

# Ultrasound-based Servoing of Manipulators for Telesurgery

Jeff Stoll<sup>a</sup>, Pierre Dupont<sup>a</sup> and Robert Howe<sup>b</sup>

<sup>a</sup>Department of Aerospace and Mechanical Engineering, Boston University, Boston MA

<sup>b</sup>Division of Engineering and Applied Sciences, Harvard University, Cambridge MA

## ABSTRACT

Certain minimally invasive surgical procedures involve the treatment of highly precise target locations within deformable tissues. While preoperative MRI and CT models can be used for surgical planning, they provide only coarse guidance during surgery due to their limited resolution and owing to tissue deformation. Ultrasound imaging is a promising means of obtaining real-time intraoperative data for target localization that is particularly well suited to minimally invasive surgery due to its portability, speed, and safety. This paper presents a system, in which ultrasound images are used to guide a manipulator to a surgical site. Electromagnetic tracking of the ultrasound probe is used to orient the images. These are then segmented in real time to determine target locations. Finally, target coordinates are used to produce control inputs to drive the manipulator to the target site. The potential of the approach is demonstrated experimentally using a manipulator arm, phantom target, and commercial ultrasound machine.

**Keywords:** Teleoperation – telesurgery – surgery – minimally invasive – ultrasound – imaging – servoing – manipulation.

## 1. INTRODUCTION

This paper presents a demonstration of a technology for real-time image-based robot guidance during minimally invasive surgical procedures. In this approach, a robotic arm is automatically guided to a target located in 3 dimensions using ultrasound images obtained by hand scanning. The technology is applicable to many possible surgical procedures involving guidance of robotic tools to deformable tissue targets. Owing to motion arising from physiological processes as well as robot motion, such targets cannot be located using only preoperative scanning and registration.

Minimally invasive techniques attempt to reduce trauma and other morbidity due to surgery by minimizing both the size of incisions and the amount of exploratory dissection. An example is laparoscopic surgery, wherein a surgeon makes three or four 25 mm incisions in the abdomen. Long, thin instruments are passed through two or three of these incisions so that the surgeon can grasp or cut tissue inside the body by manipulating the tool handles outside the body. Through the remaining port, the surgeon passes a long, thin camera (endoscope), with which he can see what is happening inside the body. This type of procedure can allow a patient to return to normal activity within a day after surgery, whereas previous techniques required weeks of convalescence. These minimally invasive techniques require fast imaging technologies, precise instruments and careful tool movement for success. The project presented combines two of these elements: imaging for surgical guidance and control of surgical robotic manipulators.

Surgical imaging comes in many forms, including but not limited to, endoscopy, magnetic resonance (MR), computed tomography (CT), x-ray, and ultrasound. Among these options, only ultrasound combines high speed, portability, high resolution, and safety<sup>11</sup>. In addition, there are a number of situations, in which the other types of imaging are simply inadequate. Off-bypass cardiac surgery, for instance, requires detailed views of tissue inside the heart. In this case, intraoperative CT, MR, and x-ray imaging have either inadequate frame rates or involve relatively high levels of harmful radiation. In addition, the nearly opaque blood flowing in the heart precludes endoscopy. Ultrasound imaging, however,

Contact information: Jeff Stoll ([jstoll@bu.edu](mailto:jstoll@bu.edu)) and Pierre Dupont ([Pierre@bu.edu](mailto:Pierre@bu.edu)): 110 Cummington St. Boston, MA 02215. Robert Howe ([howe@deas.harvard.edu](mailto:howe@deas.harvard.edu)).

is well suited for the task. It produces detailed images of fluid-filled mediums including blood and soft tissue. Therefore, analysis of heart structure is possible, as is real-time coordination of tool movement with visual feedback.

Nevertheless, ultrasound does have some important drawbacks. Most notable is the low signal to noise ratio inherent to ultrasound images. Due to the slight unpredictability of tissue structure, the statistical analysis of the reflected sound waves tends to produce noise or speckle, which can impede image analysis<sup>11</sup>. In addition, abnormally reflective objects can produce artifacts, bright reflections, which misrepresent the tissue being scanned. The difficulty of interpreting ultrasound images requires surgeons to perform a number of training cases to become proficient, which take time and concerted effort. This has proved a major hindrance to the widespread incorporation of ultrasound into surgery.

## 2. BACKGROUND

The technology presented relies on three fundamental concepts: tracked freehand ultrasound, real-time segmentation, and multi-modal registration. Freehand ultrasound systems have been used previously for the construction of 3D ultrasound image volumes<sup>2,3,6</sup>. In past studies, tracking has been accomplished mechanically, optically, and magnetically with varying levels of accuracy and expense<sup>6,7,8</sup>. The key component of every system, however, is accurate calibration between the reference frame of the tracking unit and the ultrasound image plane. The mathematical basis for calibration assumes a linear transformation (1), in which  $x_I$  represents the image-plane coordinates of an object and  $x_W$  represents the equivalent coordinates in the tracking unit frame.

$$x_W = Tx_I \quad (1)$$

A variety of calibration methods are possible involving both active scanning of single objects or surfaces and stationary images of multiple objects<sup>8,10</sup>. Objects are located both in the ultrasound images and in the tracking unit frame, and the transformation  $T$  is calculated numerically. In particular, methods employing only large-scale features such as surfaces are promising for clinical applications since they do not rely on specially-built calibration devices, which is an important part of reducing the complexity of clinical systems.

For control applications, image analysis requires the separation of images into regions corresponding to the control task. This process is known as image segmentation. There are three basic means for accomplishing this segmentation: thresholding, boundary identification, and texture identification<sup>4</sup>. The noisiness and general inconsistency of ultrasound images, however, make the standard techniques difficult to implement efficiently. In some cases, the images are filtered to eliminate the high frequency effects of speckle and better represent the large structures visible in an image. In other cases, special algorithms are devised for highlighting visible edges<sup>4,5</sup>. In all cases, the important factor for real-time segmentation is the speed of the algorithm. This concern is amplified when many algorithms are combined to fully segment an image.

The final component needed for the tracking of ultrasound images is the ability to match coordinates in these images with those of other imaging devices as well as those of robotic tools. The matching process is referred to as registration. Work on this topic has been presented in the field of frameless stereotactic navigation in which ultrasound images are registered to preoperative MR and CT images of neurosurgery patients<sup>8</sup>. Similar techniques are being applied for abdominal surgery, as well. This registration process is similar to calibration; a linear transformation must be found that both translates and rotates the base frame of the ultrasound images to the base frame of the other data set or device. Difficulties arise, however, due to the fact that distortions in either frame, e.g., due to deformations, occur in three dimensions instead of two. In addition, it can be difficult to find common points in both frames with high precision, given complex images or mechanical geometries.

Limited prior work has been done with ultrasound based servoing. Most notable is that of S. E. Salcudean, in which image servoing was used to hold a feature, such as an artery, in view while scanning its extent<sup>1,12</sup>. This servoing was performed in the plane, however, and did not deal with 3D localization of objects with finite extent. In addition, the ultrasound probe was fixed to the robot being controlled, which does not require registration.

The remainder of this paper illustrates how ultrasound-based servoing was achieved. Section 3 describes the design of the servoing system, including implementations of calibration, segmentation and registration. Section 4 describes the verification of the system’s accuracy, as well as a means by which to evaluate similar systems in the future. Section 5 summarizes the results.

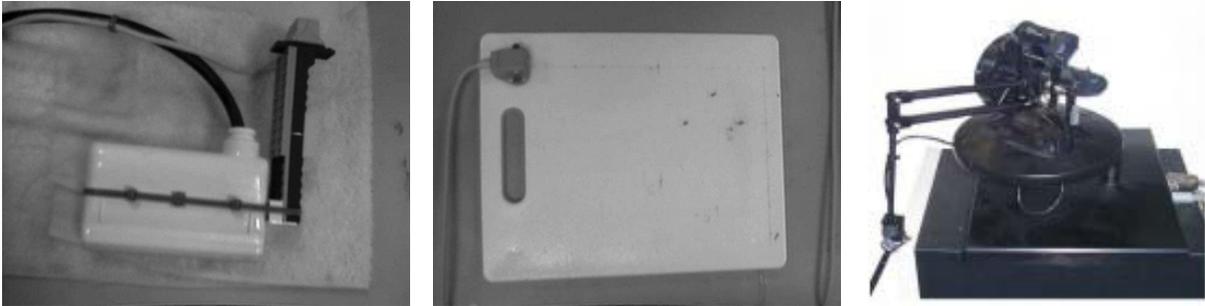
### 3. SYSTEM DESIGN

#### 3.1 Description

The system consists of two personal computers, a Hitachi EUB-310 ultrasound machine, an Ascension Technologies DC electromagnetic tracking system, and a Sensable Technologies PHANToM™ haptic interface, used as a robotic manipulator. One computer interfaces with the ultrasound machine and tracking system and performs the image processing. The second computer is dedicated to control of the PHANToM™, receiving target positions and other control variables over a unidirectional serial link running at 115,200 *bps*.

The image acquisition is performed by a Belkin VideoBusII™ USB analog video capture device acquiring approximately 30 160×240 frames per second. The user interface is written in tel/tk script, and the image segmentation and optional display is performed in C++ via the Visualization Toolkit libraries distributed by Kitware.

The electromagnetic tracking system consists of a fixed base transmitter and two remote tracking receivers. Inside both transmitter and receiver are three orthogonal coils. Each transmitter coil is pulsed individually and the resultant fields are detected at the receivers. The six degrees of freedom of each receiver (i.e., three rotations, three translations), relative to the base transmitter, are transmitted over a serial line to one of the computers. One receiver is mounted on the ultrasound scan head. It includes an offset, as shown in Figure 1(a), so that electromagnetic noise produced by the transducers does not affect the position and orientation measurements. The second receiver is initially attached to a fixture for calibration (Figure 1(b) and section 3.2), and then to the robot for registration and object tracking (Figure 1(c) and section 3.4).



(a): Ultrasound probe with offset tracking receiver attached.

(b): Calibration fixture with tracking receiver and calibration pins attached.

(c): PHANToM™ device used as robotic probe.

Figure 1

The PHANToM™ device is controlled by the second computer. Control is based on simple proportional derivative control with an error limit rule to prevent excess velocities during long motions (2).

$$F = \begin{cases} k_p e + k_v \dot{e}, & |e| \leq 1 \\ k_p e / |e| + k_v \dot{e}, & |e| > 1 \end{cases} \quad (2)$$

### 3.2 Calibration

The tracking receiver on the ultrasound probe was calibrated using the fixture shown in Figure 1(b), which consists of three markers protruding from a flat plastic board. As the marker locations are known precisely relative to a tracking receiver attached to the board, their world coordinates are easily computed from the receiver's coordinates.

To achieve calibration, the fixture was submerged in fluid and the ultrasound scanner was held with the image plane parallel to the board such that the markers appeared as bright dots in the ultrasound image as seen in Figure 2(a). The pixel coordinates of the three dots in the image were manually determined. Using the ratio of the distances between the markers in pixel and fixture coordinates, the scaling factor between world and pixel coordinates was found to be 0.406 mm per pixel. A linear transformation, consisting of a translation and rotation, was then computed to map the marker coordinates from the fixture frame to the image frame.

Thus, the total transformation,  $T_{FI}$ , from fixture to image frames consists of translation, rotation and scaling. This transformation holds for a fixed relative position and orientation of the probe and fixture. Since both have tracking receivers attached, this constant relative transformation from probe to fixture can be computed as  $T_p^{-1}T_F$  where  $T_p$  and  $T_F$  are the world coordinate transformations of the probe and fixture receivers, respectively. The calibration transformation,  $C$ , which transforms image coordinates to probe coordinates is given by:

$$C = (T_p^{-1}T_F)T_{FI} \quad (3)$$

The calibration transformation used in what follows was taken as the mean over multiple calibration scans involving a variety of orientations between the probe and fixture.

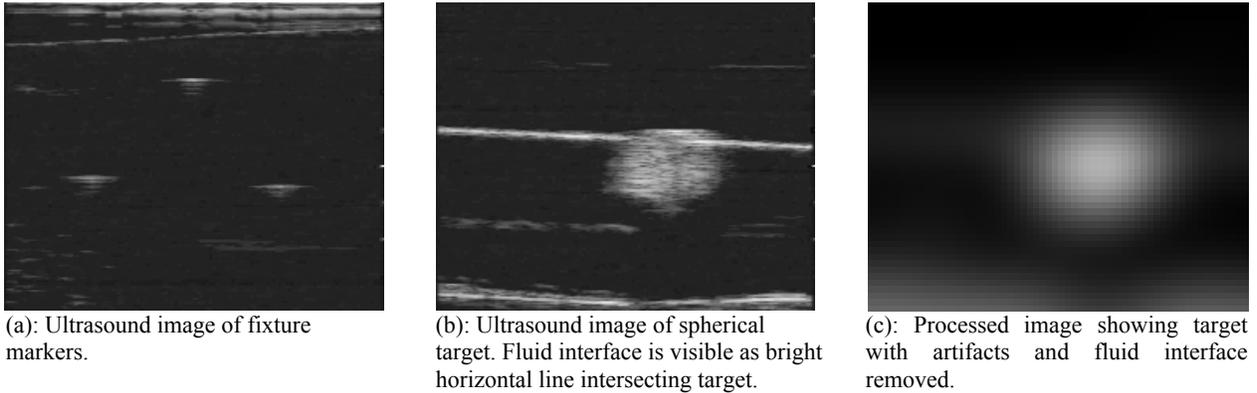


Figure 2

### 3.3 Segmentation

Real-time segmentation of the ultrasound image involved five distinct steps through which the information content of the image was reduced to arrive at the desired information, namely, the center coordinates of a circle corresponding to a cross section of the spherical target. First, the ultrasound image gain was adjusted to minimize unnecessary noise and artifacts. The result is shown in Figure 2(b). Gain adjustment, however, is not a robust means for removing image artifacts. Consequently, in the second step, the image was threshold segmented to eliminate pixels containing no information, i.e., to eliminate nearly black pixels and preserve bright pixels.

The third processing step was to subsample the image by a factor of  $\frac{1}{4}$ . This was done to reduce the computational load during the fourth step in which the image was convolved with a 2D Gaussian kernel of radius 10 and standard deviation 5 (pixels). This step increased the brightness of pixels interior to white regions in accordance with their distance to the border of the region. Thus, pixels in the center of white regions appeared brightest after convolution. To apply this technique to the spherical targets, it was necessary to use a convolution kernel as large as the image target size to achieve

a unique brightest pixel. The final segmentation step was to search the image for the pixel of maximum intensity. The coordinates  $x_i$  of this pixel were extracted and transformed to world coordinates,  $x_w$ , using the transformation

$$x_w = T_p C x_i \quad (4)$$

to yield the world coordinates of the cross section's center.

Two additional calculations were performed to achieve consistent target detection over the range of scan head movement. First, since many cross sections were taken of the target each second, the 3D coordinates of the target's center were taken as the running average of the cross sections' centers. Second, both the scan head and target were capable of moving in a large area such that the target could intermittently go in and out of view. To consistently determine the presence of the target in the ultrasound image, a threshold value of 100 was chosen such that a maximum pixel value greater than it would indicate the presence of the target. This value was determined empirically to eliminate the effect of bright artifacts. As such, it was a good method for avoiding false detections, but it did not always detect the target when it was in view with a small cross section.

### 3.4 Registration

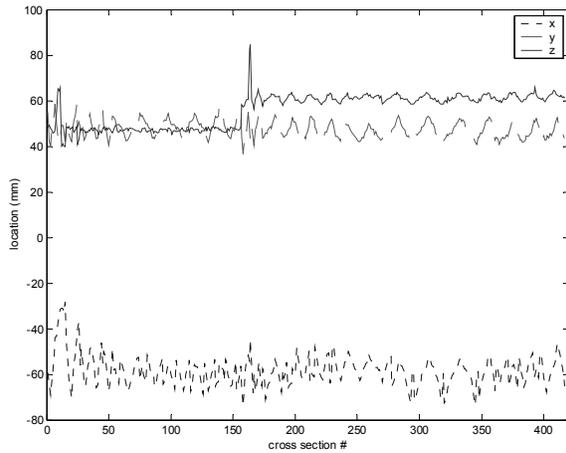
To achieve registration between the robot tip and the world coordinates as defined by the tracking system, a tracking receiver was rigidly attached to the tip of the PHANToM™. The arm was then moved through a large volume of its workspace while tip coordinates were collected using both the robot's joint sensors and the tracking system. The coordinates of the robot's base were then estimated by a least squares comparison of the tip coordinates. Due to slight errors in measurement and nonlinearities inherent in robot initialization, it was necessary to fine tune the transformation so calculated using a manual iterative process. Manual tuning was achieved by tracking a target with the ultrasound system and modifying the transformation matrix elements individually to specify correct scaling factors between the coordinate axes in each frame.

## 4. SYSTEM VERIFICATION

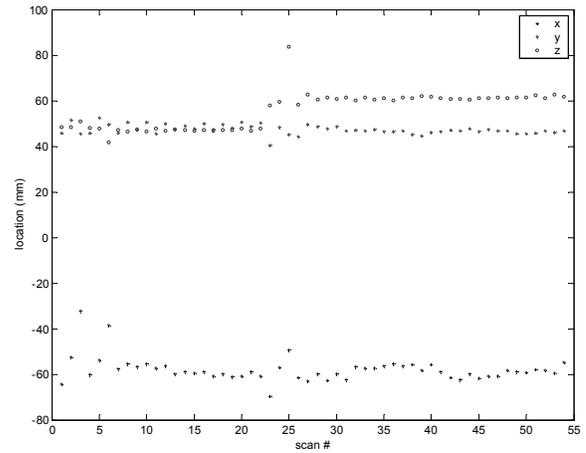
To demonstrate system potential, two experiments were carried out to assess tracking performance and targeting accuracy. Both involved position control of the robot tip using target coordinates estimated automatically from hand scanning of the target. The environment in which the scanning and targeting took place consisted of a plastic container (23×30×45 cm.) in which several cm. of vegetable oil floated on several cm. of saturated salt water at room temperature. A grape of approximately 20 mm diameter served as the roughly spherical target. Due to the density difference between the oil and salt water, and their immiscibility, the grape floated at the fluid interface, held in place by forces of gravity and buoyancy. Both fluids are good mediums for ultrasound transmission and the fluid interface provided a roughly planar surface over which the target grape could move freely.

To assess tracking performance, the robot was initialized at a distance of 280 mm from the target and commanded to move to the target's coordinates. Sixty trials were conducted and the robot touched the target 53 out of 60 times (88%), indicating a reasonable degree of repeatable accuracy. Those trials for which the robot missed the target exhibited systematic errors suggesting a robot calibration problem due to the relative joint encoders used on the PHANToM™.

To evaluate targeting accuracy, a second set of experiments was performed in which the target was held in place by a wire and the robot was commanded to coordinates offset from the target's center by 25 mm in the  $x$ -direction. In this configuration, system accuracy could be measured since the robot was guided based on image coordinates, but without disturbing the target. Robot tip position was recorded over the course of many scans. Each scan traversed the full extent of the target, such that the starting and ending images of each scan did not contain the target. In addition, the number of images in a scan that contained target cross sections (number of cross sections per scan) was varied by changing the speed at which the scan head was moved. To assess both transient and steady-state response, the target was deliberately moved about 13 mm midway through the experiment.



(a): Target coordinates versus cross section.



(b): Scan-averages target coordinates versus scan.

Figure 3

To gain a better understanding of targeting accuracy, it is worthwhile to compare the actual error with the theoretical minimum due to the finite imaging rate. The latter has the effect of discretizing the target shape in accordance with the scanning rate. Assuming the ultrasound probe is moved at a constant rate throughout the scan, and the system takes parallel cross-sectional measurements at constant intervals, the upper bound of the uncertainty is equal to half the distance,  $d$ , between each cross section. This can be seen by noting that a set of equally spaced cross sections comprising one scan can be shifted by up to  $\pm d/2$ , as shown in Figure 4.

Taking this expected uncertainty into account, the best measure of targeting accuracy is the position measurement error as a function of the distance between scan cross sections. To obtain this measurement, the actual location of the target was assumed to be the mean over all scans. Position error was computed as the distance between this global mean and the mean for each scan. Scans were sorted according to the number of cross sections they contained (and thus sorted by scan speed). The means and standard deviations for each set were computed and are plotted as shown in Figure 5. The lower bound on uncertainty is also shown in the figure. As the number of cross sections per scan increases, the lower bound on uncertainty approaches zero asymptotically. From the figure, this suggests that system targeting accuracy approaches a minimum of approximately 2 mm.

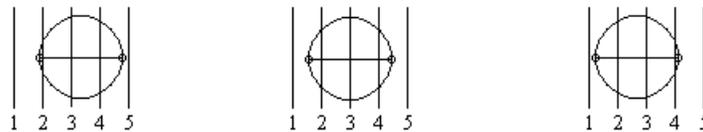


Figure 4: Illustration of lower bound on targeting accuracy due to cross section shift.

## 5. DISCUSSION AND CONCLUSIONS

The experimental results indicate that the imaging system presented can be used effectively to guide a robotic manipulator to a target. While the work reported represents initial results, it is encouraging that the accuracy of the measurements is quite good. The errors during the robot tracking trials when the robot failed to touch the target were most likely due to slight errors in the calibration and registration transformations. These elements of the system can be

easily improved. It is reasonable to assume that with more exact techniques the measurement error of the system could be considerably reduced.

The segmentation method used has also been shown to be highly effective. The simple method for finding the center of a sphere by calculating the center of its cross sections can easily be extended to all objects with symmetric convex cross sections, such as ellipsoids, rectangles, and cylinders. In addition, more complicated shapes can be addressed through minor increases in sophistication. Thus, the tracking method proposed can be applied to a wide class of target shapes.

The system presented is an encouraging first step towards ultrasound-based servoing of robotic manipulators. It was shown that a system using simple calibration, segmentation, and registration techniques is capable of targeting and approaching an object with a high degree of accuracy. Ultimately, it is hoped that such a system could aid surgeons in automatically guiding robotic devices to perform tasks. Ultrasound-based robot control during minimally invasive surgery could enhance the skills of a surgeon by achieving precise manipulator control for a wide range of tasks.

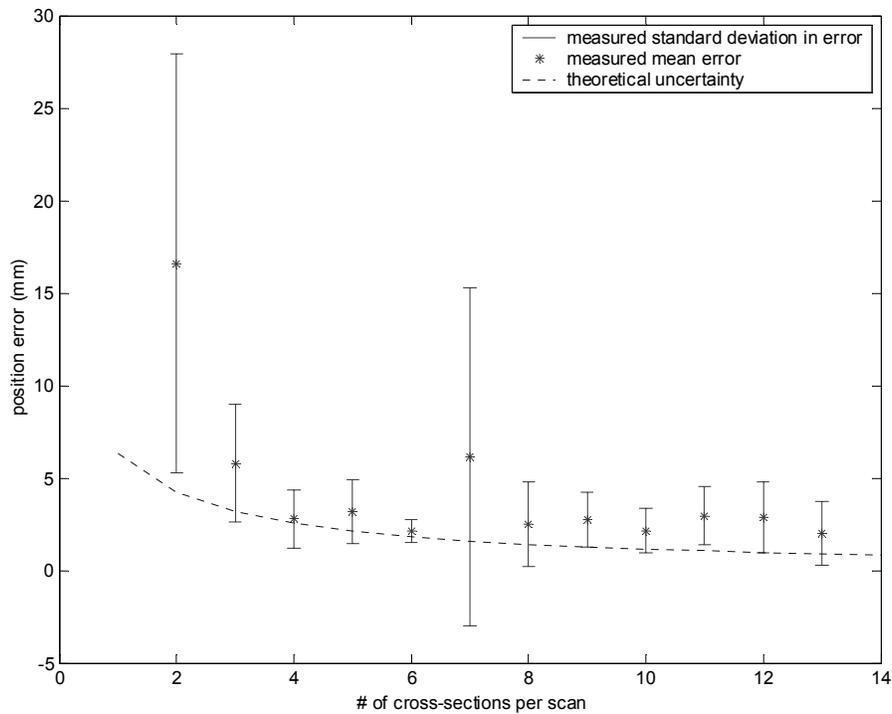


Figure 5: Scan position error versus cross section per scan.

## 6. REFERENCES

1. Abolmaesumi, P, S. Salcudean and W. Zhu. "Visual Servoing for Robot-assisted Diagnostic Ultrasound." Proc. IEEE Engineering in Medicine and Biology, pp. 2532-2535, 2000.
2. Aiger, Dror and Danial Cohen-Or. "Mosaicing Ultrasound Volumes for Visual Simulation." IEEE Computer Graphics and Applications, pp. 53-61, 2000.
3. Carr, Jonathan et al. "Design of a Clinical Free-Hand 3D Ultrasound System" SPIE 3982, pp. 14-25, 2000.
4. Heucke, Lars, Mirko Knaak and Reinhold Orglmeister. "A New Image Segmentation Method Based on Human Brightness Perception and Foveal Adaptation." IEEE Signal Processing Letters v.7 no.6, pp. 129-131, June 2000.

5. Kaspersen, Jon, T. Lango and F. Lindseth. "Wavelet-based Edge Detection in Ultrasound Images." *Ultrasound in Medicine and Biol* v.27 no.1, pp. 89-99, 2001.
6. Nelson, Thomas R. and Dolores Pretorius. "Three-dimensional Ultrasound Imaging." *Ultrasound in Medicine and Biol* v.24 no. 9, pp. 1243-1270, 1998.
7. Pagoulatos, N, R. Rohling, W. Edwards and Y. Kim. "A New Spatial Localizer based on Fiber Optics with Applications in 3D Ultrasound Imaging." *SPIE* 3976, pp. 595-602, 2000.
8. Pagoulatos, N, W. Edwards, D. Haynor and Y. Kim. "Interactive 3D Registration of Ultrasound and Magnetic Resonance Images Based on a Magnetic Position Sensor." *IEEE Trans. Info. Tech. in Biomed.* v.3 no.4, pp. 278-288, Dec. 1999.
9. Pathak, Sayan D. and Yongmin Kim. "Ultrasound Image Contrast Enhancement via Integrating Transducer Position Information." *SPIE* 3976, pp. 563-572, 2000.
10. Prager, R, R. Rohling, A. Gee, and L. Berman. "Rapid Calibration for 3D Freehand Ultrasound." *Ultrasound in Medicine and Biol.* v.24, pp. 855-869, July 1998.
11. Wells, Peter N.T. "Current Status and Future Technical Advances of Ultrasonic Imaging." *IEEE Engineering in Medicine and Biology*, pp. 14-20, Sept/Oct. 2000.
12. Zhu, W, S. Salcudean, S. Bachmann and P. Abolmaesumi. "Motion/Force/Image Control of a Diagnostic Ultrasound Robot." *ICRA*, pp. 1580-1585, April 2000.