# Video see-through in the clinical practice

**Vincenzo Ferrari**
Centro EndoCAS
Università di Pisa
+39 (0) 50 995689
vincenzo.ferrari@endocas.org

**Mauro Ferrari, Franco Mosca**
Centro EndoCAS
Università di Pisa
+39 (0) 50 995689
*name.surname*@med.unipi.it

**ABSTRACT**

In this paper, we discuss potentialities and technological limits to overcome for the introduction in the clinical practice of useful functionalities, using video see-through visualizations, created mixing virtual preoperative information, obtained by means of radiological images, with real patient live images, for procedures where the physician have to interact with the patient (palpation, percutaneous biopsy, catheterism, intervention, etc…).

**Keywords**

Mixed reality, surgical navigation, general surgery.

## INTRODUCTION

Modern CT and MRI scanners coupled with new contrast mediums allow the acquisition of volumetric datasets describing human anatomy, functionality and pathology, with high level of detail.

The detailed information contained in a volumetric dataset are fully used during the diagnostic phase, but are partially lost passing from the radiological department to the surgical department.

In fact, generally, surgeons plan interventions just using limited information provided by the radiologist and consisting in the textual diagnosis coupled with few 2D significant images selected from the volumetric dataset.

The application of the "computer assisted" model to the patient workflow, consisting of computer aided diagnosis (CAD) and computer aided surgery (CAS) technologies, allows the optimal use of medical dataset and to overcome the above cited limitations of the current clinical practice. The 3D visualization of patient specific virtual models of anatomies [23; 24], extracted from medical dataset, drastically simplifies the interpretation process of exams and provides benefits both in diagnosing and in surgical planning phases. Computer assisted technologies allow to augment real views of the patient, grabbed by means of cameras, with virtual information[26]. This augmented-reality, or in general mixed-reality techniques [20], introduces many advantages for each task where the physician have to interact with the patient (palpation, introduction of biopsy needle, catheterization, intervention, etc.) [9; 10; 25]



The next figure shows a binocular see-through mixed reality system at work implemented using a HMD (Head Mounted Display) and external cameras [8].

**Figure 1: Stereoscopic video see-through in the operative room**

To implement this kind of systems is generally required to localize the anatomy in respect to the real video source and to determine its projection model in order to coherently mix virtual and real scenarios. Localization can be done using commercial tracking systems, introducing additional costs and logistic troubles in the traditional clinical scenario, with large errors on soft tissues, while the projection model of the video source can be calculated using theoretical algorithms that impose some constrains for the real camera.

In the following is described in the detail the problem and possible solutions to avoid the need of the tracker or to improve the localization quality on soft tissues taking into account the limits of the current images source used in surgery.

## HOW TO OBTAIN A MIXED REALITY VIEW

The following picture essentially describes the video see-through concept.

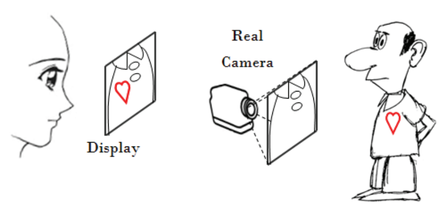


**Figure 2: Video see-through concept**

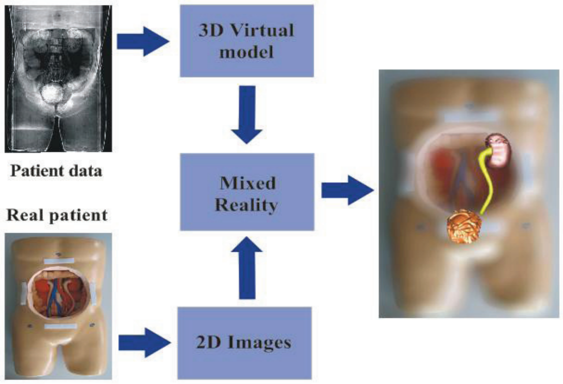


**Figure 3: Functional scheme of a surgical see-through system**

Real video frames, grabbed by of real camera/s, are mixed with virtual objects not visible in the real scene and shown on a display/s. This virtual information can be obtained using radiological images as depicted in the next figure.

The using of volumetric scanners, like CT (Computed Tomography) or MRI (Magnetic Resonance Imaging), allows to obtain a 3D virtual model of the anatomy [4; 6], which can be loaded in a virtual scene, running on a computer, rendered from a point of view coherent with the real point of view.

The mixing of the real (2D) images with the virtual (2D) rendered images can be done using a hardware video mixer or using the real images in the scene graph as foreground or background [19]. The concept and the work to do are similar: in the first case the mixing is done by external hardware after the rendering of the virtual scene, while in the second one by the GPU during the rendering. Figure 4 shows this concept. The real camera acquires video frames from the real environment (a spleen in this case). Video frames are shown as background of the virtual scene. Virtual objects are positioned in the scene (green flashes in this case) and rendered from a virtual camera.

In order obtain a coherent fusion we have to obtain a virtual scene where:

virtual camera projection model ≈ to the real one
virtual camera position ≈ to the real one
virtual objects positions ≈ to the real ones

The following paragraphs describes how to obtain the previous three conditions.

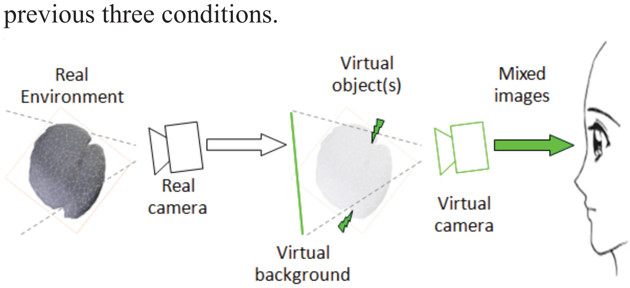


**Figure 4: Implementation of mixed-reality in a virtual scene**

**How to determine camera projection model**

Line scan and telecentric cameras are used for particular industrial applications, while for all visualization purposes, including laparoscopy, the perspective projective camera is the only used, because it offers the most similar images in respect to human vision.

Regarding the sensor, two technologies are predominantly applied: CCD (Charge Coupled Device) and CMOS (Complementary Metal Oxide Semiconductor). In each case unitary elements (pixels) are disposed on a regular grid (with fixed resolution).

Each camera, composed of a projective optics and a grid sensor, can be represented by the following model:

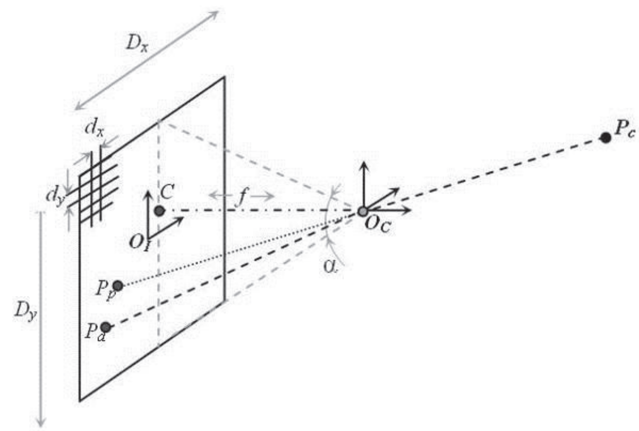


**Figure 5: Schematic representation of the pinhole camera model: the generic point Pc is ideally projected on the image sensor of the camera (the plane with origin $O_I$) through the projection center $O_C$ (where the origin of the camera reference frame is fixed)**

The perspective projection matrix Mp, mapping a generic 3D point Pc = $[x, y, z, 1]^T$, in the camera reference system, to the corresponding 2D point Pp = $[u, v, 1]^T$ in the image reference system (fixed on the center of the sensor), i.e.:

$$P_p = M_p P_c \qquad (1)$$

is defined starting from the internal camera parameters ($f$, $C_x$, $C_y$) as follows:

$$M_p = \begin{bmatrix} -f & 0 & Cx & 0 \\ 0 & -f & Cy & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad (2)$$

where $f$ is the focal distance and ($C_x$, $C_y$) are the coordinates of the projection of the $O_c$ on the image reference frame (with origin in $O_I$).

Other internal camera parameters parameterize the model of the radial distortion, introduced by common lens, by means of which the projected point $P_p$ is deviated on $P_d$.

The pixelization process is defined by the pixel dimensions $d_x$ and $d_y$ and the image sensor dimensions $D_x$ and $D_y$. These internal parameters of the camera allow to convert measurements done on the image (in pixels) in real measurements (in millimeters) and vice-versa.

All internal camera parameters can be determined in a calibration phase acquiring some images of a knowing object in different positions with fixed camera configuration (in terms of diaphragm and camera focus) and using calibration routines like described in [30].

These parameters have to be used to adjust the virtual camera to the real one.

Using traditional surgical endoscopes a new camera calibration and virtual camera adjustment is required whenever either the optic zoom or the diaphragm opening are changed. Another important source of error can be determined by the mechanical joint between the optic and the camera body. Their relative movements can determines a shift of the center of projection *C* up to tens of pixels.

**How to localize the camera**

Camera position and orientation can be obtained using a tracker able to track a sensor mounted on the camera body as shown in the following figure.

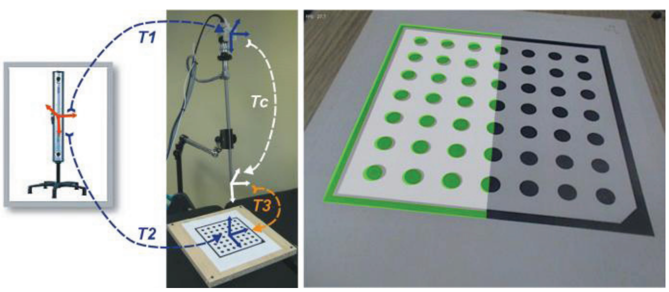


**Figure 6: Camera localization and calibration process using an optical localizer and a sensor mounted on the camera body**

The tracker offers in real time the transformation matrix $T_1$ relative to the sensor. The calibration matrix $Tc$, representing the relative transformation of the camera viewpoint with respect to the sensorized frame, necessary to determine position and orientation of the camera projection center $O_C$, can been computed using a sensorized calibration grid. During the calibration $T_1$ and $T_2$ are given by the localization system, while the transformation $T_3$ is determined using computer vision methods that allow to localize, in the camera reference frame, objects with known geometry (the sensorized calibration grid).

Another approach could be the localization using directly video frames acquired by the cameras as done in some applications. Several computer vision libraries (OpenCV or Halcon by MVTec) offers many tools for this purpose.

Using a single camera, we could localize objects with known geometry or texturing [11] as in the case of EasyOn by Seac02 *(www.seac02.it)*. The localization accuracy is enough for many applications, but requires knowing in advance the dimensions and the texture of a rigid object in the scene (or different objects rigidly linked together). Interesting monoscopic solutions have been applied using laparoscopic images: see-through systems applying on organs artificial markers [SOFT TISSUE], recovering the position of needle [29] and the pose of surgical instruments [5].

**How to register the patient**

In surgical applications, virtual objects, representing patient anatomies, are acquired in the reference frame of the radiological instrumentation just before or days before the surgical procedure, whereas the intra-operative information is related to the reference frame of the surgical room (generally defined by means of a tracking system) during the intervention.

In case of rigid objects like bones, a changing of reference frame, performed aligning fiducial points or fiducial surfaces, acquired on the radiology department and in the surgical room, can be enough [1; 3]. Deformations of the fiducial structure composed by elements, such as points of a cloud or points characterizing a surface, introduce systematic errors in the registration. In order to minimize the registration error, at least on fiducials elements, each fiducial point (or fiducial surface) in the proximity of steady element on the patient has to be chosen, and its configuration has to be as replicable as possible [19].

In case of soft tissue, further than the changing of reference frame, there are many deformation effects to avoid or to compensate, due to: changing of patient decubitus, changing in bed configuration, physiological movements (breathing, heart beating, gastrointestinal movements, etc…), constraints due to the radiological scanners (breath hold, arts positions, etc…).

To reduce these movement effects we can employ practical and useful artifices, used routinely by radiotherapists reproducing meticulously the patient settings during the treatment as in the planning room. By following their work, bed positioning and its shape, during the acquisition of medical datasets, can be chosen accordingly to the bed configuration used inside the surgical room for the specific intervention (considering the requirements of the used radiological device and the type of intervention to be performed). Furthermore during the intervention, the exact decubitus of the patient during radiological scanning requires to obtain the same relative position of the basin and the thoracic cage. A realignment of these structures needs immobilization devices and/or additional iterative work in the surgical room in order to find a perfect correspondence between pre-operative and intra-operative patient positioning [15].

The using of intra-operative imaging devices like 3D RA (Rotational Angiograph), which could be diffused in the early future, thanks to the decreasing of their price and the possibility to be portable (Ziehm Vision FD Vario 3D or Siemens ARCADIS Orbic 3D), allows to avoid the change of reference frame for each patient. These scanners, positioned in the operating room, can be easily and precisely calibrated with the localizer by means of sensors. Furthermore the acquisition of the anatomy directly on the surgical bed allows to dramatically simplify the problem, by removing error due to the change of bed and patient decubitus. This simplification will allow to obtain high precision also on soft tissues. As proven by experimental results, the application of predictive models of organs motion due to breathing, driven by simple intra-operative parameters like the trajectory of a point on the patient skin or the time over the breathing cycle, can be applied in the real surgical scenario [14; 22].

## ALTERNATIVE SOLUTIONS

### Head mounted tracker-free stereoscopic video see-through

Depth perception can be drastically increased using head mounted stereoscopic devices [17], that allow to evaluate object depth dislocation, like in the natural binocular view. The use of localized head mounted displays (HMD), like the one shown in figure 1, allows to see a synthetic scene from a point of view aligned with the real user's point of view.

For the implementation of head mounted mixed reality systems, the video see-through approach, based on the acquisition of real images by means of external cameras, is preferable to the optic see-through approach that projects virtual information on semi transparent glasses. This is due to the fact that tracking of eye movements, strictly required for optical see-through approach, is very difficult to be performed with sufficient precision [16; 18]. On the contrary, head tracking, required for video see-through approach, can be performed with high precision using external localizer based on different technologies [2; 12], like described before.

We implemented a head mounted stereoscopic video see-through system, that does not require the use of an external localizer to track head movements [8]. Our system implements mixed-reality aligning in real-time virtual and real scene just using geometric information extracted by segmenting coloured markers, attached on the patient's skin, directly from couples of camera images.

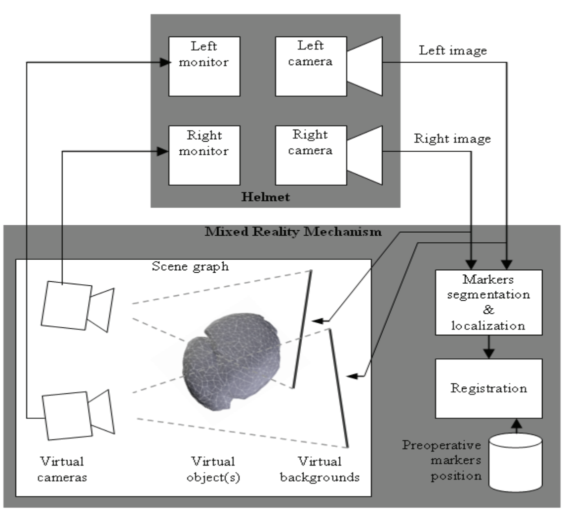


**Figure 7: Schematic representation of our stereoscopic mixed-reality system**

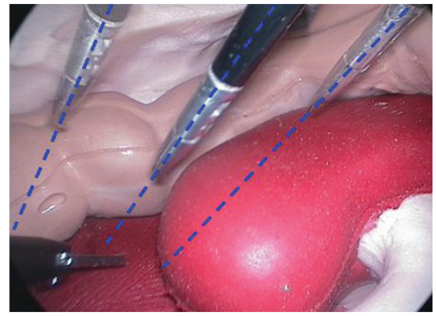

**Figure 8: Image composed by 3 frames of a laparoscopic video with fixed camera and a moving instrument. The projections of instrument axes, represented with blue lines, are constrained to pass through a point representing the projection (on the image plane) of the insertion point (on the abdominal wall)**

Figure 7 shows the functional scheme of our system, where video frames are used, not only as background of the virtual scene, but also to localize the cameras and to register the patient.

Epipolar geometry [13], using two or more cameras, allows to detect the 3D position of each conjugate points, identifiable in the images. In a stereoscopic configuration, knowing the internal camera parameters, for each marker position, in the image plane, the relative projection line in the 3D world, defined as the line $l$ passing through the camera center of projection $O_c$ and lying on the point $P_c$, is determined. These steps, performed both on left and right images, identify respectively two projection lines $l_l$ and $l_r$. Knowing the relative pose of the right camera to the left camera (expressed by a roto-traslation matrix determinable in a calibration phase), the 3D position of each marker is then defined as the intersection point between $l_l$ and $l_r$. Since $l_l$ and $l_r$ do not intersect (due to pixelization process and noise in marker identification) the 3D marker position is approximated with the position of the closest point to both projection lines. After fiducials localization a rigid registration is performed using a point based approach.

Results demonstrate that stereoscopic localization approach, adopted in our system, is enough for system usability.

### Laparoscope auto localization

As described before, localization using monoscopic cameras can be done in case of objects with known geometry or texturing. In case of laparoscopic interventions the localization of the endoscopic camera can be determined using information offered by endoscopic video images without the introduction of any artificial add-on in the scenario[7].

The position and orientation of the endoscopic camera can be determined, with respect to a reference frame fixed to the access ports configuration, elaborating video images and knowing the distances between insertion points. During laparoscopic interventions, camera movements are minor respect to instruments movements. Therefore the laparoscope can be considerate steady in a time interval, and a reference frame fixed on the camera can be used to perform measurements **[21; 28].**

The projections of instrument axis on the image plane (projection lines), which can be simply determined using HSV color space and Hough transform [27], are constrained to pass through the projection of the insertion point on the image plane [28] (figure 8).

Insertion point projection on the image plane can be calculated as the barycentre of the intersection of couples of projection lines, for each instrument. It allows (after camera calibration) to determine the direction of the insertion point in the camera reference frame (Fig. 9 Left). Therefore, versors $\boldsymbol{T_l}$ and $\boldsymbol{T_r}$, representing respectively the direction of the left and the right instrument insertion point, are determined. The versor $\boldsymbol{T_c}$, representing the direction of the camera insertion point, lies on the $Z$ axis of the camera reference frame (using 0 degree optic).

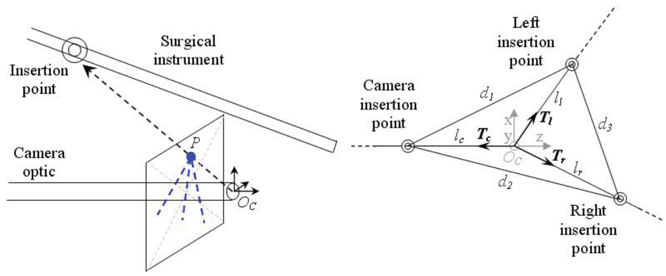


**Figure 9: (Left) The projections of instrument axes (blue lines) allow to calculate the projection of the insertion point on the image plane P, which allows to determinate the direction of the insertion point in the camera reference frame fixed on OC. (Right) Geometric relations involved in the insertion points configuration that allow to localize the laparoscope**

The geometrical relations between $\boldsymbol{T_l}$, $\boldsymbol{T_r}$, $\boldsymbol{T_c}$, and insertion points are shown on the right of figure 9. In the figure $l_c$, $l_l$ and $l_r$ represent distances of the insertion points from the camera origin, which have to be chosen in order to guaranty the distances between access ports $d_1$, $d_2$ and $d_3$. The tetrahedral configuration allows to determine univocally $l_c$, $l_l$ and $l_r$ and consequently, having $\boldsymbol{T_l}$, $\boldsymbol{T_r}$ and $\boldsymbol{T_c}$, to localize the access ports respect to the camera (and vice versa).

The localization accuracy depends on the instruments configuration and on their movements. The proposed solution allows to provide a cheap and tracker-free implementation for a class of computer assisted surgical systems that do not require extremely accurate localization. For example, offering 3D pre-operative model visualization with automatic point of view selection and remote assistance using virtual objects on the laparoscopic monitor.

## CONCLUSIONS

The development of video see-through systems is useful and possible using various approaches.

In order to reduce misalignment errors, between real and virtual world, using commercial trackers, it would be necessary, in the future, the development of endoscopic cameras taking into account the previous considerations. Endoscopes should natively integrate sensors for their localization and manufactures should take into account the stability of the joint between optic and camera body.

On the other hand it is possible the development of tracker-free implementations elaborating camera images, allowing to reduce costs and logistic troubles related to the need of sensors and the tracker in the operating room.

The using of intra-operative imaging devices like 3D RA, which could be diffused in the early future, thanks to the decreasing of their price and the possibility to be portable, will allow to obtain high precision in see-through systems also in case of soft tissues.

## REFERENCES


1. Arun, K. S., Huang, T. S., & Blostein, S. D. (1987). Least-squares fitting of two 3-D point sets. IEEE Trans. Pattern Anal. Mach. Intell., 9(5), 698-700.
2. Baillot, Y., & Julier, S. J. (2003). A tracker alignment framework for augmented reality, In Proc. Second IEEE and ACM International Symposium on Mixed and Augmented Reality (pp. 142-150): IEEE.
3. Besl, P. J., & McKay, N. D. (1992). A Method for Registration of 3-D Shapes. IEEE Trans. Pattern Anal. Mach. Intell., 14(2), 239-256.
4. Coll, D. M., Uzzo, R. G., Herts, B. R., Davros, W. J., Wirth, S. L., & Novick, A. C. (1999). 3-dimensional volume rendered computerized tomography for preoperative evaluation and intraoperative treatment of patients undergoing nephron sparing surgery. J Urol, 161(4), 1097-1102.
5. Doignon, C., Nageotte, F., Maurin, B., & Krupa, A. (2007). Model-based 3-D pose estimation and feature tracking for robot assisted surgery with medical imaging, From Features to Actions - Unifying Perspectives in Computational and Robot Vision, Workshop at the IEEE Int. Conf. on Robotics and Automation.
6. Ferrari, V., Carbone, M., Cappelli, C., Boni, L., Cuschieri, A., Pietrabissa, A., et al. (2010). Improvements of MDCT images segmentation for surgical planning in general surgery - practical examples. Proceedings of the International Congress and Exhibition. IJCARS Volume 5, Supplement 1 / June.
7. Ferrari, V., Megali, G., Pietrabissa, A., & Mosca, F. (2009). Laparoscope 3D auto-localization. Proceedings of the International Congress and Exhibition. IJCARS Volume 4, Supplement 1 / June.
8. Ferrari, V., Megali, G., Troia, E., Pietrabissa, A., & Mosca, F. (2009). A 3-D mixed-reality system for stereoscopic visualization of medical dataset. IEEE Trans Biomed Eng, 56(11), 2627-2633.
9. Freschi, C., Ferrari, V., Porcelli, F., Peri , A., Pugliese, L., Morelli, L., et al. (2010). An Augmented Reality

Navigation Guidance for High Intensity Focused Ultrasound Treatment. Paper presented at the Conf Proc ICABB, International Conference on Applied Bionics and Biomechanics 2010, Venice, Italy.

10. Freschi, C., Troia, E., Ferrari, V., Megali, G., Pietrabissa, A., & Mosca, F. (2009). Ultrasound guided robotic biopsy using augmented reality and human-robot cooperative control. Conf Proc IEEE Eng Med Biol Soc, 2009, 5110-5113.

11. Gao, C., & Ahuja, N. (2004). Single camera stereo using planar parallel plate, Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4 (pp. 108-111).

12. Genc, Y., Sauer, F., Wenzel, F., Tuceryan, M., & Navab, N. (2000). Optical see-through HMD calibration: A stereo method validated with a video see-through system, International Symposium for Augmented Reality.

13. Hartley, R., & Zisserman, A. (2004). Multiple View Geometry in Computer Vision.

14. Hawkes, D. J., Penney, G. P., Atkinson, D., Barratt, D. C., Blackall, J. M., Carter, T. J., et al. (2007). Motion and Biomechanical Models for Image-Guided Interventions, ISBI (pp. 992-995).

15. Hinson, W. H., Kearns, W. T., Ellis, T. L., Sprinkle, D., Cullen, T., Smith, P. G., et al. (2007). Reducing set-up uncertainty in the elekta stereotactic body frame using stealthstation software. Technology in cancer research & treatment, 6(3), 181-186.

16. Hua, H., Krishnaswamy, P., & Rolland, J. P. Video-based eyetracking methods and algorithms in head-mounted displays. Optics Express, 14, 4328-4350.

17. Johnson, L., Philip, E., Lewis, G., & Hawkes, D. (2004). Depth perception of stereo overlays in image-guided surgery, Medical Imaging, Proceedings of the SPIE, Volume 5372, pp. 263-272.

18. Lee, E. C., & Park, K. R. (2008). A robust eye gaze tracking method based on a virtual eyeball model. Machine Vision and Applications.

19. Megali, G., Ferrari, V., Freschi, C., Morabito, B., Cavallo, F., Turini, G., et al. (2008). EndoCAS navigator platform: a common platform for computer and robotic assistance in minimally invasive surgery. The International Journal of Medical Robotics and Computer Assisted Surgery, 4, 242 - 251.

20. Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. IEICE transactions on information and systems, 77(12), 1321-1329.

21. Naogette, F., Zanne, P., Doignon, C., & De Mathelin, M. (2006). Visual Servoing-Based Endoscopic Path Following for Robot-Assisted Laparoscopic Surgery, International Conference on Intelligent Robots and Systems (IROS).

22. Olbricha, B., Traubb, J., Wiesnerb, S., Wicherta, A., Feussnera, H., & N.Navabb. (2005). Respiratory Motion Analysis: Towards Gated Augmentation of the Liver, CARS 2005 Computer Assisted Radiology and Surgery, 19st International Congress and Exhibition.

23. Peters, T. M. (2006). Image-guidance for surgical procedures. Physics in Medicine and Biology, 51(14), R505-R540.

24. Peters, T. M. (2000). Image-guided surgery: from X-rays to virtual reality. Computer Methods in Biomechanics and Biomedical Engineering, 4(1), 27-57.

25. Pietrabissa, A., Morelli, L., Ferrari, M., Peri, A., Ferrari, V., Moglia, A., et al. (2010). Mixed reality for robotic treatment of a splenic artery aneurysm. Surg Endosc, 24(5), 1204.

26. Shuhaiber, J. H. (2004). Augmented Reality in Surgery. Archive of surgery, 139, 170-174.

27. Tonet, O., Ramesh, T. U., Megali, G., & Dario, P. (2007). Tracking endoscopic instruments without localizer: a shape analysis-based approach. Computer Aided Surgery, 12(1), 35-42.

28. Voros, S., Long, J.-A., & Cinquin, P. (2006). Automatic Localization of Laparoscopic Instruments for the Visual Servoing of an Endoscopic Camera Holder, MICCAI (1) (pp. 535-542).

29. Wengert, C., Bossard, L., Baur, C., Szekely, G., & Cattin, P. C. (2008). Endoscopic navigation for minimally invasive suturing. Comput Aided Surg, 13(5), 299-310.

30. Zhang, Z. (2000). A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 1330-1334.