(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 689 740 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
29.01.2014 Bulletin 2014/05

(21) Application number: **13187618.7**

(22) Date of filing: **08.05.2006**

(51) Int Cl.:
A61B 19/00 *(2006.01)* ***A61B 1/00*** *(2006.01)*
A61B 1/04 *(2006.01)*

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR

(30) Priority: **16.05.2005 US 130471**

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
06752425.6 / 1 893 118

(71) Applicant: **Intuitive Surgical Operations, Inc.**
Sunnyvale CA 94086 (US)

(72) Inventors:
- **Hoffman, Brian, D**
  Sunnyvale, CA California 94087 (US)
- **Larkin, David Q**
  Menlo Park, CA California 94025 (US)
- **Prisco, Giuseppe**
  Mountain View, CA California 94043 (US)
- **Zhang, Guanghua G**
  San Jose, CA California 95129 (US)
- **Kumar, Rajesh**
  Sunnyvale, CA California 94085 (US)

(74) Representative: **MacDougall, Alan John Shaw**
Mathys & Squire LLP
120 Holborn
London
EC1N 2SQ (GB)

Remarks:
This application was filed on 07-10-2013 as a divisional application to the application mentioned under INID code 62.

(54) Methods and system for performing 3-D tool tracking by fusion of sensor and/or camera derived data during minimally invasive robotic surgery

(57) Methods and system perform tool tracking during minimally invasive robotic surgery. Tool states are determined using triangulation techniques or a Bayesian filter from either or both non-endoscopically derived and endoscopically derived tool state information, or from either or both non-visually derived and visually derived tool state information. The non- endoscopically derived tool state information is derived from sensor data provided either by sensors associated with a mechanism for manipulating the tool, or sensors capable of detecting identifiable signals emanating or reflecting from the tool and indicative of its position, or external cameras viewing an end of the tool extending out of the body. The endoscopically derived tool state information is derived from image data provided by an endoscope inserted in the body so as to view the tool.

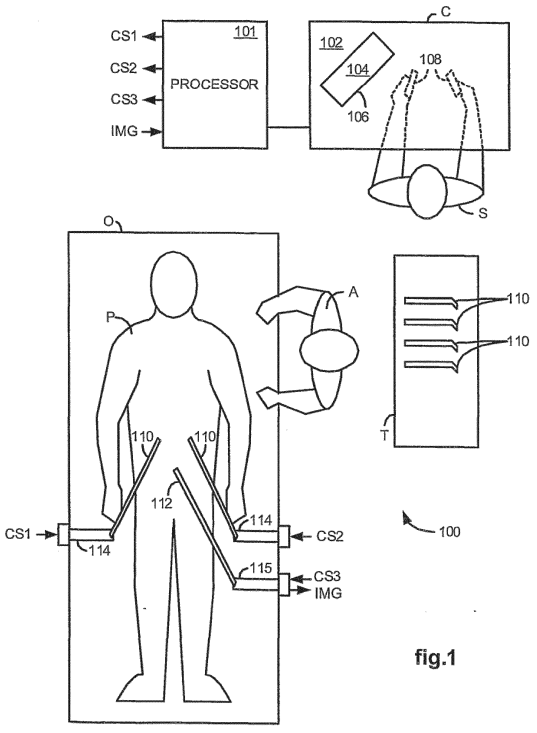

fig.1

EP 2 689 740 A1

Description

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSERED RESEARCH OR DEVELOPMENT

[0001] This invention was made with United States Government Support under cooperative agreement number: 70NANB1H3048 awarded by the National Institute of Standards and Technology (NIST). The United States government has certain rights in the invention.

FIELD OF THE INVENTION

[0002] The present invention generally relates to minimally invasive surgery and in particular, to methods and a system for performing 3-D tool tracking by fusion of sensor and/or camera derived data (e.g. tool position, velocity) during minimally invasive robotic surgery.

BACKGROUND OF THE INVENTION

[0003] Minimally invasive surgical techniques are aimed at reducing the amount of extraneous tissue that is damaged during diagnostic or surgical procedures, thereby reducing patient recovery time, discomfort, and deleterious side effects. As a consequence, the average length of a hospital stay for a standard surgery may be shortened significantly using minimally invasive surgical techniques. Also, patent recovery times, patient discomfort, surgical side effects, and time away from work may also be reduced with minimally invasive surgery.

[0004] A common form of minimally invasive surgery is endoscopy, and a common form of endoscopy is laparoscopy, which is minimally invasive inspection and surgery inside the abdominal cavity. In standard laparoscopic surgery, a patient's abdomen is insufflated with gas, and cannula sleeves are passed through small (approximately 1/2 inch or less) incisions to provide entry ports for laparoscopic surgical instruments.

[0005] The laparoscopic surgical instruments generally include a laparoscope or an endoscope (for viewing the surgical field), and working tools. The working tools are similar to those used in conventional (open) surgery, except that the working end or end effector of each tool is separated from its handle by an extension tube. As used herein, the term "end effector" means the actual working part of the surgical instrument and can include clamps, graspers, scissors, staplers, image capture lenses, and needle holders, for example.

[0006] To perform surgical procedures, the surgeon passes these working tools or instruments through the cannula sleeves to an internal surgical site and manipulates them from outside the abdomen. The surgeon views the procedure by means of a monitor that displays an image of the surgical site taken from the laparoscope. Similar endoscopic techniques are employed in, e.g., arthroscopy, retroperitoneoscopy, pelviscopy, nephroscopy, cystoscopy, cisternoscopy, sinoscopy, hysteroscopy, urethroscopy, and the like.

[0007] Minimally invasive telesurgical robotic systems are being developed to increase a surgeon's dexterity when working within an internal surgical site, as well as to allow a surgeon to operate on a patient from a remote location. In a telesurgery system, the surgeon is often provided with an image of the surgical site at a computer workstation. While viewing a three-dimensional image of the surgical site on a suitable viewer or display, the surgeon performs the surgical procedures on the patient by manipulating master input or control devices of the workstation. Each of the master input devices controls the motion of a servomechanically operated surgical instrument. During the surgical procedure, the telesurgical system can provide mechanical actuation and control of a variety of surgical instruments or tools having end effectors that perform various functions for the surgeon, e.g., holding or driving a needle, grasping a blood vessel, or dissecting tissue, or the like, in response to manipulation of the master input devices.

[0008] During the surgical procedure, however, the surgeon may manipulate the tool so that its end effector is moved outside of the endoscope's field of view, or the end effector may become difficult to see due to occlusion by fluids or other intervening objects. In such cases it would be useful to be able to provide assistance to the surgeon in locating and/or identifying the end effector on the workstation's display screen.

[0009] Various techniques have been developed for identifying the tool in a camera image. One such technique is described, for example, in Guo-Qing Wei, Klaus Arbter, and Gerd Hirzinger, "Real-Time Visual Servoing for Laparoscopic Surgery," IEEE Engineering in Medicine and Biology, Vol. 16, No. 1, pp. 40-45, 1997, wherein a color marker is used to identify a tool in a camera image so that the camera may be automatically maneuvered so as to keep the tool in the center of its visual image. Another such mechanism is described, for example, in Xiaoli Zhang and Shahram Payandeh, "Application of Visual Tracking for Robot-Assisted Laparoscopic Surgery," Journal of Robotics Systems, Vol. 19, No. 7, pp. 315-328, 2002, wherein a striped marker is used to identify a tool and its depth in a camera image so that the camera may be automatically maneuvered to visually track the tool.

[0010] These techniques, however, are not generally useful when the end effector is outside the endoscope's field of

view, or when it is occluded by fluid or some object. In either case, such markers cannot be found and tool identification is not possible. Second, simple color markers and other orientation independent markers placed on the tool do not facilitate determining the orientation or pose of the tool. Third, computational complexity in identifying and tracking the tool markers may make real-time tracking difficult. In particular, increased time spent processing each captured frame of video will reduce the frequency at which video capture can occur, possibly resulting in abrupt transitions from one image to the next. Fourth, uncorrected camera calibration and/or other system measurement errors may result in errors when determining tool positions and orientations from the camera's visual images.

OBJECTS AND SUMMARY OF THE INVENTION

[0011] To make each of multiple tools easily distinguishable to the surgeon on the workstation display screen and also to patient-side staff, a number of computer assisted techniques may be employed such as: predicting the position and orientation of the tool, and overlaying a computer aided design ("CAD") model of the tool, or other virtual mark or indicator, over the predicted position and orientation of the tool in the display screen; predicting the position and orientation of the tool, and uniquely marking each tool at its predicted position and orientation to make it distinguishable from other tools on the display screen; and predicting the position and orientation of the tool, and erasing or brushing out the shaft of the tool while highlighting its end effector in some fashion so that the end effector seemingly floats in and stands out on the display screen.

[0012] One advantage, for example, of uniquely marking each tool at its predicted position and orientation to make it distinguishable from other tools on the display screen, is that it a) makes it easier for the surgeon (who only sees the tools inside the body) to communicate with a patient-side staff (who only see the tools outside the body) about specific tools, and b) makes sure the surgeon is aware of tools that might be outside the field of view. This is especially important when there are more than two tools, since the surgeon may see only two and assume those are the two he is controlling with his two hands when in fact he is controlling (i.e. moving) another tool outside the field of view, potentially damaging tissue.

[0013] Likewise, one advantage, for example, of predicting the position and orientation of the tool, and erasing or brushing out the shaft of the tool while highlighting its end effector in some fashion so that the end effector seemingly floats in and stands out on the display screen, is that it makes more of the underlying tissue viewable by the surgeon. Note, however, that this view of the underlying tissue is by definition synthetic, as the current state of the tissue is obscured by the tool. To erase the tool shaft therefore requires storage of previous images of the operating site, such that the obscured region could be replaced with the last-known un-obscured view of that region, perhaps in a dimmed or grayed-out fashion (a.k.a. Fog-of-War), to indicate that the data is not current.

[0014] Tool tracking refers to the determination of a tool's state over time. The tool's state generally includes its position and orientation in a reference frame, as well as other related parameters such as its translational and rotational velocities. In the preferred embodiment, tool tracking is performed in the camera reference frame. Tool tracking facilitates prediction of the tool's position and orientation at a point in time when the tool is either out of view or occluded in the endoscope's view, by using position and orientation information from prior times when the tool was in view and identifiable, and/or position and orientation estimates derived from non-endoscope sources for that point in time and previous times.

[0015] Determination of the position and orientation of the tool is useful for reasons other than locating the tool on the workstation display screen. For example, tool position and orientation information may be used to generate graphical overlays containing information of interest to the surgeon. These overlays might include: telestration, the distance between tools, the distance between tools and the patient's anatomy, measurements of anatomical features in the camera reference frame, or measurements of anatomical features in another reference frame. Additionally, the position and orientation of the tool may be used to register the current tool and camera positions with pre-operative or planning data provided in a fixed or world reference frame, or to improve in general the safety and control of robotic mechanisms manipulating the tools.

[0016] As yet another example, determination of the position and orientation of the tool is also useful for the registration of data collected during a surgical procedure from sensors mounted on the tools themselves, for example, ultrasound sensors. In this case, if the position and orientation of the tool holding the sensor is known in the camera reference frame (i.e., the reference frame of the endoscope), then the position and orientation of any data collected from a ultrasound sensor mounted on the tool could be displayed, fully registered with the surgical image, on the workstation display screen to aid the surgeon during the surgical procedure.

[0017] In certain of these registration examples, it is necessary to determine the position and orientation of the tool in the fixed (or world) reference frame. Therefore, if the position and orientation of the endoscope is known in the fixed reference frame, then the position and orientation of the tool as viewed by the endoscope can be translated from the camera reference frame to the fixed reference frame. Alternatively, if the position and orientation of the tool can be determined in the fixed reference frame independently, the independent determination not only avoids this reference frame translation process, but it also provides an alternative means to determine the endoscope's position and orientation

in the fixed reference frame.

[0018] Accordingly, it is an object of aspects of the present invention to provide a method and system for performing tool tracking during minimally invasive surgery that operates even when a portion or even the entire tool is occluded in or out of the endoscope's view.

[0019] Another object of aspects of the invention is to provide a method and system for performing tool tracking that includes visual tracking using a marker that is orientation dependent so that a pose of the tool can be determined.

[0020] Another object of aspects of the invention is to provide a method and system for performing tool tracking that is accurate, reliable and/or computationally fast.

[0021] Still another object of aspects of the invention is to provide a method and system for performing tool tracking that operates real-time and minimizes abrupt transitions in determined tool positions and orientations so as to provide smooth tracking of the tool.

[0022] Yet another object of aspects of the invention is to provide a method and system for performing tool tracking that corrects or otherwise compensates for calibration errors.

[0023] These and additional objects are accomplished by the various aspects of the present invention, wherein briefly stated, one aspect is a tool tracking method comprising: tracking a tool by processing non-endoscopically derived tool state information and endoscopically derived tool state information generated while the tool is inserted and being manipulated through a minimally invasive incision in a body.

[0024] By using such a hybrid approach, when the tool is totally occluded in a frame of video from an endoscope inserted in the patient's body to view the surgical site, its state (e.g., its position, orientation, and translational and rotational velocity) may still be determined from the non-endoscopically derived tool position information such as that generated by: using system kinematics from joint positions in a robotic mechanism that is manipulating the tool; using electro-magnetic, acoustic, or other types of detectable signals emanating or reflecting from the tool (or the robotic mechanism manipulating the tool) so as to determine its position; using images generated by an external camera viewing an end of the tool extending out of the patient's body.

[0025] Also, the additional use of the non-endoscopically derived tool position information along with the endoscopically derived tool position information in determining tool position and orientation tends to minimize or at least significantly reduce any abrupt transitions in determined tool positions and orientations between frames of the image information provided the non-endoscopically derived tool position information is continuously available at a sampling rate. Further, the use of the endoscopically derived tool position information along with the non-endoscopically derived tool position information provides a redundant source of information for determining tool states, which can be utilized to determine more accurate tool tracking of tool positions and orientations over time.

[0026] Another aspect is a tool tracking method comprising: receiving sensor information indicative of a position and orientation of a tool when the tool is inserted through an incision in a body; receiving image information for the tool; and determining the position and orientation of the tool using both the sensor and the image information.

[0027] Another aspect is a minimally invasive robotic surgery system with tool tracking, comprising: one or more non-endoscopic devices providing data from which non-endoscopically derived tool state information is generated when a tool is inserted and robotically manipulated through an incision in a body; an endoscope capturing images from which endoscopically derived tool state information is generated for an area within the body when the tool is inserted therein; and a processor configured to process the non-endoscopically and endoscopically derived tool state information for tracking the state of the tool.

[0028] Another aspect is a minimally invasive robotic surgery system with tool tracking, comprising: one or more sensors providing sensor data from which non-visually derived tool state information for a tool is generated when the tool is inserted and robotically manipulated through an incision in a body; at least one camera capturing image information of the tool when the tool is inserted therein; and a processor configured to process the non-visually derived tool state information and the image information for tracking the state of the tool.

[0029] Another aspect is a tool tracking method comprising: determining a computer model of a tool; receiving a captured image including a view of the tool; determining an estimated position and orientation of the tool from the captured image, and positioning and orienting the computer model at that estimated position and orientation in reference to the captured image; and modifying the estimated position and orientation of the computer model with respect to an image of the tool in the captured image until the computer model approximately overlays the image so as to correct the estimated position and orientation of the tool for the captured image.

[0030] Another aspect is a tool tracking method comprising: determining whether sensor data indicative of a tool state is available for a point in time; determining whether image data indicative of the tool state is available for the point in time; and determining the tool state using both the sensor data and the image data if both are available for the point in time, or using only the sensor data if only the sensor data is available, or using only the image data if only the image data is available.

[0031] Another aspect is a tool tracking method comprising: determining a first estimated tool state relative to a landmark for a point in time using first sensor data indicative of the tool state at the point in time; determining an estimated

camera state relative to the landmark for the point in time using second sensor data indicative of the camera state at the point in time; determining a second estimated tool state relative to the camera for the point in time using image data generated by the camera and indicative of the tool state at the point in time; translating the first estimated tool state so as to be relative to the camera instead of the landmark; and computing an error transform between the first and the second estimated tool states so that at a subsequent point in time if image data indicative of the tool state at the subsequent point in time is not available, then the tool state is determined by applying the error transform to a third estimated tool state determined using sensor data indicative of the tool state at the subsequent point in time translated so as to be relative to the camera instead of the landmark.

[0032] Still another aspect is a tool tracking method comprising: determining non-endoscopically derived estimated state information for a tool at a given time; determining endoscopically derived estimated state information for the tool at the given time; and providing the non-endoscopically derived estimated state information and the endoscopically derived estimated state information to a Bayesian filter configured so as to generate an optimal estimate of the state of the tool.

[0033] Another aspect is a tool tracking and calibration method comprising: generating visually derived state information from image data received from a camera viewing a tool; generating state vector information by combining initial values for a set of camera parameters with the visually derived state information; and providing the state vector information to a Bayesian filter for processing so as to generate an optimal estimate of a state of the tool and corrected values for the set of camera parameters.

[0034] Another aspect is a camera tracking method comprising: determining a position of a tool in a fixed reference frame from non-visually derived tool state information generated from sensor data indicative of the position of the tool; determining a position of the tool in a camera frame moveable with a camera using visually derived tool state information generated from image data provided by the camera while viewing the tool; and determining a position of the camera in the fixed reference frame using the position of the tool in the fixed reference frame and the position of the tool in the moveable camera frame.

[0035] Another aspect is a tool tracking method comprising: determining a position of a camera in a fixed reference frame from non-visually derived camera state information generated from sensor data indicative of the position of the camera; determining a position of a tool in a camera frame moveable with the camera using visually derived tool state information generated from image data provided by the camera while viewing the tool; and determining a position of the tool in the fixed reference frame using the position of the camera in the fixed reference frame and the position of the tool in the moveable camera frame.

[0036] Still another aspect is a tool tracking method comprising: generating a plurality of estimated tool states for each point in a plurality of points in time, while the tool is inserted and being manipulated through an incision in a body; and determing an optimal estimated tool state for each point in the plurality of points in time by processing the plurality of estimated tool states using Bayesian techniques.

[0037] Additional objects, features and advantages of the various aspects of the present invention will become apparent from the following description of its preferred embodiment, which description should be taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0038] **FIG. 1** illustrates a minimally invasive robotic surgical system utilizing aspects of the present invention.

[0039] **FIG. 2** illustrates three-dimensional reference frames associated with a minimally invasive robotic surgical system utilizing aspects of the present invention.

[0040] **FIG. 3** illustrates a flow diagram of a tool tracking method for determining tool states using either or both tool sensor and tool images data, utilizing aspects of the present invention.

[0041] **FIG. 4** illustrates a flow diagram of a tool tracking method for determining tool states using both tool sensor and tool images data, utilizing aspects of the present invention.

[0042] **FIG. 5** illustrates a functional block diagram of a Kalman filter tool tracking method for determining tool states, utilizing aspects of the present invention.

[0043] **FIG. 6** illustrates a perspective projection used to relate a position in three-dimensional space to a corresponding position in two-dimensional image plane.

[0044] **FIG. 7** illustrates a projection of a three-dimensional point on a rigid body.

[0045] **FIG. 8** illustrates a first orientation-dependent tool marker useful in performing tool tracking, utilizing aspects of the present invention.

[0046] **FIG. 9** illustrates a second orientation-dependent tool marker useful in performing tool tracking, utilizing aspects of the present invention.

[0047] **FIG. 10** illustrates a flow diagram of a computer model tool tracking method utilizing aspects of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0048] **FIG. 1** illustrates, as an example, a Minimally Invasive Robotic Surgical (MIRS) system **100** including a Console ("C") utilized by a Surgeon ("S") while performing a minimally invasive diagnostic or surgical procedure, usually with assistance from one or more Assistants ("A"), on a Patient ("P") who is lying down on an Operating table ("O").

[0049] The Console includes a support **102**, a monitor **104** for displaying an image of a surgical site to the Surgeon, and one or more control devices **108.** The control devices **108** may include any one or more of a variety of input devices, such as joysticks, gloves, trigger-guns, hand-operated controllers, voice recognition devices or the like.

[0050] The Surgeon performs a procedure by manipulating the control devices **108** which in turn, cause robotic mechanisms **114** to manipulate their respective removably coupled instrument or tool assembly **110** (hereinafter simply referred to as a "tool") through a minimally invasive incision in the body of the Patient while the Surgeon views the surgical site through the monitor **104.** The number of surgical tools **110** used at one time and consequently, the number of robotic mechanisms **114** in the system **100** will generally depend on the diagnostic or surgical procedure and the space constraints within the operating room among other factors. If it is necessary to change one or more of the tools **110** being used during a procedure, the Assistant may remove the tool **110** no longer being used at the time from its robotic mechanism 114, and replace it with another tool **110** from a tray ("T") in the operating room.

[0051] The surgeon's Console is usually located in the same room as the Patient so that the Surgeon may directly monitor the procedure, be physically available if necessary, and speak to the Assistant directly rather than over the telephone or other communication medium. However, it will be understood that the Surgeon can also be located in a different room, a completely different building, or other remote location from the Patient allowing for remote surgical procedures.

[0052] Preferably, control devices **108** will be provided with the same degrees of freedom as their associated tools **110** to provide the Surgeon with telepresence, or the perception that the control devices **108** are integral with the tools **110** so that the Surgeon has a strong sense of directly controlling the tools **110.** To this end, position, force, and tactile feedback sensors (not shown) are preferably employed on the tools **110** to transmit position, force, and tactile sensations from the tools **110** back to the Surgeon's hands as he/she operates the control devices **108.**

[0053] Monitor **104** is suitably coupled to a viewing scope assembly **112,** including one or more cameras, through a processor **101,** and positioned on the support **102** of the Console such that an image of the surgical site is provided near the Surgeon's hands. Preferably, monitor **104** will display an inverted image on a display **106** that is oriented so that the surgeon feels that he or she is actually looking directly down onto the operating site. To that end, an image of the tools **110** appear to be located substantially where the operator's hands are located even though the observation points (i.e., the endoscope or viewing camera) may not be from the point of view of the image.

[0054] In addition, the real-time image is preferably transformed into a perspective image such that the operator can manipulate the end effector of a tool **110** through its corresponding control device **108** as if viewing the workspace in substantially true presence. By true presence, it is meant that the presentation of an image is a true perspective image simulating the viewpoint of an operator that is physically manipulating the tools **110.** Thus, the processor **101** (or another processor in the Console) transforms the coordinates of the tools **110** to a perceived position so that the perspective image is the image that one would see if the viewing scope assembly **112** was located directly behind the tools **110.**

[0055] The processor **101** performs various functions in the system **100.** Preferably it is used to transfer the mechanical motion of control devices **108** to robotic mechanisms **114** connected to tools **110** via control signals such as CS1 and CS2. In addition, it is preferably used to perform a tool tracking method that in turn, may be used to control movement of the viewing scope assembly **112** through its robotic mechanism **115** via control signal CS3 so that it tracks one or more of the tools **110,** as well as for other purposes such as those previously described. The processor **101** may be separate from or integrated as appropriate into the robotic mechanisms **114** and **115,** or it may be integrated in whole or in part into the Console serving as its processor or a co-processor to its processor.

[0056] The processor **101** also preferably provides force and torque feedback from the tools **110** to the hand-operated control devices **108.** In addition, it preferably performs a safety monitoring function that freezes or at least inhibits all robot motion in response to recognized conditions such as exertion of excessive force on the Patient or a "running away" of the robotic mechanisms **114** or **115.**

[0057] Although described as a processor, it is to be appreciated that the processor **101** may be implemented in practice by any combination of hardware, software and firmware. Also, its functions as described herein may be performed by one unit, or divided up among different components, each of which may be implemented in turn by any combination of hardware, software and firmware.

[0058] Additional details of the general operation and structure of the system **100** with respect to its manipulation and control of tool **110** are described, as an example, in commonly owned U.S. Pat. No. 6,346,072 entitled "Multi-Component Telepresence System and Method," which is incorporated herein by this reference.

[0059] **FIG. 2** illustrates, as examples, various reference frames associated with the MIRS system **100.** A world reference frame **201** is a fixed reference frame centered, for example, at a fixed point (i.e., a landmark) in an operating

room in which the minimally invasive diagnostic or surgical procedure is taking place. A tool reference frame **202,** on the other hand, is a moving reference frame that is centered, for example, on a fixed point on the tool and consequently, moves with the tool. Similarly, a camera reference frame **203** is also a moving reference frame that is centered, for example, on a fixed point on a camera and consequently, moves with the camera. The position and orientation of the tool reference frame **202** and the camera reference frame **203** relative to the world reference frame **201** are preferably determined from sensor data associated with robotic mechanisms used to manipulate their respective positions and orientations.

[0060] Another tool reference frame **204** represents the position and orientation of the tool reference frame as determined from the camera frame **203.** In the absence of systematic errors, tool reference frames **202** and **204** would exactly coincide. A patient reference frame **205,** is a semi-fixed reference frame that is centered, for example, on a fixed point on the patient and consequently, moves along with that point if the patient moves.

[0061] The camera defining the camera reference frame **203** is preferably a stereo camera that is calibrated so that each pair of corresponding points in its left and right image planes is mapped to a three-dimensional point in its camera reference frame **203.** The position of a tool as perceived by the camera (e.g., the tool reference frame **204)** may then be determined, for example, by first identifying the tool in the left and right image planes of the camera, then using the calibration information to determine the position of the tool in the camera reference frame **203.**

[0062] Generally, the camera reference frame **203** is associated with an endoscope inserted in the surgical site so as to be able to view the effector end of the tool during the surgical procedure. As indicated above, the endoscope in this case is preferably a stereoscopic endoscope. A second camera reference frame (not shown) may also be defined and associated with an external camera unit positioned outside of the patient, but in sufficiently close proximity so as to be able to view the rear end of the tool extending out of the patient during the surgical procedure. The external camera unit is also preferably a stereoscopic camera to facilitate 3-D determinations.

[0063] As previously described, both the tool and endoscope are preferably manipulated through incisions in the patient's body using robotic mechanisms. Each of these robotic mechanisms includes joints and linkages which allow movement of its respective tool or endoscope through five or six degrees of freedom.

[0064] The position and orientation of the tool (as depicted by the position and orientation of the tool reference frame **202)** and endoscope (as depicted by the position and orientation of the camera reference frame **203)** may be determined in the world reference frame **201** by attaching sensors to the joints and/or linkages of their respective robotic mechanisms to sense their movement. Such techniques are well known in robotics, along with the fact that their results depend upon the particular construction and operation of the robotic mechanism. Additional details may be found, for example, in John J. Craig, "Introduction to Robotics - Mechanics and Control," 2nd Ed., Addison Wesley Longman, 1986.

[0065] Another method for determining the tool and endoscope positions and orientations in the world reference frame 201 include sensing electromagnetic, acoustic, or other identifiable signals emanating or being reflected from the tool or endoscope so as to indicate its position and orientation. Still another method for determining the tool and endoscope positions and orientations in the world reference frame **201** include the use of the external stereoscopic camera described above, which may view the rear ends of the tool and endoscope extending out of the patient's body and determine its position and orientation through calculations based upon its left and right image planes.

[0066] By determining the positions and orientations of the endoscope (as depicted by the camera reference frame **203)** and tool (as depicted by the tool reference frame **202)** determined by sensors in the world reference frame **201,** the determination of the tool position and orientation in the world reference frame 201 can be checked using conventional triangulation techniques employing the endoscope position and orientation determined in the world reference frame **201** along with the tool position and orientation (as depicted by the tool reference frame **204)** determined in the camera reference frame **203** of the endoscope. Conversely, the determination of the endoscope position and orientation (as depicted by the camera reference frame **203)** in the world reference frame 201 can be checked using conventional triangulation techniques employing the tool position and orientation (as depicted by the tool reference frame **202)** determined in the world reference frame **201** along with the tool position and orientation (as depicted by the tool reference frame **204)** determined in the camera reference frame **203** of the endoscope. Having additional means for determining the tool and endoscope positions and orientations provide still more ways to check their respective positions and orientations and provide more accurate determinations.

[0067] **FIG. 3** illustrates, as an example, a flow diagram of a tool tracking method that tracks a tool by processing available sensor and image data of the tool, which were generated while the tool is inserted and being manipulated through a minimally invasive incision in a body. The sensor data in this case is from position sensors such as those employed for detecting joint positions in a robotic mechanism manipulating the tool, or those employed for detecting electromagnetic, acoustic, or other identifiable signals emanating or being reflected from the tool to indicate its position. In order to properly process the data together, both sensor and image data are preferably time stamped in some manner so that data associated with the same points in time can be processed with each other.

[0068] In **301,** a determination is made whether a tool is currently under the active control of a user, for example, by the corresponding control device being turned on by the user. If the determination in **301** is NO, then the method keeps

periodically looping through **301** until the determination in **301** results in a YES, at which time, in **302,** a determination is then made whether sensor data indicating the tool's state is available. If the determination in **302** is NO, then in **303,** a determination is made whether image data indicating the tool's state is available. If the determination in **303** is also NO, then no information is available for determining the tool's state at this time, and the method jumps back to **301** to start the process over again at another time. On the other hand, if the determination in **303** is YES, then in **304,** the state of the tool is determined using only the image data at this time, and following such determination, the method jumps back to **301** to restart the process for another time.

[0069] If the determination in **302** is YES, however, indicating that sensor data is available, then in **305,** a determination is made whether the position of the tool has changed since its last determination. It is advantageous to perform this determination at this time, because the determination is relatively easy and fast to perform, and if the tool hasn't moved, it avoids unnecessary computation to determine the new tool position and orientation.

[0070] As an example of one technique for performing the determination in **305:** (i) joint velocities are determined from sensors employed on the joints of the robotic mechanism that is manipulating the tool, (ii) the joint velocities are each squared, (iii) the squared joint velocities are summed together, and (iv) the resulting value is compared against a threshold value so that the tool is determined to have moved only if the resulting value is greater than the threshold value.

[0071] If the determination in **305** is NO, then the method jumps back to **301** to start the process over again at another time. On the other hand, if the determination in **305** is YES, then in **306,** a determination is made whether image data indicating the tool's state is available. If the determination in **306** is NO, then in **308,** the state of the tool is determined using only the sensor data at this time, and following such determination, the method then jumps back to **301** to restart the process at another time. However, if the determination in **306** is YES, then in **307,** the state of the tool is determined using both the sensor and image data at this time, and following such determination, the method then jumps back to **301** to restart the process for another time.

[0072] In the method described above in reference to **FIG. 3,** there is no assumption made regarding the availability of either the sensor or image data. Therefore, the method described therein checks independently whether both types of data are available and determines the tool's position accordingly. Although the availability of sensor data is checked before checking the availability of image data in this method, it is to be appreciated that this checking procedure can be reversed and still be within the scope of this aspect of the present invention.

[0073] The sensor data may be received from sensors or encoders positioned on joints and/or linkages on the robotic mechanism manipulating the tool during a surgical procedure, or the sensor data may be received from sensors detecting electromagnetic, acoustic, or other identifiable signals emanating or being reflected from the tool so as to indicate its position. The image data may be received from an endoscope viewing the effector end of the tool within the patient's body, or the image data may be received from an external camera viewing the exposed end of the tool extending outside of the patient's body during the surgical procedure.

[0074] **FIG. 4** illustrates, as an example, a tool tracking method for determining tool states using both tool sensor and available tool image data. In the method, **401** and **402** are generally performed off-line prior to a surgical procedure, and **403-410** are performed on-line during the surgical procedure.

[0075] In **401,** a stereoscopic endoscope is calibrated off-line so that mapping of points may be performed between its left and right 2-D image planes and the 3-D camera frame **203.** Calibration in this case includes determining the intrinsic camera parameters such as the focal length (e.g., see length "F" in **FIG. 6)**, principal point, skew, and distortion for each camera of the stereo pair. Additionally, the rotation and translation between the two cameras in the stereo pair may be determined.

[0076] Camera calibration may be performed once such as during the initial set-up of the MIRS system **100,** or it may be performed periodically such as just prior to performing a minimally invasive diagnostic or surgical procedure. One technique for performing the calibration is to capture several images of a calibration grid in a variety of positions and orientations. These images may then be fed into a commercially available (or home grown) calibration package that extracts the corner positions in the images of the calibration grid, and performs a calibration/optimization procedure to obtain the required parameters.

[0077] In **402,** initial values for error transforms are determined at the start of each surgical procedure. The error transform is defined as the position and orientation of a second estimated tool state (such as depicted as the tool reference frame **204)** determined in the camera reference frame **203** from left and right image planes of the stereoscopic camera, in the reference frame of a first estimated tool state (such as depicted as the tool reference frame **202)** determined in the world reference frame **201** from sensor data and translated to the camera reference frame **203** of a stereoscopic camera. Following the transform notation defined in "Introduction to Robotics - Mechanics and Control" previously referenced, this may be represented as ${}^{202}_{204}T$.

[0078] This procedure begins, for example, by applying an initial translation to the tool location from sensor data such that the perspective projection of this translated tool location is centered in the left and right images. The user then

moves the tool to the four corners of the stereoscopic images and clicks on the 2D tool locations in both left and right images. Combining the initial translation, the tool's 3-D sensor locations, and the tool's 2-D image locations, gives an initial value of the error transform ${}^{202}_{204}T$ between the tool location (such as depicted as the tool reference frame 202) as determined from the sensors and the tool location (such as depicted as the tool reference frame **204)** as determined from the stereoscopic image. After their determinations, the initial values are loaded into short-term memory just prior to initiating **403-410.**

[0079] In **403,** the tool state in the world reference frame 201 is determined from the sensor data. For the nominal case when the tool state consists only of position and orientation, this may be represented as the transform ${}^{201}_{202}T$ (see "Introduction to Robotics - Mechanics and Control," for details). This may be determined, for example, from system kinematics using data provided from joint or linkage position sensors associated with a robotic mechanism manipulating the tool, or by calculating the tool position in the world reference frame **201** using signals emanating or being reflected back from the tool indicating its position. Although both examples provide state information, the use of system kinematics is preferable for performing this function, because it generally provides more tool state information than the tool position indicating signals.

[0080] In **404,** the camera state in the world reference frame **201** is also determined from the sensor data. For the nominal case when the tool state consists only of position and orientation, this may be represented as the transform ${}^{201}_{203}T$ (see "Introduction to Robotics.- Mechanics and Control," for details). As in the case of the tool state determination in **403,** this may be determined, for example, from system kinematics using data provided from joint or linkage position sensors associated with a robotic mechanism manipulating the camera, or by calculating the camera position in the world reference frame **201** using signals emanating or being reflected back from the camera indicating its position.

[0081] In **405,** the estimated tool state determined in **403** is translated from the world reference frame **201** to the camera reference frame **203** of the camera using conventional reference frame transformation techniques using the estimated camera state (in the world reference frame **201)** determined in **404.**

$$ {}^{203}_{202}T = {}^{201}_{203}T^{-1}\,{}^{201}_{202}T \qquad (1) $$

[0082] In **406,** a determination is made whether image data of the tool is available for the corresponding point in time that the tool state was determined in **403.** The image data may not be available if either an image was not captured by the camera for the corresponding point in time, or the tool is not identifiable in an image captured by the camera for the corresponding point in time. The latter case may happen as the tool moves in and out of the camera's view due to manipulation of the tool by the surgeon during the surgical procedure. To assist with identifying tools in a captured image, various tool identification techniques may be used, including the use of special markers as described herein.

[0083] As a refinement to the above, even if the tool is identifiable in the image data, it may be identified as an outlier and therefore, rejected if its state falls outside of a tolerance range of a best fit curve generated from previously determined tool states at prior time points.

[0084] If the determination in **406** is YES, then in **407,** an estimate of the tool state (such as depicted as the tool reference frame **204)** in the camera reference frame **203** is determined directly. For the nominal case when the tool state consists only of position and orientation, this may be represented as the transform ${}^{203}_{204}T$. As an example of one technique for performing this task, a control point is identified on the tool in both the left and right 2-D image planes received from the stereoscopic camera, then the corresponding location of that point in the 3-D camera reference frame **203** using the previously generated calibration data is determined.

[0085] As an example of how the control point may be identified in the two image planes, a small window including the control point may be selected in the left image, and cross correlated with small windows in the right image to determine a window in that image with a highest correlation factor, which results in a match and identification of the control point in that image.

[0086] After determining the tool state in **407,** a revised error transform is calculated in **408** as the transform between the tool state determined in **405** and the tool state determined in **407,** and stored in the short-term memory replacing any initial value stored therein. The estimated tool state determined in **407** is then determined in **409** to be the tool state

for that point in time. The method then jumps back to **403** to determine the tool state in the camera reference frame **203** for another point in time.

$$ {}^{202}_{204}T = {}^{203}_{202}T^{-1}\,{}^{203}_{204}T \qquad (2) $$

[0087] On the other hand, if the determination in **406** is NO, then in **410,** the tool state (such as depicted as the tool reference frame **204)** in the camera reference frame **203** is then determined to be the estimated tool state determined in **405** adjusted by the error transform corresponding to that tool state as determined in **402.** In this case, the error transform is not updated. The method then jumps back to **403** to determine the tool state in the camera reference frame **203** for another point in time.

$$ {}^{203}_{204}T = {}^{203}_{202}T\,{}^{202}_{204}T \qquad (3) $$

[0088] Note that the error transforms may be updated during the surgical procedure in performing task **408,** because the error transforms may slowly drift over time due to a variety of factors such as errors in the initial error transform estimate, initial correlation, system kinematics, or camera calibration, as well as other factors such as external forces applied to the tool or its manipulating robotic mechanism, or hysteresis or other non-linearities in the robotic mechanism.

[0089] **FIG.** 5 illustrates, as an example, a functional block diagram for an alternative tool tracking method. The tool tracking method in this case employs an Extended Kalman Filter ("EKF"), which has the purpose of producing an optimal estimate of the state of the tool being tracked, $\hat{x}_k$, by combining one or more non-endoscopically derived tool state information, e.g., $z_{NV1\text{-}k}$ and $z_{NV2\text{-}k}$ (respectively generated in blocks **501** and **502,** by processing sensor and/or external camera data associated with the tool) with one or more endoscopically derived tool state information, such as $z_{V1\text{-}k}$, $z_{V2\text{-}k}$, and $z_{V3\text{-}k}$ (respectively generated in blocks **511, 512,** and **513,** by processing image data of the tool received from a stereoscopic endoscope using corresponding vision algorithms), and using a model of the system dynamics.

[0090] In this method, determinations of whether or not non-endoscopically derived or endoscopically-derived tool state information are available are not necessary (such as performed in **302** and **306** in **FIG. 3** for sensor data and image data), because if either type of information is unavailable for the time of computation, this fact is simply taken into account by the unavailable information consisting of measurement values remaining at a previous value (i.e., not being updated at the time). For this and other reasons, the method described in reference to **FIG**. **5** is considered a preferred technique for determining tool states.

[0091] Note that although an EKF is used in this example, other Bayesian filters or techniques such as a Kalman Filter or Particle Filter may also be employed and are fully contemplated to be within the scope of the present invention. A Bayesian filter refers to a family of statistical filters based on Bayesian estimation techniques.

[0092] Also, note that all inputs referred to as sensor data, external camera data, or endoscopic image data may be thought of as measurements, wherein the quantity being measured is some subset of the tool state (typically position and orientation) in the corresponding sensor, external camera or endoscope reference frame. In general, all the sensor measurements of the tool state include some amount of processing such as a forward kinematics computation is generally required to compute the tool state from the robot joint position sensors. Further, all measurements are generally asynchronous, but referred to an absolute clock by a time stamp.

[0093] Each measurement is generally corrupted by some random error, such as zero mean noise, and can be either unavailable at some point in time (missing) or can be totally erroneous (outlier). The EKF thus reduces the effect of the measurement noise on the tool state estimate. The nominal transformation between each sensor, external camera or endoscope reference frame is used to fuse the measurements.

[0094] Function block **501** generates non-endoscopically derived tool state information., $z_{NV1\text{-}k}$, from sensor data, and provides the non-endoscopically derived tool state information to the EKF **521** for processing. As previously described, the non-endoscopically derived tool state information may be from joint position sensors, tool position signal detectors, or external cameras. Additional function blocks, such as function block **502,** may optionally be included to generate additional non-endoscopically derived tool state information, such as $z_{NV2\text{-}k}$, from the same or other sensor data or external camera, and provide the additional non-endoscopically derived tool state information to the EKF 521 for processing.

[0095] On the other hand, function block **511** generates endoscopically derived tool state information, $z_{V1\text{-}k}$, from endoscope image data, and provides the endoscopically derived tool state information to the EKF **521** for processing.

As previously described, the endoscope image data may be the left and right image planes from a stereoscopic endoscope. Additional function blocks, such as function blocks **512** and **513,** may optionally be included to generate additional endoscopically derived tool state information, such as $z_{V2\text{-}k}$ and $z_{V2\text{-}k}$, generally from the same endoscope image data, and provide the additional endoscopically derived tool state information to the EKF **521** for processing.

[0096] The function blocks **501-502** and **511-513** perform some common tasks, as well as their individual special processing to generate their respective tool state information. As an example, each of the function blocks keeps track of time stamps placed on its received sensor or image data, which indicates when the data was sensed or captured, so that all tool state information provided by the function blocks at any given time to the EKF **521** corresponds approximately to the same time indicated on the time stamps. As another example, each of the function blocks preferably filters noise in its received sensor or image data so that the tool state information provided to the EKF **521** has approximately zero mean noise.

[0097] Following is a simplified example of how the EKF **521** is used for tool tracking during a minimally invasive diagnostic or surgery procedure. For a more thorough description of the EKF in general, see, for example, Greg Welch and Gary Bishop, "An Introduction to the Kalman Filter," TR 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, April 5, 2004.

[0098] As is well known, the EKF framework has two distinct phases, termed the "Time Update" (or "Predict") phase and the "Measurement Update" (or "Correct") phase.

[0099] In a first phase of the Kalman filter update cycle, the state estimate from the previous iteration of the filter is used to produce a prediction of the new state estimate for this cycle, $\hat{\mathbf{x}}_k^-$, based on a (potentially) nonlinear model of the system dynamics, *f*, and a forcing function, $\mathbf{u}_{k\text{-}1}$, according to equation (4) as follows:

$$\hat{\mathbf{x}}_k^- = f(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}, 0) \qquad (4)$$

wherein the '0' in equation (4) is the process noise estimate used in producing the predicted state estimate.

[0100] Linearizing the nonlinear system model, *f*, about the point $\hat{\mathbf{x}}_{k\text{-}1}$, the linearized state transition matrix, **A,** is obtained. Then, using the error covariance matrix from the previous iteration of the filter, $\mathbf{P}_{k\text{-}1,}$ a *prediction* of the new error covariance matrix is produced for this cycle, $\mathbf{P}_k^-$, according to equation (5) as follows:

$$\mathbf{P}_k^- = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{W}_k \mathbf{Q}_{k-1} \mathbf{W}_k^T \qquad (5)$$

wherein the matrix **W** represents the Jacobian of the system dynamics with respect to the process noise, **w,** and **Q** is a tunable gain matrix.

[0101] Thinking physically, if the matrix **P** describes an error ellipse having a number of dimensions equal to the system state order, then by applying equation (5), the size of the error ellipse may be expanded using our model of the system dynamics encoded in the linearized state transition matrix, **A,** and the scaled estimate of the process noise, which represents uncertainty in the system dynamics model. Increasing the size of the error ellipse is equivalent to stating that there is greater uncertainty in the estimate of the system state.

[0102] In the second phase of the Kalman filter update cycle, the predicted state estimate and predicted error covariance matrix may be corrected by taking one or more measurements of the system.

[0103] The Kalman gain is then computed. The Kalman gain, in essence, weights the contributions from one or more measurements, such that their impact on the new state estimate reflects a current estimate of their reliability. Additionally, it allows weighting of the reliance on the model vs. the measurements. In other words, the contribution from a reliable measurement may be weighted more, and an unreliable measurement less. To do so, equation (6) is applied as follows.

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{V}_k \mathbf{R}_k \mathbf{V}_k^T)^{-1} \qquad (6)$$

[0104] To explain equation (6), it is first proposed that there is a function h, known as the measurement function, which relates the quantities that are able to be measured (observed) in the system, **z,** to the actual system state, **x**.

[0105] In equation (6), the estimate of the new error covariance matrix, $\mathbf{P}_k^-$, from equation (5) is made use of. The matrix **H** is the Jacobian of the 'measurement' function, *h*, with respect to the system state, **x**. In essence, the **H** matrix describes how a change in the quantity being measured (observed), will change the actual system state.

[0106] The matrix **V** represents an estimate of the measurement noise, which includes both sensor noise, and uncertainty in the measurement function, *h*. The **R** matrix is a tunable gain.

[0107] Once the Kalman gain is computed according to equation (6), the estimate of the system state may be updated, using the predicted system state estimate, $\hat{\mathbf{x}}_k^-$, and the measurements. In this case, the predicted system state estimate is adjusted by the Kalman-gain-weighted error between the actual measurement, $\mathbf{z}_k$, and the predicted measurement, $\hat{\mathbf{z}}_k$, according to equation (7) as follows.

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - h(\hat{\mathbf{x}}_k^-, 0)) \qquad (7)$$

[0108] Finally, the estimate of the error covariance is updated according to equation (8) as follows.

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^- \qquad (8)$$

[0109] The value of $\mathbf{P}_k$ should decrease at a rate proportional to the degree to which the measurements are trusted vs. the predictions. Physically, this implies that the error ellipse enclosing the system state estimate shrinks as additional measurements are obtained.

[0110] Now that the EKF framework has been described, its formulation with respect to the present application is detailed. In particular, the system state, **x**, the system state function, *f*, the state transition matrix, **A,** the measurements, **z**, the measurement functions, ***h***, and the measurement Jacobians, **H,** are defined.

[0111] The state of the system is the position, orientation, translational velocity, and rotational velocity of the laparoscopic tool end effector, as shown in equation (9). In this way the state of a rotating rigid body is described.

$$\mathbf{x} = \begin{bmatrix} x & y & z & \theta_x & \theta_y & \theta_z & \theta_w & \dot{x} & \dot{y} & \dot{z} & \omega_x & \omega_y & \omega_z \end{bmatrix}^T_{13x1} \qquad (9)$$

[0112] Note that the orientation of the end effector, Θ**,** is represented using quaternion notation, rather than as a rotation matrix. This facilitates a smooth integration of the angular velocity to obtain the new rigid body orientation, as required by the system state update equations.

[0113] For the system update function, *f*, a rigid body in free-space is described, with no forcing function input. The state propagation is described by equation set (10). Note, however, that in the preferred embodiment, the velocities are pre-multiplied by the ΔT between samples to obtain position deltas, and simply added in equation (10).

[0114] Here, an assumption is made of no forcing function because the time step is very small, and any acceleration imparted to the rigid body can be modeled as noise in the system model.

$$\mathbf{x}_k^- = \begin{bmatrix} \hat{x}_k^- = \hat{x}_{k-1} + \hat{\dot{x}}_{k-1}\Delta T \\ \hat{y}_k^- = \hat{y}_{k-1} + \hat{\dot{y}}_{k-1}\Delta T \\ \hat{z}_k^- = \hat{z}_{k-1} + \hat{\dot{z}}_{k-1}\Delta T \\ \hat{\Theta}_k^- \\ \hat{\dot{x}}_k^- = \hat{\dot{x}}_{k-1} \\ \hat{\dot{y}}_k^- = \hat{\dot{y}}_{k-1} \\ \hat{\dot{z}} = \hat{\dot{z}}_{k-1} \\ \hat{\omega}_{xk}^- = \hat{\omega}_{xk-1} \\ \hat{\omega}_{yk}^- = \hat{\omega}_{yk-1} \\ \hat{\omega}_{zk}^- = \hat{\omega}_{zk-1} \end{bmatrix}_{13x1} \quad (10)$$

[0115] The state propagation for the orientation state, Θ, is a bit more complicated due to the use of quaternions. First, the quaternion derivative, which is a function of the current orientation and the angular velocity, is computed. To do so, the angular velocity quaternion, $\mathbf{q_1}$, and the orientation quaternion, $\mathbf{q_2}$, are described as shown in equation set (11).

$$\begin{aligned} \mathbf{q}_1 &= \left[\hat{\omega}_{xk-1}\hat{\omega}_{yk-1}\hat{\omega}_{zk-1}0\right]_{4x1}^T \\ \mathbf{q}_2 &= \left[\hat{\theta}_{x\,k-1}\hat{\theta}_{y\,k-1}\hat{\theta}_{z\,k-1}\hat{\theta}_{w\,k-1}\right]_{4x1}^T \end{aligned} \quad (11)$$

[0116] The quaternion derivative is then calculated according to equation (12):

$$d\mathbf{q} = \frac{1}{2}\mathbf{q}_2 * \mathbf{q}_1 \quad (12)$$

where the '*' operator denotes quaternion multiplication. Once the quaternion derivative has been computed, integration per equation (13) is performed.

$$\hat{\Theta}_k^- = \begin{bmatrix} \hat{\theta}_{xk-1} + dq_x\Delta T \\ \hat{\theta}_{yk-1} + dq_y\Delta T \\ \hat{\theta}_{zk-1} + dq_z\Delta T \\ \hat{\theta}_{wk-1} + dq_w\Delta T \end{bmatrix}_{4x1} \quad (13)$$

[0117] After integration, the resulting quaternion is enforced to be of unit length by normalizing $\hat{\Theta}_k^-$. Having defined

the system state function, f, equation (4) of the Kalman filter update cycle may be computed.

[0118] To satisfy equation (5) of the Kalman filter update cycle, the system state Jacobian matrix, **A,** must be defined. In the present case, if the above discussion is re-written in matrix form, the required matrix results as shown in equation (14).

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_{3x3} & \mathbf{0}_{3x4} & \Delta T \cdot \mathbf{I}_{3x3} & \mathbf{0}_{3x3} \\ \mathbf{0}_{4x3} & \mathbf{A}_q & \mathbf{0}_{4x3} & \mathbf{0}_{4x3} \\ \mathbf{0}_{3x3} & \mathbf{0}_{3x4} & \mathbf{I}_{3x3} & \mathbf{0}_{3x3} \\ \mathbf{0}_{3x3} & \mathbf{0}_{3x4} & \mathbf{0}_{3x3} & \mathbf{I}_{3x3} \end{bmatrix}_{13x13} \qquad (14)$$

[0119] The matrix $\mathbf{0}_{mxm}$ is an mxn matrix of 0's. The matrix $\mathbf{A}_q$, shown in equation (15), is the quaternion omega matrix, which encodes the quaternion multiplication operation, '*', described in equation (12).

$$\mathbf{A}_q = \frac{1}{2}\begin{bmatrix} 0 & -\hat{\omega}_{zk-1} & \hat{\omega}_{yk-1} & \hat{\omega}_{xk-1} \\ \hat{\omega}_{zk-1} & 0 & -\hat{\omega}_{xk-1} & \hat{\omega}_{yk-1} \\ -\hat{\omega}_{yk-1} & \hat{\omega}_{xk-1} & 0 & \hat{\omega}_{zk-1} \\ -\hat{\omega}_{xk-1} & -\hat{\omega}_{yk-1} & -\hat{\omega}_{zk-1} & 0 \end{bmatrix}_{4x4} \qquad (15)$$

[0120] Having defined the system state Jacobian matrix, **A,** equation (5) of the Kalman filter cycle update may now be computed.

[0121] In the present system, multiple measurement sources are provided. The first measurement as provided by function block 501, is in this case from the robot kinematics (encoders or position sensors), and is of the state directly, as shown in equation (16).

$$\mathbf{z}_{NV1-k} = \begin{bmatrix} x & y & z & \theta_x & \theta_y & \theta_z & \theta_w & \dot{x} & \dot{y} & \dot{z} & \omega_x & \omega_y & \omega_z \end{bmatrix}^T_{13x1} \qquad (16)$$

[0122] Note that the end effector orientation has been converted from a rotation matrix to a quaternion, to fit within this framework. Also, note that in case the measurement from the robot kinematics is referred to a reference frame other than the camera reference frame of the endoscope, equation (16) would need to be modified so as to accommodate transformation to the camera reference frame.

[0123] The second measurement as provided by function **511** is obtained by processing the left and right images provided by the stereoscopic endoscope. The coordinates (u_l,v_l) and (u_r,v_r) are the position of the end-effector in the left and right image planes, respectively.

$$\mathbf{z}_{V1-k} = \begin{bmatrix} u_l & v_l & u_r & v_r \end{bmatrix}^T_{4x1} \qquad (17)$$

[0124] Before combining non-endoscopically derived tool state information and endoscopically derived tool state information according to equation (18) below, it is first useful to verify that the endoscopically derived estimates of tool state are not outliers.

[0125] Assuming that they are not outliers, to form the full measurement vector, the two measurements are then stacked as shown in equation (18).

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{z}_{NV1-k} \\ \mathbf{z}_{V1-k} \end{bmatrix}_{17x1} \quad (18)$$

[0126] Note that additional non-endoscopic processing algorithms may be run, such as in block **502,** each resulting in a set of measurements, such as $\mathbf{z}_{NV2-k}$, and additional endoscopic processing algorithms on the endoscope images may be run, such as in blocks **512** and **513,** each resulting in a set of measurements, such as $\mathbf{z}_{V2-k}$ and $\mathbf{z}_{V3-k}$, all of which may in turn be stacked in the measurement vector.

[0127] To satisfy equation (7) of the Kalman filter update cycle, the measurement functions, *h,* must be defined, which relate the system state **x** to the measured (observed) quantities, **z.**

[0128] As previously noted, there are multiple measurement sources in the present system. Thus, a measurement function, *h,* is required for each source.

[0129] As the kinematic (encoder or position sensor) measurement is of the state directly, the measurement function, h_{NV1-k}, is identity, with the exception of a conversion from rotation-matrix to quaternion for the end effector orientation.

[0130] The second measurement, from the endoscope images, is the position (u,v) of the endoscope end-effector in the left and right image plane. The measurement function in this case is the perspective projection function, which relates a position in 3-D space to a position in image space, as shown in **FIG. 6.**

[0131] If a point P is assumed in the camera reference frame (of the endoscope) with coordinates (x,y,z), then its projection using a pinhole camera model is given by equation (19).

$$\begin{aligned} u_{cam} &= x/z \\ v_{cam} &= y/z \end{aligned} \quad (19)$$

[0132] If the camera is allowed to have radial lens distortion, then the distorted u,v coordinates are given by equation (20), where $r^2 = \left(\frac{x}{z}\right)^2 + \left(\frac{y}{z}\right)^2$, and k_n are the radial distortion coefficients of the camera calibration.

$$\begin{aligned} u_{cam} &= \frac{x}{z}\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \\ v_{cam} &= \frac{y}{z}\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \end{aligned} \quad (20)$$

[0133] Conversion to pixel coordinates is then performed by multiplying by the focal length, f, expressed in pixels, where c is the optical center, as given by equation (21).

$$\begin{aligned} u &= \mathrm{f}_x u_{cam} + c_x \\ v &= \mathrm{f}_y v_{cam} + c_y \end{aligned} \quad (21)$$

[0134] The subscript x or y denotes the x or y component of the focal length or optical center. Note that in this discussion the contribution of tangential or skew distortion has been ignored, which would have added extra terms to the above equations.

[0135] Having defined the measurement functions, *h,* the partial derivatives of *h* with respect to the state are taken to satisfy equation (7) of the Kalman filter update cycle.

[0136] As previously described, the kinematics measurement is of the state directly. Therefore, the measurement Jacobian for the kinematics measurement is the 13x13 identity matrix, (22).

$$\mathbf{H}_{NV1-k} = \mathbf{I}_{13x13} \tag{22}$$

[0137] Taking the partial of the vision measurement function with respect to the state, the image Jacobian is obtained, which relates end-effector translational and rotational velocities to image-space velocities.

[0138] By way of introduction, the equation for a point moving in 3-D space may be seen in equations (23).

$$\begin{aligned} \dot{u} &= \frac{\mathrm{f}_x \dot{x}}{z} - \frac{\mathrm{f}_x x \dot{z}}{z^2} \\ \dot{v} &= \frac{\mathrm{f}_y \dot{y}}{z} - \frac{\mathrm{f}_y y \dot{z}}{z^2} \end{aligned} \tag{23}$$

[0139] When extended to a stereo vision system, where the observed point is off-axis from the center of rotation, equation set (24) is obtained. For a derivation of this equation set, see, e.g., Bijoy K. Ghosh, Ning Xi, T.J. Tarn, "Control in Robotics and Automation: Sensor Based Integration", Academic Press, San Diego, 1999.

$$\mathbf{H}_{V1-k} = \begin{bmatrix} \mathbf{0}_{4x7} & \begin{matrix} H_{visk1,8} & H_{visk1,9} & H_{visk1,10} & H_{visk1,11} & H_{visk1,12} & H_{visk1,13} \\ H_{visk2,8} & H_{visk1,9} & H_{visk1,10} & H_{visk1,11} & H_{visk1,12} & H_{visk1,13} \\ H_{visk2,8} & H_{visk1,9} & H_{visk1,10} & H_{visk1,11} & H_{visk1,12} & H_{visk1,13} \\ H_{visk2,8} & H_{visk1,9} & H_{visk1,10} & H_{visk1,11} & H_{visk1,12} & H_{visk1,13} \end{matrix} \end{bmatrix}_{4x13} \tag{24}$$

where:

$$\begin{aligned} H_{visk1,8} &= \mathrm{f}_{xl}/Z_c \\ H_{visk1,9} &= 0 \\ H_{visk1,10} &= -\frac{\mathrm{f}_{xl}\left(X_c + b/2\right)}{Z_c^2} \\ H_{visk1,11} &= -\frac{\mathrm{f}_{xl}\left(X_c + b/2\right)Y_t}{Z_c^2} \\ H_{visk1,12} &= \frac{\mathrm{f}_{xl}Z_t}{Z_c} + \frac{\mathrm{f}_{xl}\left(X_c + b/2\right)\left(X_t + b/2\right)}{Z_c^2} \\ H_{vis1,13} &= -\frac{\mathrm{f}_{xl}Y_t}{Z_c} \end{aligned} \qquad \begin{aligned} H_{visk2,8} &= 0 \\ H_{visk2,9} &= \frac{\mathrm{f}_{yl}}{Z_c} \\ H_{visk2,10} &= -\frac{\mathrm{f}_{yl}Y_c}{Z_c^2} \\ H_{visk2,11} &= -\frac{\mathrm{f}_{yl}Z_t}{Z_c} - \frac{\mathrm{f}_{yl}Y_cY_t}{Z_c^2} \\ H_{visk2,12} &= \frac{\mathrm{f}_{yl}Y_c\left(X_t + b/2\right)}{Z_c^2} \\ H_{visk2,13} &= \frac{\mathrm{f}_{yl}\left(X_t + b/2\right)}{Z_c} \end{aligned} \tag{25}$$

and

$$
\begin{aligned}
H_{visk3,8} &= \mathrm{f}_{xr}/Z_c & H_{visk4,8} &= 0 \\
H_{visk3,9} &= 0 & H_{visk4,9} &= \frac{\mathrm{f}_{yr}}{Z_c} \\
H_{visk3,10} &= -\frac{\mathrm{f}_{xr}(X_c - b/2)}{Z_c^2} & H_{visk4,10} &= -\frac{\mathrm{f}_{yr}Y_c}{Z_c^2} \\
H_{visk3,11} &= -\frac{\mathrm{f}_{xr}(X_c - b/2)Y_t}{Z_c^2} & H_{visk4,11} &= -\frac{\mathrm{f}_{yr}Z_t}{Z_c} - \frac{\mathrm{f}_{yr}Y_cY_t}{Z_c^2} \\
H_{visk3,12} &= \frac{\mathrm{f}_{xr}Z_t}{Z_c} + \frac{\mathrm{f}_{xr}(X_c - b/2)(X_t - b/2)}{Z_c^2} & H_{visk4,12} &= \frac{\mathrm{f}_{yr}Y_c(X_t - b/2)}{Z_c^2} \\
H_{vis3,13} &= -\frac{\mathrm{f}_{xr}Y_t}{Z_c} & H_{visk4,13} &= \frac{\mathrm{f}_{yr}(X_t - b/2)}{Z_c}
\end{aligned}
\qquad (26)
$$

[0140] In the above equations (25) and (26), the variable 'b' refers to the baseline distance between the stereo cameras of the endoscope. The subscript 'l' or 'r' refers to the left or right camera, respectively. X_c, Y_c, Z_c refer to the origin of the coordinate frame about which the rigid body is being rotated, and X_t, Y_t, Z_t refer to the point of interest relative to the rotation origin, affixed to the rigid body, as shown in **FIG. 7** wherein X_c, Y_c, Z_c are the X,Y,Z coordinates of the center point P_c and X_t, Y_t, Z_t are the X,Y,Z coordinates of the point P_T.

[0141] Referring to **FIG. 7,** the point P_c is the center of the pin through the tool clevis. For example, a clevis surface may be painted green, and the centroid of the resulting color blob may be tracked. The centroid of that blob would then be point P_T in **FIG. 7.**

[0142] Based on the dot product of the end effector Y axis with the endoscope view vector, it can be determined which face of the clevis is being presented to the camera, and the sign of the distance Y_t may be adjusted appropriately.

[0143] Thus, the image plane velocities can be obtained by multiplying equation (24) by the state estimate prediction, as shown as in equation (27).

$$
\begin{bmatrix} \dot{u}_l \\ \dot{v}_l \\ \dot{u}_r \\ \dot{v}_r \end{bmatrix}_{4x1} = \mathbf{H}_{V1-k}\hat{\mathbf{x}}_k^- \qquad (27)
$$

[0144] To obtain the full **H** matrix for use in equations (3) and (5) of the Kalman filter update cycle, the two measurement Jacobians are stacked, as shown in equation (28).

$$
\mathbf{H}_k = \begin{bmatrix} \mathbf{H}_{NV1-k} \\ \mathbf{H}_{V1-k} \end{bmatrix}_{17x13} \qquad (28)
$$

[0145] As may be seen in **FIG. 5,** the architecture of the Kalman filter allows for several extensions, shown as dotted lines.

[0146] First, multiple endoscope video processing algorithms may be used to obtain several estimates of the tool's position in the endoscope generated image, such as shown by blocks **511-513.** These may be stacked into the measurement vector, as shown in equation (18).

[0147] The endoscope vision processing algorithm, as drawn in **FIG. 5,** finds the tool in the image by brute force. Preferably, however, the output of the endoscope vision processing algorithm may be fed back, to reduce the search

space by first looking in the area where the tool was previously found.

[0148] The state estimate output of the EKF **521** might be used, either together with the endoscope video processing output, or by itself, to reduce the search space in a similar manner. On the other hand, the endoscope video processing algorithm might simply make use of the raw kinematic input instead.

[0149] Finally, the endoscope video processing algorithm might make use of the error covariance matrix, to dynamically scale the search region based on confidence in the true tool position. This may enhance the performance of the tool tracking by bounding its search region by the error ellipse described by the error covariance matrix.

[0150] In the reverse direction, the gain matrix **R** may be updated if the vision processing system is able to provide a quality measure with regard to its measurement. Similarly, if the tool or an arm of its robotic mechanism has force sensors mounted on or embedded in it, then by observing the error between the actual and commanded positions, a quality measure for the kinematic information may be determined. Errors in the kinematic state measurement in this case may arise if applied forces result in deflection of the tool and/or the arm of its robotic mechanism. This quality measure may then be used to dynamically adjust gains in the EKF **521,** so as to appropriately weight the kinematic state measurement's contribution to the state estimate.

[0151] Note that the EKF **521,** as formulated, works in the camera reference frame **203.** However, there is no requirement that this be so. In fact, it may be advantageous to use the world reference frame **201,** especially if it is desired to register pre-operative or other synthetic image data with the camera image data.

[0152] Also note that in the above formulation, the stereo geometry of the stereoscopic endoscope is expressed explicitly in the $\mathbf{H}_{V1\text{-}k}$ matrix (which is really the partial of the perspective projection operation). Alternatively, a stereo correlation may be performed between the two images, the position and orientation of the tool in 3-D extracted, and the measurement fed back into the EKF **521** directly. The result in this case should be equivalent with the new measurement function $h_{V1\text{-}k}$ being "identity".

[0153] Following is an extension to the above formulation which allows for simultaneous tool tracking and camera calibration. In the EKF formulation above, the camera parameters are assumed known or previously calibrated. The estimate of the tool end-effector position in 3-D is based upon the knowledge of these parameters through the vision measurement function, $h_{V1\text{-}k}$, as shown in the following equation:

$$\mathbf{z}_{V1-k} = h_{V1-k}(\mathbf{x}_k) + noise \qquad (29)$$

where $\mathbf{x}_k$ is the state to be estimated, i.e., the 3-D position, velocity, orientation, and angular velocity of the tool, as shown in the following equation:

$$\mathbf{x}_k = \begin{bmatrix} x & y & z & \theta_1 & \theta_2 & \theta_3 & \theta_4 & \dot{x} & \dot{y} & \dot{z} & \omega_x & \omega_y & \omega_z \end{bmatrix}^T \qquad (30)$$

[0154] The vision measurement $\mathbf{z}_{V1\text{-}k}$ is obtained by processing the left and right image planes to extract the (u,v) coordinates of the tool, as shown in the following equation:

$$\mathbf{z}_{V1-k} = \begin{bmatrix} u_l & v_l & u_r & v_r \end{bmatrix}^T \qquad (31)$$

[0155] In the previously described formulation, errors in the intrinsic camera parameters such as:

$K_{1\text{-}n}$: The radial distortion coefficients (left & right),

f_x, f_y: The focal length (left & right),

c_x, c_y: The camera optical center (left & right), and

R, T: The relative position/orientation of the two cameras, where R is a quaternion representation of the orientation,

all contribute to the error in the tool tracking.

[0156] Calibration of these parameters can be performed by knowing the actual 3-D position of an object in the field of view, such as, for example, the tool end-effector itself. Of course, such knowledge is unfortunately not available, as it is exactly what is to be estimated.

[0157] The following method achieves simultaneous tool tracking and camera calibration based on the idea that a sequence of vision measurements, $\mathbf{z}_{V1\text{-}k}$, will be explained in the best way, i.e. with minimum statistical error over time, by the combination of true camera parameters and true tool state.

[0158] A practical implementation of this can again be accomplished using an extended Kalman filter and expanding the state vector to include the unknown camera parameters (or a subset which needs to be estimated),

$$\mathbf{x}_{cam} = \begin{bmatrix} k_{1l,r} & \cdots & k_{nl,r} & f_{xl,r} & f_{yl,r} & c_{xl,r} & c_{yl,r} & \mathbf{R} & \mathbf{T} \end{bmatrix}^T \qquad (32)$$

with the tool state $\mathbf{x}_k$ from equation (29) which is now called $\mathbf{x}_{tool_k}$:

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_{tool_k} & \mathbf{x}_{cam_k} \end{bmatrix}^T \qquad (33)$$

[0159] The state transition function for the camera states is constant, i.e.:

$$\mathbf{x}_{cam_k} = \mathbf{x}_{cam_{k-1}} \qquad (34)$$

[0160] The vision measurement function $h_{V1\text{-}k}$ is unchanged, but its dependence on the camera parameters is made explicit so that the partial derivatives of $h_{V1\text{-}k}$ with respect to the state to be used in the Kalman filter update is:

$$\mathbf{H}_{V1-k} = \begin{bmatrix} \partial h_{V1-k} / \partial \mathbf{x}_{tool} & \partial h_{V1-k} / \partial \mathbf{x}_{cam} \end{bmatrix} \qquad (35)$$

[0161] Using equations (32) and (34), the EKF can be computed in the same way as previously described. All the same extensions relative to continuous or discrete update options still apply. The initial value for $\mathbf{x}_{cam}$ is to be set the best available guess and the gain matrices should weight the $\mathbf{x}_{cam}$ states proportionally to the uncertainty in the guess.

[0162] Following is an extension to the above formulation which allows for the presence of a slowly varying systematic error in the transformation between the tool state measurement from the robot kinematics and the camera reference frame. In the EKF formulation above, the measurements from the robot position sensors and from the endoscope are assumed to be expressed in the same frame of reference or alternatively the transformation between the frame of reference of each measurement ($\mathbf{R}_{err}$, $\mathbf{T}_{err}$) is supposed to be known. In this latter case the measurement function $h_{NV1\text{-}k}$, as shown in the following equation

$$\mathbf{z}_{NV1-k} = h_{NV1-k}(\mathbf{x}_k) + noise \qquad (36)$$

is easily obtained by comparison with the expression:

$$\mathbf{z}_{NV1-k} = \mathbf{U}\,\mathbf{x}_k + \begin{bmatrix} T_x & T_y & T_z & \mathbf{0}_{1x10} \end{bmatrix} \qquad (37)$$

where

$$\mathbf{U}=\begin{bmatrix} \mathbf{R}_{err3x3} & \mathbf{0}_{3x4} & \mathbf{0}_{3x3} & \mathbf{0}_{3x3} \\ \mathbf{0}_{4x3} & \mathbf{A}_{err} & \mathbf{0}_{4x3} & \mathbf{0}_{4x3} \\ \mathbf{0}_{3x3} & \mathbf{0}_{3x4} & \mathbf{R}_{err3x3} & \mathbf{0}_{3x3} \\ \mathbf{0}_{3x3} & \mathbf{0}_{3x4} & \mathbf{0}_{3x3} & \mathbf{R}_{err3x3} \end{bmatrix}_{13x13} \qquad (38)$$

with $\mathbf{A}_{err}$ the quaternion omega matrix associated to $\mathbf{R}_{err}$ and $\mathbf{T}_{err}$ = (Tx, Ty, Tz).

[0163] As an initial estimate of ($\mathbf{R}_{err}$, $\mathbf{T}_{err}$) it is possible to compute the transformation that at one specific initialization time transforms the stereo vision measurement of the tool position into the tool position provided by the robot kinematics.

[0164] The following method achieves simultaneous tool tracking and update of ($\mathbf{R}_{err}$, $\mathbf{T}_{err}$) based on the idea that a sequence of endoscopic vision measurements, $\mathbf{z}_{V1\text{-}k}$, will be explained in the best way, i.e. with minimum statistical error over time, by the combination of coherent measurements and the true tool state.

[0165] A practical implementation of this can again be accomplished using an extended Kalman filter and expanding the state vector to include the unknown parameters of the transformation ($\mathbf{R}_{err}$, $\mathbf{T}_{err}$) (or a subset which needs to be estimated),

$$\mathbf{x}_{err}=\begin{bmatrix}\mathbf{R}_{err} & \mathbf{T}_{err}\end{bmatrix}^T \qquad (39)$$

with $\mathbf{R}_{err}$ the quaternion representation of the transformation rotation matrix and with the tool state $\mathbf{x}_k$ from equation (36) which is called $\mathbf{x}_{tool_k}$:

$$\mathbf{x}_k=\begin{bmatrix}\mathbf{x}_{tool_k} & \mathbf{x}_{err_k}\end{bmatrix}^T \qquad (40)$$

[0166] The state transition function for the $\mathbf{x}_{err}$ states is constant, i.e.:

$$\mathbf{x}_{err_k}=\mathbf{x}_{err_{k-1}} \qquad (41)$$

[0167] The measurement function $h_{NV1\text{-}k}$ is unchanged as in equation (36), but its dependence on the parameters ($\mathbf{R}_{err}$, $\mathbf{T}_{err}$) is made explicit so that the partial derivatives of $h_{NV1\text{-}k}$ with respect to the state to be used in the Kalman filter update is:

$$\mathbf{H}_{NV1-k}=\begin{bmatrix}\partial h_{NV1-k}/\partial \mathbf{x}_{tool} & \partial h_{NV1-k}/\partial \mathbf{x}_{err}\end{bmatrix} \qquad (42)$$

[0168] The initial transforms for registering kinematic estimates into the camera reference frame at the current tool location as in **FIG. 2** are stored in long term memory prepared offline. This transform is copied into a short term memory and is continuously updated over time according equation (40). Since this transform varies according to tool locations, the whole endoscope and tool operation space are sliced into multiple cubic cells and one transform is associated with each cubic cell. Since a tool can move to a location with different joint combinations, the transform is optimized, by taking this into account, from data collected from multiple joint combinations in the target cell.

[0169] As previously mentioned, to assist with identifying tools in a captured image, various tool identification techniques

may be used, including the use of special markers. **FIGS. 8** and **9** illustrate, as examples, two orientation dependent tool markers useful in performing tool tracking. The unique feature of these tool markers is that they are orientation dependent, unlike previously disclosed markers used for tracking the camera and tool(s). Therefore, if the tool has rotated along its axis, these tool markers are designed to indicate such new orientation of the tool.

[0170] Preferably, the markers are drawn or formed on the effector-end of the tool. In **FIG. 8,** four stripes **801-804** are drawn on the effector-end of the tool. Four line segments **811-814** are also drawn across corresponding stripes and spaced apart by 90 degrees around the axis of the tool so that as the tool rotates about the axis, the rotation may be determined by which of the line segments are in view at the time. Similarly, in **FIG. 9,** four stripes **901-904** are also drawn on the end of the effector-end of the tool. In this case, however, four dimples **911-914** are formed are corresponding stripes and spaced apart by 90 degrees around the axis of the tool so that as the tool rotates about the axis, the rotation may be determined by which of the dimples are in view at the time.

[0171] **FIG. 10** illustrates a flow diagram of a computer model tool tracking method. An advantage of this method over other methods is that this method is generally more robust in the face of partial occlusions such as commonly encountered in surgical procedures wherein the environment is largely unstructured. For examples, in such procedures, the tools may be partially obscured by tissue, blood, or smoke; the tools may leave the field of view entirely; lighting conditions may vary greatly with time; and specular highlights may distort the normal color of the tool. The method may be used by itself for tool tracking purposes, or as one of the vision algorithms used as input to the Kalman filter described in reference to **FIG. 5.**

[0172] In **1001,** a three-dimensional computer model of the tool is generated, for example, using well known computer-aided-design tools and techniques. In **1002,** the initial position and orientation of the tool in the three-dimensional space of a surgical site is then determined by, for example, from kinematic information provided by one or more encoders coupled to a robotic mechanism used for manipulating the tool during the procedure.

[0173] In **1003,** a frame of image information is received from a camera assembly viewing the surgical site. The frame in this case represents a scene or snapshot of the surgical site area taken by the camera assembly, which may include, for examples, a single camera (monoscopic) or a pair of calibrated cameras (stereoscopic).

[0174] In **1004,** a silhouette of the computer model is determined in the two-dimensional view of the received frame after projecting the computer model onto the view. The silhouette in this case may be determined, for example, by a change in sign of the dot-product of the view-vector with the polygon face-normal of the edge's adjacent polygons. In **1005,** the basic set of silhouette edges is then processed to remove edges which are hidden in the given tool configuration using any one of conventional hidden line removal techniques.

[0175] In **1006,** edges/contours are identified or extracted from the tool image in the received frame of image information using, for example, any one of conventional edge detection techniques.

[0176] In **1007,** the silhouette of the computer model is compared against the edges of the tool image in the frame, and its position and orientation modified until a difference between the modified silhouette and the detected edges of the tool image is minimized. For example, a quality metric may be defined as the sum of absolute differences between tool edge pixels extracted from the image information and their closest silhouette edges, and the position and orientation of the silhouette moved through a number of positions and orientations in the image to find a position and orientation where the quality metric is a minimum.

[0177] In **1008,** the modified position and orientation of the computer model resulting from **1007** is then provided as a current estimate of the tool position and orientation, and the method effectively jumps back to **1003** to receive a next frame of image information when it is available and process it through **1003-1008** as described above using the modified position and orientation of the computer model in each case as its initial position of the computer model.

[0178] Although the various aspects of the present invention have been described with respect to a preferred embodiment, it will be understood that the invention is entitled to full protection within the full scope of the appended claims.

[0179] The present application also includes the following numbered clauses:

1. A tool tracking method comprising: tracking a tool by processing non-endoscopically derived tool state information and endoscopically derived tool state information generated while the tool is inserted and being manipulated through a minimally invasive incision in a body.

2. The method according to clause 1, wherein the processing of the non-endoscopically derived tool state information and endoscopically derived tool state information is performed using a Bayesian filter.

3. The method according to clause 2, wherein the endoscopically derived tool state information is provided to the Bayesian filter less frequently than the non-endoscopically derived tool state information.

4. The method according to clause 2, wherein the non-endoscopically derived tool state information is continuously provided at a sampling rate to the Bayesian filter for processing, and the endoscopically-derived tool state information

is non-continuously provided to the Bayesian filter for processing.

5. The method according to clause 2, wherein the Bayesian filter is provided with an initial estimate of the average difference between a non-endoscopically derived tool state and an endoscopically derived tool state.

6. The method according to clause 5, wherein the Bayesian filter updates the estimate of the average difference between the non-endoscopically derived tool state and the endoscopically derived tool state while estimating the state of the tool using the non-endoscopically derived tool state and the endoscopically derived tool state.

7. The method according to clause 1, wherein the non-endoscopically derived tool state information is generated from sensor data indicative of at least a position of the tool in a fixed reference frame.

8. The method according to clause 7, wherein the sensor data is indicative of an orientation of the tool in the fixed reference frame.

9. The method according to clause 7, wherein the sensor data is provided by position sensors coupled to a mechanism for manipulating the tool through the incision in the body, and the non-endoscopically derived tool state information is generated from kinematics of the mechanism using the sensor data.

10. The method according to clause 9, wherein the processing of the non-endoscopically derived tool state information and endoscopically derived tool state information is performed using a Bayesian filter.

11. The method according to clause 11, wherein quality measures are respectively derived for the non-endoscopically and the endoscopically derived tool states, and weightings of contributions of the non-endoscopically and the endoscopically derived tool states in the Bayesian filter are determined by the quality measures.

12. The method according to clause 11, wherein the quality measure for the non-endoscopically derived tool state is determined from a difference between the position in the fixed reference frame indicated by the sensor data and a position in the fixed reference frame indicated by a command signal controlling a robotic mechanism for manipulating the tool.

13. The method according to clause 11, wherein the sensor data is provided by a detector detecting a signal indicative of the position of the tool in a fixed reference frame.

14. The method according to clause 13, wherein the signal emanates from the tool.

15. The method according to clause 13, wherein the signal reflects off of the tool.

16. The method according to clause 11, wherein the sensor data is processed so as to determine whether the tool is being manipulated at the time so that processing of the endoscopically derived tool state information is only performed if the tool is being manipulated at the time.

17. The method according to clause 1, wherein the non-endoscopically derived tool state information originates from an external camera viewing an end of the tool extending out of the body.

18. The method according to clause 1, wherein the endoscopically derived tool state information is generated from image data provided by an endoscope viewing the tool while it is inserted and being manipulated through the minimally invasive incision in the body.

19. The method according to clause 18, wherein the endoscope is a stereoscopic endoscope providing left and right image planes, and generation of the endoscopically derived tool state information includes determining at least the position of the tool in a reference frame of the endoscope by processing information of the left and the right image planes.

20. The method according to clause 19, further comprising: determining the position of the endoscope in a fixed reference frame, and determining the position of the tool in the fixed reference frame from information of the position of the tool in the reference frame of the endoscope and the position of the endoscope in the fixed reference frame.

21. The method according to clause 18, wherein the generation of the endoscopically derived tool state information includes identifying an orientation dependent tool marker in the image data, and determining a endoscopically derived estimate of at least the orientation of the tool by processing information of the orientation dependent tool marker in the image data.

22. The method according to clause 18, wherein the processing of the non-endoscopically derived tool state information and the endoscopically-derived tool state information comprises: generating a computer model of the tool positioned and oriented within an image plane defined by the image data according to the non-endoscopically derived tool state information; and modifying the position and orientation of the computer model with respect to an image of the tool in the image plane until the computer model approximately overlays the image of the tool so as to generate a corrected position and orientation of the tool.

23. The method according to clause 22, wherein the modification of the position and orientation of the computer model with respect to the image of the tool until the computer model approximately overlays the image of the tool, comprises: determining the modified position and orientation of the computer model that approximately overlays the tool image by minimizing a difference between the computer model and the image of the tool.

24. The method according to clause 2, wherein the Bayesian filter is a Kalman filter.

25. The method according to clause 24, wherein the Kalman filter is an extended Kalman filters.

26. The method according to clause 2, wherein the Bayesian filter is a particle filter.

27. A tool tracking method comprising: receiving sensor information indicative of a position and orientation of a tool when the tool is inserted through an incision in a body; receiving image information for the tool; and determining the position and orientation of the tool using both the sensor and the image information.

28. The method according to clause 27, wherein the sensor information is associated with a mechanism used for robotically manipulating the tool through the incision in the body.

29. The method according to clause 28, wherein the mechanism includes a robotic arm for manipulating the tool, and a plurality of sensors generating data associated with movement of the robotic arm from which kinematics information is derived.

30. The method according to clause 27, wherein the image information is generated by at least one camera inserted in the body so as to capture images of the tool when inserted therein.

31. The method according to clause 30, wherein the at least one camera is included in one or more endoscopes.

32. The method according to clause 30, wherein the at least one camera consists of two cameras included in a stereoscopic endoscope.

33. The method according to clause 27, wherein the image information is generated by at least one camera external to the body so as to capture images of an end of the tool extending out of the body when the other end of the tool is inserted therein.

34. The method according to clause 27, wherein the determination of the tool position and orientation comprises: determining one or more kinematic estimated positions and orientations of the tool relative to a fixed reference frame from the sensor information; determining one or more image estimated positions and orientations of the tool relative to a camera reference frame from the image information; translating the one or more kinematic estimated positions and orientations of the tool from the fixed reference frame to the camera reference frame; and processing the one or more kinematic and the one or more image estimated positions and orientations to generate the tool position and orientation relative to the camera reference frame.

35. The method according to clause 34, wherein the processing of the one or more kinematic estimated positions and orientations and the one or more image estimated positions and orientations, comprises: providing the one or more kinematic and the one or more image estimated positions and orientations to a Bayesian filter.

36. The method according to clause 35, wherein the one or more image estimated positions and orientations are determined less frequently than the one or more kinematic estimated positions and orientations.

37. The method according to clause 35, wherein the Bayesian filter is provided with an initial estimate of the average difference between a sensor derived tool state and an image derived tool state.

38. The method according to clause 37, wherein the Bayesian filter updates the estimate of the average difference between the sensor derived tool state and the image derived tool state while estimating the state of the tool using the sensor derived tool state and the image derived tool state.

39. The method according to clause 36, wherein the one or more kinematic estimated positions and orientations derive from time sampled information provided by one or more sensors coupled to a mechanism for manipulating the tool through the incision in the body, and the one or more image estimated positions and orientations derive from sampled images provided by one or more cameras so as to capture images of the tool.

40. The method according to clause 39, wherein quality measures are respectively derived for the one or more kinematic estimated positions and orientations and the one or more image estimated positions and orientations, and weightings of contributions the one or more kinematic estimated positions and orientations and the one or more image estimated positions and orientations in the Bayesian filter are determined by the quality measures.

41. The method according to clause 40, wherein the quality measure for the one or more kinematic estimated positions and orientations is determined from a difference between one of the kinematic estimated positions and a position being commanded by a command signal controlling the mechanism for manipulating the tool.

42. The method according to clause 41, wherein the one or more cameras are included in an endoscope positioned within the body to capture images of an end of the tool extended into the body.

43. The method according to clause 41, wherein the one or more cameras are positioned outside of the body to capture images of an end of the tool extending out of the body.

44. The method according to clause 37, wherein the Bayesian filter is a Kalman filter.

45. The method according to clause 44, wherein the Kalman filter is an extended Kalman filter.

46. The method according to clause 37 wherein the Bayesian filter is a particle filter.

47. The method according to clause 27, wherein the determination of the tool position and orientation includes processing the image information to identify a marker on the tool, and determine an orientation of the tool using the marker.

48. The method according to clause 27, wherein the determination of the tool position and orientation includes generating a computer model of the tool using the sensor information so as to be positioned and oriented within an image plane defined in the image information, and modifying the position and orientation of the computer model with respect to an image of the tool in the image plane until the computer model substantially overlays the image.

49. The method according to clause 48, wherein the position and orientation of the computer model in the image plane is a modified position and orientation of the computer model determined for a prior in time image plane.

50. The method according to clause 48, wherein the position and orientation of the computer model in the image plane is derived from at least the kinematic information corresponding in time to that image plane.

51. The method according to clause 49, wherein the image information is generated by an endoscope inserted in the body so as to capture images of an end of the tool inserted therein.

52. The method according to clause 51, wherein the endoscope is a stereoscopic endoscope.

53. The method according to clause 49, wherein the image information is generated by at least one camera external to the body so as to capture images of an end of the tool extending out of the body when the other end of the tool

is inserted therein.

54. A minimally invasive robotic surgery system with tool tracking, comprising:

one or more non-endoscopic devices providing data from which non-endoscopically derived tool state information is generated when a tool is inserted and robotically manipulated through an incision in a body;
an endoscope capturing images from which endoscopically derived tool state information is generated for an area within the body when the tool is inserted therein; and
a processor configured to process the non-endoscopically and endoscopically derived tool state information for tracking the state of the tool.

55. The system according to clause 54, wherein the endoscope is a stereoscopic endoscope.

56. The system according to clause 54, further comprising a mechanism used for manipulating the tool through the incision in the body, wherein the one or more non-endoscopic devices include one or more sensors providing sensor data representing kinematic information according to such manipulation.

57. The system according to clause 56, wherein the sensor data includes digitized samples of an identifiable signal emanating from or reflecting off the tool so as to indicate the position of the tool.

58. The system according to clause 54, wherein the one or more non-endoscopic devices include a camera viewing an end of the tool extending out of the body.

59. The system according to clause 54, wherein the processor is further configured to track the state of the tool by: determining one or more non-endoscopic estimated tool states from the non-endoscopically derived tool state information; determining one or more endoscopic estimated tool states from the endoscopically derived tool state information; and processing the one or more non-endoscopic and one or more endoscopic estimated tool states to generate the state of the tool.

60. The system according to clause 59, wherein the processor is further configured to use a Bayesian filter for processing the one or more non-endoscopic and one or more endoscopic estimated tool states to generate the state of the tool.

61. The system according to clause 60, wherein the Bayesian filter is a Kalman filter.

62. The system according to clause 61, wherein the Kalman filer is an extended Kalman filter.

63. The system according to clause 60, wherein the Bayesian filter is a particle filter.

64. The system according to clause 54, wherein the processor is further configured to identify a marker on the tool from the endoscope captured images, and determines an orientation of the tool using the marker while tracking the state of the tool.

65. The system according to clause 54, wherein the processor is further configured to generate a computer model of the tool positioned and oriented within an image plane defined in the endoscope captured images, and modify the position and orientation of the computer model with respect to an image of the tool in the image plane until the computer model substantially overlaps the image.

66. The system according to clause 65, wherein the processor is further configured to derive the position and orientation of the computer model in the image plane from a modified position and orientation of the computer model determined by the processor for a prior in time image plane.

67. The system according to clause 65, wherein the processor is further configured to derive the position and orientation of the computer model in the image plane from at least the non-endoscopically derived tool state information corresponding in time to the image plane.

68. A minimally invasive robotic surgery system with tool tracking, comprising:

one or more sensors providing sensor data from which non-visually derived tool state information for a tool is generated when the tool is inserted and robotically manipulated through an incision in a body;
at least one camera capturing image information of the tool when the tool is inserted therein; and
a processor configured to process the non-visually derived tool state information and the image information for tracking the state of the tool.

69. The system according to clause 68, further comprising a mechanism used for manipulating the tool through the incision in the body, wherein the sensor data represents kinematic information according to such manipulation.

70. The system according to clause 68, wherein the sensor data includes digitized samples of an identifiable signal emanating from or reflecting off the tool and indicating the position of the tool.

71. The system according to clause 68, wherein the at least one camera is included in one or more endoscopes positioned so as to view an end of the tool inserted in the body.

72. The system according to clause 68, wherein the at least one camera consists of two cameras included in a stereoscopic endoscope.

73. The system according to clause 68, wherein the processor is further configured to track the state of the tool by: determining one or more non-visual estimated states of the tool from the non-visually derived tool state information; determining one or more visual estimated states of the tool from the image information; and processing the one or more non-visual and the one or more visual estimated states to generate the state of the tool.

74. The system according to clause 73, wherein the processor is further configured to: determine an error transform from a corresponding pair of the non-visual and visual estimated states of the tool; and generate a state of the tool for a subsequent time by correcting a selected one of the non-visual or visual estimated states of the tool for the subsequent time using the error transform.

75. The system according to clause 74, wherein the processor receives the image information less frequently than the non-visually derived tool state information for processing.

76. The system according to clause 75, wherein the one or more non-visual estimated states derive from time sampled information provided by the one or more sensors, and the one or more visual estimated states derive from time sampled images provided by one or more cameras.

77. The system according to clause 75, wherein the processor is further configured to use a Bayesian filter for processing the one or more non-visual and the one or more visual estimated states to generate the state of the tool.

78. The system according to clause 77, wherein the Bayesian filter is a Kalman filter.

79. The system according to clause 78, wherein the Kalman filter is an extended Kalman filter.

80. The system according to clause 77, wherein the Bayesian filter is a particle filter.

81. The system according to clause 77, wherein the processor receives the image information less frequently than the non-visually derived tool state information for processing.

82. The system according to clause 81, wherein the one or more non-visual estimated states derive from time sampled information provided by the one or more sensors, and the one or more visual estimated states derive from time sampled images provided by one or more cameras.

83. The system according to clause 68, wherein the processor is further configured to identify a marker on the tool from the image information, and determine an orientation of the tool using the marker while tracking the state of the tool.

84. The system according to clause 68, wherein the processor is further configured to generate a computer model of the tool positioned and oriented within an image plane defined in the image information, and modify the position and orientation of the computer model with respect to an image of the tool in the image plane until the computer

model substantially overlays the image.

85. The system according to clause 80, wherein the processor is configured to derive the position and orientation of the computer model in the image plane from a modified position and orientation of the computer model determined by the processor for a prior in time image plane.

86. The system according to clause 84, wherein the processor is configured to derive the position and orientation of the computer model in the image plane from at least the non-visually derived tool state information corresponding in time to the image plane.

87. A tool tracking method comprising:

determining a computer model of a tool;
receiving a captured image including a view of the tool;
determining an estimated position and orientation of the tool from the captured image, and positioning and orienting the computer model at that estimated position and orientation in reference to the captured image; and modifying the estimated position and orientation of the computer model with respect to an image of the tool in the captured image until the computer model approximately overlays the image so as to correct the estimated position and orientation of the tool for the captured image.

88. The method according to clause 87, wherein the modification of the estimated position and orientation of the computer model with respect to the image of the tool in the captured image, comprises: determining the modified position and orientation of the computer model that approximately overlays the tool image by minimizing a difference between the computer model and the image of the tool.

89. The method according to clause 87, wherein the captured image is captured by a stereo endoscope inserted into an area of the body.

90. The method according to clause 87, further comprising: extracting a silhouette of the computer model of the tool; extracting edges of the image of the tool in the captured image; and positioning and rotating the silhouette of the computer model until the silhouette of the computer model approximately overlays the edges of the image of the tool within the captured image.

91. The method according to clause 90, further comprising: removing lines in the silhouette of the computer model of the tool corresponding to hidden lines in the image of the tool in the captured image prior to positioning and rotating the silhouette of the computer model until the silhouette of the computer model approximately overlays, the edges of the image of the tool within the captured image.

92. The method according to clause 90, wherein the positioning and rotating of the silhouette of the computer model until the silhouette approximately overlays the edges of the image of the tool within the captured image, comprises:

determining a difference between the silhouette of the computer model and the image of the tool in the captured image; and positioning and rotating the silhouette of the computer model until the difference is minimized.

93. The method according to clause 92, wherein the difference to be minimized is a sum of absolute differences between edge pixels extracted from the tool image and their closest silhouette edges.

94. The method according to clause 92, wherein the view of the area where the tool is inserted is captured as a grid of pixels by at least one camera inserted in the area.

95. A tool tracking method comprising:

determining whether sensor data indicative of a tool state is available for a point in time;
determining whether image data indicative of the tool state is available for the point in time; and
determining the tool state using both the sensor data and the image data if both are available for the point in time, or using only the sensor data if only the sensor data is available, or using only the image data if only the image data is available.

96. The method according to clause 95, further comprising: determining whether a user is manipulating the tool at the point in time; and performing the sensor data and image data availability determinations only if the user is determined to be manipulating the tool at the point in time.

97. The method according to clause 95, further comprising: determining whether the state of the tool has changed relative to a prior point in time using the sensor data; and performing the image data availability and tool state determinations only if the state of the tool is determined to have changed.

98. The method according to clause 95, wherein the tool state indicated by the sensor data includes a position and orientation of the tool in a fixed reference frame.

99. The method according to clause 98, wherein the image data is generated by a camera unit, the sensor data further indicates a position and orientation of the camera unit in the fixed reference frame, and the tool state indicated by the image data includes a position and orientation of the tool in a reference frame related to the camera unit.

100. The method according to clause 99, wherein the determination of the tool state using the sensor data and the image data, comprises: translating the position and orientation of the tool indicated by the sensor data from the fixed reference frame to the reference frame related to the camera unit before determining the tool state when using both sensor data and the image data.

101. A tool tracking method comprising:

determining a first estimated tool state relative to a landmark for a point in time using first sensor data indicative of the tool state at the point in time;
determining an estimated camera state relative to the landmark for the point in time using second sensor data indicative of the camera state at the point in time;
determining a second estimated tool state relative to the camera for the point in time using image data generated by the camera and indicative of the tool state at the point in time;
translating the first estimated tool state so as to be relative to the camera instead of the landmark; and
computing an error transform between the translated first and the second estimated tool states that at a subsequent point in time if image data indicative of the tool state at the subsequent point in time is not available, then the tool state is determined by applying the error transform to a third estimated tool state determined using sensor data indicative of the tool state at the subsequent point in time translated so as to be relative to the camera instead of the landmark.

102. The method according to clause 101, further comprising if the image data indicative of the tool state at the subsequent point in time is available: determining a fourth estimated tool state relative to the camera for the subsequent point in time using image data generated by the camera and indicative of the tool state at the subsequent point in time; and determining the tool state at the subsequent point in time as being the fourth estimated tool state.

103. The method according to clause 102, further comprising if the image data indicative of the tool state at the subsequent point in time is available: computing a second error transform between the translated third and the fourth estimated tool states so that at a subsequent, subsequent point in time if image data indicative of the tool state at the subsequent, subsequent point in time is not available, then the tool state is determined by applying the second error transform to a fifth estimated tool state determined using sensor data indicative of the tool state at the subsequent, subsequent point in time translated so as to be relative to the camera instead of the landmark.

104. A tool tracking method comprising:

determining non-endoscopically derived estimated state information for a tool at a given time;
determining endoscopically estimated state information for the tool at the given time; and
providing the non-endoscopically and endoscopically derived estimated states for the tool to a Bayesian filter configured so as to generate an optimal estimate of the state of the tool.

105. The method according to clause 104, wherein the Bayesian filter is provided with an initial estimate of an average difference between a non-endoscopically derived tool state and an endoscopically derived tool state.

106. The method according to clause 105, wherein the Bayesian filter updates the estimate of the average difference

between the non-endoscopically derived tool state and the endoscopically derived tool state while estimating the state of the tool using the non-endoscopically derived tool state and the endoscopically derived tool state.

107. The method according to clause 104, wherein the non-endoscopically derived estimated state information is generated from sensor data indicative of at least a position of the tool in a fixed reference frame at the given time.

108. The method according to clause 107, wherein the endoscopically derived estimated state information is generated from image data received from at least one camera capturing image information for an area of the body when the tool is inserted therein.

109. The method according to clause 104, wherein the Bayesian filter is a Kalman filter.

110. The method according to clause 109, wherein the Kalman filter is an extended Kalman filter.

111. The method according to clause 104, wherein the Bayesian filter is a particle filter.

112. A tool tracking and calibration method comprising:

generating visually derived state information from image data received from a camera viewing a tool;
generating state vector information by combining initial values for a set of camera parameters with the visually derived state information; and
providing the state vector information to a Bayesian filter for processing so as to generate an optimal estimate of a state of the tool and corrected values for the set of camera parameters.

113. The method according to clause 112, further comprising:

generating non-visually derived state information from sensor data received from one or more sensors and including information indicative of a position of the tool; and
generating the state vector information by combining the non-visually derived state information with the initial values for the set of camera parameters and the visually derived state information.

114. The method according to clause 113, wherein the camera is an endoscope.

115. The method according to clause 113, wherein the one or more sensors are associated with a robotic mechanism used for manipulating the tool through a minimally invasive incision in a body.

116. A camera tracking method comprising:

determining a position of a tool in a fixed reference frame from non-visually derived tool state information generated from sensor data indicative of the position of the tool;
determining a position of the tool in a camera frame moveable with a camera using visually derived tool state information generated from image data provided by the camera while viewing the tool; and
determining a position of the camera in the fixed reference frame using the position of the tool in the fixed reference frame and the position of the tool in the moveable camera frame.

117. A tool tracking method comprising:

determining a position of a camera in a fixed reference frame from non-visually derived camera state information generated from sensor data indicative of the position of the camera;
determining a position of a tool in a camera frame moveable with the camera using visually derived tool state information generated from image data provided by the camera while viewing the tool; and
determining a position of the tool in the fixed reference frame using the position of the camera in the fixed reference frame and the position of the tool in the moveable camera frame.

118. A tool tracking method comprising:

generating a plurality of estimated tool states for each point in a plurality of points in time, while the tool is inserted and being manipulated through an incision in a body; and

determining an optimal estimated tool state for each point in the plurality of points in time by processing the plurality of estimated tool states using Bayesian techniques.

119. The method according to clause 118, wherein the plurality of estimated tool states include an estimated tool state determined using only sensor data associated with a robotic mechanism for manipulating the tool, so as to be indicative of movement of the robotic mechanism.

120. The method according to clause 119, wherein the robotic mechanism includes joints and linkages, and the sensor data is indicative of movement of the joints and linkages.

121. The method according to clause 118, wherein the plurality of estimated tool states include an estimated tool state determined using only sensor data associated with the tool, so as to be indicative of a position of the tool.

122. The method according to clause 121, wherein the sensor data is further indicative of an orientation of the tool.

123. The method according to clause 118, wherein the plurality of estimated tool states include an estimated tool state determined using only image data generated by an endoscope positioned so as to view an effector end of the tool as the tool is inserted and being manipulated through the incision in the body.

124. The method according to clause 118, wherein the plurality of estimated tool states include an estimated tool state determined using only image data generated by an external camera positioned so as to view an exposed end of the tool extending out of the body as the tool is inserted and being manipulated through the incision in the body.

125. The method according to clause 118, wherein the Bayesian technique is a Kalman filtering technique.

126. The method according to clause 118, wherein the Bayesian technique is a particle filtering technique.

Claims

1. A tool tracking and calibration system comprising:

means for generating visually derived state information of a tool from image data received from a camera viewing the tool;
means for generating state vector information by combining initial values for a set of camera parameters with the visually derived state information of the tool; and
means for providing the state vector information to a Bayesian filter for processing so as to generate an optimal estimate of a state of the tool and corrected values for the set of camera parameters.

2. The system according to claim 1, further comprising:

means for generating non-visually derived state information from sensor data received from one or more sensors and including information indicative of a position of the tool; and
means for generating the state vector information by combining the non-visually derived state information with the initial values for the set of camera parameters and the visually derived state information.

3. The system according to claim 2, wherein the camera is an endoscope.

4. The system according to claim 2, wherein the one or more sensors are associated with a robotic mechanism used for manipulating the tool through a minimally invasive incision in a body.

5. A computer implemented tool tracking and calibration method comprising:

generating visually derived state information from image data received from a camera viewing a tool;
generating state vector information by combining initial values for a set of camera parameters with the visually derived state information; and
providing the state vector information to a Bayesian filter for processing so as to generate an optimal estimate of a state of the tool and corrected values for the set of camera parameters.

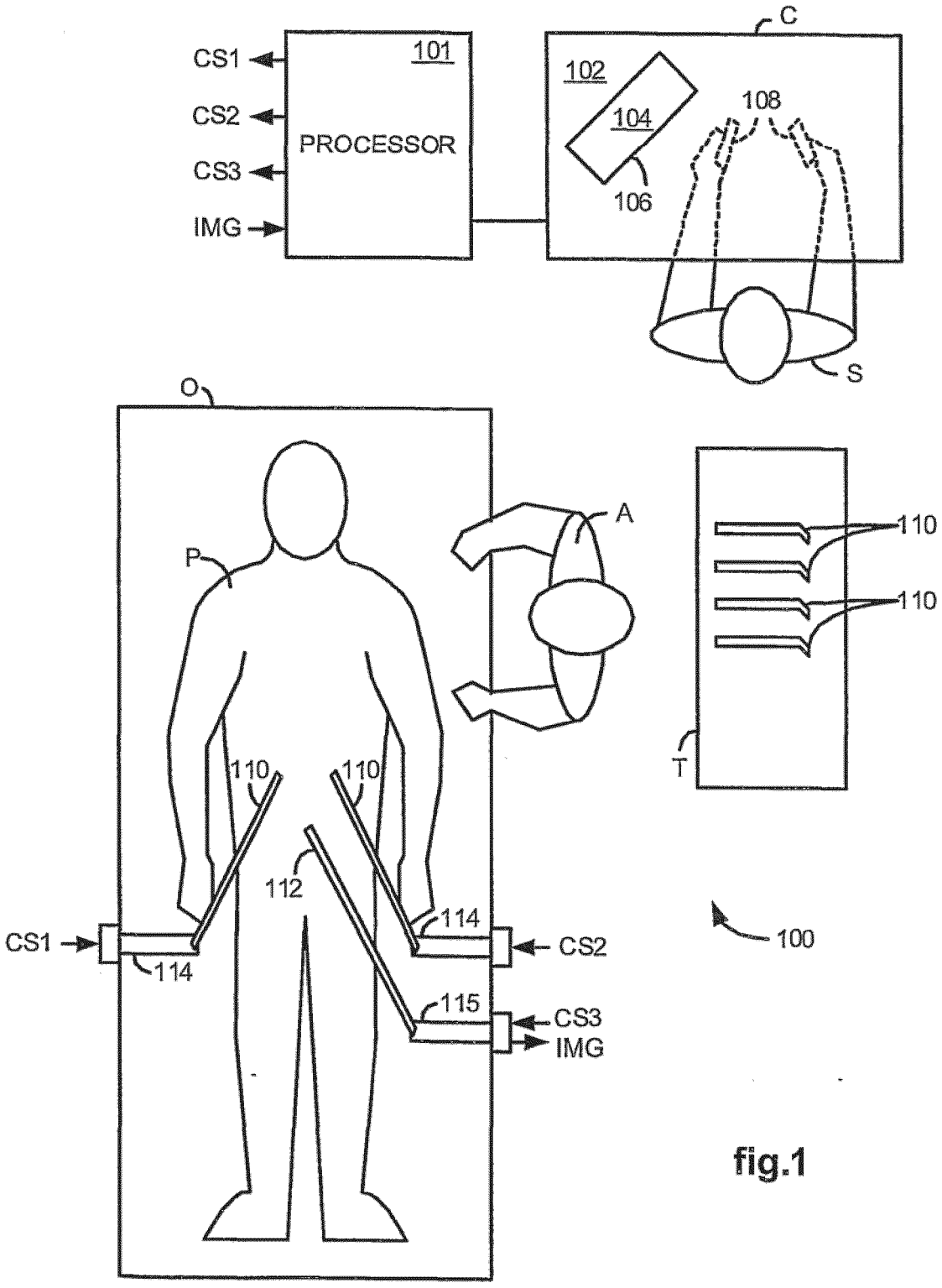
C
CS1
CS2
CS3
IMG
101
PROCESSOR
102
104
106
108
S
O
P
A
110
110
110
110
112
114
114
115
CS1
CS2
CS3
IMG
T
100

fig.1

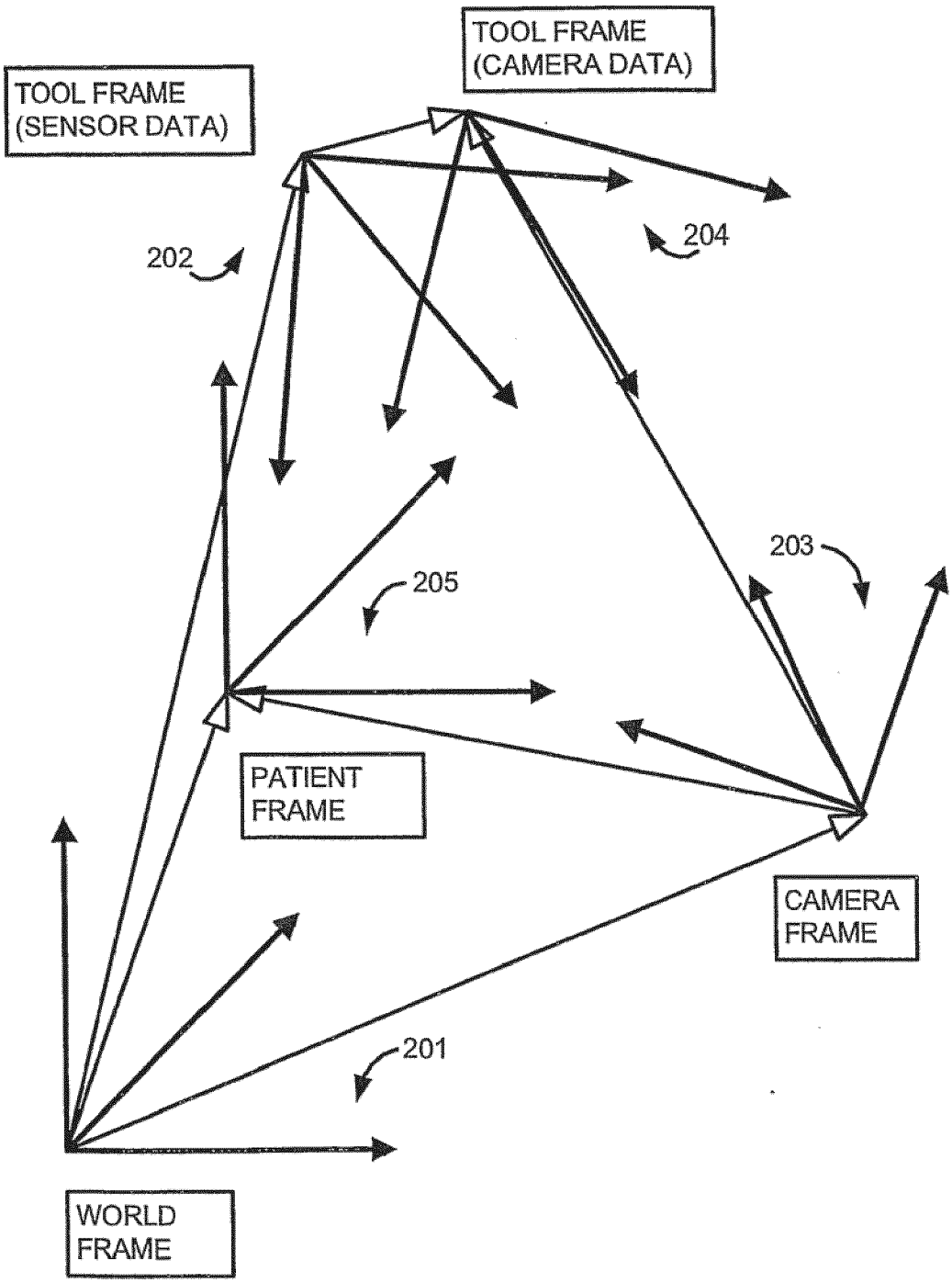
TOOL FRAME
(CAMERA DATA)
TOOL FRAME
(SENSOR DATA)
202
204
203
205
PATIENT
FRAME
CAMERA
FRAME
201
WORLD
FRAME

fig.2

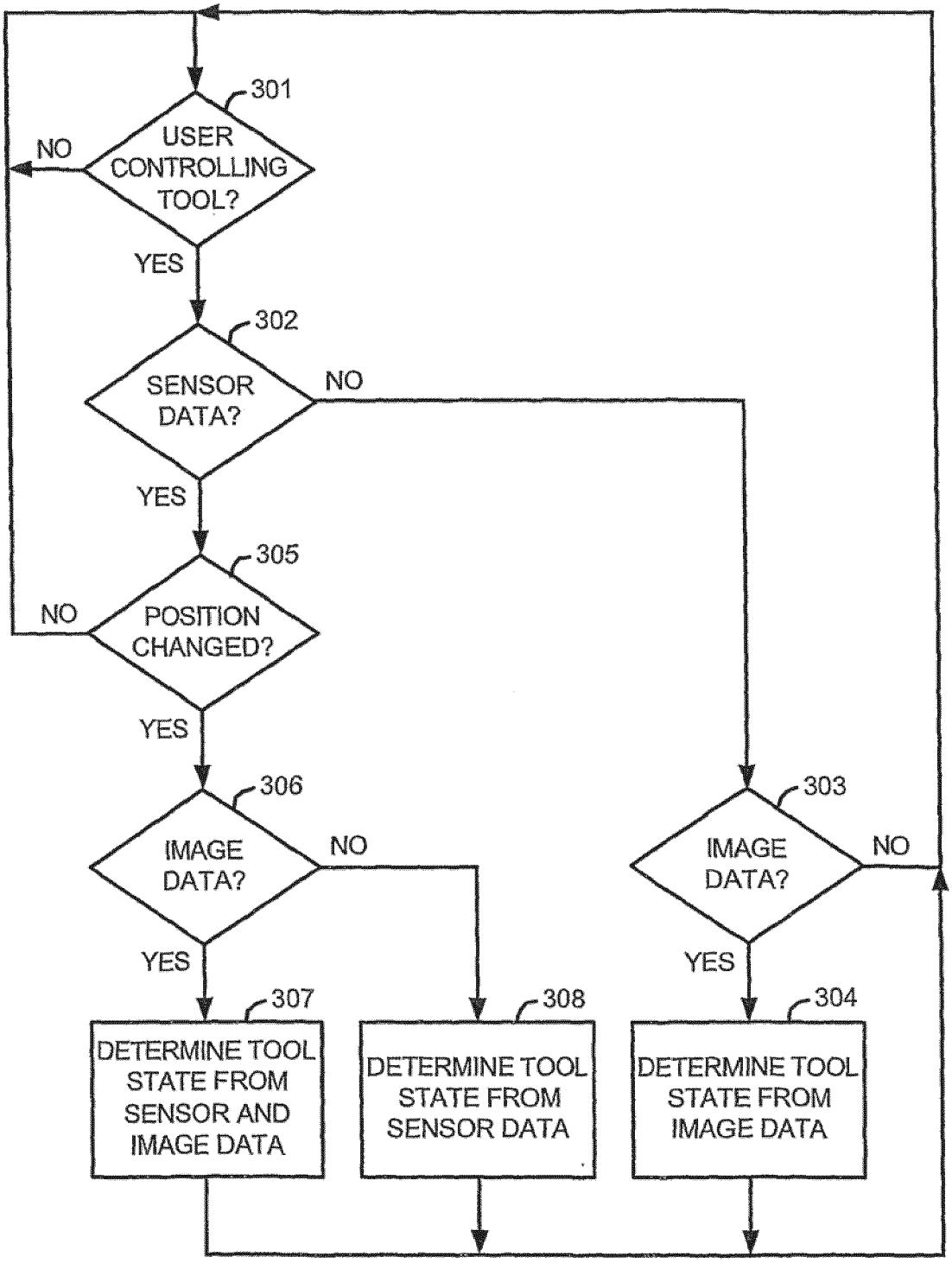
301
USER CONTROLLING TOOL?
NO
YES
302
SENSOR DATA?
NO
YES
305
POSITION CHANGED?
NO
YES
306
IMAGE DATA?
NO
YES
303
IMAGE DATA?
NO
YES
307
DETERMINE TOOL STATE FROM SENSOR AND IMAGE DATA
308
DETERMINE TOOL STATE FROM SENSOR DATA
304
DETERMINE TOOL STATE FROM IMAGE DATA

fig.3

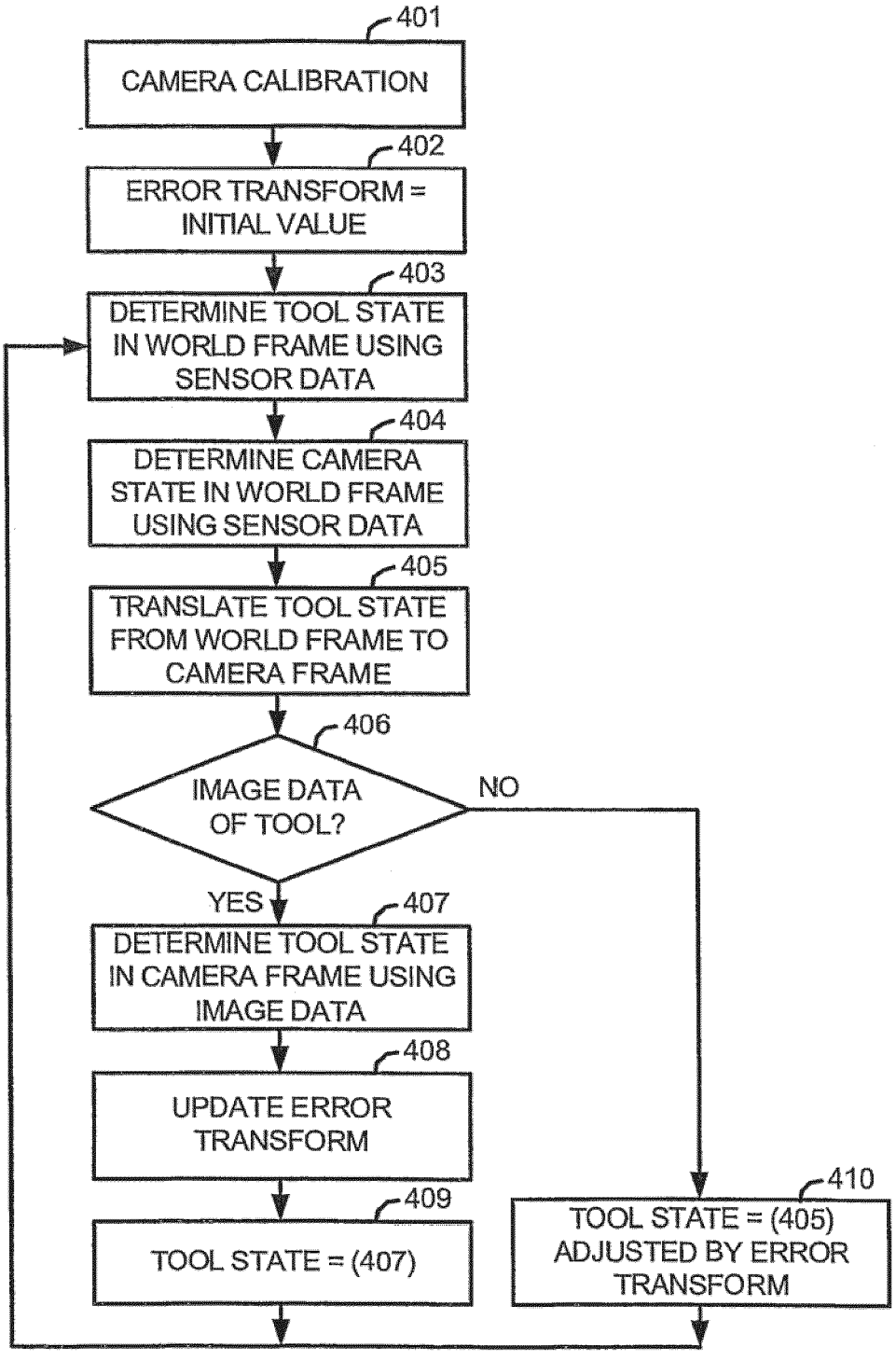
401
CAMERA CALIBRATION
402
ERROR TRANSFORM = INITIAL VALUE
403
DETERMINE TOOL STATE IN WORLD FRAME USING SENSOR DATA
404
DETERMINE CAMERA STATE IN WORLD FRAME USING SENSOR DATA
405
TRANSLATE TOOL STATE FROM WORLD FRAME TO CAMERA FRAME
406
IMAGE DATA OF TOOL?
NO
YES
407
DETERMINE TOOL STATE IN CAMERA FRAME USING IMAGE DATA
408
UPDATE ERROR TRANSFORM
409
TOOL STATE = (407)
410
TOOL STATE = (405) ADJUSTED BY ERROR TRANSFORM

fig.4

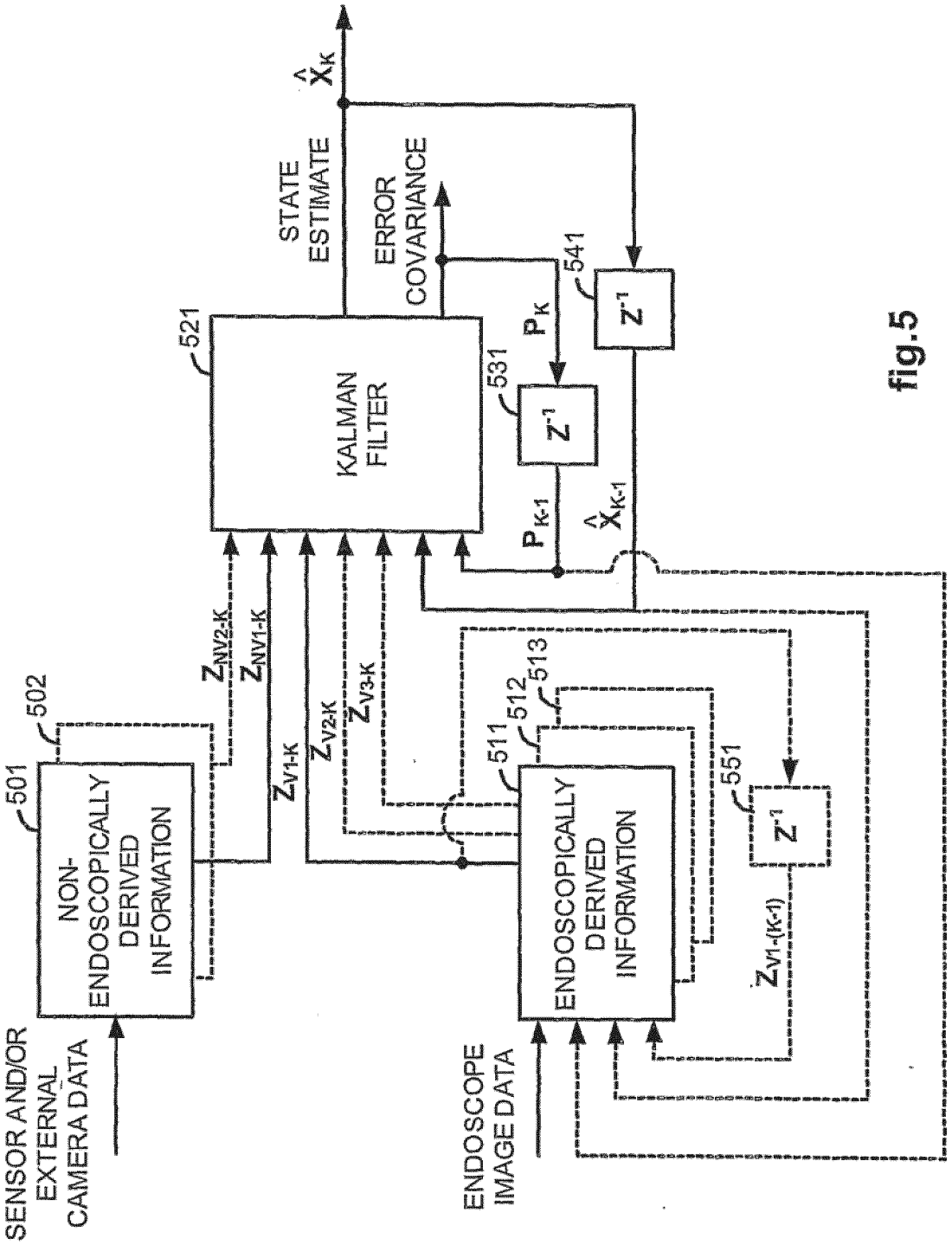
SENSOR AND/OR EXTERNAL CAMERA DATA
501
NON-ENDOSCOPICALLY DERIVED INFORMATION
502
ENDOSCOPE IMAGE DATA
ENDOSCOPICALLY DERIVED INFORMATION
511
512
513
551
Z^{-1}
$Z_{V1-(K-1)}$
Z_{V1-K}
Z_{V2-K}
Z_{V3-K}
Z_{NV1-K}
Z_{NV2-K}
521
KALMAN FILTER
531
Z^{-1}
P_{K-1}
$\hat{X}_{K-1}$
P_K
541
Z^{-1}
ERROR COVARIANCE
STATE ESTIMATE
$\hat{X}_K$

fig.5

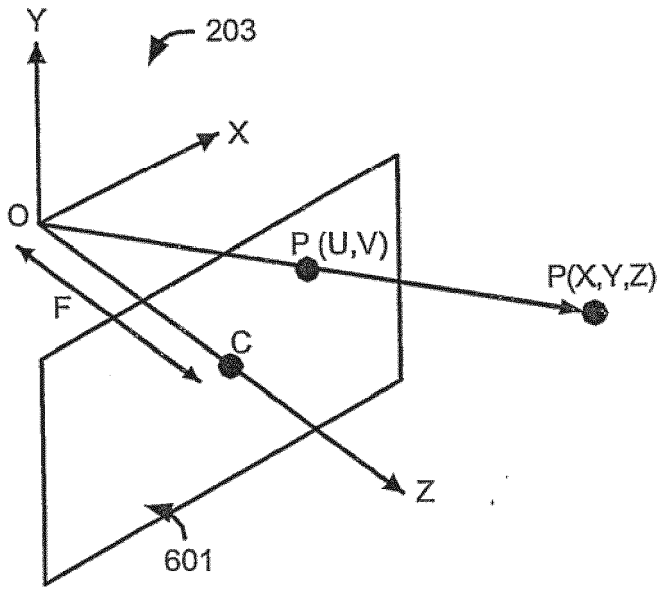

fig.6

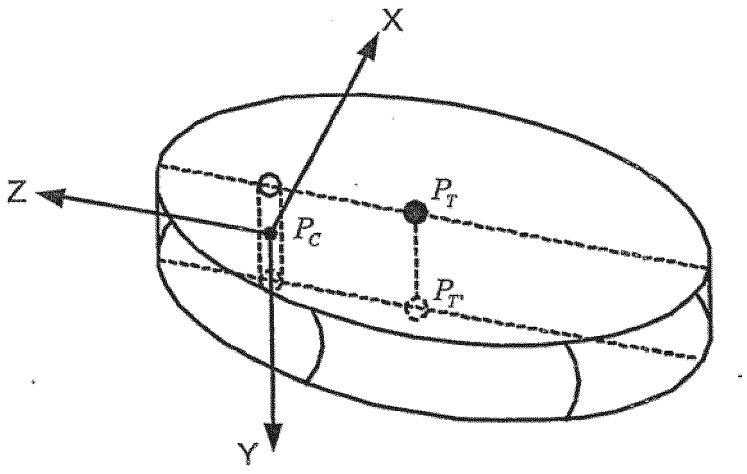

fig.7

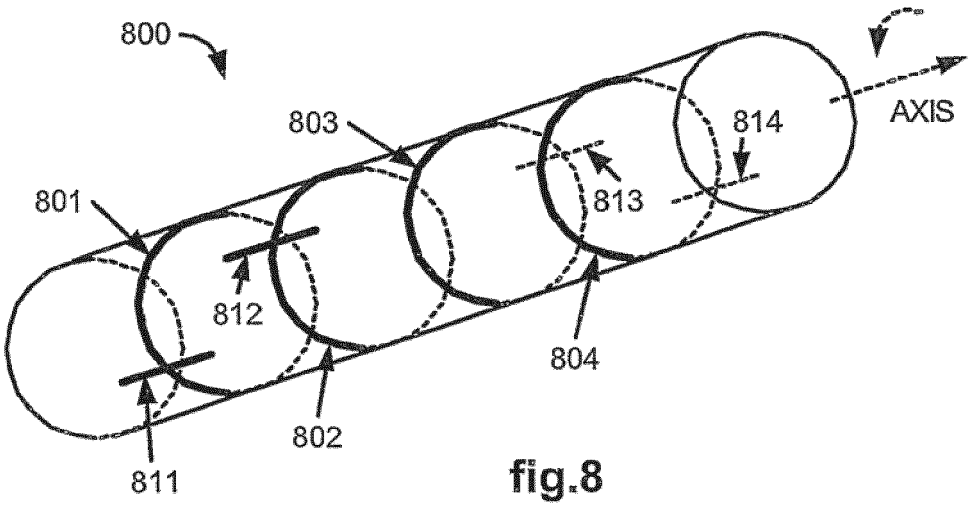
800
803
814
AXIS
801
813
812
804
802
811

fig.8

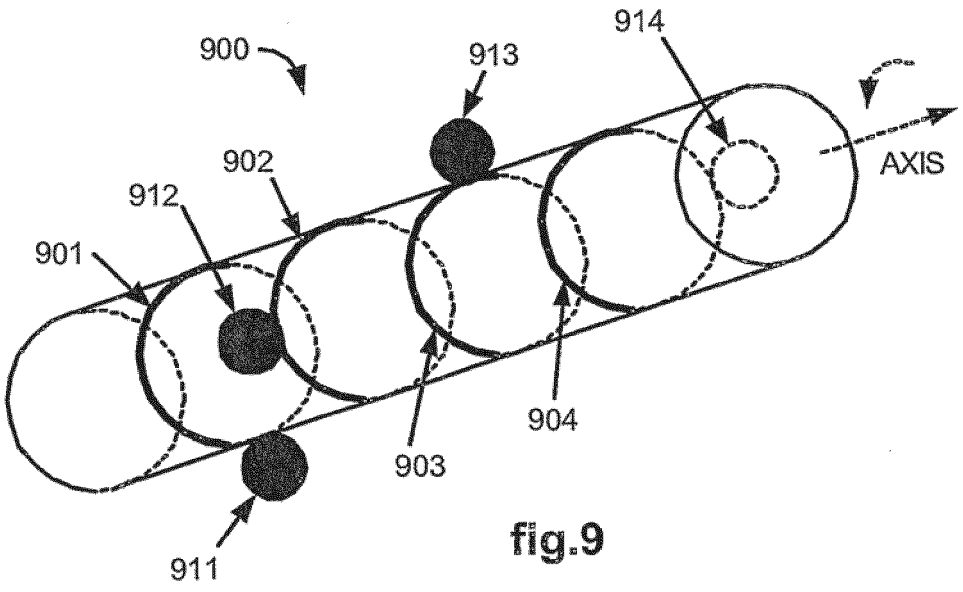
914
900
913
902
AXIS
912
901
904
903
911

fig.9

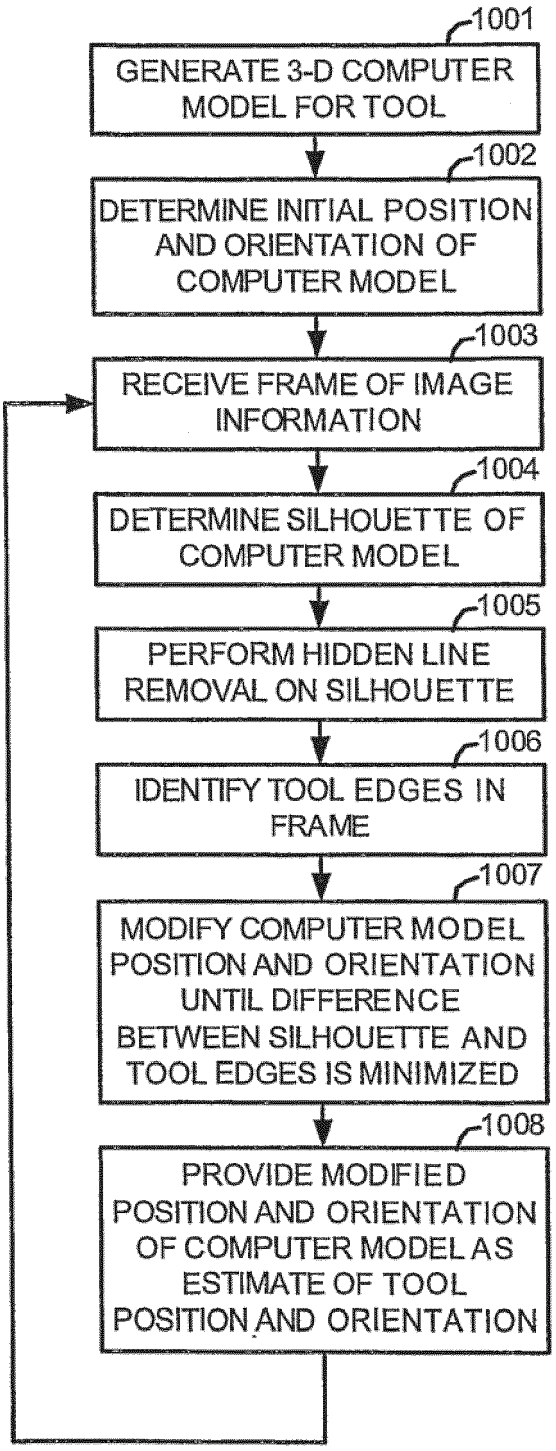
1001
GENERATE 3-D COMPUTER MODEL FOR TOOL
1002
DETERMINE INITIAL POSITION AND ORIENTATION OF COMPUTER MODEL
1003
RECEIVE FRAME OF IMAGE INFORMATION
1004
DETERMINE SILHOUETTE OF COMPUTER MODEL
1005
PERFORM HIDDEN LINE REMOVAL ON SILHOUETTE
1006
IDENTIFY TOOL EDGES IN FRAME
1007
MODIFY COMPUTER MODEL POSITION AND ORIENTATION UNTIL DIFFERENCE BETWEEN SILHOUETTE AND TOOL EDGES IS MINIMIZED
1008
PROVIDE MODIFIED POSITION AND ORIENTATION OF COMPUTER MODEL AS ESTIMATE OF TOOL POSITION AND ORIENTATION

fig.10

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 13 18 7618

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	US 6 468 265 B1 (EVANS PHILIP C [US] ET AL) 22 October 2002 (2002-10-22) * column 64, line 26 - line 37 * -----	1,5	INV. A61B19/00 A61B1/00 ADD. A61B1/04
			TECHNICAL FIELDS SEARCHED (IPC) A61B

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	11 December 2013	Angeli, Markus

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
.....
& : member of the same patent family, corresponding document

1

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 13 18 7618

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

11-12-2013

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6468265 B1	22-10-2002	US 6468265 B1	22-10-2002
		US 2003055410 A1	20-03-2003
		US 2005107808 A1	19-05-2005

EPO FORM P0459

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 6346072 B **[0058]**

Non-patent literature cited in the description

- **GUO-QING WEI ; KLAUS ARBTER ; GERD HIRZINGER.** Real-Time Visual Servoing for Laparoscopic Surgery. *IEEE Engineering in Medicine and Biology,* 1997, vol. 16 (1), 40-45 **[0009]**
- **XIAOLI ZHANG ; SHAHRAM PAYANDEH.** Application of Visual Tracking for Robot-Assisted Laparoscopic Surgery. *Journal of Robotics Systems,* 2002, vol. 19 (7), 315-328 **[0009]**
- **GREG ; WELCH ; GARY BISHOP.** An Introduction to the Kalman Filter. *TR 95-041,* 05 April 2004 **[0097]**
- **BIJOY K. GHOSH ; NING XI ; T.J. TARN.** Control in Robotics and Automation: Sensor Based Integration. Academic Press, 1999 **[0139]**