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

(54) Title: DETERMINING DEPTH FROM MULTIPLE VIEWS OF A SCENE THAT INCLUDE ALIASING USING HYPO-THESIZED FUSION
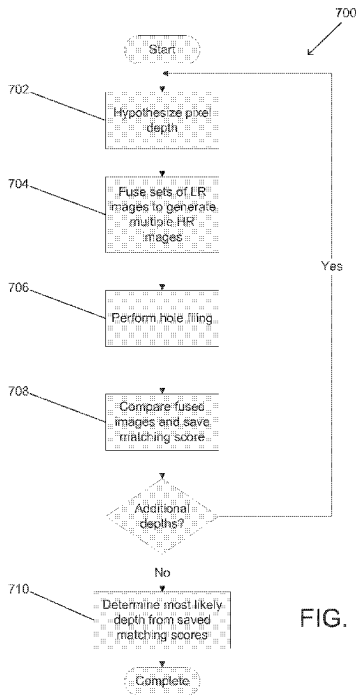
(57) Abstract: Array cameras in accordance with embodiments of the invention perform super resolution processing using images of a scene that contain aliasing. In several embodiments, the depth of pixels is determined by fusing portions of a higher resolution image at a number of hypothesized depths and determining the depth at which the portion of the higher resolution image best matches the scene captured in the lower resolution images used to fuse the higher resolution image.

FIG. 7

WO 2013/043761 A1

DETERMINING DEPTH FROM MULTIPLE VIEWS OF A SCENE THAT INCLUDE ALIASING USING HYPOTHESIZED FUSION

FIELD OF THE INVENTION

[0001]   The present invention relates to super resolution processing and more specifically to super resolution processes that determine pixel depth using multiple images of a scene captured from different viewpoints, where the captured images of the scene include aliasing.

BACKGROUND

[0002]   In a typical imaging device, light enters through an opening (aperture) at one end of the imaging device and is directed to an image sensor by one or more optical elements such as lenses. The image sensor includes pixels that generate signals upon receiving light via the optical element. Commonly used image sensors include charge-coupled device image sensors (CCDs) and complementary metal-oxide semiconductor (CMOS) sensors.

[0003]   Conventional digital cameras typically achieve color separation by performing color separation in the optical path and using a separate image sensor for the wavelengths of light corresponding to each of the primary colors (i.e. RGB), using an image sensor with color separation and multiple signal collection capability within each pixel, or by applying filters over a single sensor so that individual pixels detect wavelengths of light corresponding to one of the primary colors. Use of filters is particularly common in cameras that have a small form factor, such as cameras incorporated in mobile phone handsets and other consumer electronics devices including but not limited to, laptop computers and televisions. A common filter that is formed on image sensors is the Bayer filter, the pattern of which includes 50% green filters, 25% red filters, and 25% blue filters. The output of an image sensor to which a Bayer filter is applied can be reconstructed as a color image using interpolation techniques.

[0004]   Image sensors are subject to various performance constraints including, among others, dynamic range, signal to noise (SNR) ratio and low light sensitivity. The

dynamic range is defined as the ratio of the maximum possible signal that can be captured by a pixel to the total noise signal. The SNR of a captured image is, to a great extent, a measure of image quality. In general, as more light is captured by the pixel, the higher the SNR. The light sensitivity of an image sensor is typically determined by the intensity of light incident upon the sensor pixels. At low light levels, each pixel's light gathering capability is constrained by the low signal levels incident upon each pixel.

[0005]   A challenge associated with increasing the number of pixels in an image sensor is that the lens system is dimensioned to span the image sensor. The problem is most acute with mobile cameras, such as those used in mobile phones and consumer electronics devices, where the form factor of the lens system can significantly impact the overall form factor of the mobile device.

[0006]   In response to the constraints placed upon a traditional digital camera based upon the camera *obscura*, a new class of cameras that can be referred to as array cameras have been proposed. Array cameras are characterized in that they include multiple arrays of pixels, each having a separate lens system. Examples of 2, 3 and 4 array cameras in which each array of pixels captures light from a different band of the visible spectrum and the captured images are combined to create a full color image are disclosed in U.S. Patent 7,199,348 to Olsen et al., the disclosure of which is incorporated by reference herein in its entirety. U.S. Patent 7,262,799 to Suda, the disclosure of which is incorporated herein by reference in its entirety, discloses a 2 x 2 array camera including one sensor used to sense a red (R) image signal, one sensor used to sense a blue (B) image signal and, two sensors used to sense green (G) image signals.

SUMMARY OF THE INVENTION

[0007]   Array cameras in accordance with embodiments of the invention perform super resolution processing using images of a scene that contain aliasing. In several embodiments, the depth of pixels is determined by fusing portions of a higher resolution image at a number of hypothesized depths and determining the depth at which the portion of the higher resolution image best matches the scene captured in the lower

resolution images used to fuse the higher resolution image. One embodiment of the method of the invention includes fusing portions of the set of low resolution images to form a portion of a higher resolution image at each of a plurality of hypothesized depths, where the resolution of the portion of the higher resolution image is higher than the resolutions of the portions of the set of low resolution images used to fuse the portion of the higher resolution image, comparing the portion of the fused higher resolution image obtained at each hypothesized depth to the scene captured in the set of low resolution images, and selecting the hypothesized depth at which the portion of the fused higher resolution image is most similar to the scene captured in the set of low resolution images as the depth of at least one point in the scene captured by the set of low resolution images.

[0008] In a further embodiment of the method of the invention, comparing the portion of the fused higher resolution image to the scene captured in the set of low resolution images includes generating a set of forward mapped low resolution image portions by forward mapping the portion of the fused higher resolution image using a mapping based upon the characteristics of the cameras utilized to capture the set of low resolution images, and comparing the forward mapped low resolution image portions with corresponding portions of corresponding images in the set of low resolution images.

[0009] In another embodiment of the method of the invention, comparing the portion of the fused higher resolution image to the scene captured in the set of low resolution images includes determining the similarity of pixels in pixel stacks in the portion of the fused higher resolution image.

[0010] In a still further embodiment of the method of the invention, comparing the portion of the fused higher resolution image at a specific hypothesized depth to the scene captured in the set of low resolution images includes comparing the portion of the fused higher resolution image at the specific hypothesized depth to a portion of at least a second fused higher resolution image formed by fusing a second set of low resolution images at the specific hypothesized depth.

[0011] In still another embodiment of the method of the invention, at least one low resolution image is common to said set of low resolution images and said second set of low resolution images.

[0012] One embodiment of the invention includes a machine readable medium containing processor instructions, where execution of the instructions by a processor causes the processor to perform a process including: fusing portions of the set of low resolution images to form a portion of a higher resolution image at each of a plurality of hypothesized depths, where the resolution of the portion of the higher resolution image is higher than the resolutions of the portions of the set of low resolution images used to fuse the portion of the higher resolution image; comparing the portion of the fused higher resolution image obtained at each hypothesized depth to the scene captured in the set of low resolution images; and selecting the hypothesized depth at which the portion of the fused higher resolution image is most similar to the scene captured in the set of low resolution images as the depth of at least one point in the scene captured by the set of low resolution images.

[0013] A yet further embodiment of the method of the invention includes fusing portions of a first subset of a set of low resolution images to form a portion of a first higher resolution image at each of a plurality of hypothesized depths, where the resolution of the portion of the first higher resolution image is higher than the resolutions of the portions of the first subset of the set of low resolution images used to fuse the portion of the first higher resolution image, fusing portions of a second subset of the set of low resolution images to form a portion of a second higher resolution image at each of the plurality of hypothesized depths, where the resolution of the portion of the second higher resolution image is higher than the resolutions of the portions of the second subset of the set of low resolution images used to fuse the portion of the second higher resolution image, comparing at least the portions of the first and second higher resolution images fused at each of the plurality of hypothesized depths, and selecting the hypothesized depth at which the portions of the compared higher resolution images are most similar as the depth of at least one point in the scene imaged by pixels within portions of the first and second subsets of the set of low resolution images.

[0014]   In yet another embodiment of the method of the invention, at least one low resolution image in the set of low resolution images is common to both the first and second subsets of the set of low resolution images.

[0015]   In a further embodiment again of the method of the invention, the viewpoint of one of the low resolution images in the set of low resolution images is selected as the reference viewpoint used to fuse the portions of the first and second high resolution images at each of the plurality of hypothesized depths, and the low resolution image selected as the reference viewpoint is common to both the first and second subsets of the set of low resolution images.

[0016]   In another embodiment again of the method of the invention, low resolution images captured from viewpoints above, below, to the left, and to the right of the reference viewpoint are common to both the first and second subsets of the set of low resolution images.

[0017]   In a further additional embodiment of the method of the invention, the viewpoint of one of the low resolution images in the set of low resolution images is selected as the reference viewpoint used to fuse the portions of the first and second high resolution images at each of the plurality of hypothesized depths.

[0018]   In another additional embodiment of the method of the invention, fusing portions of a subset of the set of low resolution images to form a portion of a higher resolution image at a hypothesized depth includes: identifying pixels within the subset of the set of low resolution images based upon the hypothesized depth and the viewpoints of the low resolution images; fusing the identified pixels onto a higher resolution grid generated from a chosen reference viewpoint using known calibration information; and performing hole filing to fill holes in locations in the higher resolution grid.

[0019]   In a still yet further embodiment of the method of the invention, comparing at least the portions of the first and second higher resolution images fused at each of the plurality of hypothesized depths comprises comparing the portions of the first and second higher resolution images for matching error at each of the plurality of hypothesized depths.

[0020] In still yet another embodiment of the method of the invention, matching error is determined using at least one selected from the group of the L_1-norm and the L_2-norm of the difference of the portions of the first and second higher resolution image.

[0021] A still further embodiment again of the method of the invention also includes fusing portions of a third subset of the set of low resolution images to form a portion of a third higher resolution image at each of a plurality of hypothesized depths, where the resolution of the portion of the third higher resolution image is higher than the resolutions of the portions of the third subset of the set of low resolution images used to fuse the portion of the first higher resolution image. In addition, comparing at least the portions of the first and second higher resolution images fused at each of the plurality of hypothesized depths further comprises comparing the portions of the first, second, and third higher resolution images fused at each of the plurality of hypothesized depths.

[0022] Another embodiment includes an array camera module, including: an imager array, including a plurality of focal planes, where each focal plane comprises a two dimensional arrangement of pixels having at least two pixels in each dimension and each focal plane is contained within a region of the imager array that does not contain pixels from another focal plane, control circuitry configured to control the capture of image information by the pixels within the focal planes, and sampling circuitry configured to convert pixel outputs into digital pixel data, interface circuitry configured to transmit digital pixel data; and an optic array of lens stacks, where an image including aliasing is formed on each focal plane by a separate lens stack in the optic array of lens stacks. In addition, the array camera also includes a processor configured to receive digital pixel data from the array camera module via the interface circuitry, and memory containing an image processing pipeline application and a controller application. Furthermore, the processor is configured via the controller application to read digital pixel data from the imager array, and the image processing pipeline application configures the processor to: obtain a set of low resolution images of a scene that include aliasing by reading digital pixel data from the imager array; and synthesize a higher resolution image of the scene from a reference viewpoint using the set of low resolution images.

[0023] In a further embodiment the image processing pipeline application configures the processor to determine a depth of at least one pixel in the synthesized higher resolution image by: fusing portions of the set of low resolution images to form a portion of a higher resolution image at each of a plurality of hypothesized depths, where the resolution of the portion of the higher resolution image is higher than the resolutions of the portions of the set of low resolution images used to fuse the portion of the higher resolution image; comparing the portion of the fused higher resolution image obtained at each hypothesized depth to the scene captured in the set of low resolution images; and selecting the hypothesized depth at which the portion of the fused higher resolution image is most similar to the scene captured in the set of low resolution images as the depth of at least one point in the scene captured by the set of low resolution images.

[0024] In a still yet further embodiment, the image processing pipeline application configures the processor to compare the portion of the fused higher resolution image to the scene captured in the set of low resolution images by: generating a set of forward mapped low resolution image portions by forward mapping the portion of the fused higher resolution image using a mapping based upon the characteristics of the array camera module; and comparing the forward mapped low resolution image portions with corresponding portions of corresponding images in the set of low resolution images.

[0025] In still another embodiment, the image processing pipeline application configures the processor to compare the portion of the fused higher resolution image to the scene captured in the set of low resolution images by determining the similarity of pixels in pixel stacks in the portion of the fused higher resolution image.

[0026] In a yet further embodiment, the image processing pipeline application configures the processor to determine a depth of at least one pixel in the synthesized higher resolution image by: comparing portions of fused higher resolution images formed by fusing at least two subsets of the set of low resolution images at a plurality of hypothesized depths; and selecting the depth of at least one pixel in the synthesized higher resolution image based upon the hypothesized depth at which the compared portions of fused higher resolution images are most similar.

[0027]   In yet another embodiment, at least one low resolution image in the set of low resolution images is common to the at least two subsets of the set of low resolution images.

[0028]   In a further embodiment again, the at least two subsets of the set of low resolution images include a common low resolution image having a viewpoint that is the reference viewpoint.

[0029]   In another embodiment again, the at least two subsets of the set of low resolution images include common low resolution images captured from viewpoints above, below, to the left, and to the right of the reference viewpoint.

[0030]   In a further additional embodiment, the reference viewpoint is the viewpoint of one of the low resolution images in the set of low resolution images.

[0031]   In another additional embodiment, the image processing pipeline application configures the processor to fuse at least two subsets of the set of low resolution images at a hypothesized depth by: identifying pixels within the subset of the set of low resolution images based upon the hypothesized depth and the viewpoints of the low resolution images; fusing the identified pixels onto a higher resolution grid generated from a chosen reference viewpoint using known calibration information; and performing hole filing to fill holes in locations in the higher resolution grid.

[0032]   In a still yet further embodiment, the image processing pipeline application configures the processor to compare portions of fused higher resolution images by comparing matching error of the portions of the fused higher resolution images.

[0033]   In still yet another embodiment, matching error is determined using at least one selected from the group of the L_1-norm and the L_2-norm of the difference of the portions of the first and second higher resolution image.

[0034]   In a still further embodiment again, the pixels in the plurality of focal planes in the imager array comprise a pixel stack including a microlens and an active area, where light incident on the surface of the microlens is focused onto the active area by the microlens and the active area samples the incident light to capture image information, and the pixel stack defines a pixel area and includes a pixel aperture, where the size of the pixel apertures is smaller than the pixel area.

[0035]   In still another embodiment again, the pixel aperture is formed by a microlens that is smaller than the pixel area.

[0036]   In a still further additional embodiment, gaps exist between adjacent microlenses in the pixel stacks of adjacent pixels in a focal plane.

[0037]   In still another additional embodiment, light is prevented from entering the pixel stacks through the gaps between the microlenses by a light blocking material.

[0038]   In a yet further embodiment again, the pixel stack includes a color filter.

[0039]   In yet another embodiment again, the color filters in the pixel stacks of the two dimensional arrangement of pixels within a focal plane are the same.

[0040]   In a yet further additional embodiment, the color filters in the pixel stacks of the two dimensional arrangement of pixels within at least one focal plane form a Bayer filter pattern.

[0041]   In yet another additional embodiment, at least one of the plurality of lens stacks includes a color filter and the pixel stacks of the two dimensional arrangement of pixels within the focal plane on which said at least one of the plurality of lens stacks forms an image do not include color filters.

[0042]   In a still yet further embodiment again, the pixel aperture is formed using at least one light blocking material.

[0043]   In still yet another embodiment again, the optic array of lens stacks is constructed using wafer level optics.

[0044]   In a still yet further additional embodiment, the plurality of lens stacks include polymer optical components.

[0045]   In still yet another additional embodiment, the plurality of lens stacks include glass optical components.

BRIEF DESCRIPTION OF THE DRAWINGS

[0046]   FIG. 1 is a block diagram of an array camera in accordance with an embodiment of the invention.

[0047]   FIG. 2 is a conceptual illustration of an array camera module formed from an optic array of lens stacks and an imager array in accordance with an embodiment of the invention.

[0048]   FIG. 3 is a conceptual illustration of focal planes on an imager array in accordance with an embodiment of the invention.

[0049]   FIG. 4A is a cross-sectional view of a conventional gapless microlens pixel stack that is typical of the pixel stacks used in many conventional cameras.

[0050]   FIG. 4B is a cross-sectional view of a pixel stack including a pincushion microlens that can increase the aliasing present in a captured image relative to the gapless microlens illustrated in FIG. 4A.

[0051]   FIG. 5 illustrates an image processing pipeline in accordance with an embodiment of the invention.

[0052]   FIG. 6 is a flow chart illustrating a process for performing hypothesized fusion using forward mappings of HR images fused at different hypothesized depths in accordance with embodiments of the invention.

[0053]   FIG. 6A is a flow chart illustrating a process for performing hypothesized fusion by looking at the similarity of pixel stacks in portions of fused higher resolution images at different hypothesized depths in accordance with embodiments of the invention.

[0054]   FIG. 7 is a flow chart illustrating a process for performing hypothesized fusion in accordance with embodiments of the invention.

[0055]   FIGS. 8A illustrates an array camera module including 6 Blue cameras, 13 Green Cameras, and 6 Red cameras.

[0056]   FIGS. 8B and 8C illustrates two sets of Green cameras in the array camera module illustrated in FIG. 8A that can be utilized to perform hypothesized fusion in accordance with embodiments of the invention.

[0057]   FIG. 9A is an image generated by performing super resolution processing on images of a scene captured using an imager array having gapless microlenses.

[0058]   FIG. 9B is an image generated by performing super resolution processing on images of a scene in which aliasing has been increased in accordance with embodiments of the invention.

DETAILED DESCRIPTION

[0059]   Turning now to the drawings, systems and methods for controlling the amount of aliasing in images captured by an array camera and for synthesizing higher resolution images from the captured images using super resolution (SR) processing in accordance with embodiments of the invention are illustrated. Images exhibit aliasing when they are sampled at too low a sampling frequency, resulting in visible steps on diagonal lines or edges (also referred to as "jaggies") and artificial low frequency patterns (often referred to as Moiré). These artifacts are a product of the incorrect sampling of higher frequencies, which results in the higher frequencies folding back (being aliased) into lower frequencies. Aliasing is generally undesirable, however, array cameras in accordance with embodiments of the invention can utilize the high frequency information folded into the lower frequencies during SR processing. In a SR process, low resolution (LR) images that include sampling diversity (i.e. represent sub-pixel offset shifted views of a scene) are used to synthesize one or more higher resolution (HR) images. Each LR image samples a slightly different part of the scene and the SR process utilizes the sampling diversity to synthesize a HR image by fusing the multiple low-resolution images together on a higher resolution grid. Various SR processes are discussed in detail in U.S. Patent Application Serial No. 12/967,807, entitled "Systems and Methods for Synthesizing High Resolution Images Using Super-Resolution Processes", to Lelescu et al., the disclosure of which is incorporated by reference herein in its entirety. Due to the fact that each LR image includes a sub-pixel shifted view of the scene, aliasing present in each of the LR images is slightly different. Therefore, the aliasing in each of the images provides useful information about high frequency image content that can be exploited by the SR process to increase the overall resolution of the synthesized image.

[0060]   In a number of embodiments, increasing the amount of aliasing in the LR images captured by an array camera can increase the resolution achieved through SR processing. The amount of aliasing in the captured images can be controlled in any of a variety of ways including (but not limited to) using pixel apertures, reducing pixel pitch, and/or increasing the optical resolution of the optical channels used to form images on the pixels of a focal plane. Using pixel apertures to increase aliasing can improve the resolution increase achieved through SR processing without fundamentally changing the pixel size or design, or array configuration of an imager array in any way. Therefore, using pixel apertures to increase aliasing can be a very cost effective architectural enhancement. A pixel aperture can be created in any of a variety of different ways including by using a microlens that is smaller than the pixel pitch of a focal plane and/or by using light blocking materials.

[0061]   Increasing aliasing in captured LR images can complicate parallax detection and correction when performing SR processing. Aliasing is a result of insufficient spatial sampling frequency in each camera, and can manifest itself differently in the images captured by the different cameras. In a number of embodiments, pixel correspondence in the presence of aliasing is determined using an approach that can be referred to as "hypothesized fusion". Since aliasing effects are varied in the different images, at incorrect depths, a fused image is likely to be considerably different from the scene. At the correct depth, high resolution information obtained from the aliasing in the LR images increases the similarity between the synthesized HR image and the scene. Accordingly, hypothesized fusion processes in accordance with embodiments of the invention fuse HR images or portions of HR images using a set of LR images at a number of different hypothesized depths. The highest similarity between a fused HR image or HR image portion and a scene captured in a set of LR images is likely to be observed when correct depth hypotheses are utilized. The similarity between a fused HR image and a scene captured in a set of LR images can be determined in any of a variety of different ways. In several embodiments, similarity is determined by using a forward mapping to compare forward mappings of the fused HR image at a hypothesized depth to the captured LR images. In many embodiments, the similarity of

pixels in pixel stacks of a portion of a fused HR image are used to indicate the likely similarity of the portion of the fused HR image to the scene captured by the LR images. In a number of embodiments, multiple fused HR images or HR image portions are generated using different subsets of the captured LR images at different hypothesized depths and the multiple fused HR images or HR image portions are compared to determine the hypothesized depth at which the fused HR images or HR image portions are best matched. In several embodiments, the sets of focal planes used to fuse the HR images utilized during hypothesized fusion include focal planes that are common to two or more of the sets. In a number of embodiments, the viewpoint of one of the focal planes is used as the reference viewpoint for synthesizing an HR image and the reference focal plane is common to the sets of focal planes used during hypothesized fusion. In other embodiments, disjoint sets of focal planes are utilized.

[0062]   By analyzing the similarity of a fused HR image or image portion to the scene captured in a set of LR images at different hypothesized depths, depth information can be obtained that can be used to perform parallax correction and complete the SR processing of the LR images. Although much of the discussion that follows refers to determining depths based upon portions of fused HR images, it should be appreciated that analysis using fused HR images is computationally efficient and that additional processing can be utilized to refine the fused HR images during analysis up to an including synthesizing HR images. Accordingly, references to fused HR images in the discussion of hypothesized fusion should be understood as encompassing images obtained by simply performing raw fusion, which places captured image samples onto a higher resolution grid (possibly resulting in overlaps and missing sample positions), and encompassing images obtained by performing additional processing beyond the raw fusion.   The distinctions between obtaining an HR image through raw fusion and synthesizing an HR image using super resolution processing are explored more completely in U.S. Patent Application Serial No. 12/967,807 incorporated by reference above.

[0063]   Array cameras that control the amount of aliasing introduced into captured low resolution images using a variety of techniques including (but not limited) to using

pixel apertures and SR processes that utilize hypothesized fusion to determine pixel correspondence in the presence of aliasing in accordance with embodiments of the invention are discussed further below.

ARRAY CAMERA ARCHITECTURE

[0064] An array camera architecture that can be used in a variety of array camera configurations in accordance with embodiments of the invention is illustrated in FIG. 1. The array camera 100 includes an array camera module 110, which is connected to an image processing pipeline module 120 and to a controller 130. In many embodiments, the image processing pipeline module 120 and controller 130 are implemented using software applications and/or firmware executing on a microprocessor. In other embodiments, the modules can be implemented using application specific circuitry.

[0065] The array camera module 110 includes two or more cameras, each of which receives light using a separate optical channel. The array camera module can also include other circuitry to control imaging parameters and sensors to sense physical parameters. The control circuitry can control imaging parameters such as exposure times, gain, and black level offset. In several embodiments, the circuitry for controlling imaging parameters may trigger each camera independently or in a synchronized manner. The array camera module can include a variety of other sensors, including but not limited to, dark pixels to estimate dark current at the operating temperature. Array camera modules that can be utilized in array cameras in accordance with embodiments of the invention are disclosed in U.S. Patent Application Serial No. 12/935,504 entitled "Capturing and Processing of Images using Monolithic Camera Array with Heterogeneous Imagers" to Venkataraman et al., the disclosure of which is incorporated herein by reference in its entirety.

[0066] The image processing pipeline module 120 is hardware, firmware, software, or a combination for processing the images received from the array camera module 110. In many embodiments, the image processing pipeline module 120 is implemented using an image processing pipeline application that is stored in memory and used to configure a microprocessor. The image processing pipeline module 120 processes the

multiple LR images captured by the array camera module and produces a synthesized HR image. In a number of embodiments, the image processing pipeline module 120 provides the synthesized image data via an output 122.

[0067] The controller 130 is hardware, software, firmware, or a combination thereof for controlling various operational parameters of the imager array 110. In a number of embodiments, the controller 130 is implemented using a controller application stored in memory and used to configure a microprocessor. The controller 130 receives inputs 132 from a user or other external components and sends operation signals to control the array camera module 110. The controller 130 can also send information to the image processing pipeline module 120 to assist processing of the LR images captured by the array camera module 110.

[0068] Although a specific array camera architecture is illustrated in FIG. 1, alternative architectures that enable the capturing of LR images and application of SR processes to produce one or more synthesized HR images can also be utilized in accordance with embodiments of the invention. Array camera modules and techniques for controlling the level of aliasing in the LR images captured by array cameras in accordance with embodiments of the invention are discussed below.

ARRAY CAMERA MODULES

[0069] U.S. Patent Application Serial No. 12/935,504 (incorporated by reference above) discloses a variety of array camera modules that can be utilized in array cameras. An exploded view of an array camera module formed by combining an optic array of lens stacks with a monolithic sensor that includes a corresponding array of focal planes is illustrated in FIG. 2. The array camera module 200 includes an optic array of lens stacks 210 and a sensor or imager array 230 that includes an array of focal planes 240. The optic array of lens stacks 210 includes an array of lens stacks 220. Each lens stack creates a separate optical channel that resolves an image on a corresponding focal plane 240 on the sensor. The lens stacks may be of different types. For example, the optical channels may be used to capture images at different portions of the spectrum and the lens stack in each optical channel may be specifically optimized for

the portion of the spectrum imaged by the focal plane associated with the optical channel. More specifically, an array camera module may be patterned with "π filter groups." The term "π filter groups" refers to a pattern of color filters applied to the optic array of lens stacks of a camera module and processes for patterning array cameras with π filter groups are described in U.S. Patent Application Serial No. 61/641,164, entitled "Camera Modules Patterned with π Filter Groups", Venkataraman et al. The disclosure of U.S. Patent Application Serial No. 61/641,164 is incorporated by reference herein in its entirety. Filter patterns that can be utilized in array camera modules are disclosed further in U.S. Patent Application Serial No. 12/935,504 and U.S. Provisional Patent Application Serial No. 61/641,165.

[0070]   An optic array of lens stacks may employ wafer level optics (WLO) technology. WLO is a technology that encompasses a number of processes, including, for example, molding of lens arrays on glass wafers, stacking of those wafers (including wafers having lenses replicated on either side of the substrate) with appropriate spacers, followed by packaging of the optics directly with the imager into a monolithic integrated module.

[0071]   The WLO procedure may involve, among other procedures, using a diamond-turned mold to create each plastic lens element on a glass substrate. More specifically, the process chain in WLO generally includes producing a diamond turned lens master (both on an individual and array level), then producing a negative mould for replication of that master (also called a stamp or tool), and then finally forming a polymer replica on a glass substrate, which has been structured with appropriate supporting optical elements, such as, for example, apertures (transparent openings in light blocking material layers), and filters. Although the construction of optic arrays of lens stacks using specific WLO processes is discussed above, any of a variety of techniques can be used to construct optic arrays of lens stacks, for instance those involving precision glass molding, polymer injection molding or wafer level polymer monolithic lens processes. Any of a variety of well known techniques for designing lens stacks used in conventional cameras can be utilized to increase aliasing in captured images by improving optical resolution. Accordingly, the level of aliasing present in images captured by an array

camera module in accordance with embodiments of the invention can be determined through selection of aspects of lens stacks including (but not limited to) adding lens surfaces, changing the F# of the lens stack, and selection of materials used in construction of lens stack. Imager arrays that can capture images formed by optic arrays of lens stacks in accordance with embodiments of the invention are discussed further below.

IMAGER ARRAYS

[0072]   Imager arrays can be implemented using any of a variety of configurations in which an array of focal planes is formed on one or more sensors. A variety of imager array architectures are disclosed in U.S. Patent Application Serial No. 13/106,797, entitled "Architectures for Imager Arrays and Array Cameras" to Pain et al., the disclosure of which is incorporated by reference herein in its entirety. An imager array including multiple focal planes having independent read-out control and pixel digitization, where each focal plane has dedicated peripheral circuitry, in accordance with embodiments of the invention is illustrated in FIG. 3. The imager array 300 includes a plurality of sub-arrays of pixels or focal planes 302, where each focal plane includes a two dimensional arrangement of pixels having at least two pixels in each dimension and each focal plane is contained within a region of the imager array that does not contain pixels from another focal plane. The focal planes have dedicated row control logic circuitry 304, which is controlled by a common row timing control logic circuitry 306. Although the column circuits and row decoder are shown as a single block on one side of the focal plane, the depiction as a single block is purely conceptual and each logic block can be split between the left/right and/or top/bottom of the focal plane so as to enable layout at double the pixel pitch. Laying out the control and read-out circuitry in this manner can result in a configuration where even columns are sampled in one bank of column (row) circuits and odd columns would be sampled in the other.

[0073]   In a device including M x N focal planes, the read-out control logic includes M sets of column control outputs per row of focal planes (N). Each column sampling/read-

out circuit 308 can also have dedicated sampling circuitry for converting the captured image information into digital pixel data. In many embodiments, the sampling circuitry includes an Analog Signal Processor (ASP), which includes an Analog Front End (AFE) amplifier circuit and an Analog to Digital Converter (ADC) 310. In other embodiments, any of a variety of analog circuitry can be utilized to convert captured image information into digitized pixel information. An ASP can be implemented in a number of ways, including but not limited to, as a single ASP operating at X pixel conversion per row period, where X is the number of pixels in a row of the focal plane served by the column sampling circuit (e.g. with a pipe-lined or SAR ADC), as X ASPs operating in parallel at 1 pixel conversion per row period or P ASPs operating in parallel at X/P conversions per row (see discussion below). A common read-out control circuit 312 controls the read-out of the columns in each imager.

[0074] In the illustrated embodiment, the master control logic circuitry 314 controls independent read-out of each imager. The master control logic circuitry 314 includes high-level timing control logic circuitry to control the image capture and read-out process of the individual focal plane. In a number of embodiments, the master control portion of this block can implement features including but not limited to: staggering the start points of image read-out such that each focal plane has a controlled temporal offset with respect to a global reference; controlling integration times of the pixels within specific focal planes to provide integration times specific to the spectral bandwidths being imaged; the horizontal and vertical read-out direction of each imager; the horizontal and vertical sub-sampling/binning/windowing of the pixels within each focal plane; the frame/row/pixel rate of each focal plane; and the power-down state control of each focal plane.

[0075] The master control logic circuitry 314 can also handle collection of pixel data from each of the imagers. In a number of embodiments, the master control logic circuitry packs the image data into a structured output format. Given that fewer than M x N output ports are typically used to output the image data (e.g. there are 2 output ports), the image data is time multiplexed onto these output ports. In a number of embodiments, a small amount of memory (FIFO) is used to buffer the data from the

pixels of the imagers until the next available time-slot on the output port 316 and the master control logic circuitry 314 or other circuitry in the imager array periodically inserts codes into the data stream providing information including, but not limited to, information identifying a focal plane, information identifying a row and/or column within a focal plane, and/or information identifying the relative time at which the capture or read-out process began/ended for one or more of the focal planes. Relative time information can be derived from an on-chip timer or counter, whose instantaneous value can be captured at the start/end of read-out of the pixels from each imager either at a frame rate or a line rate. Additional codes can also be added to the data output so as to indicate operating parameters such as (but not limited to) the integration time of each focal plane, and channel gain. As is discussed further below, the host controller can fully re-assemble the data stream back into the individual images captured by each focal plane. In several embodiments, the imager array includes sufficient storage to buffer at least a complete row of image data from all focal planes so as to support reordering and or retiming of the image data from all focal planes such that the data is always packaged with the same timing/ordering arrangement regardless of operating parameters such as (but not limited to) integration time and relative read-out positions. In a number of embodiments, the imager array includes sufficient storage to buffer at least a complete line of image data from all focal planes so as to support reordering and or retiming of the image data from all focal planes such that the data is packaged in a convenient manner to ease the host's reconstruction of the image data, for example retiming/reordering the image data to align the data from all focal planes to a uniform row start position for all focal planes irrespective of relative read-out position.

[0076]   Although specific imager array implementations are discussed above with reference to FIG. 3, any of a variety of imager arrays can be utilized in an array camera including (but not limited to) the imager arrays disclosed in U.S. Patent Application Serial No. 13/106,797 as appropriate to the requirements of a specific application in accordance with an embodiment of the invention. The introduction of aliasing into images captured by an array camera using microlenses and the recovery of high

resolution information using the aliasing via super resolution processing in accordance with embodiments of the invention are discussed further below.

INTRODUCING ALIASING INTO IMAGES CAPTURED BY AN ARRAY CAMERA

[0077] From sampling theory it is known that the Nyquist frequency of the image sensor is simply one half the reciprocal of the pixel pitch. Frequencies above the Nyquist frequency cannot be sampled correctly by the image sensor and result in aliasing. Sampling theorem indicates that a judicious choice of pixel pitch (i.e. sampling rate) when sampling a bandlimited function can completely avoid aliasing, but it cannot avoid aliasing when sampling inherently non-bandlimited functions. Therefore, increasing the pixel pitch of an imager can increase the aliasing in images captured by the imager. As is discussed further below, aliasing can also be introduced into a captured image without increasing pixel pitch.

[0078] In many embodiments, the optical channel of each camera in the array camera is designed so that a predetermined amount of aliasing is present in the image resolved on the focal plane of the camera by the optical channel. As noted above, the extent to which the aliasing in the resolved image is captured by the pixels of the focal plane depends on a number of factors including the pixel pitch. In a number of embodiments, the pixels in the focal plane include pixel apertures that reduce pixel blur thereby increasing the extent to which aliasing is present in the captured image. In several embodiments, pixel apertures are created using microlenses. In many embodiments, pixel apertures are created using light blocking materials. Various techniques for increasing the aliasing present in captured images through the use of pixel apertures in accordance with embodiments of the invention are discussed further below.

USING MICROLENSES TO INCREASE ALIASING

[0079] In several embodiments, the aliasing present in images captured by an imager array is increased using microlenses that act as pixel apertures to reduce pixel blur. The manner in which microlenses can be used to increase aliasing can be

appreciated with reference to FIGS. 4A and 4B. A conventional gapless microlens pixel stack that is typical of the pixel stacks used in many conventional cameras is illustrated in FIG. 4A. Although a single pixel stack is shown in FIG. 4A, one of ordinary skill in the art will appreciate that the pixels that form a focal plane each have pixel stacks, which are similar. The pixel stack 400 includes a microlens 402, which is typically 0.3 µm at its thickest region (although this thickness can vary from company to company and process to process). The microlens sits atop an oxide layer 404, which is typically 0.3 µm thick. Beneath the oxide layer 404 is a color filter 406, which is typically 1.0 µm thick. The color filter 406 is above a nitride layer 408, which is typically 0.3 µm thick. The nitride layer 408 is above a second oxide layer 210, which is 1.0 µm thick, and sits atop the silicon 412 that includes the active area 414 of the sensor (typically a photodiode). Although specific dimensions are referenced above, the dimensions of a pixel stack are largely determined by the manufacturing processes utilized and the requirements of a specific application.

[0080]   The main task of a microlens 402 is to gather the light incident on its surface and focus that light onto the small active area 414. The top oxide layer 404 separates the microlens layer from the color filter layer 406 and provides a suitable surface for effective microlens formation. The nitride passivation layer 408 and bottom oxide layer 410 provide support and isolation for the metal interconnects that are used to connect the various parts of the sensor. The active area 414 represents a small part of the pixel stack and is responsible for sampling the light incident on it. The pixel aperture (416) is determined by the spread of the microlens, which collects the light and focuses it on the active area 414. Due to the fact that the microlens spans the pixel area, the microlens 402 can be referred to as a gapless microlens.

[0081]   The blur of the light field incident on a microlens array can be reduced by reducing the spread of the microlenses used in the pixel stacks of the focal plane. Thus, altering the microlenses in this fashion can be used to control the degree of aliasing present in captured images. The microlens acts like a low pass filter on the portion of the image resolved onto the pixel area by a lens stack.

[0082] A microlens that results in increased aliasing in a captured image relative to the image that would be captured using the gapless microlens 402 illustrated in FIG. 4A is illustrated in FIG. 4B. The pixel stack 400' includes a microlens 402', which is smaller than the pixel area (i.e. the edges of microlens do not extend to the edges of the pixel area). The microlens 402' increases the aliasing present in the captured image as compared to the gapless microlens 402 shown in FIG. 2A. The microlens thus effectively acts as a pixel aperture, controlling the amount of light that is incident on the photodiode. In many array cameras, the pixels in each focal plane are typically sensitive to only one color thus the color filters in the pixel stack can be reduced or removed entirely and placed in the optic array of lens stacks. In other embodiments, the pixels in at least one of the focal planes are patterned with a pattern of color filters corresponding to a Bayer pattern or a similar pattern appropriate to the requirements of a specific application. In the illustrated embodiment, the color filter is significantly thinner (e.g. less than 0.1 μm), which reduces the overall height of the pixel stack 200'. Although a specific pixel stack is illustrated in FIG. 4B, as will be appreciated from the description below other pixel stacks that incorporate pincushion microlenses, reduce or remove color filters, include light blocking materials to create pixel apertures that are smaller than the pixel pitch and/or have decreased pixel stack height can be utilized in imager arrays in accordance with embodiments of the invention.

[0083] An advantage of decreased pixel aperture can be increasing the amount of aliasing present in captured images and, therefore, the increase in resolution that can be recovered through SR processing. Decreasing pixel apertures can come at the expense of decreased sensitivity. Although specific pixel stacks are described above, any of a variety of pixel stacks having pixel apertures that are smaller than the pitch of the pixels within a focal plane can be utilized as appropriate to the requirements of specific application in accordance with embodiments of the invention. As is discussed further below, decreasing the size of the pixel apertures within a focal plane can increase aliasing, which can be utilized during SR processing to recover information concerning high frequency components of the image.

FACTORS INFLUENCING PIXEL STACK DESIGN

[0084]   Reducing the size of microlenses within the pixel stacks of the pixels in a focal plane can increase aliasing in images captured by the focal plane. However, reducing the size of the microlenses can also impact pixel sensitivity and crosstalk in the sensor. Any reduction in the size of the microlens relative to pixel pitch directly reduces the amount of light that is gathered (i.e. the sensitivity of the pixel). In many embodiments, each sensor in the focal plane is sensitive to only one color (e.g.: red, green, or blue). Therefore, the color filters on the pixel stack are all the same color. The absence of the ubiquitous Bayer filter implies that the pixels in a focal plane are not subject to inter-color crosstalk. This allows the use of color filters that are thinner than those in sensors with the Bayer filter, leading to correspondingly higher transmissivities. Therefore, the imagers in an array camera can have increased sensitivity compared to the sensors of a conventional camera outfitted with a Bayer color filter, which can offset the reduction in sensitivity associated with the pincushion microlens (i.e. a microlens that is smaller than the pixel pitch). In many embodiments, however, at least one focal plane in an imager array utilizes a Bayer color filter and the pixel stacks within the focal plane are configured accordingly.

[0085]   When light entering the microlens/filter of one pixel stack is directed toward a neighboring pixel, the light that is passed from one pixel stack to another is referred to as crosstalk or more specifically as optical crosstalk. Finite difference time domain simulations have shown that the amount of crosstalk in a pixel stack is directly proportional to the height of the pixel stack. Removing or reducing the thickness of the color filter in the pixel stack reduces the overall height of the pixel stack and reduces optical crosstalk. In many embodiments rearranging the color filters from the pixel stack to the optic array of lens stacks can mitigate any increase in crosstalk associated with use of a microlens in the pixel stack.

[0086]   When gaps are introduced between microlenses in a focal plane, the possibility exists that stray light can enter the pixel stack through the gaps and fall on the active area of the pixel stack, increasing the crosstalk and diminishing signal to noise ratio. In several embodiments, a light blocking material such as (but not limited

to) a photoresist can be utilized to fill the gaps between the microlenses to reduce the likelihood that stray light will enter the pixel stack.

[0087]   Although specific techniques are discussed above for increasing aliasing in the LR images through use of pixel apertures, other techniques including but not limited to techniques that utilize light blocking materials to create pixel apertures can also be utilized. Processes that can be used to recover higher resolution content from aliasing in LR images in accordance with embodiments of the invention are discussed below.

IMAGE PROCESSING

[0088]   The processing of LR images to obtain an SR image in accordance with embodiments of the invention typically occurs in an array camera's image processing pipeline. In many embodiments, the image processing pipeline performs processes that register the LR images prior to performing SR processes on the LR images. In several embodiments, the image processing pipeline also performs processes that eliminate problem pixels and compensate for parallax.

[0089]   An image processing pipeline incorporating a SR module for fusing information from LR images to obtain one or more synthesized HR images in accordance with an embodiment of the invention is illustrated in FIG. 5. In the illustrated image processing pipeline 120, pixel information is read out from the focal planes in the imager array 110 and is provided to a photometric conversion module 504 for photometric normalization. The photometric conversion module can perform any of a variety of photometric image processing processes including but not limited to one or more of photometric normalization, Black Level calculation and adjustments, vignetting correction, and lateral color correction. In several embodiments, the photometric conversion module also performs temperature normalization. In the illustrated embodiment, the inputs of the photometric normalization module are photometric calibration data and the captured LR images. The photometric calibration data is typically captured during an offline calibration process. The output of the photometric conversion module 504 is a set of photometrically normalized LR images. These

photometrically normalized images are provided to a parallax detection module 508 and to a super resolution module 514.

[0090]   Prior to performing SR processing, the image processing pipeline detects parallax that becomes more apparent as objects in the scene captured by the imager array approach the imager array. In the illustrated embodiment, parallax (or disparity) detection is performed using the parallax detection module 508.   In several embodiments, the parallax detection module 508 generates an occlusion map for the occlusion zones around foreground objects. In many embodiments, the occlusion maps are binary maps created for pairs of LR images. In many embodiments, occlusion maps are generated to illustrate whether a point in the scene is visible in the field of view of a reference imager and/or whether points in the scene visible within the field of view of the reference imager are visible in the field of view of other imagers. In order to determine parallax, the parallax detection module 508 performs scene independent geometric corrections to the photometrically normalized LR images using geometric calibration data 506 obtained via an address conversion module 502. The parallax detection module can then compare the geometrically and photometrically corrected LR images to detect the presence of scene dependent geometric displacements between LR images. Information concerning these scene dependent geometric displacements can be referred to as parallax information and can be provided to the super resolution module 514 in the form of scene dependent parallax corrections and occlusion maps. Processes for performing parallax detection are discussed in U.S. Provisional Patent Application Serial No. 61/691,666 entitled "Systems and Methods for Parallax Detection and Correction in Images Captured Using Array Cameras" to Venkataraman et al., the disclosure of which is incorporated by reference herein in its entirety. Designing a camera module to increase aliasing in captured LR images can complicate the process of determining the scene dependent parallax corrections.   Processes that can be utilized to determine appropriate scene dependent parallax corrections in the presence of aliasing are discussed further below.

[0091]   Geometric calibration (or scene-independent geometric correction) data 506 can be generated using an off line calibration process or a subsequent recalibration

process.   The scene-independent correction information, along with the scene-dependent geometric correction information (parallax) and occlusion maps, form the geometric correction information for the LR images. Once the parallax information has been generated, the parallax information and the photometrically normalized LR images can be provided to a super resolution module 314 for use in the synthesis of one or more HR images 316.

[0092]   In many embodiments, the super resolution module 314 performs scene independent and scene dependent geometric corrections (i.e. geometric corrections) using the parallax information and geometric calibration data 306 obtained via the address conversion module 302. The photometrically normalized and geometrically registered LR images are then utilized in the synthesis of a HR image. The synthesized HR image 516 may then be fed to a downstream color processing module 564, which can be implemented using any standard color processing module configured to perform color correction and/or chroma level adjustment. In several embodiments, the color processing module performs operations including but not limited to one or more of white balance, color correction, gamma correction, and RGB to YUV correction.

[0093]   In a number of embodiments, image processing pipelines in accordance with embodiments of the invention include a dynamic refocus module. The dynamic refocus module enables the user to specify a focal plane within a scene for use when synthesizing a HR image. In several embodiments, the dynamic refocus module builds an estimated HR depth map for the scene. The dynamic refocus module can use the HR depth map to blur the synthesized image to make portions of the scene that do not lie on the focal plane appear out of focus. In many embodiments, the SR processing is limited to pixels lying on the focal plane and within a specified Z-range around the focal plane.

[0094]   In several embodiments, the synthesized high resolution image 516 is encoded using any of a variety of standards based or proprietary encoding processes including but not limited to encoding the image in accordance with the JPEG standard developed by the Joint Photographic Experts Group. The encoded image can then be stored in accordance with a file format appropriate to the encoding technique used

including but not limited to the JPEG Interchange Format (JIF), the JPEG File Interchange Format (JFIF), or the Exchangeable image file format (Exif).

[0095]   Processing pipelines similar to the processing pipeline illustrated in FIG. 5 and the super resolution processing performed by such image processing pipelines are described in U.S. Patent Application Serial No. 12/967,807 (the disclosure of which is incorporated by reference above). Although specific image processing pipelines are described above, super resolution processes in accordance with embodiments of the invention can be used within any of a variety of image processing pipelines that register LR images prior to super resolution processing in accordance with embodiments of the invention. The manner in which aliasing within the LR images can be utilized by super resolution processes to increase the overall resolution of the synthesized image in accordance with embodiments of the invention are discussed further below.

SUPER RESOLUTION PROCESSING

[0096]   In a SR process, the images that are captured by the cameras that have fields of view that are at a sub-pixel offset relative to each other are used to synthesize a higher resolution image. When aliasing is introduced, the sub-pixel offsets in the fields of view of each of the cameras means that the aliasing is slightly different in each captured LR image. Therefore, the aliasing in each of the LR images provides useful information about high frequency image content that is exploited by the SR process to increase the overall resolution of the synthesized HR image. However, increasing the aliasing in the LR images can complicate parallax detection and correction.

[0097]   As is discussed in U.S. Provisional Patent Application Serial No. 61/691,666, disparity between captured LR images can be determined by searching, in some robust manner, for similar pixels in pairs of images. However, such searches can be quickly confused in flat regions and in regions with repeating patterns or textures, as the pixel (or groups of pixels) under consideration in one camera can have multiple matches in another. Such spurious matches can be disambiguated to some extent by the use of scene information (prior) and/or pyramidal search techniques. However, such techniques typically fail in the presence of aliasing effects. Aliasing is a result of

insufficient spatial sampling frequency in each camera, and can manifest itself differently in the images captured by the different cameras. As a result, pixel or patch matching (i.e. matching portions of images) using pyramidal techniques can also fail. Image processing pipelines and parallax detection processes in accordance with embodiments of the invention utilize the differences in high frequency information in each of the captured LR images, to establish pixel correspondence between the captured LR images in a way that accommodates the aliasing in each captured image. In a number of embodiments, pixel correspondence in the presence of aliasing is determined using an approach that can be referred to as "hypothesized fusion". Hypothesized fusion in accordance with embodiments of the invention is discussed further below.

HYPOTHESIZED FUSION

[0098] In many embodiments, hypothesized fusion processes are utilized to determine pixel correspondences in the presence of aliasing. Based upon the pixel correspondences, super resolution processes (such as those described in U.S. Patent Application Serial No. 12/967,807) can be performed that extract higher frequency content from the aliased frequencies in the images. However, in the absence of sub-pixel registration information, this can be non-trivial. As mentioned earlier, in the presence of aliasing, it is difficult to recover the depth of each pixel to obtain pixel correspondences among the various images. To circumvent this problem, multiple HR images or HR image patches can be obtained by fusing some or all of the captured LR images at various hypothesized depths. Only at the correct depth, will the fused HR image (or part of it) represent an image of the captured scene. Therefore, the depth of a point in a scene can be determined by fusing portions of LR images that contain the point in the scene at different hypothesized depths and selecting the depth of the point in the scene as the hypothesized depth at which the fused HR image most closely matches the scene. As is discussed further below, an array camera typically does not possess a baseline or ground truth concerning the scene. Therefore, the extent to which the fused HR image corresponds to the scene can be determined in any of a

variety of ways. In several embodiments, the fused HR image is forward mapped using a process similar to that utilized using SR processing and the resulting LR forward mapped images compared to the captured LR images. In many embodiments, the variance in pixel stacks in the fused HR image is used to indicate the similarity of the fused HR image to the scene. In a number of embodiments, multiple subsets of the captured LR images are fused to create multiple fused HR images or HR image portions and the multiple fused HR images or HR image portions are compared. In other embodiments, any of variety of techniques can be used to evaluate the similarity of one or more fused HR images or fused HR image portions to the scene captured in the LR images.

[0099] A process for performing hypothesized fusion in accordance with an embodiment of the invention is illustrated in FIG. 6. Initially, the process 600 involves capturing (602) multiple images of a scene. Due to the difficulty of ascertaining the parallax depth in aliased regions the process hypothesizes (604) a depth "d". Different embodiments may involve hypothesizing the depth 'd' in aliased regions using different measures and metrics. In one embodiment, the hypothesized depth may be an ordered list based on the 'n' closest pixels with high confidence depth maps. Given the depth "d", the pixel correspondence between registered captured images becomes fixed. The pixels from the corresponding images can then be mapped to a high resolution grid in an attempt to fuse (606) or synthesize an HR image. As is discussed in U.S. Patent Application Serial No. 12/967,807, the resulting HR image is a representation of an image of the scene that is being captured. The process of capturing images using the focal planes in the imager array can be considered to involve transformation of the scene (which should correspond to the HR image of the scene) through a forward mapping process that can be modeled using appropriate transfer functions. The specific forward mapping (608) is dependent upon the construction of the array camera module. The LR image estimates produced by applying the forward mapping to the synthesized HR image can be compared (609) to the captured LR images and a matching score determined and saved. The process then determines (610) whether there are additional hypothesized depths to test. If there are, then a new hypothesized

depth is used and the process repeats. When matching scores have been obtained at all of the hypothesized depths, the hypothesized depth that yields the best matching score can be selected as the final depth estimate. In many embodiments, additional termination conditions can be utilized when comparing the forward mapped LR images to the captured LR images. For example, if a good match of the aliased region is found, then the loop can terminate and the hypothesized depth "d" determined to be a good depth and the pixel correspondence between the captured images is confirmed.

[00100]   An alternative to performing forward mapping of a fused or super-resolved HR image at different hypothesized depths to determine the extent to which the fused or super-resolved HR image corresponds with the scene captured by a set of LR images is to instead look at the characteristics of the pixels that are fused to create a portion of the HR image. In many embodiments, the number of captured LR images is sufficiently large so as to result in pixel stacks in the fused HR image. At the correct depth, the same or very similar pixels should be fused onto the same pixel locations in the higher resolution grid. Therefore, the similarity of the pixels in pixel stacks can be utilized as a measure of the similarity of a portion of a fused HR image to the scene captured in a set of LR images. In many embodiments, the extent to which a portion of a fused HR image at a specific hypothesized depth matches the scene captured in a set of LR images is determined based upon the variance of pixels within pixels stacks in the portion of the fused HR image.

[00101]   A process for determining the depth of a point in a scene captured by a set of LR images by fusing portions of HR images at different hypothesized depths and selecting a depth based upon the variance in the pixel stacks of the fused HR image portions at each hypothesized depth in accordance with an embodiment of the invention is illustrated in FIG. 6a. The process 650 includes selecting (652) a hypothesized pixel depth, fusing (654) portions of LR images to generate a portion of an HR image at the selected hypothesized pixel depth, and determining (656) a matching score based on at least the variance of pixels within pixels stacks within the fused HR image portion. The process repeats until a determination (658) is made that a matching score has been determined at each hypothesized depth of interest and then the hypothesized depth at

which the fused HR image portion best matches the scene captured by the LR images is determined based upon the hypothesized depth that yields the highest matching score. In many embodiments, additional termination conditions can be utilized when comparing the forward mapped LR images to the captured LR images. For example, if a matching score exceeding a predetermined threshold is obtained then the process can terminate and the hypothesized depth that yielded the matching score is selected as the appropriate depth.

[00102] Although a specific processes are illustrated in FIGS. 6 and 6a, any of a variety of processes can be utilized to determine pixel correspondence between multiple images of a scene that include aliasing in accordance with embodiments of the invention. In many embodiments, the process of verifying hypothesized depths can be performed by generating multiple fused images using LR images captured by different sets of focal planes within the array camera module. Processes for verifying hypothesized depths using multiple fused images in accordance with embodiments of the invention are discussed further below.

HYPOTHESIZED FUSION USING HR IMAGE COMPARISON

[00103] When performing hypothesized fusion, determining the "correctness" of the fused image would be greatly assisted by knowledge of the scene content or a ground truth aliasing-free image as reference. In the absence of either, array cameras in accordance with many embodiments of the invention partition the focal planes in the imager array into at least two sets, and form a separate fused HR image using the LR images captured by each set of focal planes. Since aliasing effects are varied in the different images, at incorrect depths, the fused HR images are likely to be considerably different. At the correct depth, high resolution information obtained from the aliasing in the LR images increases the similarity between a synthesized HR image and the scene. Accordingly, two or more HR images fused using correct depths will have a significantly higher level of similarity relative to sets of HR images fused using incorrect depth assumptions. By comparing HR images formed at various depths from the LR images captured by each set of focal planes, their mutual similarities form a good measure of

32

"correctness" of the fused images, thus providing a depth estimate for the aliased regions.

[00104] A process for performing hypothesized fusion involving the generation of two fused images in accordance with embodiments of the invention is illustrated in FIG. 7. The process 700 utilizes LR images captured by two sets of focal planes in the imager array. In several embodiments, the two sets of focal planes are non-distinct (i.e. some focal planes are included in both sets) such that there is considerable spatial overlap between the camera viewpoints across the two sets of LR images. Assuming (702) a particular depth d for one or more points in the scene, the corresponding pixels in each set of LR images are fused (704) onto a higher resolution grid generated from a chosen reference viewpoint using known camera calibration information. A simple hole filling mechanism can then be used to fill (706) holes at locations in the fused images where no pixels converge. The two HR images are then compared (708) and a robust cost $C(i, d)$ is computed as the error in pixel match (at each location indexed by i) in the two fused images formed using a hypothesized depth d. For each pixel i, the depth is determined (710) as a depth d_i for which the error in matching $C(i, d)$ is the least over all the sampled depths. This provides a depth map for all pixels in the image, even in regions with significant variation amongst the LR images due to aliasing. In other embodiments, any of a variety of termination conditions can be utilized including (but not limited to) terminating the analysis of different hypothesized depths when a matching score exceeding a threshold is determined.

[00105] Although specific processes for determining the most likely hypothesized depth of a pixel in a HR image are discussed above with reference to FIG. 7, any of a variety of processes can be utilized involving comparison of two or more fused HR images generated using LR images captured by sets of focal planes within an imager array in accordance with embodiments of the invention. Processes for partitioning the focal planes in an imager array into sets of focal planes when performing hypothesized fusion in accordance with embodiments of the invention are discussed further below.

PARTITIONING FOCAL PLANES IN AN ARRAY CAMERA MODULE

[00106] In an array camera module, cameras can be partitioned into two or more groups of cameras (irrespective of the particular frequency range captured by the cameras). In a number of embodiments, cameras capturing information in a particular frequency range (color) can be partitioned into two or more groups of cameras for the purpose of performing hypothesized fusion. In many embodiments, the complexity of hypothesized fusion can be reduced by only considering the Green cameras. The spatial shift between cameras causes phase shifts in the aliased frequencies resulting in aliased regions in images that appear considerably different between captured LR images of a scene. As discussed earlier, the goal of selecting two or more sets of cameras is to exploit these dissimilarities to fuse similar images at the correct depth hypothesis for each pixel or region within a HR image. That is to say, when pixels from each of the sets of LR images are placed on a higher resolution grid, the reduction in aliasing effects enables the reconstruction of very similar images for each set (assuming the correct depth hypotheses are used). The difference between the fused images is typically primarily related to the errors in the depth hypotheses.

[00107] In many embodiments, each set of LR images is used to generate a HR image from the viewpoint of a common reference camera. Depending upon the specific imager array used to capture the LR images, the reference camera may be located at the center of the camera or in a location offset from the center of the camera. In many embodiments, the HR images can be synthesized from a virtual viewpoint. The manner in which the focal planes in an array camera module are partitioned typically depends upon the number of focal planes in the array camera module. In many embodiments, the hypothesized fusion process attempts to construct fused images that are as similar as possible. Since the fusion step attempts to achieve sub-pixel accuracy, errors in camera calibration can lead to differences between the HR images created by fusing the LR images captured by the sets of focal planes. In many embodiments, the partitioning is performed so that a number of focal planes are common to each set.

[00108] Hypothesized fusion can be performed by partitioning the focal planes in an array camera module into two *disjoint* sets. In a number of embodiments, the focal

planes are partitioned into two sets that divide the focal planes in the camera array horizontally, vertically or in any of the diagonal directions. Forming two disjoint sets of cameras, however, is likely to result in a smaller number of focal planes in each set of focal planes (compared to forming overlapping sets of focal planes). With few cameras, a smaller level of magnification can be achieved without leaving large holes that need to be interpolated. Therefore, depending upon the number of focal planes in the imager array, partitioning the focal planes into disjoint sets can result in retaining much of the aliasing effects.

[00109] The location of the reference camera or viewpoint can also be an important consideration when partitioning the focal planes in an imager array. Since the fused images are formed using the reference camera as the viewpoint, it is useful to select the viewpoint of an actual camera (as opposed to a virtual viewpoint) as the reference the reference camera in both sets. Additionally, advantages can be obtained by selecting the sets of focal planes such that each set has considerable coverage on any given side (above, below, left and right) of the reference camera. Therefore, in many embodiments each set includes a focal plane above, below, to the left, and to the right of the reference camera. This increases the likelihood that the pixels placed on the higher resolution grid are similarly distributed in either set. At the same time, selecting the sets in this way can minimize errors associated with occlusion zones and increase the likelihood that whatever errors remain are present in each of the fused HR images.

[00110] The selection of cameras based on the above considerations, typically depends upon the number and grid configuration of the focal planes in an imager array (including the location of color filters within the array camera module), the scene being imaged, and the requirements of specific applications. In several embodiments, a determination is made dynamically concerning the manner in which to partition the focal planes based upon sensor information, and/or the captured LR images. In other embodiments, predetermined partitions are utilized.

[00111] The partitioning of focal planes in an array camera module in accordance with embodiments of the invention is illustrated in FIGS. 8A - 8C. A grid of cameras in an array camera module is conceptually illustrated in FIG. 8A. The array camera module

includes a 5x5 configuration of cameras including 6 Blue cameras, 13 Green cameras, and 6 Red cameras. The Green cameras in the array camera module illustrated in FIG. 8A can be partitioned into the two sets illustrated in FIGS. 8B and 8C. The five central Green cameras are common to both sets and the additional four cameras in each set are distinct to the set. Both sets include the central Green camera, which is typically used as a reference camera when synthesizing high resolution images using a 5x5 array. In arrays that don't have a central imager, a Green camera proximate the center of the array is typically utilized as the reference camera.

[00112] Although the partitioned sets illustrated in FIGS. 8B and 8C include the same numbers of cameras, in many embodiments the sets can include different numbers of cameras. Furthermore, hypothesized fusion is not limited to using a single type of camera from within the array camera module. The manner in which hypothesized fusion can be performed using LR images captured by partitioned sets of focal planes within an imager array in accordance with embodiments of the invention is discussed further below.

FUSION WITH DEPTH HYPOTHESIS

[00113] Based upon the partitioning of focal planes, the LR images captured by the focal planes in each set can be fused onto a high resolution grid. The process of fusing the LR images onto the high resolution grid utilizes camera calibration data determined a priori. Placing pixels from a focal plane to a high resolution grid from the viewpoint of the reference camera involves accounting for the relative baselines of the focal planes, the focal length (can be assumed fixed) as well as the depth of the point whose projected pixel is being considered. Initially, the actual depth or distance of the point from the camera plane is unknown. To solve for depth, a list of possible depth hypotheses can be utilized. For each depth d in the list, pixels from each LR image in the appropriate set can be placed onto the higher resolution grid taking into account the magnification amount being considered. Since the focal planes within each set capture the same image from a slightly shifted viewpoint, the pixels may be sub-pixel shifted and, hence, may be placed at a slightly different location in the HR grid. The final fused

image is then formed by some form of interpolation that produces a regularly sampled grid from the randomly sampled observations. Given enough samples arising out of a sufficient number of LR images in each set, simple interpolation schemes such as kernel regression can be employed.

[00114] Performing fusion using each set of LR images provides HR images or portions of HR images for each hypothesized depth. At incorrect depths, the pixel placements can be expected to be erroneous. As a result, images are obtained that are not really representative of the captured scenes. This is especially true of regions including higher levels of aliasing. As mentioned earlier, the measure of "correctness" is calculated by the similarity of the HR images produced by each of the sets of LR images. Where focal planes are common to both sets, the LR images captured by the common focal planes contribute equally to both the HR images at the correct as well as incorrect hypothesized depths. However, at incorrect depths, a sufficiently large number of distinct cameras with quite varied aliasing effects produce a lower match between image regions obtained from different sets of images. Since each of these cameras sample differently shifted images, the samples at any given incorrect depth hypothesis are inaccurately placed on the HR grid, increasing the differences between the fused HR images created using each set. At the correct depth hypothesis for any given region, when the samples can be expected to be placed correctly, both sets produce images that are sufficiently free of aliasing to achieve a threshold level of similarity. Employing a proper measure of similarity over the entire hypothesized depths, it is possible to estimate the actual depth of an aliased region in the captured LR images. Different locations in the image may produce best matches at different depth hypotheses. Choosing the best matches at each location allows the production of a depth estimate for each pixel, and, hence, a depth map for the entire image. In many embodiments, the computational complexity of performing hypothesized fusion can be reduced by only performing hypothesized fusion in regions where a threshold level of aliasing is present. In several embodiments, the computational complexity of performing hypothesized fusion can also be reduced by utilizing information concerning the depth of an adjacent region or pixel to commence the search for a correct depth.

When a threshold level of similarity is achieved, a depth can be selected. In this way, the number of depth hypotheses tested by the process can be reduced in regions having similar depths. Processes for determining the similarity of regions of fused HR images when performing hypothesized fusion in accordance with embodiments of the invention are discussed further below.

DETERMINING SIMILARITY

[00115]   Estimating the correct depth at any given location during hypothesized fusion relies on an ability to accurately measure the similarity between fused HR images. In a number of embodiments, image patches are used to determine local similarities. Comparisons of portions or patches of images provide robustness to the presence of noise, which can be expected to corrupt the captured LR images. Further, uses of small image portions can account for the local content surrounding a pixel under consideration when comparing fused HR images, thereby avoiding spurious matches. For each pixel (or a selection of pixels) in the HR grid where a pixel from the reference camera is positioned, a *MxN* patch can be formed with the pixel of interest at its center. This is done for images from each set. Where the focal planes are partitioned into two sets, these image portions $p_{1,i}$ and $p_{2,i}$ are then compared using a measure of their difference. In a number of embodiments, an L_1 or an L_2 norm is used to measure the difference although other measures appropriate to the requirements of a specific application can also be utilized. Note that this can be done for each portion in the HR image, and for each hypothesized depth. Mathematically, the cost $C(i, d)$ is obtained using the following expression:

$$C(i,d) = \left\| p_{1,i} - p_{2,i} \right\|_2 \tag{2}$$

[00116]   Once such a cost is computed for all hypothesized depths, the depth at each location of interest (e.g. each pixel from the reference camera) can be computed as

$$\hat{d}_i = \arg\min_d C(i,d). \tag{3}$$

[00117]   Although there exists many cost functions that may be utilized to compare two image portions or patches for similarity, the L_2 norm provides a simple yet effective

measure of similarity appropriate to many applications. Experimental evidence that illustrates the recovery of high frequency information from LR images in which aliasing is introduced using pixel apertures using hypothesized fusion processes, similar to the hypothesized fusion processes described above, is discussed below.

RECOVERY OF HIGH FREQUENCY INFORMATION

[00118] SR processes in accordance with embodiments of the invention can utilize aliasing within captured images of a scene to recover high frequency information. The increase in resolution of the resulting synthesized images can be appreciated by comparing images synthesized from images captured using imager arrays constructed to minimize aliasing and images synthesized from images captured using imager arrays constructed to introduce aliasing into the captured image (e.g. by using microlenses that create pixel apertures).

[00119] A simulation of the resolution improvement that can be obtained by introducing aliasing into captured images in accordance with embodiments of the invention is illustrated in FIGS. 9A and 9B. FIG. 9A is a first SR image 900 synthesized using images captured by a 5 x 5 array of VGA resolution imagers, where each pixel in the imager has a pixel pitch of 2.2μm. FIG. 9B is a second SR image 902 synthesized by simulating images captured by a 5 x 5 array of VGA resolution images, where each pixel has a 2.2μm pixel pitch, but is sampled in a 1.4μm area due to the use of microlenses that do not span the full 2.2μm pixel pitch to create a pixel aperture. The use of microlens pixel apertures in this way increases the aliasing captured in the LR images. As can be appreciated in a corresponding region 904 in each image, the ability of the SR processing to recover high frequency information is enhanced due to the increase in aliasing in the captured images caused by the pixel apertures. Super resolution processes may, however, rely on hypothesized fusion to provide correct depth information.

[00120] The size of the microlenses and therefore the implied sizes of the gaps between the microlenses is a tradeoff between increased aliasing and decreased sensitivity. Assuming square pixels and therefore square microlenses the area covered

by the microlenses is proportional to the square of the pixel pitch. An additional consideration is the reduction of the pixel stack and therefore a concomitant reduction in the cross-talk between adjacent pixels. The removal of the color filter from the pixel stack has the potential of reducing the pixel stack height by 30%. In many embodiments, the consequent reduction in cross-talk through pixel stack height reduction results in increased sensitivity and can be traded-off for increased aliasing. Similar issues are faced by pixels including microlenses that are not square.

[00121]   Although specific imager resolutions and pixel sizes are described above, as can readily be appreciated the imager resolution, the pixel sizes, and the apertures used to introduce aliasing into the captured images can be selected as appropriate to the requirements of a specific application in accordance with embodiments of the invention.

[00122]   While the above description contains many specific embodiments of the invention, these should not be construed as limitations on the scope of the invention, but rather as an example of one embodiment thereof. Accordingly, the scope of the invention should be determined not by the embodiments illustrated, but by the appended claims and their equivalents.

WHAT IS CLAIMED IS:

1.   A method of determining a depth of a point in a scene using light field image data comprising a set of low resolution images that capture the scene, the method comprising:

        fusing portions of the set of low resolution images to form a portion of a higher resolution image at each of a plurality of hypothesized depths, where the resolution of the portion of the higher resolution image is higher than the resolutions of the portions of the set of low resolution images used to fuse the portion of the higher resolution image;

        comparing the portion of the fused higher resolution image obtained at each hypothesized depth to the scene captured in the set of low resolution images; and

        selecting the hypothesized depth at which the portion of the fused higher resolution image is most similar to the scene captured in the set of low resolution images as the depth of at least one point in the scene captured by the set of low resolution images.

2.   The method of claim 1, wherein comparing the portion of the fused higher resolution image to the scene captured in the set of low resolution images comprises:

        generating a set of forward mapped low resolution image portions by forward mapping the portion of the fused higher resolution image using a mapping based upon the characteristics of the cameras utilized to capture the set of low resolution images; and

        comparing the forward mapped low resolution image portions with corresponding portions of corresponding images in the set of low resolution images.

3.   The method of claim 1, wherein comparing the portion of the fused higher resolution image to the scene captured in the set of low resolution images comprises determining the similarity of pixels in pixel stacks in the portion of the fused higher resolution image.

4.   The method of claim 1, wherein comparing the portion of the fused higher resolution image at a specific hypothesized depth to the scene captured in the set of low resolution images comprises comparing the portion of the fused higher resolution image at the specific hypothesized depth to a portion of at least a second fused higher resolution image formed by fusing a second set of low resolution images at the specific hypothesized depth.

5.   The method of claim 4, wherein at least one low resolution image is common to said set of low resolution images and said second set of low resolution images.

6.   A machine readable medium containing processor instructions, where execution of the instructions by a processor causes the processor to perform a process comprising:

        fusing portions of the set of low resolution images to form a portion of a higher resolution image at each of a plurality of hypothesized depths, where the resolution of the portion of the higher resolution image is higher than the resolutions of the portions of the set of low resolution images used to fuse the portion of the higher resolution image;

        comparing the portion of the fused higher resolution image obtained at each hypothesized depth to the scene captured in the set of low resolution images; and

        selecting the hypothesized depth at which the portion of the fused higher resolution image is most similar to the scene captured in the set of low resolution images as the depth of at least one point in the scene captured by the set of low resolution images.

7.   A method of determining a depth of a point in a scene using light field image data comprising a set of low resolution images that capture the scene, the method comprising:

fusing portions of a first subset of a set of low resolution images to form a portion of a first higher resolution image at each of a plurality of hypothesized depths, where the resolution of the portion of the first higher resolution image is higher than the resolutions of the portions of the first subset of the set of low resolution images used to fuse the portion of the first higher resolution image;

fusing portions of a second subset of the set of low resolution images to form a portion of a second higher resolution image at each of the plurality of hypothesized depths, where the resolution of the portion of the second higher resolution image is higher than the resolutions of the portions of the second subset of the set of low resolution images used to fuse the portion of the second higher resolution image;

comparing at least the portions of the first and second higher resolution images fused at each of the plurality of hypothesized depths; and

selecting the hypothesized depth at which the portions of the compared higher resolution images are most similar as the depth of at least one point in the scene imaged by pixels within portions of the first and second subsets of the set of low resolution images.

8.   The method of claim 7, wherein at least one low resolution image in the set of low resolution images is common to both the first and second subsets of the set of low resolution images.

9.   The method of claim 8, wherein:

the viewpoint of one of the low resolution images in the set of low resolution images is selected as the reference viewpoint used to fuse the portions of the first and second high resolution images at each of the plurality of hypothesized depths; and

the low resolution image selected as the reference viewpoint is common to both the first and second subsets of the set of low resolution images.

43

10.   The method of claim 9, wherein low resolution images captured from viewpoints above, below, to the left, and to the right of the reference viewpoint are common to both the first and second subsets of the set of low resolution images.

11.   The method of claim 7, wherein the viewpoint of one of the low resolution images in the set of low resolution images is selected as the reference viewpoint used to fuse the portions of the first and second high resolution images at each of the plurality of hypothesized depths

12.   The method of claim 7, wherein fusing portions of a subset of the set of low resolution images to form a portion of a higher resolution image at a hypothesized depth comprises:
        identifying pixels within the subset of the set of low resolution images based upon the hypothesized depth and the viewpoints of the low resolution images;
        fusing the identified pixels onto a higher resolution grid generated from a chosen reference viewpoint using known calibration information; and
        performing hole filing to fill holes in locations in the higher resolution grid.

13.   The method of claim 7, wherein comparing at least the portions of the first and second higher resolution images fused at each of the plurality of hypothesized depths comprises comparing the portions of the first and second higher resolution images for matching error at each of the plurality of hypothesized depths.

14.   The method of claim 13, wherein matching error is determined using at least one selected from the group of the L_1-norm and the L_2-norm of the difference of the portions of the first and second higher resolution image.

15.   The method of claim 7, further comprising:
        fusing portions of a third subset of the set of low resolution images to form a portion of a third higher resolution image at each of a plurality of hypothesized depths,

44

where the resolution of the portion of the third higher resolution image is higher than the resolutions of the portions of the third subset of the set of low resolution images used to fuse the portion of the first higher resolution image;

wherein comparing at least the portions of the first and second higher resolution images fused at each of the plurality of hypothesized depths further comprises comparing the portions of the first, second, and third higher resolution images fused at each of the plurality of hypothesized depths.

16.   An array camera, comprising:

an array camera module, comprising:

an imager array, comprising:

a plurality of focal planes, where each focal plane comprises a two dimensional arrangement of pixels having at least two pixels in each dimension and each focal plane is contained within a region of the imager array that does not contain pixels from another focal plane;

control circuitry configured to control the capture of image information by the pixels within the focal planes; and

sampling circuitry configured to convert pixel outputs into digital pixel data;

interface circuitry configured to transmit digital pixel data;

an optic array of lens stacks, where an image including aliasing is formed on each focal plane by a separate lens stack in the optic array of lens stacks; and

a processor configured to receive digital pixel data from the array camera module via the interface circuitry; and

memory containing a image processing pipeline application and a controller application;

wherein the processor is configured via the controller application to read digital pixel data from the imager array;

wherein the image processing pipeline application configures the processor to:

obtain a set of low resolution images of a scene that include aliasing by reading digital pixel data from the imager array; and

synthesize a higher resolution image of the scene from a reference viewpoint using the set of low resolution images.

17.   The array camera of claim 16, wherein the image processing pipeline application configures the processor to determine a depth of at least one pixel in the synthesized higher resolution image by:

fusing portions of the set of low resolution images to form a portion of a higher resolution image at each of a plurality of hypothesized depths, where the resolution of the portion of the higher resolution image is higher than the resolutions of the portions of the set of low resolution images used to fuse the portion of the higher resolution image;

comparing the portion of the fused higher resolution image obtained at each hypothesized depth to the scene captured in the set of low resolution images; and

selecting the hypothesized depth at which the portion of the fused higher resolution image is most similar to the scene captured in the set of low resolution images as the depth of at least one point in the scene captured by the set of low resolution images.

18.   The array camera of claim 17, wherein the image processing pipeline application configures the processor to compare the portion of the fused higher resolution image to the scene captured in the set of low resolution images by:

generating a set of forward mapped low resolution image portions by forward mapping the portion of the fused higher resolution image using a mapping based upon the characteristics of the array camera module; and

comparing the forward mapped low resolution image portions with corresponding portions of corresponding images in the set of low resolution images.

19. The array camera of claim 17, wherein the image processing pipeline application configures the processor to compare the portion of the fused higher resolution image to the scene captured in the set of low resolution images by determining the similarity of pixels in pixel stacks in the portion of the fused higher resolution image.

20. The array camera of claim 16, wherein the image processing pipeline application configures the processor to determine a depth of at least one pixel in the synthesized higher resolution image by:

comparing portions of fused higher resolution images formed by fusing at least two subsets of the set of low resolution images at a plurality of hypothesized depths; and

selecting the depth of at least one pixel in the synthesized higher resolution image based upon the hypothesized depth at which the compared portions of fused higher resolution images are most similar.

21. The array camera of claim 20, wherein at least one low resolution image in the set of low resolution images is common to the at least two subsets of the set of low resolution images.

22. The array camera of claim 21, wherein the at least two subsets of the set of low resolution images include a common low resolution image having a viewpoint that is the reference viewpoint.

23. The array camera of claim 22, wherein the at least two subsets of the set of low resolution images include common low resolution images captured from viewpoints above, below, to the left, and to the right of the reference viewpoint.

24.   The array camera of claim 20, wherein the reference viewpoint is the viewpoint of one of the low resolution images in the set of low resolution images.

25.   The array camera of claim 20, wherein the image processing pipeline application configures the processor to fuse at least two subsets of the set of low resolution images at a hypothesized depth by:

identifying pixels within the subset of the set of low resolution images based upon the hypothesized depth and the viewpoints of the low resolution images;

fusing the identified pixels onto a higher resolution grid generated from a chosen reference viewpoint using known calibration information; and

performing hole filing to fill holes in locations in the higher resolution grid.

26.   The array camera of claim 20, wherein the image processing pipeline application configures the processor to compare portions of fused higher resolution images by comparing matching error of the portions of the fused higher resolution images.

27.   The array camera of claim 26, wherein matching error is determined using at least one selected from the group of the L_1-norm and the L_2-norm of the difference of the portions of the first and second higher resolution image.

28.   The array camera of claim 17, wherein:

the pixels in the plurality of focal planes in the imager array comprise a pixel stack including a microlens and an active area, where light incident on the surface of the microlens is focused onto the active area by the microlens and the active area samples the incident light to capture image information; and

the pixel stack defines a pixel area and includes a pixel aperture, where the size of the pixel apertures is smaller than the pixel area.

29.   The array camera of claim 28, wherein the pixel aperture is formed by a microlens that is smaller than the pixel area.

30.   The array camera of claim 29, wherein gaps exist between adjacent microlenses in the pixel stacks of adjacent pixels in a focal plane.

31.   The array camera of claim 30, wherein light is prevented from entering the pixel stacks through the gaps between the microlenses by a light blocking material.

32.   The array camera of claim 28, wherein the pixel stack includes a color filter.

33.   The array camera of claim 32, wherein the color filters in the pixel stacks of the two dimensional arrangement of pixels within a focal plane are the same.

34.   The array camera of claim 32, wherein the color filters in the pixel stacks of the two dimensional arrangement of pixels within at least one focal plane form a Bayer filter pattern.

35.   The array camera module of claim 28, wherein at least one of the plurality of lens stacks includes a color filter and the pixel stacks of the two dimensional arrangement of pixels within the focal plane on which said at least one of the plurality of lens stacks forms an image do not include color filters.

36.   The array camera module of claim 28, wherein the pixel aperture is formed using at least one light blocking material.

37.   The array camera module of claim 16, wherein the optic array of lens stacks is constructed using wafer level optics.

38. The array camera module of claim 16, wherein the plurality of lens stacks include polymer optical components.

39. The array camera module of claim 16, wherein the plurality of lens stacks include glass optical components.

FIG. 1
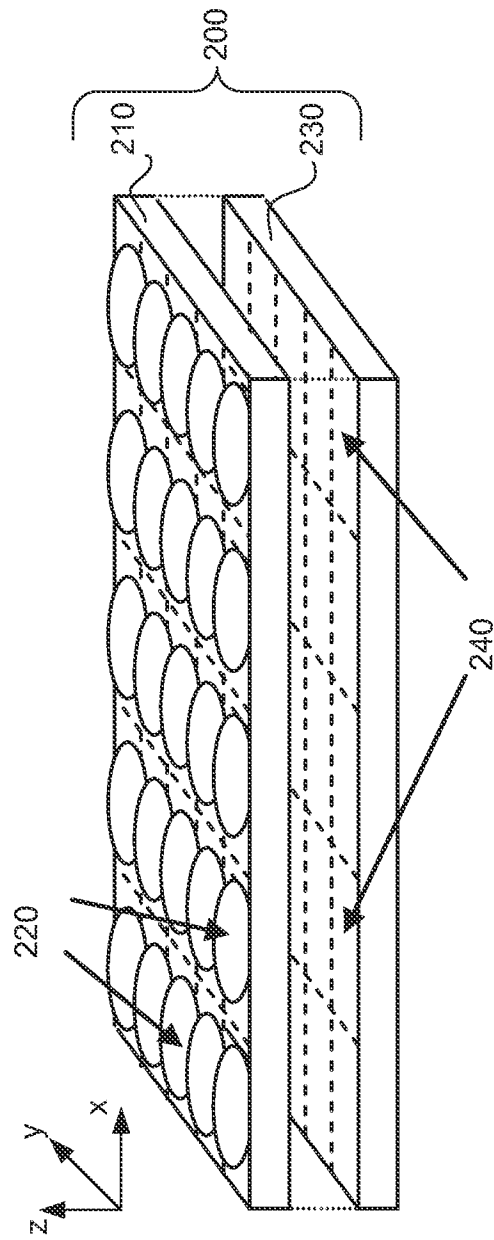
FIG. 2
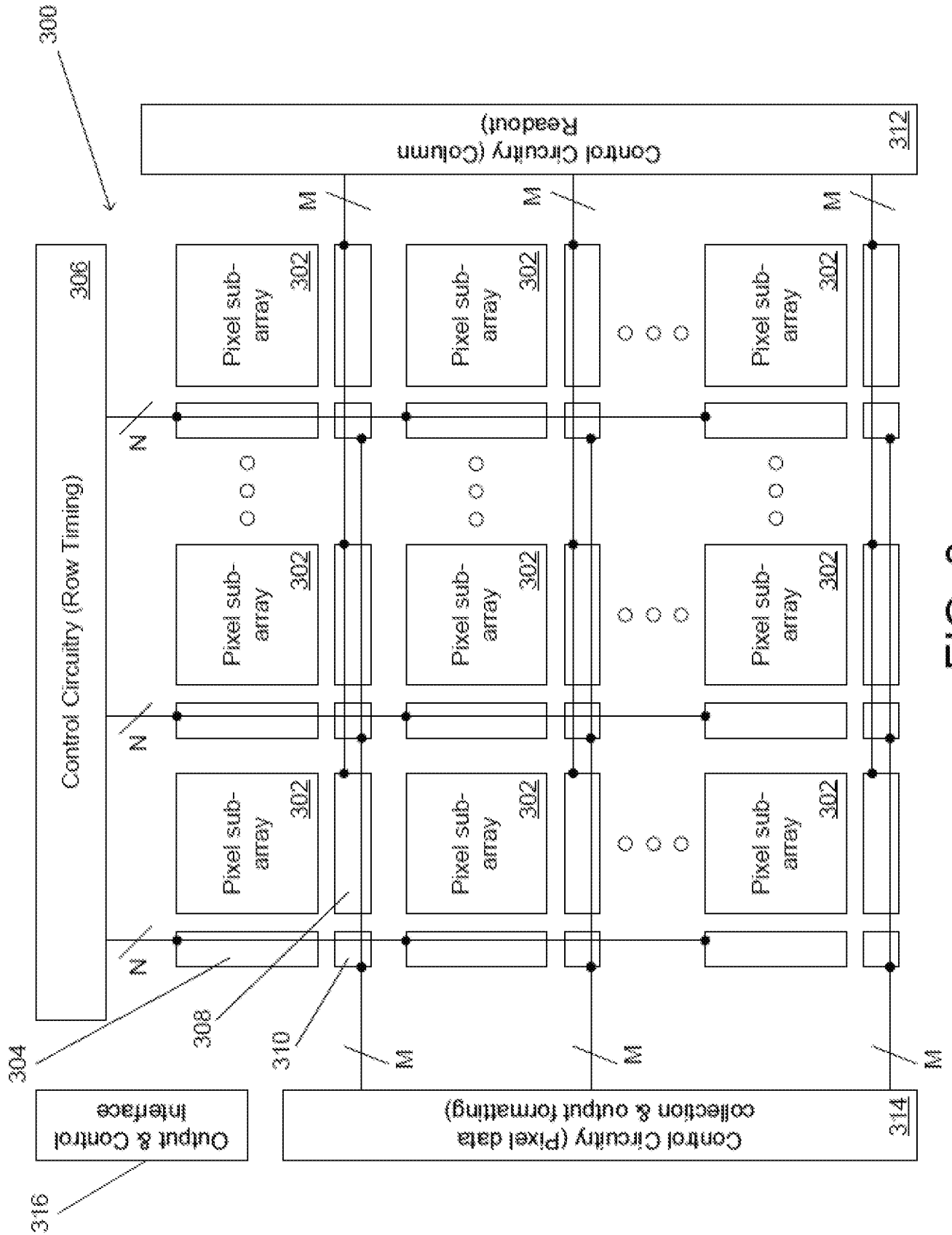
FIG. 3
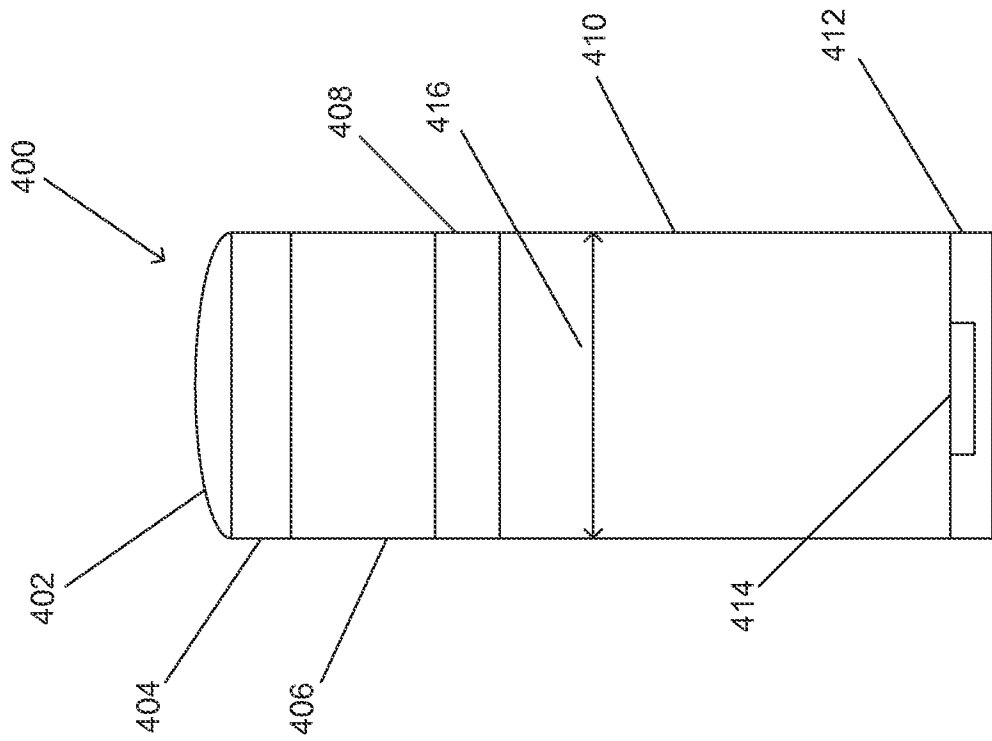
FIG. 4B
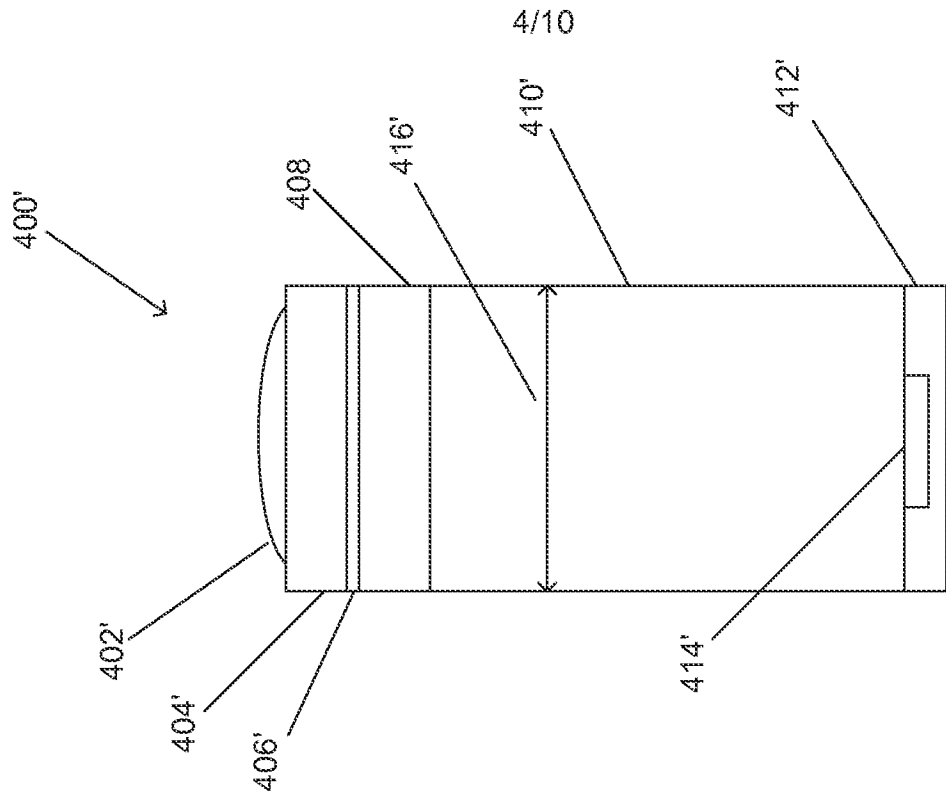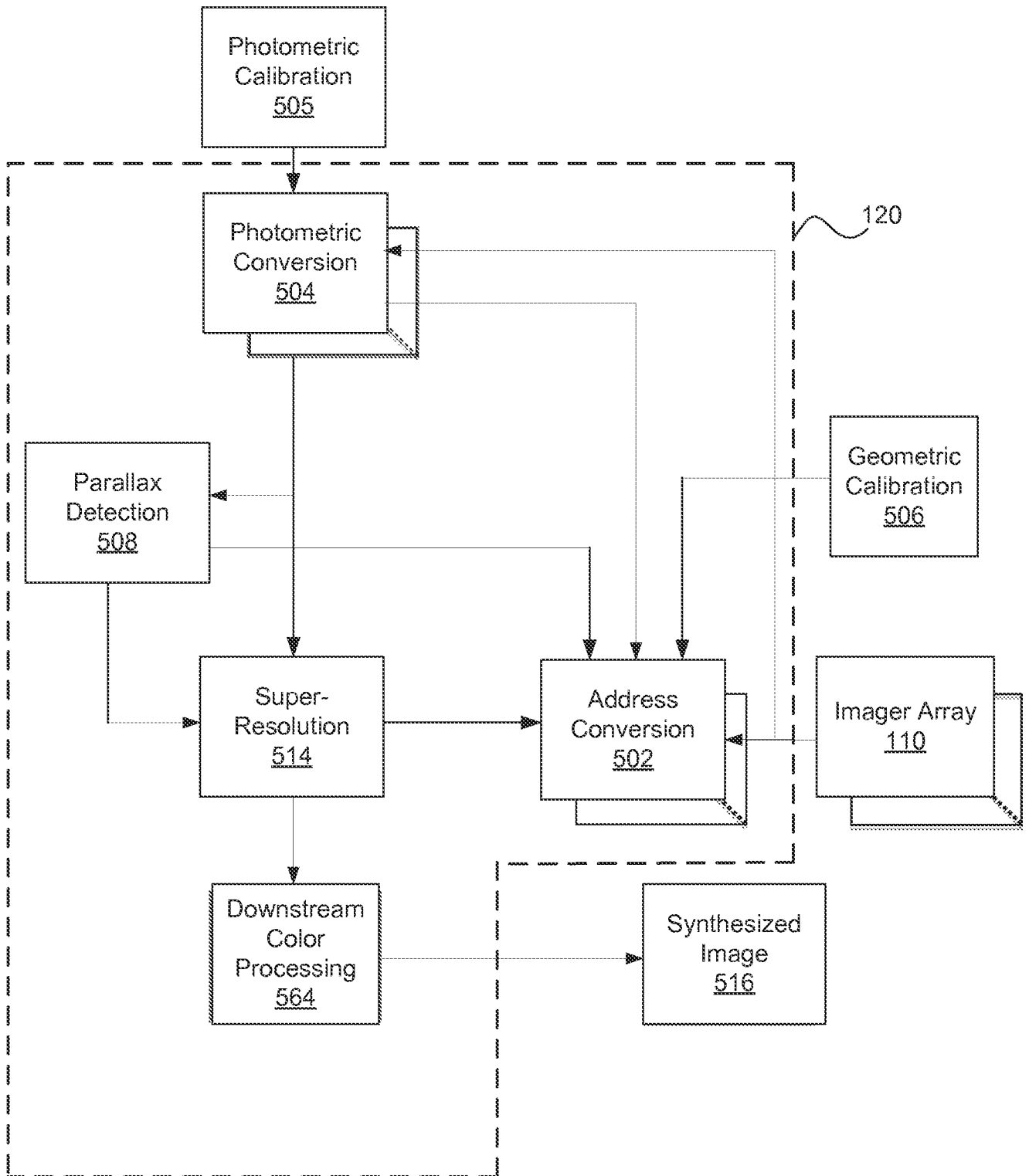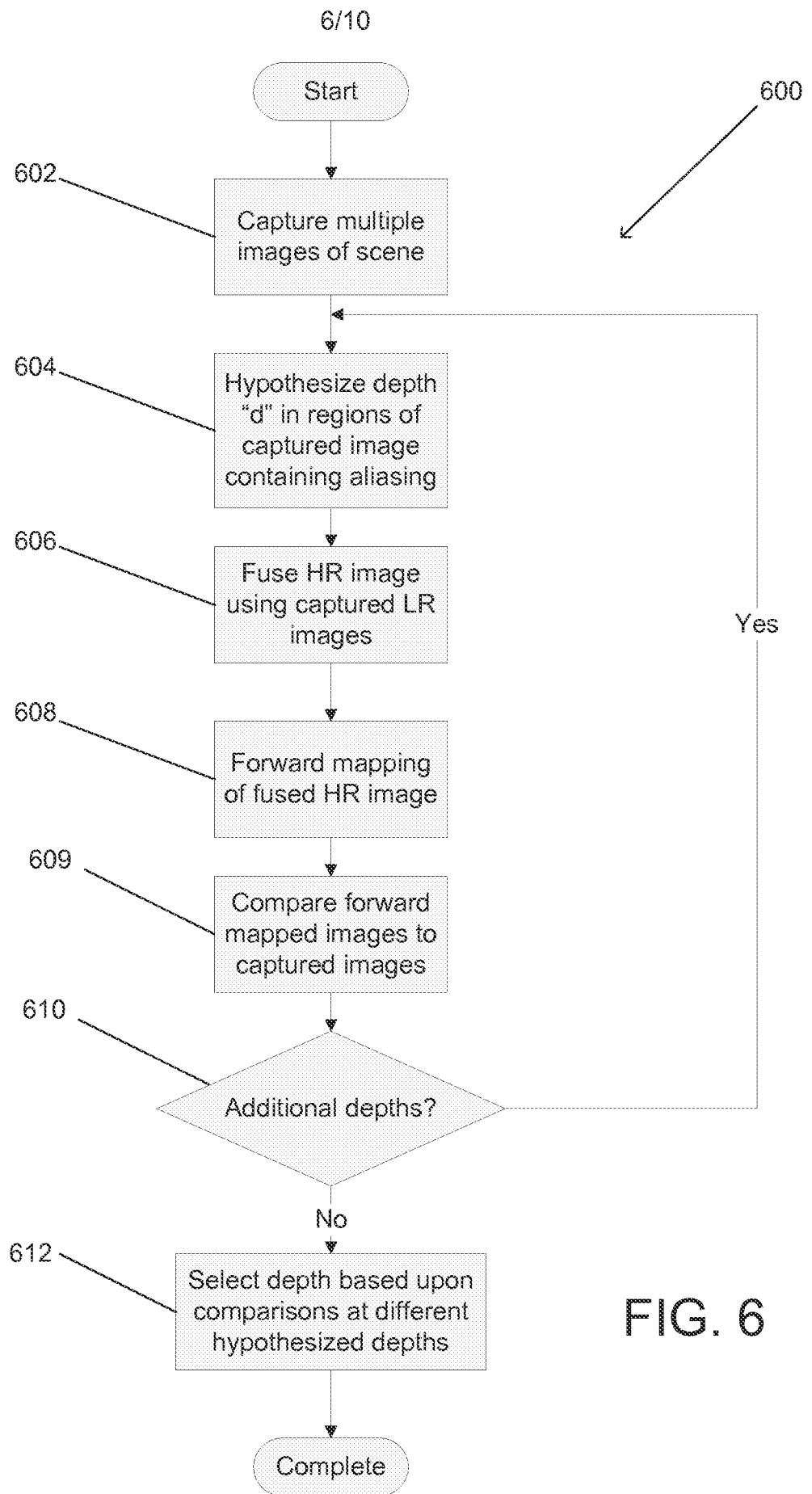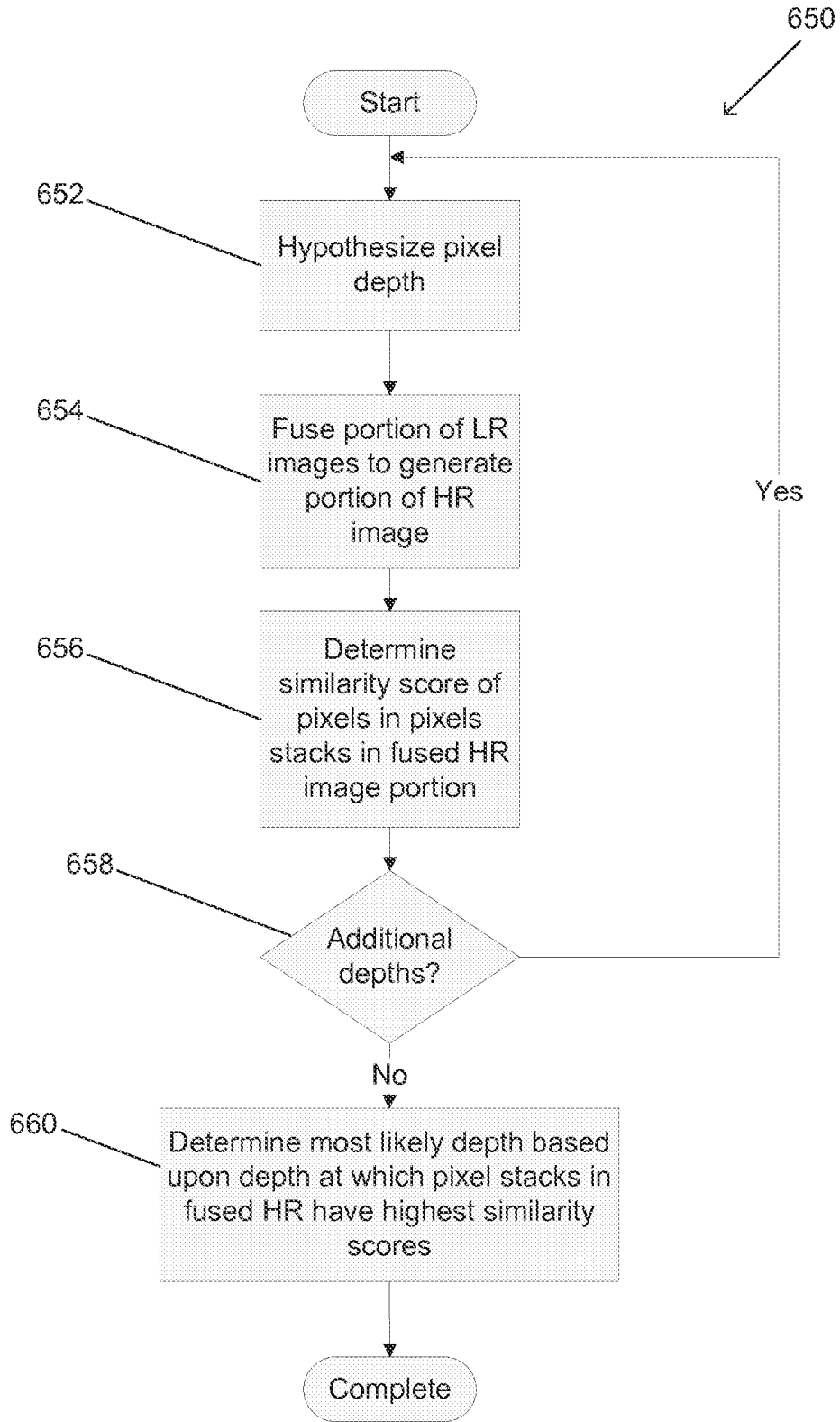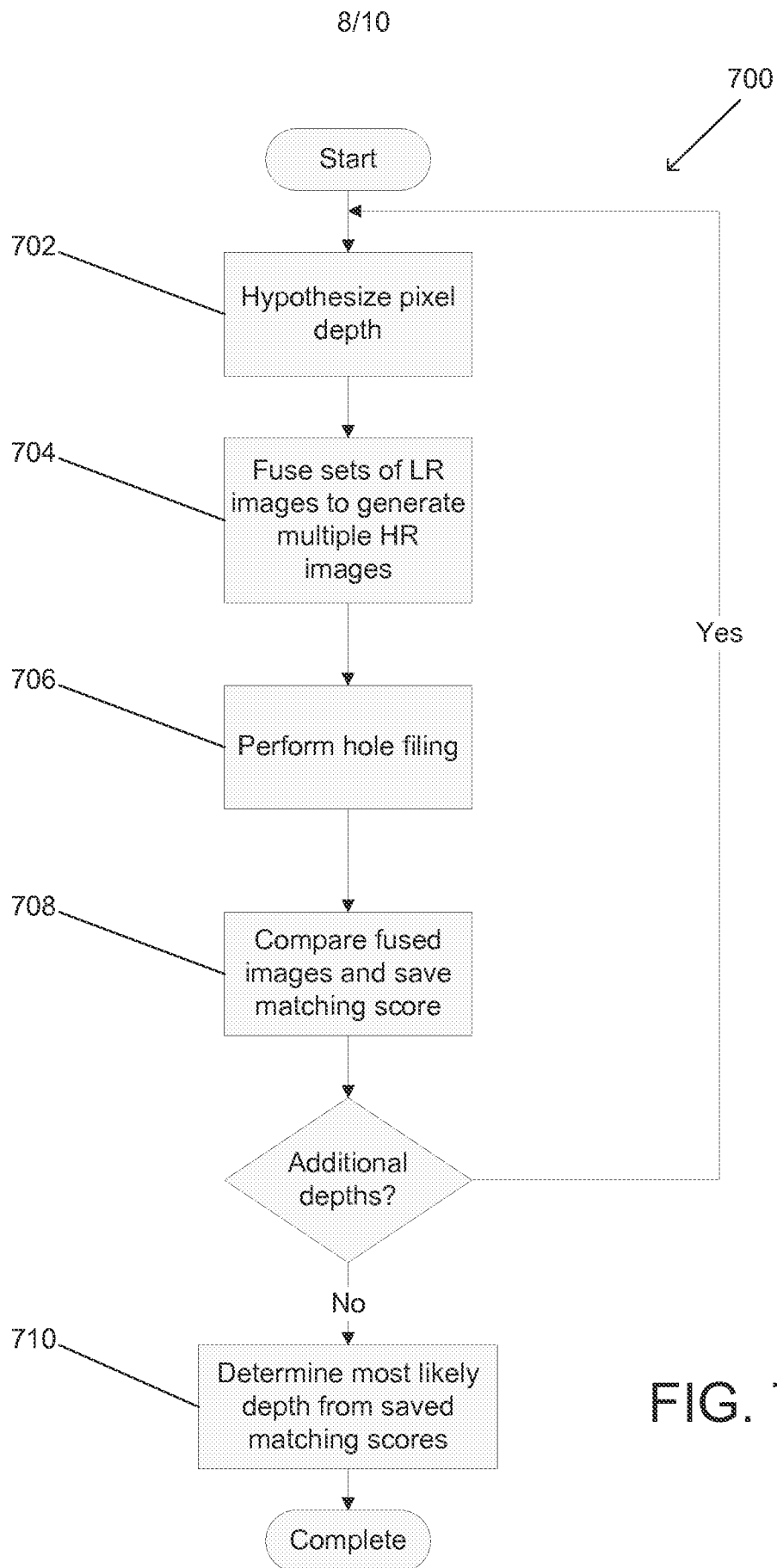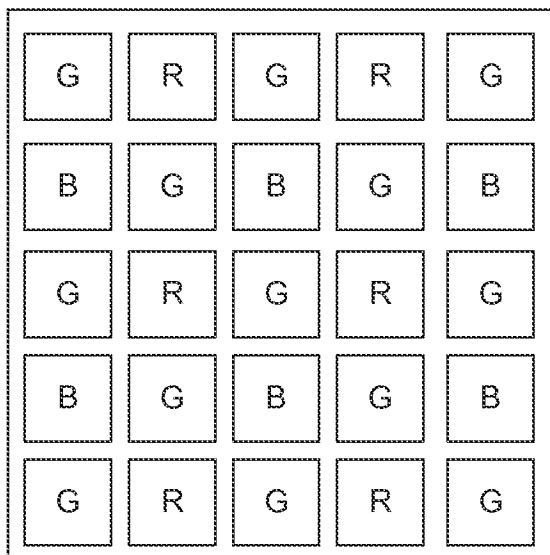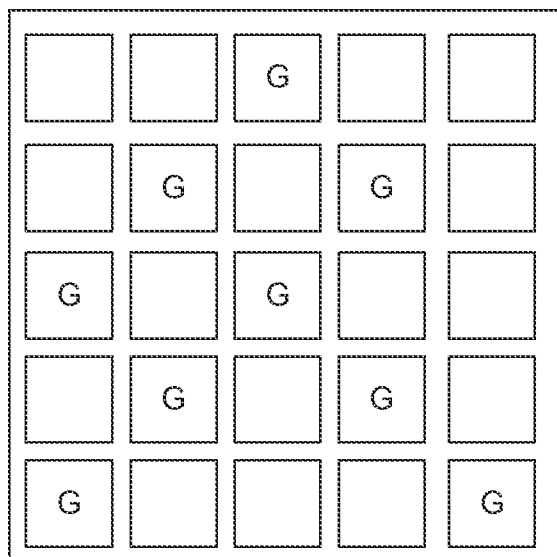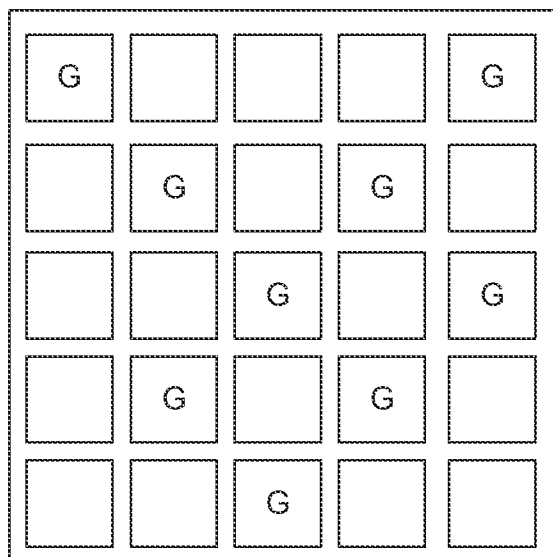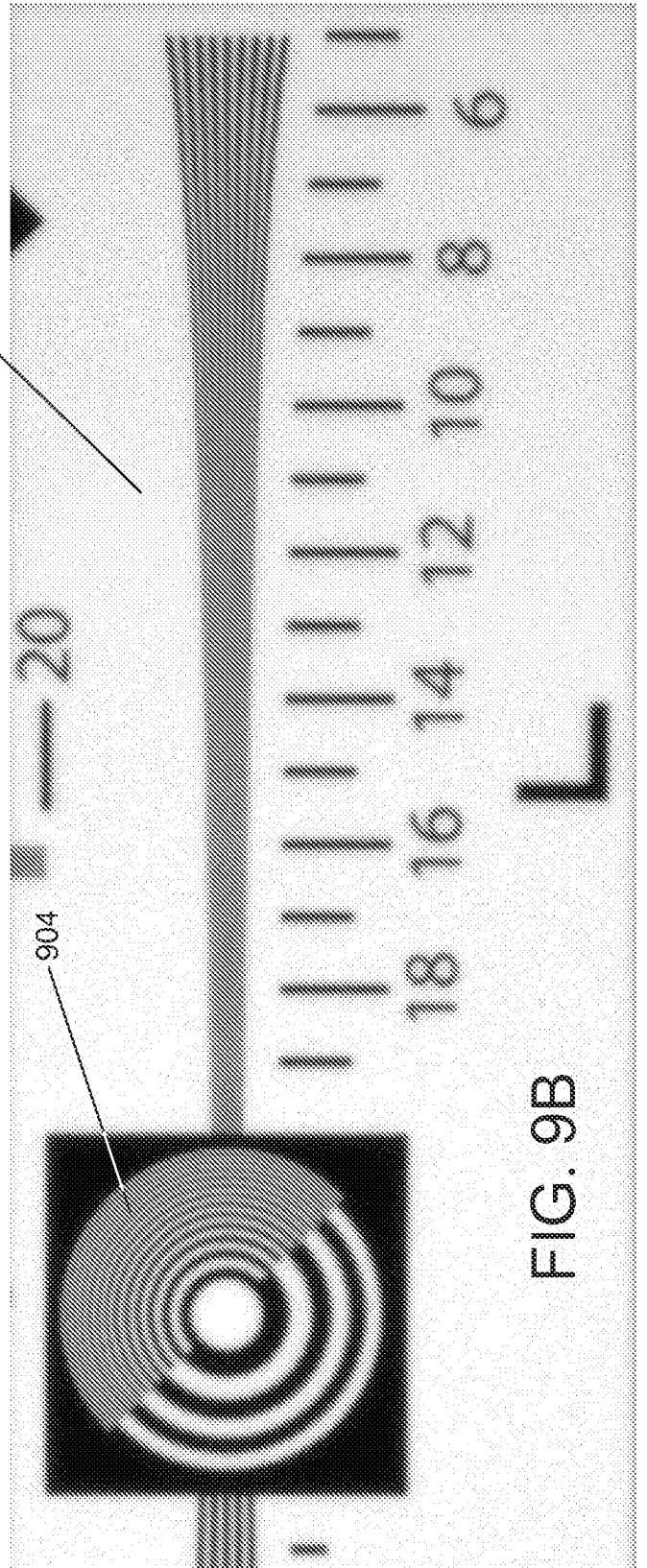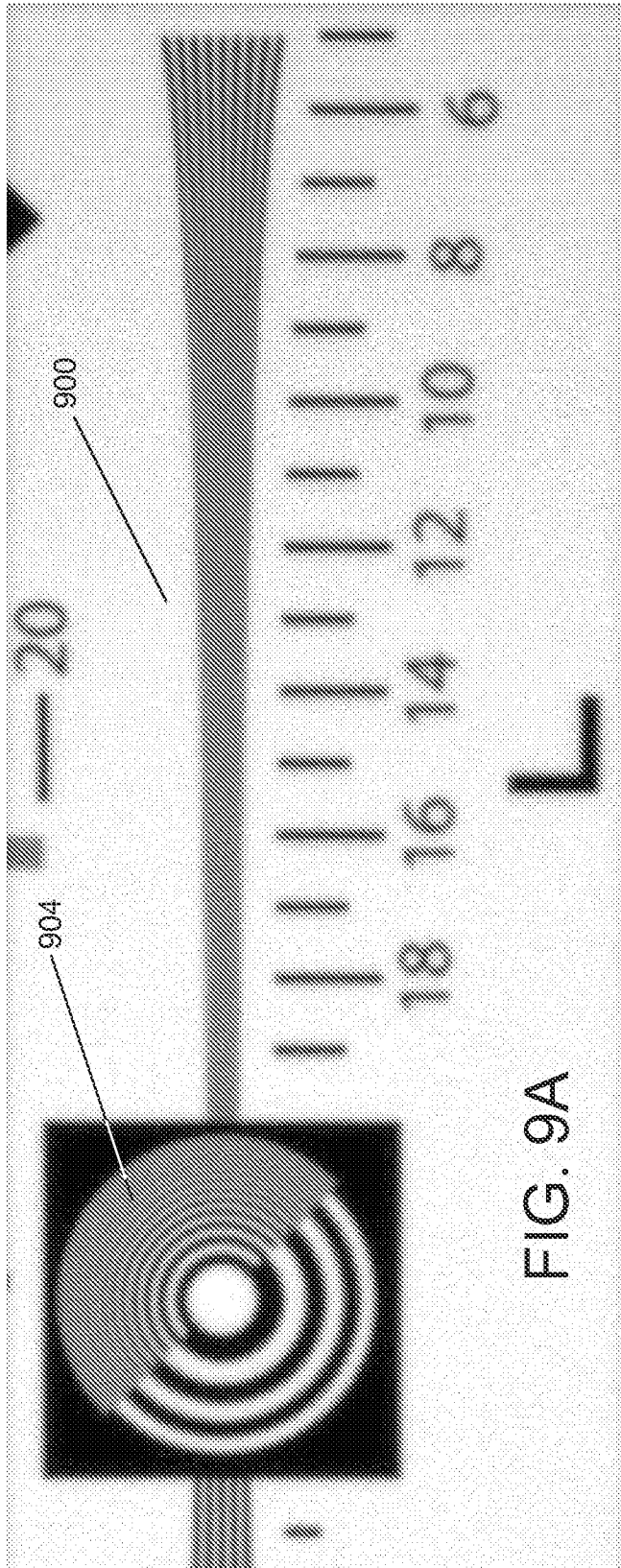
FIG. 4A

FIG. 5

FIG. 6

FIG. 6a

700

Start

702 — Hypothesize pixel depth

704 — Fuse sets of LR images to generate multiple HR images

706 — Perform hole filing

708 — Compare fused images and save matching score

Additional depths?

Yes

No

710 — Determine most likely depth from saved matching scores

Complete

FIG. 7

9/10

FIG. 8A

FIG. 8B

FIG. 8C

FIG. 9A

FIG. 9B

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - H04N 5/225 (2012.01)
USPC - 348/218.1
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)
USPC: 348/218.1
IPC(8): H04N 5/225 (2012.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 348/207.99, 218.1, 335, 340 (Keyword limited; terms below)
IPC(8): H04N 5/225 (2012.01) (Keyword limited; terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST(PGPB, USPT, EPAB, JPAB); Google (Scholar, Patents, Web); PatBase(All).
Terms used: determine depth light field fuse low resolution images form higher compare hypothesized select most similar forward mapped camera characteristics pixel stack common image "reference viewpoint" etc.

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---------- Y	US 2004/0056966 A1 (Schechner et al.), 25 March 2004 (25.03.2004), entire document, especially Abstract, para [0022]-[0023], [0078]-[0081], [0085], [0094], [0099], [0107], [0113], [0121], [0139], [0143], [0154], [0157]-[0158], [0169], [0178]-[0179], [0186]-[0187], [0190], [0192], [0200], [0202], [0208], [0225], [0233], [0237], [0252], [0254]	1-30, 32-35, 37-39 ---------------------- 31, 36
Y	US 2011/0043661 A1 (Podoleanu), 24 February 2011 (24.02.2011), entire document, especially Abstract, para [0036]	31, 36
A	US 2004/0114807 A1 (Lelescu et al.), 17 June 2004 (17.06.2004), entire document	1-39
A	US 2005/0151759 A1 (Gonzalez-Banos et al.), 14 July 2005 (14.07.2005), entire document	1-39
A	US 2007/0268374 A1 (Robinson), 22 November 2007 (22.11.2007), entire document	1-39
A	US 2010/0208100 A9 (Olsen et al.), 19 August 2010 (19.08.2010), entire document	1-39
A	Capel. "Image Mosaicing and Super-resolution" [online]. Dated 2001. Retrieved on 10 November 2012 (10.11.2012). Retrieved from the Internet at URL:<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.226.2643&rep=rep1&type=pdf> , entire document	1-39
A	Fanaswala. "Regularized Super-Resolution of Multi-View Images" [online]. Dated August 2009. Retrieved on 10 November 2012 (10.11.2012). Retrieved from the Internet at URL:<http://www.site.uottawa.ca/~edubois/theses/Fanaswala_thesis.pdf>, entire document	1-39

☐ Further documents are listed in the continuation of Box C. ☐

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

Date of the actual completion of the international search	Date of mailing of the international search report
10 November 2012 (10.11.2012)	2 0 NOV 2012

Name and mailing address of the ISA/US	Authorized officer:
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450	Lee W. Young
Facsimile No. 571-273-3201	PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)