Development and Human Factors Analysis of Neuronavigation vs. Augmented Reality

Abhilash Pandya\textsuperscript{1, 3}, Mohammad-Reza Siadat\textsuperscript{2}, Greg Auner\textsuperscript{1, 2}, Mohammad Kalash\textsuperscript{1}, R. Darin Ellis\textsuperscript{3, 4}

\textsuperscript{1}Electrical and Computer Engineering Department, \{apandya, gauner, mkalash\}@ece.eng.wayne.edu
\textsuperscript{2}Computer Science Department msiadat@cs.wayne.edu
\textsuperscript{3}Biomedical Engineering rdellis@wayne.edu
\textsuperscript{4}Industrial and Mfg. Engineering
Wayne State University, Detroit, MI 48201, USA

This paper is focused on the human factors analysis comparing a standard neuronavigation system with an augmented reality system. We use a passive articulated arm (Microscribe, Immersion technology) to track a calibrated end-effector mounted video camera. In real time, we superimpose the live video view with the synchronized graphical view of CT-derived segmented object(s) of interest within a phantom skull. Using the same robotic arm, we have developed a neuronavigation system able to show the end-effector of the arm on orthogonal CT scans. Both the AR and the neuronavigation systems have been shown to be within 3mm of accuracy. A human factors study was conducted in which subjects were asked to draw craniotomies and answer questions to gauge their understanding of the phantom objects. The human factors study included 21 subjects and indicated that the subjects performed faster, with more accuracy and less errors using the Augmented Reality interface.

1.0 Introduction

Currently Neuronavigation and robotics systems like the Neuromate (Integrated Surgical Systems Inc) \cite{1} provide primarily three 2-D views (coronal, axial and sagittal views) to gain awareness of the patient’s geometry. The surgeon has to perform the 2D (image) to 3D transformation in their minds and also project the envisioned data on the view of the patient. We believe that Augmented Reality (AR) generation is a natural extension for the surgeon because it does both the 2D to 3D transformation and projects the views directly on the patient view. To illustrate the difficulty to interpret 2D slices, in figure 1(a), a simple 3D shape like the cube is represented by a triangle in the coronal slice and a vessel is represented by two dots in the axial view. Currently, the surgeon must convert these views to a 3D representation and merge it with what he physically sees. This scene represents a virtual reality scene where the actual view is not presented. Figure 1(b) represents a live video view of a phantom skull with the models of interest displayed directly on the view. This represents an Augmented Reality view because the real world view is presented and is augmented with additional geometrical information.

It is a combination of the real scene viewed by the user and a virtual scene generated from 3D geometry/segmentation accurately co-registered on the display that augments the scene with additional information. Recently the real-time video processing and computer graphics have provided us with the capability of augmenting the video stream with geometrical replicas of the actual objects or sensor data. In neurosurgery, for instance, critical objects of interest within the patient’s brain determine exactly the size, shape, location and orientation of the craniotomy to be performed. The objects of interest can be tumor, major vessels, or anatomically/physiologically important brain structures. Before the course of surgery, there are usually several sets of image data (different modalities of MRI, CT, SPECT, Functional MRI, etc.) scanned. The image data provides very important information about the spatial arrangements and the functional importance of objects of interest within the brain in the image space. Through the process of segmentation and model extraction, these objects/regions can be represented by 3D computer graphics models that can be used as the input to the AR system.
2.0 Related Work

A landmark paper on registration methods for Image Guided Surgery and Enhanced Visualization was presented by Grimson et al. [2]. In this paper, an augmentation scene of a static camera system was produced using a laser range scanner and a video camera. In their paper, they state that augmentation of a stationary video camera is relatively straightforward, however, [dynamic] tracking of the camera is “more relevant and more challenging”. They report a registration RMS error of their system at 1.6mm, but, admit that this error is the error of data fitting and that it was difficult to ascertain the actual registration error and that their future work would include “some kind of phantom [accuracy] study”.

Raya et al. [3] have proposed an AR prototype to replace the traditional optical microscope view with a digital one. For their AR prototype, they use, a) an infrared tracking system, b) two video cameras tracked by infrared LEDs. They consider two types of error for their prototype error analysis: 1) object space error, 2) camera calibration error. Neither of which can provide a reliable foundation for error analysis. No comparative study with current neuronavigation systems is provided.

Hattori et al. [4] have developed a data fusion system for the robotics surgery system “da Vinci,” composed of an optical 3D location sensor and a digital video processing system. In a clinical situation, however, the proposed system needs to be calibrated/registered to calculate the transformation from the optical marker to the camera. This extra step should be taken each time before the course of surgery. In their paper, they don’t present a comprehensive results section and any accuracy study of the da Vinci system and an infrared tracking system to generate an AR scene.

Khamene et al. [5] have developed an AR system for MRI-guided needle biopsy. Their main goal was to reduce or completely remove the need for the interventional scanning (by a high field closed magnet MRI) as well as the need for an open MRI scanner from the biopsy procedure. Their system consists of: 1) one video-see-through head mounted display (HMD), 2) two video cameras attached to the HMD to provide a stereo view of the scene, 3) a third video camera for tracking, 4) a set of optical markers attached to the patient’s bed, 5) a set of optical markers attached to the needle. In their analysis, they overlay the model of the skin of the patient on the patient. It seems to us that having the patient’s skin overlaid by a needle does not really augment any useful information for a biopsy procedure. What may be essential for this application is mostly the needle in the MRI image space during the course of the surgery, which does not need AR. They have reported an accuracy study as good as 1 mm for the whole system. Our experience suggests that typical tracking devices alone are on this order of error. They have pointed out that for a small number of cases where the accuracy is substantially larger than 1 mm, the error was most likely caused by needle bending. No study comparing standard neuronavigation system to the AR system is provided.
In their paper on a data fusion environment for multimodal neuronavigation, Jannin et al [6] briefly experimented with AR techniques as applied to the Zeus Microscope. They used projected 2D contours in the focal plane of the right ocular of the microscope. The main limitation that they noted was that no information about structures before or behind this plane were visible and that different contours could not be visualized with different colors, line widths, or labels. No error estimates were provided for the augmentation technique they used, neither was it compared to Neuronavigation.

None of the researchers have compared the utility of their technology to standard neuronavigation systems using human subjects. To our knowledge, this is the first study of its kind. We feel it is important that technology development be followed with (or be developed in conjunction with) subject testing and evaluation. This process will not only prove (or disprove) the utility of the technology, but, help in the development of optimal user interfaces and assist with the selection of relevant data and the appropriate display formats.

3.0 Materials and Methods

Augmentation represents an extension to image-guided surgery by integrating an end-effector mounted camera. Image-guided surgery utilizes a method that transforms tracked tools trajectories such that they can be displayed to provide accurate and reliable image-to-surgical space guidance. The methodology involves three components: image acquisition with definition of coordinated space from one or several imaging modalities, planning or simulation of the surgical procedure, and intraoperative surgical procedures, which include the determination of the spatial relationship between the image and the surgical coordinate space (patient registration). In figure 1a, we show an image guided system we have integrated in which a socket interface to the Microscribe allows multiple clients on the internet to view the tracking of the device end-point on the 3D slices and graphical representation of the phantom in 3Dslicer. The critical component in interactive image-guided surgery is the use of an intraoperative localizer system or a digitizer, which ultimately provides the surgeon useful navigational information usually in the form of position and/or orientation of surgical instruments in the image space. Infrared tracking is one of the most popular methods used in stereotactic neurosurgery. Robotics (as in the case of the Neuromate System) is another very good method. The key advantages of robots are that they can effectively position, orient, and manipulate surgical tools in 3D space with a high level of accuracy. Using the well-developed and now standard techniques of image guidance (Figure 2a) and adding a photographic camera, camera registration software, and video/graphics mixing software (Figure 2b), an AR scene can be generated.(Figure 2c).

The steps taken to generate a medical AR scene are illustrated in Figures 2 and 3. After image data collection, segmentation of the objects of interest in the image space and generating 3D graphical models out of the segmentations called virtual objects are needed for augmentation. This step can be performed for image guidance also, but, it is not a necessary step since the slices of the imaging modality can suffice for navigation (and frequently are the only ones used for navigation). The key differences between Neuronavigation and Augmented Reality is the camera parameter estimation, the camera mounting and the mixing of the live video and graphics. Camera parameter estimation (camera calibration) to build up a virtual camera that closely models the actual camera and the mounting and pose estimation of the camera needs to be done only once for the system, but, can be done periodically in order to verify the camera system. These two steps will be covered in detail. When generating the graphical view of the virtual objects the relative spatial positions and orientations of the virtual camera and virtual objects at this step are the same as those of the actual camera and objects of interest. Furthermore, the virtual camera and virtual objects very accurately model their actual replicas. This allows the graphical view of the virtual objects to be seen from the actual camera point of view on the live video if the objects of interest are visible. This mixing of the live video from the actual camera with the graphical view from the virtual camera, enables the surgeon to see the segmented objects of interest projected on the video view from any perspective and at any depth.

3.1 Imaging, Segmentation, and Model Creation

Creating the virtual objects is a necessary step for generating an Augmented Reality environment based on imaging. For medical applications, each patient has a unique set of objects that can be used for augmentation. Usually these objects are tumors, skin surfaces, or a set of major vessels, and relevant normal or abnormal structures in the brain. We define our object model by a segmentation procedure that uses medical imaging data (See Figure 1). Another form of augmentation can come from the augmentation of sensor data for which both the position and sensor values are known.

Segmentation toolkits are a routine portion of almost every commercial product that is categorized as an image guided surgery package (Stryker, Zeiss, Neuromate, SNN, Stealth etc.). Most of the software packages are based on a thresholding method and use an interactive procedure to determine the best threshold value. Moreover, they provide the user with tools to perform improvements on the segmented slices.
There is an enormous and important research area for segmentation that can be leveraged for improving the accuracy and ease of segmentation. Introduction of knowledge-based systems that simultaneously incorporates information about anatomical boundaries (shape) and tissue signature (gray scale) using \textit{a priori} knowledge are starting to be developed. Other techniques proposed make use of a combination of gray scale and edge-detection algorithms and some \textit{a priori} knowledge to provide an unsupervised segmentation. We have chosen the 3D slicer software that is freely available, open-source [7]. It has been developed in collaboration between the MIT’s Artificial Intelligence Lab and the Surgical Planning Lab at Brigham & Women's Hospital, an affiliate of Harvard Medical School. 3D slicer can generate an output file in modeling VTK format for which we have written parsers within our software (See Figure 1). It has several important tools available for segmentation.

3.2 Robotic-based Tracking of Camera System

3.2 Implementation of Augmented Reality System

As neuronavigation techniques have now become standard, we focus our attention on the implementation of an Augmented Reality system. Augmented Reality systems represent an extension of neuronavigation systems. We have researched view augmentation of a camera that can be mounted at the end-effector of a robot. The live views from the robot end-effector

Figure 2: (a) We use the Microscirbe as a tool tracker. The position and orientation of the end-effector is shown on the orthogonal slices and 3D model of the phantom skull. (b) after adding a calibrated and registered camera, (c) we can generate a Robotic Augmentation Scene.

Figure 3: These are the steps needed to generate both a Neuro Navigation (NN) System and an Augmented Reality system. Note that AR represents an extension of Neuronavigation and can be performed simultaneously with NN.
camera will have synthetically generated 3D graphics structures overlaid on them. As a testbed for the robotic-based systems, we have mounted a miniature camera (Watec LCL-628 7.5 mm diameter 330 TV Lines 3.9mm lens f2.8 (H-52º, V-40º) on the end-effector of the Microscribe device. This device can be considered a passive robot (articulated arm). Its geometric and transformation structure similarities make it an inexpensive and useful analog to current robotic systems in terms of experimentation with AR technology. It has the advantage of being readily accessible and amenable to quick prototype development and evaluation. We have integrated the serial interface to the Microscribe and can get a precise position and orientation of the camera to the augmented reality interface and have generated a prototype using this device that can easily be translated for use with active robotic systems or other tracking system.

In figure 4, a forward kinematics solution (i.e. the end-effector of the robot coordinates in terms of the base coordinates) would be defined as the concatenation of all the individual joint transformation matrices. Each individual transform specifies how the first joint is related to the second joint. The combination of the matrices, define the position and orientation of the end-effector in the base coordinate system.

\[
T_{B-EE} = T_{B-J0} \times T_{J0-J1} \times T_{J1-J2} \times T_{J2-J3} \times T_{J3-EE}
\]  

(1)

Additional transformations are needed to compute the needed AR transformation. A relationship between the camera coordinate system and the object that has to be augmented \((T_{O-C})\) needs to be derived. This transformation is computed by knowing the transformation from the object to the base of the robot \((T_{O-B})\), the measured transformation from the base of the robot to the end-effector \((T_{B-EE})\) and also the transformation from the end-effector of the robot to the CCD of the camera \((T_{EE-C})\) as follows (see figure 4):

\[
T_{O-C} = T_{O-B} \times T_{B-EE} \times T_{EE-C}
\]

(2)

The computed relationship of equation 2 will allow the alignment of the actual camera coordinates with the coordinate system of virtual graphics camera. One of the key registrations that need to be done is object (patient) registration \((T_{O-B})\). This transformation is computed using a pair-point matching that is also done in standard neuronavigation systems (See Figure 2a). Knowing a pair of matched points between the actual patient (phantom) and the tomographic image space, an iterative algorithm to optimize the transformation parameters can result in the needed transformation. The Levenberg-Marquardt optimization method has been shown to provide a fast convergence and we have used this algorithm in our implementation of pair-point matching. The transformation \((T_{EE-C})\) utilizes methods from camera parameter estimation (based on the pattern-based estimation of the extrinsic camera parameters). Some details of this computation will be provided next.

Camera calibration is a protocol for the computation of both the camera geometric and optical characteristics (intrinsic parameters) and the 3-D position and orientation (extrinsic parameters). Using the camera parameters estimated by the calibration technique described (in the next section) the camera viewpoint is calculated relative to the real world marker (pattern) shown in Fig 3. In general, camera parameter estimation involves moving the camera viewpoint relative to a geometrically well-defined pattern. Both the intrinsic (camera model) parameters and the extrinsic (pose) parameters are estimated. Once the intrinsic parameters are known, the extrinsic parameters (pose) relative to any pattern view can be readily computed in one step. Details of the camera calibration procedure can be found in numerous references [7].

In this work, we take advantage of the estimation procedure to calculate the transformation matrix from the pattern to the camera coordinates systems \((T_{P-C})\). First, the pattern used for camera calibration, was rigidly fixed relative to the base of the Microscribe. Then the location and orientation of the pattern’s coordinate system (relative to the robot’s base) was calculated using three collected points on the pattern (digitized with the Microscribe). The first point \((P_0)\) as the origin of the coordinate system, the second point \((P_x)\) was a point along the x-axis and the third point \((P_y)\) was a point along the y-axis of the pattern coordinate system. At this point, the vectors in the x and y directions \((V_x)\) and \((V_y)\) were computed as follows:

\[
V_x = \frac{P_x - P_0}{|P_x - P_0|}
\]

(3)
\[ V_y = \frac{P_y - P_0}{|P_y - P_0|} \tag{4} \]

The vector in the z-direction was computed simply as the following cross product.

\[ V_z = V_x \times V_y \tag{5} \]

These three vectors along with the point \( P_0 \) form the 4x4 homogeneous transformation matrix from the base coordinate to the pattern coordinate systems (\( T_{BP} \)) in the following way:

\[ T_{B,P} = \begin{bmatrix} V_x & V_y & V_z & P_0 \end{bmatrix} \tag{6} \]

The above transformation matrix can be described in terms of the projection of the unit vectors of the pattern coordinate system onto the base coordinate system and a transition vector from origin of the pattern to the origin of the camera coordinate systems. Knowing the above transformation from the base of the robot to the pattern (\( T_{B,P} \)), knowing also the transformation from the base of the robot to the end-effector (\( T_{B,EE} \)) and also knowing the transformation from the pattern to the camera coordinates (\( T_{P,C} \)) via camera extrinsic parameter determination, the needed transform from the end-effector of the robot to the camera (\( T_{EE,C} \)) can be derived as follows:

\[ T_{EE,C} = T_{B,EE}^{-1} \times T_{B-P} \times T_{P-C} \tag{7} \]

3.3 The Human Factors Study

For this study, there were 21 subjects, 7 female and 14 males, ranging in age from 25 - 45. There were 3 surgeons involved in the study (subjects 14, 15, 16). The subjects were seated in front of the microscribe arm and a phantom skull. They were asked to manipulate the arm in and around the phantom skull and given detailed instructions on how to work the device. First the subjects were given time to train on both the neuronavigation system and the augmented reality system. The training sessions consisted of a series of two trials and lasted as long as the subjects required getting familiar with the system; typically about 10 minutes each (although it took longer for the neuronavigation). The actual tests were conducted in 2 trials with a 5 minute rest between each trial. The order of administration of the AR portion and the NN portion for both the training session and the actual test were counterbalanced (the orders were alternated). The total test lasted about 1 hour.
During each of the actual trials, each subject was first asked to locate certain objects inside the skull and asked to then draw an opening (craniotomy) for that object from a particular location on the skull surface. This is typically the first step during neurosurgeries. For the neuronavigation system, they had to determine the extent of the objects in each of the orthogonal slices and then draw the outline of the object from that information. In the display for the NN, they had both the 3D view with all the relevant objects segmented, along with orthogonal 3D CT slices. In the AR system, they were to position the camera away from the skull, locate the overlaid object in the video view, place their marker in the video view and draw the object on the surface of the skull. After the drawing session, the subject was asked a set of questions. He was asked to use the visualization tool at hand (either Augmented Reality or Neuronavigation) to determine if they were able to understand the relationships of each of the objects that were inside the skull. The objects inside of the skull were fabricated to be neutral objects with no anatomical basis (cubes, bolts, cylinders, and tubes). This allowed us to test non-surgeons as well as surgeons and anatomical knowledge did not give an advantage for the objects in question. Examples of these questions are as follows:

1. Where is the pyramid?

2. Where is the cylinder(facing the skull)

3. Is the bolt touching the vessel?
4. Is the bolt above or below the Cube?

Each subject was timed on how long it took for the entire test, and the questions were graded to determine how many errors were made. The movements of the robot arm were also recorded for later analysis. In addition, a questionnaire was given at the end of the test to get some subjective and comparative impressions of the two systems.

4.0 Results and Discussion

Accuracy of both the neuronavigation and AR is an important concern for medical use. AR error was measured at an average of 2.75mm with a max error of 5.2mm and a standard deviation of 0.80mm. This error is on the borderline of acceptability for neurosurgery applications (one of the most demanding in terms of accuracy requirements). Details of our error study will be up-coming in our journal paper on this topic and will be presented in our presentation. The overall accuracy of the system can be improved in several ways. A higher fidelity tracking system could be used (the microscribe was measured at 0.80mm error). The tracking device in this case contributes about one-third of the error because it is involved not only in the tracking of the camera, but also for the registration of object space relative to its base. There is substantial research in this area where surface fitting and other methods are used to get improved registration. The process of camera calibration can also be improved and is an area of intensive research. For this study, we have built a prototype that approaches the error requirements imposed by Neurosurgery. We are confident that error improvements will be made such that these systems will be able to be used with confidence in the operating room.

Currently, Neuronavigation systems primarily use three 2-D views (coronal, axial and sigittal views) to gain awareness of the patient geometry. The surgeon has to perform the 2D (image) to 3D transformation in their minds and also project the envisioned data on the view of the patient. In addition, oblique/non-orthogonal craniotomies are very difficult to do in standard neuronavigation systems (and was not even tested). It is relatively easy to do in an AR system. We believe that Augmented Reality generation is a natural extension for the surgeon because it does both the 2D to 3D transformation and projects the views directly onto the patient view. We have already illustrated the difficulty to interpret 2D slices. In figure 1, a simple 3D shape like the cube is represented by a triangle (which was confused by some of the subjects with the pyramid present in the skull). The bolt is viewed as a circle in the axial view and a disjoint line in the sagittal view. This does not seem like the natural method of visualization. This was part of the motivation for this work.

Our Human Factors study indicated that neuronavigation took a statistically significant longer time than did augmented reality. In addition, (although on the border of statistical significance (p value of 0.068)), neuronavigation did have on average a greater number of errors. (see figures 4, 5 and tables 1 and 2). All of the surgeons tested made more errors using the NN system and took longer than the AR sessions. We have also given details of the implementation of this prototype such that it can be recreated. We conjecture that medical robotic devices of the future should be able to use this technology to directly link these systems to patient data and provide the optimal visualization of that data for the surgical team. The design and methods of this prototype device can, we believe, be extrapolated for current medical robotics systems and neuronavigation systems[8].
Table 1: Errors: Average, Std and computer t-statistic p value.

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>STD</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>1.04</td>
<td>1.02</td>
<td>0.06846</td>
</tr>
<tr>
<td>NN</td>
<td>1.7</td>
<td>1.27</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Time (minutes): Average, Std and t-statistic p value

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>STD</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>5.38</td>
<td>2.74</td>
<td>0.00013</td>
</tr>
<tr>
<td>NN</td>
<td>8.47</td>
<td>1.90</td>
<td></td>
</tr>
</tbody>
</table>

5.0 References