

A Computer Based Simulator for Ultrasound Guided Needle Insertion Procedures

D. Magee¹ and D. Kessel²

¹ School of Computing, University of Leeds, Leeds, UK

² St James's University Hospital, Leeds, UK

Keywords: simulation, registration, calibration, ultrasound-guidance, needle-insertion

Abstract

A system for the simulation of needle insertion procedures under ultrasound guidance is presented. This is a core skill in interventional radiology, currently acquired on an apprenticeship basis. The system is designed to improve the early stages of training, and detect people with poor natural ability. The simulator is based on a volumetric data set which is non-rigidly aligned to a physical mannequin surface. Magnetic 3D position sensors allow the generation of 'virtual ultrasound' images based on the position of a mock ultrasound probe and needle. The calibration procedures required for this are presented. Metrics for skill measurement are presented and evaluated, and it is shown that a set of 10 novice users increase their skill levels after structured practice.

1 Introduction

Needle placement under ultrasound guidance is a core skill in interventional radiology. In such procedures a needle (or similar device) is inserted into a patient with real-time feedback from a 2D ultrasound display. Examples of such procedures include:

- Drainage of abscess
- Drainage of fluid collections in the abdomen (ascites) and chest (pleural effusion)
- Relief of blockages affecting the kidneys or the liver
- Insertion of catheter tubes into blood vessels. (NICE guidelines recommend use of ultrasound whenever possible)
- Biopsy of deep tissues such as abdominal organs
- Treatment of cancers by radiofrequency ablation

The spatial reasoning and hand eye co-ordination skills required to perform such procedures are non-trivial to acquire. Interventional radiological training is conventionally performed on the basis of observation of many procedures, followed by subsequent supervised practice (known in medical circles as the *see one, do one, teach one* paradigm [13]). With the increasing number of trainees coming through the system this approach is of increasing concern. Tafra [12] compares the training of pilots to surgeons; "It is frightening to speculate

the level of mortality that might ensue if pilots received the same level of training that surgeons are currently subjected to in learning new surgical procedures."

One solution to the training problem is training simulators. Physical ultrasound breast phantoms suitable for biopsy training exist (e.g. from SIEL Ltd.). However, these are insufficient as it is necessary to train on a range of abnormalities and anatomy, even for simple procedures [12]. Such phantoms also require the use of an expensive ultrasound machine. An alternative is computer based simulations. These fall into two general categories; i) Virtual reality simulators, in which the complete environment is simulated within a computer, and ii) Augmented reality simulators, in which part of the simulation is physical, and part computer generated. The system presented in this paper falls into the latter class. Keeping as much as possible in the physical domain improves the correspondence between the simulation and the actual procedure. Past experience also suggests that there has been considerable uptake of mannequin based simulators for non-invasive procedures in a clinical training setting (for example commercial products from METI and Laerdal), while there is relatively little use of computer/VR only solutions.

Our initial prototype system is based around a physical (mannequin based) simulation, augmented with computer generated 'virtual' ultrasound imagery. Images are derived from 3D scans of human subjects which may be interchanged to give a wide range of abnormalities and anatomies for training. An initial evaluation has been performed using a set of unskilled individuals and a set of trained clinicians. This has allowed the development of metrics for the measurement of skill/skill acquisition for these procedures. Potentially such a system could be used to identify individuals with little aptitude for these tasks at an early stage, and redirect them to other areas of clinical practice (clinical feedback suggests there are a significant minority of such individuals). These metrics could also be used to measure skill development prior to practice on live patients.

2 Background

Computer based clinical simulators have been under development for over ten years. However, the realism and purpose of these simulators varies from one instance to another. Many simulators are entirely based around a conventional computer setup (CPU, keyboard, screen, mouse). These do not aim to recreate the entire procedure, rather certain visual aspects of the procedure, such as the radioactive

seed implantation modelled by [1]. Such simulators have quite a wide scope for use in planning of needle insertion [14] and other procedures, but are of limited use for training beyond the early stages. More complex setups, based around virtual environments, haptic feedback devices and/or mannequins, can be used to provide more realistic experience for training. However, the uptake of such systems in interventional surgical/radiological training is not generally high. The principal success of simulation for invasive procedures has been in endoscopy related procedures (e.g. [8, 6], and various commercial products). This is primarily due to the original procedure using video display technology as the primary visual feedback, and the physical interface being reasonably simple. To quote [6]; “The benefits of endoscopic training through computer simulation, rather than the traditionally performed one-to-one apprenticeship schemes, are now well accepted in the community” (although supervised practice is still an essential part of training). Similarly simulators for various teleoperated procedures, such as eye procedures [11], are available. Simulation of more complex procedures (in terms of the perceptual environment) is still an active research topic. Holbrey et al. [10] present a system based on a haptic feedback device, stereo display, and Finite Element Models for vascular suturing (stitching) training. A variety of simulators exist for a range of minimally invasive procedures. Gorman *et al.* [9] present a system incorporating a mannequin and a haptic feedback device for the simulation of lumbar (back/spine) puncture procedure. This procedure relies on tactile feedback and offline CT data, and so has no real-time visual feedback (as part of the realistic part of the simulation). Simulators for fluoroscopy (real-time, or near real-time, X-ray based 2D imaging) guided procedures [13, 4] (see also PERC Mentor: www.simbionix.com) are closer in nature to the ultrasound guided simulation we present. However, these differ fundamentally in the way the imaging is steered (or not) physically. Direction of imaging is far less important with fluoroscopy than with ultrasound due to the absorbency/reflection properties of ultrasound. However, ultrasound is inherently safe, whereas exposure to X-ray should be minimised. The commercial product UltraSim (www.medsim.com) simulates ultrasound imaging without any interventional procedure. This simplifies the implementation somewhat. UltraSim is primarily used for diagnosis training (rather than interventional training). DiMaio and Salcudean [5] present a fast Finite Element model for the haptic modelling of needle insertion that could be added to our system (see discussion). We do not know of any simulator for ultrasound guided needle insertion other than the 2D computer only simulation mentioned previously [1], which does not model probe motion or any 3D aspect of the task.

3 Simulator Implementation

Our simulator (figure 1) is based around a relatively simple, low cost, hardware setup consisting of a standard PC (Dual P3 3GHz, 2GB ram, running Microsoft Windows XP™), a

full size plastic torso mannequin, and a pair of Ascension PCBirds™ magnetic 3D position/orientation sensors. One of the sensors is rigidly attached to a mock ultrasound probe, and the other attached to a wooden needle via a sliding mechanism. Software is implemented in C++ using *OpenGL*. Figure 2 shows the system in use.

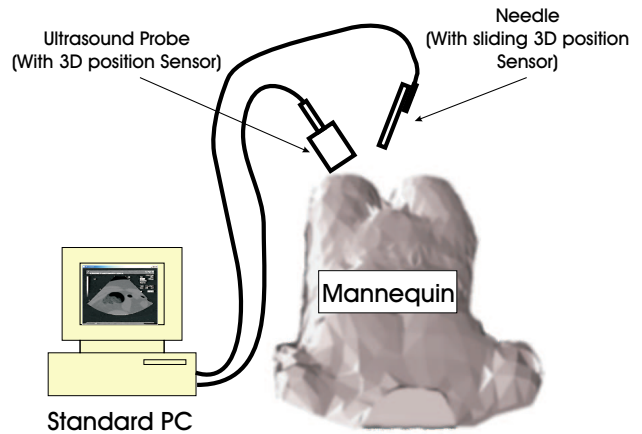


Figure 1: Overview of hardware setup



Figure 2: The simulator in use

The software produces a ‘virtual’ ultrasound image on a dedicated screen reconstructed from volumetric data obtained from a human subject (registered to the surface of the mannequin, section 3.1). The image is produced by raster scanning through each pixel in the ultrasound portion of the image (figure 3), and projecting each point to the corresponding point in 3D space based on the position sensor output. The 2D image formed from the corresponding 3D position greylevels, and a background image, is rendered using *OpenGL*.

The projection of the x, y image pixel location into 3D space is given by equation 1.

$$X_{3d} = S_{pos} + K_{p2d}(x - O_x)S_{dir} + K_{p2d}(y - O_y)S_{or} \quad (1)$$

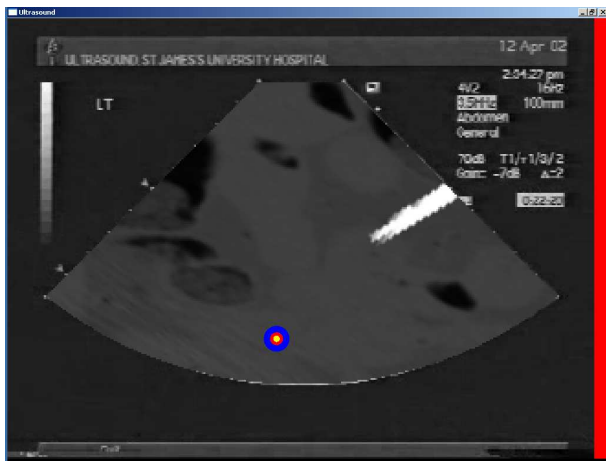


Figure 3: The main ultrasound window

Where S_{pos} is the sensor position in 3D space, S_{dir} is the sensor direction (unit) vector, S_{or} is the sensor orientation (unit) vector, K_{p2d} is a constant relating pixels to inches, and O_x, O_y is the image point relating to the sensor position. S_{pos} , S_{dir} and S_{or} are the output from a calibration function applied to the sensor outputs, rather than the raw sensor outputs (see section 3.2). The greylevel value at the image pixel corresponding to x, y is calculated by projecting X_{3d} into the volume space (section 3.1), and performing linear interpolation from the 8 neighbouring voxel greylevels.

The needle is rendered directly into (a copy of) the volumetric data. Voxels relating to the needle are identified by first defining a line in 3D space based on the output of the second position sensor, as defined as in equation 2.

$$X = S_{pos2} + dS_{dir2} \quad (2)$$

Where $0 < d < \text{needle length}$. As with equation 1, S_{pos2} and S_{dir2} are the output from a calibration function applied to the sensor outputs, rather than the raw sensor outputs (see section 3.2). The line is projected into the volume space (section 3.1) and the greylevel of voxels within R (the needle radius) of the line (measured in a direction normal to the line) are set to maximum (i.e. white). Rendering into the volume (rather than the 2D image) has two advantages; i) Rendering need not be at the same update rate as the image update, and ii) The calculation of the 2D projection is trivial, as it is done as part of the projection of the volume (i.e. no additional effort). The true benefit of ii) would be observed if a more complex shape, such as a biopsy needle (rather than simply a cylinder) was rendered, as we plan in the future.

3.1 Registration of Volumetric Data to Mannequin Surface

As stated previously, the simulator renders ‘virtual ultrasound’ images based on volumetric data acquired from real humans. We currently use CT data from the Visible Human dataset [2],

although any volumetric data (CT, MRI or 3D Ultrasound) could be used. This data is registered to the surface of the mannequin in a 3 stage process. Firstly, surfaces are extracted from the volumetric data and mannequin as point clouds. The surface on the volumetric data is extracted by 2D hand annotation of slices (figure 4), and the surface of the mannequin is extracted by passing a (calibrated) motion sensor over the surface of the mannequin (figure 5).

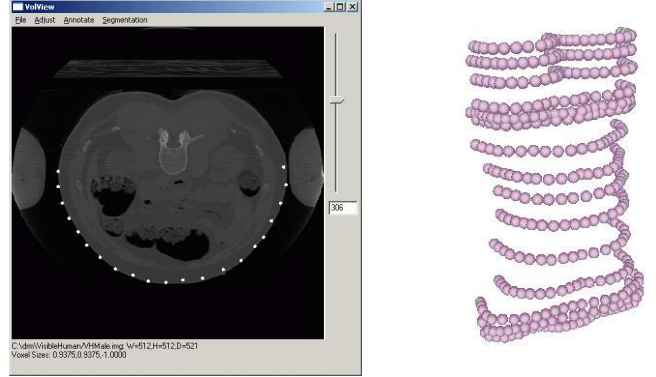


Figure 4: Hand annotation of surface in CT volume

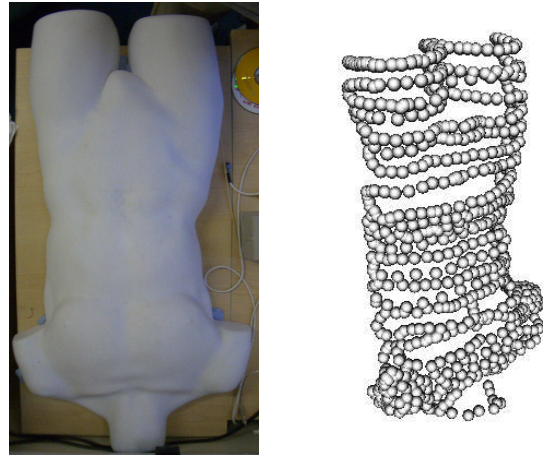


Figure 5: A Point cloud surface of the mannequin in formed by motion of the 3D sensor over the surface

The second stage of registration is a rigid alignment of the two data sets. To do this we use the Iterative Closest Point Algorithm (ICP) [3]. This is an iterative local optimisation algorithm, the success of which depends on the initial starting alignment. ICP locally minimises mean-square distance between two point sets. To ensure success (in practice), it is usual to use multiple starting configurations [3], to obtain a globally optimal registration. We take this approach. Firstly we align the mean of the volumetric data set to the mean of the mannequin data set and scale the volumetric data set as in equation 3.

$$X_{vnew_i} = K \frac{\frac{1}{N_m} \sum_{j=1}^{N_m} |X_{m_j} - \bar{X}_m|}{\frac{1}{N_v} \sum_{j=1}^{N_v} |X_{v_j} - \bar{X}_v|} (X_{v_i} - \bar{X}_v) + \bar{X}_m \quad (3)$$

Where $X_{v_{new_i}}$ is the transformation of the volume point X_{v_i} , N_m is the number of points making up the mannequin surface, N_v is the number of points making up the volume surface, X_{m_j} is a point on the mannequin surface, and \bar{X}_m and \bar{X}_v are the means of the mannequin and volume surfaces respectively. K is a constant (around 2.5) that ensures the transformed volume surface is larger than the mannequin surface. This (in general) prevents the ICP algorithm shrinking the volumetric surface to fit just a subset of the mannequin surface. The ICP algorithm is applied multiple times with the transformed volumetric surface data, further transformed by a random rotation, displacement and scaling, as a starting point. Of the multiple registration solutions, the solution with the lowest scale normalised mean-squared error is selected. Scale normalised mean-squared error (E_{snmse}) is calculated as in equation 4.

$$E_{snmse} = \frac{1}{S \times N_v} \sum_{i=1}^{N_v} |X_{r_i} - C_m(X_{r_i})|^2 \quad (4)$$

Where S is the scale calculated by the ICP algorithm, X_{r_i} is a point on the volume surface transformed by the transformation calculated by the ICP algorithm, and $C_m(X_{r_i})$ is the closest point on the mannequin surface to X_{r_i} . The use of this (scale normalised) metric rejects any solutions where the ICP algorithm has significantly shrunk the volumetric surface (we also explicitly reject solutions with a scale below a threshold). A typical registration result of this algorithm is given in figure 6. In our experience 50 random starting points is sufficient to robustly give results of this quality.

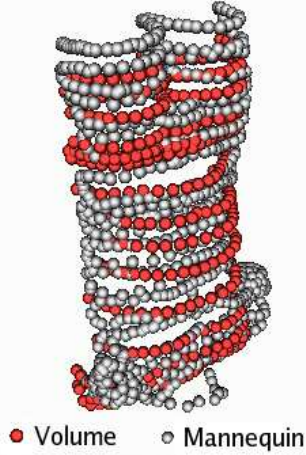


Figure 6: Result of rigid registration using multiple start ICP

The ICP based rigid registration algorithm described aligns the volumetric and mannequin surfaces reasonably well. However, for the surfaces to be completely aligned a non-rigid registration process is also necessary due to differences in the shape of the surfaces (particularly around the belly). We formulate this non-rigid registration as the estimation of a quadratic mapping function that maps the volumetric surface on to the mannequin surface as in equation 5.

$$\begin{pmatrix} x_{new} \\ y_{new} \\ z_{new} \end{pmatrix} = K_c + K_x \begin{pmatrix} x \\ y \\ z \end{pmatrix} + K_{xx} \begin{pmatrix} x^2 \\ y^2 \\ z^2 \end{pmatrix} + K_{xy} \begin{pmatrix} xy \\ xz \\ yz \end{pmatrix} \quad (5)$$

Where K_c is a 3×1 matrix, and K_x , K_{xx} and K_{xy} are 3×3 matrices. Equation 5 defines a 3D warp that may be applied to volumetric data. If a set of point correspondences ($\gg 10$ correspondences) are known, the values in K_c , K_x , K_{xx} and K_{xy} may be estimated using a least-squares approach (applied separately for each dimension (x,y,z)). However, we desire a smooth mapping from a (relatively) small number of correspondences. To achieve this we add some smoothness constraints to the least squares set of equations. For the x dimension mapping:

$$\frac{dx_{new}}{dx} = 1, \quad \frac{dx_{new}}{dy} = 0, \quad \frac{dx_{new}}{dz} = 0 \quad (6)$$

and for the y dimension mapping:

$$\frac{dy_{new}}{dx} = 0, \quad \frac{dy_{new}}{dy} = 1, \quad \frac{dy_{new}}{dz} = 0 \quad (7)$$

and for the z dimension mapping:

$$\frac{dz_{new}}{dx} = 0, \quad \frac{dz_{new}}{dy} = 0, \quad \frac{dz_{new}}{dz} = 1 \quad (8)$$

These constraints are selected to force the transform to be as close as possible to an unrotated rigid mapping. The constraints are implemented as soft constraints by including 3 equations relating to each volumetric data point in the set of equations presented to the least-squares solver (i.e. 4 simultaneous equations per point are used, one for the mapping and 3 for the constraints).

To obtain the pairs of correspondences required for the non-rigid estimation process described we use an approach based on the RanSaC algorithm [7]. This commonly used algorithm in model fitting and 3D geometry is based on a three part sample, estimate and verify approach. A small set of correspondences is sampled from the set of possible correspondences based on some probabilistic matching criteria. This set is then used to estimate the model parameters (the quadratic function in our case). The model is then verified by application to the entire data set using a ‘‘Consensus function’’. The process is repeated many times, and the model with the highest ‘‘consensus’’ selected. The model may then be re-estimated based on a larger set of correspondences calculated using the selected model. Our RanSaC correspondence sampling probability density is based on the euclidean distance between potential correspondences normalised by the minimum distance between the volume point and any point on the mannequin surface, as in equation 9.

$$D_{cmatch}(v, m) = \frac{E_d(v, m)}{E_d(v, C_m(v))} \quad (9)$$

Where $E_d(x, y)$ is the euclidean distance between points x and y . A $N_v \times N_m$ surface point similarity matrix is formed based on equation 9, which is binarised using a threshold (T_d , typically 1.3), as in equation 10.

$$P_{cmatch}(v, m) = \begin{cases} 1 & \text{if } D_{cmatch}(v, m) < T_d \\ 0 & \text{if } D_{cmatch}(v, m) \geq T_d \end{cases} \quad (10)$$

This excludes a large proportion of possible poorly matching correspondences. Around 20-30 correspondences per trial are selected randomly from the non-excluded (true) matches, in order to calculate the mapping (equation 5). Typically 100 trials are performed in order to select the best mapping.

The consensus function (C) used to select the final mapping from the set of candidate mappings is based on the proportion of transformed volume surface points that are within a fixed threshold of a mannequin point (P_t).

$$C = \frac{\# \text{ points where } |X_{nr_i} - C_m(X_{nr_i})| < T_c}{N_v} \quad (11)$$

Where X_{nr_i} is the i 'th non-rigidly registered volume point, and T_c is a threshold (typically 0.05 times the mean distance of the mannequin points from the mannequin centroid). The solution with the highest value for C is selected, and the mapping re-calculated based on a larger set of correspondences made up of the closest mannequin surface matches to the transformed volume surface points where the distance between correspondences is below the threshold T_c . In practice this (generally) has the effect of increasing the consensus value (C) for the final solution (if not the original mapping is used). Typical results of this procedure are given in figure 7.

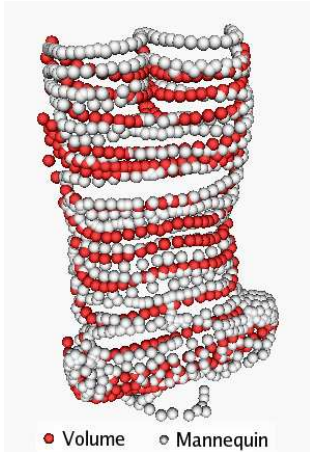


Figure 7: Results of non-rigid registration

The quadratic function is used to warp the volumetric data offline to a new volume (with arbitrarily sized cubic voxels). The warp is performed by calculating the corresponding position in the original volume to each voxel in the new volume using i) a fixed scaling and offset derived from the choice of voxel size/range in the new volume, ii) The inverse of the quadratic function, and iii) the inverse rigid mapping. To calculate ii) we initially calculate a quadratic approximate inverse by swapping the (rigidly-aligned) volume

and mannequin surface point sets in the last stage of the non-rigid estimation process. The output of this function is then used as an initial value for the calculation of an accurate inverse of the forward function using Newtons method (typically only 2 or 3 iterations are required). Once the warped volume has been formed on-line calibration is simply matter of a rigid scale and offset between the volume co-ordinate system and the world co-ordinate system.

3.2 Calibration of Probe and Needle

The motion sensors used output 3D position, orientation and rotation information. Due to the practical constraints on sensor placement, it is not always possible to place the sensor at the position that we want to measure (e.g. the centre of the end of the probe), or orient it in the direction we want to measure. However, it is possible to rigidly attach the sensors and estimate a calibration function to map the sensor values to the desired values. The raw output of the sensors may be visualised as a point and two vectors as shown in figure 8.

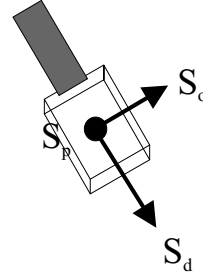


Figure 8: Raw sensor output as a point and two vectors

Probe position calibration is performed as the estimation of an offset from S_p in the frame of reference defined by S_d and S_o . This is in fact a 3 degree of freedom estimation problem, thus we introduce the variable θ , an angle used to rotate the S_o vector about S_d (thus allowing the calibration point to be off the plane defined by S_o and S_d). Equation 12 relates the sensor outputs to the calibration point (C_p).

$$C_p = S_p + d_1 S_d + d_2 R_\theta(S_o) \quad (12)$$

Where $R_\theta()$ is a rotation of θ about S_{dir} , as in equation 13 (effectively defining a third axis using the cross product).

$$R_\theta \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \cos \theta + (S_{d_y} z - S_{d_z} y) \sin \theta \\ y \cos \theta + (S_{d_z} x - S_{d_x} z) \sin \theta \\ z \cos \theta + (S_{d_x} y - S_{d_y} x) \sin \theta \end{pmatrix} \quad (13)$$

Substituting 13 into 12 and rearranging gives:

$$\begin{pmatrix} 1, 0, 0, -S_{p_x}, -S_{o_x}, -(S_{d_y}S_{o_z} - S_{d_z}S_{o_y}) \\ 0, 1, 0, -S_{p_y}, -S_{o_y}, -(S_{d_z}S_{o_x} - S_{d_x}S_{o_z}) \\ 0, 0, 1, -S_{p_z}, -S_{o_z}, -(S_{d_x}S_{o_y} - S_{d_y}S_{o_x}) \end{pmatrix} \begin{pmatrix} C_{p_x} \\ C_{p_y} \\ C_{p_z} \\ d_1 \\ d_2 \cos \theta \\ d_2 \sin \theta \end{pmatrix} = \begin{pmatrix} S_{p_x} \\ S_{p_y} \\ S_{p_z} \end{pmatrix} \quad (14)$$

Equation 14 is in a form where the unknowns (C_p , d_1 , $d_2 \cos \theta$ and $d_2 \sin \theta$) may be estimated from multiple unique examples using a standard least-squares solver (e.g. the *dgels* routine in the LAPACK library). To obtain a set of values for the least squares solver a set of (around 30) samples are taken with the calibration point resting on a fixed point, as illustrated in figure 9.

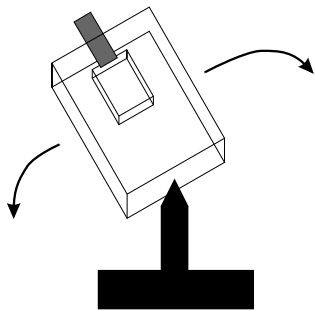


Figure 9: Sensor calibration data obtained by fixing the calibration point and moving the sensor

The two probe directions C_d and C_o (analogous to S_d and S_o) are estimated by moving the probe in these directions, taking samples at approximately uniform spatial intervals. Each pair of samples gives an estimate of C_d (or C_o) in world co-ordinates (V_n , a unit direction vector). This may be projected into the co-ordinate system defined by S_d, S_o and S_3 (the cross product of S_d and S_o) for the first of these samples. We perform this projection using the least-squares solution to the system of equations given by:

$$V_n = (K_d, K_p, K_3)(S_{d_n}, S_{o_n}, S_{3_n}) \quad (15)$$

Where K_d , K_p and K_3 are scalar unknowns that define C_d or C_o w.r.t. the co-ordinate system defined by vectors S_{d_n} , S_{o_n} and S_{3_n} .

Needle calibration is initially based on the same mathematical approach as for the probe position calibration. Figure 10 illustrates the data acquisition in this case.

Needle calibration is performed multiple times (typically 5) with different values of D_n . These calibrations are projected

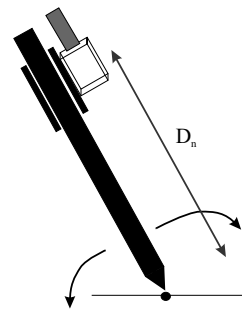


Figure 10: Needle calibration data obtained by fixing the calibration point (and D_n) and moving the sensor

into the space defined by S_d , S_o and S_3 , and Principal Components Analysis performed to extract the eigenvector with the highest variance. This corresponds to the needle direction (analogous to C_d in the probe case). Each calibration is projected onto this principal eigenvector (using the dot product of the calibration and the vector), and the resulting vector (projection \times eigenvector) subtracted from each calibration. The mean of these resultant calibrations is taken as the final needle calibration. N.B. The rotation of the needle is not relevant for this application.

4 Evaluation of Skill Acquisition

One of the principal aims of the system described in this paper is the measurement of operator skill. In order to measure skill qualitatively and automatically we have developed a set of metrics for the evaluation of sessions representing a single attempt at finding and hitting a target with a needle. The aim of these metrics is to measure absolute skill, for skill level determination, and relative skill, for learning evaluation. If it is possible to differentiate between skilled interventional radiologists and novice users, this is also a partial validation of the realism of the simulator.

The simulator allows the recording of sessions for later analysis. The first stage of this analysis is the identification of the time period that represents the needle insertion event. This is performed using a simple line-surface intersection algorithm, where the skin surface is represented by a triangular mesh (figure 11).

We define a set of 34 metrics based on a number of principals suggested by skilled clinicians (table 1). 5 sessions from 3 consultant radiologists were taken to give an estimate of the distribution of these metrics for such skilled operators. A set of 10 novices, with no experience of real or simulated needle insertion procedures, was recruited. After brief period of familiarisation with the ideas behind the procedures (but no practice) 4 sessions from each novice (with a different target position for each session) were recorded. The metrics were calculated for each novice and expert session and evaluated

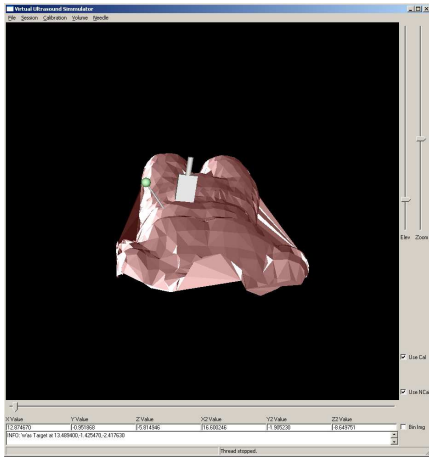


Figure 11: Visualisation of a recorded session with skin surface represented as a triangular mesh

based on their ability to differentiate between the novice and expert sessions (using the t-test). Table 2 lists the 10 metrics where there was a statistically significant difference ($P < 0.05$) between the expert and novice sessions.

Principal	Metrics
The needle should travel in a straight line between its entry point and the target	<ul style="list-style-type: none"> • Mean/max/std. dev. of needle tip distance from entry/mean/optimal line • Mean/max/std. dev. of needle angle from entry/mean/optimal line
The needle should be visible in the ultrasound image	<ul style="list-style-type: none"> • % time needle tip visible • Mean/max/std. dev. of needle tip distance from ultrasound probe plane at needle entry / mean ultrasound plane • Mean/max/std. dev. of needle angle from ultrasound probe plane at needle entry / mean ultrasound probe plane
The ultrasound probe should not move significantly after needle insertion	<ul style="list-style-type: none"> • Mean/max/std. dev. of angle between ultrasound plane and ultrasound plane at needle entry / mean ultrasound plane
The target should be visible in the image while the needle is inserted	<ul style="list-style-type: none"> • % time target in view • Mean/max/std. dev. of distance between target and ultrasound probe plane

Table 1: Principals and Metrics for the comparison of performance

The set of novices were allowed a 20 minute practice period with multiple new targets (as requested), before being re-tested on 4 new targets. The complete session took around 40 minutes per novice (including the two testing periods). An identical session was conducted after a gap of a week, to simulate a realistic training schedule. For brevity we present only an analysis of the improvement in performance from the beginning of the first session to the end of the second session w.r.t. the metrics selected (table 3).

8 of the 10 novices showed a statistically significant improvement ($P < 0.05$) w.r.t. at least two of the metrics. Table 3 shows between 1 and 5 novices showed a statistically

Metric	P	Novice mean [var]	Expert mean [var]
Closest dist. of n.tip to target	$1.86e-11$	0.792 [0.247]	0.247 [0.00099]
Mean n.tip dist. from optimal line	$2.89e-7$	0.783 [0.219]	0.316 [0.0042]
Mean n.angle from US plane	$1.11e-6$	0.271 [0.054]	0.061 [0.00062]
Mean n.tip dist from US plane	$5.21e-6$	0.527 [0.123]	0.168 [0.0067]
S.D. n.tip dist from US plane	$2.21e-5$	0.283 [0.025]	0.126 [0.0018]
Max. n.angle from US plane	$6.14e-4$	0.451 [0.108]	0.238 [0.0046]
S.D. n.tip dist from optimal line	0.0021	0.316 [0.041]	0.209 [0.0010]
Mean n.angle from optimal line	0.011	0.383 [0.104]	0.206 [0.011]
Max. n.angle from optimal line	0.011	0.509 [0.117]	0.337 [0.010]
S.D. n.angle from US plane	0.019	0.073 [0.0054]	0.044 [0.00020]

Table 2: Metrics for the comparison of performance (ordered by t-test statistical significance P). Distances are in inches, angles are in radians.

Metric	# Subjects improved	# Subjects deteriorated
Closest dist. of n.tip to target	4	0
Mean n.tip dist. from optimal line	3	0
Mean n.angle from US plane	3	0
Mean n.tip dist from US plane	5	0
S.D. n.tip dist from US plane	3	0
Max. n.angle from US plane	2	0
S.D. n.tip dist from optimal line	3	0
Mean n.angle from optimal line	2	0
Max. n.angle from optimal line	1	0
S.D. n.angle from US plane	3	2

Table 3: Statistically significant (t-test, $p < 0.05$) improvement/deterioration of performance for 10 subjects w.r.t. metrics

significant improvement for each metric. For only one metric (standard deviation of needle angle from the ultrasound probe plane) was there a statistically significant deterioration in performance (for two novices only). This is the metric we have least confidence in (table 2). As other related metrics show an improvement in performance, this leads us to believe this metric may not be 100% reliable. However, the results of the t-tests do not give the complete picture. The variance in the initial data sets is high, as there is a significant random element to the novice users approach initially. This, combined with the relatively small sample size (4 per novice), means a relatively large improvement is required to be deemed 'significant' (at 95% confidence). Table 4 gives statistics for improvement in mean metric values, and reduction in metric variances over the same experiment. The distribution of the t-test P value is also given for subjects where the mean metric value improved over the experiment.

Every novice showed an improved metric mean for at least 6 metrics, with 7 novices showing an improvement in 9 or 10 metrics. Similar results we obtained for metric variance reduction. These results show more subtle improvements may be quantified. It should be noted that, although a measurable increase in performance was made by the novices, the expert group were still significantly better than the novices. This is to be expected, as medical training comprises of rather more than

Metric	# Subjects Improved Mean	# Subjects Reduced Variance	P [sd] where mean improved
Closest dist. of n.tip to target	10	8	0.125 [0.123]
Mean n.tip dist. from optimal line	10	9	0.151 [0.139]
Mean n.angle from US plane	9	9	0.116 [0.084]
Mean n.tip dist from US plane	10	7	0.130 [0.144]
S.D. n.tip dist from US plane	8	8	0.086 [0.080]
Max. n.angle from US plane	8	9	0.148 [0.126]
S.D. n.tip dist from optimal line	9	9	0.160 [0.150]
Mean n.angle from optimal line	9	7	0.141 [0.103]
Max. n.angle from optimal line	6	7	0.162 [0.150]
S.D. n.angle from US plane	5	7	0.207 [0.150]

Table 4: Means/variance improvement for 10 subjects w.r.t. metrics

two 40 minute sessions.

5 Discussion and Future Work

A system for the simulation of needle insertion procedures in interventional radiology has been presented, and its use in training discussed. It has been shown that it is possible to measure skill levels. However, further evaluation based around real procedures is still required to determine if these skills transfer to real world procedures. A number of realism improvements are planned including; deformation modelling, haptic feedback and improved image realism. We plan to evaluate the necessity of each of these w.r.t. skill acquisition.

6 Acknowledgements

The authors would like to thank Andy Bulpitt, Roy Ruddle and Ken Brodlied for useful discussions and help.

References

- [1] R. Alterovitz, J. Pouliot, R. Taschereau, I. Hsu, and K. Goldberg. Simulating needle insertion and radioactive seed implantation for prostate brachytherapy. In *Proc. Medicine Meets Virtual Reality 12*, pages 19–25, 2003.
- [2] R. Banvard. The visible human project. In *Proc. CODATA 2002: Frontiers of Scientific and Technical Data*, 2002.
- [3] P. Besl and N. McKay. A method for registration of 3d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
- [4] Z. Chui, J. Anderson, X. Ma, W. Peng, Y. Cai, and Y. Nowinski. Computer environments for interventional neuroradiology procedures. *Simulation and Gaming*, 32:3:404–419, 2001.
- [5] S. DiMaio and S. Salcudean. Needle insertion modelling for the interactive simulation of percutaneous procedures. In *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 253–260, 2002.
- [6] M. ElHelw, A. Chung, A. Darzi, and G. Yang. Image-based modelling of soft tissue deformation. In *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 83–90, 2003.
- [7] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6):381–395, 1981.
- [8] G. Gopalakrishnan and V. Devarajan. Strapsim: A virtual reality-based stapling simulator for laparoscopic herniorrhaphy. In *Proc. Medicine Meets Virtual Reality 12*, pages 111–113, 2004.
- [9] P. Gorman, T. Krummel, R. Webster, M. Smith, and D. Hutchens. A prototype haptic lumbar puncture simulator. In *Proc. Medicine Meets Virtual Reality*, pages 106–109, 2000.
- [10] R. Holbrey, A. Bulpitt, K. Brodlied, M. Walkley, and J. Scott. A model for virtual suturing in vascular surgery. In *Proc. EurographicsUK: Theory and Practice of Computer Graphics*, pages 50–58, 2004.
- [11] M. Sagar, D. Bullivant, G. Mallinson, P. Hunter, and I. Hunter. A virtual environment and model of the eye for surgical simulation. In *Proc. 21st Annual Conf. on Computer Graphics and Interactive Techniques*, pages 205–212, 1994.
- [12] L. Tafta. The learning curve and sentinel node biopsy. *The American Journal of Surgery*, 182:347–350, 2001.
- [13] Y. Wang, C. Chui, H. Lim, Y. Cai, and K. Mak. Real-time interactive simulator for percutaneous coronary revascularization procedures. *Computer Aided Surgery*, 3:211–227, 1998.
- [14] J. Zeng, C. Kaplan, J. Xuan, I. Sesterhenn, J. Lynch, M. Freedman, and S. Mun. Optimizing prostate needle biopsy through 3d simulation. In *Proc. SPIE Conference on Medical Imaging*, pages 488–497, 1998.