

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2013266488 B2**

(54) Title
System and method for managing spatiotemporal uncertainty

(51) International Patent Classification(s)
G06K 9/00 (2006.01) **H04N 7/16** (2006.01)
H04N 7/12 (2006.01)

(21) Application No: **2013266488** (22) Date of Filing: **2013.05.21**

(87) WIPO No: **WO13/177125**

(30) Priority Data

(31)	Number	(32)	Date	(33)	Country
	13/476,712		2012.05.21		US

(43) Publication Date: **2013.11.28**

(44) Accepted Journal Date: **2018.08.16**

(71) Applicant(s)
Help Lightning, Inc.;UAB Research Foundation

(72) Inventor(s)
Dillavou, Marcus W.;Shum, Phillip Corey;Guthrie, Barton L.;Shenai, Mahesh B.;Deaton, Drew Steven;May, Matthew Benton

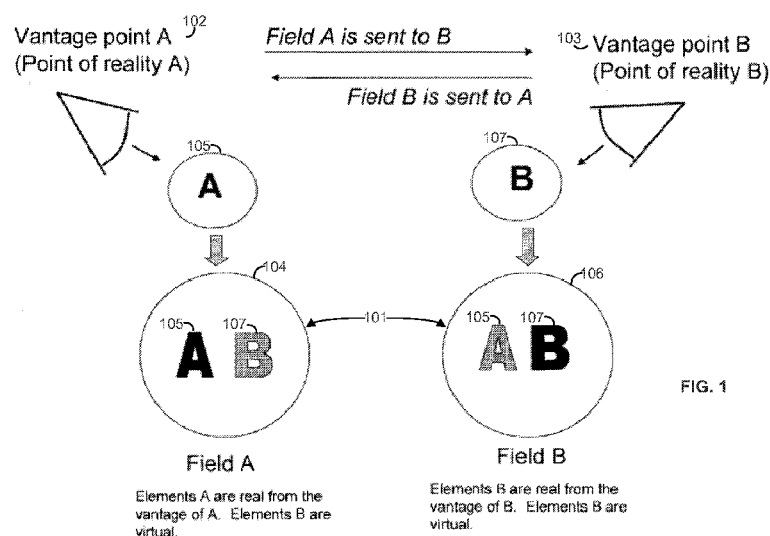
(74) Agent / Attorney
Phillips Ormonde Fitzpatrick, L 16 333 Collins St, Melbourne, VIC, 3000, AU

(56) Related Art
YONEKURA T ET AL, "Peer-to-Peer Networked Field-type Virtual Environment by Using AtoZ", CYBERWORLDS, 2004 INTERNATIONAL CONFERENCE ON TOKYO, 2004,IEEE, (20041118), doi:10.1109/CW.2004.49, ISBN 978-0-7695-2140-4, pages 241 - 248
CHONG N Y ET AL, "A collaborative multi-site teleoperation over an ISDN", MECHATRONICS, PERGAMON PRESS, OXFORD, GB, (20031001), vol. 13, no. 8-9, doi:10.1016/S0957-4158(03)00010-2, ISSN 0957-4158, pages 957 - 979



- (51) **International Patent Classification:**
H04N 7/12 (2006.01) H04N 7/16 (2011.01)
- (21) **International Application Number:**
PCT/US2013/041967
- (22) **International Filing Date:**
21 May 2013 (21.05.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
13/476,712 21 May 2012 (21.05.2012) US
- (71) **Applicants:** VIPAAR, LLC [US/US]; 1500 First Avenue North, Birmingham, AL 35203 (US). UAB RESEARCH FOUNDATION [US/US]; 1120G Administration Building, 1530 3rd Avenue South, Birmingham, AL 35294-0111 (US).
- (72) **Inventors:** DILLAVOU, Marcus, W.; 2301 1st Avenue N, Apt. 202, Birmingham, AL 35203 (US). SHUM, Philip, Corey; 423 33rd St S, Birmingham, AL 35205 (US). GUTHRIE, Barton, L.; 2106 Flowerwood Drive, Birmingham, AL 35244 (US). SHENAI, Mahesh, B.; 518 Bristol Lane, Birmingham, AL 35226 (US). DEATON, Drew, Steven; 3016 Woodbridge Road, Birmingham, AL 35203 (US). MAY, Matthew, Benton; 2928 Pawnee Avenue S, Apt. 3, Birmingham, AL 35205 (US).
- (74) **Agents:** RIESEN, Michael, J. et al.; Ballard Spahr LLP, 999 Peachtree Street, Suite 1000, Atlanta, GA 30309 (US).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— with international search report (Art. 21(3))

[Continued on next page]

(54) **Title:** SYSTEM AND METHOD FOR MANAGING SPATIOTEMPORAL UNCERTAINTY

(57) **Abstract:** Provided herein are methods and systems for managing spatiotemporal uncertainty in image processing. A method can comprise determining motion from a first image to a second image, determining a latency value, determining a precision value, generating an uncertainty element based upon the motion, the latency value, and the precision value, and rendering the uncertainty element.



(88) Date of publication of the international search report:

27 February 2014

SYSTEM AND METHOD FOR MANAGING SPATIOTEMPORAL UNCERTAINTY

BACKGROUND

- [0001] Spatiotemporal misalignments are often due to latency between the capture of an image at a remote source and the display of the captured image locally. Spatiotemporal misalignment or uncertainty may lead to spatial inaccuracies that compromise the fidelity of the composite image visualized by either participant in a real-time, shared first-person perspective session. In scenarios with significant motion and latency, for example, spatiotemporal uncertainty can be significant. While this uncertainty may not be very meaningful in low-risk contexts where pixel-level accuracy is unimportant, in a technical field such as microsurgery the accurate quantification of this error is crucial. It is therefore desirable to develop a system and method for managing the quantification and visualization of spatiotemporal uncertainty.

SUMMARY

- [0002] Disclosed are systems and methods for managing spatiotemporal uncertainty. Disclosed are systems and methods for managing the quantification and visualization of spatiotemporal uncertainty. In the domain of real-time, dual first-person perspective videoconferencing, spatiotemporal uncertainty may arise due to network and processing-based latency. The determination and visualization of spatiotemporal uncertainty has the potential to mitigate some of the risk of performing precise tasks in contexts with latency.
- [0003] Broadly, a method can comprise determining motion from a first frame to a second frame, determining an uncertainty factor based on determined motion, and generating an uncertainty element based on the uncertainty factor.
- [0004] Another method can comprise determining motion from a first image to a second image, determining a latency value, determining a precision value, and generating an uncertainty element based upon the motion, the latency value, and the precision value.
- [0005] Another method can comprise determining a latency value, determining a latency offset for a first image based upon the latency value, locating a second image based upon the latency offset from the first frame, and generating a motion ghost

overlaying the first frame, wherein the motion ghost comprises at least a portion of the second frame.

[0005a] According to one aspect of the invention, there is provided a method for providing a visualization of uncertainty in a collaborative video environment, the method comprising: determining motion of an element in a video sequence from a first image to a second image in the video sequence; generating a motion vector directed from the first image to the second image, wherein the motion vector is generated based upon the determined motion; determining a latency value comprising a network latency iteratively sampled over a pre-defined time period and a local latency indicating a difference in time from capture of an image to display of the image via a local processor; determining a precision value comprising a scalar value configured to be multiplied by the generated motion vector or a transformed motion vector, wherein the precision value comprises a pre-determined acceptable error or a user-provided acceptable error, and is dependent on context; generating an uncertainty element based upon the motion vector or the transformed motion vector, the latency value, and the precision value, involving multiplying the precision value by the generated motion vector or the transformed motion vector, the uncertainty element comprising an uncertainty region of pixels projecting from one or both of the motion vector and the transformed motion vector, wherein a size of the uncertainty region is proportional to a magnitude of the motion vector or the transformed motion vector; and rendering, to at least one of a plurality of collaborating users, the uncertainty element as a visualization comprising one or both of blur /or and colorization of pixels.

[0005b] There is also disclosed herein a method comprising: determining, by a computing device, motion from a first image to a second image; determining, by a computing device, a latency value; determining, by a computing device, a precision value; generating, by a computing device, an uncertainty element based upon the motion, the latency value, and the precision value, wherein the uncertainty element is generated based upon a transformed motion vector; and causing, by a computing device, rendering of the uncertainty element.

[0006] Additional advantages will be set forth in part in the description which follows or may be learned by practice. The advantages will be realized and attained by means of the elements and combinations particularly pointed out in the appended inventive concepts. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not to be considered restrictive.

[0006a] Where any or all of the terms "comprise", "comprises", "comprised" or "comprising" are used in this specification (including the claims) they are to be interpreted as specifying the presence of the stated features, integers, steps or components, but not precluding the presence of one or more other features, integers, steps or components.

[0006b] A reference herein to a patent document or any other matter identified as prior art, is not to be taken as an admission that the document or other matter was known or that the information it contains was part of the common general knowledge as at the priority date of any of the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments and together with the description, serve to explain the principles of the methods and systems provided:

Figure 1 illustrates virtual interactive presence; Figure 2A illustrates virtual interactive presence;

Figure 2B illustrates a local expert assisting a remote user;

Figure 3 illustrates an exemplary system;

Figure 4A illustrates an exemplary method;

Figure 4B illustrates an exemplary method;

Figure 4C illustrates an exemplary method;

Figure 4D illustrates an exemplary method;

Figure 4E illustrates an exemplary method;

Figure 4F illustrates an exemplary method;

Figure 4G illustrates an exemplary block diagram representing frame latency;

Figure 4H illustrates an exemplary graph of opacity;

Figure 4I illustrates a visualization of an uncertainty;

Figure 4J illustrates a visualization of an uncertainty;

Figure 5 illustrates an exemplary virtual presence system;

Figure 6 illustrates exemplary processes performed within a graphics server;

Figure 7 illustrates exemplary processes performed within a network server;

Figure 8 illustrates a side view of an exemplary VIP display;

Figure 9 illustrates a user's view of an exemplary VIP display;

Figure 10 illustrates a user's view of an exemplary VIP display;
Figure 11 illustrates an exemplary method;
Figure 12 illustrates another exemplary method;
Figure 13 illustrates virtual presence in a remote surgical environment;
Figure 14 illustrates merging of medical imaging with an operative field; and
Figure 15 illustrates an exemplary operational environment.

DETAILED DESCRIPTION

- [0008] Before the present methods and systems are disclosed and described, it is to be understood that the methods and systems are not limited to specific synthetic methods, specific components, or to particular compositions, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.
- [0009] As used in the specification and the appended inventive concepts, the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise.
- [0010] Ranges may be expressed herein as from “about” one particular value, and/or to “about” another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent “about,” it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint.
- [0011] “Optional” or “optionally” means that the subsequently described event or circumstance may or may not occur, and that the description includes instances where said event or circumstance occurs and instances where it does not.
- [0012] Throughout the description and claims of this specification, the word “comprise” and variations of the word, such as “comprising” and “comprises,” means “including but not limited to,” and is not intended to exclude, for example, other additives, components, integers or steps. “Exemplary” means “an example of” and is not intended to convey an indication of a preferred or ideal embodiment.

[0013] Disclosed are components that can be used to perform the disclosed methods and systems. These and other components are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these components are disclosed that while specific reference of each various individual and collective combinations and permutation of these may not be explicitly disclosed, each is specifically contemplated and described herein, for all methods and systems. This applies to all aspects of this application including, but not limited to, steps in disclosed methods. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific embodiment or combination of embodiments of the disclosed methods.

[0014] The present methods and systems may be understood more readily by reference to the following detailed description of preferred embodiments and the Examples included therein and to the Figures and their previous and following description.

[0015] Disclosed are methods and systems for managing spatiotemporal uncertainty. The disclosed methods and systems can utilize virtual reality. Virtual reality (VR) refers to a computer-based application which provides a human-computer interface such that the computer and its devices create a sensory environment which is dynamically controlled by the actions of the individual, so that the environment appears “real” to the user. With VR, there is communication between a computer system and a user. The computer creates a sensory environment for the user to experience which may be, in one aspect, multisensory (although this is not essential) and the computer creates a sense of reality in response to user inputs.

[0016] In one exemplary aspect, the system disclosed can utilize at least two types of VR, Immersive and Non-immersive. Immersive VR creates the illusion that the user is actually in a different environment. In one aspect, the system accomplishes this through the use of such devices as Head Mounted Displays (HMD's), earphones, and input devices such as gloves or wands. In another aspect, in order to enhance to realism of the experience, a plurality of Degrees of Freedom (DOF's) are utilized, which the software can simulate. Generally, the more the DOF's, the better the realism of the experience. Exemplary DOF's include, without limitation: X,Y,Z, roll, pitch, and yaw.

[0017] Non-immersive VR creates an environment that is differentiable from the user's surrounding environment. It does not give the illusion that the user is

transported to another world. Non-immersive VR works by creating a 3-dimensional image and surround sound through the use of stereo projection systems, computer monitors, and/or stereo speakers. Non-immersive VR can be run from a personal computer without added hardware.

[0018] In one aspect, movement in Immersive VR can be realized by a system through the use of optical, acoustical, magnetic, or mechanical hardware called trackers. Preferably, the input devices have as many of these trackers as possible, so that movement can be more accurately represented. For instance, virtual gloves can have up to 3 trackers for each index, and more for the palm and wrist, so that the user can grab and press objects. In one aspect, the trackers can be equipped with positioning sensors, that tell a computer which direction the input is facing and how the input device is tilted in all directions. This gives a sensor with six degrees of freedom.

[0019] In another aspect, the system disclosed can utilize augmented reality (AR). AR can refer to a computer-based application which provides a human-computer interface such that the computer and its devices create an altered experience for the user through the inclusion of elements both real and virtual. An example of augmented reality that can be used in the present systems and methods includes, without limitation, the superimposition of computed tomography (CT) or magnetic resonance (MRI) data onto an image of a patient. As another example, the use of AR can include the superimposition of patient biometric data on an image of a surgical field. Other fields can make use of the disclosed systems and methods.

[0020] Output devices bring the user to the virtual world. An example of an output device that can be used in the present system include, without limitation, head mounted displays (HMD) in the form of glasses or goggles, which allow a user to wear a display system on their head. One approach to the HMD is to use a single Liquid Crystal Display (LCD), wide enough to cover both eyes. Another approach is to have two separated displays – one for each eye. This takes somewhat more computer power, since the images displayed are different. Each display has a separate image rendered from the correct angle in the environment. Eye-tracking can be combined with HMDs. This can allow, for example, surgeons to move their eyes to the part of an image they want to enhance.

[0021] Another example of an output device that can be used in an embodiment of the present system is shuttered glasses. This device updates an image to each eye

every other frame, with the shutter closed on the other eye. Shuttered glasses require a very high frame rate in order to keep the images from flickering. This device is used for stereo monitors, and gives an accurate 3-d representation of a 2-d object, but does not immerse the user in the virtual world.

[0022] Another output device that can be used in an embodiment of the present system is a screen with multiple projectors. The screen can be either a plane or bent. A challenge when using multiple projectors on the same screen is that there can be visible edges between the projections. This can be remedied by using a soft-edge system wherein the projection goes more and more transparent at the edges and the projections overlap. This produces an almost perfect transition between the images. In order to achieve a desired 3D effect, shuttered glasses can be used. Special glasses can be used, that alternate between making the glass either completely opaque or completely transparent. When the left eye is opaque, the right one is transparent. This is synchronized to the projectors that are projecting corresponding images on the screen.

[0023] In another aspect, a Cave Automatic Virtual Environment (CAVE) can also be used in the present system. A CAVE can use mirrors in a cube-shaped room to project stereo images onto the walls, giving the illusion that you are standing in a virtual world. The world is constantly updated using trackers, and the user is allowed to move around almost completely uninhibited.

[0024] Disclosed are methods and systems for image registration. Such methods and systems can render a number of elements/participants virtually present into a field of interest in a manner such that the users can interact for any given purpose, such as the delivery of remote expertise. A field of interest can comprise varying amounts of “real” and “virtual” elements, depending on a point of view. Elements can include any “real” or “virtual” object, subject, participant, or image representation. Various components of the disclosed methods and systems are illustrated in **FIG. 1**.

[0025] A common field of interest **101** can be a field within which elements are physically and/or virtually present. Point of Reality (or Point of View) can refer to the vantage of the element/participant that is experiencing the common field of interest. In **FIG. 1**, exemplary points of reality, or points of view, are shown at **102** and **103**, representing displays. The common field of interest **101** can appear similar from both vantages, or points of view, but each comprises differing combinations of local (physical) and remote (virtual) elements/participants.

[0026] Local elements can be elements and/or participants which are physically present in the common field of interest. In **FIG. 1**, element A **105** is a local element for field A **104** and is physically present in field A **104**. Element B **107** is a local element for field B **106** and is physically present in field B **106**. It is understood that virtual elements (not shown) can be inserted or overlaid in field A **104** and/or field B **106**, as desired.

[0027] Remote elements can be elements and/or participants that are not physically present in the common field of interest. They are experienced as “virtually present” from any other local vantage point. As shown in **FIG. 1**, element B **107** is a remote element to field A **104** and is virtually present in field A **104**. Element A **105** is a remote element in field B **106** and is virtually present in field B **106**.

[0028] Methods for rendering a virtual interactive presence by combining local and remote elements and/or participants can comprise one or more of the following steps. A common local field can be rendered in a manner that reflects the presence of the field, elements and/or participants. As shown in **FIG. 2A**, Participant A can experience real elements in field A through a viewer. The common local field can be rendered such that it is experienced remotely in a manner that enables remote participants to experience it similarly to the local persons. As shown in **FIG. 2A**, this is illustrated by Participant A experiencing element B as virtually present in field A.

[0029] Remote persons can insert themselves and/or interact with the virtual field as rendered to them. For example, Participant A can insert hands, instruments, etc. into field A and interact with the virtual element(s) B. Viewer B can view a ‘virtual compliment’ to this, with Viewer B’s real elements interacting with Participant A’s virtual elements.

[0030] The common local field can be continuously updated such that the presence of the remote participants can be rendered in real time. For example, the remote scene can be the most up-to-date available with the time lag between the remote capture and the local render kept as low as possible. Conversely, if there is a need to introduce a timing difference, this can be accomplished as well.

[0031] The common local field can be scaled to a size and depth to meaningfully match the local scene. And the common local field can be configurable, such that remote elements can be made more or less transparent, removed entirely, or otherwise altered to suit the needs of the local user.

[0032] Each field is captured by a digital camera. The resulting image is physically distorted from its reality, based upon the physical characteristics of the camera. A processor, therefore, receives and displays a “physically” distorted version of the local reality. Likewise, a digital camera also captures the remote field(s), but the incoming stream is relayed through a transmission device and across a network. The processor, therefore, receives the remote stream that contains both physical and transmission-based distortion. The processor must then apply a series of transformations that removes the physical and transmission-based distortion from the common local field.

[0033] The local participants can experience the virtually present participants in a manner that enables continuous interaction in the common local field. **FIG. 2B** illustrates a local expert assisting a remote user. The hands of the local expert **201** are slightly transparent and superimposed into the field that is viewed by the remote user. The remote user can view the local expert’s hands, the remote user’s hands and a puzzle located at the remote user’s location. The local expert is assisting the remote user in assembling a puzzle.

[0034] **FIG. 3** illustrates an exemplary image processing system **300**. As shown, the system **300** can comprise a first display **302** and a second display **304** configured for displaying one or more of an image, a video, a composite video/image, and a common field of interest, for example. However, it is understood that any number of displays can be included in the system **300**. In certain aspects, the second display **304** can be disposed remotely from the first display **302**. As an example, each of the first display **302** and the second display **304** can be configured to render the common field of interest thereon. As a further example, each of the first display **302** and the second display **304** can be configured to render at least one of the local field and the remote field thereon. In certain aspects, at least one of the first display **302** and the second display **304** can be a VIP display, as described in further detail herein. However, it is understood that each of the first display **302** and the second display **304** can be any type of display including a monoscopic display and a stereoscopic display, for example. It is understood that any number of any type of display can be used.

[0035] A first sensor **306** can be in signal communication with at least the first display **302** and can be configured for obtaining image data such as a virtual presence data, for example. In certain aspects, the first sensor **306** can be one or

more of a camera, an infrared sensor, a light sensor, a RADAR device, a SONAR device, a depth scan sensor, and the like. It is understood that the first sensor **306** can be any device or system capable of capturing/obtaining an image data representative of at least one of a “real” element” and a “virtual” element.

[0036] A second sensor **308** can be in signal communication with at least the second display **304** and can be configured for obtaining image data such as virtual presence data, for example. In certain aspects, the second sensor **308** can be one or more of a camera, an infrared sensor, a light sensor, a RADAR device, a SONAR device, a depth scan sensor, and the like. It is understood that the second sensor **308** can be any device or system capable of capturing/obtaining an image data representative of at least one of a “real” element” and a “virtual” element. It is further understood that any number of sensors can be used.

[0037] A plurality of processors **310, 312** can be in direct or indirect signal communication with at least one of the first display **302**, the second display **304**, the first sensor **306**, and the second sensor **308**. Each of the processors **310, 312** can be configured to render the image data collected by the sensors **306, 308** onto at least one of the displays **302, 304**. It is understood that the processors **310, 312** can be configured to modify the image data and the resultant image for transmission and display. It is further understood that any number of processors can be used, including one. In certain aspects, the system **300** comprises only the processor **310, 312** in data communication with each other.

[0038] In certain aspects, each of the displays **302, 304** can comprise an associated one of the processors **310, 312** for rendering images onto the displays **302, 304**. Each of the processors **310, 312**, or another system comprising a processor, can communicate with each other through a network connection. For example, remote sites can connect via the Internet or other network. Tasks can be divided amongst each of the processors **310, 312**. For example, one of the processors **310, 312** can be configured as a graphics processor or graphics server and can gather images from one of the sensors **306, 308** and/or a network server, perform an image composition tasks, and drive one or more of the displays **302, 304**.

[0039] In an aspect, one or more of the processors **310, 312** can be configured to render an image. As an example, one or more of the processors **310, 312** can be configured to render a common field of interest that reflects a presence of a plurality of elements based upon the image data obtained by at least one of the sensors **306,**

308. As a further example, at least one of the elements rendered in the common field of interest can be a remote element physically located remotely from another of the elements. The processors **310, 312** can also be configured to render/output the common field of interest to at least one of the displays **302, 304**. As an example, the processors **310, 312** can render interaction between a remote user and a local user in the common field of interest. As a further example the presence of the remote element can be rendered in real time to the local user and the presence of a local element can be rendered in real time to the remote user.

[0040] **FIG. 4A** illustrates exemplary process **400** that can be performed with at least one of the processors **310, 312**. Other processor and/or computing devices can be used to perform the process **400**. In step **402**, motion can be determined. As an example, a plurality of images can be rendered as a sequence of frames. As a further example, motion of elements and/or subjects represented in the images can be determined from a first frame to a second frame. In an aspect, the first frame can be a current frame and the second frame can be a previous frame of a series or sequence of images.

[0041] In an aspect, motion represented by a plurality of images or frames can be determined (e.g., calculated, estimated, retrieved) using a motion estimation algorithm (e.g., block matching algorithm). Motion estimation algorithms can be used in streaming video applications to eliminate the transmission of redundant video data. As an example, an encoder can estimate the motion in the current frame with respect to a previous reference frame by examining the similarity between pixel regions in a current image with pixel regions in a previous reference image. As another example, phase correlation, optical flow, and Bayesian estimator techniques can be utilized to perform motion estimation. As a further example, statistical functions can be employed to perform motion estimation.

[0042] In step **404**, an uncertainty factor can be determined. In an aspect, the uncertainty factor can be based on the motion determined in step **402**. As an example, the uncertainty factor can be a value, measurement, extrapolation, data point, or other aspect that can affect an uncertainty of position of any pixel, object, subject, element, or the like. As a further example, the uncertainty factor can comprise one or more of a network latency, jitter, a local latency, a processing latency, a frame latency, an acceptable error, and a tolerance.

[0043] In step **406**, an uncertainty element can be generated. In an aspect, the uncertainty element can be generated based on one or more of the motion determined in step **402** and the uncertainty factor(s) determined in step **404**. As an example, the uncertainty element can be rendered (e.g., transmitted, presented, etc.) to a device and/or user. As a further example, the uncertainty element can be rendered as an audio, visual, and/or tactile feedback to a user.

[0044] As an example, **FIG. 4B** illustrates an exemplary process for motion estimation. Other motion estimation techniques and/or algorithms can be used. In step **408**, an image region can be determined. As an example, the image region can be a macroblock with standard dimensions set forth in a video compression standard such as H.263 or H.264. As another example, an image region can be determined by selecting an arbitrarily sized area in an image (e.g., a first image or frame). As a further example, the image region can be determined with user-provided input.

[0045] In step **410**, a step size can be determined. As an example, the step size can be a numerical value dependent on context of the content of one or more images. As a further example, the larger the step size, the more accurate the motion estimation. In step **412**, the image region in the first frame can be compared to an image region in a second frame. As an example, an image region in a second frame that is correspondingly located to the image region in the first frame can be tested for comparison with the image region in the first frame. As another example, eight image regions in a second frame that are a distance equal to the step size from the image region in the second frame can be tested for comparison with the image region in the first frame.

[0046] In an aspect, a cost function can be employed to perform a comparison between the luminance of image regions. As another example, a cost function can be utilized to perform a comparison between the color of image regions. In an exemplary aspect, the cost function can be the mean absolute difference (MAD). In another aspect, the cost function can be the mean squared difference (MSD). In a further aspect, the cost function can be a cross-correlation function. In step **414**, the cost function can be minimized. As an example, the minimization of a cost function can comprise selecting a new center point for comparison.

[0047] In step **416**, the step size can be reduced. As an example, the step size can be reduced by subtracting one from the step size. As another example, the step size can be reduced by halving the step size. In step **418**, the step size can be checked with

conditional logic (e.g., IF/THEN statements). In an exemplary aspect, if the step size is larger than one, then the process returns to step **410**; if not, the process moves forward to step **420**.

[0048] In step **420**, a motion vector can be generated. In an aspect, the motion vector can allow a representation of the estimated motion of image regions across multiple frames or images. As an example, a motion vector can be generated from the image region in a current frame to a center point of an image region in a previous frame with minimum distortion as determined by the output of a cost function. As another example, a motion vector can be generated from the image region in a previous frame to the center point of an image region in a current frame with minimum distortion as determined by the output of a cost function. As a further example, a motion vector can be generated for each image region in an image or portion of an image, which in aggregate can be referred to as a vector map. In an aspect, the motion vector can be transformed (e.g., inverted, divided, multiplied) and/or applied to a visualization technique in order to render an uncertainty due to the motion. Various techniques can be used to process the motion vector to estimate and/or visualize the motion and uncertainty represented by the motion vector.

[0049] Returning to **FIG. 4A**, in step **404**, an uncertainty factor can be determined. In an aspect, the uncertainty factor can be based on the motion determined in step **402**. As an example, the uncertainty factor can be a value, measurement, extrapolation, data point, or other aspect that can affect an uncertainty of position of any pixel, object, subject, element, or the like. As a further example, the uncertainty factor can comprise one or more of a network latency, jitter, a local latency, a processing latency, a frame latency, an acceptable error, and a tolerance.

[0050] As an illustrative example, **FIG. 4C** shows a method for determining an uncertainty factor. In step **426**, network latency can be determined. As an example, a network latency value can be retrieved (e.g., sampled) from an external device (e.g., hardware codec). As another example, network latency can be sampled from a software program. As a further example, network latency can be iteratively sampled based upon a pre-defined time period (e.g., one second). As an additional example, network latency can be retrieved by taking a rolling average of observed latencies over a time period. In an aspect, the network latency can comprise jitter.

[0051] In step **428**, local latency (e.g., capture, process, draw cycle) can be determined. In an aspect, one or more images (e.g., frames of a plurality of images)

can be time stamped from moment captured to moment displayed. In another aspect, local latency can be iteratively sampled from the processors **310, 312 (FIG. 3)**. Accordingly, a difference from capture to display of one or more frames and/or images can be defined as the local latency.

[0052] In step **430**, the uncertainty factor can be determined based upon one or more of the network latency determined in step **426** and the local latency determined in step **428**. As an example, the uncertainty factor can comprise a cumulative or total latency comprising one or more of the network latency and local latency. As a further example, the uncertainty factor can comprise a worst-case latency value determined by summing latency and/or average latency and/or jitter to one standard deviation above the mean.

[0053] As an illustrative example, **FIG. 4D** shows a method for determining an uncertainty factor. In step **432**, a precision value can be determined. In an aspect, the precision value can comprise a pre-determined acceptable error. As an example, the precision value can comprise a user-provided acceptable error (e.g., number of pixels). The precision value can be dependent on context, wherein the lower the precision value, the less accurate the rendered images. In an aspect, the precision value is a scalar quantity that can be multiplied by one or more motion vectors in a vector map. In another aspect, the user-provided acceptable error is a scalar quantity that can be multiplied by one or more motion vectors in a vector map. In step **434**, the uncertainty factor can be determined based upon one or more of the precision value, the network latency, jitter and the local latency.

[0054] Returning to **FIG. 4A**, in step **406**, an uncertainty element can be generated. In an aspect, the uncertainty element can be generated based on one or more of the motion determined in step **402** and the uncertainty factor(s) determined in step **404**.

[0055] As an illustrative example, **FIG. 4E** shows a method for generating an uncertainty element. In step **436**, an uncertainty region (e.g., plurality of pixels, region of pixels surrounding at least a portion of a periphery of an element, object, subject in an image) can be generated. As an example, the uncertainty element can be visualized as an uncertainty region. In an aspect, the uncertainty region can be one or more pixels substantially projecting from one or more of a motion vector and a transformed motion vector (e.g., inverse of a motion vector generated at step **420**). As an example, a size of the uncertainty region can be based upon the latency value. As a further example, a size of the uncertainty region is proportional to a magnitude

of the vector. As a further example, a size of the uncertainty region is proportional to the precision value. As an additional example, a size of the uncertainty region is proportional to the acceptable error.

[0056] In an aspect, the uncertainty element can comprise an indicator or graphic that communicates the level of uncertainty to a user. As an example, the indicator can be a meter, gauge, or scale, a color spectrum, and/or a quantitative representation of the uncertainty (e.g., uncertainty factor) and/or latency. As another example, the indicator can be rendered as a visualization when other visualization methods are not utilized. As a further example, the indicator can represent an uncertainty in an image as a reactionary (e.g., uncertainty of past position or and/or motion) or predictive element (e.g., uncertainty of future position or and/or motion). In an aspect, the indicator can reflect variation over time in the summed magnitude of one or more vectors scaled by an uncertainty factor (e.g., derived from step 404) in a vector map. In another aspect, the indicator can reflect variation in an aggregate sum of motion for one or more frames or images.

[0057] In step 438, a visualization can be rendered. In an aspect, the visualization can represent the uncertainty element. As an example, the visualization can comprise a colorization of pixels, an icon, an indicator, and/or a blurring of pixels. As another example, software such as OpenGL can be utilized to perform the visualization. As a further example, the visualization can be rendered in a portion (e.g., polygon) of a frame or image.

[0058] As an illustrative example, FIG. 4F shows a method for determining an uncertainty factor. In step 440, a latency value can be determined. As an example, the latency value can comprise one or more of network latency, local latency, and jitter. As a further example, a network latency value can be retrieved (e.g., sampled) from an external device or program (e.g., codec). As a further example, network latency can be iteratively sampled based upon a pre-defined time period (e.g., one second). In an aspect, network latency can be retrieved by taking a rolling average of observed latencies over a time period. In an aspect, the network latency can comprise jitter. In an aspect, local latency (e.g., capture, process, draw cycle) can be determined. In an aspect, one or more images (e.g., frames of a plurality of images) can be time stamped from moment captured to moment displayed.

[0059] In step 442, frame offset or latency offset can be determined. As an example, the frame latency can be defined by a frame offset based upon one or more of the

local latency and the network latency. As a further example, a frame offset can be determined by multiplying the total latency by a rendering rate (e.g., frames per second) of the rendered images. In step **444**, a frame or image can be located. As an illustrative example, **FIG. 4G** illustrates a series or sequences of frames (e.g., as stored in system memory **1512**). Typically, frames are processed with respect to a timeline **R**. As an example, a plurality of images can be rendered as a sequence of frames **X, Y, Z, U**. As a further example, each of the frames can comprise an identifier such as time stamps **A, B, C, D** and sequential frame identifiers **E, F, G, H**. In an aspect, a first frame can be a current frame **U** and a second frame **X** can be a previous frame of a series or sequence of images. In an aspect, a frame offset **L** can be used to locate a previous image **X** with respect to a currently rendered image **U** by subtracting the frame offset from the frame identifier **H** of frame **U**. In another aspect, a frame latency **N** can be determined by subtracting the timestamp of the image **X** from the timestamp of the currently rendered image **U** (e.g., frame).

[0060] Returning to **FIG. 4F**, in step **446**, an uncertainty element can be generated. In an aspect, the uncertainty element can comprise a rendering of the frame located at step **444**. As an example, the located frame can be merged with a local image. In an aspect, the uncertainty element can comprise an overlay or a ghosted image of the located frame. As an example, a ghosted image can be a processed image merged with a local image.

[0061] In an aspect, an opacity of the ghosted image can be variable. For example, the opacity of the ghosted image can change based upon the graph shown in **FIG. 4H**. As shown in **FIG. 4H**, the opacity of a currently rendered image frame **U** may be set at 1. The opacity of previous frames **Y, Z** may be decreased linearly, and then increased toward a located frame **X**. However, the opacity of the ghosted image can be changed based upon any function or along any plot.

[0062] As an illustrative example, a visualization can represent a user moving a pen or other object or tool, as illustrated in **FIG. 4I**. Accordingly, a currently rendered frame **448** can be ghosted with motion artifacts from a previous frame **448'**, effectively allowing the user to visualize the frames currently being seen and reacted to by a remote user. As a further example, a sequence of frames between the currently rendered frame **448** and a previous frame **448'** can be ghosted. As shown in the previous **448'**, a user's right hand **450** with pen **452** plans an incision on a surgical field **454**. A remote user **456** provides instruction. In the currently rendered

frame 448, as a user's right hand 450 with pen 452 goes into motion, motion artifacts 458, 459 allow visualization of the frames currently being seen by a remote user. This can alert the user of the disparity between what is being reacted to and what is currently happening. In an aspect, the visualization can include a blurring of the pixels representing the object of uncertainty or an area near the object of uncertainty. As an example, a colorization or alert color can be used to visually represent that the image being rendered comprises some level of uncertainty.

[0063] As an illustrative example, FIG. 4J shows a bidirectional capture, process, and display of images onto displays (e.g., displays 302, 304 (FIG. 3)). Typically, frames are processed with respect to a timeline 460. A first frame 462 is shown with an element 464 (e.g., captured by sensor 306 and visualized on display 302). A second frame 466 is shown with an element 468 (e.g., captured by sensor 308 and visualized on display 304). As the element 464 moves along a downward vertical motion, a ghost 469 (e.g., derived from step 446) of the first frame 462 can be visualized in a third frame 462' (e.g., on display 302). Additionally, element 468 can be rendered along with an uncertainty region 470 (e.g., derived from step 436).

[0064] On one display (e.g., display 304), the element 468 is shown in the second frame 466. As the element 468 moves leftward, a ghost 472 (e.g., derived from step 446) of the second frame 466 can be visualized on a fourth frame 466' (e.g., on display 304). Additionally, the element 464 can be rendered in the fourth frame 466' along with an uncertainty region 474 (e.g., derived from step 436).

[0065] In an aspect, the example illustrated in FIG. 4J can allow a viewer of displays 302, 304 to visualize the local images captured by sensors 306, 308 that are being reacted to by a remote user. Further, the example illustrated in FIG. 4J can allow a viewer of displays 302, 304 to view an estimation of the current image being captured by a remote sensor 306, 308. In this manner, the effects of local and network latency can be decreased from the standpoint of a viewer.

[0066] FIG. 5 illustrates an exemplary virtual presence system. One such system can be used by each remote participant that is to join the same session. Each system can communicate with each other through a network connection. For example, remote sites can connect via the internet. Tasks can be divided amongst a plurality of computers in each system. For example, one computer (a graphics server) can gather images from local cameras and a network server, perform the stereo image

composition tasks, and drive a local stereoscopic display system. As a further example, the processor(s) **310** of system **300** can be embodied by the graphics server.

[0067] **FIG. 6** illustrates exemplary processes that can be performed with the graphics server. Images can be gathered into local data structures (frame rings). Local images can be gathered from a plurality of cameras, for example two cameras. Remote images can be provided by the network server via a high-speed remote direct memory access (RDMA) connection, for example. These images can be combined so that the remote user and the local user can be seen in the same scene (as in **FIG. 3**). This composite result can be transmitted to a local stereoscopic display system. A second computer can act as the network server, which can perform network encoding/decoding tasks as well as depth map generation, for example.

[0068] **FIG. 7** illustrates exemplary processes that can be performed with the network server. Local images gathered from the graphics server via the RDMA connection can be analyzed and mapped with depth information, encoded for efficient network transmission, and sent to an external network connection to be received by a corresponding network server at the remote site. Simultaneously, encoded images and depth maps can be received from the remote site, decoded, and provided to the local graphics server via the RDMA connection.

[0069] The system can be user-controlled by a control terminal connected to the network server; the user can then access and control the graphics server via the dedicated network connection to the network server.

[0070] Parameters of virtual interactive presence can be configured depending on the system used. Configurable parameters include, but are not limited to, size of virtual elements, presence of virtual elements (opaque, translucent, etc.), time of virtual presence (time can be configured to be delayed, slowed, increased, etc.), superimposition of elements such that any combination of virtual and real can be superimposed and/or 'fitted' over one another, and the like.

[0071] **FIG. 8** illustrates a side view of an exemplary VIP display. **FIG. 9** illustrates a user's view of an exemplary VIP display. **FIG. 10** illustrates a user's view of an exemplary VIP display.

[0072] As used herein, a "local" field of interest can refer to a local physical field and local user, thus making every other field remote. Each field can be local to its local physical user, but remote to other users. The composite of the fields can be a common field of interest. This is distinct from common "virtual worlds" in that there

can be components of “real” within the local rendering of the common field of interest and interactions can be between actual video (and other) renderings of physical objects and not just graphic avatars representing users and objects. The methods and systems provided allow for virtual interactive presence to modify/optimize a physical domain by the interplay of real and virtual.

[0073] In an aspect, illustrated in **FIG. 11**, provided are methods for virtual interactive presence comprising rendering a common field of interest that reflects the physical presence of a remote user and a local user at **1101**, rendering interaction between the remote user and the local user in the common field of interest at **1102**, and continuously updating the common field of interest such that the presence of the remote user is rendered in real time to the local user and the presence of the local user is rendered in real time to the remote user at **1103**.

[0074] The common field of interest can be rendered such that the remote user experiences the common field of interest similarly to the local user. The local user can experience the remote user’s physical presence in a manner that enables continuous interaction in the common field of interest with the remote user. The methods can further comprise rendering the physical presence of a local object in the common field and rendering interaction between the local user and the local object in the common field. The methods can further comprise rendering the physical presence of a local object in the common field of interest and rendering interaction between the remote user and the local object in the common field of interest.

[0075] In another aspect, illustrated in **FIG. 12**, provided are methods for virtual interactive presence comprising rendering a local field of interest that reflects the physical presence of a local object, a volumetric image of the local object, and a local user at **1201**, rendering interaction between the local object, the volumetric image, and the local user in the local field of interest at **1202**, and continuously updating the local field of interest such that the presence of the local object and the volumetric image of the local object is rendered in real time to the local user at **1203**.

[0076] The local object can be, for example, a patient and the volumetric image of the local object can be, for example, a medical image of a part of the patient. However, the local object can be any object of interest and the image of the local object can be any accurate rendering of that object. For example, could be an automobile engine and a 3D graphic of the engine, etc.

- [0077] The medical image can be, for example, one of, an x-ray image, an MRI image, or a CT image. The methods can further comprise superimposing, by the local user, the volumetric image onto the local object. The superimposition can be performed automatically by a computer.
- [0078] The methods can further comprise adjusting, by the local user, a property of the volumetric image. The property can be one or more of transparency, spatial location, and scale.
- [0079] The methods can further comprise rendering a local tool in the local field of interest. The methods can further comprise rendering the local tool in accurate spatial relation to the rendering of the local object. The tool can be any type of tool, for example, a surgical tool.
- [0080] In another aspect, provided are systems for virtual presence, comprising a virtual presence display, configured for displaying a common field of interest, a local sensor, configured for obtaining local virtual presence data, a network interface, configured for transmitting local virtual presence data and receiving remote virtual presence data, and a processor, coupled to the virtual presence display, the local sensor, and the network interface, wherein the processor is configured to perform steps comprising, rendering a common field of interest that reflects the physical presence of a remote user and a local user based on the local virtual presence data and the remote virtual presence data, rendering interaction between the remote user and the local user in the common field of interest, continuously updating the common field of interest such that the presence of the remote user is rendered in real time to the local user and the presence of the local user is rendered in real time to the remote user, and outputting the common field of interest to the virtual presence display.
- [0081] The virtual presence display can be one or more of a stereoscopic display, a monoscopic display (such as a CRT, LCD, etc.), and the like. The sensor can be one or more of a camera, an infrared sensor, a depth scan sensor, and the like. The common field of interest can be rendered such that the remote user experiences the common field of interest similarly to the local user. The local user can experience the remote user's physical presence in a manner that enables continuous interaction in the common field of interest with the remote user.
- [0082] The processor can be further configured to perform steps comprising rendering the physical presence of a local object in the common field of interest and

rendering interaction between the local user and the local object in the common field of interest.

[0083] The processor can be further configured to perform steps comprising rendering the physical presence of a local object in the common field of interest and rendering interaction between the remote user and the local object in the common field of interest.

[0084] Further provided are systems for virtual presence, comprising a virtual presence display, configured for displaying a local field of interest, a local sensor, configured for obtaining local virtual presence data, a processor, coupled to the virtual presence display and the local sensor, wherein the processor is configured to perform steps comprising, rendering a local field of interest that reflects the physical presence of a local object and a local user based on the local virtual presence data and a volumetric image of the local object, rendering interaction between the local object, the volumetric image, and the local user in the local field of interest, continuously updating the local field of interest such that the presence of the local object and the volumetric image of the local object is rendered in real time to the local user, and outputting the local field of interest to the virtual presence display.

[0085] The virtual presence display can be one or more of a stereoscopic display, a monoscopic display (such as a CRT, LCD, etc.), and the like. The sensor can be one or more of a camera, an infrared sensor, a depth scan sensor, and the like.

[0086] The local object can be, for example, a patient and the volumetric image of the local object can be, for example, a medical image of a part of the patient. The medical image can be, for example, one of, an x-ray image, an MRI image, or a CT image. However, the local object can be any object of interest and the image of the local object can be any accurate rendering of that object. For example, could be an automobile engine and a 3D graphic of the engine, etc.

[0087] The processor can be further configured to perform steps comprising superimposing, by the local user, the volumetric image onto the local object. The processor can be further configured to perform steps comprising adjusting, by the local user, a property of the volumetric image. The property can be one or more of transparency, spatial location, and scale.

[0088] The processor can be further configured to perform steps comprising rendering a local tool in the local field of interest. The processor can be further

configured to perform steps comprising rendering the local tool in accurate spatial relation to the rendered local object.

[0089] The disclosed methods and systems can have broad applications. For example, surgery, gaming, mechanics, munitions, battle field presence, instructional efforts (training) and/or any other situation where interaction is part of the scenario.

[0090] Also disclosed are methods and systems that enable a remote expert to be virtually present within a local surgical field. Virtual interactive presence can be used to enable two surgeons remote from each other to interactively perform a surgical procedure. The methods and system enable two or more operators to be virtually present, and interactive, within the same real operative field, thus supporting remote assistance and exporting surgical expertise.

[0091] The methods and systems can also be used to superimpose imaging data of the operative anatomy onto the anatomy itself for guidance and orientation (augmented reality). The methods and systems can be used for training of students. The methods and systems augment and enhance the field of robotics by virtually bringing an expert into the robotic field to guide the robot operator. The methods and systems are applicable to endoscopic procedures by inserting the expert's hands directly into the endoscopic field for guidance. The methods and systems expand remote surgery by providing the assistance of a remote expert to an actual local surgeon, whose basic skills can handle emergencies, and who will learn from the virtual interaction. The methods and systems can be used at trauma sites and other medical environments. The methods and systems can be used to provide remote assistance in other areas such as engineering, construction, architecture, and the like. The methods and systems disclosed can be used to transmit expertise to a remote 'site of need', merge contemporary imaging directly into the surgical field, and train surgical students

[0092] An exemplary remote surgical assistance system for transmitting surgical maneuvers of a local expert to a remote surgeon for the purpose of guiding/assisting the remote surgeon is illustrated in **FIG. 13**. The remote surgical field can be viewed by the remote surgeon with a binocular video system. The video system can show the field with his hands and instruments performing the procedure. The viewing system can be referred to as a surgical videoscope.

[0093] The binocular video rendering of the remote field can be transmitted to the local expert), who can view the (now virtual) stereoscopic rendering of the procedure

through a second surgical videoscope system. The local expert can insert his hands into the virtual field, thus seeing his real hands within the virtual field.

[0094] The video image of the local expert's hands can be transmitted back to the remote surgeon's surgical videoscope system superimposed into the real field. The remote surgeon can then see the expert's virtual hands within his surgical field in a spatially/anatomically relevant context. With this system, the local expert can use his hands to show the remote surgeon how to perform the case.

[0095] Exemplary elements of the system can comprise a remote station where the remote surgeon can perform the operative procedure, a remote surgical videoscope system comprised of, for example, a fixed stereoscopic videoscope that may resemble a mounted microscope. This apparatus can be used by the remote surgeon to view the operative field. Any other type of suitable VIP display can be used. The system can project the binocular video image to a similar local surgical videoscope at a local station. The local surgical videoscope can receive the binocular video image of the remote procedure and allow the local expert to view it. The local videoscope can view the local surgeon's hands as they move within the virtual remote field as viewed through the local videoscope. The local videoscope can then transmit the local expert's hands back to the remote videoscope so that the remote surgeon can see the expert's virtual hands within the real field.

[0096] With this system, the local expert can show the remote surgeon the appropriate maneuvers that result in successful completion of the case. The remote surgeon can have a basic skill set to carry out the new procedure. Therefore, the local expert can simply demonstrate to the remote surgeon new ways to apply the skill set. This system does not have to supplant the remote surgeon, but can be used to enhance his/her capability. The remote surgeon can be on hand to rapidly deal with any emergencies. Time delay is minimized because the remote surgeon can use his/her own hands to perform the task, eliminating the need for the local expert to manipulate remote robotic apparatuses.

[0097] Also disclosed are methods and systems for merging contemporary medical imaging onto an operative field. A volume image can be obtained of the operative field. For example, a volume MRI of the head, prior to the surgical procedure. The image data can be reconstructed into a three dimensional rendering of the anatomy. This rendering can be transmitted to the surgical videoscope that will be used to view the operative field. Through the videoscope, the surgeon can view this 3D rendering

in a translucent manner superimposed onto the surgical field. In this case, the surgeon would see a rendered head superimposed on the real head. Using software tools in the surgical videoscope interface, the surgeon can rotate and scale the rendered image until it “fits” the real head. The videoscope system can allow the surgeon to differentially fade the rendered head and real head so that the surgeon can “look into” the real head and plan the surgery.

[0098] Exemplary elements of the system can comprise a surgical videoscope viewing system through which the surgeon views the surgical field. A computer for reconstruction of a volume-acquired MRI/CT (or other) image with sufficient resolution to enable matching it to the real surgical anatomy. The volume rendered image can be displayed through the videoscope system so that the surgeon can see it stereoscopically. A software interface can enable the surgeon to vary the translucency of the rendered and real anatomy so that the rendered anatomy can be superimposed onto the real anatomy. The surgeon can “open up” the rendered anatomy to view any/all internal details of the image as they relate to the real anatomy. Surgical tools can be spatially registered to the rendered anatomy so that behavior can be tracked and applied to the image.

[0099] As shown in **FIG. 14**, an example of such a task is placing small objects inside a jar of dark gelatin so that they are not visible to the surgeon. The task is for the surgeon to use a long forceps to reach into the gelatin and touch or grasp the objects. The Surgical Videoscope system can obtain a volume scan of the gelatin jar and render the jar in three dimensions and display a binocular rendering through the videoscope. The surgeon can view the rendering and the real jar through the scope system and fit the rendered jar onto the real jar. By differentially adjusting translucency, the surgeon can reach into the real jar with a forceps and grasp a selected object, while avoiding other designated objects.

[00100] The grasping instrument can be spatially registered onto the volumetric rendering of the surgical field, thereby allowing a graphic of the tool to be displayed on the rendering of the surgical field in appropriate anatomic orientation. This can provide enhanced guidance. This can be implemented by touching designated landmarks on the real object (jar) with a digitizer that communicates with the image rendering system, thus defining the object/probe relationship. Because the object (jar) is registered to the image of the jar by superimposition, a graphic of the probe can be displayed in relation to the image of the jar enabling virtual surgery.

[00101] There are many situations in which the present system can be used. For example, remote surgery, medical training, and tele-medicine, which can be used for third world countries or in a military situation. Surgeons remotely located from patients can assist other surgeons near the patient, can assist medics near the patient, and can perform surgical operations when coupled to a robotic surgery system. Other examples include, augmented or enhanced surgery - normal surgery using virtual environments, an example of which is endoscopic surgery. Surgical procedures can also be simulated. Surgeons located remote from each other may plan and practice a procedure before carrying out the operation on a real patient.

[00102] Other applications include the preparation of patient before surgery, medical therapy, preventative medicine, exposure therapy, reducing phobias, training people with disabilities and skill enhancement, and the like.

[00103] The viewer then views the projection through passive stereoscopic polarized glasses (similar to sunglasses) that route the left-eye image to the left eye, and the right-eye image to the right eye. This provides an illusion of stereopsis when the correctly-offset images are properly rendered by the software. The system can be replaced by other types of stereoscopic displays with no functional detriment to the system. The stereoscopic display can comprise at least two display projectors fitted with polarizing lenses, a back-projection screen material that maintains light polarization upon diffusion, special glasses that restrict each eye to see only light of a particular polarization, and the viewer. The image to be viewed can be rendered with two slightly different view transformations, reflecting the different locations of the ideal viewer's two eyes. One projector displays the image rendered for the left eye's position, and the other projector displays the image rendered for the right eye's position. The glasses restrict the light so that the left eye sees only the image rendered for it, and the right eye sees only the image rendered for it. The viewer, presented with a reasonable stereoscopic image, will perceive depth.

[00104] **FIG. 15** is a block diagram illustrating an exemplary operating environment for performing the disclosed methods. This exemplary operating environment is only an example of an operating environment and is not intended to suggest any limitation as to the scope of use or functionality of operating environment architecture. Neither should the operating environment be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in the exemplary operating environment.

- [00105] The methods can be operational with numerous other general purpose or special purpose computing system environments or configurations. Examples of well known computing systems, environments, and/or configurations that may be suitable for use with the system and method include, but are not limited to, personal computers, server computers, laptop devices, and multiprocessor systems. Additional examples include set top boxes, programmable consumer electronics, network PCs, minicomputers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.
- [00106] The methods may be described in the general context of computer instructions, such as program modules, being executed by a computer. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. The system and method may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote computer storage media including memory storage devices.
- [00107] The methods disclosed herein can be implemented via one or more general-purpose computing devices in the form of a computer **1501**. The components of the computer **1501** can include, but are not limited to, one or more processors or processing units **1503**, a system memory **1512**, and a system bus **1513** that couples various system components including the processor **1503** to the system memory **1512**.
- [00108] The system bus **1513** represents one or more of several possible types of bus structures, including a memory bus or memory controller, a peripheral bus, an accelerated graphics port, and a processor or local bus using any of a variety of bus architectures. By way of example, such architectures can include an Industry Standard Architecture (ISA) bus, a Micro Channel Architecture (MCA) bus, an Enhanced ISA (EISA) bus, a Video Electronics Standards Association (VESA) local bus, and a Peripheral Component Interconnects (PCI) bus also known as a Mezzanine bus. This bus, and all buses specified in this description can also be implemented over a wired or wireless network connection. The bus **1513**, and all buses specified in this description can also be implemented over a wired or wireless network connection and each of the subsystems, including the processor **1503**, a

mass storage device **1504**, an operating system **1505**, application software **1506**, data **1507**, a network adapter **1508**, system memory **1512**, an Input/Output Interface **1510**, a display adapter **1509**, a display device **1511**, and a human machine interface **1502**, can be contained within one or more remote computing devices **1514a,b,c** at physically separate locations, connected through buses of this form, in effect implementing a fully distributed system.

[00109] The computer **1501** typically includes a variety of computer readable media. Such media can be any available media that is accessible by the computer **1501** and includes both volatile and non-volatile media, removable and non-removable media. The system memory **1512** includes computer readable media in the form of volatile memory, such as random access memory (RAM), and/or non-volatile memory, such as read only memory (ROM). The system memory **1512** typically contains data such as data **1507** and/or program modules such as operating system **1505** and application software **1506** that are immediately accessible to and/or are presently operated on by the processing unit **1503**.

[00110] The computer **1501** may also include other removable/non-removable, volatile/non-volatile computer storage media. By way of example, **FIG. 15** illustrates a mass storage device **1504** which can provide non-volatile storage of computer code, computer readable instructions, data structures, program modules, and other data for the computer **1501**. For example, a mass storage device **1504** can be a hard disk, a removable magnetic disk, a removable optical disk, magnetic cassettes or other magnetic storage devices, flash memory cards, CD-ROM, digital versatile disks (DVD) or other optical storage, random access memories (RAM), read only memories (ROM), electrically erasable programmable read-only memory (EEPROM), and the like.

- [00111] Any number of program modules can be stored on the mass storage device **1504**, including by way of example, an operating system **1505** and application software **1506**. Each of the operating system **1505** and application software **1506** (or some combination thereof) may include elements of the programming and the application software **1506**. Data **1507** can also be stored on the mass storage device **1504**. Data **1507** can be stored in any of one or more databases known in the art. Examples of such databases include, DB2®, Microsoft® Access, Microsoft® SQL Server, Oracle®, MySQL, PostgreSQL, and the like. The databases can be centralized or distributed across multiple systems.
- [00112] A user can enter commands and information into the computer **1501** via an input device (not shown). Examples of such input devices include, but are not limited to, a keyboard, pointing device (e.g., a “mouse”), a microphone, a joystick, a serial port, a scanner, tactile input devices such as gloves, and other body coverings, and the like. These and other input devices can be connected to the processing unit **1503** via a human machine interface **1502** that is coupled to the system bus **1513**, but may be connected by other interface and bus structures, such as a parallel port, game port, or a universal serial bus (USB).
- [00113] A display device **1511** can also be connected to the system bus **1513** via an interface, such as a display adapter **1509**. A computer **1501** can have more than one display adapter **1509** and a computer **1501** can have more than one display device **1511**. For example, a display device can be a monitor, an LCD (Liquid Crystal Display), or a projector. In addition to the display device **1511**, other output peripheral devices can include components such as speakers (not shown) and a printer (not shown) which can be connected to the computer **1501** via Input/Output Interface **1510**.
- [00114] The computer **1501** can operate in a networked environment using logical connections to one or more remote computing devices **1514a,b,c**. By way of example, a remote computing device can be a personal computer, portable computer, a server, a router, a network computer, a peer device or other common network node, and so on. Logical connections between the computer **1501** and a remote computing device **1514a,b,c** can be made via a local area network (LAN) and a general wide area network (WAN). Such network connections can be through a network adapter **1508**. A network adapter **1508** can be implemented in both wired and wireless

environments. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets, and the Internet **1515**.

[00115] One or more VIP displays **1516a,b,c,d,e** can communicate with the computer **1501**. In one aspect, VIP display **1516e** can communicate with computer **1501** through the input/output interface **1510**. This communication can be wired or wireless. Remote VIP displays **1516a,b,c** can communicate with computer **1501** by communicating first with a respective remote computing device **1514a,b,c** which then communicates with computer **1501** through the network adapter **1508** via a network such as the Internet **1515**. Remote VIP display **1516d** can communicate with computer **1501** without the need for a remote computing device. Remote VIP display **1516d** can communicate via a network, such as the Internet **1515**. The VIP displays **1516a,b,c,d,e** can communicate wireless or through a wired connection. The VIP displays **1516a,b,c,d,e** can communicate individual or collectively as part of a VIP display network.

[00116] For purposes of illustration, application programs and other executable program components such as the operating system **1505** are illustrated herein as discrete blocks, although it is recognized that such programs and components reside at various times in different storage components of the computing device **1501**, and are executed by the data processor(s) of the computer. An implementation of application software **1506** may be stored on or transmitted across some form of computer readable media. Computer readable media can be any available media that can be accessed by a computer. By way of example, and not limitation, computer readable media may comprise “computer storage media” and “communications media.” “Computer storage media” include volatile and non-volatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules, or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by a computer.

[00117] Unless otherwise expressly stated, it is in no way intended that any method set forth herein be construed as requiring that its steps be performed in a specific

order. Accordingly, where a method claim does not actually recite an order to be followed by its steps or it is not otherwise specifically stated in the inventive concepts or descriptions that the steps are to be limited to a specific order, it is no way intended that an order be inferred, in any respect. This holds for any possible non-express basis for interpretation, including: matters of logic with respect to arrangement of steps or operational flow; plain meaning derived from grammatical organization or punctuation; the number or type of embodiments described in the specification.

[00118] It will be apparent to those skilled in the art that various modifications and variations can be made in the present methods and systems without departing from the scope or spirit. Other embodiments will be apparent to those skilled in the art from consideration of the specification and practice disclosed herein. It is intended that the specification and examples be considered as exemplary only, with a true scope and spirit being indicated by the following claims.

CLAIMS

What is claimed is:

1. A method for providing a visualization of uncertainty in a collaborative video environment, the method comprising:
 - determining motion of an element in a video sequence from a first image to a second image in the video sequence;
 - generating a motion vector directed from the first image to the second image, wherein the motion vector is generated based upon the determined motion;
 - determining a latency value comprising a network latency iteratively sampled over a pre-defined time period and a local latency indicating a difference in time from capture of an image to display of the image via a local processor;
 - determining a precision value comprising a scalar value configured to be multiplied by the generated motion vector or a transformed motion vector, wherein the precision value comprises a pre-determined acceptable error or a user-provided acceptable error, and is dependent on context;
 - generating an uncertainty element based upon the motion vector or the transformed motion vector, the latency value, and the precision value, involving multiplying the precision value by the generated motion vector or the transformed motion vector, the uncertainty element comprising an uncertainty region of pixels projecting from one or both of the motion vector and the transformed motion vector, wherein a size of the uncertainty region is proportional to a magnitude of the motion vector or the transformed motion vector; and
 - rendering, to at least one of a plurality of collaborating users, the uncertainty element as a visualization comprising one or both of blur and/or colorization of pixels.
2. The method of claim 1, wherein the motion is determined from a motion estimation Algorithm.
3. The method of claim 1, wherein the motion is determined from a block matching algorithm.
4. The method of claim 1, wherein the first image is a current frame and the second image is a previous frame of the video sequence.

5. The method of claim 4, wherein the motion vector is generated from an image region in the current frame to a center point of an image region in the previous frame.
6. The method of claim 1, wherein the uncertainty element is generated based upon a transformed motion vector comprising a directional inverse of the motion vector.
7. The method of claim 6, wherein the uncertainty element is generated by multiplying the precision value and the directional inverse of the motion vector.
8. The method of claim 1, wherein the uncertainty region is rendered as an alert color indicating a level of uncertainty.
9. The method of claim 1, wherein the first image is a current frame and the second image is a previous frame of a series of images.
10. The method of claim 9, further comprising generating a motion vector directed from the current frame to the previous frame.
11. The method according to any one of the preceding claims, wherein a size of the uncertainty region is based upon the latency value.
12. The method according to any one of the preceding claims, wherein a size of the uncertainty region is proportional to the precision value.
13. The method according to any one of the preceding claims, wherein the latency value is determined based upon one or more of a local latency, a network latency, and a jitter.

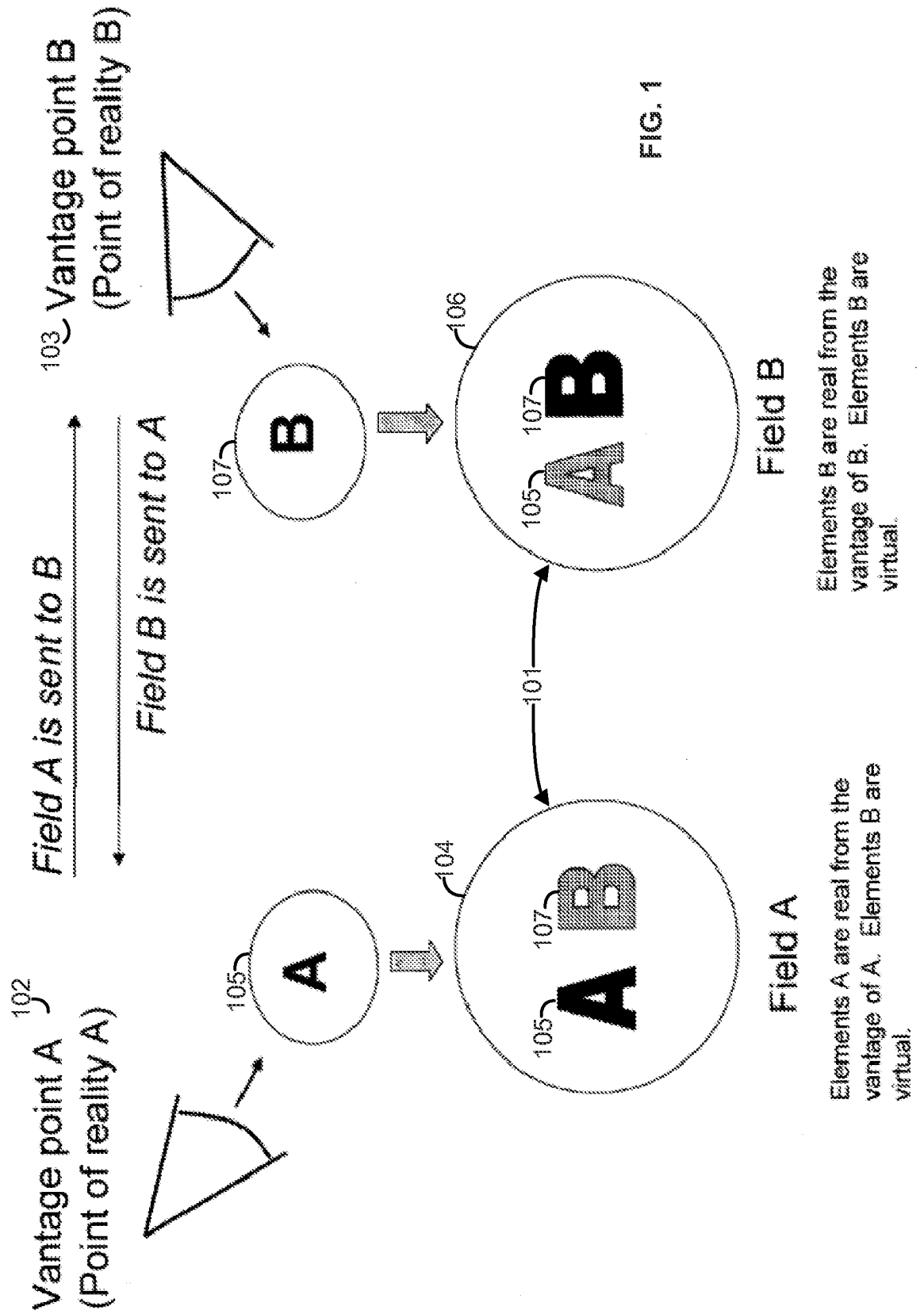


FIG.1

FIG. 2A

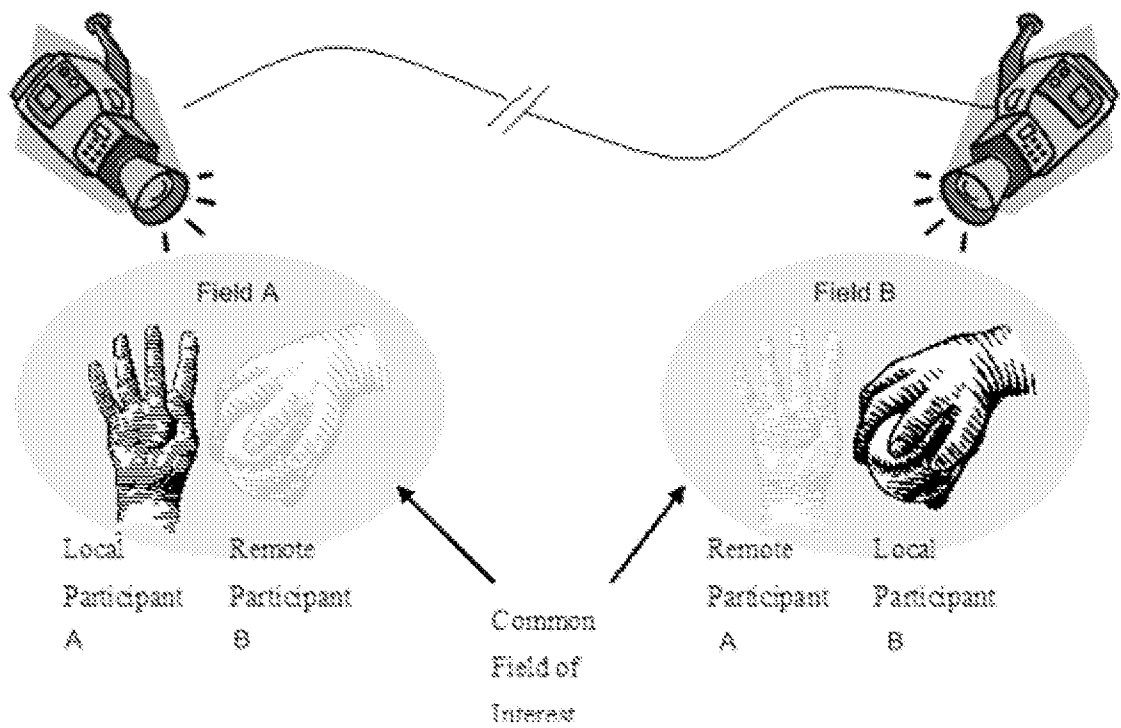


FIG. 2B



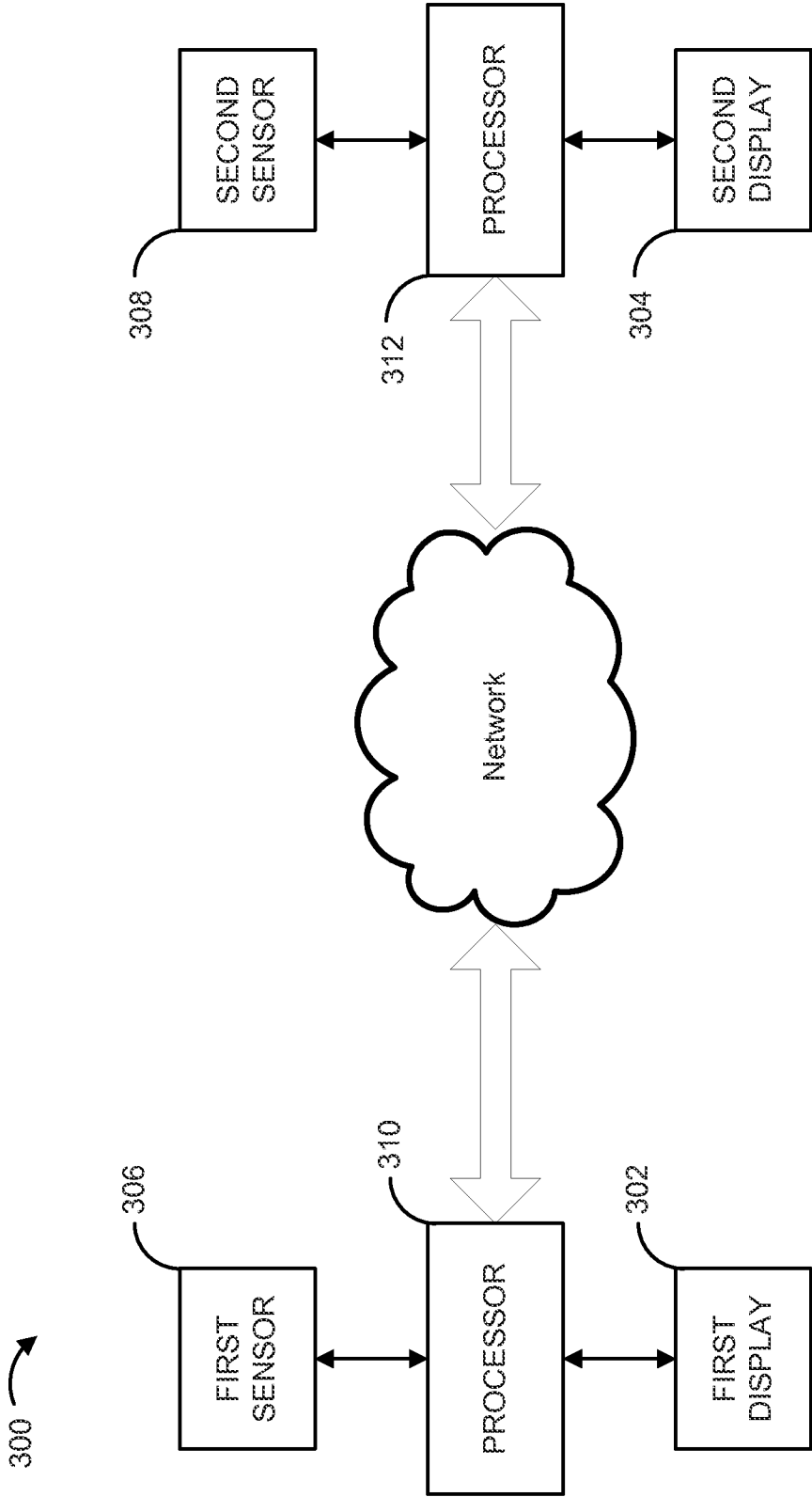


FIG. 3

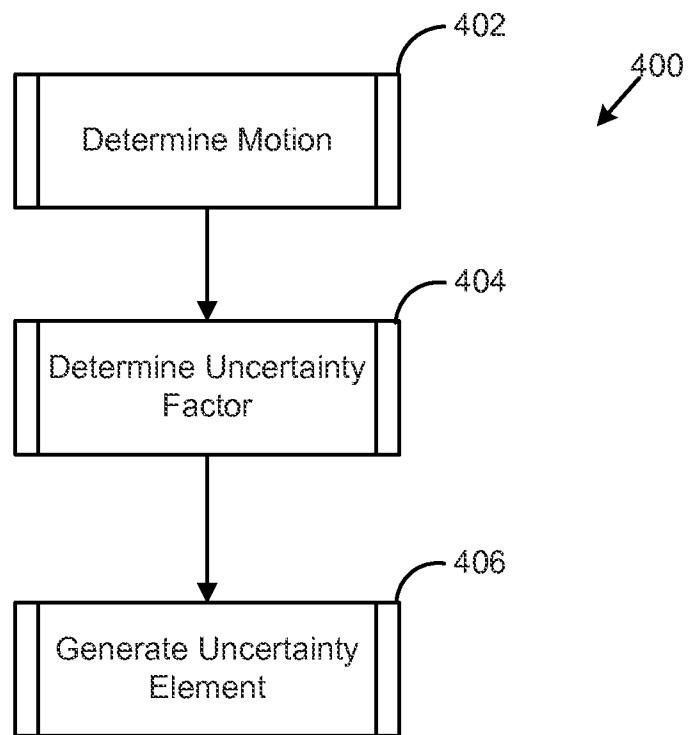


FIG. 4A

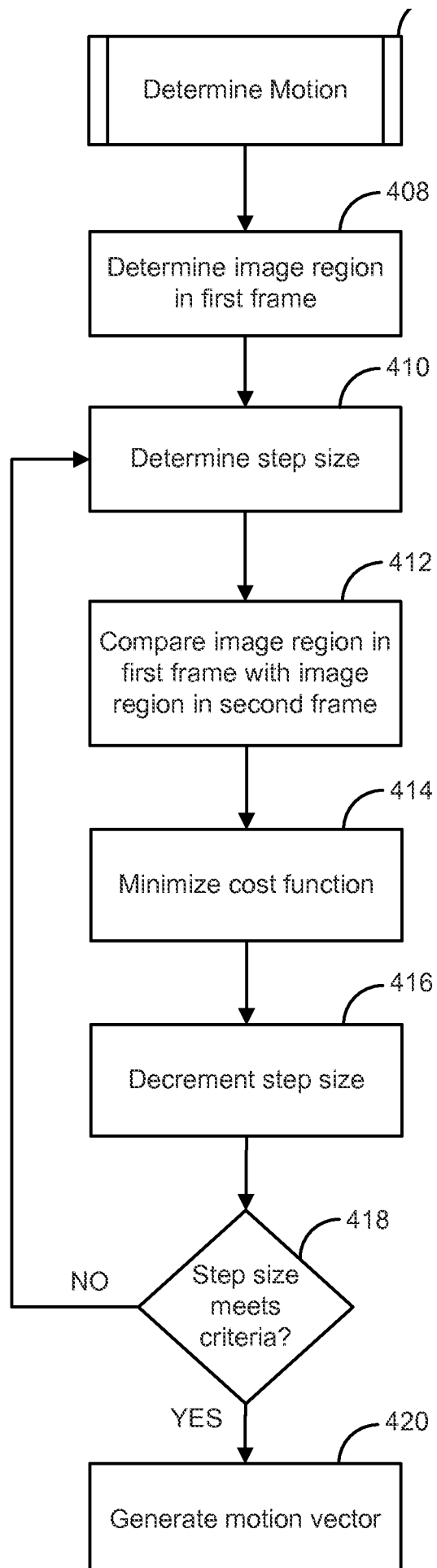


FIG. 4B

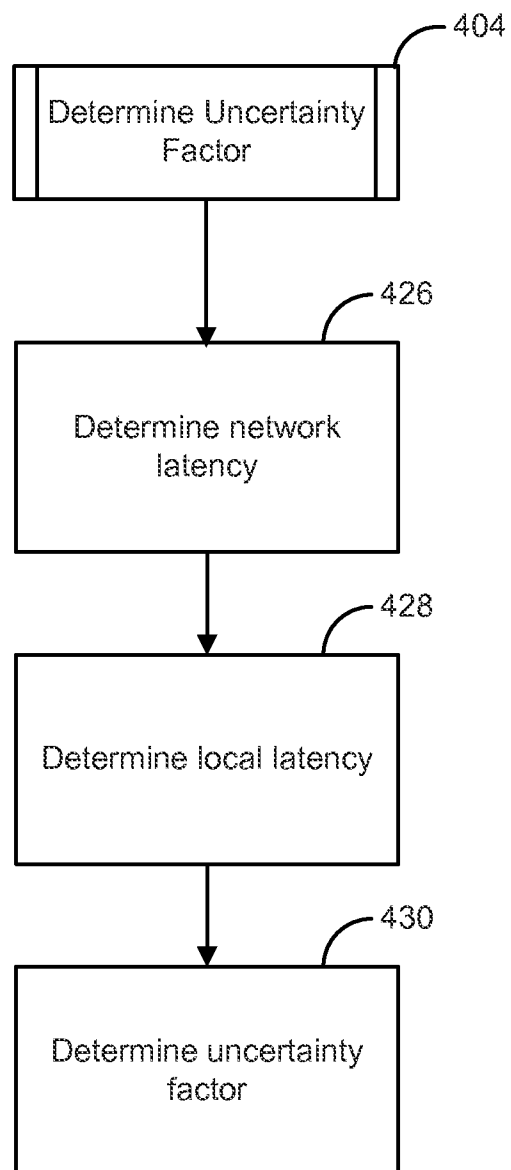


FIG. 4C

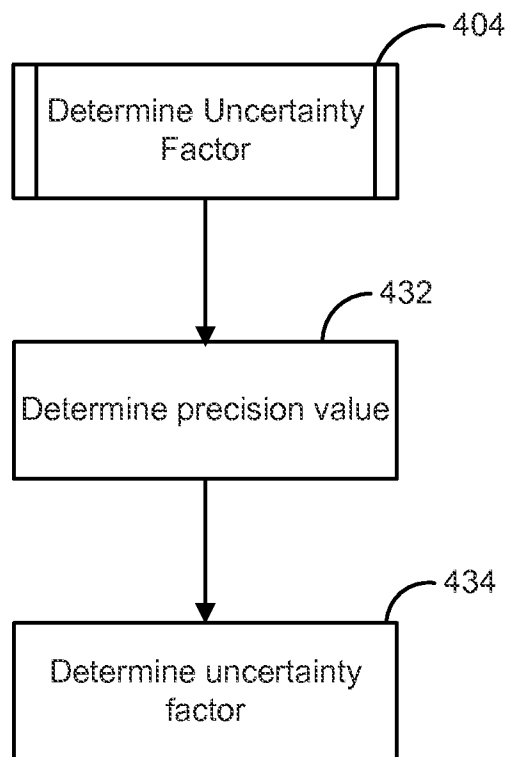


FIG. 4D

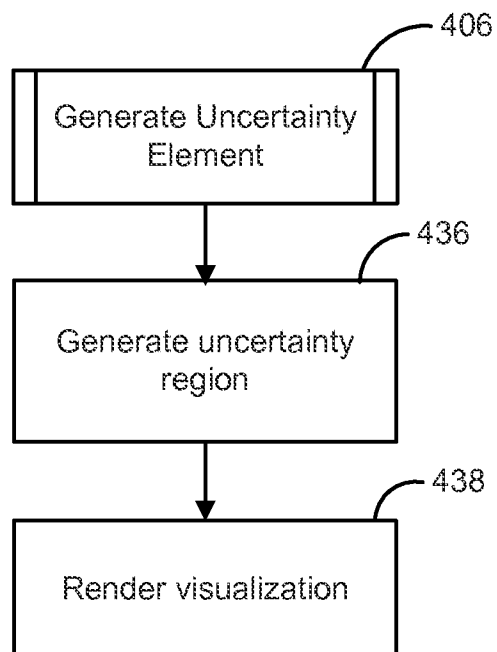


FIG. 4E

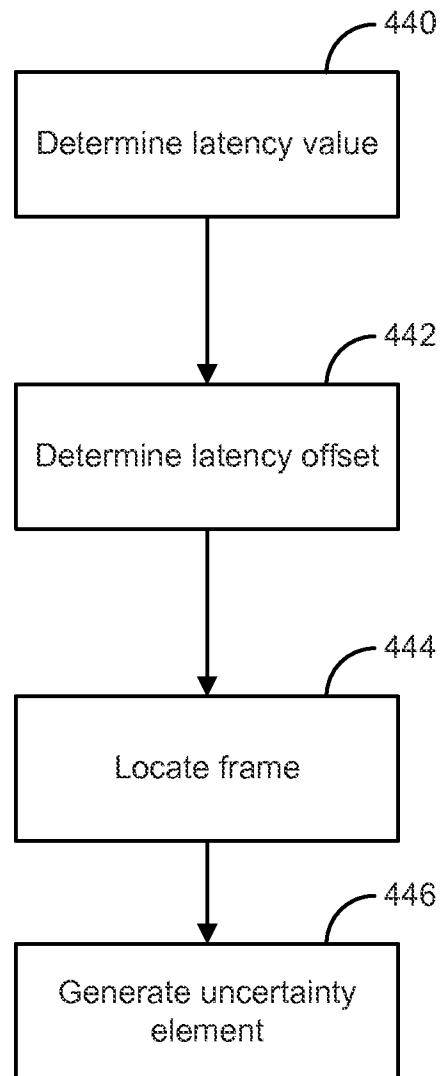


FIG. 4F

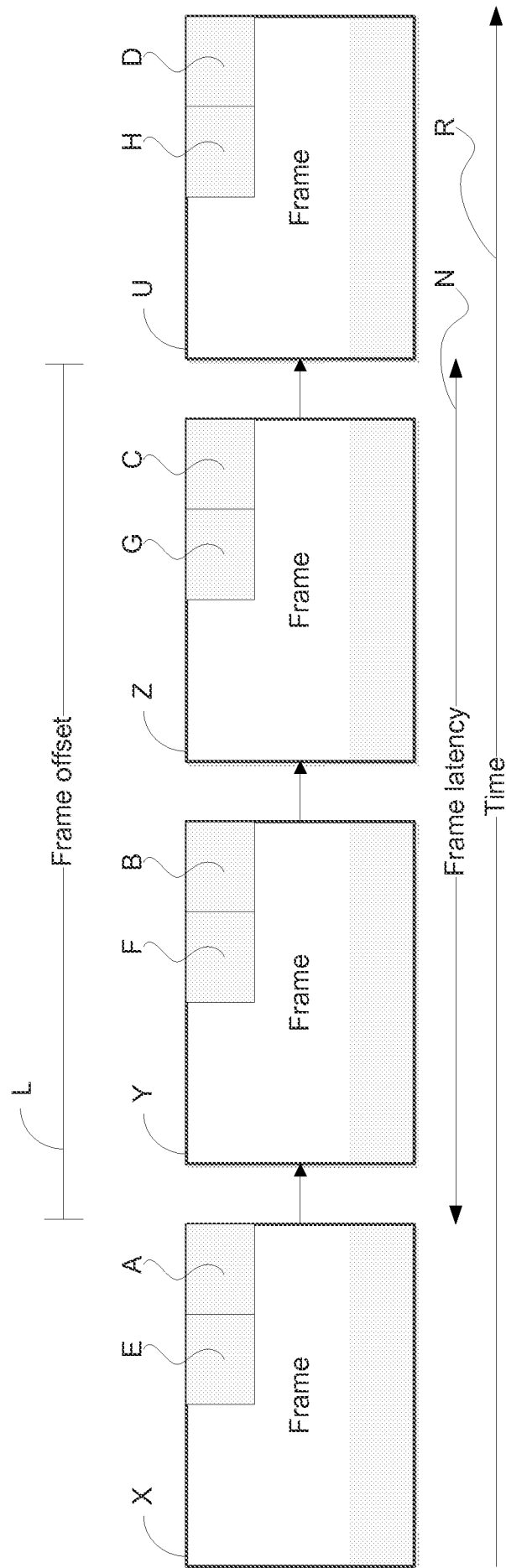


FIG. 4G

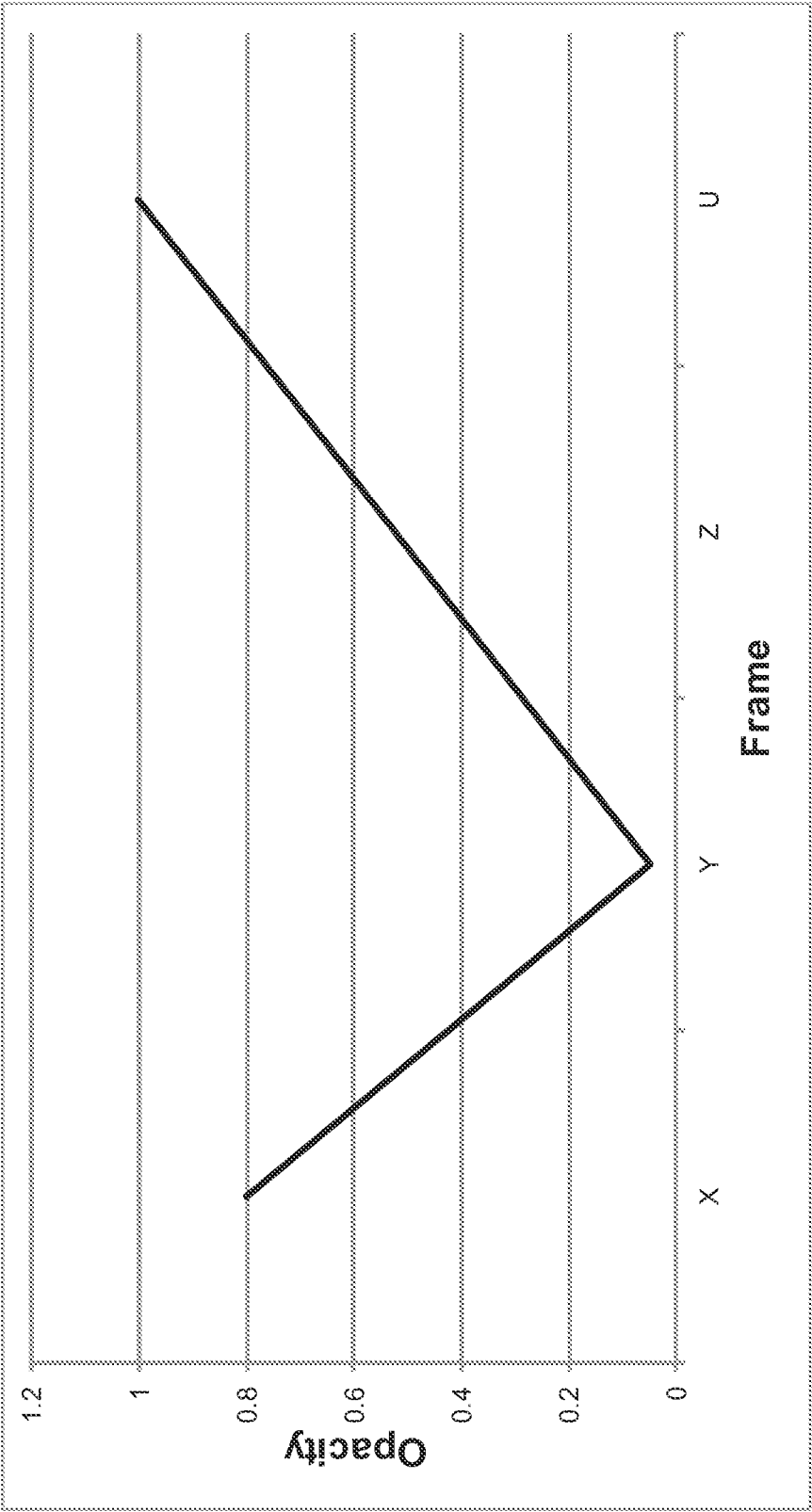


FIG. 4H

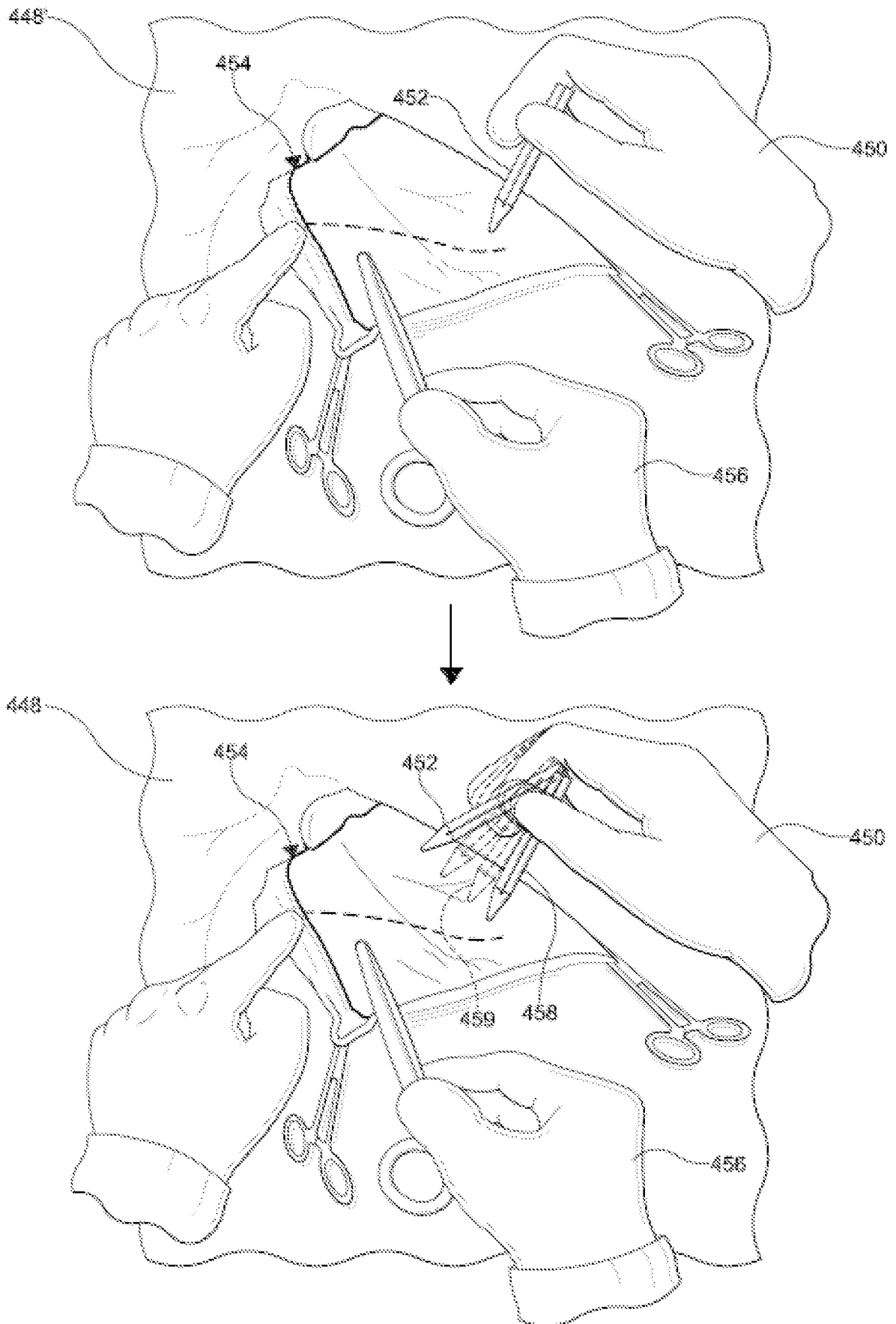


FIG. 4I

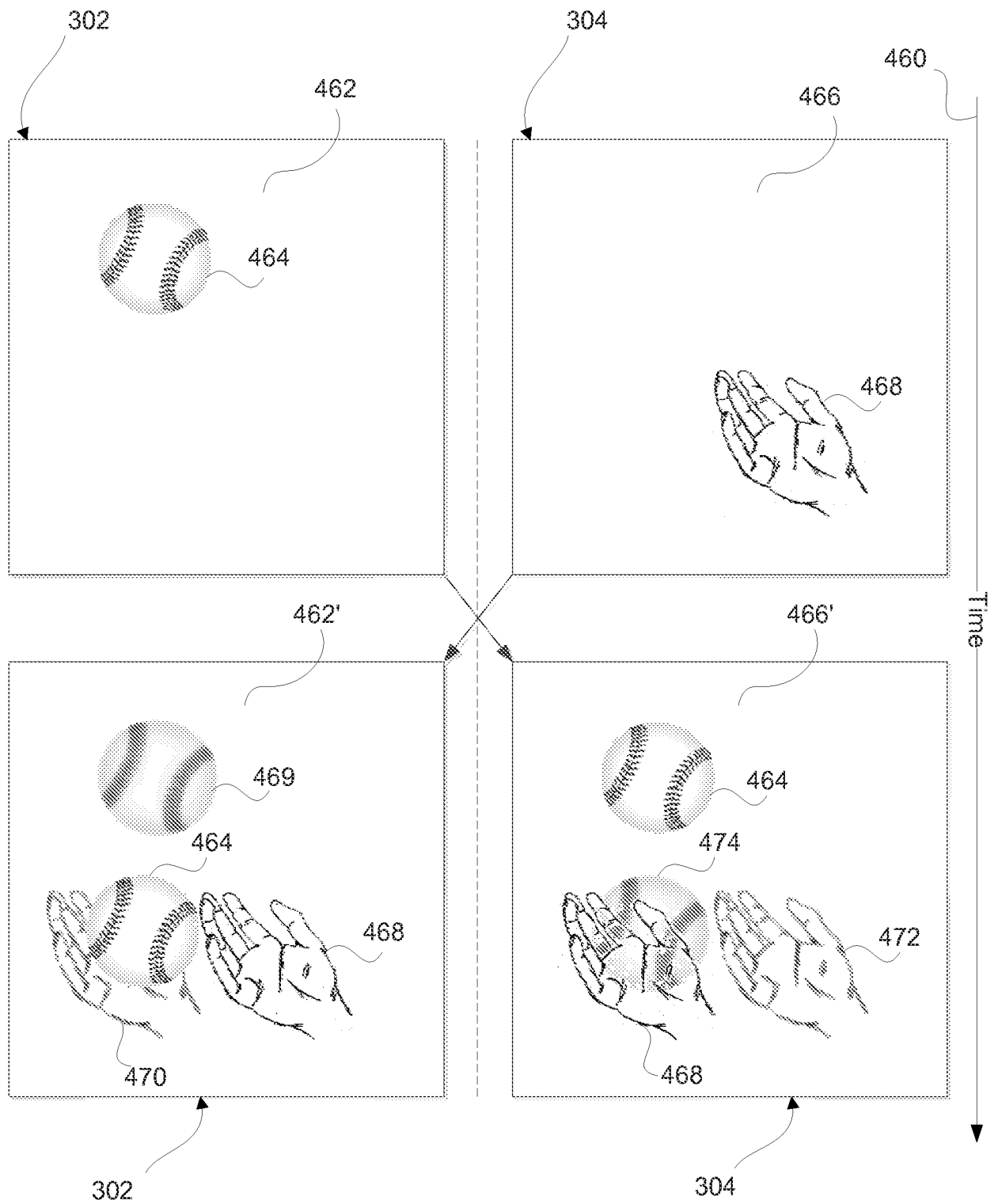
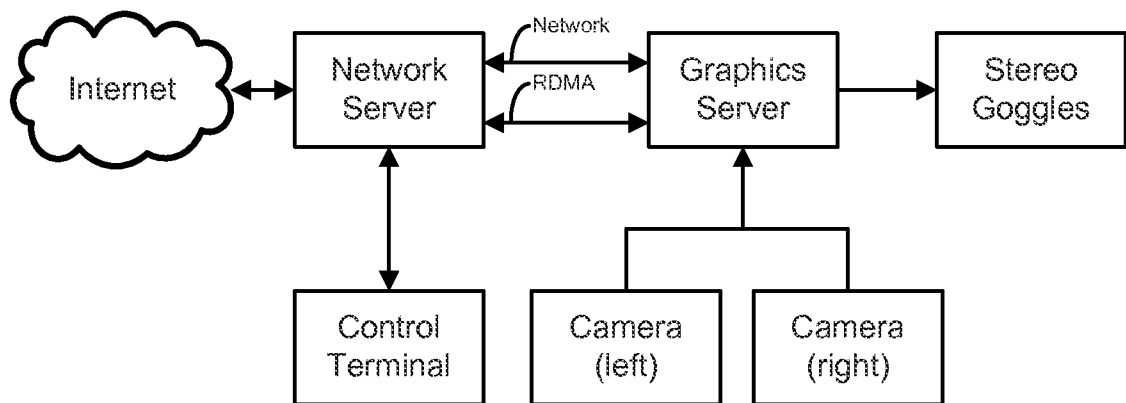


FIG. 4J

FIG. 5



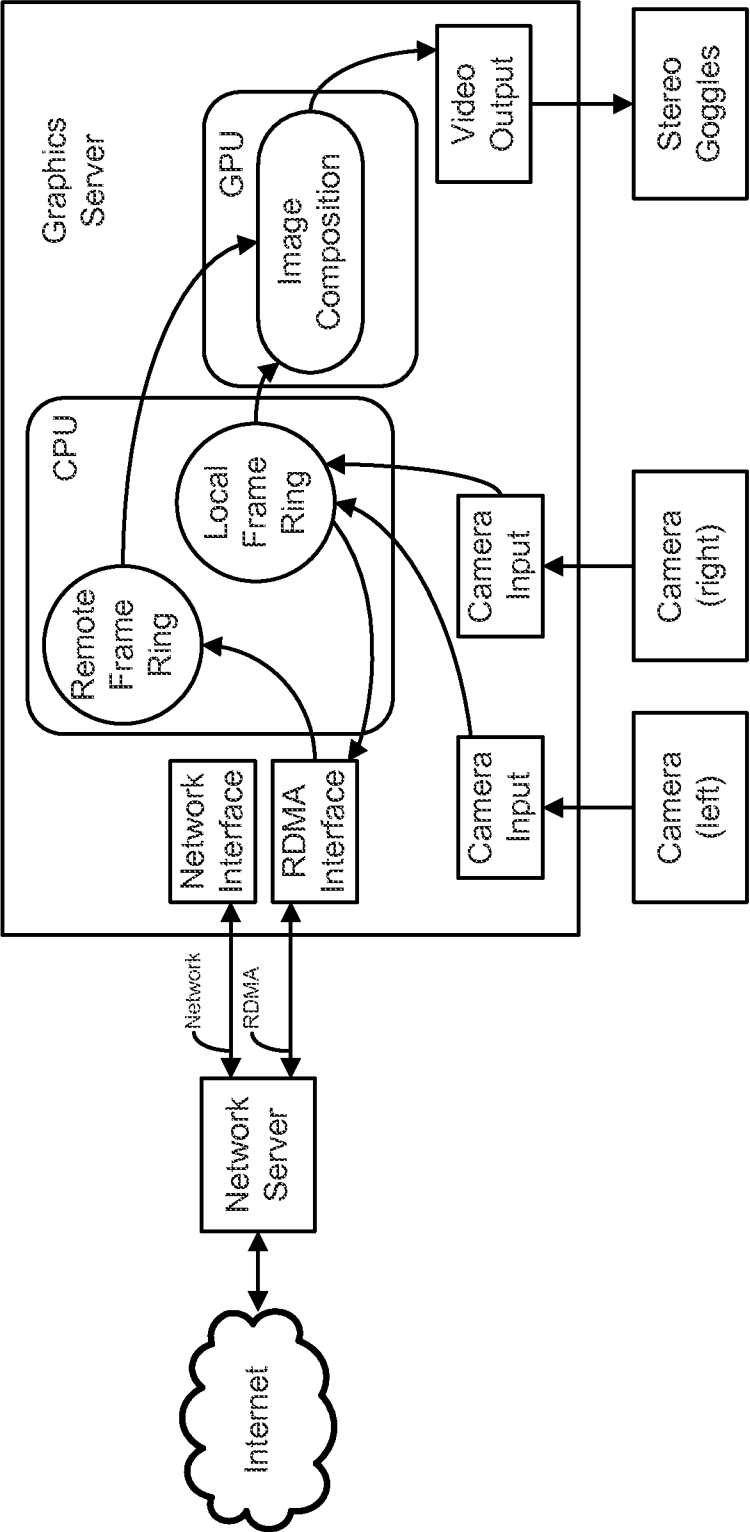


FIG. 6

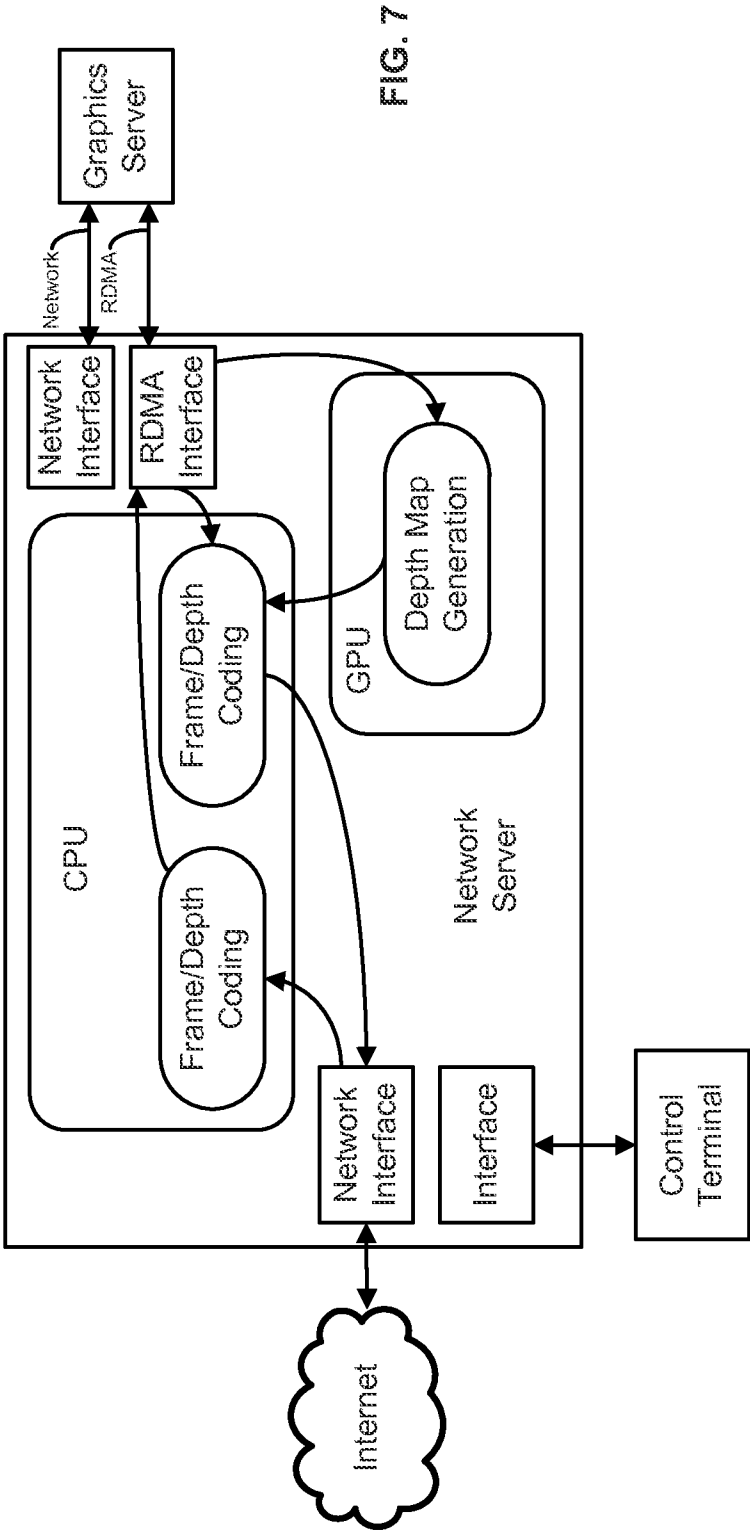


FIG. 8



FIG. 9

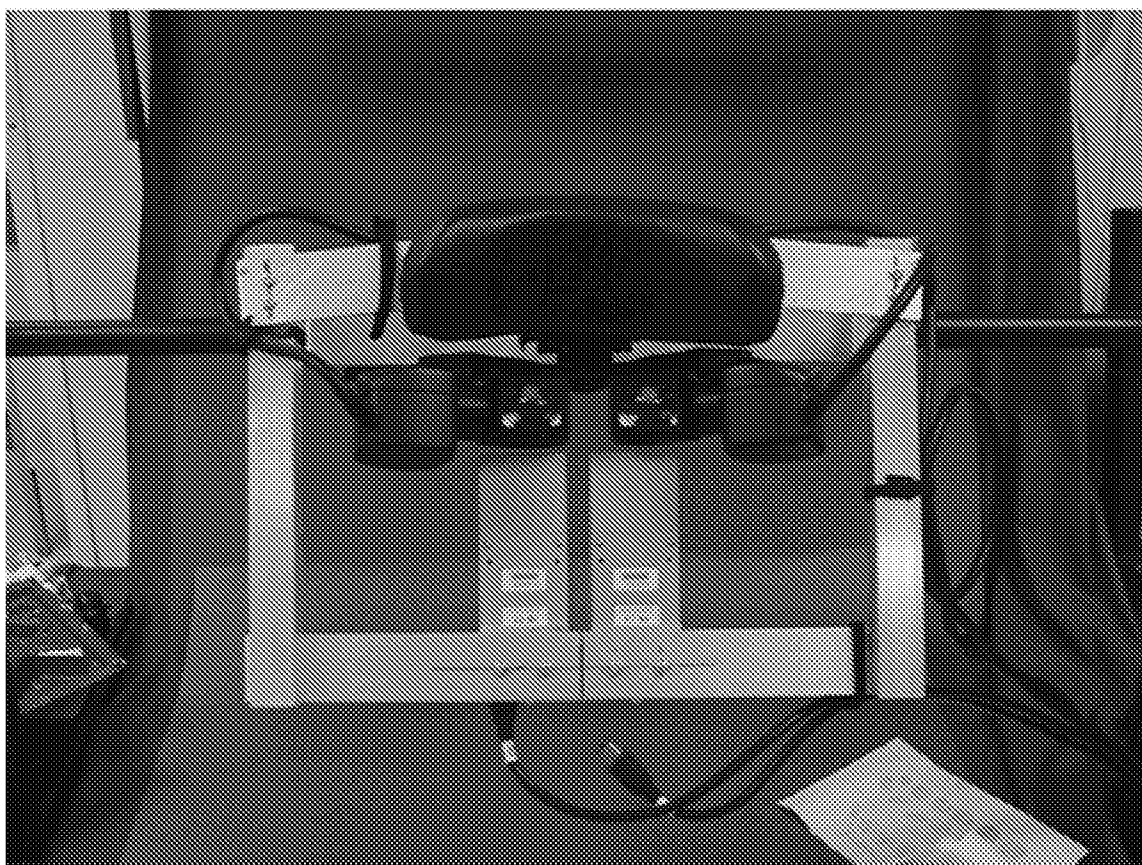


FIG. 10



FIG. 11

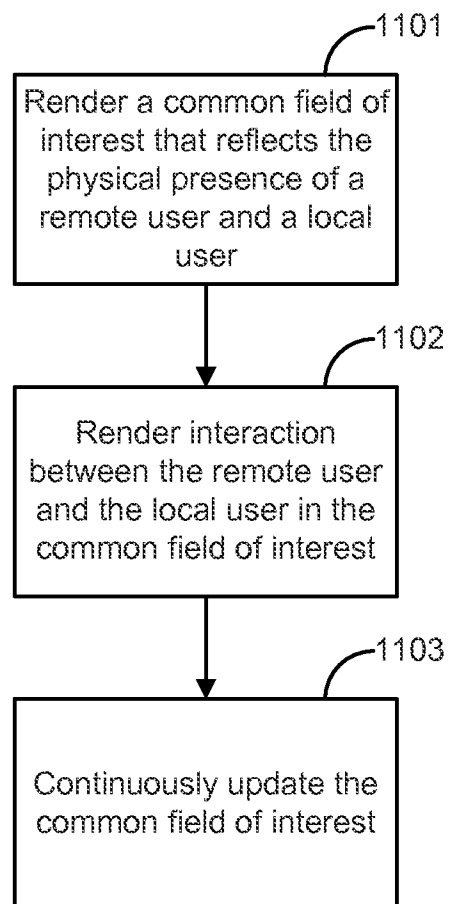
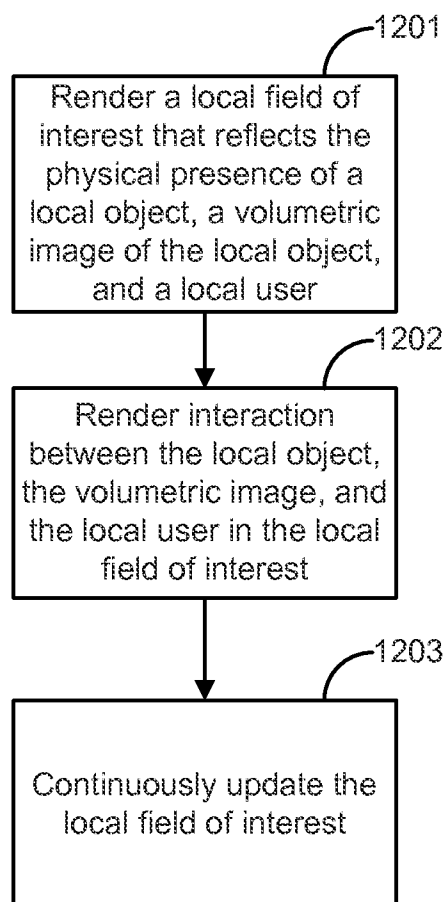
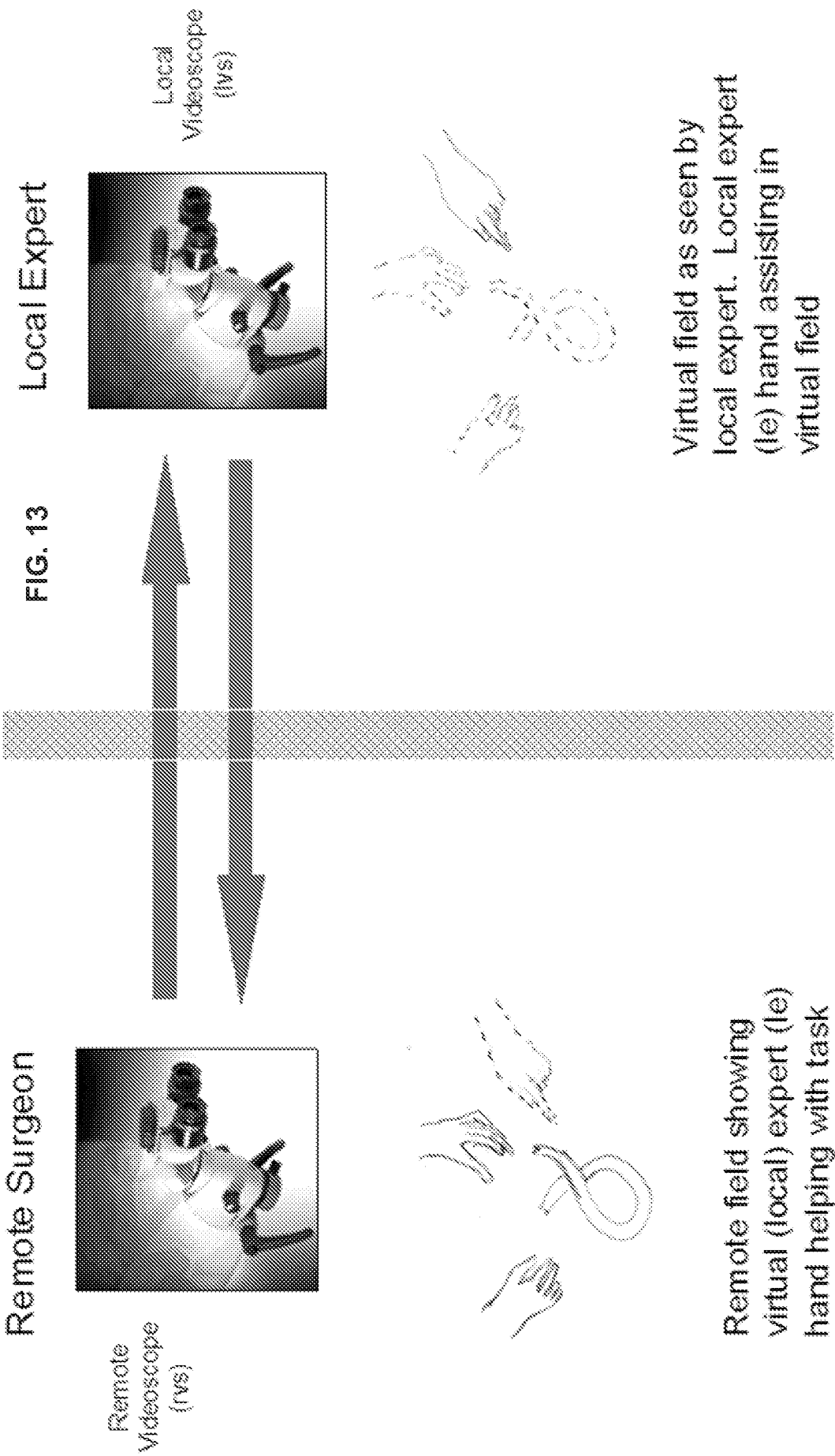
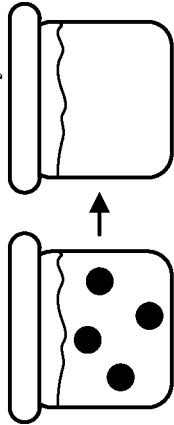


FIG. 12

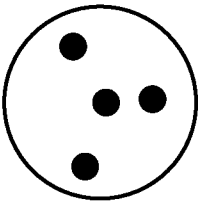




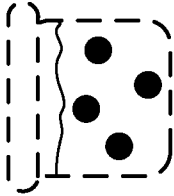
Objects suspended in beaker of opaque gelatin. Objects are not visible to the eye.



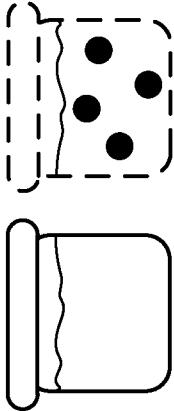
Slices of MRI scan of beaker of object.



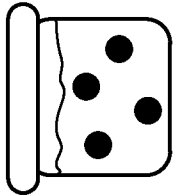
Computer volume rendering of "virtual beaker."



Real and virtual beaker as viewed through a videoscope.



User interface within the videoscope enables user to scale/rotate virtual beaker so that it is superimposed onto real beaker.



Once merged, the user can use the virtual beaker to navigate within the real beaker.

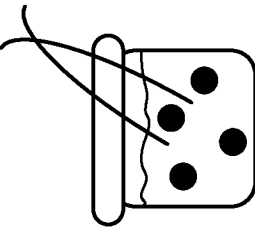


FIG. 14

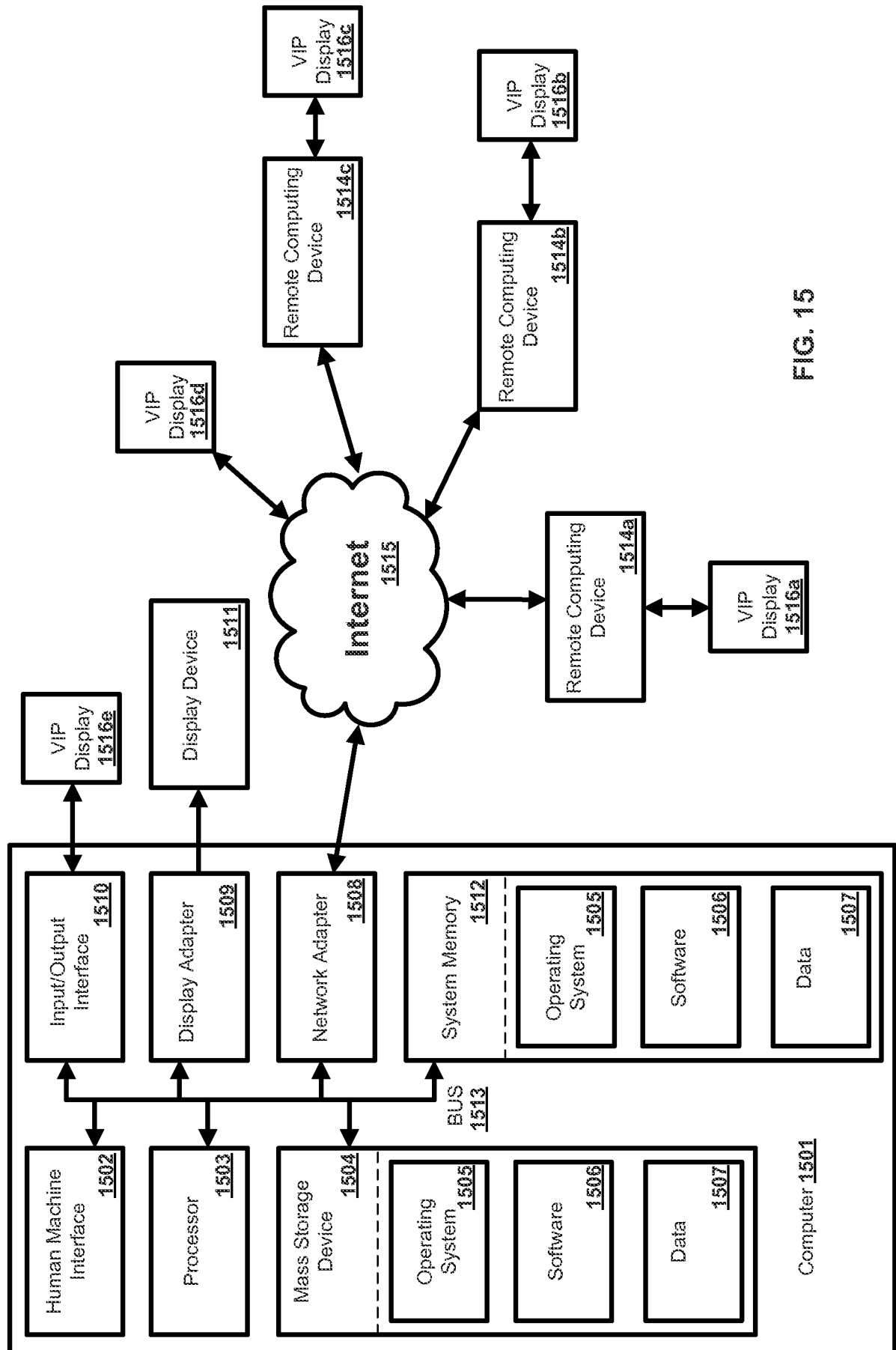


FIG. 15