Navigation with Local Sensors in Handheld 3D Ultrasound-Initial in-vivo Experience

Article *in* Proceedings of SPIE - The International Society for Optical Engineering · March 2011

DOI: 10.1117/12.878901

4 authors, including:

Philipp J Stolka
ADVITOS GmbH
35 PUBLICATIONS **264** CITATIONS
SEE PROFILE

Gregory D. Hager
Johns Hopkins University
520 PUBLICATIONS **24,868** CITATIONS
SEE PROFILE

Emad M. Boctor
Johns Hopkins University
248 PUBLICATIONS **2,925** CITATIONS
SEE PROFILE

Navigation with Local Sensors in Handheld 3D Ultrasound – Initial in-vivo Experience

Philipp J. Stolka[a], Xiang Linda Wang[a], Gregory D. Hager[a], and Emad M. Boctor[a,b]

[a]The Johns Hopkins University, Baltimore, MD, USA;
[b]Johns Hopkins Medical Institutions, Baltimore, MD, USA

ABSTRACT

Handheld ultrasound is useful for intra-operative imaging, but requires additional tracking hardware to be useful in navigated intervention settings, such as biopsies, ablation therapy, injections etc. Unlike common probe-and-needle tracking approaches involving global or local tracking, we propose to use a bracket with a combination of very low-cost local sensors - cameras with projectors, optical mice and accelerometers - to reconstruct patient surfaces, needle poses, and the probe trajectory with multiple degrees of freedom, but no global tracking overhead. We report our experiences from a first series of benchtop and in-vivo human volunteer experiments.

Keywords: Navigation, local sensors, handheld, 3D ultrasound, vision, needle interventions, guidance

1. INTRODUCTION

1.1 Motivation

Image guidance is a key enabling technology for millions of interventional procedures in medicine per year. Biopsies, ablative procedures, energy-based therapies, as well as minimally invasive and open procedures all rely on real-time feedback from imaging in order to be performed safely and effectively. In current practice, this feedback is predominantly provided by ultrasound (US) as it offers safe, low-cost, and real-time imaging. Despite its advantages, ultrasound often has poor image quality and is difficult to manage, particularly for interventions requiring high precision. In many cases, the target of interest, e.g. a tumor visible in MRI, is at best poorly visible in US.

In particular, needle-based interventions (biopsies, ablation therapy, epidural injections etc.) are regularly guided by ultrasound. To alleviate the above-mentioned issues, tracking systems (e.g. optical or electro-magnetic) navigate the needle towards the target, supported by the ultrasound view (Fig. 1). Such systems are bulky in the operating room, cumbersome to set up and use, and expensive. Precise localization of the needle w.r.t. the targets usually requires extensive and lengthy calibrations, which needs to be repeated for each needle, inhibiting e.g. quick needle replacement for changing requirements. Finally, the display of ultrasound imaging and tool tracking on external screens requires strong hand-eye coordination. (Alternatives include e.g. video or augmented reality overlays.[7–9])

1.2 Approach

The Intuitive Fusion system implements two approaches to local-sensor-based tracking: first, it replaces external probe tracking with local sensors, and second, it visually tracks patient, needle, and probe relative to each other.

We substitute absolute tracking with a solution that is centered on using local sensors for generation of limited-DoF (relative, as opposed to absolute) localization data. This data from a set of several such sensors can then be combined to provide complete position information through iterative global trajectory reconstruction. This results in a very small footprint in computational complexity and additional hardware. Unlike previous approaches (e.g. using optical trackers and gyroscopes,[3] or ultrasound RF data itself[5]), our approach limits strong long-term orientational drift with a setup that is algorithmically, numerically, and physically simpler and does not suffer from these drawbacks.

Further author information: Corresponding author Ph. J. Stolka (e-mail pjs@jhu.edu, telephone +1-410-516-5396). E. M. Boctor: e-mail eboctor@jhmi.edu.

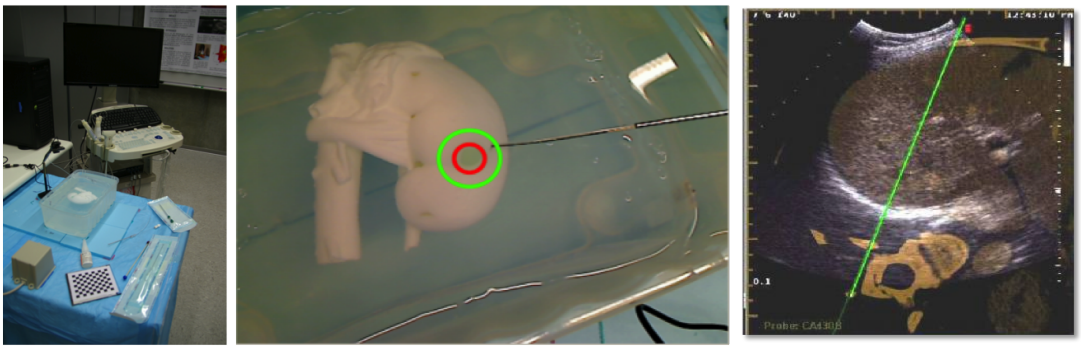

Figure 1. Ultrasound-navigated needle interventions: Experimental system setup for navigated partial nephrectomies – ultrasound cart, external EM tracking, variety of tools (left[9]); video overlay of guidance target symbols onto endoscopic view (center[9]); needle trajectory overlay onto registered ultrasound/CT view on screen (right, from `www.medcom-online.de`/MedCom NaviSuite®).

A second thrust of the presented work addresses co-registration of patient and tools. Tracking is necessary to achieve satisfactory targeting. Again, unlike common approaches involving external tracking (suffering from hardware and setup complexity), we propose to track probe, patient, and tools relative to each other, using low-cost local sensing methods. This promises to substantially reduce the footprint, simplify the workflow, and reduce the cost of ultrasound-guided interventions; furthermore, it allows the real-time detection of needle bending and tissue deformation.

We report our experiences with the Intuitive Fusion approach from two series of experiments, including initial experiences from a first series of human volunteer experiments, concentrating on the usability of the opto-inertial tracking approach with a first prototype, as well as visual needle detectability in realistic environments.

2. METHODS

2.1 System

The Intuitive Fusion technology as an embedded system consists of hardware and software features. The described clip-on brackets (Fig. 2) carry an array of vision and tracking components, and can be adapted to a wide range of handheld ultrasound probes. In particular, cameras and a projector can be attached to the passive holder for vision-based surface and needle tracking, and accelerometers and optical mouse tracking units for opto-inertial probe tracking. Algorithms are described in detail in Section 2.3.

The computer vision modules include functionality for surface reconstruction using a structured light system (SLS)[4] using two standard web cams (Logitech QuickCam Deluxe for Notebooks). Similar to existing approaches for camera-based needle tracking,[2] our approach is more general as it integrates the cameras with a mini projector for display of predefined patterns, helping with surface detection. Needle detection is also performed using these stereo cameras.

Tracking of the probe is done using opto-inertial sensors, in the presented system with an accelerometer (Nintendo Wii Remote) and optical mouse tracking units (Microsoft Optical Mouse) attached to the probe for incremental 4/5-DoF probe trajectory reconstruction, which then in turn is combined[11] with a 2D ultrasound image stream to perform 3D ultrasound reconstruction.[10]

2.2 Workflow Comparison

Conventionally, the operator has to go through the following steps to perform a navigated ultrasound intervention:

1. Introduce a tracking system into the intervention field. Depending on the type of tracker (optical, electromagnetic, mechanical), line-of-sight or avoidance of field disturbance has to be maintained.

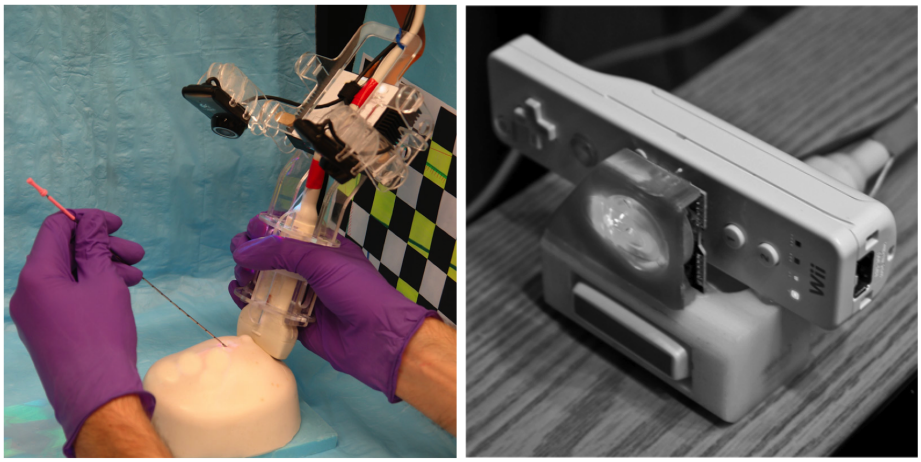

Figure 2. Intuitive Fusion needle/surface-tracking device (left); Intuitive Fusion opto-inertial probe-tracking device (right)

2. Calibrate the ultrasound probe. This establishes the transformation between the ultrasound imaging array and the attached tracker sensor.
3. Calibrate the tools. Again, the tracker sensors attached to needles, pointers etc. have to be fixed, and procedures to establish the relative transformations have to be followed.
4. (Define target, or register with pre-operative image data and plans. This requires some operator interaction.)
5. Perform intervention. The operator is watching a screen, showing relevant tracking/guidance information, and transforms motions from screen to the intervention site.
6. Recalibrate tools after every tool change. For every needle or other tool, separate sensors and calibration procedures are necessary.

In contrast, our Intuitive Fusion system (Fig. 2) requires no external tracking, calibrations, or mental acrobatics:

1. Clip the Intuitive Fusion bracket onto the ultrasound probe. This fixes camera and projection units relative to the ultrasound imaging array.
2. (Define target region in ultrasound image and camera images. This requires some operator interaction.)
3. Perform ultrasound-guided intervention, with needle and probe tracking performed by the probe bracket. For tracking, no specific requirements apply to the operator; guidance requires manual alignment according to symbols shown on-screen.

A second setup will incorporate surface-based registration with pre-operative image data and planning (cf.[1]):

1. Perform pre-operative planning on CT/MRI data.
2. Clip the Intuitive Fusion bracket on the ultrasound probe.
3. Acquire one ultrasound image, simultaneously reconstruct the surface. Optionally, use probe tracking to acquire 3D-US.
4. Register ultrasound & surface to pre-operative image. These two steps can be repeated online to maintain registration.
5. Perform ultrasound-guided intervention, with probe, patient, and needle tracking performed by the probe bracket.

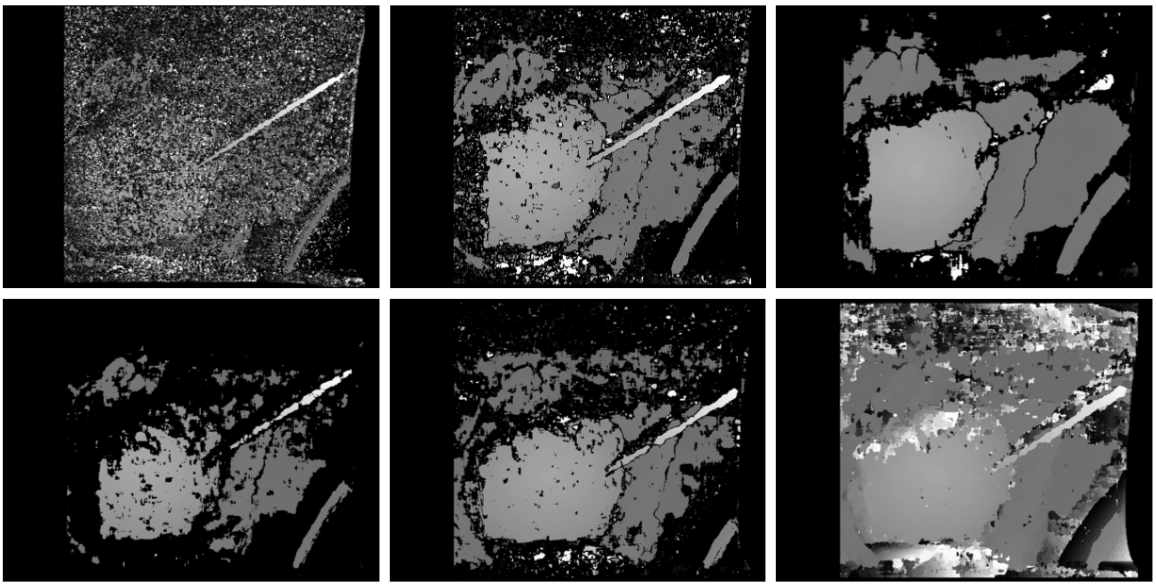

Figure 3. Surface reconstruction using structured light: increasing block matching window size (first row); decreasing uniqueness threshold (second row). The region under projection can be clearly seen partially overlapping the phantom.

2.3 Algorithms

To allow the described setups to work, a variety of hard- and software components need to be set up and calibrated. The relevant algorithmic steps are described in the following.

2.3.1 Surface Reconstruction

The challenges in reconstructing the patient skin surface in the given application are similar to the ones encountered e.g. in mobile robotics, such as a wide range of distances, surfaces shapes, and surface (skin) types. Peculiar to this handheld device application, though, is the very wide range of stereo disparities to deal with, due to the very close minimum distance of objects to the stereo camera (e.g. the needle can be placed essentially almost between the single camera lenses, and the surfaces may curve up towards the cameras as well). Compounding these difficulties are smooth skin (and surrounding) surfaces without sufficient features such as hair, wrinkles etc.

After calibration, stereo video frames can be collected. In the setup used for the experiments described later on, resolution and frame rate were limited by the hardware (stock web cams, VGA 640×480 resolution at 15fps each). Concurrently with frame acquisition, image rectification and disparity estimation are taking place. Using a block matching algorithm (`cvFindStereoCorrespondenceBM()`), a frame rate of about 15fps can be sustained online.

However, it is necessary to augment the stereo camera's view to reach a sufficiently high surface match ratio. Problems for surface reconstruction arise mainly from smooth surface patches, where block matching fails. To alleviate this, structured-light patterns are projected into the camera fields of view, supplying features onto the surface (Fig. 3). A locally-unique structured light overlay overcomes the problem of potentially very high disparities without aliasing the pattern.

In addition, the pattern projection has to be synchronized with frame acquisition on a frame-by-frame basis for algorithmic reasons – the patterns interfere with needle detection later on. In the described implementation, for surface acquisition the pattern is projected, the stereo camera acquires one frame, and pattern projection is stopped.

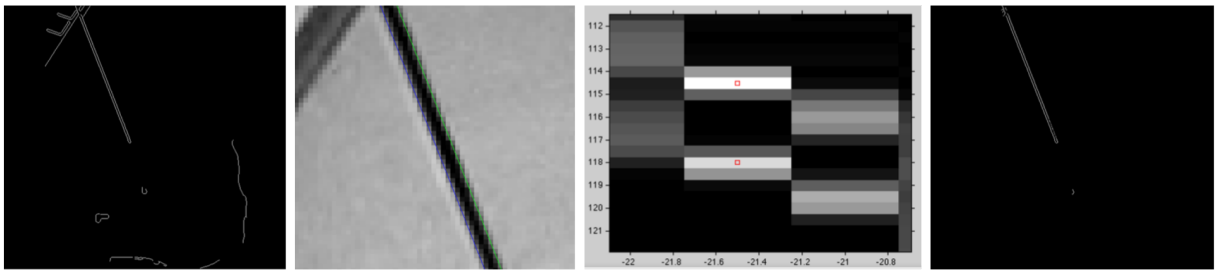

Figure 4. 2D needle detection: Canny edges, Hough transform lines, "twin peaks" in the transform, cropped input for possible 2nd pass (left to right)

2.3.2 Needle Detection

The envisioned system shall be able to find generic needles during standard procedures, it shall perform completely automatically, and run in real time ("online"). These requirements do not allow the addition of special tracking markers to the needle, nor to require the operator to perform special needle calibration steps or motions, or to modify the standard procedure. In short, any needle introduced into the field of view of the sensors on top of the background features shall be picked up by the tracking system and its pose in stereo camera coordinates be computed with a frame rate sufficiently high to allow interactive needle guidance. As can be seen in Fig. 3, the disparity map alone is in most cases not sufficient to uniquely identify a needle, even under good visibility conditions.

The proposed algorithm can be summarized in three steps: (1) find needle candidates, (2) prune candidate set, and (3) generate a 3D result pose to be returned.

1. Find needle candidates:

   In spite of the required general, unconstrained environment, the (any) needle can be found by looking for the longest, most prominent, straight line segment in the image.

   Strategies to find those features are well-established; the most straightforward might be edge-detection on the rectified grey-scale image, computation of the Hough transform on the resulting binary image, and selection of the n strongest peaks in the transform as needle candidates. This process is performed on both left and right images separately. Using the probabilistic Hough transform, one finds the presumed end points of detected 2D lines.

   In addition, Fig. 4 indicates ways to improve the line detection result in future work. For a well-visible needle, its image in the edge detection frame consists of two nearly-parallel lines (corresponding to the edges of said needle's image, they extend in parallel over at least part of the length of the needle). This can be used to increase the robustness against false-positive needle detections. Furthermore, limited to the vicinity of the two peaks of the Hough transform corresponding to these parallel lines, the transform can be robustly recomputed with a higher resolution, thus increasing the final precision of the needle tracking.

2. Prune candidate set:

   The procedure above results in a potentially large number of line segments all over the two rectified stereo images, most of which do not correspond to any actual needle. In the next step, this set is pruned to find probable needle line candidates for further processing.

   Iterating over all combinations (i, j) of line segments ${}^{L}s_i$ and ${}^{R}s_j$ from the left and right images I_L and I_R, in a first step the combinations with horizontally non-overlapping segments (i.e. the segment in the left image is completely above or below the one in the right) are rejected. Second, combinations with line segments shorter than a threshold l_{min} (heuristically set to $l_{min} = 30px$) are rejected as well.

   In the third step, stereo information is used for further checks (Fig. 5). For each pixel ${}^{L}p_i(y)$ on one line segment ${}^{L}s_i$ in the left image I_L, the corresponding pixel ${}^{R}p_j(y)$ on the line segment ${}^{R}s_j$ at the same height

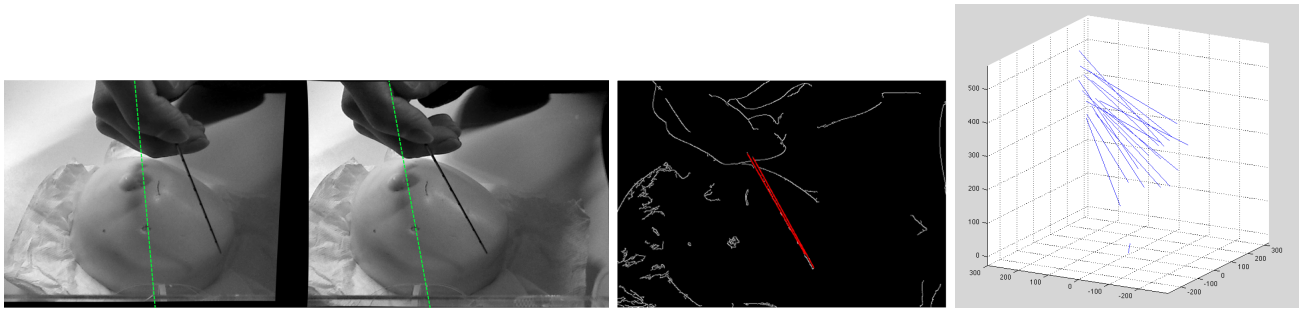

Figure 5. 3D needle detection: Input stereo images (right, with overlaid guidance markers); detected needle line (center); 3D line representations (right, in stereo coordinates [mm]).

y in the other image I_R is computed (if necessary, extrapolating the line segment ${}^{R}s_j$ to a line), and their disparity $d_{i,j}(y) =^{L} p_i(y).x -^{R} p_j(y).x$ determined. This disparity is then compared against the known disparity map D, to verify that the 3D line ${}^{C}l_{i,j}$ (implicitly defined by the two current 2D line segments in camera coordinates C) is properly positioned in front of ("above") the surface in the background. As the needle usually disrupts block matching in the surface reconstruction, leaving noise in its wake, only the neighborhood of the segment ${}^{L}s_i$ in the disparity map can be compared against $d_{i,j}(y)$. If more than a certain fraction (set to 10%) of the checked disparities indicate the reconstructed line point to lie behind the background surface, this line combination (i, j) is rejected.

During all those checks, the combination corresponding to the longest – valid – line segment ${}^{L}s_i$ encountered so far is saved, if it exists.

3. 3D result generation:

   Finally, the corresponding 3D line ${}^{C}l$ is returned. The line representation is computed using stereo calibration info, providing two distinct points ${}^{C}p_{upper}$ and ${}^{C}p_{lower}$ on the line. As the needle is detected with 4 degrees of freedom (DoF) – the axis-rotational and axis-translational DoF are not recoverable – this representation is sufficient.

2.3.3 Calibrations

In particular, two calibrations have to be performed for the vision components: intrinsic calibration of and extrinsic stereo calibration between the two single cameras $C_{1,2}$, as well as calibration between ultrasound and stereo camera ${}^{C}M_{US}$.

The first step is to calibrate the cameras – first their intrinsic parameters, then their stereo (extrinsic) parameters. For this (and all subsequent computer-vision-related steps), the OpenCV toolkit (via the OpenCV 2.0 C interface) was used, based on the standard checkerboard calibration approach. Undistortion lookup tables are computed and stored in this step.

More interesting is the $C-US$ calibration, as the two modalities cover completely disparate feature spaces and fields of view. The transformation between both coordinate systems was determined by holding the probe/bracket combination over an ultrasound phantom, imaging a needle using both modalities in different poses, covering approximately the whole combined surveilled 4-DoF space (Fig. 7). In stereo camera space, two points ${}^{C}p_{i1,i2}$ on the needle line are stored for each pose i, in ultrasound space the needle intersection location ${}^{US}p_{iX}$ is provided manually (Fig. 6). The calibration procedure exploits the required collinearity between the three points ${}^{C}p_{i1}$,${}^{C}p_{i2}$, and ${}^{US}p_{iX}$, as it determines the transformation ${}^{C}M_{US}$ which minimizes the error in

$$\overline{{}^{C}p_{i1},{}^{C}p_{i2}} \times \overline{{}^{C}p_{i2},{}^{C}M_{US} \cdot {}^{US}p_{iX}} = 0. \quad (1)$$

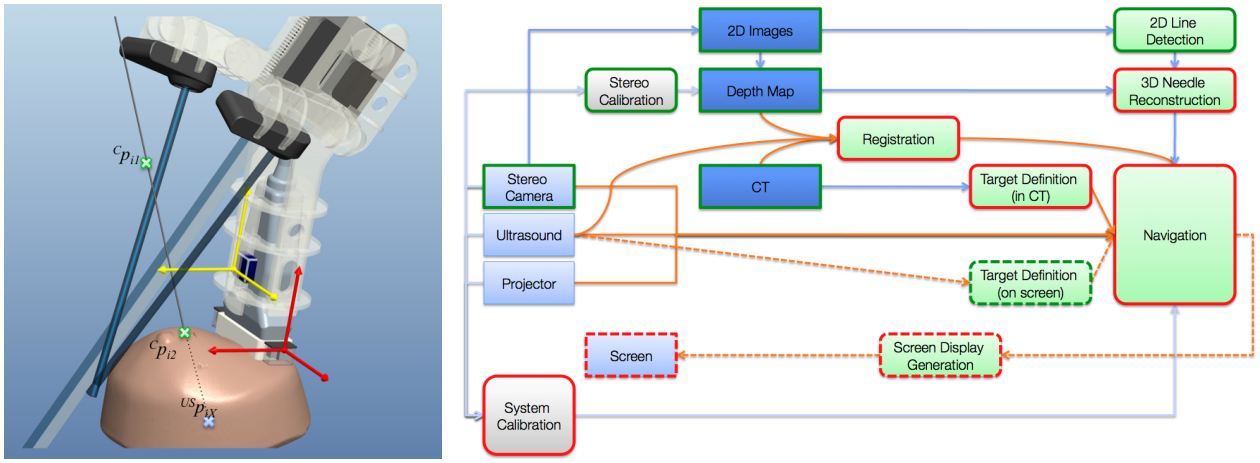

Figure 6. Camera-Ultrasound calibration (left): three points in two modalities along the needle shaft are collinear. Intuitive Fusion data flow (right).

2.3.4 Guidance

Guidance of the user holding the needle, attempting to reach a target, will be provided in two different ways: "end-point guidance" by overlaying the anticipated needle intersection with the ultrasound plane onto the live ultrasound stream, or as "full-needle" guidance by overlaying needle shapes onto the live stereo camera views, asking the user to align the visible needle with the shapes in both views simultaneously.

For "end-point guidance", given the needle pose ^{C}l and the calibration transformation $^{US}M_{C}$ from stereo camera to ultrasound coordinates, one can compute the intersection point $^{US}p_{X}$ of the needle line $^{US}l = ^{US}M_{C} \cdot ^{C}l$ and the (thin) ultrasound frame I_{US} in ultrasound coordinates. The intersection point and/or the needle is drawn onto the live ultrasound image whenever a needle is detected (similar to Fig. 1). This form of guidance obviously does not require any form of user interaction, and can be considered "passive" in that it works transparently and without imposing any particular approach. Furthermore, it allows free motion of both the needle and the probe.

In contrast, the second method of "full-needle guidance" requires to define both a target $^{US}p_{X}$ in ultrasound as well as a skin entry point $^{C}p_{E}$ in (one of) the video views. Both together define a 3D line $^{C}l_{X,E}$, which is then projected into the 2D views I_{L} and I_{R} of the left and right cameras, respectively (concept shown in Fig. 5). In the current system, this requires the probe to be held still, as the position of the defined target is not tracked.

3. RESULTS

Preliminary results with both vision-based and opto-inertial systems can be summarized as follows:

1. Skin surfaces (phantom and human) generally do not exhibit enough texture information to allow stereo disparity estimation on their own. Locally-unique structured light projection (SLS) remedies this – but interferes with needle tracking, requiring synchronization.

2. Typical scenes allow the immediate detection of generic needles (using Canny edge detection, Hough line transform, and lots of stereo line combinations) against general backgrounds without preparation, but this has to be anti-synchronized with SLS. Pruning of the 3D lines set resulting from typical stereo views requires additional steps (plausibility checks, comparison against disparity map etc.). Within the relevant field of view, even simple web cams achieve $0.25mm/px$ resolution, which we expect to transfer into 3D tracking precision, as the described approach allows for sub-pixel accuracy of line detection.

3. Camera-ultrasound calibration using the described linear-needle-alignment approach returns known ground-truth transformations based on perfect simulation data, even for as few as five needle poses. As soon as errors are introduced into the system, the accuracy decreases (e.g. for stereo-space errors of $\pm 1.6mm$ and

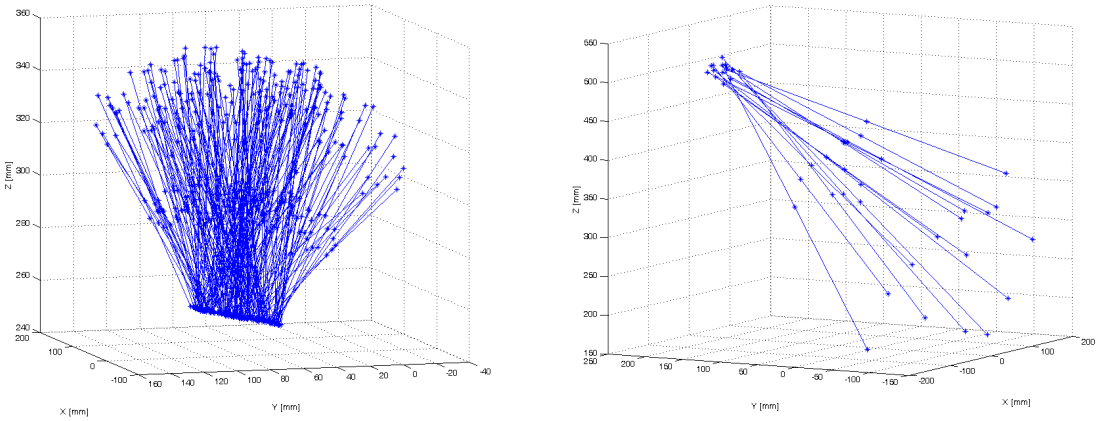

Figure 7. Camera-Ultrasound calibration results using simulated data (200 needle poses, left), real data (15 needle poses, right). Each line is comprised of the two 3D points from stereo vision and one 2D point from the ultrasound intersection location, projected into 3D camera space using the recovered calibration transformation.

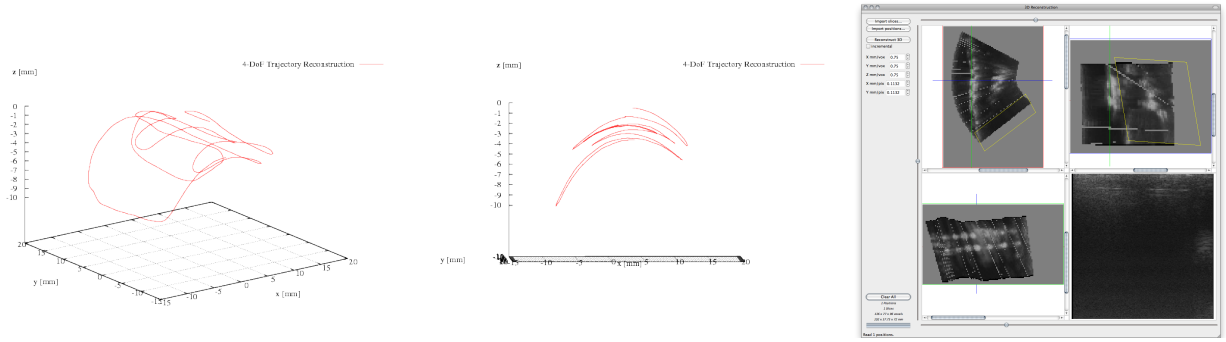

Figure 8. Intuitive Fusion opto-inertial probe trajectory reconstruction on a cylindrical phantom (right); 3D ultrasound volume reconstruction (left).

ultrasound-space errors of $\pm 0.8mm$, the resulting transformation error was below approx. $6mm$ and $5°$). With twice the error range on the input data, simulation-data calibrations still stabilized when using more than 100 needle poses.

4. Opto-inertial 4-DoF probe tracking may suffer from drift ($< 10\%$) and intermittent tracking loss on skin (Fig. 8). This may be attributed to both featureless skin surfaces, but also the interference of ultrasound gel with the optical tracking units.

   Nevertheless, our in-vivo experience (in an IRB-approved protocol) was positive in that the operator did not report inconvenience from the use of the additional opto-inertial sensor bracket (in spite of its large prototype-hardware footprint), and usability was not impaired by the additional cabling.

   Under laboratory conditions, 3D ultrasound volumes can be collected reliably. In general, the most relevant impediment to clinical use of opto-inertial tracking in actual interventional settings is the required sustained surface contact to prevent tracking loss. However, at least two functions appear to be useful for future development: support for one-time ("sweep") 3D ultrasound volume acquisition (plus surface data[1]) prior to registration with pre-operative imaging data, and concurrent orientation measurements, which in turn might support needle tracking and potentially registration.

4. CONCLUSIONS

The Intuitive Fusion system – in the form of a sensor-augmented bracket for handheld ultrasound probes – captures probe trajectories, patient surface data, and tracks needles and ultrasound probes in 3D and real time, without necessary calibration of tools or lengthy registration procedures. It does so without the need for external references such as tracking system emitters or cameras. Upcoming work will concentrate on the achievable precision and robustness of the presented methods.

ACKNOWLEDGMENTS

The authors receive Maryland TEDCO funding (03/2011-) for the described project. Furthermore we gratefully acknowledge the valuable input of Messrs. Robert de Jong and Dr. Clifford Weiss.

REFERENCES

[1] Seth Billings, Ankur Kapoor, Bradford J. Wood, Emad Boctor: *A hybrid surface/image based approach to facilitate ultrasound/CT registration.* SPIE Medical Imaging 2011, Orlando, FL, USA

[2] Candice Chan, Felix Lam, And Robert Rohling: *A needle tracking device for ultrasound guided percutaneous procedures.* Ultrasound in Med. & Biol., Vol. 31, No. 11, pp. 14691483, 2005

[3] A. M. Goldsmith, P. C. Pedersen, T. L. Szabo: *An Inertial-Optical Tracking System for Portable, Quantitative, 3D Ultrasound.* IEEE International Ultrasonics Symposium IUS 2008, Beijing/China

[4] G. Hager and E. Wegbreit: *Acquisition of three-dimensional images by an active stereo technique using locally unique pattern.* US Patent 7103212.

[5] R. James Housden, Andrew H. Gee, Graham M. Treece and Richard W. Prager: *Sensorless Reconstruction of Unconstrained Freehand 3D Ultrasound Data.* Ultrasound in Medicine & Biology, Volume 33, Issue 3, March 2007, Pages 408-419

[6] Mahmoud M. Ismail, Emad Boctor, Katsuyuki Taguchi, Jingyan Xu, Benjamin M. Tsui: *3D-guided CT reconstruction using time-of-flight camera.* SPIE Medical Imaging 2011, Orlando, FL, USA

[7] Robert Krempien, Harald Hoppe, Lüder Kahrs, Sascha Däuber, Oliver Schorr, Georg Eggers, Marc Bischof, Marc W. Munter, Jürgen Debus, Wolfgang Harms: *Projector-Based Augmented Reality For Intuitive Intraoperative Guidance In Image-Guided 3D Interstitial Brachytherapy.* Int. J. Radiation Oncology Biol. Phys., Vol. 70, No. 3, pp. 944952, 2008

[8] F. Sauer, A. Khamene: *Video-assistance for ultrasound guided needle biopsy.* US Patent 6612991

[9] Philipp J. Stolka, Matthias Keil, Georgios Sakas, Elliot R. McVeigh, Russell H. Taylor, Emad M. Boctor: *A 3D-elastography-guided system for laparoscopic partial nephrectomies.* SPIE Medical Imaging 2010 (San Diego, CA/USA)

[10] Philipp J Stolka, Hyun-Jae Kang, Michael Choti, Emad M Boctor: *Multi-DoF Probe Trajectory Reconstruction with Local Sensors for 2D-to-3D Ultrasound*, IEEE International Symposium on Biomedical Imaging 2010, Rotterdam, The Netherlands)

[11] Philipp J Stolka, Hyun-Jae Kang, Emad M Boctor: *The MUSiiC Toolkit: Modular Real-Time Toolkit for Advanced Ultrasound Research.* MICCAI 2010, International Workshop on Systems and Architectures for Computer Assisted Interventions 2010 (Beijing/China)