured image, and demonstrates that the extracted images can be a representation of image composition.

[0084] A Hough transform can be applied 306 to identify lines in the two images of the potential stereoscopic pair. Lines that are non-horizontal, non-vertical, and hence indicate some perspective in the image can be compared between the two images to search for perspective changes between the two views that may indicate a perspective change or excessive toe-in during capture of the pair.

[0085] The aforementioned criteria may be applied to scaled versions of the original images for reducing computational requirements. The results of each measurement may be gathered, weighted, and combined to make a final decision regarding the probable quality of a given image pair as a stereoscopic image pair.

[0086] The method of Figures 3A - 3D includes identifying 314 a valid stereoscopic pair. For example, Figures 8A and 8B illustrate a flow chart of an exemplary method for determining whether an image pair is a valid stereoscopic pair and which image is left and right according to embodiments of the present invention. This method may be implemented, for example, by the image generator function 114 shown in Figure 1. Referring to Figures 8A and 8B, the method

includes defining 800 a candidate stereoscopic image pair. In this example, two images with indices  $m$  and  $m+1$  are examined. The method includes performing 802 a quick analysis to identify stereoscopic pairs.

[0087] At step 804, color segmentation is performed on the objects. At step 806, the bounding box of  $8 \times 8$  blocks for each object in each image may be identified. At step 810, images may be partitioned into  $N \times N$  blocks. At step 812, blocks with high information content may be selected. At step 813, the method includes performing motion estimation on blocks in L relative to R image (accumulate motion vectors for IVR determination. These steps may be considered Techniques 1, 2, and 3.

[0088] At step 814, edge detection may be performed on left/right images. Next, at step 816, vertical and horizontal lines in left / right images may be identified and may be classified by length, location, and slope. At step 818, a Hough transform may be performed on the left / right images. Next, at step 820, the method includes analyzing Hough line slope for left / right images and identifying non-vertical and non-horizontal lines.

[0089] Referring to Figures 8A and 8B, LDIAG represents the set of lines that have been identified as non-vertical or non-horizontal using the Hough transform. LHV represents lines that have been classified as either vertical or horizontal. MVY are the luminance motion vectors, and MVCRB the chrominance motion vectors, for each segmented object or  $N \times N$  block. Similarly, MVYM is the mean luminance motion vector measurement, and MYCRBM the mean chrominance motion vectors. BMAD is the mean accumulated best match difference. ORG is the measurement of how well origins of horizontal / vertical lines match. LEN is the measurement of how well lengths of horizontal / vertical lines match. SLP is the measurement of how well slopes of horizontal / vertical lines match. TIN is the measurement of how well slopes of diagonal lines match.

[0090] At step 822, the following calculations may be performed for all objects or blocks of interest and lines:

$$MVYM = \sum_{i,j} MVY_{i,j,T_n},$$

where  $i$ : object ID,  $j$ : motion vector ID,  $m$ : image ID

$$BMAD = \sum_i \overline{abs[block_{0,ij,m} - block_{mv,ij,m+1}]},$$

where  $i$ : object ID,  $j$ : motion vector ID,  $m$ : image ID,  $mv$ : best match vector for block]

$$\begin{aligned}
 MVCRBM_i J &= \sum_i^{V_m} MVCRB_{ijrn} \\
 ORG &= \sum_i^V abs[origin(LHV_{im}) - origin(LHV_{im+1})] \\
 LEN &= \sum_i^Y abs[length(LHV_{im}) - length(LHV_{im+1})] \\
 SLP &= \sum_i^Y abs[slope(LHV_{im}) - slope(LHV_{im+1})] \\
 TIN &= \sum_i abs[slope(LDIAG_{Um}) - slope(LDIAG_{Um+1})]
 \end{aligned}$$

[0091] At step 824, a weighted average of the above measures may be performed to determine whether images are a pair or not. Next, at step 826, average motion vector direction may be used to determine left / right images.

[0092] Referring again to Figures 3A - 3D, the method can next include determining which image of the stereoscopic pair represents the left view image and which image represents the right view image. This aspect can be important in many applications since, for example, a user can capture a plurality of images moving to the left or right. First, image segmentation 308 can be performed to identify objects within the two captured views. The motion estimation step that has been defined before saves the motion vectors of each object or block with high information content. If the general motion of segmented objects is to the right for one view relative to the other, it is indicative of a left view image, and vice versa. Since the process of motion estimation of segmented objects is also used in stereoscopic pair evaluation, left/right image determination can be performed in parallel.

[0093] For a stereo pair of left and right view images, the method of Figures 3A - 3D includes rectification point selection 316, rectification 318, and region of interest identification 320. For example, interest points for stereo correspondence, rectification and registration can be identified. According to embodiments of the present invention, the left view image, sized N x M, is broken into a number, N, of smaller n x m sub-images. Each sub-image can be filtered to find junction points, or interest points, within and between objects in view. Interest points can be identified, for example, by performing horizontal and vertical edge detection, filtering for strong edges of a minimum length, and identifying crossing points of these edges. Interest point determination can be assisted by Hough transform line analysis when determining the dominant

edges in a scene. Interest points may not be selected from areas identified as occluded in the initial analysis of a stereo pair. Interest points can span the full image.

[0094] For a stereo pair of left and right view images with a set of identified interest points, rectification 318 may be performed on the stereo pair of images. Using the interest point set for the left view image, motion estimation techniques (as described in stereo pair identification above) and edge matching techniques are applied to find the corresponding points in the right view image. Figure 9 depicts an example of applying this technique. Referring to Figure 9, the N corresponding points in the left and right view images are made into a 3 x N set of point values, for example:

$$right_{pts} = \left\{ \begin{array}{ccc} x1_r & x2_r & x3_r \\ y1_r & y2_r & y3_r \\ 1 & 1 & 1 \end{array} \right\} \text{ and}$$

$$left_{pts} = \left\{ \begin{array}{ccc} x1_l & x2_l & x3_l \\ y1_l & y2_l & y3_l \\ 1 & 1 & 1 \end{array} \right\},$$

and the fundamental matrix equation

$$right_{pts}^T * F * left_{pts} = 0$$

is solved or approximated to determine the 3 x 3 fundamental matrix, F, and epipoles, e1 and e2. The camera epipoles are used with the interest point set to generate a pair of rectifying homographies. It can be assumed that the camera properties are consistent between the two captured images. The respective homographies are then applied to the right and left images, creating the rectified images. The overlapping rectangular region of the two rectified images is then identified, the images are cropped to this rectangle, and the images are resized to their original dimensions, creating the rectified image pair, right\_r and left\_r. The rectified image pair can be defined by the following equations:

$$right\_r = \text{cropped}(F * right)$$

$$left\_r = \text{cropped}(F * left)$$

For the stereo pair of "left\_r" and "right\_r" images, registration is next performed on the stereo pair. A set of interest points is required, and the interest point set selected for rectification (or a subset thereof) may be translated to positions relative to the output of the rectification process by applying the homography of the rectification step to the points. Optionally, a second set of interest points may be identified for the left\_r image, and motion estimation and edge matching techniques may be applied to find the corresponding points in the right\_r image. The interest point selection process for the registration operation is the same as that for rectification. Again, the N corresponding interest points are made into a 3 x N set of point values as set forth in the following equations:

$$right\_r_{pts} = \begin{Bmatrix} x'1_r & x'2_r & x'3_r \\ y'1_r & y'2_r & y'3_r \dots \\ 1 & 1 & 1 \end{Bmatrix} \text{ and}$$

$$left\_r_{pts} = \begin{Bmatrix} x'1_l & x'2_l & x'3_l \\ y'1_l & y'2_l & y'3_l \dots \\ 1 & 1 & 1 \end{Bmatrix},$$

and the following matrix equation

$$left\_r_{pts} = Tr * right\_r_{pts}$$

is approximated for a 3 x 3 linear conformal transformation, Tr, which may incorporate both translation on the X and Y axes and rotation in the X/Y plane. The transform Tr is applied to the right\_r image to create the image "Right'" as defined by the following equation:

$$Right' = Tr * right\_r,$$

where right\_r is organized as a 3 x N set of points (xi\_r, yi\_r, 1) for i = 1 to image\_rows \* image\_cols.

[0095] Finally, the second set of interest points for the left\_r image may be used to find correspondence in the Right' image, the set of points as set forth in the following equations:

$$Right'_{pts} = \begin{Bmatrix} x'1_r & x'2_r & x'3_r \\ y'1_r & y'2_r & y'3_r \dots \\ 1 & 1 & 1 \end{Bmatrix} \text{ and}$$

$$left\_r_{pts} = \left\{ \begin{array}{ccc} x'1_l & x'2_l & x'3_l \\ y'1_l & y'2_l & y'3_l \dots \\ 1 & 1 & 1 \end{array} \right\},$$

is identified and composed, and the equation

$$Right'_{pts} = T1 * left\_r_{pts}$$

is approximated for a second linear conformal transformation, T1. The transform T1 is applied to the left\_r image to create the image "Left'", as defined by the following equation:

$$Left' = T1 * left\_r$$

"Right'" and "Left'" images represent a rectified, registered stereoscopic pair.

[0096] The method of Figures 3A - 3D includes an overall parallax, or disparity, calculation 332. According to embodiments of the present invention, for a stereoscopic pair of registered "Left'" and "Right'" images, a pixel-by-pixel parallax, or disparity, map is created. This can be performed, for example, by using a hierarchical motion estimation operation between the Left' and Right' images, starting with blocks sized N x N and refining as necessary to smaller block sizes. During the estimation process, only horizontal displacement may be considered, limiting the search range. After each iteration of the process, the best match position is considered for pixel-by-pixel differences, and the next refinement step, if needed, is assigned by noting the size of the individual pixel differences that are greater than a threshold, Tp. Regions of the image previously identified as occluded in one image are assigned the average parallax value of the pixels in the surrounding neighborhood. Regions of an image that are not known to be occluded from previous steps in the process, and for which an appropriate motion match cannot be found (pixel differences are never < Tp) are assigned to the maximum possible parallax value to allow for simple identification in later steps of the stereo composition process. In the example of Figures 3A - 3D, the method includes correspondence point selection 322, correspondence 324 and registration transform to generate the Right' image 326. In addition, the method includes correspondence 328 and registration transform to generate the Left' image 330.

[0097] Figure 10 illustrates a flow chart of an exemplary method for determining pixel disparities according to embodiments of the present invention. The method may be implemented,

for example, by the image generator function 114 shown in Figure 1. Referring to Figure 10, the method includes receiving 1000 a transformed stereoscopic image pair, including a left and right image. The method includes dividing 1002 the images into blocks of  $N \times N$  pixels. For every block, the method includes performing 1004 motion estimation between left and right to determine a best match vector. Next, for every pixel in each block, the method includes calculating 1006 the differences between left and right for the best match vector.

[0098] The method of Figure 10 includes determining 1008 whether the best match difference is less than the threshold  $T_p$ . If the best match difference is less than the threshold  $T_p$ , the disparity of the pixel is set equal to the best match vector 1010. Otherwise, if the best match difference is not less than the threshold  $T_p$ , the method includes determining 1012 whether the pixel is occluded. If the pixel is determined to be occluded, the disparity of the pixel is set equal to the best match vector 1010. If the pixel is determined not to be occluded, the method includes grouping pixels in an  $M \times M$  block and performing a new analysis with  $M \times M$  refinement 1014.

[0099] After steps 1010 and 1014 of Figure 10, the method includes determining 1016 whether there are more pixels in the current block being processed. If there are more pixels, the method returns to step 1006. Otherwise, the method determines 1018 whether there are more blocks to be processed. If not, the method exits 1020. If there are more blocks, the method returns to step 1004.

[0100] Returning now to Figures 3A - 3D, the method includes applying 334 a parallax analysis. For example, for a stereoscopic pair of registered "Left" and "Right" images, the maximum and minimum pixel parallax values can be analyzed to decide whether the maximum or minimum parallax is within the ability of a viewer to resolve a three-dimensional image. If it is determined that the parallax is within the ability of a viewer to resolve the three-dimensional image, the method proceeds to step 342. If not, the method proceeds to step 336. Occluded regions and pixels with "infinite" parallax are not considered in this exemplary method.

[0101] For a stereoscopic pair of registered "Left" and "Right" images, the screen plane of the stereoscopic image can be altered 336, or relocated, to account for disparities measured as greater than a viewer can resolve. This is performed by scaling the translational portion of transforms that created the registered image views by a percent offset and re-applying the transforms to the original images. For example, if the initial left image transform is as follows:

$$Tl = \begin{Bmatrix} S * \cos \theta & S * \sin \theta & Tx \\ -S * \sin \theta & S * \cos \theta & Ty \\ 0 & 0 & 1 \end{Bmatrix}$$

for scaling factor S, X/Y rotation angle  $\theta$ , and translational offsets Tx and Ty, the adjustment transform becomes

$$Tl_{alt} = \begin{Bmatrix} S * \cos \theta & S * \sin \theta & Tx * Xscale \\ -S * \sin \theta & S * \cos \theta & Ty * Yscale \\ 0 & 0 & 1 \end{Bmatrix}$$

where Xscale and Yscale are determined by the desired pixel adjustment relative to the initial transform adjustment, i.e.,

$$Xscale = 1 + \frac{\{desired\_pixel\_adjustment\}}{Tx}.$$

Only in rare occurrences will Yscale be other than zero, and only then as a corrective measure for any noted vertical parallax. Using the altered transform, a new registered image view is created, e.g. the following:

$$Left' = Tl_{alt} * left\_r$$

Such scaling effectively adds to or subtracts from the parallax for each pixel, effectively moving the point of now parallax forward or backward in the scene. The appropriate scaling is determined by the translational portion of the transform and the required adjustment.

[0102] At step 338 of Figures 3A - 3D, it is determined whether the parallax is within the ability of a viewer to resolve the three-dimensional image. If it is determined that the parallax is within the ability of a viewer to resolve the three-dimensional image, the method proceeds to step 342. If not, the method proceeds to step 340. For a stereoscopic pair of registered "Left" and "Right" images, the pixel-by-pixel parallax for pixels of segmented objects may also be adjusted 340, or altered, which effectively performs a pseudo-decrease (or increase) in the parallax of individual segmented objects for objects that still cannot be resolved after the screen adjustments

above. This process involves the same type of manipulation and re-application of a transform, but specific to a given region of the picture, corresponding to the objects in question.

[0103] Since moving an object region in the image may result in a final image that has undefined pixel values, a pixel-fill process is required to ensure that all areas of the resultant image have defined pixel values after object movement. An exemplary procedure for this is described below. Other processes, both more or less complex, may be applied.

[0104] Figure 11 illustrates a flow chart of an exemplary method for adjusting parallax of segmented, moving objects according to embodiments of the present invention. Further, Figure 12 illustrates an exemplary diagram of a method for adjusting parallax of moving, segmented objects according to embodiments of the present invention. The method may be implemented, for example, by the image generator function 114 shown in Figure 1. Referring now to Figure 11, the method includes identifying 1100 a segmented object in an image I to relocate. The method of Figure 11 also includes defining a bounding rectangle R around the object and defining left / right bounds of a region M for left / right motion 1102. In an example of defining the bounding rectangle R, the segmented region to be moved may be identified as a rectangular region of pixels, R, in the left\_r or right\_r image (whichever is to be altered), sized X columns by Y rows with the following coordinates:

$R_{ul} = (x_i, y_u)$ ; the upper left coordinate

$R_{ul} = (x_i, y_i)$ ; the lower left coordinate

$R_{ur} = (x_r, y_u)$ ; the upper right coordinate

$R_{ir} = (x_r, y_i)$ ; the lower right coordinate

For a large or complex object, multiple rectangular regions may need to be defined and moved, but the process executes identically for each region.

[0105] In an example of defining left / right bounds of a region M for left / right motion, the region M is the region to which the altered transform can be applied. This process first assesses the direction of movement to occur and defines one side of region M. If the intended movement is to the right, then the right bounding edge of region M is defined by the following coordinate pair in the appropriate left\_r or right\_r image (whichever is to be adjusted):

$$M_{ur} = (x_r + P, y_u); \text{ upper right}$$

$$M_{lr} = (x_r + P, y_l); \text{ lower right}$$

If movement is to the left, the left bounding edge of region M is defined as:

$$M_{ul} = (x_l - P, y_u); \text{ upper left}$$

$$M_{ll} = (x_l - P, y_l); \text{ lower left}$$

P is an extra number of pixels for blending purposes. The scaled version of the registration transform matrix  $T_{alt}$  is provided 1104. The inverse of the altered transform (assumed already calculated as above for movement of the screen plane for the whole image) may then be applied 1106 to the opposite edge of the region R to get the other edge of region M. For the sake of example, assume that the movement of R is intended to be to the right, and that the left image is to be altered (meaning  $T_{alt}$  has been created for the intended movement). Since the right side of M is already known, the other side can now be determined as:

$$M_{ul} = T_{alt}^{-1} * R_{ul} + (P, 0); \text{ upper left}$$

$$M_{ll} = T_{alt}^{-1} * R_{ul} + (P, 0); \text{ lower left}$$

Again, P is an extra number of pixels for blending, and  $T_{alt}^{-1}$  is the inverse transform of  $T_{alt}$ . Note that P is added after the transform application, and only to the X coordinates. The region to be moved is now defined as the pixels within the rectangle defined by M.

[0106] The method also includes applying 1108 the inverse transform of  $T_{alt}$  to the image to be transformed for blocks in the region M. For example, from this point, one of two operations can be used, depending on a measurement of the uniformity (texture) or the area defined by the coordinates  $M_{ul}$ ,  $M_{ll}$ ,  $R_{ul}$ , and  $R_{ur}$  (remembering again that the region would be using other coordinates for a movement to the left). Uniformity is measured by performing a histogram analysis on the RGB values for the pixels in this area. If the pixel variation is within a threshold, the area is deemed uniform, and the movement of the region is affected by applying the following equation:  $Left' = T_{alt} * left\_r$ , for  $left\_r \subset M$ . This is the process shown in the example method of Figure 12. Alternatively, if the area is not uniform, movement of the object is applied to the smaller area:

$$Left' = T_{alt} * left\_r, \text{ for the } left\_r \text{ region defined by } R_{ul}, R_{ur}, M_{ul}, \text{ and } M_{lr}.$$

The method of Figure 11 includes overwriting 1110 pixels within the defined rectangle in the transformed image with the newly transformed pixels.

[0107] The method of Figure 11 includes interpolating the outer P pixels on each side of the area with existing data. For example, the area in Left' defined by the coordinates  $M_{ul}$ ,  $M_u$ ,  $R_{ur}$ , and  $R_l$  will be empty, but is filled with a linear gradient fill between points on each horizontal line in the region. The fill-in process first determines the following distance d:

$$d = R_{ul}(x) - M_{ul}(x)$$

for the x-coordinates of  $R_{ul}$  and  $M_{ul}$ , and then proceeds to determine an interpolated gradient between the two pixel positions to fill in the missing values. For simplicity of implementation, the interpolation is always performed on a power of two, meaning that the interpolation will produce one of 1, 2, 4, 8, 16, etc. pixels as needed between the two defined pixels. Pixel regions that are not a power of two are mapped to the closest power of two, and either pixel repetition or truncation of the sequence is applied to fit. As an example, if  $R_{ul}(x) = 13$  and  $M_{ul}(x) = 6$ , then  $d = 7$ , and the following intermediate pixel gradient is calculated for a given row, j, in the region:

$$p1 = \frac{7}{8} * (x_{6,j}) + \frac{1}{8} * (x_{13,j})$$

$$p2 = \frac{6}{8} * (x_{6,j}) + \frac{2}{8} * (x_{13,j})$$

$$p3 = \frac{5}{8} * (x_{6,j}) + \frac{3}{8} * (x_{13,j})$$

$$p4 = \frac{4}{8} * (x_{6,j}) + \frac{4}{8} * (x_{13,j})$$

$$p5 = \frac{3}{8} * (x_{6,j}) + \frac{5}{8} * (x_{13,j})$$

$$p6 = \frac{2}{8} * (x_{6,j}) + \frac{6}{8} * (x_{13,j})$$

$$p7 = \frac{1}{8} * (x_{6,j}) + \frac{7}{8} * (x_{13,j})$$

$$p8 = (x_{13,j})$$

Since only 7 values are needed, p8 would go unused in this case, such that the following assignments would be made:

$$(x_6, y_j) = p1$$

$$(x_7, y_j) = p2$$

$$(x_8, y_j) = p3$$

$$(x_9, y_j) = p4$$

$$(x_{10}, y_j) = p5$$

$$(x_{11}, y_j) = p6$$

$$(x_{12}, y_j) = p7.$$

This process can repeat for each row in the empty region.

[0108] A weighted averaging the outer P "extra" pixels on each side of the rectangle with the pixel data currently in those positions is performed to blend the edges.

[0109] As an alternative to the procedure of applying movement and pixel blending to alter the parallax of an object, the disparity map calculated using the two views, "Left" and "Right," can be altered for the region M to reduce the disparity values in that region, and then applied to one of the "Left" or "Right" single image views to create a new view (e.g., "Left\_disparity"). The result of this process is a new stereo pair (e.g., "Left" and "Left\_disparity") that recreates the depth of the original pair, but with lesser parallax for the objects within the region M. Once created in this manner, the "disparity" view becomes the new opposite image to the original, or for example, a created "Left\_disparity" image becomes the new "Right" image.

[0110] Returning to Figures 3A - 3D, the method includes performing 342 depth enhancements. For example, for a stereoscopic pair of registered "Left" and "Right" images, the screen plane of the stereoscopic image may be relocated to allow a viewer to emphasize or de-emphasize object depth in the three-dimensional image. This relocation may be implemented to enhance the subjective quality of the displayed image or to create three-dimensional effects that involve changing object depth over time to simulate motion. The process for this uses the same

procedures as for general readjustment of the screen plane, and for segmented object specific adjustments, but is performed voluntarily for effect, rather than necessarily for correction.

[0111] The method of Figures 3A - 3D includes removing 344 moving objects. For example, for a stereoscopic pair of registered "Left" and "Right" images, disparity differences can be identified which indicate object motion within, into, or out of the image frame for one image. These areas can be identifiable as those which have "infinite" parallax assignments from the disparity map step of the process. Areas indicating such motion are replicated or removed using data from the other image view and/or other views captured between the "Left" and "Right" images. Without any loss of generality, we will assume that first picture taken is the leftmost, and the last picture taken is the rightmost. In actuality, the opposite can occur. In the following description the following definitions apply:

- First picture: the first picture captured in the sequence (1)
- Last picture: the last picture captured in the sequence (N)
- Leftmost pictures: any set of pictures from 1<sup>st</sup> to N-1
- Rightmost pictures: any set of pictures from 2<sup>nd</sup> to Nth
- Left target picture: any of the leftmost pictures or a modified version of all captured pictures that will be used during the 3D generation process as left picture
- Right target picture: any of the rightmost pictures or a modified picture that will be used during the 3D generation process as right picture

The method of identifying and compensating for moving objects consists of the following steps. For a given sequence of pictures captured between two positions, divide each picture into smaller areas and calculate motion vectors between all pictures in all areas. Calculate by a windowed moving average the global motion that results from the panning of the camera. Then subtract the area motion vector from the global motion to identify the relative motion vectors of each area in each picture. If the motion of each area is below a certain threshold, the picture is static and the first and last picture, or any other set with the desired binocular distance, can be used as left and right target pictures to form a valid stereoscopic pair that will be used for registration, rectification, and generation of a 3D picture. If the motion of any area is above an empirical threshold, then identify all other areas that have zero motion vectors and copy those areas from any of the leftmost pictures to the target left picture and any of the rightmost pictures to the target right picture.

[0112] For objects where motion is indicated and where the motion of an object is below the acceptable disparity threshold, identify the most suitable image to copy the object from, copy the object to the left and right target images and adjust the disparities as shown in the attached figure. The more frames that are captured, the less estimation is needed to determine the rightmost pixel of the right view. Most of occluded pixels can be extracted from the leftmost images. For an object that is moving in and out of the scene between the first and last picture, identify the object and completely remove it from the first picture if there is enough data in the captured sequence of images to fill in the missing pixels.

[0113] For objects where motion is indicated and where the motion is above the acceptable disparity, identify the most suitable picture from which to extract the target object and extrapolate the proper disparity information from the remaining captured pictures.

[0114] The actual object removal process involves identifying  $N \times N$  blocks, with  $N$  empirically determined, to make up a bounding region for the region of "infinite" parallax, plus an additional  $P$  pixels (for blending purposes), determining the corresponding position of those blocks in the other images using the parallax values of the surrounding  $P$  pixels that have a similar gradient value (meaning that high gradient areas are extrapolated from similar edge areas and low gradient areas are extrapolated from similar surrounding flat areas), copying the blocks/pixels from the opposite locations to the intended new location, and performing a weighted averaging of the outer  $P$  "extra" pixels with the pixel data currently in those positions to blend the edges. If it is determined to remove an object, fill-in data is generated 346. Otherwise, the method proceeds to step 348.

[0115] Figures 13A, 13B, and 13C illustrate an exemplary process for disparity interpolation according to embodiments of the present invention. Referring to Figures 13A, 13B, and 13C, positions 1300 and 1302 are positions of a camera (not shown) when capturing images of object 1304 at different times. The image captured at position 1300 was captured prior to the image captured at position 1302. A view of the object 1304 from position 1300 is indicated by lines 1306, and a view of the object 1304 from position 1302 is indicated by lines 1308. As shown by direction arrow 1310 in Figure 13B, the object 1304 is moving from left to right across the camera's view. Between the image capture times, the object 1304 has moved from a position 1304A (shown in a broken line) to the position 1304 shown in Figure 13B) as shown in Figure 13B.

[0116] The movement of the object **1304** is such that the disparity is unacceptable and should be corrected. In this example, the image obtained from position **1300** can be utilized for creating a three-dimensional image, and the image obtained from position **1302** can be altered for use together with the other image in creating the three-dimensional image. To correct, the object **1304** may be moved to the left (as indicated by direction arrow **1312** in Figure 13C) in the image captured from position **1302**. The object **1304** may be moved to the left to a position of a desired left view (i.e., the positioning of the object **1304** within the view from position **1302** as indicated by lines **1314** shown in Figure 13C). The desired left image for the three-dimensional image may be composed by reducing visibility of the left-most pixel from RLa to RLd; and by increasing visibility of the right-most pixel by interpolating the [RRd, RRa] area from pixels found in the right-most of RRa.

[0117] Another example of a process for adding/removing objects from a single image is illustrated in Figures 14 and 15. Figure 14 illustrates a flow chart of an exemplary method for adding/removing objects from a single image according to embodiments of the present invention. Referring to Figure 14, the method includes creating parallax maps for stereoscopic images I1 and I2 and defining the area of image I1 to change (step **1400**). The method of Figure 13 also includes defining **1402** 8 x 8 blocks in image I1 to cover the intended area plus P pixels. Using the parallax map, the corresponding data in image I2 is found (step **1404**). The corresponding data is copied from image I2 to image I1 (step **1406**). Next, the method includes applying a weighted average of the outer P pixels of the copy (step **1408**).

[0118] Referring to Figure 15, the figure is a diagram of an exemplary method for adding/removing objects from a single image according to embodiments of the present invention. An original "Left" image **1500** and an original "Right" image **1502** are provided. The images may be paired to form a three-dimensional image in accordance with embodiments of the present invention described herein. The images **1500** and **1502** both show objects, which are designated **1504L** and **1506L**, respectively, in the "Left" image **1500**, and designated **1504R** and **1506R**, respectively, in the "Right" image **1502**. The parallax of these objects is such that three-dimensional display of these objects in the three-dimensional image **1504** would be satisfactory to a viewer.

[0119] Referring to Figure 15, the images of another object (designated 1508L in the "Left" image 1500, and designated 1508R in the "Right" image 1502) were captured while the object was moving at such a speed such that the parallax disparity of the object in the "Left" image 1500 and the "Right" image 1502 makes viewing the three-dimensional image 1504 of the object unsatisfactory to a viewer. For this reason, the moving object may be removed from the "Left" image 1500 and the "Right" image 1502. A new "Left" image 1510 without the moving object may be generated by bounding a region 1512L to be corrected in the original "Left" image 1500 for removing the moving object (i.e., an area including the moving object in the "Left" image 1500). A corresponding area in the original "Right" image 1502 may be copied and used for replacing the bounded region 1512L in the original "Left" image 1500 to render the new "Left" image 1510. In a similar manner, a new "Right" image 1514 without the moving object can be rendered. The new "Left" image 1510 and the new "Right" image 1514 can then be paired for rendering a new three-dimensional image 1516 without the moving object.

[0120] As an alternative to the procedure of identifying bounding regions of 8 x 8 blocks around objects to be added or removed in a view, the disparity map calculated using multiple views, "Left", "Right", and/or the images in between, can be applied to one of the "Left" or "Right" single image views to create a new view (e.g., "Left\_disparity"). The result of this process is a new stereo pair (e.g., "Left" and "Left\_disparity") that effectively recreates the depth of the original pair, but without object occlusions, movement, additions, or removals. Once created in this manner, the "disparity" view becomes the new opposite image to the original, or for example, a created "Left\_disparity" image becomes the new "Right" image. Effectively, this procedure mimics segmented object removal and/or addition, but on a full image scale.

[0121] Returning to Figures 3A - 3D, the method includes applying 348 color correction to the images. For example, for a plurality of images, a pixel-by-pixel color comparison may be performed to correct lighting changes between image captures. This is performed by using the parallax map to match pixels from Left' to Right' and comparing the luminance and chrominance values of those pixels. Pixels with both large luminance and chrominance discrepancies are ignored, assuming occlusion. Pixels with similar luminance and variable chrominance are altered to average their chrominance levels to be the same. Pixels with similar chrominance and variable luminance are altered to average their luminance values to account for lighting and reflection changes.

[0122] For a finalized, color corrected, motion corrected stereoscopic image pair, the "Left" and "Right" images are ordered and rendered to a display as a stereoscopic image. The format is based on the display parameters. Rendering can require interlacing, anamorphic compression, pixel alternating, and the like.

[0123] For a finalized, color corrected, motion corrected stereoscopic image pair, the "Left" view may be compressed as the base image and the "Right" image may be compressed as the disparity difference from the "Left" using a standard video codec, differential JPEG, or the like.

[0124] The method of Figures 3A - 3D includes displaying 350 the three-dimensional image on a stereoscopic display. For example, the three-dimensional image may be displayed on the display 112 of the device 100 or a display of the computer 108. Alternatively, the three-dimensional image may be suitably communicated to another device for display.

[0125] When a video sequence is captured with lateral camera motion as described above, stereoscopic pairs can be found within the sequence of resulting images. Stereoscopic pairs are identified based on their distance from one another determined by motion analysis (e.g., motion estimation techniques). Each pair represents a three-dimensional picture or image, which can be viewed on a suitable stereoscopic display. If the camera does not have a stereoscopic display, the video sequence can be analyzed and processed on any suitable display device. If the video sequence is suitable for creating three-dimensional content (e.g., one or more three-dimensional images), it is likely that there are many potential stereoscopic pairs, as an image captured at a given position may form a pair with images captured at several other positions. The image pairs can be used to create three-dimensional still images or re-sequenced to create a three-dimensional video.

[0126] When creating three-dimensional still images, the user can select which images to use from the potential pairs, thereby adjusting both the perspective and parallax of the resulting images to achieve the desired orientation and depth. Figure 15 illustrates an exemplary process for creating three-dimensional still images from a standard two-dimensional video sequence by identifying stereoscopic pairs in accordance with embodiments of the present invention. Referring to Figure 16, this process can be used to create content for multi-view stereoscopic displays by creating a set of three-dimensional images of a subject with the same parallax but captured from slightly different positions. A three-dimensional video sequence can be created using one of the following methods. The first method is to select stereoscopic pairs with a constant positional offset,

and sequence them in the same relative order in which they were captured. The user can select the offset to achieve the desired depth. During playback this method creates the effect of camera motion the same as occurred during capture, while the depth of the scene remains constant due to the fixed parallax. Figure 17 illustrates an exemplary process for creating three-dimensional video from a standard two-dimensional video sequence according to embodiments of the present invention.

[0127] Another method of creating a three-dimensional sequence includes creating stereoscopic pairs by grouping the first and last images in the sequence, followed by the second and next-to-last images, and so on until all images have been used. During playback this creates the effect of the camera remaining still while the depth of the scene decreases over time due to decreasing parallax. The three-dimensional images can also be sequenced in the opposite order so that the depth of the scene increases over time. Figure 18 illustrates an exemplary process of creating three-dimensional video with changing parallax and no translational motion from a standard two-dimensional video sequence in accordance with an embodiment of the subject matter disclosed herein. The camera or other display device can store a representation of the resulting three-dimensional still images or video in an appropriate compressed format. For more efficient storage of still images, one of the images in the stereoscopic pair can be compressed directly, while the other image is represented by its differences with the first image. For video sequences, the first stereoscopic pair in the sequence can be stored as described above for still images, while all images in other pairs are represented by their differences with the first image.

[0128] The generation and presentation, such as display, of three-dimensional images of a scene in accordance with embodiments of the present invention may be implemented by a single device or combination of devices. In one or more embodiments of the present invention, images may be captured by a camera such as, but not limited to, a digital camera. The camera may be connected to a personal computer for communication of the captured images to the personal computer. The personal computer may then generate one or more three-dimensional images in accordance with embodiments of the present invention. After generation of the three-dimensional images, the personal computer may communicate the three-dimensional images to the camera for display on a suitable three-dimensional display. The camera may include a suitable three-dimensional display. Also, the camera may be in suitable electronic communication with a high-

definition television for display of the three-dimensional images on the television. The communication of the three-dimensional images may be, for example, via an HDMI connection.

[0129] In one or more other embodiments of the present invention, three-dimensional images may be generated by a camera and displayed by a separate suitable display. For example, the camera may capture conventional two-dimensional images and then use the captured images to generate three-dimensional images. The camera may be in suitable electronic communication with a high-definition television for display of the three-dimensional images on the television. The communication of the three-dimensional images may be, for example, via an HDMI connection.

[0130] The subject matter disclosed herein may be implemented by a digital still camera, a video camera, a mobile phone, a smart phone, phone, and the like. In order to provide additional context for various aspects of the disclosed subject matter, Figure 19 and the following discussion are intended to provide a brief, general description of a suitable operating environment 1900 in which various aspects of the disclosed subject matter may be implemented. While the invention is described in the general context of computer-executable instructions, such as program modules, executed by one or more computers or other devices, those skilled in the art will recognize that the disclosed subject matter can also be implemented in combination with other program modules and/or as a combination of hardware and software.

[0131] Generally, however, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular data types. The operating environment 1800 is only one example of a suitable operating environment and is not intended to suggest any limitation as to the scope of use or functionality of the subject matter disclosed herein. Other well known computer systems, environments, and/or configurations that may be suitable for use with the invention include but are not limited to, personal computers, handheld or laptop devices, multiprocessor systems, microprocessor-based systems, programmable consumer electronics, network PCs, minicomputers, mainframe computers, distributed computing environments that include the above systems or devices, and the like.

[0132] With reference to Figure 19, an exemplary environment 1900 for implementing various aspects of the subject matter disclosed herein includes a computer 1902. The computer 1902 includes a processing unit 1904, a system memory 1906, and a system bus 1908. The system bus 1908 couples system components including, but not limited to, the system memory 1906 to the processing unit 1904. The processing unit 1904 can be any of various available processors. Dual

microprocessors and other multiprocessor architectures also can be employed as the processing unit 1904.

[0133] The system bus 1908 can be any of several types of bus structure(s) including the memory bus or memory controller, a peripheral bus or external bus, and/or a local bus using any variety of available bus architectures including, but not limited to, 11-bit bus, Industrial Standard Architecture (ISA), Micro-Channel Architecture (MCA), Extended ISA (EISA), Intelligent Drive Electronics (IDE), VESA Local Bus (VLB), Peripheral Component Interconnect (PCI), Universal Serial Bus (USB), Advanced Graphics Port (AGP), Personal Computer Memory Card International Association bus (PCMCIA), and Small Computer Systems Interface (SCSI).

[0134] The system memory 1906 includes volatile memory 1910 and nonvolatile memory 1912. The basic input/output system (BIOS), containing the basic routines to transfer information between elements within the computer 1902, such as during start-up, is stored in nonvolatile memory 1912. By way of illustration, and not limitation, nonvolatile memory 1912 can include read only memory (ROM), programmable ROM (PROM), electrically programmable ROM (EPROM), electrically erasable ROM (EEPROM), or flash memory. Volatile memory 1910 includes random access memory (RAM), which acts as external cache memory. By way of illustration and not limitation, RAM is available in many forms such as synchronous RAM (SRAM), dynamic RAM (DRAM), synchronous DRAM (SDRAM), double data rate SDRAM (DDR SDRAM), enhanced SDRAM (ESDRAM), Synchlink DRAM (SLDRAM), and direct Rambus RAM (DRRAM).

[0135] Computer 1902 also includes removable/nonremovable, volatile/nonvolatile computer storage media. Figure 19 illustrates, for example, disk storage 1914. Disk storage 1914 includes, but is not limited to, devices like a magnetic disk drive, floppy disk drive, tape drive, Jaz drive, Zip drive, LS-100 drive, flash memory card, or memory stick. In addition, disk storage 1024 can include storage media separately or in combination with other storage media including, but not limited to, an optical disk drive such as a compact disk ROM device (CD-ROM), CD recordable drive (CD-R Drive), CD rewritable drive (CD-RW Drive) or a digital versatile disk ROM drive (DVD-ROM). To facilitate connection of the disk storage devices 1914 to the system bus 1908, a removable or non-removable interface is typically used such as interface 1916.

[0136] It is to be appreciated that Figure 19 describes software that acts as an intermediary between users and the basic computer resources described in suitable operating environment 1900. Such software includes an operating system 1918. Operating system 1918, which can be stored on disk storage 1914, acts to control and allocate resources of the computer system 1902. System

applications 1920 take advantage of the management of resources by operating system 1918 through program modules 1922 and program data 1924 stored either in system memory 1906 or on disk storage 1914. It is to be appreciated that the subject matter disclosed herein can be implemented with various operating systems or combinations of operating systems.

[0137] A user enters commands or information into the computer 1902 through input device(s) 1926. Input devices 1926 include, but are not limited to, a pointing device such as a mouse, trackball, stylus, touch pad, keyboard, microphone, joystick, game pad, satellite dish, scanner, TV tuner card, digital camera, digital video camera, web camera, and the like. These and other input devices connect to the processing unit 1904 through the system bus 1908 via interface port(s) 1928. Interface port(s) 1928 include, for example, a serial port, a parallel port, a game port, and a universal serial bus (USB). Output device(s) 1930 use some of the same type of ports as input device(s) 1926. Thus, for example, a USB port may be used to provide input to computer 1902 and to output information from computer 1902 to an output device 1930. Output adapter 1932 is provided to illustrate that there are some output devices 1930 like monitors, speakers, and printers among other output devices 1930 that require special adapters. The output adapters 1932 include, by way of illustration and not limitation, video and sound cards that provide a means of connection between the output device 1930 and the system bus 1908. It should be noted that other devices and/or systems of devices provide both input and output capabilities such as remote computer(s) 1934.

[0138] Computer 1902 can operate in a networked environment using logical connections to one or more remote computers, such as remote computer(s) 1934. The remote computer(s) 1934 can be a personal computer, a server, a router, a network PC, a workstation, a microprocessor based appliance, a peer device or other common network node and the like, and typically includes many or all of the elements described relative to computer 1902. For purposes of brevity, only a memory storage device 1936 is illustrated with remote computer(s) 1934. Remote computer(s) 1934 is logically connected to computer 1902 through a network interface 1938 and then physically connected via communication connection 1940. Network interface 1938 encompasses communication networks such as local-area networks (LAN) and wide-area networks (WAN). LAN technologies include Fiber Distributed Data Interface (FDDI), Copper Distributed Data Interface (CDDI), Ethernet/IEEE 1102.3, Token Ring/IEEE 1102.5 and the like. WAN technologies include, but are not limited to, point-to-point links, circuit switching networks like Integrated Services

Digital Networks (ISDN) and variations thereon, packet switching networks, and Digital Subscriber Lines (DSL).

[0139] Communication connection(s) 1940 refers to the hardware/software employed to connect the network interface 1938 to the bus 1908. While communication connection 1940 is shown for illustrative clarity inside computer 1902, it can also be external to computer 1902. The hardware/software necessary for connection to the network interface 1938 includes, for exemplary purposes only, internal and external technologies such as, modems including regular telephone grade modems, cable modems and DSL modems, ISDN adapters, and Ethernet cards.

[0140] The various techniques described herein may be implemented with hardware or software or, where appropriate, with a combination of both. Thus, the methods and apparatus of the disclosed embodiments, or certain aspects or portions thereof, may take the form of program code (i.e., instructions) embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other machine-readable storage medium, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. In the case of program code execution on programmable computers, the computer will generally include a processor, a storage medium readable by the processor (including volatile and non-volatile memory and/or storage elements), at least one input device and at least one output device. One or more programs are preferably implemented in a high level procedural or object oriented programming language to communicate with a computer system. However, the program(s) can be implemented in assembly or machine language, if desired. In any case, the language may be a compiled or interpreted language, and combined with hardware implementations.

[0141] The described methods and apparatus may also be embodied in the form of program code that is transmitted over some transmission medium, such as over electrical wiring or cabling, through fiber optics, or via any other form of transmission, wherein, when the program code is received and loaded into and executed by a machine, such as an EPROM, a gate array, a programmable logic device (PLD), a client computer, a video recorder or the like, the machine becomes an apparatus for practicing the invention. When implemented on a general-purpose processor, the program code combines with the processor to provide a unique apparatus that operates to perform the processing of the present invention.

[0142] While the embodiments have been described in connection with the preferred embodiments of the various figures, it is to be understood that other similar embodiments may be used or modifications and additions may be made to the described embodiment for performing the

same function without deviating therefrom. Therefore, the disclosed embodiments should not be limited to any single embodiment, but rather should be construed in breadth and scope in accordance with the appended claims.

## CLAIMS

### What is Claimed:

1. A method for creating a three-dimensional image of a scene, the method comprising:  
using at least one processor for:  
receiving a plurality of images of a scene;  
determining attributes of the plurality of images; and  
generating, based on the attributes, a pair of images from among the plurality of images for use in creating a three-dimensional image.
2. The method of claim 1 wherein the pair of images comprises different perspective views of the scene.
3. The method of claim 1 further comprising:  
comparing color attributes of the images;  
determining images in which the color attributes are within a predetermined color threshold;  
and  
modifying the images within the predetermined color threshold for creating the three-dimensional image.
4. The method of claim 1 further comprising:  
identifying similar objects in the plurality of images; and  
comparing images including the similar objects for occlusion.
5. The method of claim 1 further comprising:  
identifying similar objects in the plurality of images; and  
comparing vertical displacement of the similar objects with respect to a predetermined threshold level.
6. The method of claim 1 further comprising comparing color differences between images.
7. The method of claim 1 further comprising applying horizontal and vertical edge line analysis to images.
8. The method of claim 1 further comprising:  
applying a Hough transform to identify lines within the images; and  
determining perspective view changes between the images based on the identified lines.
9. The method of claim 1 wherein the three-dimensional image includes a stereoscopic pair representing a left view image and a right view image.

10. The method of claim 9 further comprising automatically determining the left view image and the right view image for creating the stereoscopic pair.
12. The method of claim 9 further comprising determining interest points within the left view image and the right view image for one or more of rectification and registration.
13. The method of claim 12 wherein determining interest points comprises:
  - performing horizontal and vertical edge detection;
  - filtering for strong edges of a minimum length; and
  - identifying crossing points of the strong edges.
14. The method of claim 9 further comprising:
  - determining a parallax disparity between the left view image and the right view image;
  - determining whether the parallax disparity meets a predetermined criteria; and
  - if the parallax disparity does not meet the predetermined criteria, adjusting an attribute of at least one pixel in one of the left view image and the right view image such that the parallax disparity meets the predetermined criteria.
15. The method of claim 9 further comprising cropping one of the left view image and the right view image.
16. The method of claim 9 further comprising:
  - determining a parallax disparity between at least a portion of an object in the left view image and the right view image;
  - determining whether the parallax disparity is greater than a predetermined threshold level;
  - and
  - if the parallax disparity is greater than the predetermined threshold level, removing the object from the left view image and the right view image.
17. The method of claim 1 further comprising using an image capture device for capturing the plurality of images of the scene.
18. The method of claim 17 wherein using an image capture device comprises using at least one of a digital still camera, a video camera, a mobile phone, and a smart phone for capturing the plurality of images of the scene.
19. The method of claim 1 further comprising generating a three-dimensional image of the scene using the pair of images.

20. The method of claim 19 wherein generating a three-dimensional image comprises one or more of registration, rectification, color correction, matching edges of the pair of images, transformation, depth adjustment, motion detection, and removal of moving objects.
21. The method of claim 1 further comprising displaying the three-dimensional image on a suitable three-dimensional image display.
22. The method of claim 21 wherein displaying the three-dimensional image comprises displaying the three-dimensional image on a device such as a digital still camera, a computer, a video camera, a digital picture frame, a set-top box, phone, or a high-definition television.
23. The method of claim 1 further comprising:  
capturing the plurality of images at a camera;  
communicating the plurality of images to a computer;  
at the computer, the step of determining the attributes of the plurality of images;  
at the computer, creating the three-dimensional image; and  
at the computer communicating the three-dimensional image to the camera.
24. A system for creating a three-dimensional image of a scene, the system comprising:  
a memory having stored therein computer-executable instructions;  
a computer processor that executes the computer-executable instructions;  
an image generator configured to:  
receive a plurality of images of a scene;  
determine attributes of the plurality of images; and  
determine, based on the attributes, a pair of images from among the plurality of images for use in generating a three-dimensional image.
25. A computer-readable storage medium having stored thereon computer executable instructions for performing the following steps:  
receiving a plurality of images of a scene;  
determining attributes of the plurality of images; and  
generating, based on the attributes, a pair of images from among the plurality of images for use in generating a three-dimensional image.
26. A method for generating a three-dimensional video of a scene, the method comprising:  
using at least one processor for:  
receiving a video sequence comprising a plurality of still images of a scene;  
determining attributes of the plurality of still images; and

determining, based on the attributes, a plurality of stereoscopic image pairs from among the plurality of still images for use in generating a three-dimensional video.

27. The method of claim 26 further comprising adjusting at least one of a parallax and perspective of each image pair such that at least one of a predetermined orientation and depth results.

28. The method of claim 26 further comprising sequencing stereoscopic image pairs in the three-dimensional video according to an order of the video sequence of the still images.

29. The method of claim 26 further comprising compressing the image pairs of the three-dimensional video.

30. A system for generating a three-dimensional video of a scene, the system comprising:  
a memory having stored therein computer-executable instructions;  
a computer processor that executes the computer-executable instructions;  
an image generator configured to:

receive a video sequence comprising a plurality of still images of a scene;

determine attributes of the plurality of still images; and

determine, based on the attributes, a plurality of stereoscopic image pairs from among the plurality of still images for use in generating a three-dimensional video.

31. A computer-readable storage medium having stored thereon computer executable instructions for performing the following steps:

receiving a video sequence comprising a plurality of still images of a scene;

determining attributes of the plurality of still images; and

determining, based on the attributes, a plurality of stereoscopic image pairs from among the plurality of still images for use in generating a three-dimensional video.

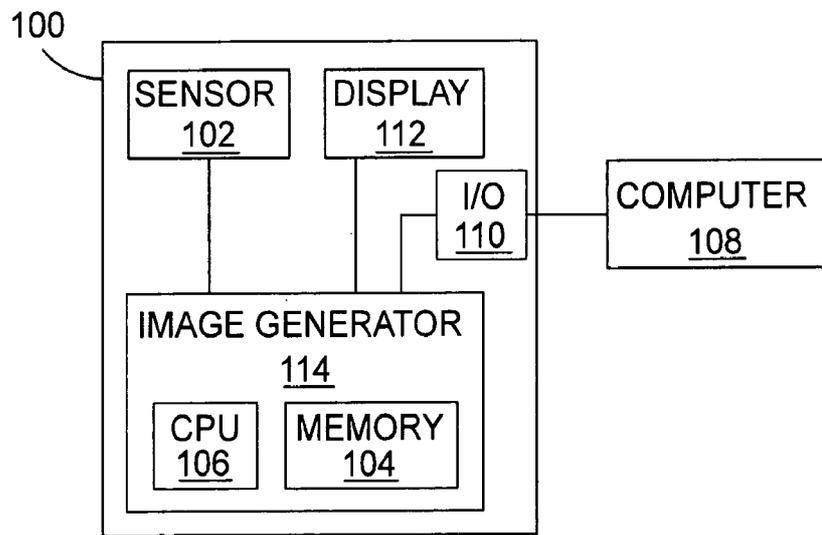


Figure 1

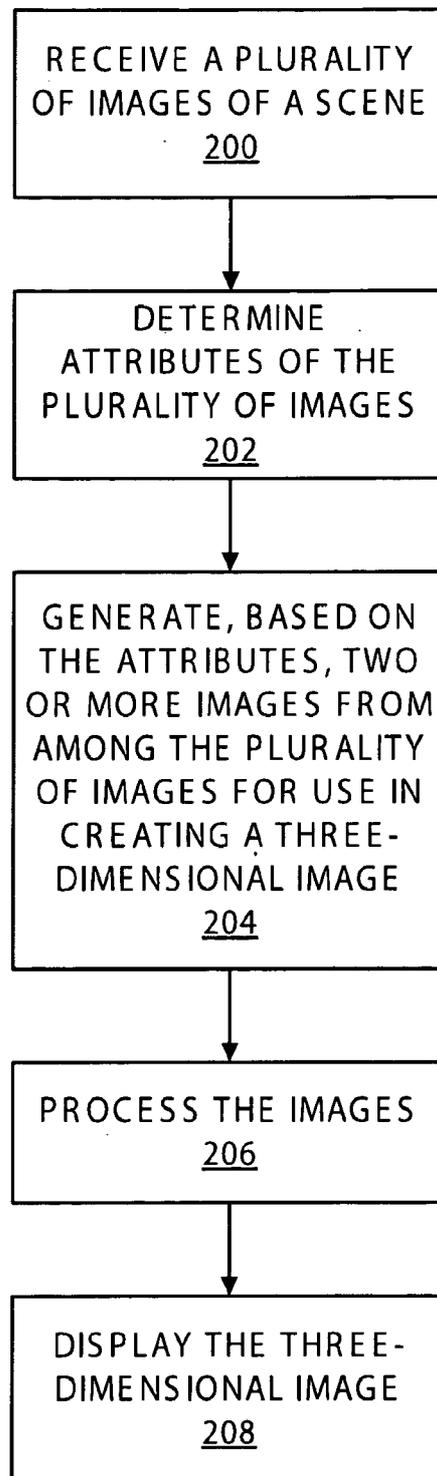


Figure 2

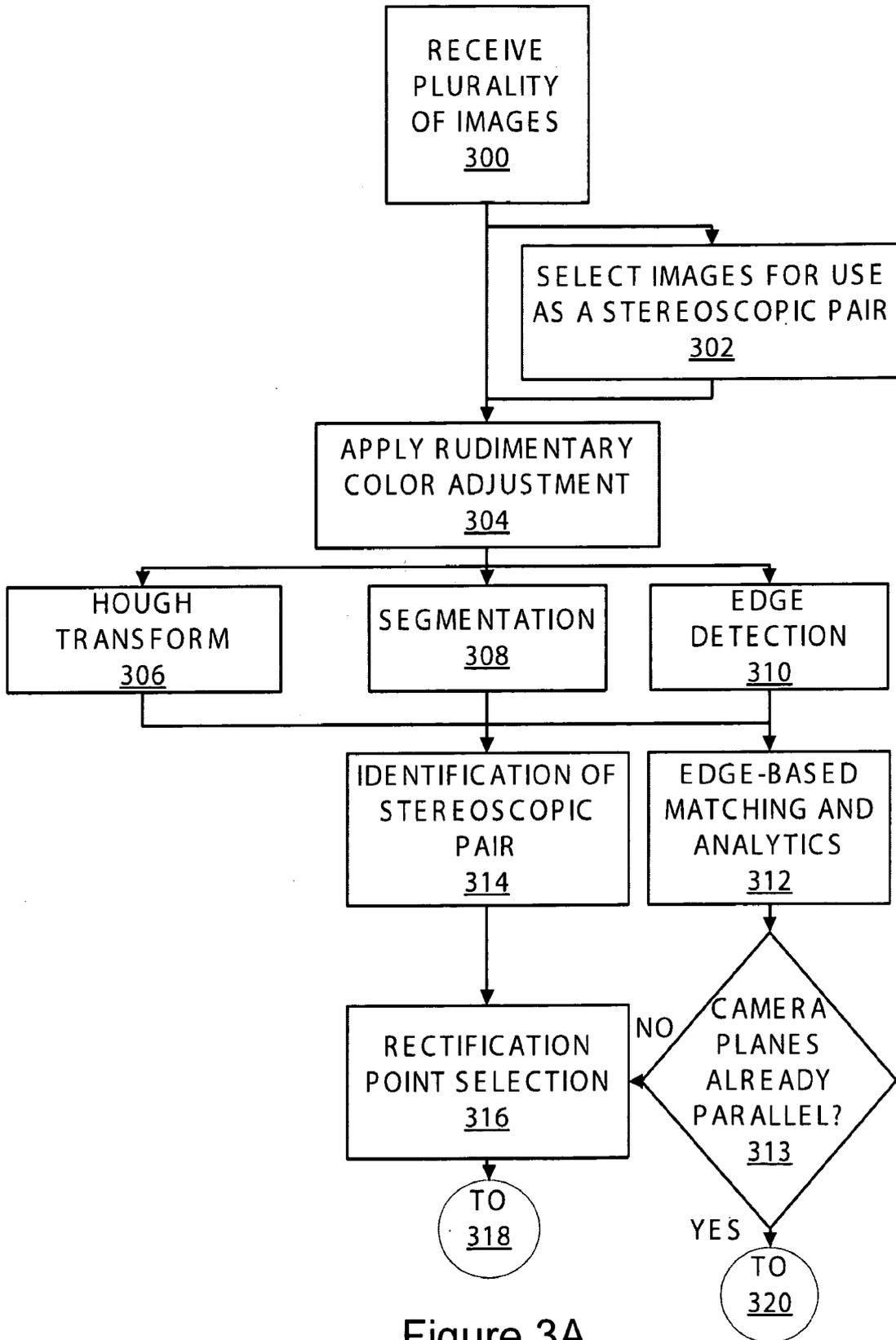


Figure 3A

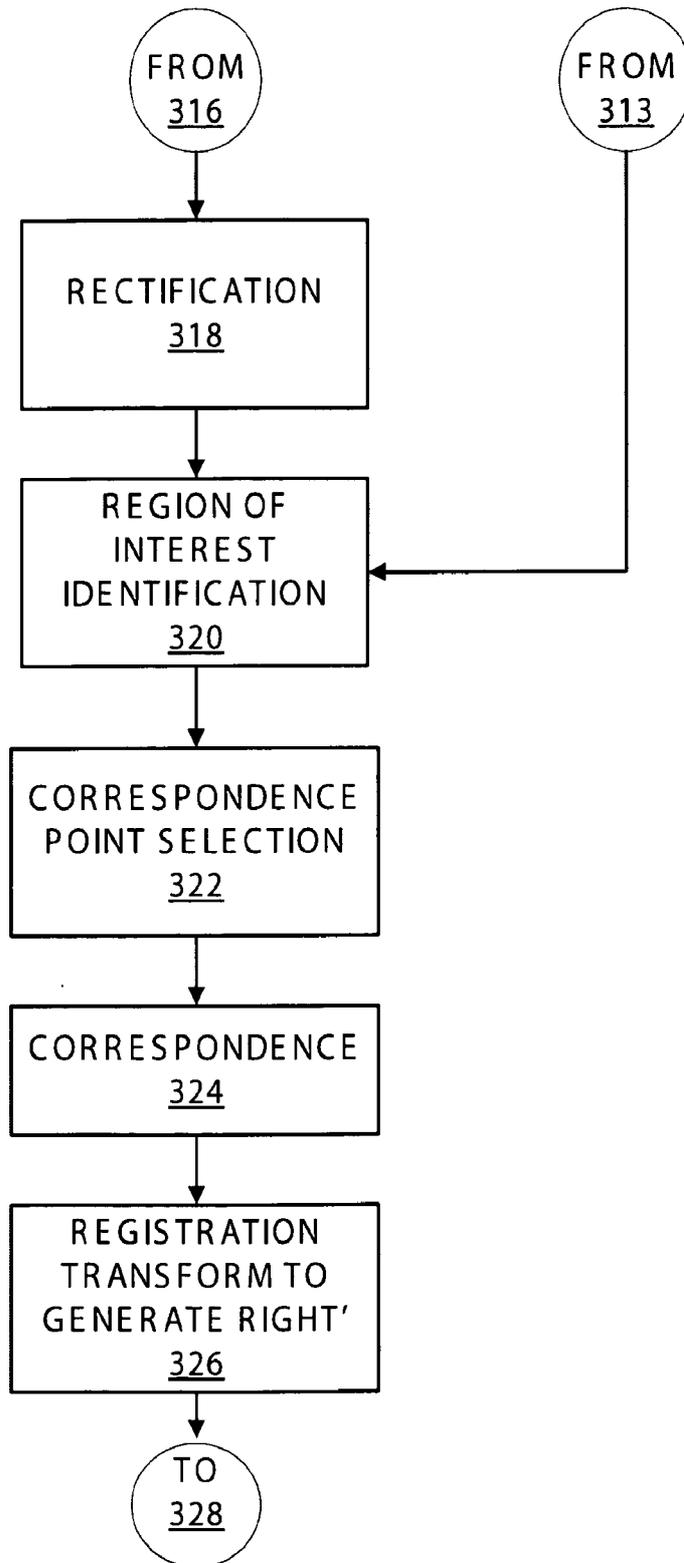


Figure 3B

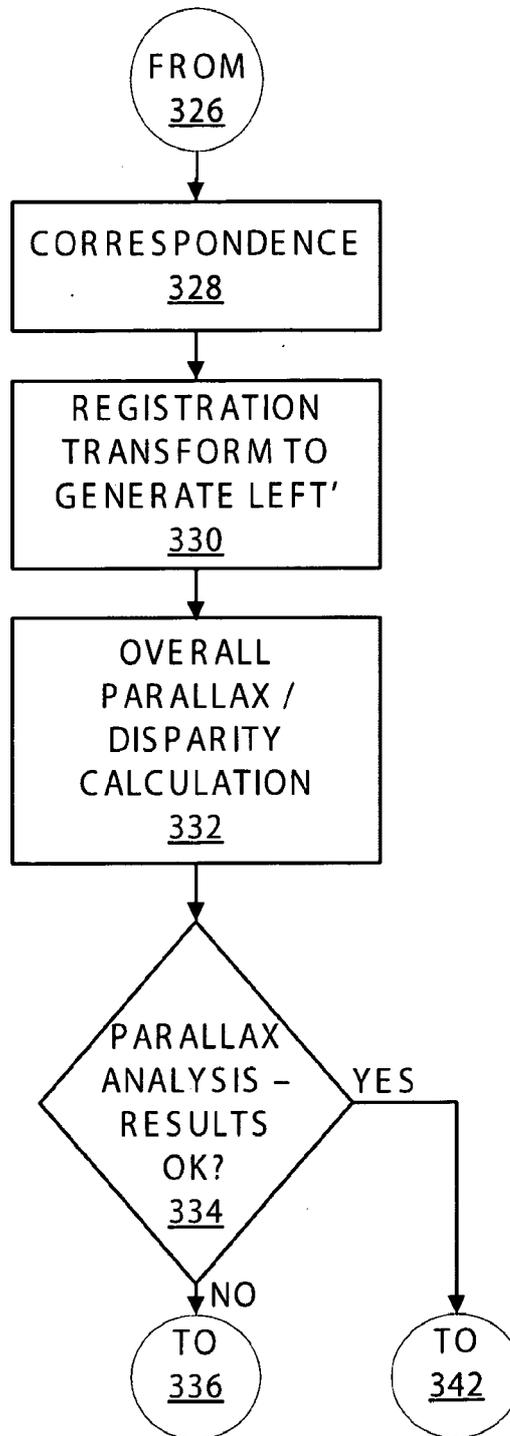


Figure 3C

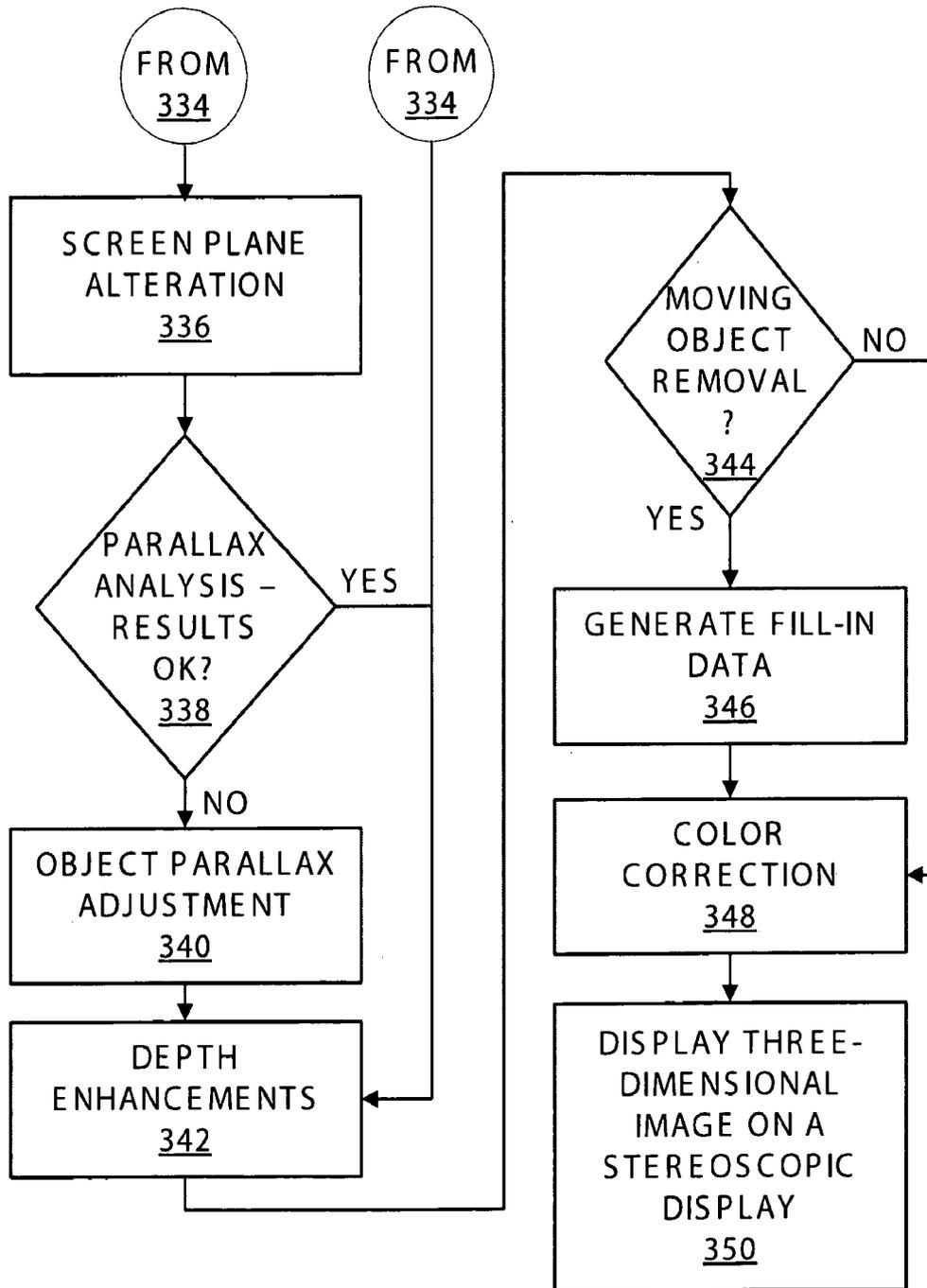
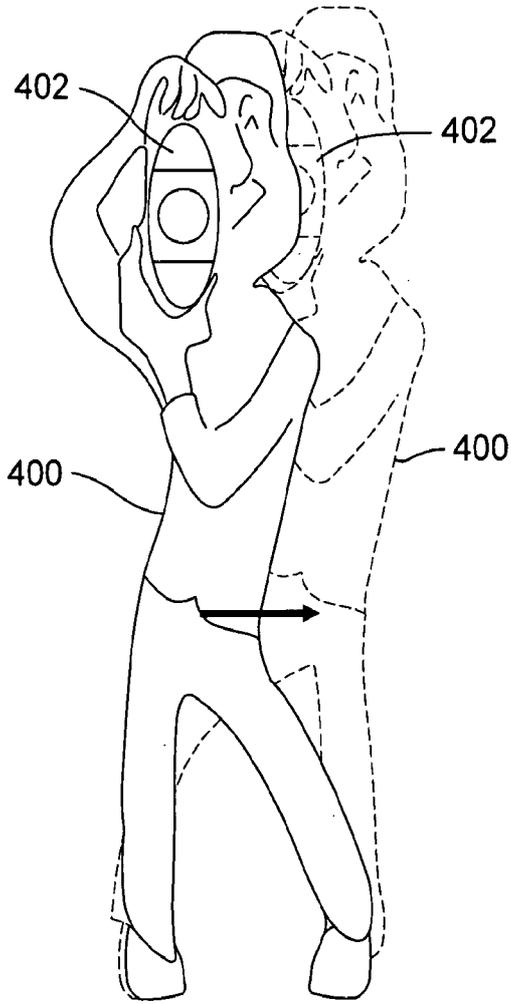
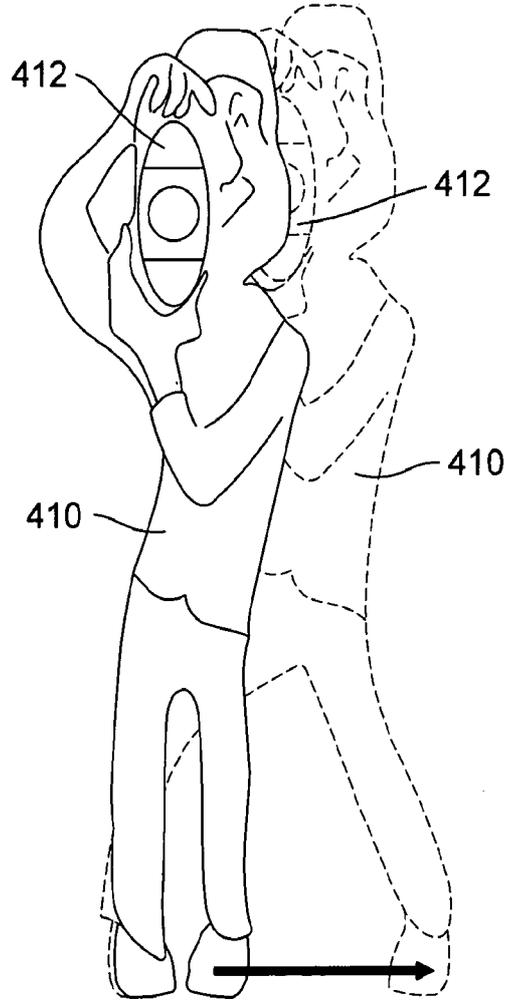


Figure 3D



ACTUAL BINOCULAR DISTANCE = DISTANCE BETWEEN FEET

Figure 4A



ACTUAL BINOCULAR DISTANCE = 1/2 DISTANCE BETWEEN FEET

Figure 4B

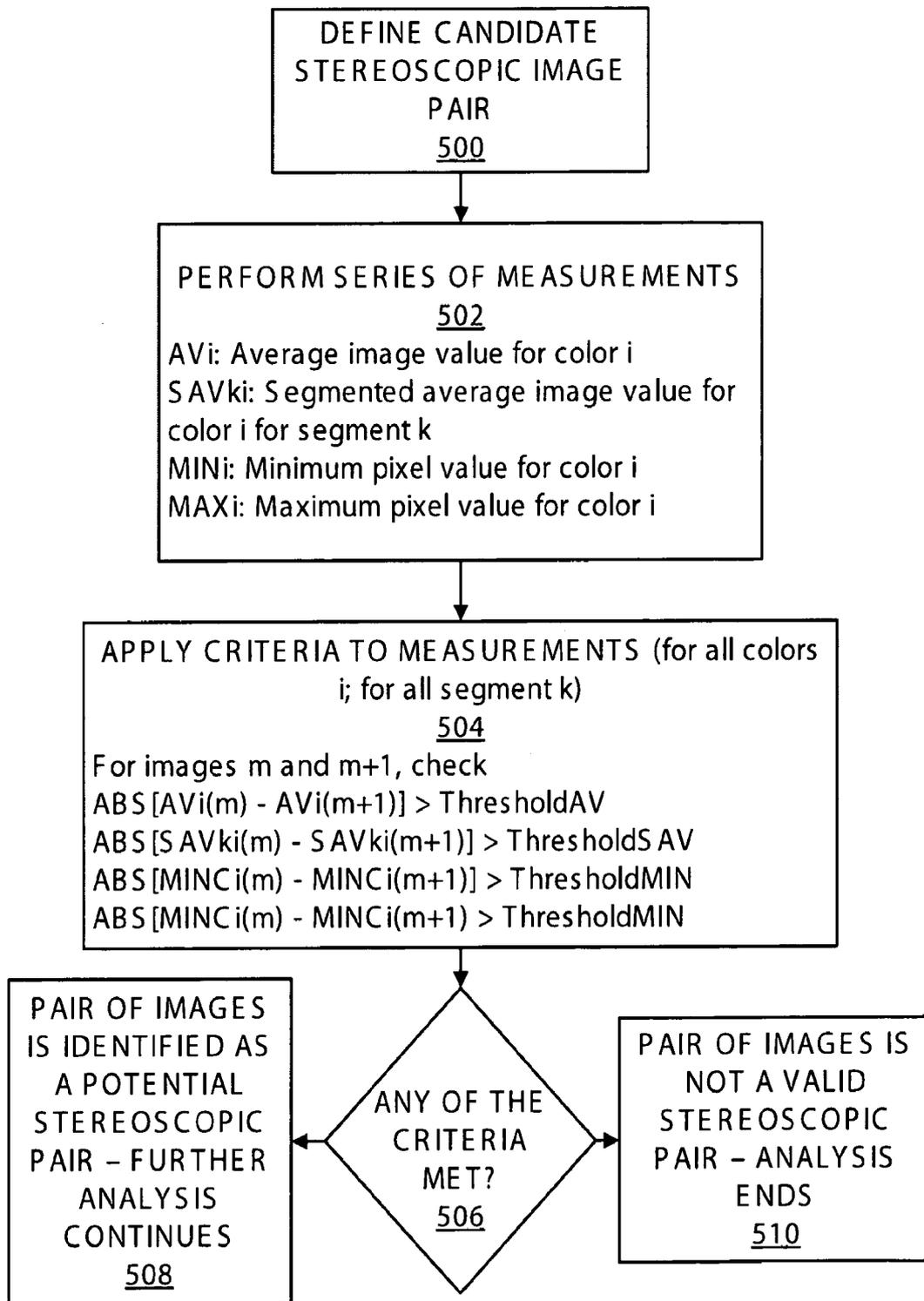


Figure 5

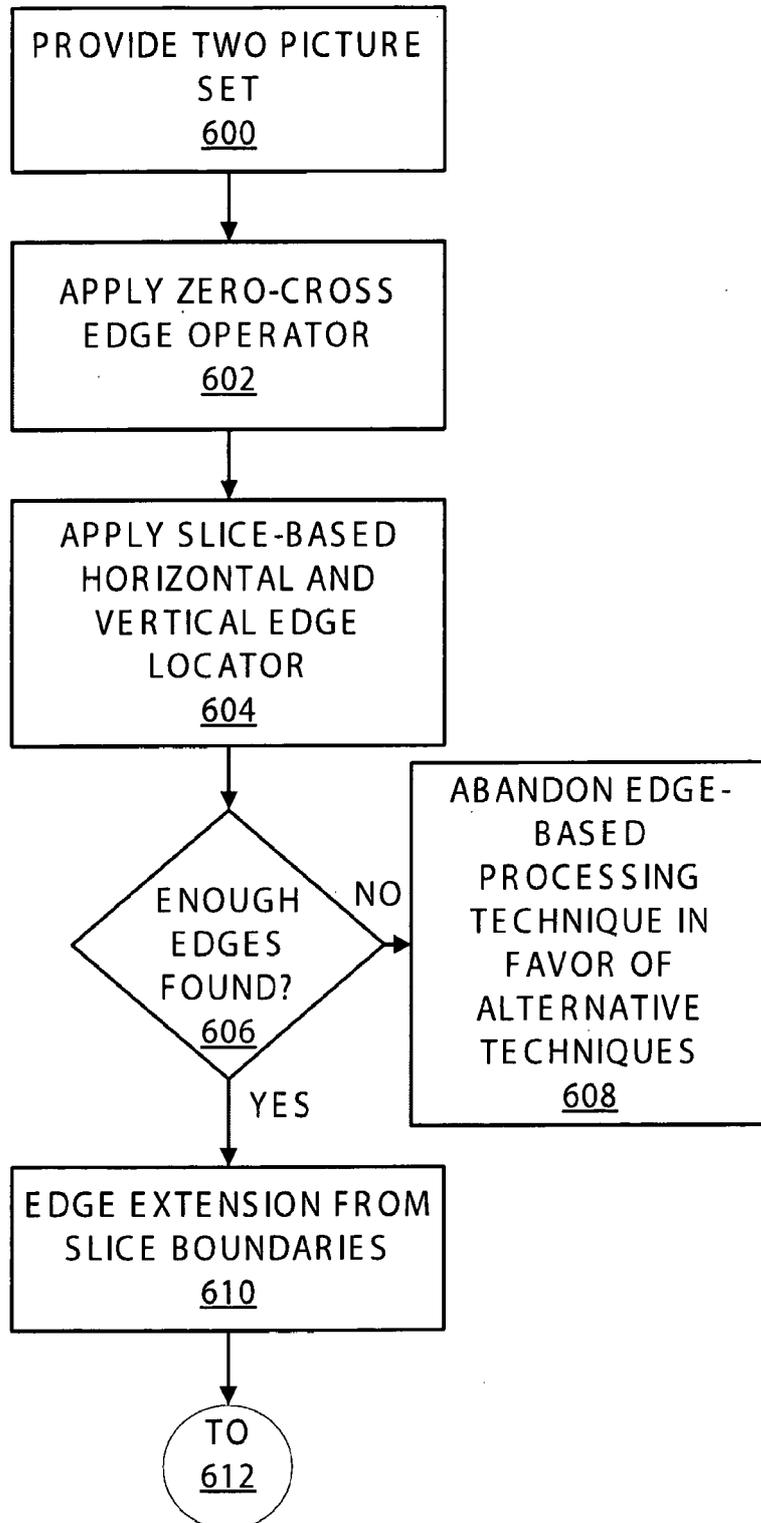


Figure 6A

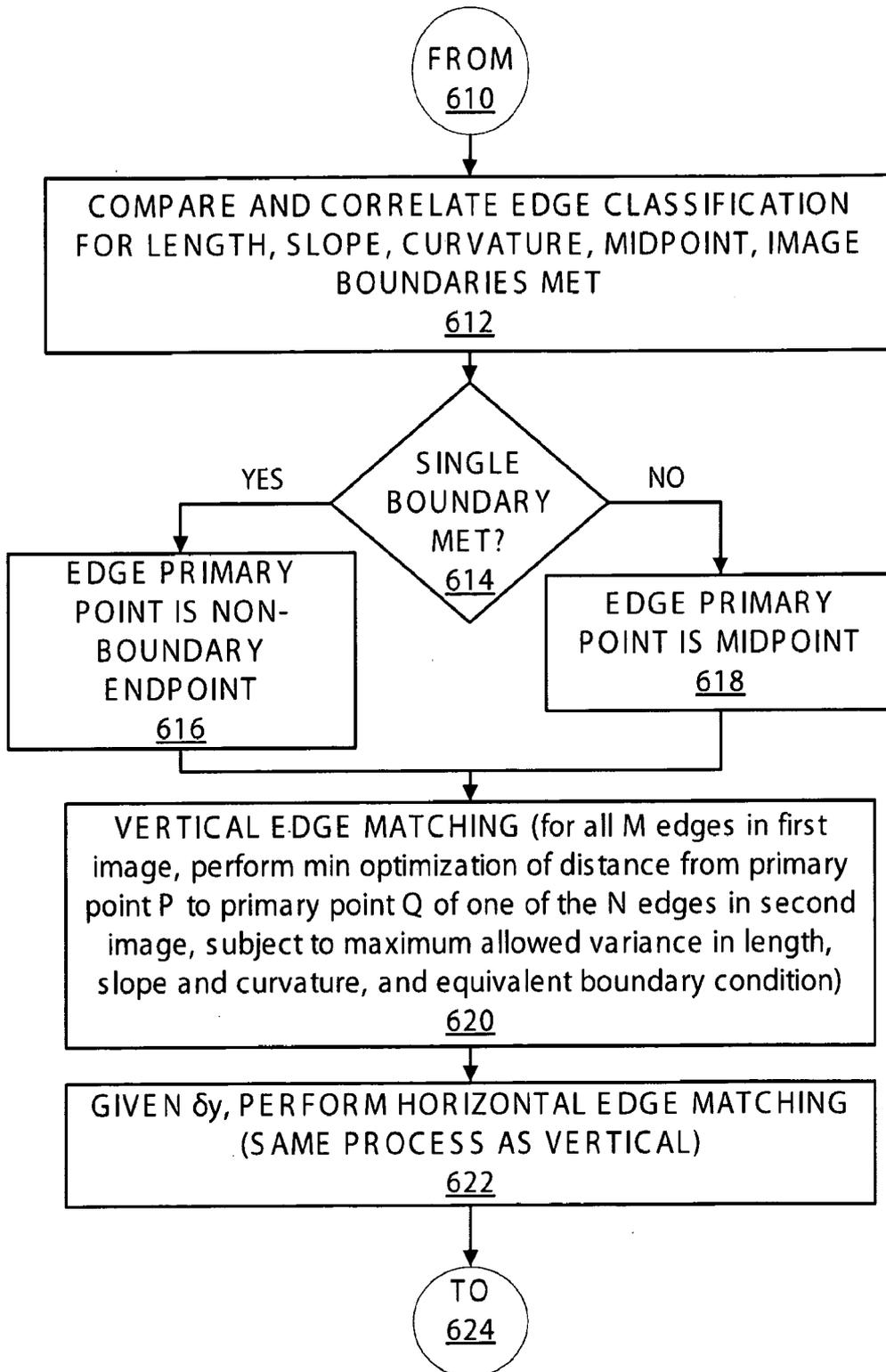


Figure 6B

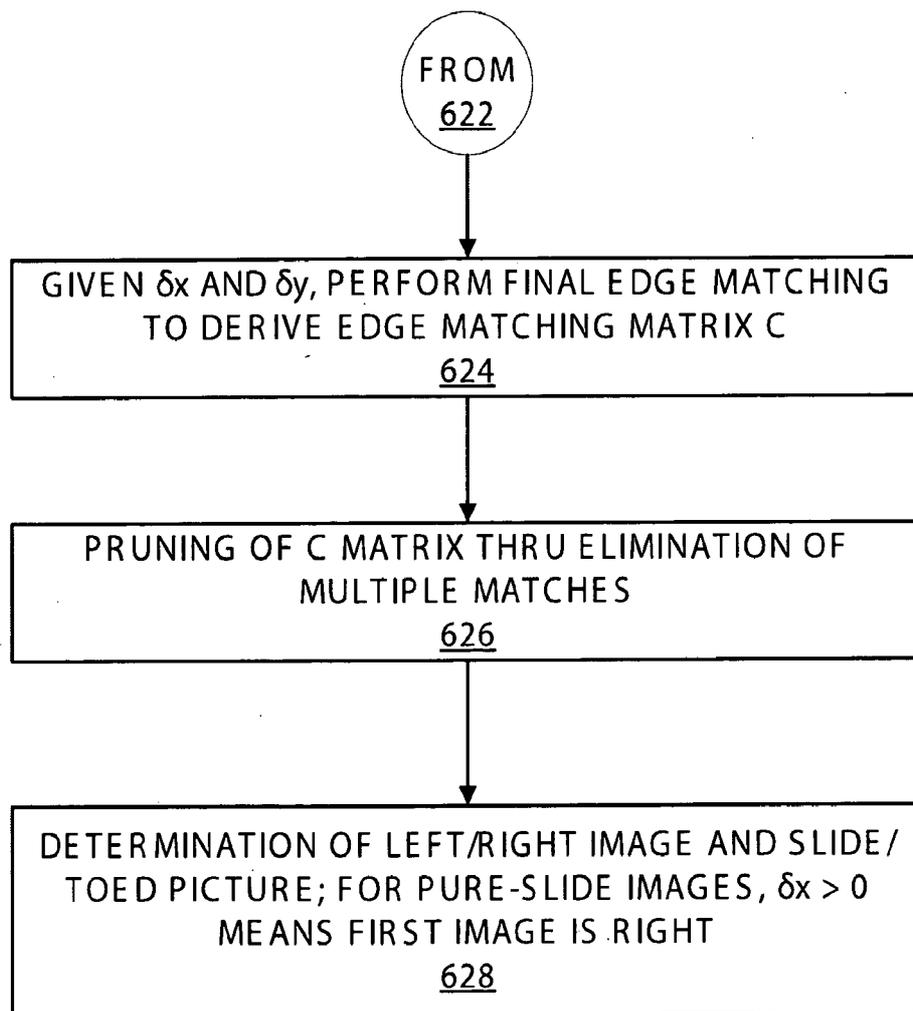


Figure 6C

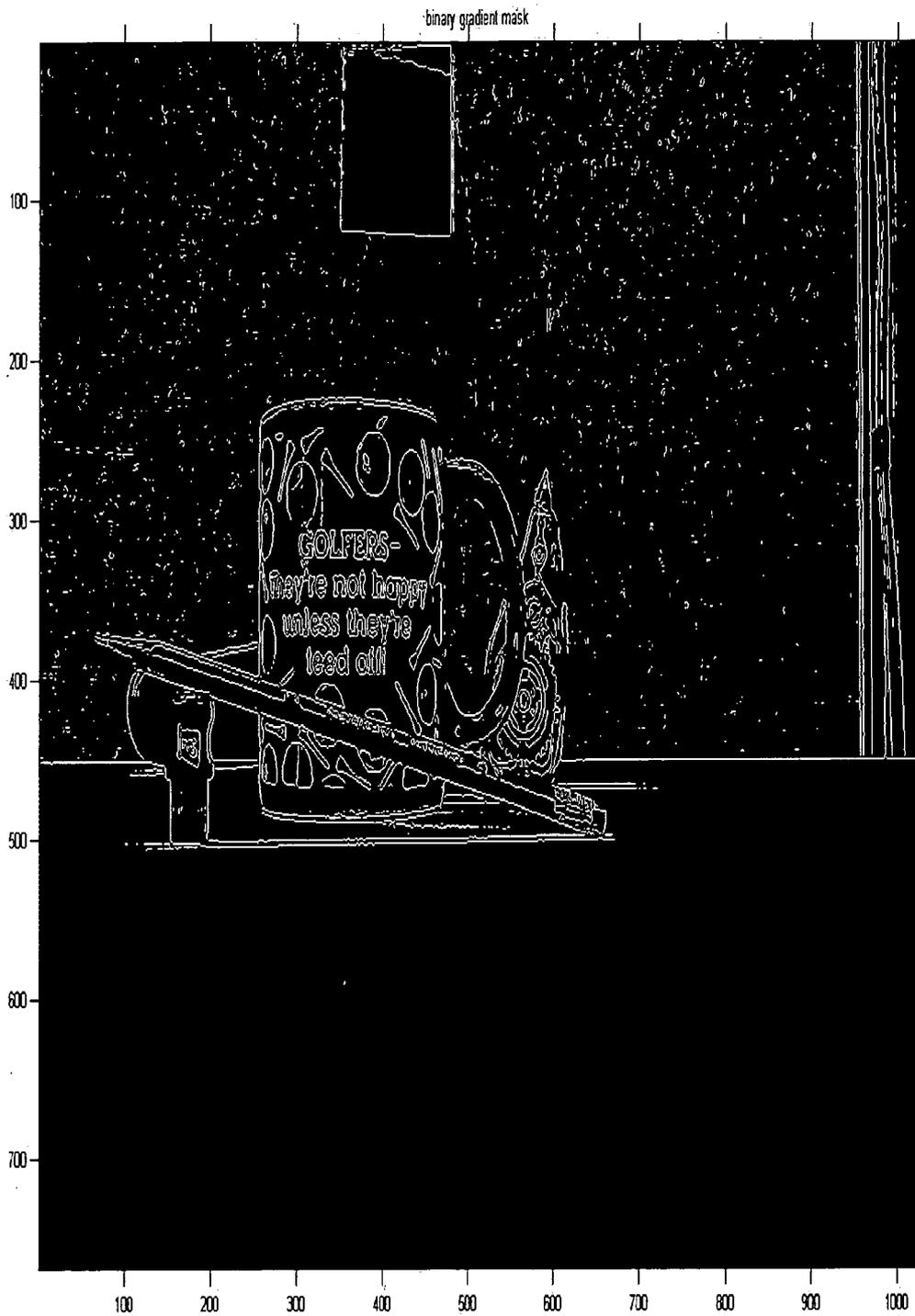


Figure 7

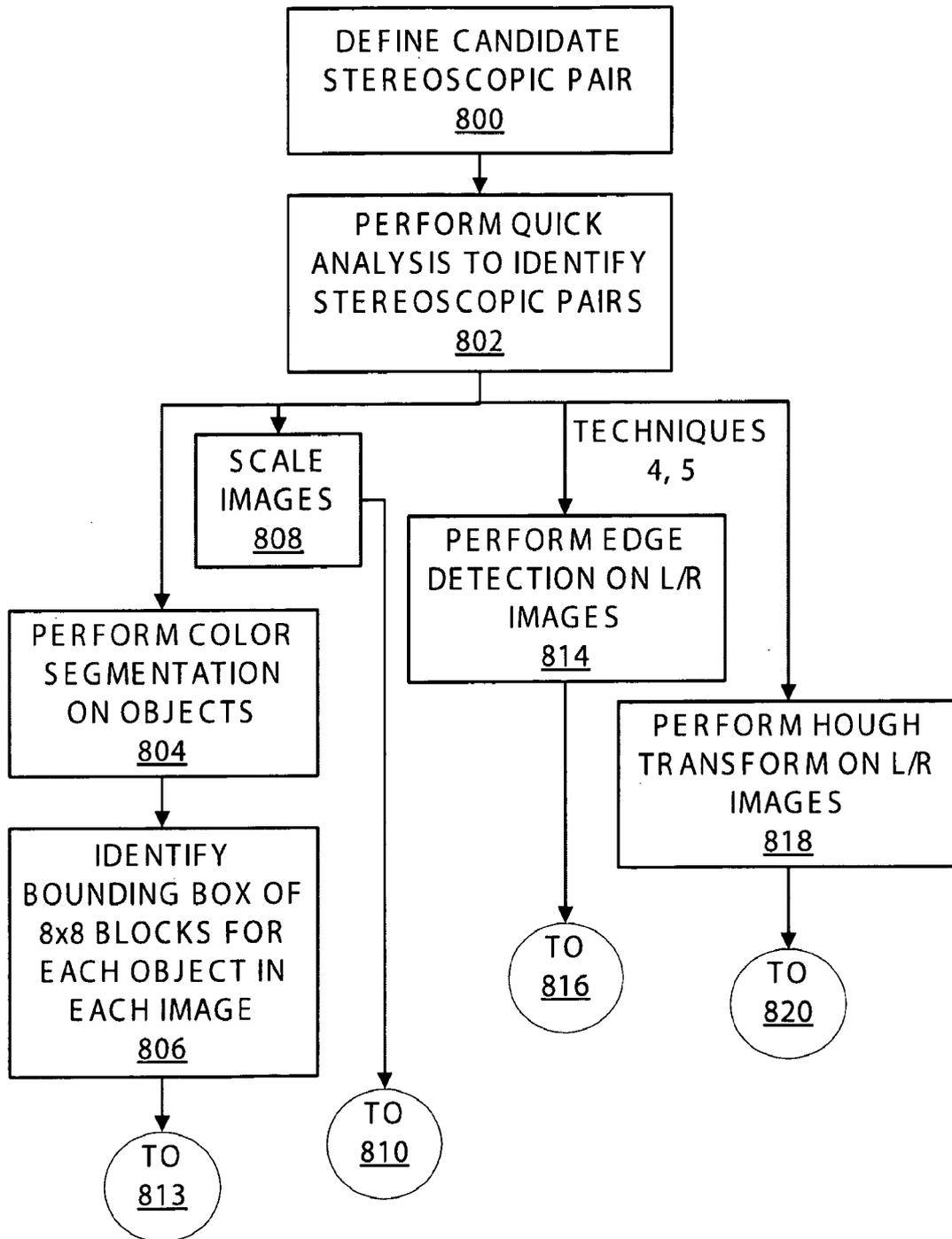


Figure 8A

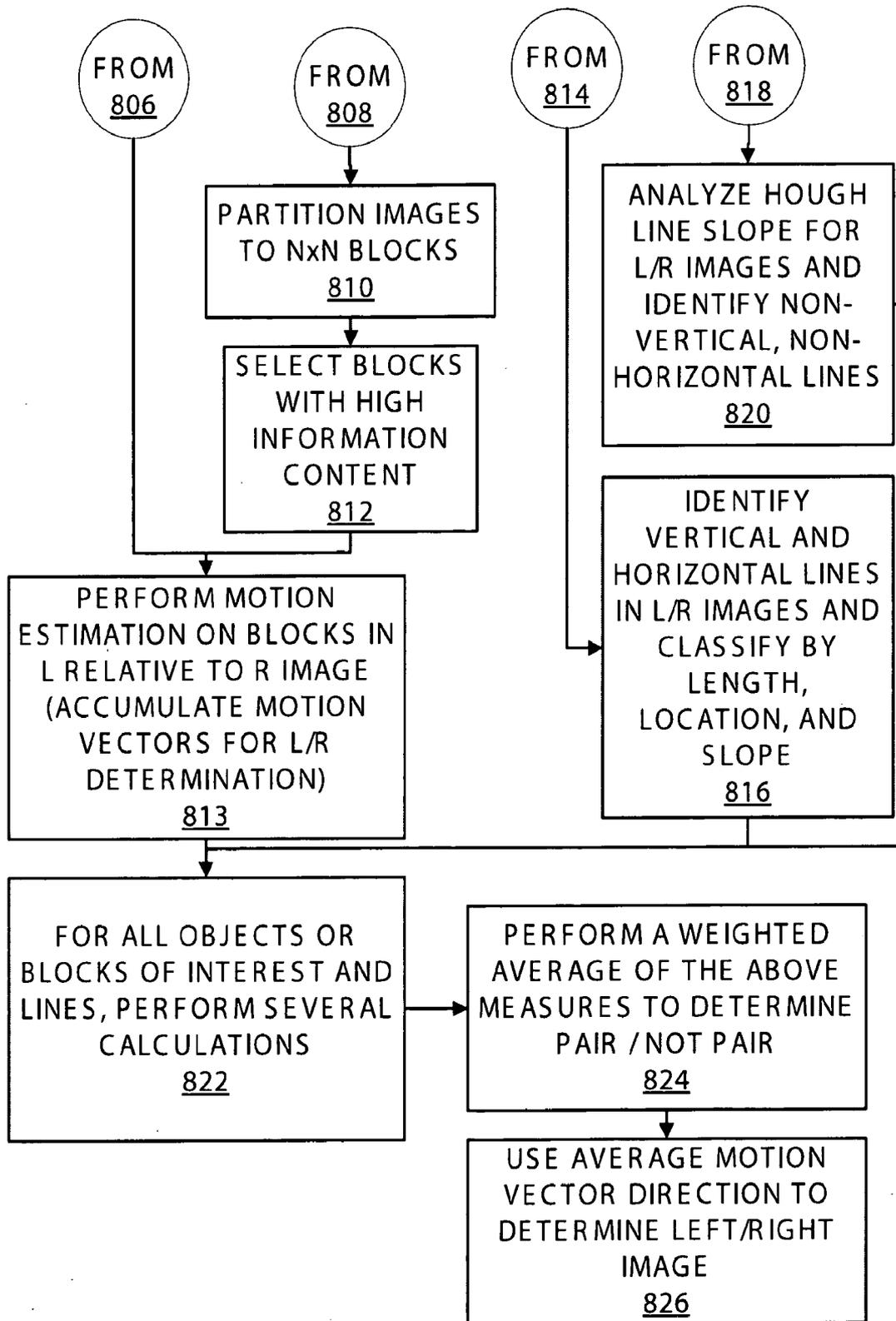


Figure 8B

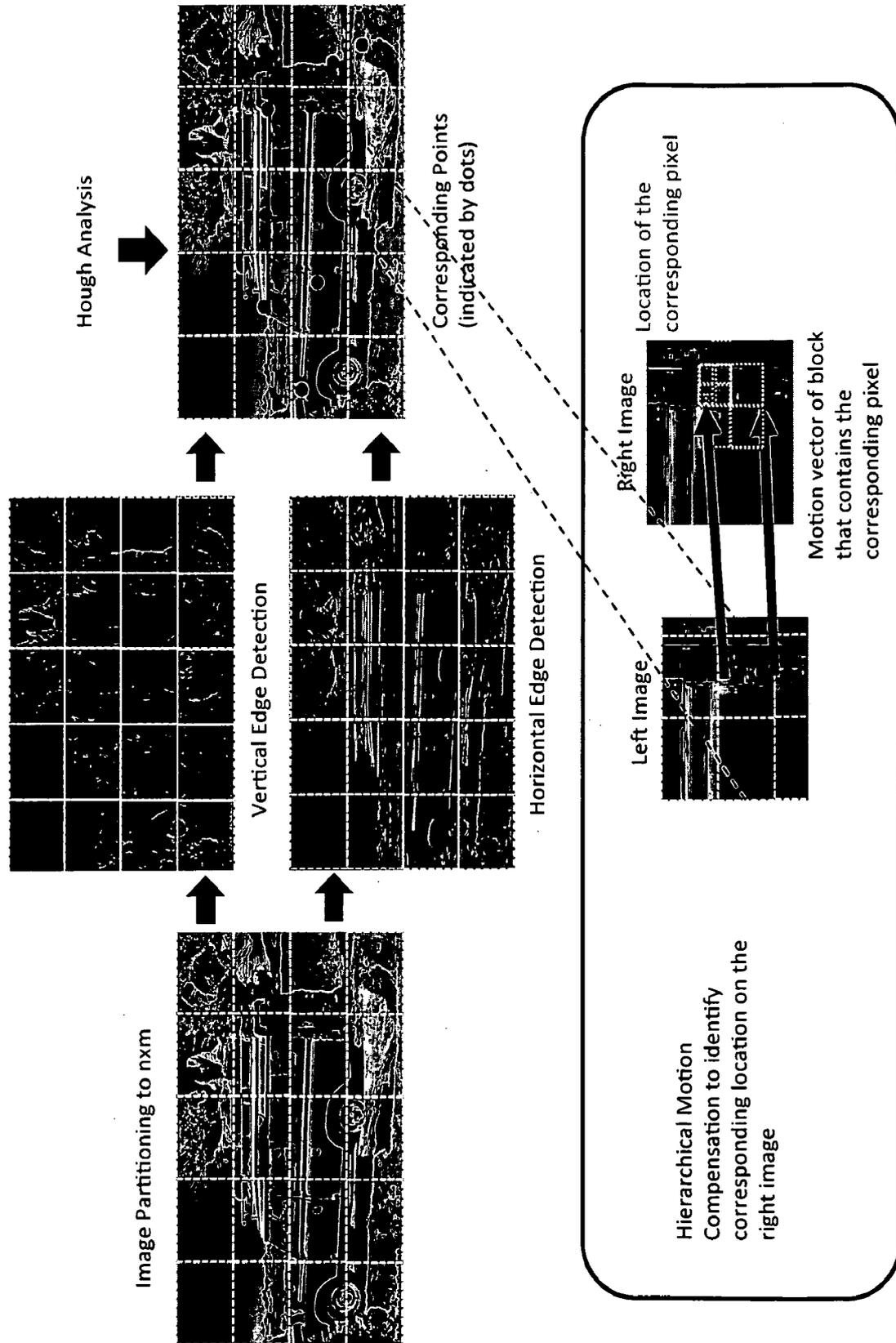


Figure 9

Figure 10

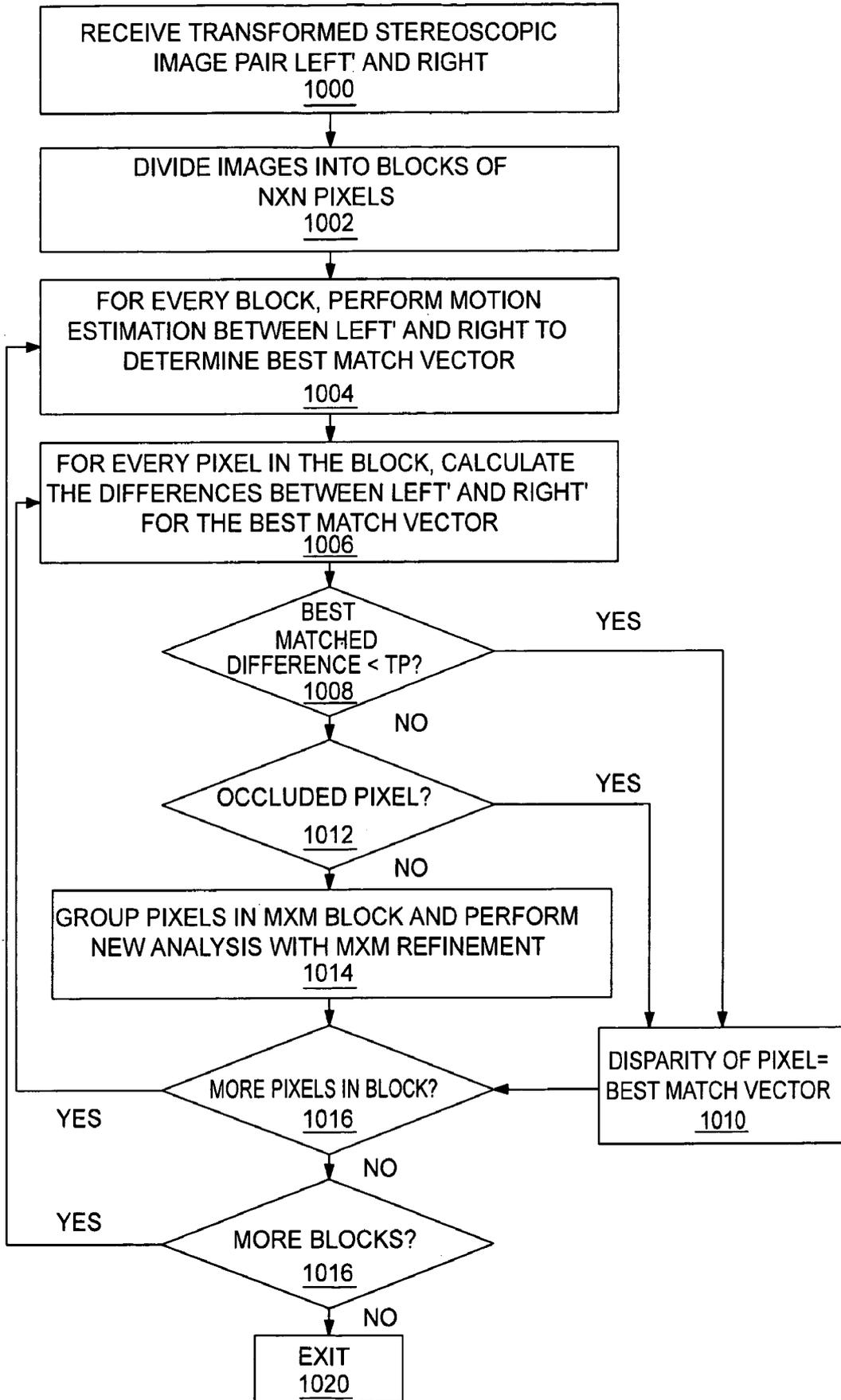
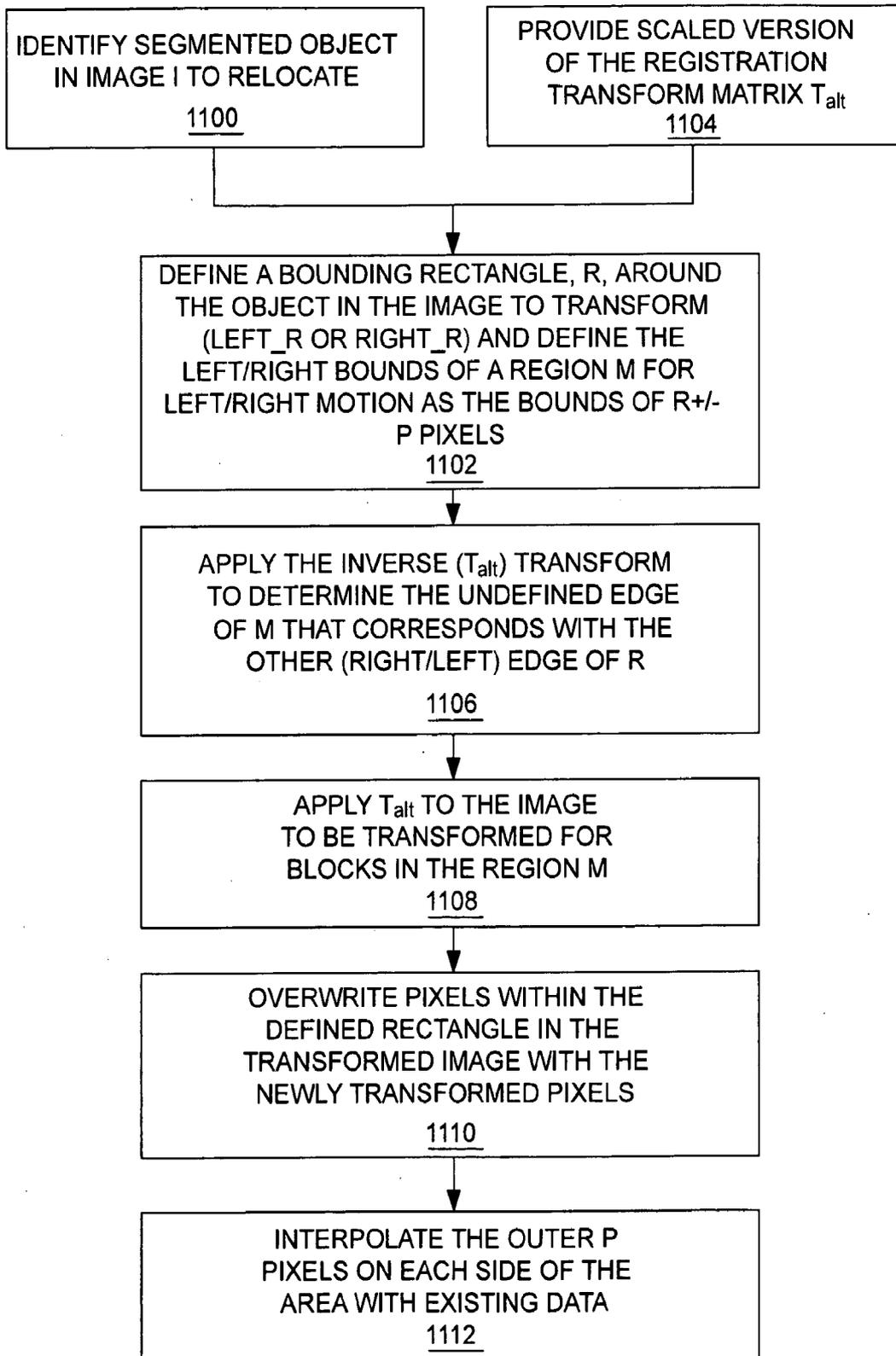


Figure 11



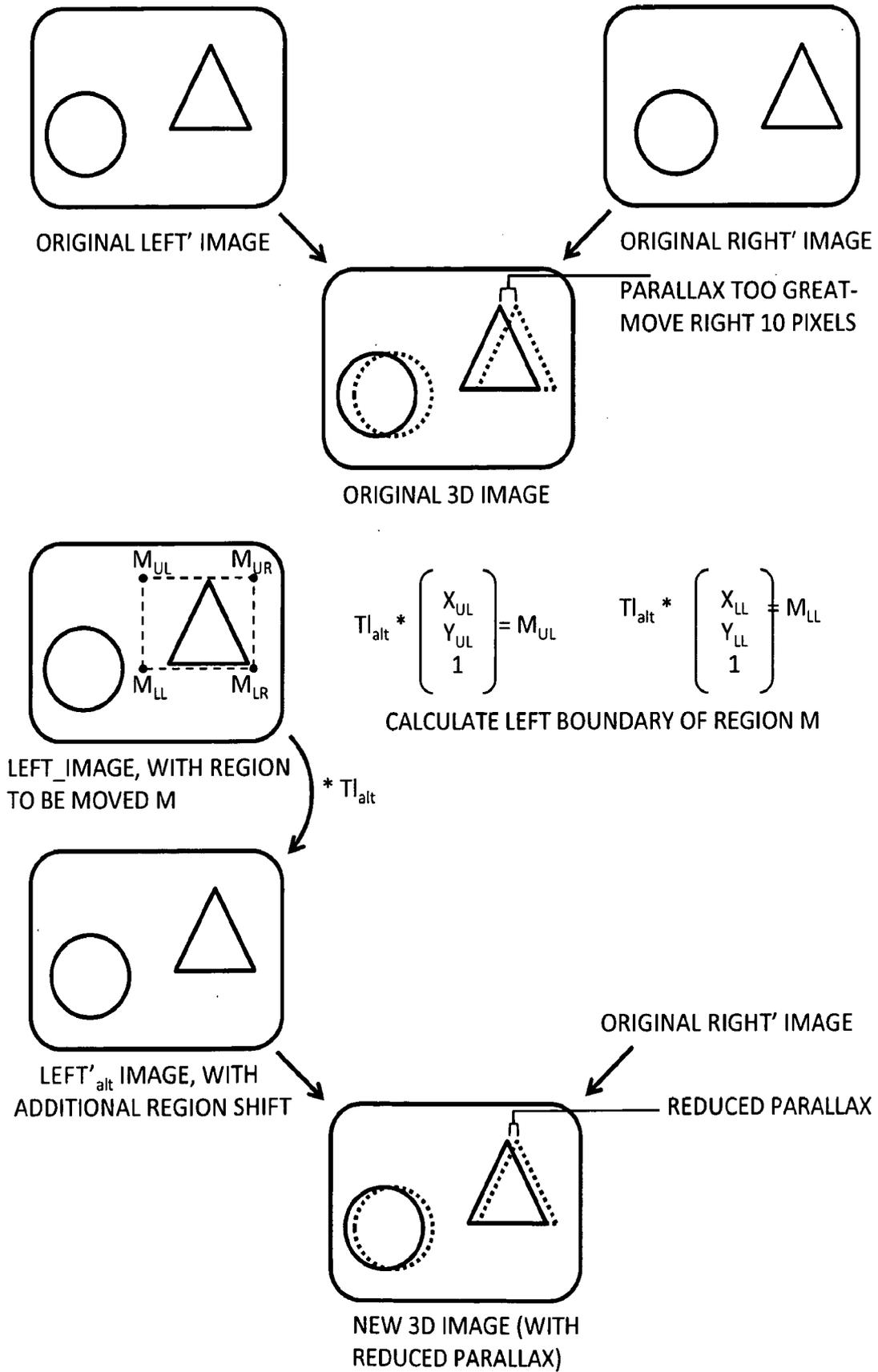


Figure 12

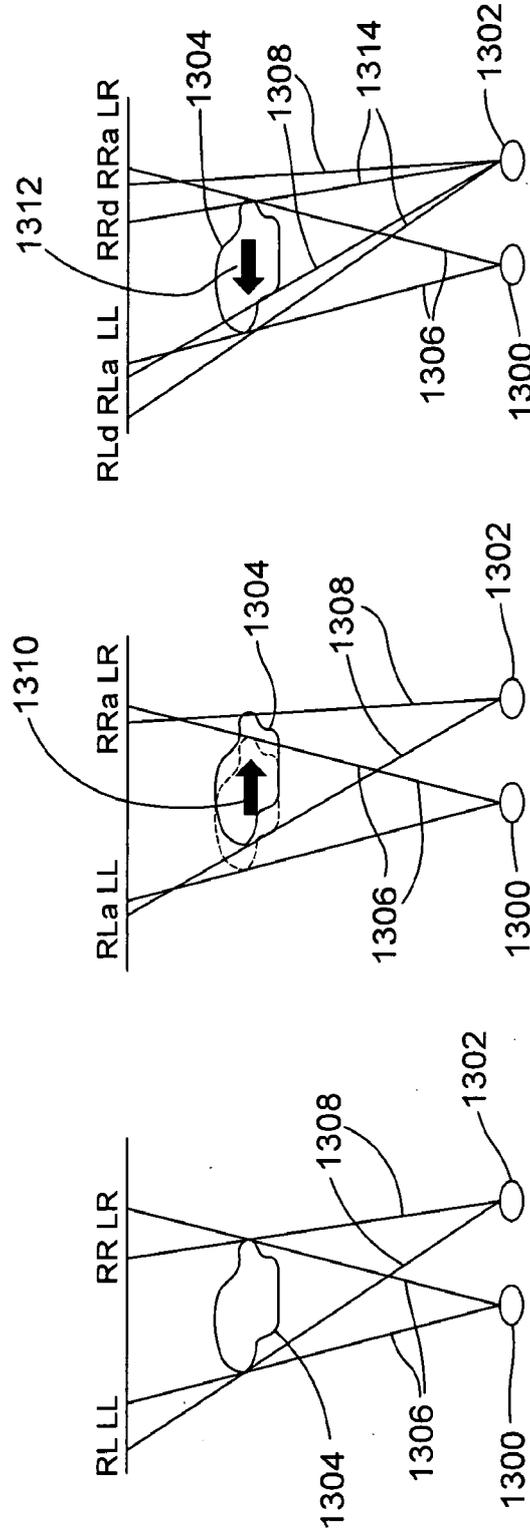


Figure 13C

Figure 13B

Figure 13A

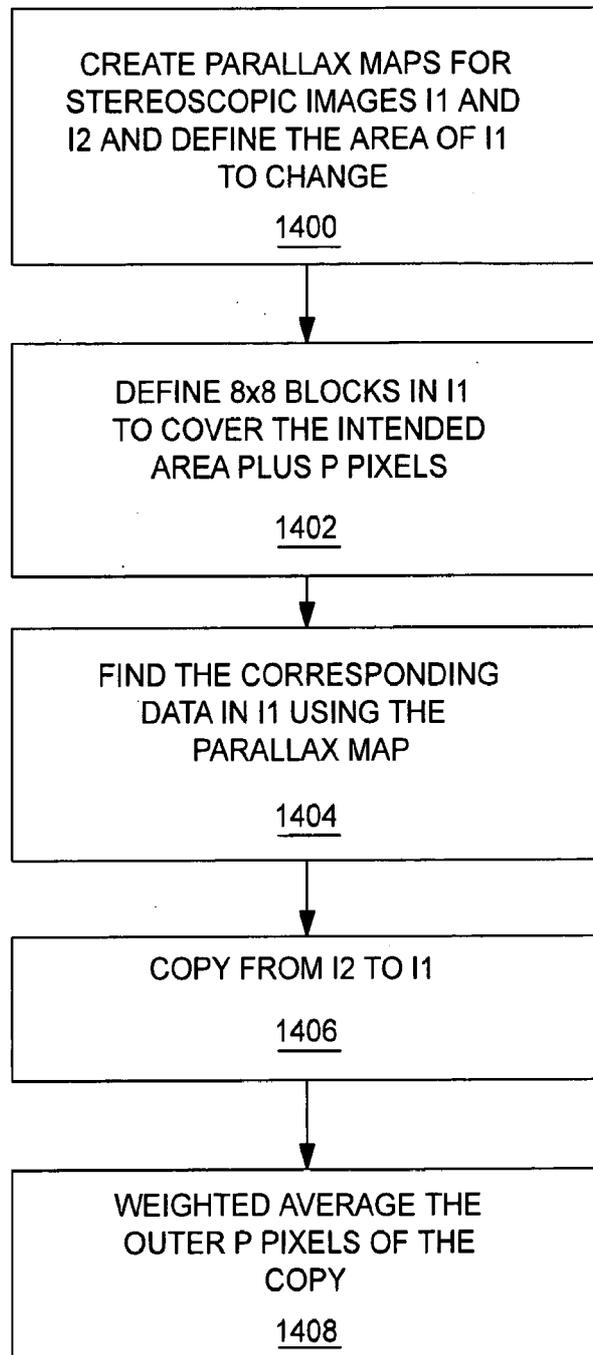


Figure 14

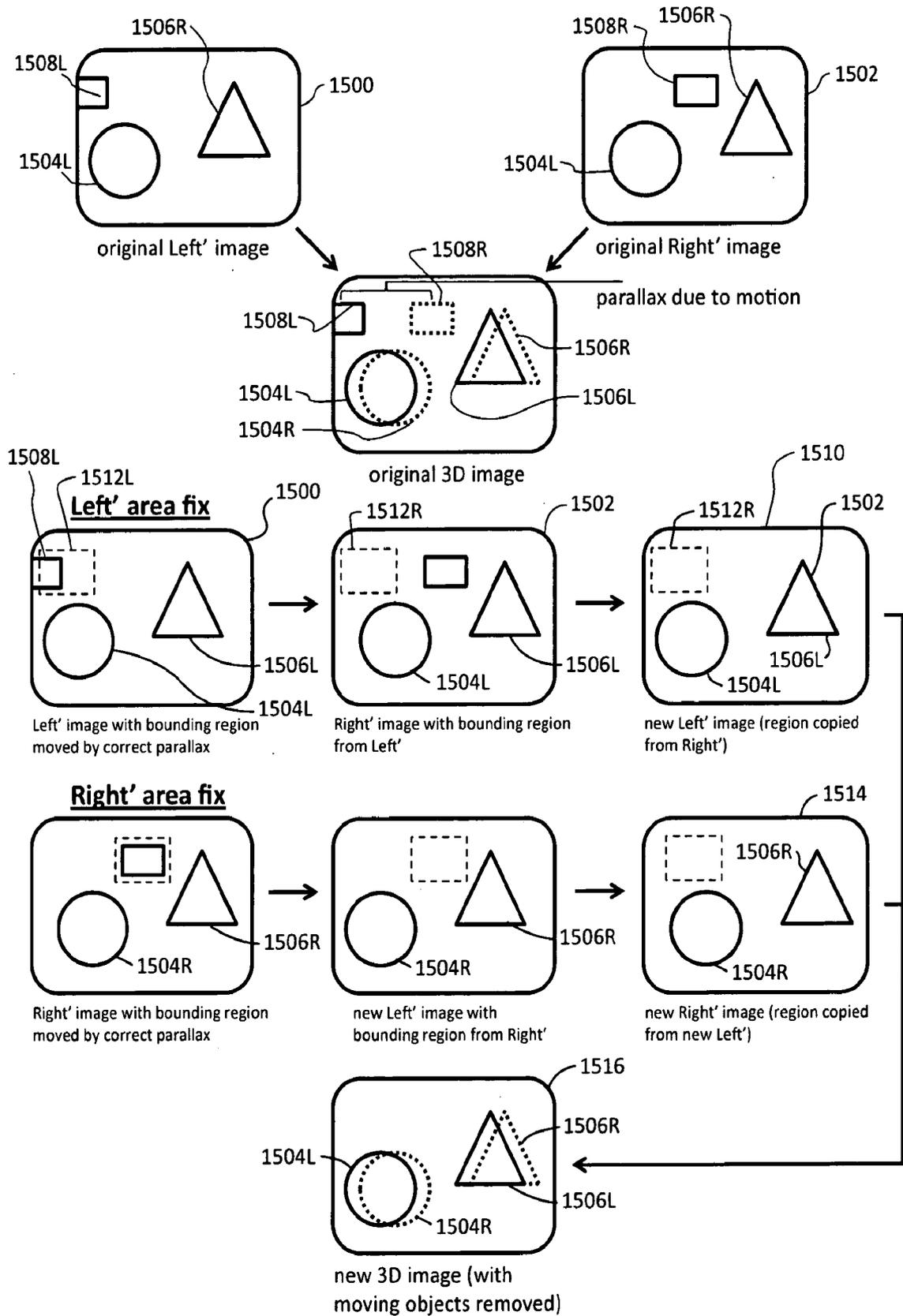


Figure 15

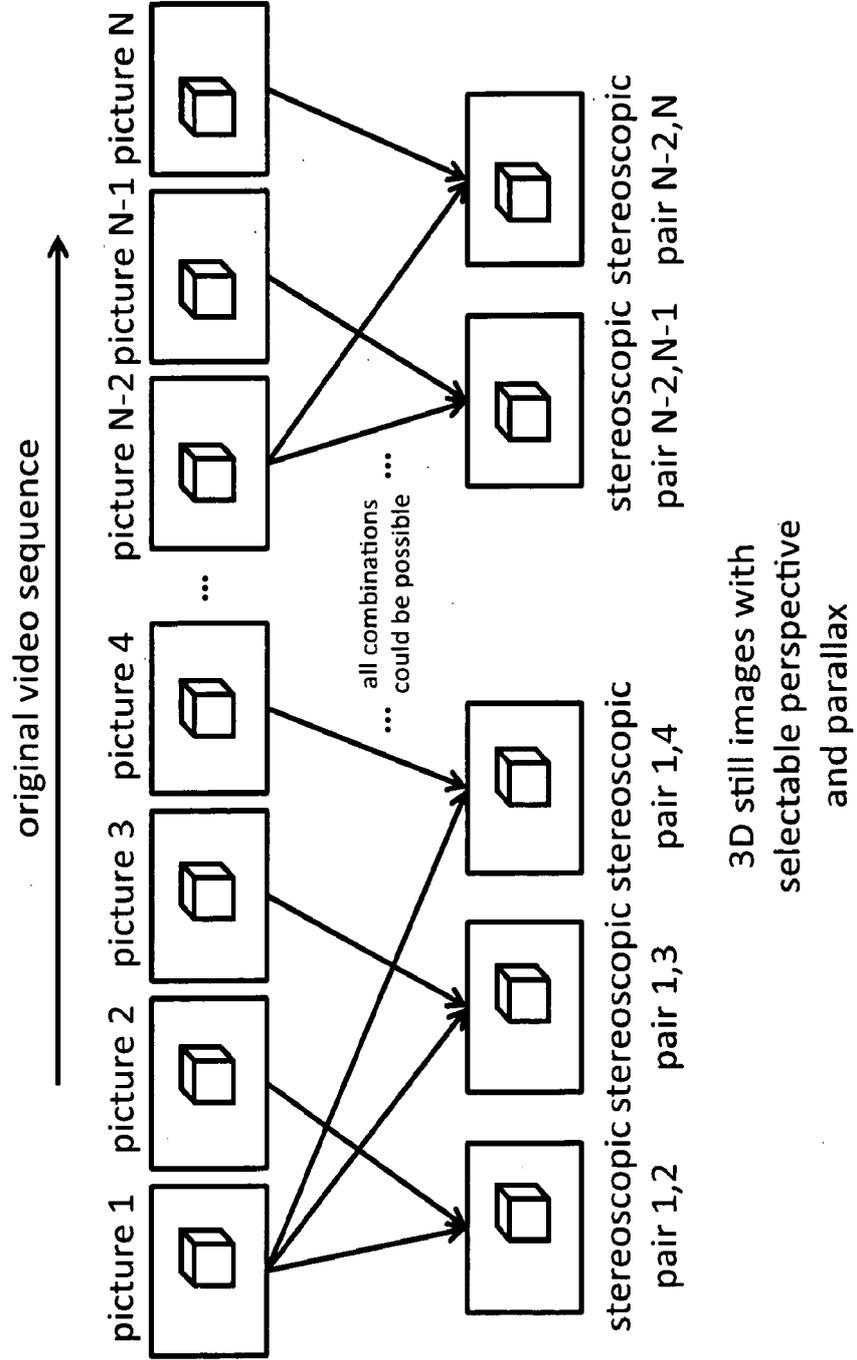


Figure 16

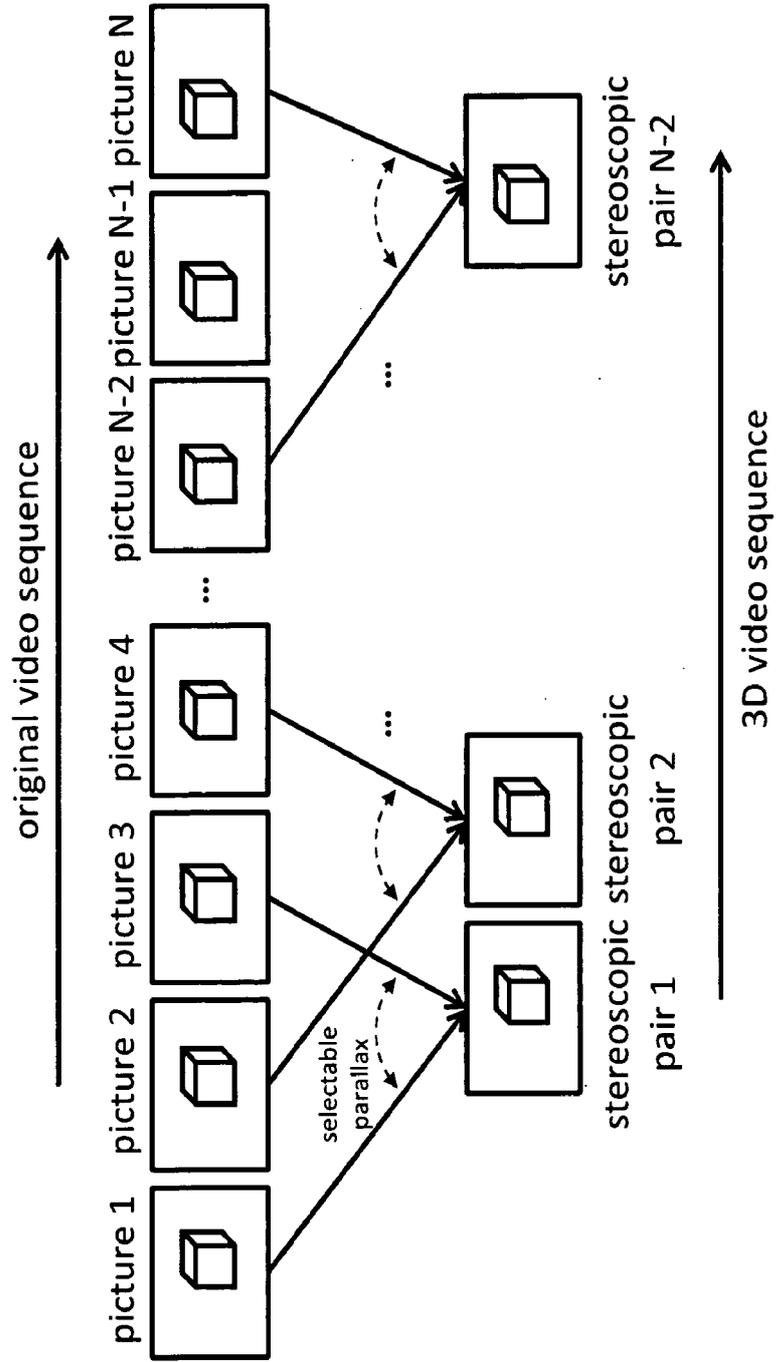


Figure 17

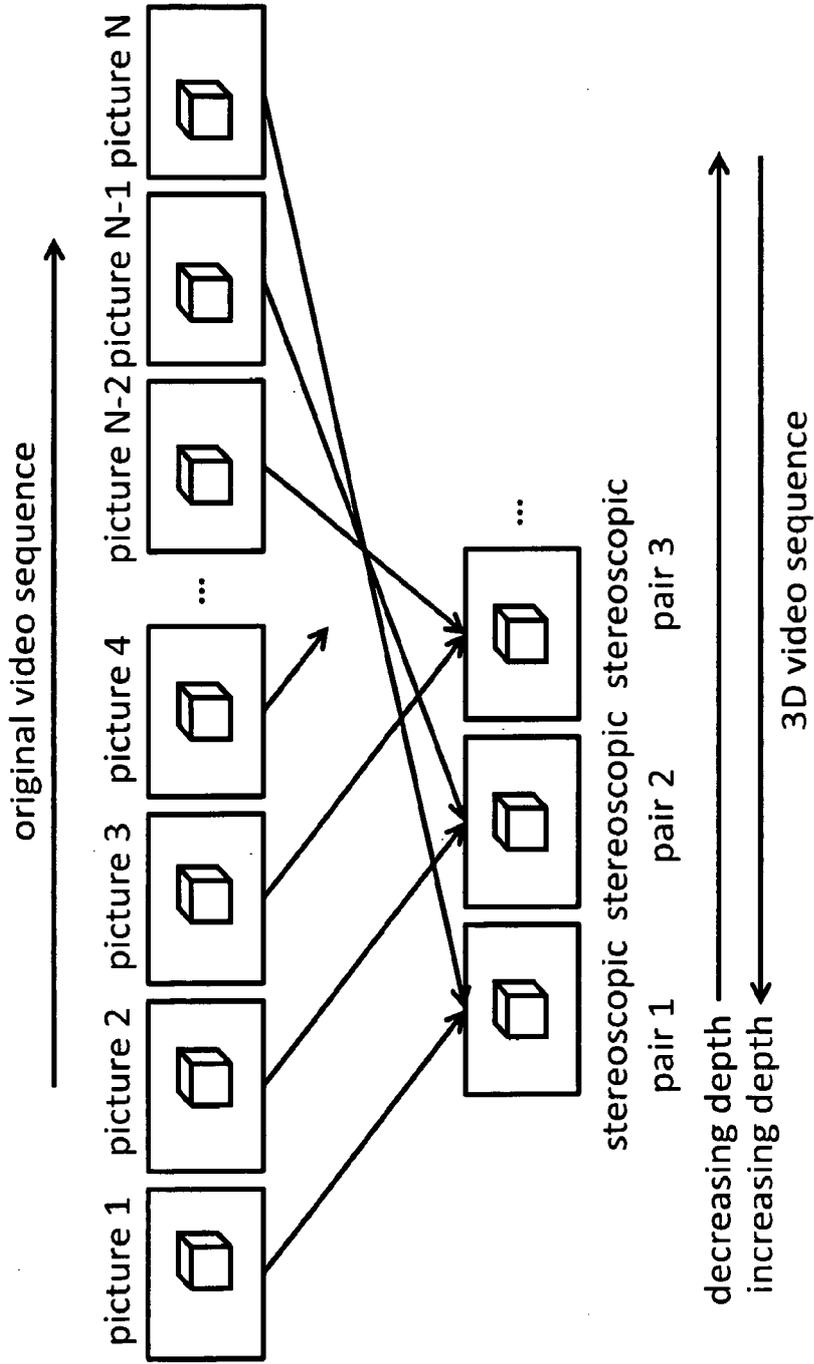


Figure 18

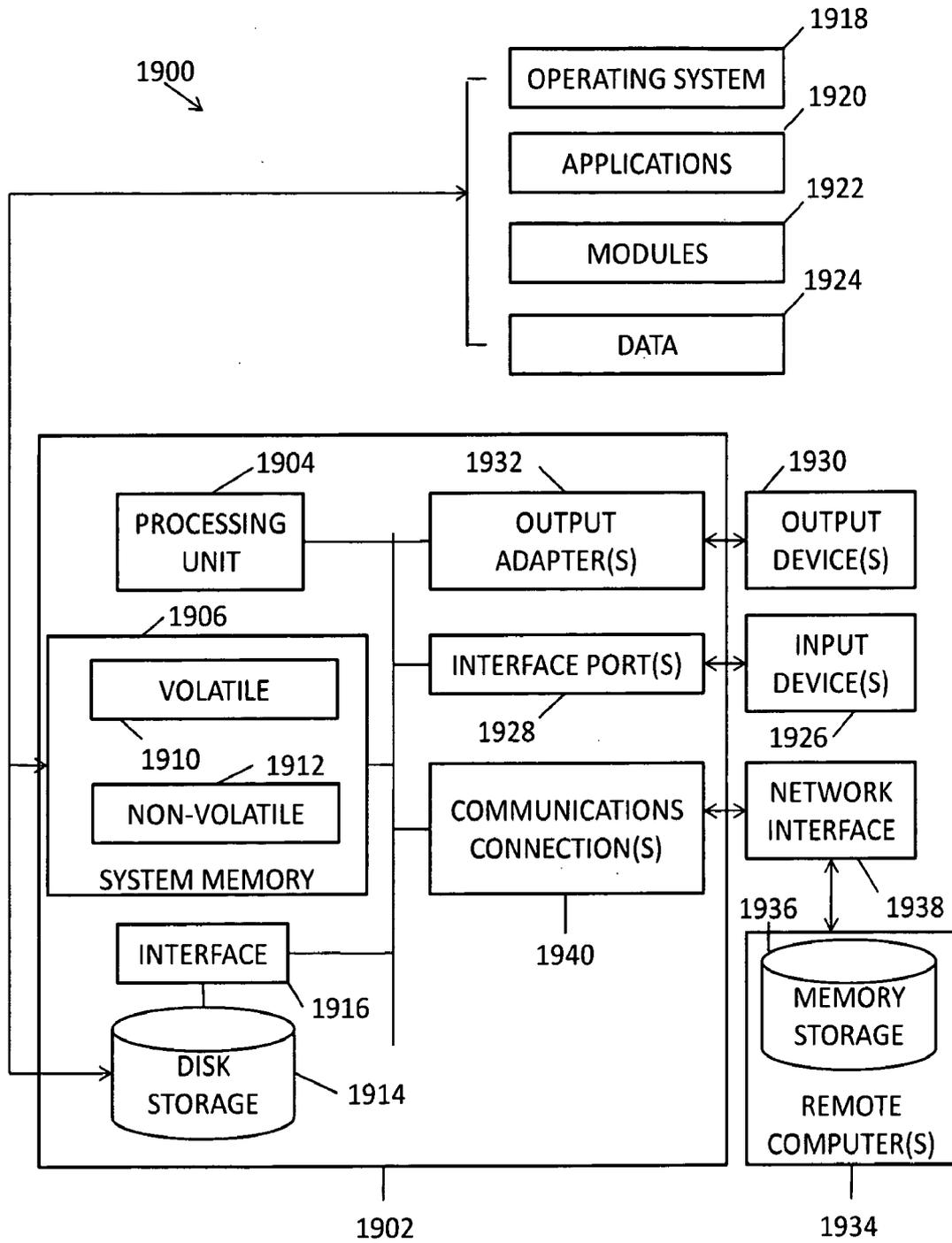


Figure 19

# INTERNATIONAL SEARCH REPORT

International application No <b>PCT/US2010/043022</b>
--

**A. CLASSIFICATION OF SUBJECT MATTER**

INV . H04N13/00  
ADD .

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

**B. FIELDS SEARCHED**

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

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

Electronic data base consulted during the international search (name of data base and, where practical, 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
<b>X</b>	US 2002/191841 A1 (HARMAN PHILIP VICTOR [AU]) 19 December 2002 (2002-12-19) paragraph [0087] - paragraph [0183] -----	1, 3, 6, 9

Further documents are listed in the continuation of Box C

See patent family annex

\* Special categories of cited documents

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document 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

9 November 2010

Date of mailing of the international search report

16/11/2010

Name and mailing address of the ISA/

European Patent Office, P B 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel (+31-70) 340-2040,  
Fax (+31-70) 340-3016

Authorized officer

Penchev, Petyo

FURTHER INFORMATION CONTINUED FROM POT/ISA/ 210

Continuation of Box II.2

Claims Nos.: 2, 4, 5, 7, 8, 10-31(completely); 1, 3, 6, 9(partially)

The initial phase of the search revealed a very large number of documents relevant to the issue of novelty. So many documents were retrieved that it is impossible to determine which parts of the claim 1 may be said to define subject-matter for which protection might legitimately be sought (Article 6 PCT). The present application contains 11 claims dependent on claim 1, and they are drafted in such a way that the claims as a whole are not in compliance with the provisions of clarity and conciseness of Article 6 PCT, as they create a smoke screen in front of the skilled reader when assessing what should be the subject-matter to search. The non-compliance with the substantive provisions is to such an extent, that the search was performed taking into consideration the non-compliance in determining the extent of the search (PCT Guidelines 9.19). The extent of the search was consequently limited to claims 1, 3, 6, 9 and paragraph [0072] in combination, which appears to comprise a reasonable definition of what is understood to be the invention for which protection is sought.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.2), should the problems which led to the Article 17(2) declaration be overcome.

INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2010/043022

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons

- 1  Claims Nos \_\_\_\_\_ because they relate to subject matter not required to be searched by this Authority, namely \_\_\_\_\_
  
- 2  Claims Nos 2, 4, 5, 7, 8, 10-31 (completely) ; 1, 3, 6, 9 (partially) because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically see FURTHER INFORMATION sheet PCT/ISA/210
  
- 3  Claims Nos \_\_\_\_\_ because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows

- 1  All required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
  
- 2  As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees
  
- 3  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos \_\_\_\_\_
  
- 4  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims, it is covered by claims Nos \_\_\_\_\_

**Remark on Protest**

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation
- No protest accompanied the payment of additional search fees

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2010/043022

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2002191841	A1	NONE	