Phantomless Auto-Calibration and Online Calibration Assessment for a Tracked Freehand 2D Ultrasound Probe

Matthew Toews* and William M. Wells III

Abstract—This paper presents a method for automatically calibrating and assessing the calibration quality of an externally tracked 2D ultrasound (US) probe by scanning arbitrary, natural tissues, as opposed a specialized calibration phantom as is the typical practice. A generative topic model quantifies the posterior probability of calibration parameters conditioned on local 2D image features arising from a generic underlying substrate. Auto-calibration is achieved by identifying the maximum a-posteriori (MAP) image-to-probe transform, and calibration quality is assessed online in terms of the posterior probability of the current image-to-probe transform. Both are closely linked to the 3D point reconstruction error (PRE) in aligning feature observations arising from the same underlying physical structure in different US images. The method is of practical importance in that it operates simply by scanning arbitrary textured echogenic structures, e.g. in-vivo tissues in the context of US-guided procedures, without requiring specialized calibration procedures or equipment. Observed data take the form of local scale-invariant features that can be extracted and fit to the model in near real-time. Experiments demonstrate the method on a public data set of in vivo human brain scans of 14 unique subjects acquired in the context of neurosurgery. Online calibration assessment can be performed at approximately 3Hz for US images of 640x480 pixels. Auto-calibration achieves an internal mean point reconstruction error (PRE) of 1.2mm and a discrepancy of [2mm,6mm] in comparison to calibration via a standard phantom-based method.

I. INTRODUCTION

Freehand ultrasound (US) is a cheap, safe and portable imaging modality for guidance and visualization in interventional settings. Although 3D ultrasound sensors exist, B-mode or 2D probes remain widely used in clinical practice due to their relatively high image quality and low cost [1]. For effective visualization and navigation, 2D US slices acquired using a hand-held probe must be related to the underlying 3D anatomy of the patient, including other relevant information such as diagnostic 3D MR or CT scans. This is typically accomplished via an external probe tracking system, e.g. optical or electromagnetic, that provides the rigid pose of a tracking target fixed to the US probe within the 3D world coordinate system of the patient.

* M. Toews is with the École de Technologie Supérieure of the University of Quebec, Montréal, QC, Canada (e-mail: matthew.toews@etsmtl.net).

W. M. Wells III is with the Brigham and Women's Hospital of the Harvard Medical School, Boston, MA, USA (email: sw@bwh.harvard.edu)

A crucial component of such US-guided navigation systems is the similarity transform mapping coordinates in the 2D US image plane to the 3D coordinate system of the probe tracker. This transform is determined by the geometry of the US image plane relative to the tracking target, and is typically estimated via a procedure referred to as probe calibration. A variety of calibration procedures have been proposed in the literature, see [2], [1] for detailed surveys. The overwhelming majority operate by scanning a specialized 3D object with known geometrical properties, referred to as a *phantom*, that may consist of Lego bricks [3], wedges [4], wires [5], [6], points [7], [8], planes [9], [10], gelatin casts [11], etc. Although phantom-based calibration procedures are widely used, they have several drawbacks, particularly in clinical settings. Calibration must be computed each time the image-to-probe geometry may have changed, e.g. following instrument sterilization, variations in imaging parameters such as depth-of-field, etc. Even relatively simple calibration procedures require additional equipment, procedures and expertise, complicating the clinical work flow. Finally, calibration may degrade or fail over the course of a procedure, due to factors including physical accidents, partially obscured or occluded tracking targets, etc., resulting in inaccurate US-guidance and a safety risk to the patient.

Motivated by these factors, a body of research has examined automatic calibration and quality assessment based directly on intra-procedural US image data [12], [13], [14], [15]. In general, these approaches adopt image registration techniques to establish alignment between US data acquired from different probe positions, thereby providing constraints for calibration. US image registration is relatively straightforward in the case of 3D probes [14]; however, registering 2D US slices separated by out-of-plane movements during acquisition is a non-trivial challenge. Boctor et. al. describe image-based methods for self-calibration [16] and online calibration assessment [12], but report an inability to recover out-of-plane calibration parameters in the case of 2D US data. Wein et. al. propose an image-based calibration procedure in which a special acquisition protocol is used to help in solving the alignment problem: two perpendicular US sweeps are acquired by rotating the US probe by 90 degrees about a fixed 3D point [13]. The acquisition protocol may not always be easily applicable due to physical constraints, e.g. awkward patient positioning may prevent acquisition of perpendicular sweeps. Barratt et. al. propose self-calibration based on registering segmented bone surfaces in US with a pre-operative model [15]. Surface segmentation is a major challenge, however, and experiments use

an off-line manual segmentation, and surface alignment may not be applicable in domains other than orthopedic surgery.

We propose a novel method for online quality assessment and auto-calibration for a tracked 2D US probe system that operates directly on images of arbitrary tissues exhibiting echogenic texture, e.g. in-vivo patient anatomy. It can thus be used directly in interventional settings to increase the robustness of US-guidance and to improve patient safety. Our method adopts a generative model of local image features arising from a generic underlying substrate, that quantifies the posterior probability of calibration parameters conditioned on features observed in multiple US sweeps. It operates on freehand 2D US sweeps acquired according to a relatively unconstrained protocol, e.g. with US data acquired from similar imaging plane orientations, and it is thus applicable in practical clinical settings. Experiments demonstrate the effectiveness of the method on a set of public US data from 14 subjects acquired in the context of image-guided neurosurgery. To the best of our knowledge, alternative, practically applicable methods for image-based auto-calibration and online assessment do not currently exist.

II. MATERIALS AND METHODS

The context for our method is as follows: US images of an arbitrary echogenic substrate are acquired via smooth sweeps of a 2D US probe, where the rigid six degree-of-freedom pose of the probe is provided by means of a tracking target fixed to the probe. The substrate may generally exhibit non-rigid mechanical properties, e.g. in the case of soft biological tissues, however it is assumed to be unmoving and thus modeled as rigid over the course of an US sweep. Furthermore, the substrate is assumed to contain an arbitrary configuration of echogenic macro structures, i.e. regions of differing acoustic impedance such as internal organs, brain structures such as sulci and ventricles, etc. These give rise to distinctive, localizable patterns or features in the US image that can be detected in different US images acquired from similar, yet not identical, probe positions.

Our method operates by continuously aligning each newly acquired US image to sequences of previously acquired US data stored in memory. Alignment is achieved by identifying image-to-image correspondences between US slices, which then serve as the basis for auto-calibration and quality assessment. In order to effectively constrain the image-to-probe transform, correspondences must be identified between pairs of US images acquired from non-trivial differences in probe position, including differences in both out-of-plane rotation and translation. Although precisely identical tissues cannot be observed in 2D US slices differing by out-of-plane rotation, image similarity measures can be effective at identifying similar image patches in US images acquired with minor differences in probe position. For example, image intensity correlation is known to degrade smoothly with out-of-plane displacement due to non-zero US slice thickness, and this principle is used in speckle-decorrelation methods [17], [18], [19] that determine elevational separation between US images as a function of an intensity decorrelation profile.

Using a similar principle here, we formulate calibration in terms of correspondences between distinctive 2D features extracted from two or more US image sequences. Unlike speckle decorrelation methods, we do not attempt to measure out-of-plane displacement directly, but rather to identify corresponding locations where it is close to zero. The assumption is that although corresponding feature pairs may not arise from precisely the same tissue, their difference in 3D location can be modeled as zero-mean Gaussian noise. A sufficient number of such corresponding feature pairs, identified across multiple US slices, can thus be used as the basis for our auto-calibration and line quality assessment procedures. The advantage of this approach is that calibration can be performed with a simple acquisition protocol using data acquired directly from the patient, and thus does not require separate calibration procedures or equipment.

We present our method in this section as follows. Section "Local Feature Correspondence" describes the correspondence method forming the basis of our method. Section "Generative Model" presents our probabilistic model including all random variables, their distributions and their parameters. Finally, Section "Estimation" describes the process of estimating calibration, including a training phase to initialize model parameters from an initial US sequence and a fitting phase in which new US data is used to compute the image-to-probe calibration transform.

A. Local Feature Correspondence

A fundamental operation in our method is to identify correspondences between distinctive, generic image structures in a 2D US slice and a sequence of previously acquired slices. To be effective in an arbitrary interventional setting, correspondence computation must cope with out-of-plane geometrical variations, as US data are generally acquired from different slices through the 3D patient anatomy. Furthermore, it must account for in-plane geometrical variations such as translation, rotation and scaling, e.g. in case of variation in probe position or depth settings. Finally, although auto-calibration makes use of 3D probe tracking information, the correspondence process must operate in the case where probe tracking is unavailable, e.g. in the case of unknown or degraded calibration. To this end, we consider correspondences between informative local patterns or features that can be reliably detected and localized in multiple intersecting US planes.

Local feature detection and matching has a rich history in the computer vision field. Early approaches focused on identifying patterns such as corners that can be localized within the image plane [20], [21]. Scale-space theory demonstrated that the notion of an image pattern is intimately linked to the scale at which it is observed [22]. This lead to the development of so-called scale-invariant feature detection approaches, which identify the scale of distinctive image features in a manner independent of the image resolution [23], [24], in addition to location and orientation, e.g. the widely used scale-invariant feature transform (SIFT) approach. Further developments included affine-invariant features that are effective for locally approximating geometrical image deformations due to planar

projection [25], however, these are less relevant in the case of US data which are not formed by projection.

In this paper, we adopt the scale-invariant feature representation for computing image correspondence, in which a feature is characterized by geometry and image appearance. Feature geometry $S = \{\bar{u}, \theta, \sigma\}$ is an oriented image patch including 2D pixel location $\bar{u} = \{u, v\}$, orientation θ and scale σ. Appearance is a descriptor vector I encoding the image intensity pattern associated with S, and is used in computing correspondence. The most effective encodings are typically based on histograms of local gradient orientations [23], [26], [27], [28], [29] popularized by the SIFT approach. Figure 3 a) illustrates an example of correspondences between scale-invariant features extracted in different images.

A key advantage of correspondence via local scale-invariant features is that a single corresponding feature pair is sufficient to estimate a four-parameter similarity transform aligning two images, i.e. global image rotation, magnification and translation. Each correspondence thus provides an independent hypothesis as to the underlying alignment solution, and hypotheses can be clustered to identify valid correspondences robustly despite the presence of outliers and incorrect correspondences. For example, techniques such random sample consensus (RANSAC) [30] and the Hough transform are commonly used to identify valid correspondences between pairs of projective images [23], [26], [31], [32] in the context of computer vision.

In comparison to computer vision problems, however, US image data bears distinct challenges, necessitating an alternative approach. To illustrate, relatively dense correspondences can be identified between arbitrary pairs of 2D projective images of the same 3D scene (i.e. photographs), where projections of the same unoccluded 3D point can generally be observed in images acquired from different camera viewpoints. In contrast, US data are not formed via perspective projection but rather as slices through 3D space. As manually acquired US slices do not generally pass through the same 3D point at the same angle, dense one-to-one correspondences cannot generally be identified between an arbitrary pair of US images.

B. Generative Model

US probe calibration seeks to map points in 2D US images to 3D locations relative to the probe. We adopt a generative model describing the process by which US features are produced, including probabilistic uncertainty. Let $\bar{x} = [x, y, z, 1]^T$ represent the location of a 3D point in the world in homogenous coordinates, and let $\bar{u} = [u, v, 0, 1]^T$ represent the pixel location of the same point within a 2D US image plane. Points $\bar{u}$ and $\bar{x}$ are related by the following equation:

$$\bar{x} = T_w T_p T_s \bar{u} + \bar{\xi}, \qquad (1)$$

where $\bar{\xi}$ is a zero-mean additive noise vector quantifying the uncertainty of 3D point localization due to both the tracking system and calibration, and T_w, T_p and T_s are 4×4 homogenous transform matrices. T_w is a rigid transform from the US probe to the 3D world, provided for each US image by the tracking mechanism. $T_s = \mathrm{diag}[s_u, s_v, 1, 1]$ is a

diagonal scaling matrix converting US spatial units (pixels) to world distance units (mm). Scaling parameters (s_u, s_v) may in principle be provided by the US system, however they vary with changes in US settings such as depth, and are treated as unknown here. Finally, T_p is the unknown rigid transform from the US image plane to the US probe, with six intrinsic parameters: three rotation angles and three coordinate translations. For simplicity, let $T_{ps} \triangleq T_p T_s$ represent the unknown image-to-probe transform matrix. Figure 1 illustrates the transform from US image to world coordinates.
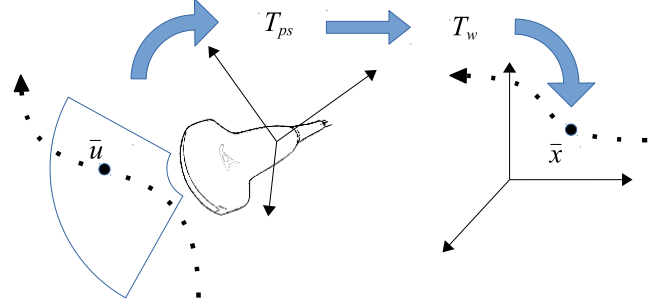

Fig. 1. Illustrating the transform from an US image pixel coordinate $\bar{u}$ to a 3D world coordinate $\bar{x}$ via the image-to-probe transform T_{ps} (determined by calibration) and the probe-to-world T_w provided by a probe tracking system.

We propose a novel generative model that can be used to infer T_{ps} from observed feature and tracker data, $\{S_{ij}, I_{ij}\}$ and T_{wj} respectively, where j indexes US images in a sequence and i indexes features within an image. The Bayes network for our model is shown in Figure 2. At a high level, it consists of a topic model similar to probabilistic latent semantic analysis (PLSA) [33] that characterizes features observed in new US image in terms of those observed in a prior training or reference sequence, as shown in Figure 3 (a), and a parts-based appearance model [34] characterizing 2D feature layout, and calibration / pose factors relating 2D image locations to 3D. Descriptions of model variables and probability expressions are as follows.
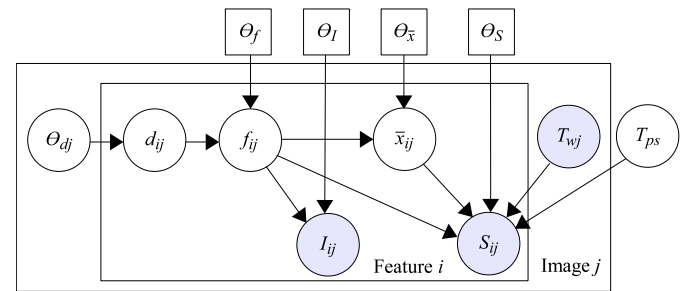

Fig. 2. A Bayes network representation for ultrasound calibration. Arrows represent statistical dependencies between variables, which are either random variables (circles) or distribution/density parameters (squares). Observed features are modeled as conditionally independent and identically distributed. The inner plate (rectangle) replicates over individual features indexed by i, and the outer plate over images indexed by j. Shaded circles indicate observed variables, i.e. image features I_{ij}, S_{ij} and probe position T_{wj}.

T_{ps}: a 3D similarity transform mapping 2D US image coordinates to the 3D coordinates of the tracked probe, including a scale factor converting pixel to mm units. $p(T_{ps})$ is a prior density over transform parameters.

T_{wj}: a 3D rigid transform mapping 3D probe coordinates at image j to 3D world coordinates, provided by a probe tracking system.

f_{ij}: a categorical random variable taking on values $[0, \ldots, K]$, indicating the latent echogenic structure or *model feature* associated with observed feature i in image j. The discrete probability distribution $p(f_{ij}|d_{ij}; \Theta_f)$ is conditioned on the training image d_{ij} associated with feature ij, and parameterized by Θ_f, i.e. discrete probability mass values. Let f_{ij}^k denote the event $f_{ij} = k$, i.e. that observed feature i of image j is an observation of model feature k. Furthermore, let f_{ij}^0 or $f_{ij} = 0$ denote a spurious background feature that is not associated with any specific structure that will play a role later in parameter estimation.

d_{ij}: a categorical random variable taking on discrete values $[1, \ldots, M]$, indicating the index of a reference or training image from which observed feature ij is sampled. Let d_{ij}^m denote the event $d_{ij} = m$, i.e. that the model feature f_{ij} associated with observed feature i of image j came from training image m. The probability of d_{ij} is quantified by the discrete distribution $p(d_{ij}|\Theta_{dj})$ conditional on parameters Θ_{dj} defined next.

Θ_{dj}: parameters of discrete distribution $p(d_{ij}|\Theta_{dj})$ over the discrete training image index associated with feature data observed in new image j. Our model assumes that the acquisition plane of image j is approximately parallel to images in a reference or training sequence, acquired with a smooth sweep of US image probe. In this setting, features observed in image j can be assumed to arise from tissues also observed in a concentrated, contiguous set of images in the training sequence, as shown in Figure 3. We thus represent the distribution $p(d_{ij}|\Theta_{dj})$ as a discrete Gaussian-like probability mass function defined by mean and concentration parameters $\Theta_{dj} = \{\Theta_{dj\mu}, \Theta_{dj\sigma}\}$:

$$p(d_{ij} = m; \Theta_{dj}) \propto e^{-\frac{1}{2}\left(\frac{m - \Theta_{dj\mu}}{\Theta_{dj\sigma}}\right)^2}. \quad (2)$$

Note that this represents a departure from typical topic models that consider sets of unstructured training data, e.g. text corpora [33]. The prior probability of parameters Θ_{dj} is quantified by the density $p(\Theta_{dj})$. This density can generally be defined according to prior assumptions regarding mean and concentration parameters, here it is considered uniform or uninformative for simplicity.

$\bar{x}_{ij} \triangleq [x, y, z, 1]^T$: a homogenous coordinate vector representing the 3D location associated with model feature f_{ij}. $p(\bar{x}_{ij}|f_{ij}; \Theta_x)$ is an isotropic Gaussian density conditioned on f_{ij} with mean and variance parameters Θ_x. Note that $\bar{x}_{ij}$ is related to 2D location $\bar{u}_{ij}$ via Equation (1), i.e. $\bar{x}_{ij} = T_{wj}T_{ps}\bar{u}_{ij} + \bar{\xi}$.

I_{ij}: a vector representing a 2D image feature appearance descriptor, derived from US image intensities within patch S_{ij}. $p(I_{ij}|f_{ij}; \Theta_I)$ is an isotropic Gaussian density over appearance descriptor elements conditioned on f_{ij} with mean and variance parameters Θ_I.

$S_{ij} \triangleq (\bar{u}_{ij}, \theta_{ij}, \sigma_{ij})$: the 2D image feature geometry including a 2D pixel location $\bar{u}_{ij} \triangleq [u_{ij}, v_{ij}, 0, 1]^T$ expressed in homogenous coordinates, an in-plane orientation θ_{ij} and scale σ_{ij}. $p(S_{ij}|T_{wj}, T_{ps}, \bar{x}_{ij}, f_{ij}; \Theta_S)$ is a conditional density over the in-plane geometry of f_{ij} factored as

$$p(S_{ij}|T_{wj}, T_{ps}, \bar{x}_{ij}, f_{ij}; \Theta_S) = p(\bar{u}_{ij}|T_{wj}, T_{ps}, \bar{x}_{ij}, f_{ij}; \Theta_S)$$
$$p(\theta_{ij}|T_{wj}, T_{ps}, f_{ij}; \Theta_S)p(\sigma_{ij}|T_{wj}, T_{ps}, f_{ij}; \Theta_S). \quad (3)$$

In Equation (3), $p(\bar{u}_{ij}|T_{wj}, T_{ps}, \bar{x}_{ij}, f_{ij}; \Theta_S)$ is an isotropic Gaussian density conditioned on transforms (T_{wj}, T_{ps}) and 3D location $\bar{x}_{ij}$. $p(\theta_{ij}|T_{wj}, T_{ps}, f_{ij}; \Theta_S)$ is a Von Mises density, a circular analog of the Gaussian density, conditioned on transforms (T_{wj}, T_{ps}) and model feature f_{ij}. $p(\sigma_{ij}|T_{wj}, T_{ps}, f_{ij}; \Theta_S)$ is a Gaussian density over $\log \sigma_{ij}$ conditioned on transforms (T_{wj}, T_{ps}) and model feature f_{ij}. Densities are parameterized by mean and variance/variability parameters Θ_S.
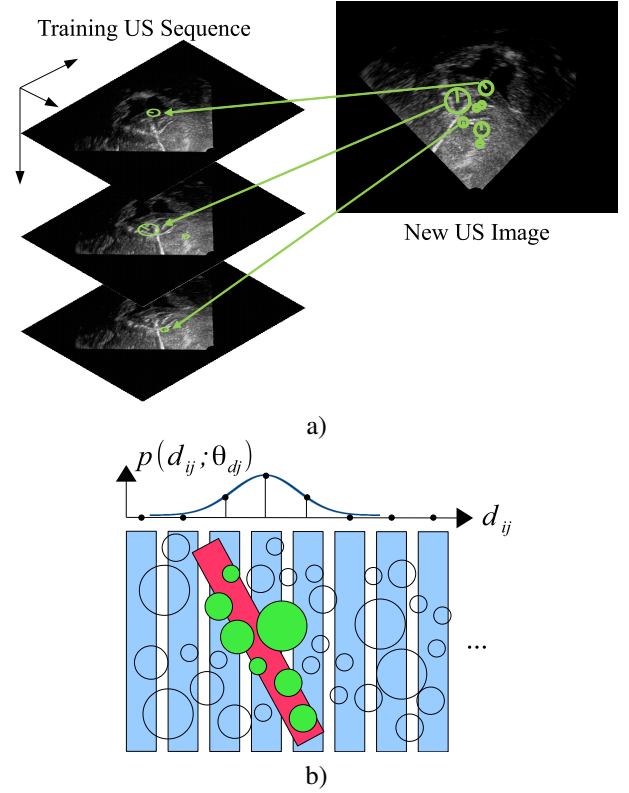

Fig. 3. Illustrating the relationship between features observed in a new US image and a training or reference sequence. In a) scale-invariant features (green circles) observed in a new testing US image (upper right) generally correspond with features observed across multiple images in a reference or training sequence (left). As shown in b), a new image (diagonal box) is acquired from an orientation that is similar but not identical to images in a training or reference sequence (vertical boxes) through a substrate containing observable local structure (circles). Structure observable in the new US image (colored circles) is thus modeled as arising from a concentrated set of training images via $p(d_{ij}|\Theta_{dj})$. Note that while training frames need not be sampled regularly and may overlap, the expected 3D point localization error is lower bounded by the Nyquist rate or interval, i.e. half the spatial displacement between US frames.

C. Estimation

Our method seeks to evaluate the posterior probability of calibration parameters T_{ps} based on features extracted from two or more US sweeps through an arbitrarily textured echogenic substrate. Observable data for each US image j consists of probe transform T_{wj} along with the appearance and 2D geometry of $\{I_{ij}, S_{ij}\}$ extracted features. Specifically, we consider the joint posterior probability of $(T_{ps}, \{\Theta_{dj}\})$, that based on the graphical model in Figure 2 is:

$$p(T_{ps}, \{\Theta_{dj}\}|\{S_{ij}, I_{ij}\}, \{T_{wj}\}) \propto$$
$$p(T_{ps}) \prod_j p(\Theta_{dj}) \prod_i \left(\sum_{k,m} p(d_{ij}^m; \Theta_{dj}) p(f_{ij}^k|d_{ij}^m; \Theta_f) p(\bar{x}_{ij}|f_{ij}^k; \Theta_x) p(I_{ij}|f_{ij}^k; \Theta_I) p(S_{ij}|f_{ij}^k, \bar{x}_{ij}, T_{wj}, T_{ps}; \Theta_S) \right). \quad (4)$$

Note that in Equation (4), d_{ij} and f_{ij} are considered as nuisance variables and are marginalized. Variables $\{\Theta_{dj}\}$ are not generally of interest and would also typically be marginalized in a Bayesian approach, however they are estimated jointly here with T_{ps}. Joint modeling is less computationally intensive than marginalization and is consistent with the assumption that each US image maps to a single distinct position in the underlying substrate being imaged.

An important note regarding Equation (4) is that inner summation takes the form of a mixture model. In the case of distinctive image features [35], there is typically at most one highly probable fit or *correspondence* f_{ij}^k between an observed image feature (ij) and a model feature k. Intuitively, this is because the probability of observing the same distinctive pattern at multiple image locations is low for natural images, as opposed to artificial images containing repetitive structure such as a chessboard. The inner marginalization sum is thus dominated either by the mixture component associated with this correspondence or by the contribution of the background model feature f_{ij}^0, which exhibits broad variance in appearance and geometry and is taken to be a nominally low, constant probability. Note that a new image feature may fit to a set of similar model features arising from a similar physical structure whose extent spans several adjacent images in a US sequence, however one of these represents the most likely fit.

Estimation thus seeks to efficiently approximate Equation (4) from a set of sparse, highly probable correspondences from feature data. The process consists of three phases: 1) training where the model is initialized, 2) fitting where 2D feature correspondences are identified between a new image sequence and the model and finally 3) quality assessment and auto-calibration, where the posterior probability in Equation (4) is evaluated and maximized based on correspondences.

Training initializes model parameters for fitting and estimation where Θ_X, Θ_I and Θ_S are means and (co)variance parameters for conditional densities, and Θ_f are feature occurrence probabilities. In general, maximum likelihood (ML) parameter estimates can be obtained by clustering features based on appearance and geometry, where each cluster consists of observations of a specific model feature k, i.e. f_{ij}^k. Here we consider initialization where the model is defined with respect to a single reference or training sweep through the underlying tissue. Aside from the practical advantage of requiring only a single rather than multiple sweeps, the assumption of 1:1 correspondence between training and testing data removes the need for ML estimation to identify feature clusters.

First, values of d_{ij} are defined as image indices in the training sequence, and a new unique model feature $f_{ij} = k$ is assigned to each feature observed in the training sequence. Defined this way, $p(d_{ij}|\Theta_{dj})$ and $p(f_{ij}|d_{ij}, \Theta_f)$ are delta functions and parameters Θ_f are initialized to reflect the specific features observed in each image of the training sequence. Density mean values are taken to be observed values I_{ij}, S_{ij}. (Co)variance estimates are not initially available, and are initialized as follows. Appearance variance Θ_I is feature-specific and estimated as the minimum squared Euclidean distance $||I_{ij}, I'||^2 = \sum_k (I_{ij}[k] - I'[k])^2$ between descriptor I_{ij} and a descriptor I' extracted from a set of unrelated US data (e.g. acquired from arbitrary non-brain tissues). Variance estimates for Θ_S are set to fixed values that apply to all features, except for in-plane translation variance of density $p(\bar{u}_{ij}|T_{wj}, T_{ps}, \bar{x}_{ij}, f_{ij}; \Theta_S)$ which is taken to be proportional to feature scale σ_{ij}^2.

Fitting seeks to identify a set of highly probable image-to-model correspondences between model features and 2D features extracted in a new US sequence. These can then be used to estimate or maximize the joint posterior in Equation (4). Correspondence is challenging because T_{ps} is unknown, and thus constraints on 3D feature position cannot be used. We thus propose a correspondence procedure for the case where a new image intersects a subset of training images at nearly parallel 3D acquisition planes. In this case, similar features can be observed in both new and training images, and 2D correspondence constraints can be brought to bear. To illustrate, let $T = \Phi T'$, where Φ is a 3D similarity transform relating the transform of a new image T to that of a training image T'. In the case of approximately parallel acquisition planes,

$$\Phi \approx \begin{bmatrix} \alpha \cos\phi & -\alpha \sin\phi & 0 & du \\ \alpha \sin\phi & \alpha \cos\phi & 0 & dv \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (5)$$

where (α, ϕ, du, dv) are the parameters of a 2D in-plane similarity transform.

Fitting thus proceeds by identifying image-to-model correspondences independently for each new US image as follows: (1) Appearance descriptors from features extracted in the new image are matched to nearest neighbor mean descriptors in

the model using Euclidean distance, providing a set of ML image-to-model feature correspondence candidates f_{ij}^k, i.e.:

$$f_{ij}^k = \underset{f_{ij}}{\arg\max} \left\{ p(I_{ij}|f_{ij}; \Theta_I) \right\}. \qquad (6)$$

This process can be computed efficiently using fast approximate nearest neighbor methods, e.g. using the K-D tree data structure [36]. (2) As 2D feature geometry is described by 4 parameters, each image-to-model correspondence is sufficient to provide an ML estimate of the approximate in-plane similarity transform Φ aligning feature data in the testing image to the model. This can be viewed as maximizing the likelihood associated with in-plane geometry, substituting unknown 3D coordinates $\bar{x}_{ij}$ with in-plane 2D coordinates $\bar{u}_{ij}$:

$$\Phi_{ij} = \underset{\Phi}{\arg\max} \left\{ p(S_{ij}|f_{ij}^k, \bar{x}_{ij} = u_{ij}, T_{wj} = \mathbf{I}, T_{ps} = \Phi; \Theta_S) \right\}. \qquad (7)$$

A search is then performed to identify a dense set of hypotheses $\{\Phi_{i,j}\}$ that differ by less than thresholds on in-plane translation, rotation and scaling, in a manner similar to the Hough transform. Correspondences associated with these hypotheses are consistent with an in-plane similarity transform between the new image and the reference sequence, and thus represent potentially valid image-to-model correspondences, or *inlier correspondences*. Finally, (3) ML estimates of mean and concentration parameters Θ_{dj} are estimated from the reference image indices associated with image-to-model correspondences f_{ij}^k:

$$\Theta_{dj}^* = \underset{\Theta_{dj}}{\arg\max} \left\{ p(d_{ij}|\Theta_{dj}) \right\}. \qquad (8)$$

Having identified a set of correspondences f_{ij}^k and estimated parameters Θ_{dj}, the posterior expression in Equation (4) can now be evaluated over values of T_{ps}. 3D locations of observed features $\bar{x}_{ij}$, in addition to mean parameters Θ_x for conditional density $p(\bar{x}_{ij}|f_{ij}^k; \Theta_x)$, are estimated from 2D feature locations via Equation (1), e.g. $\bar{x}_{ij} = T_{wj}T_{ps}\bar{u}_{ij}$. These can be viewed as ML estimates:

$$\bar{x}_{ij}^* = \underset{\bar{x}_{ij}}{\arg\max} \left\{ p(\bar{u}_{ij}|T_{wj}, T_{ps}, \bar{x}_{ij}, f_{ij}^k; \Theta_S) \right\}. \qquad (9)$$

Given that the inner marginalization sum of the joint posterior in Equation (4) is dominated by a single non-zero term, its logarithm takes the form:

$$\log\, p(T_{ps}, \{\Theta_{dj}\}|\{S_{ij}, I_{ij}\}, \{T_{wj}\}) \propto$$
$$C + \log\, p(T_{ps}) - \sum_{f_{ij}^k \in Inliers} \|T_{wj}T_{ps}\bar{u}_{ij} - T_{wk}T_{ps}\bar{u}_k\|^2, \qquad (10)$$

where C is a constant, $\bar{u}_k$ is the mean 2D location of model feature k, and $p(T_{ps})$ is a prior probability density over T_{ps} that can be specified according to prior knowledge or left as an uninformative or uniform distribution.

Calibration Quality Assessment: From Equation (10), it can be seen that the posterior probability of T_{ps} is largely determined by the prior probability $p(T_{ps})$ and the sum of squared errors between predicted 3D locations of the model

features k and new image features (ij). In the case of an informative prior $p(T_{ps})$ over T_{ps}, e.g. following standard phantom-based calibration, we propose using average point reconstruction error (PRE) as an online indicator of calibration quality:

$$PRE(\,T_{ps}\,) = \frac{1}{N} \sum_{ij} \|T_{wj}T_{ps}\bar{u}_{ij} - T_{wk}T_{ps}\bar{u}_k\|. \qquad (11)$$

The PRE represents the average error in image correspondences given an informed prior $p(T_{ps})$. It is low in the case of accurate calibration, however it increases with degradation of tracking or calibration, or rigid patient motion. Significant non-linear tissue deformation or resection is indicated by either increased PRE or by a lack of image-to-model correspondences, both lead to a low posterior probability.

Auto-Calibration: In the case where quality assessment reports unacceptably high PRE or a lack of correspondences, re-calibration may be necessary to restore image guidance. Traditional phantom-based calibration is difficult to perform in an intra-operative setting, however, for previously mentioned reasons. Auto-calibration is performed by estimating the calibration matrix T_{ps}^* maximizing the log posterior expression in Equation (10) by optimizing over T_{ps}.

Practically, T_{ps} is a scaled rigid transform defined by eight intrinsic parameters: a 3D displacement vector $\bar{t} = \{t_1, t_2, t_3\}$, two positive scaling factors (s_u, s_v) that are treated here as unconstrained values in the log domain $(\log(s_u), \log(s_v))$, and three rotation parameters constrained by the orthonormality requirements of a 3×3 rotation matrix. Rotation parameter constraints complicate optimization of Equation (11), here we adopt a Rodrigues parameterization in which a 3D rotation is represented as scalar angular rotation about a 3D axis. Specifically, a 3D vector of unconstrained values $\bar{\omega} = \{\omega_1, \omega_2, \omega_3\}$ is adopted, where the unit vector $\frac{\bar{\omega}}{\|\bar{\omega}\|}$ defines the rotation axis and the norm $\|\bar{\omega}\|$ defines the rotation angle.

Optimization thus seeks to identify a parameter vector $\{\log(s_u), \log(s_v), \bar{t}, \bar{\omega}\}$ maximizing Equation (10). A variety of non-linear optimization methods could be brought to bear, here we adopt the Nelder-Mead simplex method [37] which does not require explicit gradient computation and converges reliably in optimization scenarios involving small numbers of parameters. Additionally, rather than maximizing over all correspondences, optimization considers the error of the 75% of correspondences with minimum error, in order to reduce the influence of a small percentage of outlier correspondences. Figure 4 provides an illustration of the algorithmic flow for the sake of clarity[1].

III. RESULTS

A. Data

Experiments are designed to test auto-calibration and quality assessment in the context of image-guided neurosurgery, with freehand US sweeps of the same human brain acquired prior to major resection. The public MNI BITE data is used [38], consisting of 14 US data sets from different neurosurgical

---

[1]An implementation of the algorithm in C++ code is available at https://github.com/mt3000/TMI-2017-Ultrasound-Toews
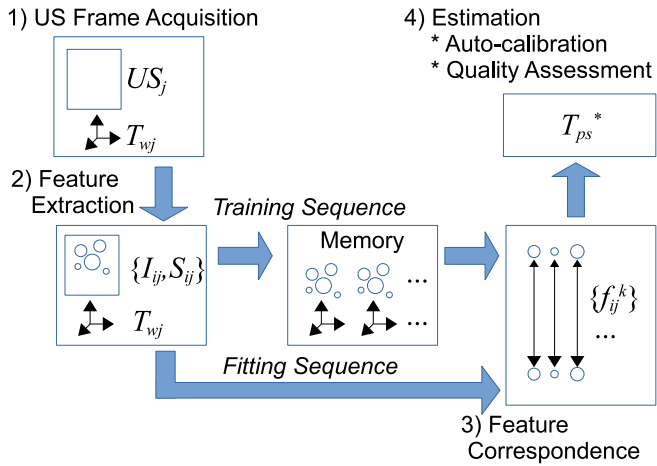
Fig. 4. Illustrating the algorithmic flow of parameter estimation based on two sequences of tracked US data (training and fitting). 1) US frames and tracking coordinates $\{T_{wj}\}$ are acquired, and 2) local scale-invariant features $\{S_{ij}, I_{ij}\}$ are extracted from each US frame. 3) Correspondences $\{f_{ij}^k\}$ are computed between features extracted from the training sequence stored in memory and a new fitting sequence. 4) Finally auto-calibration T_{ps}^* and quality assessment are estimated from feature-to-feature correspondences.

patients. All data are acquired and distributed with patient consent according to an institutionally approved ethics protocol. The US probe was tracked using a Polaris infrared optical system (Northern Digital, Waterloo, ON, Canada), using reflective spheres rigidly attached to a TA003 tracker (Traxtal Technologies, Toronto, ON, Canada). The ultrasound scanner is a ATL HDI 5000 (Philips, Bothell, WA), using a P7-4 MHz phased array transducer probe with depth settings of 6.5 cm and 8 cm. US image data consist of screen captures at 0.3mm pixel resolution, these are used without additional post-processing, note however that the scanner may be performing typical internal processing operations, e.g. log compression of intensities. Note that image enhancement techniques such as speckle reduction may improve the reliability of feature detection and correspondence, this is an avenue for future investigation.

The BITE data are provided as pre-calibrated, where a standard phantom-based calibration method has been used to determine a unique calibration matrix T_{ps} independently for each of the 14 patient cases. This phantom-based calibration method is based on a configuration of Z-fiducials described in [39] and a reported error range of $[0.49mm, 0.74mm]$ [38], and it serves as an independent baseline against which the accuracy of our method is evaluated.

The BITE data were acquired for the purpose of 3D image reconstruction from 2D freehand US, with 2-5 manual sweeps per patient. Each sweep consists of a unidirectional linear motion perpendicular to the US plane with approximately constant probe orientation, and contains approximately 500 US frames with an average probe displacement of approximately 0.25mm. The out-of-plane orientation variation between sweeps ranges from 0-23 degrees, correspondences are reliably identified across this range. Each US frame is provided with a single linear transform $T_j = T_{wj}T_{ps}$ combining

the tracker pose and the standard phantom-based calibration, mapping the 3D probe position to a 3D reference coordinate system. Note that these data were not acquired for the purpose of quality assessment or auto-calibration, and they thus provide an effective demonstration of our method in the context of freehand US acquisition.

B. Procedure

Prior to auto-calibration, training and fitting are performed between all pairs of sweeps for each subject, recall that fitting computation does not require 3D information, e.g. probe tracking or calibration. A number of scale-invariant feature detectors exist, here we adopt the computationally efficient 'SURF' algorithm [26]. Briefly, salient image regions $(\bar{u}, \theta, \sigma)$ are identified based on the local Hessian determinant, and local intensity I is encoded via a 64-element descriptor of Haar wavelet responses. Images are processed at size 640×480 with 0.2mm resolution, where each image results in ≈ 800 features. A mask is applied to filter features occurring near the border of the US fan, as these may characterize the shape of the US fan rather than underlying tissue content. Figure 5 illustrates features identified in a US frame, which generally arise from distinctive US image structure. Note that features here typically arise from patterns of anatomical structure ranging in scale from several mm to several cm, and because correspondences are located on US slices, the minimum expected error in 3D feature localization is proportional to the spatial displacement between US frames.
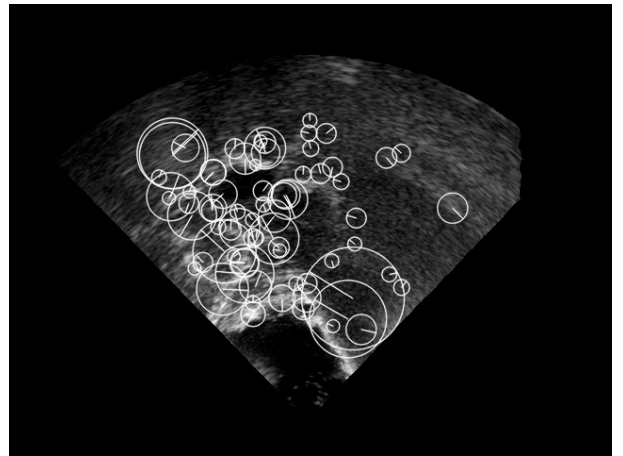

Fig. 5. Features extracted in a single US frame. White circles overlaying the image represent in-plane location and scale, radial lines represent orientation. To reduce visual clutter, only 1/5 of extracted features are shown.

A three-stage procedure is adopted, simulating a clinical scenario including: (1) initial calibration, (2) quality assessment following perturbation and (3) auto-calibration. For each stage, the point reconstruction error (PRE) is estimated via Equation (11). Initial calibration evaluates the 3D error of feature correspondences given the original phantom-based BITE calibration method. To test calibration quality assessment, the tracking matrix T_j for each US frame j is perturbed by a randomly generated image-to-probe matrix T_{ps}', thereby simulating online loss of calibration due to a physical

Stage	Mean	Median	Min	Max
(1) Initial calibration	1.36	1.12	0.58	2.95
(2) Random perturbation	23.2	21.1	3.11	71.4
(3) Auto-calibration	1.18	1.05	0.49	2.91

TABLE I

3D POINT RECONSTRUCTION ERROR (PRE), MM UNITS, FOR 10 RANDOM TRIALS ON 14 SUBJECTS. NOTE THAT HIGH PRE DUE TO RANDOM PERTURBATION (2) IS AUTOMATICALLY FLAGGED DURING QUALITY ASSESSMENT AND CORRECTED FOR BY AUTO-CALIBRATION (3).

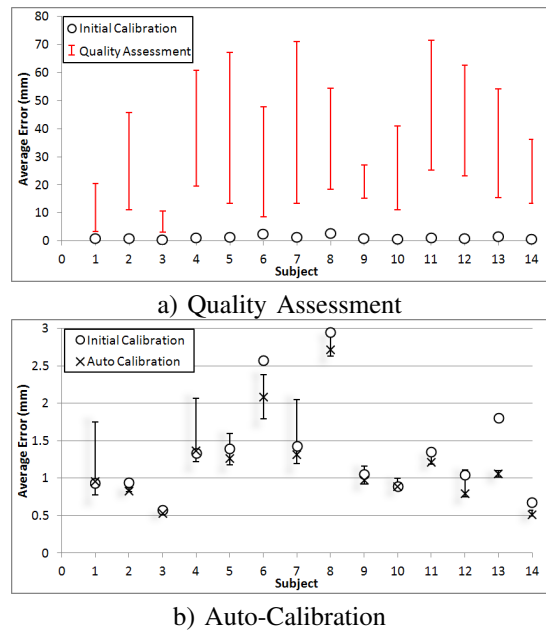a) Quality Assessment

b) Auto-Calibration

Fig. 6. PRE during (a) quality assessment and (b) auto-calibration across 14 subjects. In both graphs, $PRE(T_{ps})$ for initial phantom-based calibration is shown (circles). Graph (a) plots the range of $PRE(T_{ps})$ following 10 random perturbations (red vertical bars), which is noticeably higher than that of initial calibration. Graph (b) plots the range of $PRE(T_{ps}*)$ for 10 randomized auto-calibration trials per subject, which is similar to initial calibration.
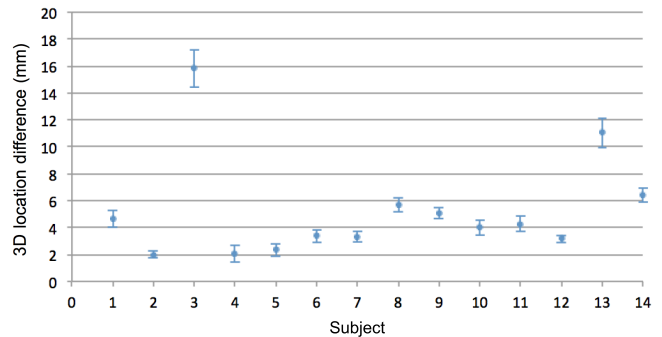

Fig. 7. Discrepancy between 3D point locations as predicted by auto-calibration and original phantom-based calibration. High discrepancy in the cases of subjects 3 and 13 appear to be due to significant deformation between US acquisitions.

accident. Calibration quality is assessed in terms of the PRE based on perturbed tracking coordinates $T'_j = T_j T'_{ps}$. Finally, auto-calibration seeks to recover the inverse image-to-probe transform T'^{-1}_{ps}, and to thereby recover calibration.

The three-stage procedure is repeated in ten independent trials for each of the 14 BITE subjects. Random image-to-probe transform parameters are generated over ranges typical of hand-held tracked ultrasound probes: $[-200, 200]$mm for 3D displacements $t_i \in \bar{t}$, $[0, 2\pi]$ for Rodrigues parameters $\omega_i \in \bar{\omega}$ and $[0.05, 10]$ for isotropic pixel scale (s_u, s_v). For optimization of Equation (11) via the simplex method, parameters are initialized as $\{\log(s_u), \log(s_v), \bar{d}, \bar{\omega}\} = \{0, 0, \{0, 0, 0\}, \{2\pi, 0, 0\}\}$.

C. Discussion of Results

Figure 6 displays the PRE for initial calibration, random perturbation and auto-calibration for the 14 subjects. Results for each data set are reported for the pair of US sweeps with the highest number of correspondences, as this generally results in the minimum PRE. Statistics for PRE across experimental stages are reported in Table I, note that the PRE values for initial phantom-based calibration and auto-calibration are virtually identical, indicating that both methods lead to similar reconstruction accuracy levels. Note also that there is no overlap between the PRE ranges for quality assessment following random perturbation and calibration. Optimization converges to low PRE values for all random perturbation trials, demonstrating the robustness of the algorithm.

Note that the PRE following auto-calibration in Table I row (3) is necessarily low, as it is the quantity minimized by our method, and does not necessarily reflect the absolute error of auto-calibration. In order to estimate the accuracy against an independent reference, we compare the absolute 3D locations of features as determined by auto-calibration trials vs. the standard phantom-based calibration provided with the BITE data. Figure 7 graphs the average Euclidean distance between 3D point locations as predicted by auto-calibration vs phantom-based calibration, based on 2D point locations throughout US images. In 12 of 14 cases, auto-calibration and phantom-based calibration map 2D image locations to 3D with an average discrepancy ranging from [2mm,6mm], which would result in useful recovery of intra-operative US-based neuronavigation considering the 1mm resolution of diagnostic MRI. In 2 of 14 cases, there is an average discrepancy of 10-15mm, this appears to be primarily due to noticeable brain shift between acquisitions. In such cases, US sequences acquired in closer succession may help to limit the effect of deformation due to brain shift.

We investigate the reasons for the discrepancy between auto-calibration and phantom-based calibration via simulated results. We reconstruct an ultrasound volume from sequence pairs, reslice the volume via each of the sequence geometries respective such that sequence data comes from precisely the same underlying geometry. Auto-calibration from resliced data results in very low PRE and discrepancy from phantom-based calibration, 0.4mm+- 2mm. Thus discrepancy between auto-calibration and phantom-based calibration must be due to factors external the system, e.g. brain shift and/or deformation between acquisitions. Auto-calibration assumes US data acquired from a rigid underlying substrate, the brain is a non-rigid object that deforms over the course of neurosugery. Deformations may be due to a variety of factors, including
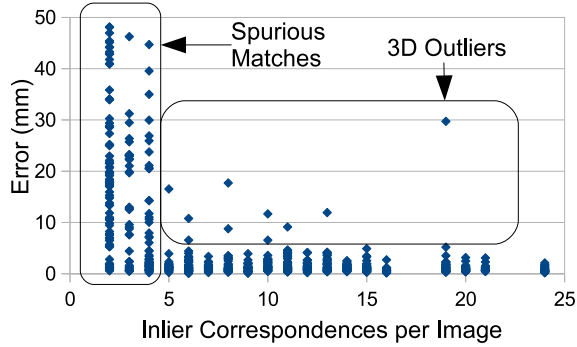
Fig. 8. A graph of 3D PRE error vs. inlier correspondence count reveals low error for inlier counts greater than 4, with the exception of a small number of 3D outliers. Note that high errors due to spurious matching and 3D outliers are automatically filtered by our method and have a negligible impact on quality assessment or auto-calibration.
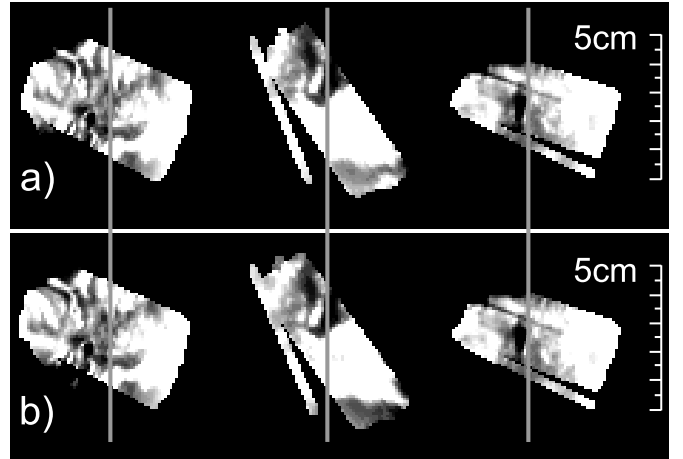

Fig. 9. US volumes reconstructed using (a) phantom-based calibration (b) auto-calibration. Images show axial, sagittal and coronal cross sections, vertical guidelines are provided to highlight similar structure across images. Note the high degree of visual similarity. Reconstruction is performed via trilinear interpolation of US slices within a volume.

contact with the US probe during acquisition, brain shift due to gravity and loss of cerebrospinal fluid, etc., and shifts of up to 8mm have been reported prior to opening the dura lining [40]. Such non-rigid deformations cannot be captured via a linear calibration model and result in error, which we quantify here via the discrepancy between auto-calibration and initial phantom-based calibration.

Figure 8 provides insight into the accuracy of individual image-to-model correspondences following the fitting phase of estimation, for a typical pair of US sequences. The graph shows the Euclidean distance between 3D locations associated with pairs of corresponding image features, where 3D locations are computed from phantom-based calibration T_{ps}. Error is high for slices where the number of inlier correspondences is low (i.e. less than 5), these are generally incorrectly aligned due to insufficient overlap of US content. This is expected, since the volumetric overlap between the two sweeps here is $\approx 78\%$. Error is low for slices with 5 inliers or higher, with the exception of a small number (2.2%) of *3D Outliers*, i.e. coincidental correspondences between different 3D tissues. In total, 41% of testing slices contain 5 or more inlier correspondences, exhibiting an average error of 1.3mm. Note that the number of inliers can be used to identify US frames with no valid correspondences, potentially indicating failure of image guidance.

To visually compare the result of phantom-based calibration and auto-calibration, we reconstruct 3D volumes from single set of 2D slices, using calibration parameters T_{ps} for the respective methods. Figure 9 illustrates axial, sagittal and coronal views of a sample US volume, reconstructed at a relatively coarse 1mm isotropic voxel resolution to reduce 3D artifacts such as holes due to missing US data. These data were acquired from using two 900-slice tracked US sequences, acquired prior to major resection in the context of neurosurgery via arbitrary sweeps along similar trajectories, consisting of primarily of translation with minor 3D rotations. The reconstructed volumes are nearly identical.

IV. CONCLUSIONS

This work presents a novel image-based method for autocalibration and online calibration quality assessment for a tracked 2D ultrasound probe system. The method is based on a novel generative model of local image features arising from US images of an arbitrary echogenic substrate containing distinct, localizable texture patterns. It operates on sequences of US data acquired from approximately parallel imaging planes, where correspondences between different US images can be identified despite minor out-of-plane rotation and in-plane similarity transforms. The model fitting process is efficient and an unoptimized C++ implementation runs at approximately 3 frames per second on a 2.5GHz processor for 640x480 pixel US slices. Computation time is largely due to feature extraction and could be reduced significantly via graphics processing unit (GPU) parallelization. Source code for this work is available at (www.sourcecode.com).

Both auto-calibration and calibration quality assessment and are shown to be closely linked to the point reconstruction error $PRE(T_{ps})$, the average discrepancy between the estimated 3D locations of 2D features given the image-to-probe transform T_{ps}. US-guidance failure can be signaled by either a lack of inlier correspondences at the model fitting stage or by a PRE that is significantly higher than that of baseline calibration. To our knowledge, this is the first calibration quality assessment method applicable to full 3D image-to-probe transforms from 2D US data. The essential operation for quality assessment is fitting feature data to the model, which at a frame rate of 3Hz is fast enough for online monitoring in a clinical setting, e.g. via a running average PRE.

Auto calibration from arbitrary tissues offers the capability of recovering from US-guidance failure during a procedure, without difficulties associated with phantom-based calibration, e.g. intra-operative sterilization, additional equipment and procedures, etc. The primary advantage of our method is that the data acquisition protocol is relatively unrestric-

tive, requiring only two manual probe sweeps following a similar 3D trajectory. Optimization based on an average of 1800 correspondences requires approximately 1-2 seconds on a 2.5GHz processor, which is insignificant in comparison to time requirements for manual data acquisition and fitting/correspondence. Experiments demonstrate that calibration can be recovered from arbitrary random perturbations with a PRE range of $[0.50mm, 2.9mm]$, similar to that provided by initial phantom-based calibration $[0.58mm, 3.0mm]$. These results are particularly compelling given that the US data used here were not acquired for the purpose of either auto-calibration nor calibration quality assessment.

Several aspects of our method merit discussion. First, successful auto-calibration requires that the anatomy of interest contain distinctive, localizable image structure from which local features can be extracted. Qualitatively, we have noted effective correspondence for US images of diverse organs including muscle and abdominal organs. Furthermore, US frames must be acquired from similar, approximately parallel acquisition planes in order to establish feature correspondences. We note that accurate correspondences can be achieved across plane normal vector differences of up to 20 degrees. Future work will include quantitative investigations of auto-calibration in other domains, e.g. the abdomen, and explore the potential 2D feature correspondences for intra-operative registration and navigation.

It is also important to note that while our method treats variability in feature appearance and geometry in terms of a uniform zero-mean additive noise model, uncertainty in ultrasound imaging is generally non-stationary and varies according to both the tissues present and their position in the field of view. Specific factors include tissue-dependent variations in the speed of sound, which is typically taken to be constant in standard US imaging systems, and diffraction effects relating to the transducer aperture, which depend on the US beam focus and lead to a spatially varying point spread function (PSF). These factors introduce systematic distortions in the appearance and geometry of tissues in US images acquired from different probe positions, and thus potentially introduce bias into auto-calibration. Various methods can potentially be used to correct for systematic distortions prior to auto-calibration, and to thereby increase the number and accuracy of feature correspondences and the quality of auto-calibration. For example, prior knowledge of tissues present in the image could be used to account for the variable speed of sound, and deconvolution could be used to correct for distortions due to a spatially varying PSF in US image formation [41]. Finally, tracking uncertainty may introduce error or bias into estimation is thus a practical concern. In practice, we note that tracking uncertainty may arise from errors or variations in synchronization between tracking and imaging measurements, and may be reduced by reducing the US probe velocity during acquisition.

In practice, our model is partially tolerant of systematic deformations, as feature correspondences operative in auto-calibration arise from imaging conditions in which distortions are relatively minor. For example, correspondences are identified between sets of near-parallel images with relatively minor differences in acquisition geometry, primarily from image content near the center of the field of view. In these conditions, deviations from our assumption of zero mean additive noise are minimal. In the case where features arising from a specific tissue region are deformed to the extent that a set of geometrically consistent correspondences cannot be achieved, they are effectively ignored as outliers and have no impact on auto-calibration. Finally, the deviation from the assumption of a constant speed of sound is relatively minor for brain tissues, e.g. [1506-1565m/s] for cerebral spinal fluid and white matter, in comparison to general body tissue, e.g. [1440-3500m/s] for fat and bone [42].

Thus while our paper provides a baseline evaluation on a set of standard, publicly available neurosurgical ultrasound data, an important direction of future research will be investigating correction procedures to more accurately account for spatially varying uncertainty, particularly in adapting auto-calibration to general anatomies. One possible approach is to assess distortions via standard medical ultrasound quality assurance phantoms, e.g. in terms of estimated volume, detection of small targets, etc. Bridging the gap between phantoms typically consisting of regularly arranged geometrical structures and natural tissues exhibiting distinctive local image patterns required by our method may be a challenge.

V. Acknowledgements

References

[1] R. W. Prager, U. Z. Ijaz, A. H. Gee, and G. M. Treece, "Three-dimensional ultrasound imaging," *Journal of Engineering in Medicine*, vol. 224, pp. 192–223, 2010.

[2] L. Mercier, T. Lang, F. Lindseth, and D. L. Collins, "A review of calibration techniques for freehand 3-d ultrasound systems," *Ultrasound in Medicine & Biology*, vol. 31, no. 4, pp. 449–471, 2005.

[3] Y. Xiao, C. X. B. Yan, S. Drouin, D. De Nigris, A. Kochanowska, and D. L. Collins, "User-friendly freehand ultrasound calibration using lego bricks and automatic registration," *International journal of computer assisted radiology and surgery*, pp. 1–9, 2016.

[4] M. Najafi, N. Afsham, P. Abolmaesumi, and R. Rohling, "A closed-form differential formulation for ultrasound spatial calibration: multi-wedge phantom," *Ultrasound in medicine & biology*, vol. 40, no. 9, pp. 2231–2243, 2014.

[5] J. W. Trobaugh, W. D. Richard, K. R. Smith, and R. D. Bucholz, "Frameless stereotactic ultrasonography: method and applications," *Computerized Medical Imaging and Graphics*, vol. 18, no. 4, pp. 235–246, 1994.

[6] J. Kowal, C. A. Amstutz, M. Caversaccio, and L. P. Nolte, "On the development and comparative evaluation of an ultrasound b-mode probe calibration method," *Computer Aided Surgery*, vol. 8, no. 3, pp. 107–119, 2003.

[7] D. Muratore and R. Galloway, "Beam calibration without a phantom for creating a 3-d freehand ultrasound system," *Ultrasound Med Biol.*, vol. 27, no. 11, pp. 1557–1566, 2001.

[8] A. Khamene and F. Sauer, "A novel phantom-less spatial and temporal ultrasound calibration method," in *MICCAI*, 2005, pp. 65–72.

[9] R. Prager, R. Rohling, A. Gee, and L. Berman, "Rapid calibration for 3-d freehand ultrasound," *Ultrasound in medicine & biology*, vol. 24, no. 6, pp. 855–869, 1998.

[10] G. M. Treece, A. H. Gee, R. W. Prager, C. J. Cash, and L. H. Berman, "High-definition freehand 3-d ultrasound," *Ultrasound in medicine & biology*, vol. 29, no. 4, pp. 529–546, 2003.

[11] J. M. Blackall, D. Rueckert, C. R. Maurer Jr, G. P. Penney, D. L. Hill, and D. J. Hawkes, "An image registration approach to automated calibration for freehand 3d ultrasound," in *MICCAI*, 2000, pp. 462–471.

[12] E. Boctor, I. Iordachita, G. Fichtinger, and G. Hager, "Ultrasound self-calibration," in *SPIE Medical Imaging*, 2006, pp. 61 412N–61 412N.

[13] W. Wein and A. Khamene, "Image-based method for in-vivo freehand ultrasound calibration," in *SPIE Medical Imaging*, vol. 6920, 2008, p. 69200K.

[14] Y. L. Ma, K. S. Rhode, G. Gao, A. P. King, P. Chinchapatnam, T. Schaeffter, D. Hawkes, R. Razavi, and G. P. Penney, "Ultrasound calibration using intensity-based image registration: For application in cardiac catheterization procedures," in *SPIE Medical Imaging*, vol. 6918, 2008, p. 69180O.

[15] D. C. Barratt, G. P. Penney, C. S. Chan, M. Slomczykowski, T. J. Carter, P. J. Edwards, and D. J. Hawkes, "Self-calibrating 3d-ultrasound-based bone registration for minimally invasive orthopedic surgery," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 3, pp. 312–323, 2006.

[16] E. M. Boctor, I. Iordachita, G. Fichtinger, and G. D. Hager, "Real-time quality control of tracked ultrasound," in *MICCAI*, 2005, pp. 621–630.

[17] R. J. Housden, A. H. Gee, G. M. Treece, and R. W. Prager, "Sensorless reconstruction of unconstrained freehand 3d ultrasound data," *Ultrasound in medicine & biology*, vol. 33, no. 3, pp. 408–419, 2007.

[18] C. Laporte and T. Arbel, "Learning to estimate out-of-plane motion in ultrasound imagery of real tissue," *MIA*, vol. 15, no. 2, pp. 202–213, 2011.

[19] N. Afsham, M. Najafi, P. Abolmaesumi, and R. Rohling, "Out-of-plane motion estimation based on a rician-inverse gaussian model of rf ultrasound signals: speckle tracking without fully developed speckle," in *Proc. of SPIE Vol*, vol. 8320, 2012, pp. 832 017–1.

[20] H. P. Moravec, "Visual mapping by a robot rover," in *Proc. of the 6th International Joint Conference on Artificial Intelligence*, 1979, pp. 598–600.

[21] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, 1988, pp. 147–151.

[22] T. Lindeberg, "Feature detection with automatic scale selection," *IJCV*, vol. 30, no. 2, pp. 79–116, 1998.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[24] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *IJCV*, vol. 60, no. 1, pp. 63–86, 2004.

[25] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *IJCV*, vol. 65, pp. 43–72, 2005.

[26] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," *CVIU*, vol. 110, no. 3, pp. 346–359, 2008.

[27] M. Toews and W. M. Wells III, "Sift-rank: Ordinal descriptors for invariant feature correspondence," in *CVPR*, 2009, pp. 172–177.

[28] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *European conference on computer vision*. Springer, 2010, pp. 778–792.

[29] J. Dong and S. Soatto, "Domain-size pooling in local descriptors: Dsp-sift," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5097–5106.

[30] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[31] M. Toews and T. Arbel, "Detection, localization and sex classification of faces from arbitrary viewpoints and under occlusion," *IEEE TPAMI*, vol. 31, no. 9, pp. 1567–1581, 2009.

[32] D. Mirota, A. Uneri, S. Schafer, S. Nithiananthan, D. Reh, M. Ishii, G. Gallia, R. Taylor, G. Hager, and J. Siewerdsen, "Evaluation of a system for high-accuracy 3d image-based registration of endoscopic video to c-arm cone-beam ct for image-guided skull base surgery," *IEEE TMI*, vol. 32, no. 7, pp. 1215–1226, 2013.

[33] T. Hofmann, "Probabilistic latent semantic indexing," in *ACM SIGIR*. ACM, 1999, pp. 50–57.

[34] M. Toews and T. Arbel, "A statistical parts-based appearance model of anatomical variability," *IEEE TMI*, vol. 26, no. 4, pp. 497–508, 2007.

[35] M. Toews and W. Wells III, "Efficient and robust model-to-image alignment using 3d scale-invariant features," *Medical Image Analysis*, vol. 17, no. 3, pp. 271–282, 2013.

[36] M. Muja and D. Lowe, "Scalable nearest neighbour algorithms for high dimensional data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2227–2240, 2014.

[37] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–313, 1965.

[38] L. Mercier, R. F. Del Maestro, K. Petrecca, D. Araujo, C. Haegelen, and D. L. Collins, "Online database of clinical mr and ultrasound images of brain tumors," *Medical Physics*, vol. 39, p. 3253, 2012.

[39] D. G. Gobbi, R. M. Comeau, and T. M. Peters, "Ultrasound probe tracking for real-time ultrasound/mri overlay and visualization of brain shift," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1999, pp. 920–927.

[40] M. Letteboer, P. Willems, M. Viergever, and W. Niessen, "Brain shift estimation in image-guided neurosurgery using 3-d ultrasound," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 2, pp. 268–276, 2005.

[41] U. R. Abeyratne, A. P. Petropulu, and J. M. Reid, "Higher order spectra based deconvolution of ultrasound images," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 42, no. 6, pp. 1064–1075, 1995.

[42] P. Hasgall, E. Neufeld, M. Gosselin, A. Klingenböck, and N. Kuster, "IT'IS database for thermal and electromagnetic parameters of biological tissues," *IT'IS Foundation website*, 2015.