

# Integrating Diagnostic *B*-Mode Ultrasonography Into CT-Based Radiation Treatment Planning

Wolfgang Wein\*, Barbara Röper, and Nassir Navab

**Abstract**—This paper presents methods and a clinical procedure for integrating *B*-mode ultrasound images tagged with position information with a planning computed tomography (CT) scan for radiotherapy. A workflow is described that allows the integration of these modalities into the clinic. A surface mapping approach provides a preregistration of the ultrasound image borders onto the patient's skin. Successively, a set of individual ultrasound images from a freehand sweep is chosen by the physician. These images are automatically registered with the planning CT scan using novel intensity-based methods. We put a particular focus on deriving an appropriate similarity measure based on the physical properties and artifacts of ultrasound. A combination of a weighted mutual information term, edge correlation, clamping to the skin surface, and occlusion detection is able to assess the alignment of structures in ultrasound images and information reconstructed from the CT data. We demonstrate the practicality of our methods on five patients with head and neck tumors and cervical lymph node metastases and provide a detailed report on the conducted experiments, including the setup, calibration, acquisition, and verification of our algorithms. The mean target registration error on nine data sets is 3.9 mm. Thus, the additional information about intranodal architecture and fulfillment of malignancy criteria derived from a high-resolution ultrasonography of lymph nodes can be localized and visualized in the CT scan coordinate space and is made available for further radiation treatment planning.

**Index Terms**—Computed tomography (CT), fusion, radiation therapy, registration, ultrasound.

## I. INTRODUCTION

### A. Motivation

CONVENTIONAL *B*-mode ultrasound imaging produces cross-sections of human anatomy of limited size, based on the acoustic properties of tissue. It is a cost-effective and flexible imaging modality, and therefore widely used in almost all areas of medicine and healthcare. Modalities like X-ray computed tomography (CT) and magnetic resonance imaging (MRI) acquire a stack of cross-sectional slices, which yield a 3-D data set of the desired anatomy. A spatial alignment, i.e., *registration*, of the images taken from these devices provides access to inher-

ently complementary information. Hence, it can be beneficial for diagnosis, planning, and treatment of various medical procedures.

In the particular application of radiation treatment planning for inoperable head and neck cancer, the identification of metastatic neck lymph nodes is mandatory for the correct target volume delineation. This can be achieved with a reported accuracy of 80%–95% using high-frequency ultrasound [1]–[3]. In direct comparison with ultrasonography, diagnostic CT was equally predictive in revealing lymph node size but achieved lower performance in depicting internal nodal architecture, leading to a lower sensitivity and specificity than ultrasonography [4]. In planning CTs for radiotherapy, contrast medium is usually omitted, and consequently their diagnostic properties are particularly poor.

Omitting the CT and performing the treatment planning and execution solely based on ultrasound has its place only in brachytherapy (e.g., in prostate cancer), where a small treatment volume can be visualized together with the radiation source(s) by ultrasound without artifacts from bone or air, the dose distribution being dependent on distance rather than on tissue properties. However, in the context of external-beam radiotherapy delivered from different angles, axial sections of the whole body are mandatory and cannot be provided by ultrasound.

Thus, for external-beam radiotherapy, CT remains the base imaging modality of choice for treatment planning and simulation, as it provides a coordinate systems with stable geometric fidelity and the necessary electron density information for the computation of the accurate dose distribution within the CT anatomy. Consequently, the wish to integrate the diagnostic properties of other imaging modalities into the planning process has usually been met by registration with CT, based on either external markers or anatomy information (e.g., registration of MRI and CT), or with dedicated combined systems like positron emission tomography (PET)-CT. Similar solutions of registration between freehand ultrasonography and planning CT seem to be a valuable goal, as the additional ultrasound (US) information may enable the radiation oncologist to refine the target volume definition and individualize treatment planning. For example, in studies on integrating PET into the treatment planning process of brain or lung tumors significant changes in gross tumor volume (GTV), delineation could be demonstrated (review by [5]). It is quite possible to have a similar impact of ultrasound information for GTV definition in head and neck cancer.

### B. Related Work

In the context of radiotherapy, tracked ultrasound has been used mainly to quantify and reduce daily setup errors so far.

Manuscript received December 3, 2006; revised February 6, 2007. This research was supported by Siemens Corporate Research, Princeton, NJ, under an academic grant. *Asterisk indicates corresponding author.*

\*W. Wein is with Computer Aided Medical Procedures, TU Munich, Munich, Germany, and Siemens Corporate Research, Inc., Princeton, NJ 08540 USA (e-mail: wein@cs.tum.edu).

B. Röper is with the Clinic and Policlinic of Radiation Oncology, Klinikum Rechts der Isar, TU Munich, Munich, Germany (e-mail: barbara.roeper@lrz.tum.de).

N. Navab is with Computer Aided Medical Procedures, TU Munich, Munich, Germany (e-mail: navab@cs.tum.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2007.895483

The commercial NOMOS *B*-mode acquisition and targeting system<sup>1</sup> allows one to integrate the CT planning contours with tracked ultrasound acquisitions prior to each treatment fraction. A description of the system, as well as a comparative study of prostate localization based on BAT and CT, can be found in [6]. Though the largest experience has been gathered with prostate cancer [7], some data on other sites such as the upper abdomen are available as well [8]. A similar commercial solution is SonArray [7], [9] by Varian Medical Systems.<sup>2</sup> The RESTITU system<sup>3</sup> uses two calibrated (3-D) ultrasound devices, in the CT and treatment room, respectively [10]. Hence, the problem of aligning planning information within the coordinate system of the treatment room is reduced to a monomodal registration of the fractional 3-D US acquisition with the reference 3-D US scan taken in the CT room.

In the field of automatic image registration, much research has been carried out using features extracted from the ultrasound images in order to align them with corresponding structures in other modalities. In the context of thermal ablation for liver metastases, the most significant structures in the ultrasound images are the liver surface and major vessels inside the liver. In [11], those structures are segmented in an MRI scan, and points on the surface of the same structures are manually picked in the ultrasound images. A method based on the iterative closest point (ICP) algorithm is then used to compute the transformation between the two modalities.

Bone structures can be well identified using ultrasonic imaging, as they produce a strong (though also specular and hence position-dependent) reflection and full occlusion behind the reflection. Thus it is feasible to use such structures in a more automated manner for registration [12], [13]. In [12], a modified ICP algorithm is used in order to register the bone surface extracted from CT with points in the ultrasound images likely to reflect the bone surface. This can be considered as a combined registration and segmentation approach.

The use of color Doppler ultrasound aids the automatic feature extraction of vessel structures, which can in turn be used for feature-based registration [14]–[16]. In [14], liver vessel features are extracted from both preoperative MRI/CT data and intraoperative 3-D power-Doppler ultrasound data. The registration is initiated with a few manually selected landmarks, and a rigid transformation is then estimated using a modified ICP method, which takes the vessel topology into account. Eventually, the registration is further improved using a transformation grid modeled with deformable *B*-splines. Slomka *et al.* evaluated rigid registration of power-Doppler 3-D ultrasound with MRI scans of the carotid artery [16].

Pure intensity-based registration with other modalities has been performed mainly for 3-D ultrasonic data. Roche *et al.* [17] use an adapted correlation ratio similarity measure in order to register the ultrasonic data simultaneously to both the intensity and the gradient information of an MRI scan. In [18], the Kullback–Leibler distance is minimized for the registration of 3-D US and MRI. A registration involving an automatic mapping of MRI and ultrasound data to “vessel probability values” and suc-

cessive registration of this information is proposed in [19]. This work is extended in [20] in order to incorporate a deformable model generated from MRI acquisitions of the liver at different stages in the breathing cycle. In [21], a similar approach uses bone probability values generated from the CT intensity and gradient in order to perform a rigid registration of CT with tracked ultrasound. Single tracked ultrasound slices can hence be registered to the preoperative information in a nonrigid manner. In [22], the image phase is proposed as underlying data for a mutual information based registration. It includes experiments on 2-D data with brain MRI and simulated ultrasound images, where early results on deformable registration are presented.

To our knowledge the published data on true image-based registration of CT with ultrasound are sparse. A mutual information-based registration of CT with tracked 3-D ultrasound for prostate localization is presented in [23]. It is aided by the manual segmentation of prostate and bladder and the application of preprocessing steps on both modalities. In [24], in the CT data of a kidney, the intensity values are enhanced with strong edges from the gradient. Successively, an automatic registration with freehand 3-D ultrasound is performed.

We described a part of our method, as well as initial experiments with magnetically tracked ultrasound on three patients, in [25]. We opt for an intensity-based approach; however, we use a combination of different information and physical properties of both modalities to introduce a more stable measure for automatic registration.

## II. PROPOSED MEDICAL WORKFLOW

For the described application of radiotherapy planning, a workflow (Steps 1–8) is needed to integrate the diagnostic ultrasound with the current treatment planning processes. The planning CT is usually performed with the patient fixed onto a head pad with a thermoplastic mask individually molded. This minimizes the spatial deviation between the anatomy depicted in the CT scan and the daily radiation treatment delivery.

- 1) Just before ultrasonography, the patient is immobilized likewise to ensure the same reclination of the head. Then, the mask is carefully removed while the patient is told to keep his position for the following examination.
- 2) The diagnostic ultrasound is performed by slow freehand sweeps of the probe in lateral and craniocaudal directions along the patient’s neck and chin with special regards to the lymph node regions (i.e., submental, submandibular, digastric, jugular chain, spinal accessory, and occipital nodes). In our scenario, a position sensor is attached to the ultrasound probe. All diagnostic sweeps are recorded as a combination of videos containing the actual ultrasound images, alongside the tracker readings. An offline calibration step allows us to place all recorded ultrasound images in the spatial context of the tracking world coordinate system (Fig. 1). Hence, 3-D ultrasound information is obtained with full coverage of the patient’s neck and chin.
- 3) Our system now applies an automatic preregistration step by aligning the lines denoting the top of the numerous ultrasound images in 3-D space with the skin surface extracted from the CT data set; see Section III for details.

<sup>1</sup>BAT; [www.nomos.com](http://www.nomos.com)

<sup>2</sup>[www.varian.com](http://www.varian.com).

<sup>3</sup>Resonant Medical; [www.resonantmedical.com](http://www.resonantmedical.com).

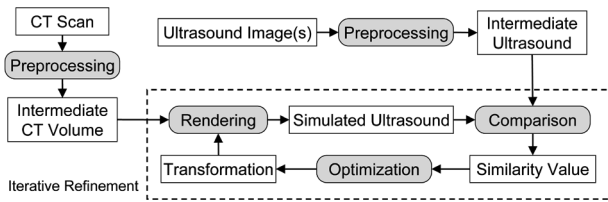


Fig. 1. Workflow for image-based ultrasound-CT registration.

- 4) From all sweeps with relevant lymph nodes, the physician picks some particular frames containing structures that are identifiable in both modalities, like the carotid artery (maybe with individual calcifications), internal jugular vein (preferably with little or no compression), certain muscles, or thyroidal/salivary gland tissue.
- 5) These frames serve for reviewing the first results from global skin surface registration.
- 6) In case of excessive deviation, manual drag and drop of the stack of ultrasound pictures into a closer range of corresponding CT anatomy is performed. Any frame can be selected and its relative location and orientation modified, while the spatial relation to all other frames (given by the position sensor) is maintained. This allows a convenient navigation, as in-plane and out-of-plane transformation parameters can be manipulated separately. Either a 3-D display of the data or side-by-side visualization is used, as shown in Fig. 6(b) and (d).
- 7) This allows an automatic image-based registration to be performed on these images successively, as will be described in Section IV.
- 8) The results from image-based registration are evaluated and may serve for target volume delineation. In particular, our method overlays pathological lymph node areas, which need to be identified and irradiated with a certain dose, but are barely visible in and extractable from the CT data alone.

An adapted treatment planning software (which is not the focus of this paper but should be addressed in the future) should allow the physician to scroll through the registered ultrasound frames, displayed alongside corresponding multiplanar reconstructions interpolated from the CT data, and vice versa through the original axial CT slices overlaid with the registered 3-D US data in order to account for all additional information during target volume delineation.

In the following sections, we present algorithmic details of the proposed registration steps and describe the setup and results of experiments used for validation.

### III. GLOBAL SKIN SURFACE REGISTRATION

The set of all ultrasound images recorded from a patient including the submental and submandibular regions as well as both sides of the neck (Robbins level I–VI) can be used to derive an approximate registration to the CT scan. In each ultrasound image, the horizontal line denoting the top of the image roughly lies on the skin surface. As the images are continuously recorded and tagged with position information, all frames acquired from a patient (up to several thousand images) yield a complete coverage of the neck surface in the tracking coordinate

system. Thus, a surface-based registration with the 3-D skin surface segmented from CT will provide a fairly good global initial registration.

The CT scan is automatically segmented with a region-growing method [26], using a seed point in the surrounding air and inverting the result afterwards. Successively, a border detection algorithm is run, and a Chamfer distance map transformation [27] is applied to the surface. It results in a distance volume where the voxel values reflect the closest approximate Euclidean distance to the surface. Ten equal-spaced points on the top of each ultrasound image form the 3-D point set. The global rigid transformation is initialized with a  $90^\circ$  rotation that brings the tracking world coordinate system in the same orientation than the CT volume and a translation that consists of the subtracted mean of all points (hence the point set is centered in the CT volume). For each estimate of the global rigid transformation, the distance of the transformed points to the CT surface can be efficiently looked up in the distance volume. The transformation parameters are iteratively refined until the error converges. Fig. 2 shows the recorded ultrasound images in a 3-D visualization, as well as the 3-D point set derived from them, overlaid onto a volume rendering of the CT scan.

### IV. INTENSITY-BASED REGISTRATION

#### A. General Considerations

For intensity-based registration, the quality of anatomical alignment between the two modalities is a function of the image gray values. This prohibits closed-form solutions that can often be defined for feature-based methods. Such registration methods can be considered as a search for an optimal image alignment, which is an iterative process, as outlined in Fig. 1. A tomographic data set will usually enclose the respective ultrasound data, be it 2-D or 3-D. Hence, data from the CT volume, according to the spatial extent of the ultrasound data (a slice or cone with specific thickness, etc.), can be extracted at a location where the anatomy contained in the ultrasound images is presumed. This should establish a representation that can be compared with the respective ultrasound data. In the ideal case, it would be a realistic reconstruction of ultrasound image intensities, i.e., a *simulation* of ultrasound from CT. The comparison usually yields a scalar value denoting the quality of anatomical alignment at the current spatial configuration. A nonlinear optimization method then changes this configuration (i.e., typically the transformation parameters for simulation) iteratively, until an optimal alignment is found.

Any computation step that brings the modalities closer together in terms of comparability is appreciated. If an intermediate representation is established for comparison, it might be a simulation of ultrasound from CT, but just as well a simulation of CT from ultrasound, as well as anything in between. An ultrasound image depicts the strength of echoes whose magnitude increases at boundaries between different types of tissue. This can be related to CT data as the gradient magnitude of the X-ray attenuation (from a simplified point of view, see, e.g., [28, Section 2.2]). Hence, the derivative of the CT voxel intensity yields information that can be compared to ultrasound. On the other

hand, numerical integration of the edge information depicted in ultrasound unfortunately does not result in values that can be put in relation to the CT data, as we are missing directional information.

As a consequence, the resemblance of ultrasound images to CT can only be increased by reducing imaging artifacts; however, a simulation of CT intensities from ultrasound is not possible solely based on the image data (but might be feasible with learning-based approaches and methods for incorporating prior information).

### B. Simulation From CT

A realistic simulation of ultrasound from CT is not possible for several reasons. Both the acoustic impedance and speed of sound in a given tissue can only be roughly estimated, as the CT Hounsfield units represent the attenuation of a polychromatic X-ray source within a voxel and therefore only approximately can be related to tissue density. Regardless of tissue interfaces, the echogeneity within a tissue depends on the microscopic inhomogeneities (i.e., in muscle fibers, fatty tissue, etc.), which is not revealed by CT.

Instead of a realistic simulation of ultrasound, we hence need an intelligent and efficient intermediate representation of the CT data at arbitrary cut-planes, such that an iterative registration can be performed in an acceptable time. These slices have multiple components containing intensity, gradient, and edge information, which are used to derive various parts of a similarity metric, so that the correspondence of anatomy contained therein with structures in 2-D *B*-mode ultrasound images can be determined.

The use of gradient and edge information is justified, as medical ultrasound mainly depicts tissue interfaces, caused by ultrasonic reflections at a boundary of different acoustic impedances. Those will also produce different intensities in CT and hence a visible edge. The strength of such a boundary gradient can, however, not be put into a mathematical relation to the strength of the respective ultrasound reflection. On the other hand, the original CT attenuation values are of importance as well, as they reveal different types of human soft tissue, which in turn cause different effects in ultrasound, such as attenuation, speckle characteristics, reflection, etc.

In our approach, first the 3-D gradient vector values are computed from the CT data set by convolution with a Sobel filter cube. They are stored in a four-channel volume together with the original voxel intensity (i.e., channels 1–3 contain the gradient vector, channel 4 the CT Hounsfield intensity). The interpolated slices contain four channels as well. For each pixel, the four-vector is computed from the volume using trilinear interpolation. In the first channel of the slice, the original CT intensity is stored. The 3-D gradient vector is scalar multiplied with each of the vectors indicating the horizontal and vertical slice plane directions, respectively. The resulting values, corresponding to the 2-D gradient of the CT intensity within the slice, are stored in the second and third channel.

The 2-D slice gradient values are then used to perform Canny edge-detection on the slice data, and the result is stored in the fourth channel. The most time-consuming steps within the Canny algorithm are the computation of the 2-D gradients, as well as filtering them with a sufficiently large Gaussian kernel

for smoothing. As we compute the 2-D gradients directly from the precomputed 3-D gradient values, we do not need to run a 2-D filtering for gradient computation. In addition, those gradients are fairly smooth, as they originate from a 3-D Sobel filter using a 27-neighborhood. This makes further Gaussian filtering unnecessary. The two remaining steps for the Canny algorithm, nonmaxima suppression, and hysteresis thresholding can each be performed in one traversal of the 2-D slice. The horizontal gradient is weighted with a user-defined factor between zero and one, as the ultrasound data tend to show mainly edges along the lateral direction, parallel to the transducer array.

Thus we are able to construct intermediate slices from the CT data at estimated transformations of the US scan plane in very little time (1.1 ms for a 128<sup>2</sup> pixel slice, interpolated from a 512<sup>2</sup> · 100 CT/gradient volume, on an AMD Opteron 2.2 GHz machine). The individual components of the slice pixels are then used to compute a similarity metric with the ultrasound data.

### C. Occlusion Handling

If an ultrasonic pulse hits bony structures, all image intensities in the ultrasound image further along the specific ray are occluded, and mainly determined by noise. Therefore, all ultrasound intensity values on a ray below such an occlusion should be disregarded in the registration method. In our implementation, we scan the US image from bottom to top, updating the variances for all ultrasonic pulse rays. Where they exceed a threshold  $\sigma_y^2$  (which is easily determined in the user interface), the first pixel to be considered is defined. Thus, our region of interest (ROI)  $\Omega$  is expressed by the following equations:

$$\Omega = \{(x, y) \mid (y < y_{\text{top}}) \wedge (y \geq b(x))\} \quad (1)$$

$$b(x) = \min \left\{ y : \frac{1}{y} \sum_{i=0}^{y-1} U(x, i)^2 - \left( \frac{1}{y} \sum_{i=0}^{y-1} U(x, i) \right)^2 < \sigma_y^2 \right\} \quad (2)$$

where  $x$  is the lateral (increasing to the right) and  $y$  the axial (increasing upward) pixel index of an ultrasound image  $U$ . By applying a median filter on the bottom function  $b(x)$ , discontinuities are removed before defining the ROI. In addition, we discard all pixels that are located above  $y_{\text{top}} = (9/10)\text{size}_y$ , as we observed that the anatomy is highly compressed there due to the probe pressure on the patient's skin. This compressed region is very distinct from the remaining anatomical structures, its size (3.6 mm) being consistent on all data we obtained from patients [Fig. 6(a)]. This ROI definition is similar to those used in [19] and [24]. It can be adapted to curved-array transducers as well by accumulating (2) along the actual ultrasound pulse rays (which do not coincide with the image columns as in our case; hence some interpolation would be required). However, using a sector probe will markedly enhance all remaining problems relating to tissue compression and distortion, as there is no simple function available for correcting the tissue shift.

### D. Similarity Measure

Deriving a similarity measure for image-based registration of ultrasound with CT, based on tissue attenuation values and

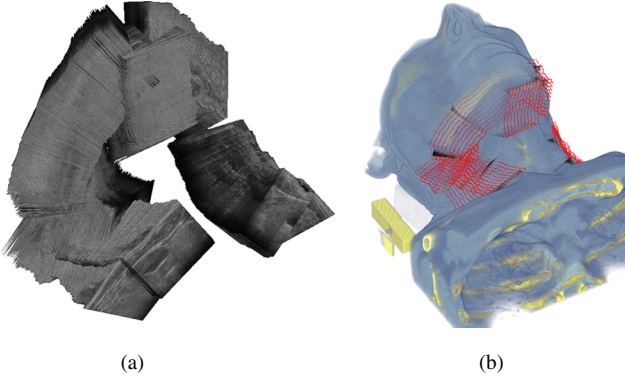


Fig. 2. (a) Spatial overview of all ultrasound images of Patient 3. (b) Extracted surface points (red) with volume rendering of CT and some individual ultrasound slices.

their edges reconstructed from CT, is a demanding issue. Based on both the physical properties of the imaging modalities, as well as the visible appearance of their images, we introduce several components for a similarity measure, which can in turn be weighted to define a global cost function with respect to the transformation parameters.

1) *Skin Surface Clamping*: In the compressed fraction of the ultrasound image, the interpolation from CT is done with six times the vertical scaling [see Fig. 6(a) on top, the 6x factor has been manually optimized]. As a result, the interface between skin and air always has to be within that region, producing a large vertical gradient in the interpolated slice. When all vertical gradient pixels are summed to  $t$ , high and low thresholds  $t_h$ ,  $t_l$  can be defined in order to decide if the skin surface lies inside, outside, or close to the compressed region

$$f(t) = \begin{cases} 1, & \text{if } t > t_h \\ 0, & \text{if } t < t_l \\ (t - t_l)/(t_h - t_l), & \text{otherwise} \end{cases} \quad (3)$$

$$S_0 = 3f(t)^2 - 2f(t)^3.$$

A cubic polynomial is used instead of the linear one in order to avoid discontinuities. As a component in a cost function,  $S_0$  penalizes transformations that are physically impossible, as the patient's skin is always located at the top of the ultrasound images.  $t_h$  and  $t_l$  have been manually defined by displaying the value of  $t$  while gradually moving the transformation towards the patient, and visually assessing the skin surface alignment on sweeps from all patients.

Note that another option for evaluating the distance to the patient's skin would be to perform lookups in the distance volume that was used for the global preregistration in Section III. However, we prefer the described method, as we do not have to load another volume (or another intensity channel) into the computer's main memory, which is already largely occupied by the CT/Gradient volume. Furthermore, the above computations can be performed with little additional computational cost.

2) *Edge Alignment*: As we have detected the edges in the simulated images, we would like to derive a similarity estimate based on the distance to edge structures in the ultrasound images. The straightforward approach would be to 1) compute an

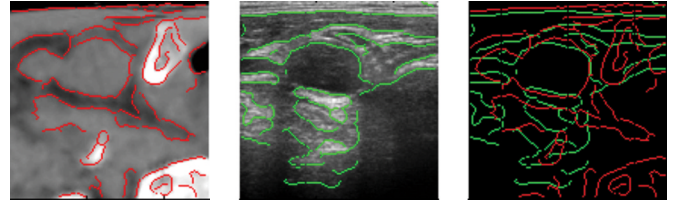


Fig. 3. Registered image pair and edge detection result (red = CT, green = US).

edge-detection for the ultrasound images, 2) compute a 2-D distance map for those edges, and 3) sum over the distance map values at the locations indicated by the edges of the simulated data. Steps 1) and 2) need to be performed once for each ultrasound slice, while 3) establishes a similarity metric and thus has to be computed for each simulated slice during pose estimation.

However, due to the different nature of CT and ultrasound data, the detected edges do not correspond in general, as shown by Fig. 3. We therefore propose to skip the ultrasound edge detection, using the original ultrasound intensity just as an indicator for edges instead.

Given a binary edge image, the distance of an image point  $\vec{x}$  to the edge structures  $Y = \{\vec{y}_i\}$  is defined as  $d(\vec{x}) = \min_i |\vec{x} - \vec{y}_i|$ . Instead of the Euclidean distance, we can also express the proximity to edges by using a Gaussian expression, which allows us to adjust the sensitivity of the cost function value with respect to the distances, using  $\sigma^2$

$$p(\vec{x}) = \max_i \exp -\frac{(\vec{x} - \vec{y}_i)^2}{\sigma^2}. \quad (4)$$

Taking into account that we do not have precise edge information, a proximity value can be defined as

$$p(\vec{x}) = \sum_i u_i \exp -\frac{(\vec{x} - \vec{y}_i)^2}{\sigma^2} \quad (5)$$

where  $u_i \in [0, \dots, 1]$  is the probability of the image pixel  $\vec{y}_i$ 's being an edge. Assuming that the ultrasound image intensity directly scales with the edge probability (i.e.,  $u_i \propto U(\vec{y}_i)$ ), a 2-D proximity function  $p(\vec{x})$  can be computed according to (5) by just convoluting the ultrasound image with a large Gaussian kernel. From this, we define a similarity measure component  $S_1 = (\overline{p_e} - \overline{p})/\sigma_p$ , where  $\overline{p}$  is the mean of all values in the proximity image,  $\overline{p_e}$  is the mean of just the pixels at locations where an edge is present in the simulated image, and  $\sigma_p$  is the standard deviation of the proximity image values.

3) *Statistical Correspondence*: In addition to tissue interfaces, different tissues by themselves cause different ultrasonic scattering characteristics, in particular characteristic speckle patterns and echogeneity. These are in turn reflected in the average ultrasound image intensities for a given tissue type. It is, therefore, applicable to assess the statistical dependence of the CT intensities, which classify the tissue according to the X-ray attenuation property, with the intensity in the ultrasound image. We use mutual information (the most commonly used statistical similarity measure; see [29] for a survey) on the CT and ultrasound intensities. The normalized mutual information (NMI) term uses the entropies of the combined and individual

images, which are computed with the Shannon entropy from probability distributions of the image intensities

$$\begin{aligned} \text{NMI}(U, C) &= 2 - 2H(U, C)/(H(U) + H(C)) \\ H(U) &= - \sum_j p_u(j) \log p_u(j) \\ H(C) &= - \sum_i p_c(i) \log p_c(i) \\ H(U, C) &= - \sum_i \sum_j p(i, j) \log p(i, j). \end{aligned}$$

Here,  $U$  denotes an ultrasound image and  $C$  the corresponding simulated image, i.e., the slice interpolation of CT attenuation values. The probability distributions can be estimated using histogram information from the images

$$p_u(i) = \frac{1}{n_\Omega} |\{(x, y) \in \Omega | U(x, y) = i\}| \quad (6)$$

$$p_c(j) = \frac{1}{n_\Omega} |\{(x, y) \in \Omega | C(x, y) = j\}| \quad (7)$$

$$p(i, j) = \frac{1}{n_\Omega} |\{(x, y) \in \Omega | U(x, y) = i \wedge C(x, y) = j\}|. \quad (8)$$

Here, we assume that each intensity value is mapped into one histogram bin and  $n_\Omega$  is the number of pixels in the ROI

$$n_\Omega = |\Omega| = \sum_{x=0}^{n_x-1} (y_{\text{top}} - b(x)). \quad (9)$$

An equivalent formulation for constructing the probability distribution from a histogram can be written using a binary count function  $c_u$

$$p_u(i) = \frac{1}{n_\Omega} \sum_{(x,y) \in \Omega} c_u(x, y, i) \quad (10)$$

$$c_u(x, y, i) = \begin{cases} 1, & \text{if } U(x, y) = i \\ 0, & \text{otherwise} \end{cases}. \quad (11)$$

Due to the various physical effects in ultrasound imaging, both the chance that an image intensity reflects the anatomy (due to refraction, reverberation, and other artifacts) and the signal-to-noise ratio decrease with the distance from the ultrasound transducer. Thus we would like to give more weight to image pixels that are closer to the probe, i.e., with higher  $y$  values. In our approach, we introduce an integer weighting for estimating the probability distribution

$$p'_u(i) = \frac{1}{n'_\Omega} \sum_{(x,y) \in \Omega} (y + c_0) c_u(x, y, i) \quad (12)$$

$$n'_\Omega = \sum_{x=0}^{n_x-1} \sum_{y=b(x)}^{y_{\text{top}}-1} (y + c_0). \quad (13)$$

Every intensity value is inserted  $y + c_0$  times into the histograms and the joint histogram.  $c_0$  is a shifting constant that affects the amount of weighting. We set  $c_0 = n_y$  ( $n_y$  being the number of image rows); hence the top of the image is weighted twice as

much as the bottom. For  $c_0 \rightarrow \infty$ , the original mutual information notation is obtained. Our weighted mutual information component NMI' of the similarity measure is assembled by inserting all used ultrasound slice images and the corresponding simulations into one histogram, as it increases the statistical significance of the derived entropy terms. See Appendix I for details on the weighting of mutual information.

4) *Cost Function*: The final similarity measure from a set of  $n$  ultrasound slices  $\{U_i\}$  and their CT simulations  $\{C_i\}$  is

$$cf = w_0 \frac{1}{n} \sum_{i=1}^n S_0(C_i) + w_1 \frac{1}{n} \sum_{i=1}^n S_1(U_i, C_i) + w_2 \text{NMI}'(\{U_i\}, \{C_i\}) \quad (14)$$

where  $w_0, w_1, w_2$  are fractional weights of the individual measure components adding to one ( $w_0 + w_1 + w_2 = 1$ ) and  $S_i$  are the measure components as defined in the previous sections.

### E. Registration

For automatic registration, a nonlinear optimization method maximizes the cost function  $cf$  iteratively with respect to the parameters of a rigid transformation (six degrees of freedom, translation, and Euler angles), which is initialized with zero and affects the current location of all slices. We used three optimization schemes: Simple Hill-Climbing (aka Best Neighbor Search), Powell–Brent, and Exhaustive Hill Climbing. The latter one evaluates all combinations of [forward, keep, backward] for all parameters, using the best result of all  $3^6 = 729$  evaluations as estimate for the next iteration. When the optimization terminates, the resulting relative transformation is multiplied onto the overall registration transformation.

## V. EXPERIMENTS

### A. Data Acquisition

We conducted a study on five patients. For the freehand ultrasound acquisitions, we decided to install both optical and magnetic tracking. However, for the evaluation of our algorithms, we used solely the optical tracking data. In the future, information from both tracking technologies will allow interesting comparative studies. All ultrasound examinations were performed with a GE Logiq 500-scanner and a 8.5-11MHz transducer with a 40 mm linear probe (LA39, GE Healthcare Technologies, Waukesha, WI). An optical tracking target, consisting of four infrared marker balls, was mounted on the handle of the ultrasound probe using cable ties and a layer of Varihesive bandage-aid between probe and target [see Fig. 4(a)]. We tested the tracking setup extensively to assure that the target does not hinder the physicians flexibility when scanning a patient, in particular that the optical target does not touch the patient's skin. On the other hand, the target should be recognized by the tracking cameras at all times.

The final setup contained four optical ARTtrack2 cameras<sup>4</sup> arranged behind the head of the patient [Fig. 4(c)]. A two-camera system would have been sufficient if mounted somewhat higher. As we had four cameras at our disposal, we used them to obtain a symmetric setup, so that the expected

<sup>4</sup>A.R.T. GmbH, Weilheim, Germany; www.ar-tracking.de



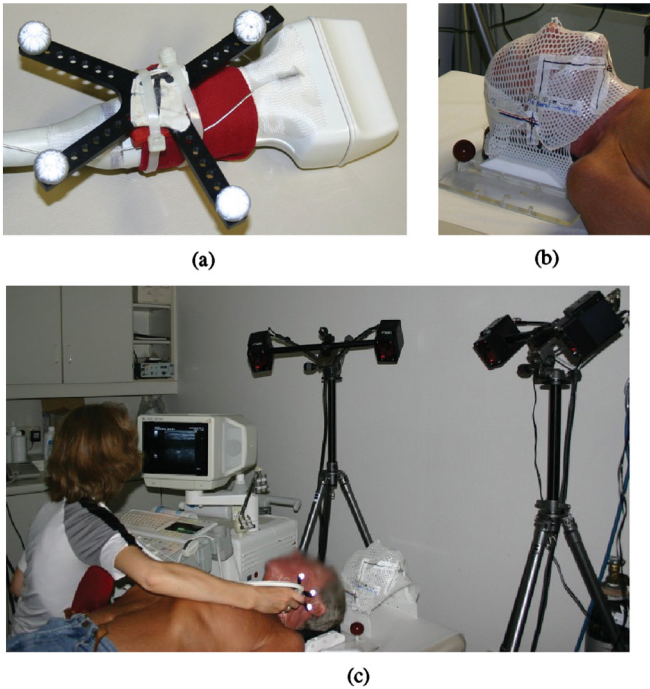


Fig. 4. Setup for 3-D freehand ultrasound acquisitions. (a) Ultrasound transducer with optical and magnetic tracking targets. (b) Positioning with mask. (c) Actual patient scanning.

tracking precision is comparable on both sides of the patient's neck. The ultrasound data were recorded using an IDS FALCON frame grabber card.<sup>5</sup> The ART tracking data were received over user datagram protocol network transmission from the separate PC running the ART DTrack software. We used the CAMPAR framework for medical augmented reality applications [30], which was developed by colleagues in our group, for the acquisition. We implemented the recording of sequences by storing the gray-valued video region containing the actual ultrasound image (size  $454 \times 454$ ) on disk, along with its timestamp and the corresponding  $4 \times 4$  tracking matrices. The resulting performance was the frame grabber's full speed of 25 frames/s; 5 MB/s are written to disk. On mouse clicks, the corresponding 2-D position in the ultrasound video was logged in an additional text file, along with the current tracker poses. We included an additional pose of a tracked tool, whose tip intersected the ultrasound image, for calibration purposes (see Section V-B). For the actual freehand ultrasound examinations, patients were positioned supine on the examination couch with the head pad supporting the neck and the individual thermoplastic immobilization mask covering the face, as shown in Fig. 4(b). Then the mask was removed, taking care not to move the head that was still stabilized by the preformed pad beneath, and the patient was told to keep this posture. From each patient, 11–16 ultrasound sequences, up to 45 s each, were recorded. It resulted in 5000–11 000 ultrasound images per patient.

Right after the examination, we took a CT scan of each patient in the radiology department. The device was a Siemens Sensation Cardiac 64, used with the standard neck scanning protocol and reconstruction with 0.6 mm slice thickness.

<sup>5</sup>IDS Imaging Development Systems GmbH, Obersulm, Germany; www.ids-imaging.com.

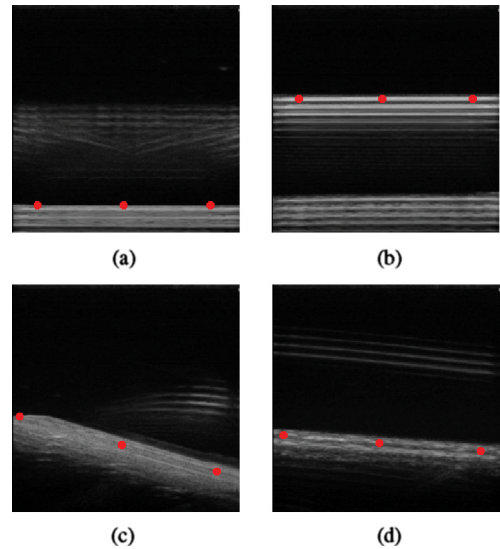


Fig. 5. Ultrasound images from the water bath calibration scan; red dots show points used for the floor reconstruction. (a) Straight, deep. (b) Straight, shallow. (c) Rotated probe. (d) Tilted probe.

## B. Ultrasound Calibration

In order to obtain the position and orientation of the ultrasound image plane with respect to the tracker's world coordinate system, a calibration is necessary to derive the transformation that relates the image plane to the tracking sensor coordinates. There has been a great amount of research on freehand ultrasound calibration methods in the last decade (see [31] for a detailed review). We adapted the single-wall calibration technique [32] to our experimental setting in order to yield an easy yet sufficiently accurate calibration, as described in the following.

For any point  $\vec{p} = (u, v, 0, 1)^T$  in the ultrasound imaging plane, the corresponding point  $\vec{p}_w$  in the tracker's world coordinate system is obtained by

$$\vec{p}_w = T_T T_C \vec{p}. \quad (15)$$

The calibration transformation  $T_C$  maps a point from the ultrasound coordinate system to the tracking target, while  $T_T$  in turn transforms it to the tracker's world coordinate system. The ultrasound probe is dispensed in a shallow water bath, and moved in a variety of locations and orientations. The floor then appears as a line in the ultrasound image, and the user clicks on positions of that line in the video image. One hundred to two hundred of those 2-D positions in the ultrasound image (two to three positions per still frame) are recorded alongside the tracking matrices in a file. Using a water-level device, we assured that both the floor of the water bath and the rim of the container are exactly parallel to the water surface. By means of the ART DTrack room calibration, the tracker world coordinate system could be defined such that the water bath floor is exactly in the  $x - y$  plane.

Thus, the calibration transformation  $T_C$  can be recovered by minimizing the variance of the  $z$ -component of the points  $\{p_w\}$  reconstructed with (15). The Matlab Simplex optimizer was used to estimate the Euler angles parameterization of  $T_C$ . The resulting standard deviation of the reconstructed points in the  $z$ -direction is about 1.1 mm for multiple calibration runs. We

TABLE I  
DESCRIPTION OF DATA SETS AND TARGET REGISTRATION ERRORS

Pat.-Sweep	side	anatomy	target LN / US diameter	surface TRE	auto. TRE
1-6	left	LNL IV-II, CA, IJV with some compression	normal / 5 mm	28.9 mm	4.1 mm
2-1	left	LNL IV-II, CA, IJV, chain of LN metastases	metastasis / 12 mm	54.5 mm	3.3 mm
2-10	right	LNL IV-II, CA, IJV, sternocleidoid muscle	normal / 7 mm	53.4 mm	2.8 mm
3-6	right	LNL III-II, larynx, bifurcation, submandibular gland	suspect / 12 mm	4.9 mm	3.5 mm
3-7	left	LNL IV-II medial, trachea, thyroid, larynx, CA	normal / 5 mm	4.9 mm	3.8 mm
4-2	right	LNL IV-III, thyroid, CA, IJV	metastasis / 12 mm	4.1 mm	3.0 mm
4-6	right	IB-IIA-IIB-V, bulky lymph node disease	metastasis / > US probe	7.7 mm	6.1 mm
5-5	right	IIA-craniocaudal, parotis, submandibular gland	suspect / 19 mm	6.1 mm	4.6 mm
5-6	right	IA-IIA, chin, jaw, digastric muscle, sm.gland	normal / 10 mm	10.7 mm	4.0 mm

Abbrev: Pat.=patient number, Sweep=number of ultrasound sequence, LN=lymph node, US diameter=maximal nodal diameter according to US sections, surface TRE=Target registration error after skin surface registration, auto. TRE=Target registration error left after additional image-based registration, LNL=lymph node level (Robbins), CA=carotid artery, IJV=internal jugular vein

manually selected points on the line representing the floor in the ultrasound image. Depending on the angle of the probe, the appearance of the floor becomes manifold and blurred rather than showing a sharp line (Fig. 5). Using careful adjustment of a single focal zone, the floor plane can be determined even for difficult angles. In particular, when slowly moving the probe from a perpendicular to a steep position, one can see how the first characteristic sharp line of the floor merges into a more complex structure, while the actual position of the floor is still identifiable. This approach allowed us to avoid manufacturing calibration hardware such as the Cambridge phantom [32], which overcomes this problem by providing an obstacle always perpendicular to the image plane. Furthermore, as the calibration did not need to be done on a regular basis by physicians, we did not consider more automated approaches involving complex algorithms for estimating the correct floor points, as in [33]. For single-wall calibration with automatic floor line extraction, a reconstruction precision of 3.4 mm is reported in [32] with magnetic tracking and 2.7 mm in [33] using optical tracking. In a more recent study [34], these two methods are quantitatively evaluated with optical tracking. The authors report mean 3-D point localization errors of 2.7 and 1.67 mm, respectively.

Medical ultrasound devices usually assume a constant speed of sound of 1540 m/s, which is the average in human soft tissue. In our water bath at room temperature, however, the speed is approximately 1485 m/s. Objects visible in the images are thus closer to the probe than indicated by the rulers on the image borders. Hence for our calibration, the vertical direction of the ultrasound image was scaled accordingly with a factor of  $1485/1540 = 0.964$  (compare [31, p. 458 (7)]).

We used a second tracked tool to intersect the ultrasound image plane in the water bath. By means of a 3-D visualization of the tool and the ultrasound image we visually confirmed the calibration was correct (i.e., the optimization had not minimized the floor plane to a wrong local optimum). It also allowed us to assess the latency of the ultrasound image with respect to the tracker readings by means of periodic pointer movement. We concluded that no systematic temporal calibration is necessary. One would expect the ultrasound image to be up to one frame older than the tracker readings (due to the image processing and reconstruction on the ultrasound machine as well as frame grabber delay). This equals to a maximum of 40 ms delay, or 0.2 mm error if a continuous sweep with the typical speed of 5 mm/s is considered.

### C. Results

For each of the five patients, all of the recorded tracked ultrasound sequences were registered onto the skin surface extracted from the CT scan, as described in Section III. This registration always converged, with computation times being 1–5 s. The mean residual point distances (based on the Chamfer distance lookups) for the five patients were 2.36, 3.67, 2.41, 2.31, and 2.70 mm, respectively.

For studying the accuracy and robustness of the intensity-based registration methods (Section IV), one or two ultrasound sweeps were selected from each patient, nine in total. Together, the selections should cover the whole range of the expected clinical applications, including the typical lymph node sites on both sides of the neck and displaying pathological as well as normal lymph nodes of different sizes. Of each sweep in turn, four to seven frames were picked for automatic image-based registration. For validation purposes, the physician carefully established a ground-truth transformation for each set of images using anatomical landmarks such as calcifications in the carotid artery, well-defined lymph nodes or gland tissue, etc. For each set, one particular lymph node of interest was selected as the target; its deviation after the automatic registration with respect to the manually defined ground truth represented the target registration error (TRE). The Hill-Climbing optimizer was used with an initial step size of  $5 \text{ mm}/5^\circ$ , and the similarity measure weightings were  $w_0 = w_1 = w_2 = 0.33$ . Before registration, the ultrasound images were downsampled to  $128 \times 128$  resolution, as their original high spatial sampling rate is not given in the CT data set and hence cannot be exploited for registration. In addition, downsampling the ultrasound images positively affects the computation speed.

Table I depicts the registration accuracy with respect to the selected target for all nine sequences, after surface registration (column 5, surface TRE) and image-based registration (column 6, auto TRE). In the first two patients, large translational misalignments between pure surface registration and ground truth were noted, mainly in the lateral or superior/inferior dimension. In the next three patients the deviations for surface registration ranged between 4.1 and 10.7 mm. After subsequent image-based automatic registration (in case of the first three sequences after some manual adjustment in order to lie within the capture range), the remaining mean target registration error was 3.9 mm. Apart from sweep 4–6, where the size of the lymph node metastasis exceeded the 40 mm linear ultrasound probe and little



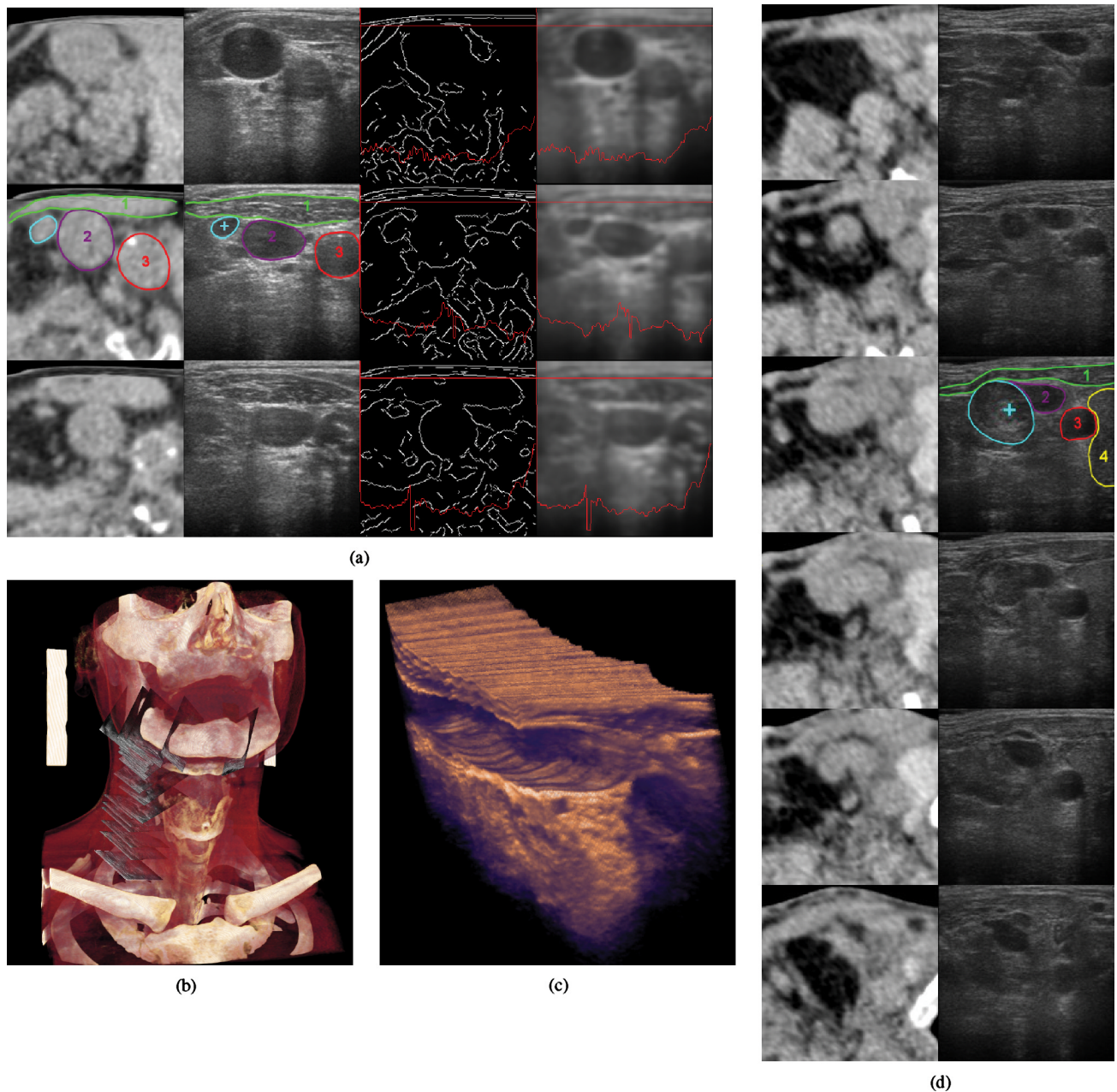


Fig. 6. Image data from different patients used for the evaluation. (a) Three corresponding CT (first column) and ultrasound slices (second column) from Patient 2, right neck, with edges from CT (third column) and blurred ultrasound for edge alignment (fourth column), positioned above, in the middle and below a normal target lymph node (blue outline, cross denotes center for TRE computation). Examples for corresponding anatomy: 1—sternocleidomastoid muscle, 2—internal jugular vein, 3—carotid artery. ROI is delineated in red on the third and fourth column. The physical image size is  $4 \times 4$  cm<sup>2</sup>. (b) Fused visualization of several registered sequences of Patient 5. (c) Volume rendering of compounded ultrasound sweep from Patient 2, clipped to show a profile of the internal jugular vein (note the ripple artifacts due to blood pulsation). (d) Images used for registration from Patient 4 (sweep 2), six consecutive ultrasound and CT slice pairs; a malignant target lymph node (blue outline and TRE cross; 1: sternocleidomastoid muscle, green; 2: internal jugular vein: purple; 3: carotid artery; red; 4: thyroid, yellow).

anatomic distinction was given in the CT data, the target registration error after image-based fusion remained below 5 mm in all cases.

Based on the findings above, the physician would have to expect the following clinical workload and estimated additional time consumption per patient to enable registration of ultrasonography with planning CT (with the setup of the camera system/calibration of optical tracking provided and planning-CT performed routinely):

TABLE II  
ROBUSTNESS AND SPEED OF DIFFERENT OPTIMIZERS, EVALUATED BY REPEATEDLY REGISTERING FROM DISPLACED STARTING ESTIMATES

Optimizer	$\sigma_t$ (mm)	$\sigma_r$ (mm)	TRE(mm)	time(s)
Hill Climbing	1.74	2.92	2.78	4.70
Powell-Brent	1.96	2.90	3.04	23.38
Exhaustive H.C.	1.11	3.14	2.26	336.30

Abbrev:  $\sigma_t$ =translational deviation,  $\sigma_r$ =rotational deviation, TRE=mean Target Registration Error, time=computation time

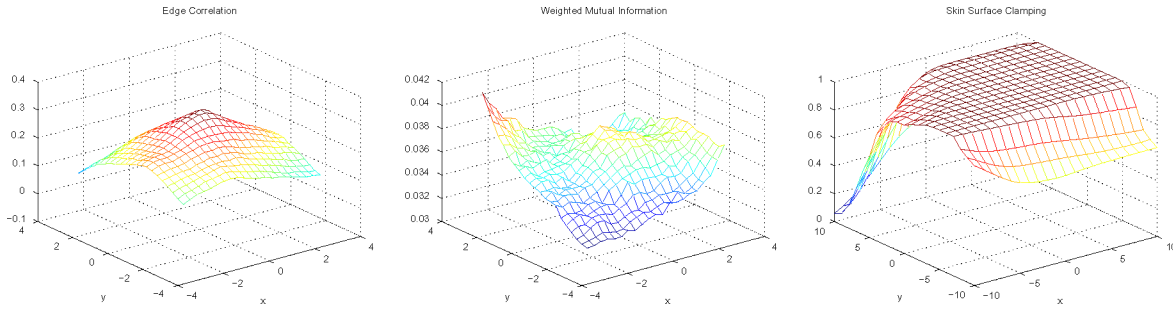


Fig. 7. Plots of the individual similarity measure components against two translational transformation parameters.

- 1) patient positioning with mask and head pad on the ultrasound examination table, removing the mask (preferable following the diagnostic ultrasound examination): 1 min;
- 2) acquisition of one to six freehand ultrasound sweeps (each 20–30 s) containing all pathological lymph nodes on both sides of the neck: 4 min;
- 3) computation of US-CT global skin surface registration: 5 s;
- 4) selection of four to seven US frames per sweep for image-based registration: 4 min;
- 5) visual control, if global skin surface registration is good enough to allow for image-based registration: 3 min;
- 6) if not, manual drag and drop of the  $> 1$  cm deviating sweep(s) into the range of successful image-based registration: 5 min;
- 7) computation of image-based registration: 30 s;
- 8) reevaluation: 2 min.

Thus, in case the global surface registration deviates substantially, performing all steps would amount to approximately 20 min additional working time for the physician (patients 1 and 2), which could be cut down to 15 min by leaving out step 6) in case of sufficient global surface registration to allow for a meaningful image-based registration (patients 3–5).

In order to evaluate the robustness of the different optimization strategies, 130 registrations were launched from initial transformations randomly displaced up to 4 mm/4° in each parameter around the ground truth pose. It was executed for all three optimizers on a set of seven ultrasound images from sequence 10 of Patient 2, three of which are shown in Fig. 6(a). Table II denotes the variance of the error in the translational and rotational components of the resulting transformation in Euler angle parameterization, as well as the mean TRE for a lymph node [Fig. 6(a)] picked as target and the computation time (on a 2.2 GHz AMD Opteron machine with 2 GB RAM running Linux). The Hill-Climbing optimizer is stable while requiring the least number of cost function evaluations, resulting in a very small computation time. Due to the expensive line minimizations, the Powell–Brent optimizer takes significantly more time but does not outperform the Hill-Climbing method. Exhaustive Hill-Climbing achieves the highest robustness, as it evaluates all combinations of possible search directions in each iteration, and hence always reaches the closest local optimum in the parameter space. As its computation time is very high though, we prefer to use the Hill-Climbing strategy as the best tradeoff between speed and robustness. To show the contribution of the various similarity measure components

to the overall registration process, Fig. 7 plots their value against changes in two translational parameters, with respect to the ground-truth transformation. The edge correlation term provides a smooth component, globally converging towards the optimum and hence driving the stability of the algorithm. Weighted mutual information (WMI) contributes with a local peak denoting the highest statistical correspondence. However, for larger translations, the individual image entropies used for normalization might change significantly, as the image contents become different. This can affect the mutual information value in a way that it produces wrong local optima (left side on the center plot). In order not to compromise the global convergence range, we have to limit the influence of the WMI term with respect to edge correlation. As its values are much smaller (0.01–0.1) compared to edge correlation (0.02–0.35), an equal weighting  $w_1 = w_2 = 0.33$  represents a good balance between the two components. The skin surface clamping plot depicts a rim of value 1, dropping smoothly to zero when some of the ultrasound images drift away from the patient’s skin. In fact, for our weight  $w_0 = 0.33$ , the optimization never allows any values other than one for this component.

## VI. DISCUSSION

The described setup allows acquisition of freehand ultrasound data from all sides and orientations required for thorough examination of the patient’s neck while knowing the location of each image in 3-D-space. However, several sources for error might affect a registration of this information to the corresponding CT data.

### A. Tracking Inaccuracy

In our particular case, this error is very small due to the use of the high-end ART tracking system with four cameras.<sup>6</sup>

### B. Calibration Error

Our calibration method yielded a reconstruction plane standard deviation of 1.1 mm (see Section V-B). Considering that we perform sweeps in a continuous motion (without many angulations) and register them individually, we expect the calibration accuracy to have limited impact on the registration. However, we did not perform a complete calibration accuracy study in the scope of this paper.

<sup>6</sup>A 0.4 mm root mean square position error over the whole measurement volume, as stated at [www.ar-tracking.de](http://www.ar-tracking.de)

### C. Patient Immobilization

Even though the patient was placed on the immobilization head pad also used for treatment, slight movements of the patient's head might have happened during the examination, as the thermoplastic mask was removed after the initial positioning. In addition, involuntary movements are possible in the context of counterbalancing the mild ultrasound probe pressure. The head pad was not fixed on the examination couch; thus only gravity and friction prevented its displacement. Similarly, the examination couch itself was a mobile design with parking brakes in use that did not completely prevent translational movement if leaning against it (e.g., examination of the patient's left neck from his right side). These latter two aspects could be responsible for the large deviations in patients 1 and 2 and should be addressed in subsequent studies by using a stationary examination couch and firm fixation of the head pad. The former aspects are more difficult to counteract: leaving the thermoplastic mask in place would impede access to two-thirds of the lymph nodes under question; cutting it according to the requirements of ultrasonography would lead to an unstable mask no longer sufficient for subsequent radiotherapy. Special face masks that gain stability from an individually moulded mouth piece are poorly tolerated by head and neck cancer patients due to hypersalivation and tumorous obstruction of airways. Because of these problems, the global skin surface registration of the whole exam may serve as initial estimation, but a more precise registration should be performed for individual sweeps, which is the case in our approach.

### D. Internal Tissue Movement

The heartbeat causes the vessels to pulsate with a characteristic rhythm over time, with a single high-pressure pulse per heartbeat for arteries and a low-pressure double pulse for veins resulting in tissue movements of 1–3 mm; see Fig. 6(c). In addition, some of the patients had to swallow during the examination, resulting in a strong shift of internal structures in the whole neck region. If deep breath-taking happens, other deformations occur. Hence we asked the patients for quiet, regular breathing throughout the examination, where only minor changes in the anatomy are expected.

### E. Ultrasound Image Distortion

Speed-of-sound variations can cause geometric distortion in the ultrasound images. The speed of sound varies slightly in soft tissue (fat 1476 m/s, water at body temperature 1529 m/s, muscle 1568 m/s, blood 1570 m/s), while the ultrasound machine assumes 1540 m/s for the image reconstruction. The maximum deviation here is about 4%; hence, e.g., a 1 cm fat layer would cause the tissue below to appear 0.4 mm displaced. Further distortion can appear due to refraction. In the given data, a lateral shadowing effect on the carotid artery was a frequent finding; however, we do not expect implications for the registration. The shadowing itself is excluded by our algorithm (see Section IV-C), and the slightly displaced echoes below the carotid artery contribute with a low weight to the similarity measure, as they are on the bottom of the image.

### F. Ultrasound Probe Pressure

In any ultrasound exam, some minimum pressure has to be applied in order to keep direct contact with the patient's skin. Where the skin surface is convex, the linear ultrasound probe will flatten the surface underneath, and the compression should be strongest in the middle of the probe. This causes deformation of the skin, subcutaneous fat, muscular layers, and vessels. Placing and removing the ultrasound probe repeatedly at neighboring locations will thus lead to different distortion of the same structures. In particular, vessel structures, as well as lymph nodes close to the skin, are affected: Due to different inherent blood pressures between the arterial (60–140 mm-Hg) and venous (2–10 mm-Hg) blood system, the external compression effect imposed by the ultrasound probe is much more pronounced in veins; see anatomy contours in Fig. 6(a). In some of the sweeps, we could observe that the internal jugular vein would be totally collapsed due to compression in some parts of the sweep, while it was visible with little or no deformation in others. Nonfixed lymph nodes may be shifted slightly sideways by the ultrasound probe, depending on the consistency of surrounding tissue. In the study reported in this paper, we handled these issues by manual selection of ultrasound image frames well suited for registration, as opposed to a 3-D/3-D registration of all information available from a sweep with the CT scan. Thus, when two sweeps of freehand ultrasound with overlapping structures of interest are recorded, one cannot expect to have an exact match of the corresponding anatomy in both sequences. Again, this supports our supposition that a precise image-based coregistration of anatomic structures to CT is restricted to individual freehand ultrasound sweeps.

Some of the issues mentioned above suggest the integration of deformable models into our registration methods, which we consider an important topic for future research. However, this will require a high degree of distinction between different tissue types with different deformability (arteries, veins, subcutaneous layers, muscle, fat, etc.) and thus in turn will depend on segmentation algorithms and feature-based methods. Furthermore, validating the correctness of all registered CT-ultrasound data in terms of a deformable mapping would be a very time-consuming effort, left with many uncertainties concerning the relation of physiological flexible soft tissues and rigid tumor tissue.

Considering a feature-based approach as an option for automatic registration between US and CT, the carotid artery would be a structure of choice, as it appears well in the ultrasound images, is less prone to artifacts from probe pressure as opposed to veins, and is sited in close proximity to several lymph node regions. But for CT, it is angiographic acquisitions with contrast agent that allow one to precisely extract the tubular shape of the carotids [35]. However, standard planning CT scans for radiotherapy are performed without contrast agent; thus we made use of native CT scans in order to minimize the additional workload in the treatment planning process and to enhance the acceptance for the clinical routine. Our similarity measure approach is derived from physical properties of ultrasound and CT imaging, rather than from particular anatomy. Therefore, the algorithms

should also be applicable on other parts of the human body accessible to ultrasonography. The intensity-based registration itself is performed within a few seconds and may be beneficial as well for endoscopic ultrasound or interventional strategies such as radio-frequency ablation of liver metastases.

According to our time estimates for the proposed procedure, the physician faces a maximum of 20 min additional workload in case of the global surface registration's not meeting the requirements for subsequent image-based registration (patients 1–2). Avoiding pitfalls like an unstable head pad fixation and examination couch along with careful patient positioning will supposedly lead to a sufficient global surface registration like in our patients 3–5, in which case step 6 is dispensable, resulting in 15 min workload per patient. In view of the complexity of planning process in intensity-modulated radiation therapy (IMRT), usually involving the radiation oncologist and medical physicist for more than an hour, this may be deemed acceptable if substantial sparing of tissue irradiation in noneffected lymph node sites and dose escalation in clearly identified nodal metastases due to on-site ultrasound visualization is to be the gain.

Still, from the clinician's point of view, further reduction of personal involvement is desirable. As a next step, the manual selection of US frames should be replaced by automation (e.g., every twentieth frame of a sweep). In fact, in this paper, we select frames not only for being less distorted than others but also for being helpful to establish the ground truth validation. If image-guided registration is possible on frames selected automatically as well, preceded by a stable skin surface registration, steps 4 and 5 will become dispensable as well, cutting the manual involvement time down to below 8 min, which should be the aim of coming research.

Altogether, we found a TRE of 3–5 mm for all lymph nodes other than bulky disease, which is in accordance with the literature for the BAT-system compared to gold marker verification of the prostate [6]. In view of the internal movements and tissue deformation mentioned above, this seems to be acceptable. Studies on positioning head and neck patients with a thermoplastic mask reported on a daily setup variation of around 3 mm, but the reported measurements usually relied on osseous landmarks in portal films. Translating the traditional three-point laser setup error of mean 3.33 mm in any single direction into a mean composite vector, deriving from a high-precision optically guided patient localization system and considering all six degrees of freedom (like in our TRE), the offset was 6.97 mm with a standard deviation of 3.63 mm [36]. In fact, the planning CT alone, imaging the neck tissue at a certain time point, pretends to resemble an accurate picture of a region that is in fact in motion all the time due to pulsating vessels, breathing, swallowing, etc. These effects are mirrored in the ultrasound examination, which gives a more realistic imaging along the time scale, e.g., when watching the pulsation without moving the probe. Thus, part of the registration error cannot be attributed to pressure effects from the ultrasound probe but rather relies on inherent tissue movement. This will not be overcome by optimizing positioning and should be allowed for when defining the margins (additional zones around GTV) for planning target volume definition. There is a growing interest on this issue of soft tissue movement, as the widths of margins determine to what extent dose escalation and

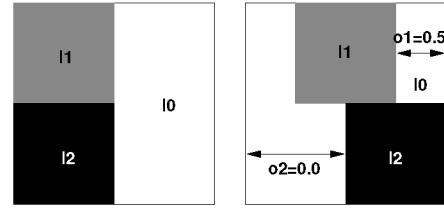


Fig. 8. Test images containing three intensities. The boxes  $I_1$  and  $I_2$  are displaced in the moving image, resulting in relative overlaps  $o_1$  and  $o_2$ . An overlap of 1 means the boxes are at identical positions; 0 corresponds to no overlap.

normal tissue sparing is possible in high-precision radiotherapy. The magnitude of TRE values derived from CT-US registration may help to define the minimal margin to ensure full dose coverage in macroscopic tumor surrounded by soft tissue.

## VII. CONCLUSION

We developed methods for the integration of diagnostic ultrasound of cervical lymph nodes into the radiation therapy planning context of head and neck cancer, where target volume definition is performed on a native CT. This includes a tracked free-hand ultrasound exam and a global preregistration to the skin surface derived from the CT data to allow an automatic intensity-based registration of a set of ultrasound slices to the corresponding CT data. For the ultrasound acquisition, it is crucial to control external movements as far as possible to ensure spatial consistency for the global surface registration. Internal tissue movements (swallowing, pulsating vessels) and flattening of surface tissue by the linear ultrasound probe are inevitable though. From a clinician's point of view, 3–5 mm uncertainty about the exact site of each image voxel seems acceptable for radiotherapy planning purposes in soft tissue. Given the tracking system is set up and calibrated, the expected additional workload for the physician amounts to 15 min per patient with adequate fixation. This might be halved by automation of frame selection for image-based registration, which should be addressed in studies to come. In principle, ultrasonography findings have been made available for treatment planning, where benefits for target volume definition might be expected.

## APPENDIX I

### WEIGHTED MUTUAL INFORMATION

1) *Basics*: In order to study the effect of weighted image histograms for mutual information, we consider the case of simple test images [Fig. 8]. Half of the fixed (reference) image  $A$  has intensity  $I_0$ , and each one-fourth is covered with a box of intensity  $I_1$  and  $I_2$ , respectively. In the moving image  $B$ , those boxes are shifted to the right, resulting in an overlap  $o_1, o_2 \in [0 \text{dots}, 1]$  with respect to the boxes in the fixed image. The entropy in the images is independent of the overlaps

$$p(I_0) = \frac{1}{2}; \quad p(I_1) = \frac{1}{4}; \quad p(I_2) = \frac{1}{4}$$

$$H_a = H_b = - \sum_i p(i) \log p(i) \quad (16)$$



TABLE III  
JOINT PROBABILITY DISTRIBUTION

$\mathbf{P}_{ab}$	$\mathbf{I}_0$	$\mathbf{I}_1$	$\mathbf{I}_2$
$\mathbf{I}_0$	$\frac{1}{4}(o_1 + o_2)$	$\frac{1}{4}(1 - o_1)$	$\frac{1}{4}(1 - o_2)$
$\mathbf{I}_1$	$\frac{1}{4}(1 - o_1)$	$\frac{1}{4}o_1$	0
$\mathbf{I}_2$	$\frac{1}{4}(1 - o_2)$	0	$\frac{1}{4}o_2$

TABLE IV  
WEIGHTED JOINT PROBABILITY DISTRIBUTION

$\mathbf{P}_{ab}$	$\mathbf{I}_0$	$\mathbf{I}_1$	$\mathbf{I}_2$
$\mathbf{I}_0$	$\frac{1}{2}(o_1 w_1 + o_2 w_2)$	$\frac{1}{2}(1 - o_1)w_1$	$\frac{1}{2}(1 - o_2)w_2$
$\mathbf{I}_1$	$\frac{1}{2}(1 - o_1)w_1$	$\frac{1}{2}o_1 w_1$	0
$\mathbf{I}_2$	$\frac{1}{2}(1 - o_2)w_2$	0	$\frac{1}{2}o_2 w_2$

$$= - \left( \frac{1}{2} \log \frac{1}{2} + 2 \frac{1}{4} \log \frac{1}{4} \right) = \frac{3}{2} \log 2. \quad (17)$$

The joint probability distribution in Table III lists the probability of every pair of intensities in the left and right images, respectively, and depends on the overlap values.

The joint entropy computed from it results in

$$H_{ab} = -\frac{1}{4} \left( o_1 \log \frac{o_1}{4} + 2(1 - o_1) \log \frac{1 - o_1}{4} + o_2 \log \frac{o_2}{4} + 2(1 - o_2) \log \frac{1 - o_2}{4} + (o_1 + o_2) \log \frac{o_1 + o_2}{4} \right). \quad (18)$$

2) *Histogram Weighting*: If we are more confident that the overlap of the upper box reflects the image alignment than the lower one, we would like to weight the image intensities depending on the vertical position. We therefore multiply the histogram values with two integer count variables  $c_1$  and  $c_2$ . They correspond to weights  $w_1, w_2$  adding to one

$$w_1 = \frac{c_1}{c_1 + c_2}; \quad w_2 = \frac{c_2}{c_1 + c_2}. \quad (19)$$

The marginal image entropies are still independent of the overlap. However, they depend on the weights

$$p(I_0) = \frac{1}{2}; \quad p(I_1) = \frac{w_1}{2}; \quad p(I_2) = \frac{w_2}{2} \quad (20)$$

$$H_a = H_b = - \sum_i p(i) \log p(i) \quad (21)$$

$$= -\frac{1}{2} \left( w_1 \log \frac{w_1}{2} + w_2 \log \frac{w_2}{2} + \log \frac{1}{2} \right). \quad (22)$$

The joint entropy is now affected by both the overlap and the weight values Table IV

$$l(x) = -x \log x \quad (23)$$

$$H_{ab} = l \left( \frac{1}{2} o_1 w_1 \right) + 2l \left( \frac{1}{2} (1 - o_1) w_1 \right) + l \left( \frac{1}{2} o_2 w_2 \right) + 2l \left( \frac{1}{2} (1 - o_2) w_2 \right) + l \left( \frac{1}{2} (o_1 w_1 + o_2 w_2) \right). \quad (24)$$

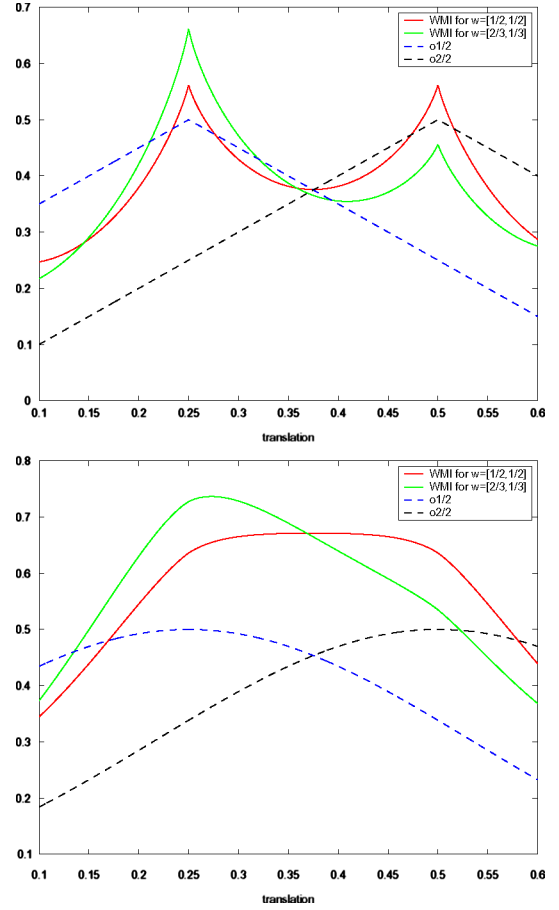


Fig. 9. WMI versus translation of the moving image, for different overlap functions and weightings. The dashed lines show the relative overlaps  $o_1, o_2$  for the test images (upper diagram) and a Gaussian overlap function (lower diagram).

Fig. 9 plots the WMI term versus a translation to the left of the moving image from Fig. 8. The upper diagram illustrates that WMI contains a higher peak at the alignment of the upper box for an overlap function rising linearly with respect to the translation (derived from the test images). For a smooth overlap function that causes a single similarity measure peak in the center (lower diagram), that peak is shifted towards the alignment of the upper box with our weighting. Here not the linear overlap  $o_1, o_2$  with respect to translation is used but a Gaussian function. The mutual information term used is  $2 - (2H_{ab}) / (H_a + H_b)$ .

#### ACKNOWLEDGMENT

The authors would like to thank Dr. M. Dobritz from the Radiology Department for his support with the CT acquisitions. The authors would like to thank in particular F. Sauer and A. Khamene for their continuous support.

#### REFERENCES

- [1] A. Ahuja and M. Ying, "Sonography of neck lymph nodes, Part II: Abnormal lymph nodes," *Clin. Radiol.*, vol. 58, pp. 359–366, May 2003.
- [2] Q. Wang, S. Takashima, F. Takayama, J. Wang, S. Kawakami, A. Saito, T. Matsushita, and S. Sone, "Detection of occult metastatic lymph nodes in the neck with gray-scale and power Doppler US," *Acta. Radiol.*, vol. 42, pp. 312–319, May 2001.

- [3] K. Yonetsu, M. Sumi, M. Izumi, M. Ohki, S. Eida, and T. Nakamura, "Contribution of doppler sonography blood flow information to the diagnosis of metastatic cervical nodes in patients with head and neck cancer: Assessment in relation to anatomic levels of the neck," *AJNR Amer. J. Neuromdiol.*, vol. 22, pp. 163–169, Jan. 2001.
- [4] M. Sumi, M. Ohki, and T. Nakamura, "Comparison of sonography and CT for differentiating benign from malignant cervical lymph nodes in patients with squamous cell carcinoma of the head and neck," *AJR Amer. J. Roentgenol.*, vol. 176, pp. 1019–1024, Apr. 2001.
- [5] A. Grosu, M. Piert, W. Weber, B. Jeremic, M. Picchio, U. Schratzenstaller, F. Zimmermann, M. Schwaiger, and M. Molls, "Positron emission tomography for radiation treatment planning," *Strahlenther Onkol.*, vol. 181, pp. 483–499, Aug. 2005.
- [6] H. McNair, S. Mangar, J. Coffey, B. Shoulders, V. Hansen, A. Norman, J. Staffurth, S. Sohaib, A. Warrington, and D. Dearnaley, "A comparison of CT- and ultrasound-based imaging to localize the prostate for external beam radiotherapy," *Int. J. Radiol. Oncol. Biol. Phys.*, vol. 65, pp. 678–687, Jul. 2006.
- [7] A. Fung, K. Ayyangar, D. Djajaputra, R. Nehru, and C. Enke, "Ultrasound-based guidance of intensity-modulated radiation therapy," *Med. Dosimetry*, vol. 31, pp. 20–29, 2006.
- [8] M. Fuss, B. Salter, S. Cavanaugh, C. Fuss, A. Sadeghi, C. Fuller, A. Ameduri, J. Hevezi, T. Herman, and C. T. Jr., "Daily ultrasound-based image-guided targeting for radiotherapy of upper abdominal malignancies," *Int. J. Radiol. Oncol. Biol. Phys.*, vol. 59, pp. 1245–1256, Jul. 2004.
- [9] T. Scarbrough, N. Golden, J. Ting, C. Fuller, A. Wong, P. Kupelian, and C. T. Jr., "Comparison of ultrasound and implanted seed marker prostate localization methods: Implications for image-guided radiotherapy," *Int. J. Radiol. Oncol. Biol. Phys.*, vol. 65, pp. 378–387, Jun. 2006.
- [10] F. Cury, G. Shenouda, L. Souhami, M. Duclos, S. Faria, M. David, F. Verhaegen, R. Corns, and T. Falco, "Ultrasound-based image guided radiotherapy for prostate cancer—Comparison of cross-modality and intramodality methods for daily localization during external beam radiotherapy," *Int. J. Radiol. Oncol. Biol. Phys.*, vol. 66, pp. 1562–1567, 2006.
- [11] G. P. Penney, J. M. Blackall, D. Hayashi, T. Sabharwal, A. Adam, and D. J. Hawkes, "Overview of an ultrasound to CT or MR registration system for use in thermal ablation of liver metastases," *Proc. Med. Image Understand. Anal.*, pp. 65–68, 2001.
- [12] D. Amin, T. Kanade, A. M. D. Gioia, and B. Jaramaz, "Ultrasound registration of the bone surface for surgical navigation," *Comput. Aided Surg.*, vol. 8, no. 1, pp. 1–16, Jan. 2003.
- [13] B. Brendel, S. Winter, A. Rick, M. Stockheim, and H. Ermert, "Registration of 3D CT and ultrasound datasets of the spine using bone structures," *Comput. Aided Surg.*, vol. 7, pp. 146–155, 2002.
- [14] T. Lange, S. Eulenstein, M. Hunerbein, H. Lamecker, and P.-M. Schlag, "Augmenting intraoperative 3D ultrasound with preoperative models for navigation in liver surgery," in *Proc. MICCAI2004 Part II*, 2004, p. 534ff.
- [15] B. Porter, D. Rubens, J. Strang, S. Totterman, and K. Parker, "Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers," *IEEE Trans. Med. Imag.*, vol. 20, no. 4, pp. 354–359, Apr. 2001.
- [16] P. Slomka, J. Mandel, D. Downey, and A. Fenster, "Evaluation of voxel-based registration of 3-D power Doppler ultrasound and 3-D magnetic resonance angiographic images of carotid arteries," *Ultrasound Med. Biol.*, vol. 27, pp. 945–955, Jul. 2001.
- [17] A. Roche, X. Pennec, G. Malandain, and N. Ayache, "Rigid registration of 3D ultrasound with MR images: A new approach combining intensity and gradient information," *IEEE Trans. Med. Imag.*, vol. 20, pp. 1038–1049, Oct. 2001.
- [18] G. Stippel, J. Ellsmere, S. Warfield, W. Wells, and W. Philips, "A new technique for multi-modal 3D image registration," in *WBIR 2003 Proceedings*. New York: Springer, 2003, vol. 2717, Lecture Notes Computer Science, pp. 244–253.
- [19] G. Penney, J. Blackall, M. Hamady, T. Sabharwal, A. Adam, and D. Hawkes, "Registration of freehand 3D ultrasound and magnetic resonance liver images," *Med. Image Anal.*, vol. 8, pp. 81–91, Mar. 2004.
- [20] J. Blackall, G. Penney, A. King, and D. Hawkes, "Alignment of sparse freehand 3-D ultrasound with preoperative images of the liver using models of respiratory motion and deformation," *IEEE Trans. Med. Imag.*, vol. 24, no. 11, pp. 1405–1416, Nov. 2005.
- [21] G. Penney, D. Barratt, C. Chan, M. Slomczykowski, T. Carter, P. Edwards, and D. Hawkes, "Cadaver validation of intensity-based ultrasound to CT registration," in *Proc. MICCAI Part II*, Oct. 2005, pp. 1000–1007.
- [22] M. Mellor and M. Brady, "Phase mutual information as a similarity measure for registration," *Med. Image Anal.*, vol. 9, pp. 330–343, Sep. 2005.
- [23] C. Castro-Pareja, V. Zagrodsky, L. Bouchet, and R. Shekhar, "Automated prostate localization in external-beam radiotherapy using mutual information-based registration of treatment planning CT and daily 3D ultrasound images," in *Int. Congr. Ser.*, 2005, vol. 1281, pp. 435–440.
- [24] A. Leroy, P. Mozer, Y. Payan, and J. Troccaz, "Rigid registration of freehand 3D ultrasound and CT-scan kidney images," in *Proc. MICCAI*, Sep. 2004, p. 837ff.
- [25] W. Wein, B. Roper, and N. Navab, "Automatic registration and fusion of ultrasound with CT for radiotherapy," in *MICCAI 2005 Proceedings*. New York: Springer, 2005, vol. 3750, Lecture Notes in Computer Science, pp. 303–311.
- [26] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 6, pp. 641–647, Jun. 1994.
- [27] G. Borgefors, "Distance transforms in arbitrary dimensions," *Comput. Vision, Graph. Image Process.*, vol. 27, pp. 321–345, 1984.
- [28] A. Roche, X. Pennec, G. Malandain, N. Ayache, and S. Ourselin, "Generalized correlation ratio for rigid registration of 3D ultrasound with MR images INRIA Tech. Rep., Jul. 2000.
- [29] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual information based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
- [30] T. Sielhorst, M. Feuerstein, J. Traub, O. Kutter, and N. Navab, "CAM-PAR: A software framework guaranteeing quality for medical augmented reality," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 1, pp. 29–30, Jun. 2006, Supplement 1.
- [31] L. Mercier, T. Langø, F. Lindseth, and D. Collins, "A review of calibration techniques for freehand 3-D ultrasound systems," *Ultrasound Med. Biol.*, vol. 31, pp. 449–471, 2005.
- [32] R. Prager, R. Rohling, A. Gee, and L. Berman, "Rapid calibration for 3-D freehand ultrasound," *Ultrasound Med. Biol.*, vol. 24, pp. 855–869, 1998.
- [33] F. Rousseau, P. Hellier, and C. Barillot, "Confusius: A robust and fully automatic calibration method for 3D freehand ultrasound," *Med. Image Anal.*, vol. 9, pp. 25–38, Sep. 2005.
- [34] F. Rousseau, P. Hellier, M. Letteboer, W. Niessen, and C. Barillot, "Quantitative evaluation of three calibration methods for 3-D freehand ultrasound," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1492–1501, Nov. 2006.
- [35] M. Sanderse, H. Marquering, E. Hendriks, A. van der Lugt, and J. Reiber, "Automatic initialization algorithm for carotid artery segmentation in CTA images," in *MICCAI 2005 Proceedings*. New York: Springer, 2005, vol. 3750, Lecture Notes Computer Science, pp. 846–853.
- [36] T. Hong, W. Tome, R. Chappell, P. Chinnaiyan, M. Mehta, and P. Harari, "The impact of daily setup variations on head-and-neck intensity-modulated radiation therapy," *Int. J. Radiol. Oncol. Biol. Phys.*, vol. 61, pp. 779–788, Mar. 2005.