Fast, Intuitive, Vision-Based: Performance Metrics for Visual Registration, Instrument Guidance, and Image Fusion

Ehsan Basafa(✉), Martin Hoßbach, and Philipp J. Stolka

Clear Guide Medical, Baltimore, MD 21211, USA
{basafa,hossbach,stolka}@clearguidemedical.com

Abstract. We characterize the performance of an ultrasound+ computed tomography image fusion and instrument guidance system on phantoms, animals, and patients. The system is based on a visual tracking approach. Using multi-modality markers, registration is unobtrusive, and standard instruments do not require any calibration. A novel deformation estimation algorithm shows externally-induced tissue displacements in real time.

Keywords: Ultrasound · Computed tomography · Image fusion · Instrument guidance · Navigation · Deformable modeling · Computer vision · Metrics

1 Introduction

For many ultrasound (US) operators, the main difficulty in needle-based interventions is keeping hand-held probe, target, and instrument aligned at all times after initial sonographic visualization of the target. In other cases, intended targets are difficult to visualize in ultrasound alone – they may be too deep, occluded, or not echogenic enough. To improve this situation, precise and robust localization of all components – probe, target, needle, and pre- or intra-procedural 3D imaging – in a common reference frame and in real time can help. This allows free motion of both target and probe, while continuously visualizing targets. Easy-to-use image fusion of high resolution 3D imaging such as magnetic resonance (MR) and computed tomography (CT) with real-time ultrasound data is the key next stage in the development of image-guided interventional procedures.

The Clear Guide SCENERGY (Clear Guide Medical, Inc., Baltimore, MD) is a novel CT-US fusion system aiming to provide such user-friendly and accurate guidance. Its main differentiator is the intuitive provision of such fusion and guidance capabilities with only minor workflow changes. The system is cleared through FDA 510(k), CE Mark, and Health Canada license.

2 Image Fusion and Guidance System

The Clear Guide SCENERGY provides CT and US fusion for medical procedures, as well as instrument guidance to help a user reach a target in either modality

R. Shekhar et al. (Eds.): CLIP 2016, LNCS 9958, pp. 9–17, 2016.
DOI: 10.1007/978-3-319-46472-5_2

Fig. 1. (a) Clear Guide SCENERGY system, with touchscreen computer, hand-held SuperPROBE (ultrasound probe with mounted Optical Head), connected to a standard ultrasound system. (b) User interface in Fusion Mode, with registered US and CT and overlaid tracked instrument path.

(Fig. 1(a)). Using skin-attached markers (Clear Guide VisiMARKERs) that are visible both optically and radiographically, the system tracks the hand-held US probe pose in real time relative to the patient, and extracts the corresponding CT slice for overlaid display with the current live US slice (Fig. 1(b)). Instrument and target (if selected) are overlaid onto the live CT/US fused view for guidance.

2.1 System

The Optical Head is rigidly attached to standard ultrasound probes via probe-specific brackets, all of which is collectively called the Clear Guide SuperPROBE. Stereo cameras in the Optical Head observe the field of view next to the SuperPROBE, and detect both instruments and markers. Infrared vision and illumination enable this even in low-light environments.

The touchscreen computer provides the user interface and performs all computations. Ultrasound image acquisition and parameterization happens through the user's existing ultrasound and probe system, to which the system is connected through a video connection, capturing frames at full frame rate and resolution. Imaging geometry (depth and US coordinate system) is extracted by real-time pattern matching against known pre-calibrated image modes.

The system receives CT volumes in DICOM format via network from a Picture Archive and Communication System (PACS) or USB mass storage.

3 Interventional Workflow

The clinical workflow (Fig. 2(a)) consists of two functional modes: Registration and Imaging. The system starts in Registration mode (Sect. 3.1) to allow the user to import CT data, and to perform a visual-sweep registration. The operator then switches into Imaging mode (Sect. 3.2), where fused US+CT images and instrument guidance are displayed in real time.

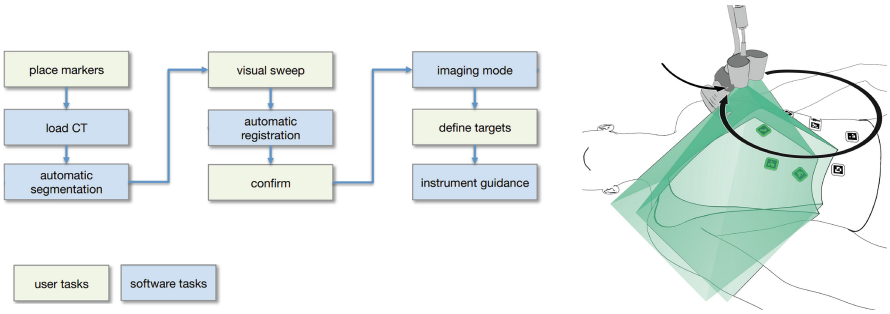

Fig. 2. (a) Workflow for complete image-guided procedure using the SCENERGY system. (b) Example SuperPROBE motion during Visual Sweep Registration showing cameras' fields of view.

3.1 Registration

CT Scan with VisiMARKERs. The registration between pre-procedural CT and the patient relies on multi-modality markers placed on the skin, and their locations' exact reconstruction by the cameras. Thus, it is important to ensure that at least some markers will be visible during the entire procedure. Registration is more robust when marker placement and spacing is irregular and non-symmetric.

In a typical clinical workflow, 5–15 fiducial markers are added to the patient prior to the pre-procedural scan. During loading of that scan, these "early markers" are automatically segmented based on shape and radiopacity. However, the clinician has the option of adding further "late markers" before registration. These provide additional points of reference for later tracking to improve tracking robustness, but do not affect registration. After registration, the system does not differentiate between early and late markers, treating all markers as ground truth for tracking.

The system also segments out the patient skin surface from the CT volume using the Otsu algorithm [5]. This surface is used for three purposes: user reference, aiding in registration, and creating a deformable model (Sect. 3.2).

Visual Tracking. The system continuously scans the stereo camera images for the markers' visual patterns [4] and, through low-level pattern detection, pattern interpretation, stereo reconstruction, and acceptance checking, provides the 6-DoF marker pose estimation for each marker. After registration, the probe pose estimation is based on observations of (subsets of) the markers.

Visual Sweep Registration. "Registration" (the pairing of real-time optical data and the static CT dataset) is performed in two steps: first, visual marker observations are collected to create a 3D marker mesh, and second, image data and observations are automatically matched by searching for the best fit between them. Though this process is not new in itself, the implementation results in a simplification of the user workflow compared to other systems.

After loading the static data, the user performs a "visual sweep" of the region of intervention, smoothly moving the SuperPROBE approximately 15 cm to 20 cm above the patient over each of the markers in big loops (Fig. 2(b)). The sweeps collect neighboring markers' poses and integrate them into a 3D marker mesh, with their position data improving with more observations. The software automatically finds the best correspondence between the observed and segmented markers based on the registration RMS error, normal vector alignment, and closeness to the segmented patient surface. The continuously updated Fiducial Registration Error (FRE) helps in assessing the associated registration accuracy. Misdetected, shifted, or late markers do not contribute to the FRE or the registration itself, if they fall more than 10 mm from their closest counterpart in the other modality. However, note that the commonly used FRE is not directly correlated to the more clinically relevant Target Registration Error (TRE) [2]. No operator interaction (e.g. manual pairing of segmented and detected markers) is required for automatic registration.

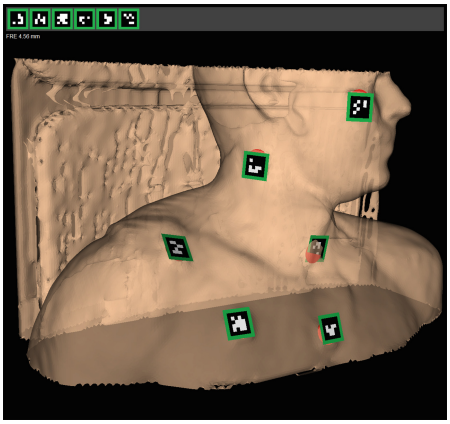

Fig. 3. Visual Sweep registration result, showing markers matched (green) to CT-segmented locations (red). (Color figure online)

As markers are detected, their relative positions are displayed and mapped onto the segmented patient skin surface according to the best found registration (Fig. 3). This marker mesh is the ground truth for future probe pose estimation.

3.2 Imaging

Fusion Image Guidance. The system constantly reconstructs CT slices from the static volume and overlays them on the US image (Fig. 4) using the current probe pose relative to the observed marker mesh (based on real-time ongoing registration of current observations to the ground truth mesh) and the current US image geometry as interpreted from the incoming real-time US video stream.

Dynamic Targeting. The operator may define a target by tapping on the live US/CT image. Visual tracking allows continuous 3-D localization of the target point relative to the ultrasound probe, fixed in relation to the patient. This "target-lock" mechanism enhances the operator's ability to maintain instrument alignment with a chosen target, independent of the currently visualized slice. During the intervention, guidance to the target is communicated through audio and on-screen visual cues (Fig. 4).

Deformation Modeling. Pressing the ultrasound probe against a patient's body, as is common in most ultrasound-enabled interventions, results in

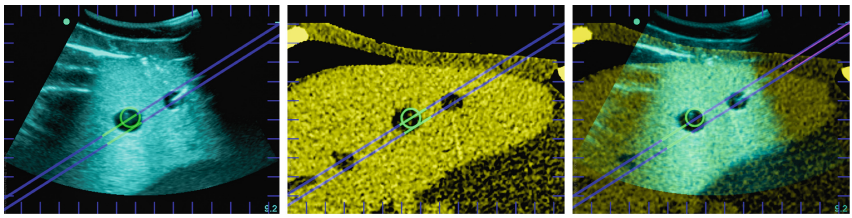

Fig. 4. (a) Live US image, (b) corresponding registered CT slice, (c) fusion image of both modalities (all images showing overlaid instrument and target guidance, with magenta lines indicating PercepTIP [6] needle insertion depth). Note the CT deformation modeling matching the actual US image features. (Color figure online)

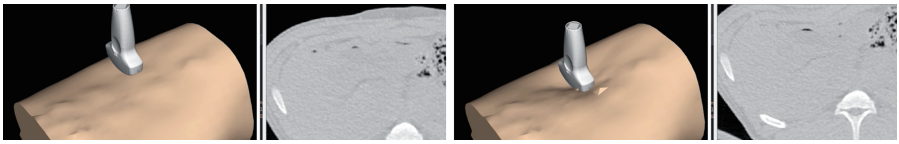

Fig. 5. Surface segmented from CT with tracked probe in-air (a), with probe pressing down on the surface (b).

deformation seen in the real-time ultrasound image. When using image fusion, the static image would then be improperly matched to the ultrasound scan if this effect were not taken into account. Based on probe pose, its geometry, and the patient surface, the system thus estimates collision displacements and simulates the corresponding deformation of the CT slice in real time (Figs. 4 and 5). The underlying non-linear mass-spring-damper model approximates the visco-elastic properties of soft tissues, and is automatically generated and parameterized by the CT's Hounsfield values at the time of loading and segmenting the CT data [1].

4 Performance Metrics

Conventionally, interventional image guidance systems are described in terms of fiducial registration error (FRE, which is simple to compute at intervention time) and target registration error (TRE, which is more relevant, but harder to determine automatically). In addition to that, we also break down the performance evaluation of the presented system into several distinct metrics as follows.

4.1 Segmentation Accuracy and FRE

Distances between hand-selected centers of markers ("gold standard") and those from the automated Clear Guide SCENERGY algorithm indicate segmentation accuracy. Because the automated system considers all voxels of marker-like intensity for centroid computation, we believe the system actually achieves higher

precision than manual "ground truth" segmentation which was based on merely selecting the marker corners and finding the center point by 3D averaging.

Segmentation error (automatic segmentation compared to manual center determination) was (0.58 ± 0.4) mm ($n = 2$ pigs, $n = 2$ patients, $n = 5$ phantoms; $n = 64$ markers total, $6 \ldots 11$ markers each), taking approx. 5 s for one complete volume.

Fiducial registration error (FRE) is the RMS between segmented CT and observed camera marker centers. It was (2.31 ± 0.94) mm after visual-sweep registration ($n = 2$ breathing pigs, $n = 7$ breathing patients, $n = 5$ phantoms; $4 \ldots 11$ markers registered for each; all at 0.5 mm CT slice spacing).

No instances of incorrect marker segmentation or misregistration (i.e. resulting wrong matches) were observed (100 % detection rate; $FP = FN = 0$).

4.2 Fusion Accuracy (TRE)

Fusion accuracy was measured as *Tissue Registration Error (TRE)* (in contrast to its conventional definition as *Target Registration Error*, which constrains the discrepancy to just a single target point per registration). It depends on registration quality (marker placement and observations) and internal calibration (camera/US). Fused image pairs (collected by a novice clinical operator; $n = 2$ breathing pigs, $n = 7$ breathing patients, $n = 5$ phantoms) were evaluated to determine fusion accuracy. As tens of thousands of image pairs were collected in every run, we manually selected pairs with good anatomical visualization in both US and CT; however not selecting for good registration, but only for good visibility of anatomical features. To ensure a uniform distribution of selected pairs, we systematically chose one from each block of $m = 350 \ldots 500$ consecutive pairs ($4 \ldots 94$ pairs per run).

Discrepancy lines were manually drawn on each image pair between apparently corresponding salient anatomical features, evenly spaced (approx. 10 lines per pair; $59 \ldots 708$ lines per run) (Fig. 6(a)). After extreme-outlier removal (truncation at 3× interquartile range; those correspond to clearly visible mismatches) and averaging first within (i.e. instantaneous accuracy) and then across pairs per run (i.e. case accuracy) to reduce sampling bias, the resulting *Tissue Registration Error (TRE)* was 3.75 ± 1.63 mm.

4.3 Systematic Error

Systematic error is the cumulative error observed across the entire system, which includes the complete chain of marker segmentation, sweep-based registration, probe tracking, CT slicing, and instrument guidance errors. This performance metric is a "tip-to-tip" distance from the needle point shown in registered ground-truth CT to the same needle point shown by overlaid instrument guidance (Fig. 6(b)). It represents the level of trust one can place in the system if no independent real-time confirmation of instrument poses – such as from US or fluoro – is available. (Note that this metric does not include User Error, i.e. the influence of suboptimal needle placement by the operator.) This metric is sometimes

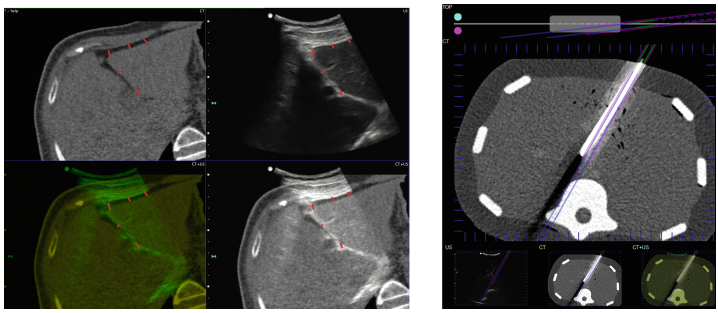

Fig. 6. (a) Tissue Registration Error computation based on discrepancy lines (red). (b) Systematic Error computation based on difference between needle in CT and overlaid instrument guidance. (Color figure online)

referred to as "tracking error" – "the distance between the 'virtual' needle position computed using the tracking data, and the 'gold standard' actual needle position extracted from the confirmation scan" [3]. The total systematic error was found to be (3.99 ± 1.43) mm ($n = 9$ phantoms with FRE (1.23 ± 0.58) mm; with results averaged from $2 \ldots 12$ reachable probe poses per registered phantom). The tracked CT is displayed at $15 \ldots 20$ fps, and instrument guidance at 30 fps.

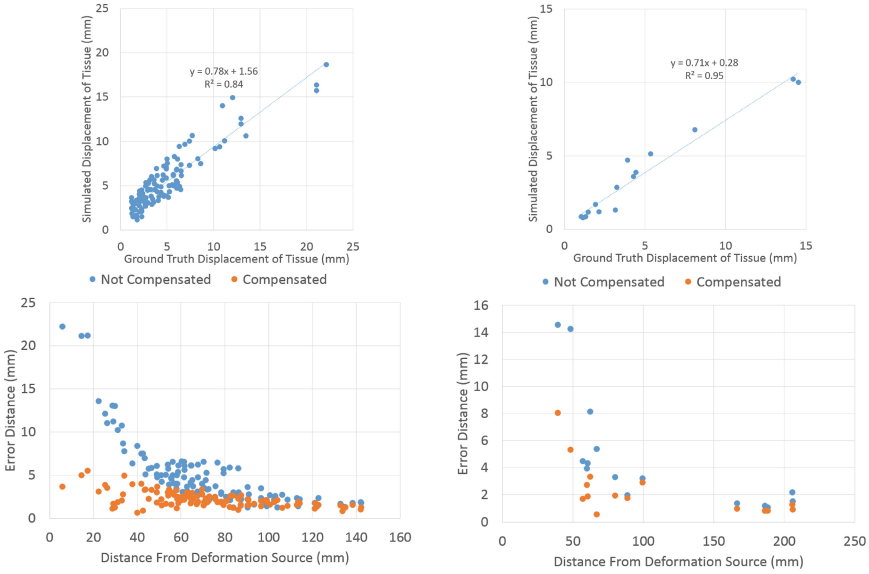

Fig. 7. Deformation simulation results: displacement recovery (top) and residual error (bottom), for ex-vivo liver (left) and in-vivo pig (right)

4.4 Deformation Accuracy

The system simulates deformation of the static CT image to compensate for compression error caused by pressing the probe onto the patient tissue. Performance testing measured the estimated recovery (i.e. simulated displacement for each target divided by original compression displacement) and the residual error (Fig. 7). In a silicone liver dataset [7], recovery was estimated at 78.2 % (n = 117 BB targets; R^2 = 0.84), whereas an in-vivo porcine dataset yielded 71.4 % (n = 15 BB targets; R^2 = 0.95) recovery; with the simulation running at 50 fps and a settling time of 1 . . . 2 s. The deformation model thus demonstrates a clear benefit as compared to no deformation model.

5 Conclusion

We described a novel US+CT image fusion and instrument guidance system, based on inside-out visual tracking from hand-held ultrasound probes. It simplifies the user workflow compared to the state of the art, as it provides automatic patient and marker segmentation, allows for rapid "visual sweep" patient/CT registration, works with nearly all standard instruments, and naturally does not suffer from the usual line-of-sight or EM-field-disturbance drawbacks of conventional tracking systems.

A variety of experiments characterized the performance of all workflow steps under a wide range of conditions (lab, veterinary, and clinical). The results show the system to have an accuracy comparable to established systems (e.g. Philips PercuNav [3]). Therefore, we believe, the system can be readily adopted by physicians for user-friendly, intuitive fusion and instrument guidance.

One limitation of this study is the relatively low number of live patient/animal trial runs. Work is underway to increase this number and provide more robust statistical inferences. The number of phantom experiments was kept low in order to not skew the results towards better accuracy inherent in tests involving stationary, non-breathing phantoms. Future work will focus on the compensation of patient-breathing-induced errors using the same visual tracking technology.

References

1. Basafa, E., Farahmand, F.: Real-time simulation of the nonlinear visco-elastic deformations of soft tissues. Int. J. Comput. Assist. Radiol. Surg. **6**(3), 297–307 (2011)
2. Fitzpatrick, J.M.: The role of registration in accurate surgical guidance. Proc. Inst. Mech. Eng. H **224**(5), 607–622 (2010)
3. Krucker, J., Xu, S., Venkatesan, A., Locklin, J.K., Amalou, H., Glossop, N., Wood, B.J.: Clinical utility of real-time fusion guidance for biopsy and ablation. J. Vasc. Interv. Radiol. **22**(4), 515–524 (2011)
4. Olson, E.: AprilTag: a robust and flexible visual fiducial system. In: Proceedings of the IEEE ICRA, pp. 3400–3407 (2011)

5. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
6. Stolka, P.J., Foroughi, P., Rendina, M., Weiss, C.R., Hager, G.D., Boctor, E.M.: Needle guidance using handheld stereo vision and projection for ultrasound-based interventions. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part II. LNCS, vol. 8674, pp. 684–691. Springer, Heidelberg (2014)
7. Suwelack, S., Röhl, S., Dillmann, R., Wekerle, A.L., Kenngott, H., Müller-Stich, B., Alt, C., Speidel, S.: Quadratic corotated finite elements for real-time soft tissue registration. In: Nielsen, P.M.F., Wittek, A., Miller, K. (eds.) Computational Biomechanics for Medicine: Deformation and Flow, pp. 39–50. Springer, New York (2012)

http://www.springer.com/978-3-319-46471-8

Clinical Image-Based Procedures. Translational Research in Medical Imaging
5th International Workshop, CLIP 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Proceedings
Shekhar, R.; Wesarg, S.; González Ballester, M.Á.; Drechsler, K.; Sato, Y.; Erdt, M.; Linguraru, M.G.; Oyarzun Laura, C. (Eds.)
2016, IX, 85 p. 40 illus., Softcover
ISBN: 978-3-319-46471-8