Visual Tracking for Multi-Modality Computer-Assisted Image Guidance

Ehsan Basafa, Pezhman Foroughi, Martin Hossbach, Jasmine Bhanushali, Philipp Stolka
Clear Guide Medical, 3400 Clipper Mill Rd, Baltimore, MD, USA 21211

ABSTRACT

With optical cameras, many interventional navigation tasks previously relying on EM, optical, or mechanical guidance can be performed robustly, quickly, and conveniently. We developed a family of novel guidance systems based on wide-spectrum cameras and vision algorithms for real-time tracking of interventional instruments and multi-modality markers. These navigation systems support the localization of anatomical targets, support placement of imaging probe and instruments, and provide fusion imaging. The unique architecture – low-cost, miniature, in-hand stereo vision cameras fitted directly to imaging probes – allows for an intuitive workflow that fits a wide variety of specialties such as anesthesiology, interventional radiology, interventional oncology, emergency medicine, urology, and others, many of which see increasing pressure to utilize medical imaging and especially ultrasound, but have yet to develop the requisite skills for reliable success. We developed a modular system, consisting of hardware (the Optical Head containing the mini cameras) and software (components for visual instrument tracking with or without specialized visual features, fully-automated marker segmentation from a variety of 3D imaging modalities, visual observation of meshes of widely-separated markers, instant automatic registration, and target tracking and guidance on real-time multi-modality fusion views). From these components, we implemented a family of distinct clinical and pre-clinical systems (for combinations of ultrasound, CT, CBCT, and MRI), most of which have international regulatory clearance for clinical use. We present technical and clinical results on phantoms, ex- and in-vivo animals, and patients.

Keywords: Image fusion, multi-modality registration, instrument tracking, navigation, augmented reality

1. INTRODUCTION

Medical interventions such as biopsies, ablations, catheterization, nerve blocks, drainages etc. are nowadays primarily performed under image guidance, i.e. by using a medical imaging modality to visualize relevant anatomy, target areas, and the instruments used to perform the intervention. Such imaging modalities include ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), cone-beam CT, fluoroscopy, and others. To achieve clinically required accuracy and to reduce complications, medical navigation systems enable navigation on different imaging modalities, perform image fusion between static and dynamic modalities, and allow instrument guidance. Enabling aspects of such navigation systems are **tracking** (ongoing localization of relevant entities such as instruments, imaging, and organs) and **registration** (establishing of geometric relationships between different physical components of the system to correlate locations between them).

Current solutions are predominantly based on **electromagnetic** (EM) and – to a lesser extent – **optical tracking**, both of which have limitations that hamper their clinical acceptance. EM tracking is susceptible to external disturbing fields, requires wired sensors, and needs a large tracking base placed near to the interventional area. Optical tracking requires a clear line of sight from the overhead tracking base and large reflective instrument attachments. Nonetheless, they are widely commercially available as options to premium ultrasound systems (GE XDclear, Philips PercuNav, Siemens eSie Fusion, …), or as stand-alone equipment (Brainlab navigation, Medtronic StealthStation, …). Both technologies enjoy a modicum of commercial success, but are widely reviled by clinicians for all but the most complex interventions due to (combinations of) equipment footprint, setup overhead, system complexity, training requirements, availability, and disposables cost. Clinical consequences of this situation include limited or overly complicated use of state-of-the-art techniques, unnecessary complications [1], and excessive imaging and radiation exposure for patients and operators.

We report on the development and clinical use of real-time vision information in aiding image-guided percutaneous targeting procedures such as biopsy or ablative therapies. The specific focus was initially on instrument guidance for ultrasound-based procedures, with a long-term vision to create a family of general-purpose software modules that can be configured to work with a wide variety of medical imaging modalities and clinical workflows. Among the key technical capabilities developed are camera-based instrument tracking and reliable, accurate, dynamic registration of ultrasound

(US) imagery to static 3D image data, to permit real-time guidance from pre-operative imaging. The guidance hardware is based on an "Optical Head" attached rigidly to standard ultrasound probes, providing a stereo view of the intervention area to the CORE navigation computer (Figure 1). There, computer vision algorithms track pertinent objects, such as instruments and patient markers. Using this information, all modalities can be rendered as matched ("fused") images, with additional instrument guidance overlaid.

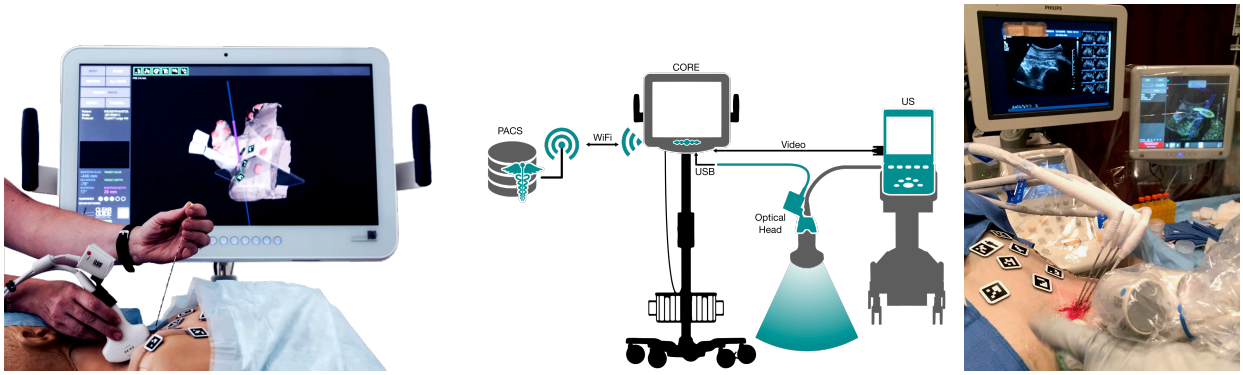

Figure 1: Clear Guide SCENERGY with handheld SuperPROBE; Clear Guide ONE/SCENERGY system schematic; US/CT fusion image guidance for hepatic cryoablation intervention using Clear Guide SCENERGY

2. METHODS

2.1 Visual Tracking

The developed visual tracking approach is based on real-time analysis of **one or multiple video streams from calibrated cameras**, with each frame is searched for two types of objects: straight instruments [3] and visual markers (stick-on VisiMARKER and TipTAG fiducials, both with unique visual QR-code-like patterns ("April Tags") that allow their 6-DoF localization in camera coordinates [2]). To achieve this, we developed an "**Optical Head**" stereo-camera attachment that can be mounted directly onto handheld ultrasound probes. In its standard configuration, it observes an interventional volume of approx. 5–30 cm usable depth at +/-45º width in either in-plane or out-of-plane configurations (Figure 2).

Figure 2: Optical Head tracking hardware on standard ultrasound probes (left: in-plane, right: out-of-plane configuration)

2.2 Instrumentation

Within the cameras' tracking volume, **standard straight instruments** – such as core biopsy needles, RF/cryo/microwave ablation instruments, aspiration needles etc. of >4cm length and 14G–21G gauge – are continuously and instantly 3D-reconstructed for real time instrument tracking (in 4 degrees of freedom/DoF) [3]. **Proprietary PercepTIP needles** exhibit an additional pseudo-random binary sequence patterning (PRBS) [4] that allows 5-DoF extrapolation of the needle tip location from only short observed shaft segments [5]. We have furthermore introduced a marker-based instrument tracking approach – "**TipTAG instruments**" – for full 6-DoF tracking with minimal setup and calibration, using a variation of the clinical-grade Clear Guide VisiMARKER patient fiducials that are attached freely onto interventional instruments. Both PercepTIP and TipTAG support multiple, arbitrary instruments, and are in regular

clinical use. Instrument tracking is thus compatible with nearly any pre-existing tool (Figure 3, left), and does not encumber the tools, unlike other tracking technologies that limit tool choice, range of approach directions or lateral motions, number of simultaneously trackable tools (Figure 3, right), or that require the attachment of proprietary, expensive, and tool-specific sensors.

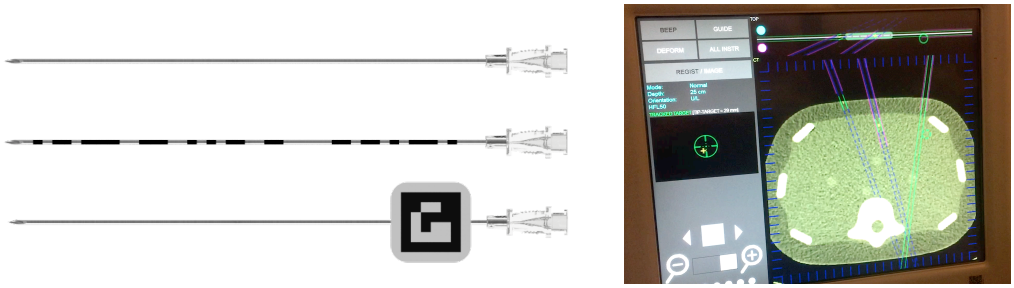

Figure 3: Supported instrument types: standard, PercepTIP, and TipTAG instruments (left top-to-bottom); Clear Guide SCENERGY tracking multiple instruments (right)

2.3 Patient Markers and Registration

We then extended the capabilities of the ultrasound-only Clear Guide ONE instrument guidance system to support **ultrasound+CT image fusion** on the same underlying hardware platform. The workflow includes a) pre-interventional imaging of the patient with skin-attached multi-modality Clear Guide VisiMARKER fiducials, b) import of this data into the system and automated pre-processing, c) "Visual Sweep" automatic registration, and d) instrument guidance on these fused ultrasound+CT images.

Automatic 3D volume loading and patient segmentation is a crucial step towards user-friendly image fusion. All support for manipulation of volumetric images is achieved via well-supported libraries (DCMTK and GDCM) that correctly sort and load DICOM volumes, which are imported via a network DICOM connection (e.g. from a PACS) or directly via USB. We use a statistical method (Otsu thresholding) [6] to separate the patient and marker volume from the background, and then create a high resolution polygonal surface model (Marching Cubes), both within ~10s. This method works regardless of the modality, since the patient body typically has intensity values clearly different from the background. The CT volume is later resliced in real time (>30fps) based on the probe pose estimation.

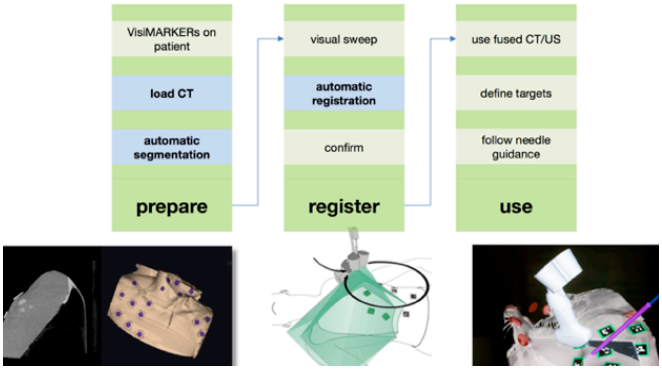

Figure 4: Clear Guide SCENERGY workflow

The Clear Guide VisiMARKER fiducials' relative poses are reconstructed from a "**visual sweep**" (moving the cameras relative to the marker-adorned object; Figure 4, center). This observation of skin-attached patient markers is used for automatic registration with CT, CBCT, or MRI volumes of the patient. "Early markers" are those present in pre-interventional 3D imaging, and are fully-automatically segmented by the system based on shape and intensity.

The skin marker observations are accumulated into a **3D marker mesh**, and the mesh is then **rigidly registered** onto the previously segmented markers (via ICP; Figure 5). However, this may fail – i.e. result in shifted imaging – if the patient has moved during CT imaging or did not follow breath-hold instructions, if the skin markers have shifted between imaging and intervention, if internal organs have moved, if the patient pose during intervention is not similar enough to the one during imaging, or if the operator has performed the visual sweep too fast or has not observed enough markers.

In these cases, the system allows for intra-operative **manual registration refinement** by unlocking the spatial relationship temporarily and align the images correctly (Figure 6).

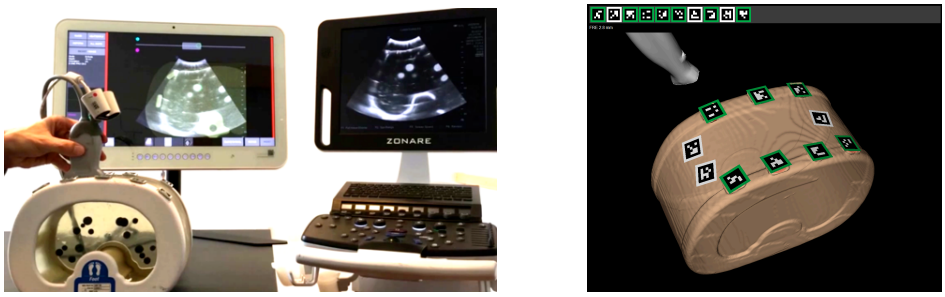

Figure 5: Performing a Visual Sweep registration on a multi-modality phantom with attached VisiMARKERs (left); registration between segmented and camera-observed markers (shown on auto-segmented phantom surface, right)

The registration method between observed (camera) and segmented (3D imaging) markers is robust against "**missing markers**" (i.e. which have disappeared or shifted substantially since imaging). Conversely, additional "**late markers**" can be attached to areas that are relevant to an intervention, but that did not receive sufficient numbers of "early markers" to ensure robust tracking there [7]. These added markers are integrated into the observed marker mesh, but since they do not match to pre-existing segmented markers in a low-FRE registration, they are excluded from contributing to the registration process. While a minimum of three observed early markers is needed for initial automatic registration, late markers do not improve registration, but make tracking more robust against occlusions and deformations. Interventional tracking can be done from a single observed marker, although more markers increase robustness.

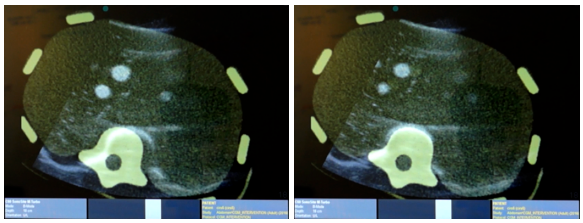

Figure 6: Ultrasound (gray) and CT (yellow) fusion alignment before and after Manual Refinement

The same visual localization of one or more markers by the Optical Head during the intervention also allows real-time tracking of the ultrasound probe pose relative to the patient. The system provides guidance to the operator about where to put the probe to achieve correct and immediate target visualization in the US image, and shows in real time a corresponding variable-orientation slice from the CT overlaid onto ultrasound, according to that probe pose. The system also presents dynamic targeting information on screen that indicates alignment directions for instrument placement. These visuals replace iterative targeting/imaging with direct target indication instead. All these pieces of information – ultrasound, 3D imaging, targets, and instruments – are co-displayed on screen in both a 3D schematic visualization and a 2D fused guidance view (Figure 7).

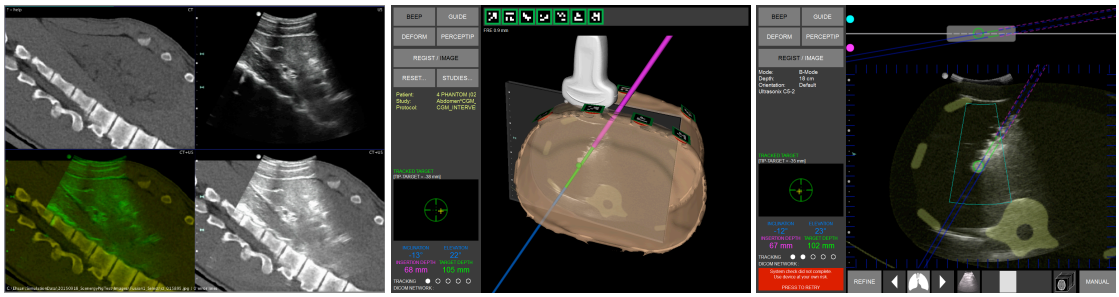

Figure 7: Fusion images from pig trial, with virtual CT, ultrasound and fusion with two different color codings (left); SCENERGY user interface: intra-procedural registration and schematic guidance view (center); fused US+CT view with overlaid tracked instrument (right)

2.4 Tissue Deformation Modelling

When using a modality such as CT with well-characterized intensity-to-tissue relationships, real-time deformation estimation [8] allows the modelling of tissue displacements from the application of external objects – here, the deformation from the ultrasound probe's pressure on the skin. A non-linear mass-spring-damper deformation model is automatically derived from the input image volume. The real-time probe pose, tracked by the stereo cameras via the skin markers, and a realistic 3D probe model allow to compute the estimated collision volume between probe and patient, which in turn constitutes the boundary condition for the real-time deformation estimation, which is propagated into the simulated tissue (Figure 8). As a result, anatomical structures and defined targets deflect under external forces, and experimental validation showed a 70%– 80% displacement recovery rate. Naturally, the largest displacements occur close to the surface. To our knowledge, no other clinically available navigation system includes this functionality.

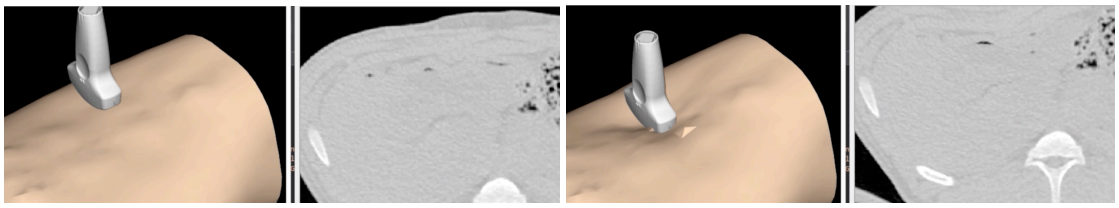

Figure 8: Real-time deformation modeling based on CT data, probe pose, and probe model (segmented skin and US probe; left: before probe contact, right: under probe contact)

2.5 Research Systems

All tracking can be performed from one, two, or more camera streams. The Optical Head contains two robustly calibrated miniature RGB+IR cameras and additional fallback infrared illumination. These track VisiMARKERs and TipTAGs as well as needle instruments. Monocular setups can do the same, although they cannot reliably track arbitrary instruments (while TipTAG instruments are possible). An example, non-clinical variation on this technology is the ZOOMLANDER, a software-only implementation running on a handheld Microsoft Surface Pro 3 tablet without the need for an external-camera Optical Head. It uses the same software components and workflows for 3D-image registration and TipTAG instrument tracking as the clinical systems, in spite of the differences in hardware. Current versions support 3D volume visualization as augmented-reality views with overlaid internal anatomy and targets, or DRRs (digitally reconstructed radiographs) derived from CT or CBCT data (Figure 9) at real-time frame rates.

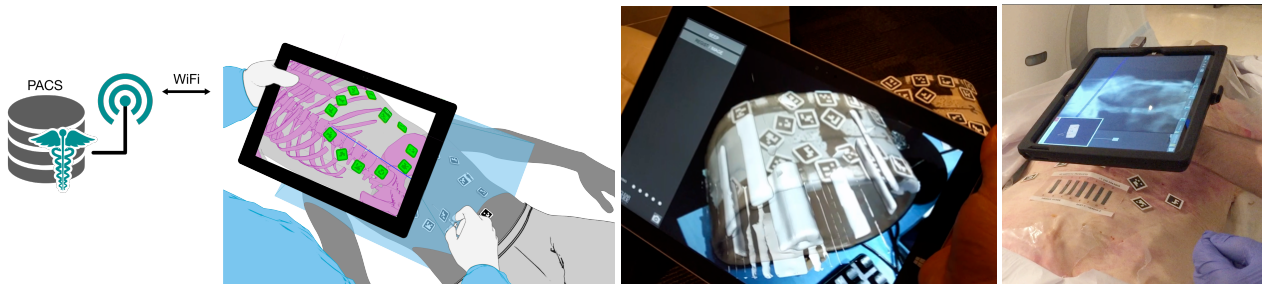

Figure 9: Handheld tablet PC guidance system (left); augmented reality video with internal-anatomy overlay and dynamic DRR navigation views based on monocular marker tracking of CT volume in phantom and cadaver trials (center and right)

3. RESULTS

In support of regulatory clearance, **in-vivo animal studies** on pigs were performed to collect accuracy measurements [7]. Specifically, these animal tests demonstrated the robustness and reliability of automatic segmentation and registration algorithms – which is of particular concern, as this aspect is novel compared to other technologies. Results showed RMS error of 0.58 ± 0.43mm and fiducial registration error (FRE) of 2.31 ± 0.94mm for automated marker segmentation and Visual Sweep registration, respectively, with 100% detection accuracy for segmentation.

Additionally, system-level errors (i.e. including tissue shift and total tracking error) were measured across phantoms, in-vivo animals, and patients. Tissue Registration Error (TRE) and Systematic Error (instrument tracking relative to fused modalities) were 3.75 ± 1.63mm and 3.99 ± 1.43mm, respectively [7].

The Clear Guide ONE and SCENERGY products [9] have been used clinically for a wide range of **ultrasound- and fusion-based percutaneous procedures** (incl. liver biopsy, liver ablation, renal biopsy, renal ablation, PCNL, pancreatic mass biopsy, chest mass biopsy, midline catheterization, A-line catheterization, nerve blocks, peripheral nerve blocks, fluid drainage, chest fluid drainage), and across private, public, and military institutions, both in the U.S. and abroad. Both systems have received FDA 510(k), CE Mark, and Health Canada clearances for clinical use (ONE with ultrasound, SCENERGY with ultrasound+CT).

Clinical benefits include an increase of computer-assisted image guidance (CAIG) availability for any given procedure, a wider applications field, and the reduction of geographic disparities (with the described visual technology being smaller, lower-cost, and simpler to use than current solutions) and skills disparities (e.g. accuracy improvement from 19.7mm to 5.1mm for non-expert operators placing instruments without vs. with US/CT fusion in phantoms [10]). Specifically, supplementation of regular ultrasound guidance with CAIG enhances procedural efficacy and decreases risk of damage to adjacent tissue (e.g. 79% procedure time reduction and 64% needle passes reduction [11]; successful 1st-attempt cannulation 89% with vs. 43% without CAIG in in-vivo pig study, with 95% vs. 78% final success rate, respectively [12]).

4. DISCUSSION

4.1 Conclusions

We developed a novel camera-based tracking technology with flexible miniature monocular-, stereo-, or multi-camera optics, which can be used across a wide range of imaging modalities. The technology supports instant and highly flexible automatic registration, which is commonly the make-or-break aspect of similar approaches. Multi-modality stick-on markers can be applied to instruments or the patient, providing robust 6-DoF tracking. Instruments with a proprietary pattern allow for 5-DoF tracking, and generic needle-like instruments are supported with 4-DoF.

We described a range of systems implemented on this basis: Clear Guide ONE for ultrasound-based intervention guidance with instrument and patient tracking, SCENERGY for CT/CBCT/MRI registration with fusion imaging, and ZOOMLANDER for tablet-based augmented-reality visualization (DRR or segmented structures). Both ONE and SCENERGY are commercially available and presently in clinical use.

4.2 Future Work

The described technology – the hardware and software modules – has matured into a clinical-grade platform for interventional multi-modality guidance. It is based on visual recognition and tracking, an area that is currently experiencing explosive growth in the context of artificial intelligence and autonomous computing. It is establishing itself as the only novel technology besides the incumbent electromagnetic and infrared-optical trackers. Clear Guide Medical has developed a small family of clinical, cleared navigation systems on top of this tracking approach.

An immediate next step is extending the range of medical imaging modalities available for clinical use to include MRI, CBCT, and PET imaging, based on clinical feedback. We are also pursuing shrinking the hardware footprint to address other clinical specialties, such as emergency medicine and pediatrics, where there is less scope for deliberate use of auxiliary devices.

REFERENCES

[1] Cavanna, L. et al., “Ultrasound guidance reduces pneumothorax rate and improves safety of thoracentesis in malignant pleural effusion: report on 445 consecutive patients with advanced cancer”, World Journal of Surgical Oncology, 12:139 (2014)

[2] Olson, E., “AprilTag: A robust and flexible visual fiducial system”, Proc IEEE ICRA. pp. 3400–3407 (2011)

[3] Stolka, P. J., and Boctor, E. M., “Navigation with local sensors in handheld 3D ultrasound: initial in-vivo experience”, Proc. SPIE Medical Imaging (2011)

[4] Petriu, E. M., “Absolute position measurement using pseudo-random binary encoding”, IEEE-IM Vol. 1 Issue 3 (1987)

[5] Stolka, P. J., Foroughi, P., Rendina, M., Weiss, C., Hager, G., Boctor, E., “Needle Guidance using Handheld Stereo Vision and Projection for Ultrasound-based Interventions”, Proc. MICCAI pp. 684–691 (2014)

[6] Otsu, N., “A Threshold Selection Method from Gray-Level Histograms”, IEEE Transactions on Systems, Man, and Cybernetics 9(1), 62–66 (1979)
[7] Basafa, E., et al., “Fast, Intuitive, Vision-Based: Performance Metrics for Visual Registration, Instrument Guidance, and Image Fusion”, Proc. MICCAI (2016)
[8] 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)
[9] http://www.clearguidemedical.com/models-overview
[10] Thai, J. N., Seifabadi, R., Xu, S., Stolka, P. J., Basafa, E., Hossbach, M., Levy, E. B., Wood, B. J., “Novel US-CT Fusion with Stereo Cameras Mounted on Ultrasound: Phantom and Early Clinical Experience” (submitted)
[11] Pechman D., et al: “Computer-Assisted Instrument Guidance: Enhanced Procedural Efficacy and Safety”, Critical Care Medicine, Volume 44 - Issue 12 (2016)
[12] Karaliou, V., et al., “Evaluation of the Clear Guide ONE System vs. Conventional Ultrasound-Guided Vascular Cannulation in Swine”, Proc. ELSO (2016)