Inter-slice Correspondence for
2D Ultrasound-guided Procedures

Matthew Toews[1], Alexandra J. Golby[1], and William M. Wells III[1]

Brigham and Women's Hospital, Harvard Medical School
mt@bwh.harvard.edu, AGOLBY@partners.org, sw@bwh.harvard.edu

Abstract. This paper reports on a new computational methodology, inter-slice correspondence (ISC), for robustly aligning sets of 2D ultrasound (US) slices during image-guided medical procedures. Correspondences are derived from distinctive, local scale-invariant features, which are used in one-to-many matching of US slices in near real-time despite out-of-plane rotation, in addition to global in-plane similarity transforms, occlusion, missing tissue, US plane mirroring, changes in US probe depth settings. Experiments demonstrate that ISC can align manually-acquired US slices without probe tracking information in the context of image-guided neurosurgery, with an accuracy of 1.3mm. A novel reconstruction-without-calibration application based on ISC is proposed, where 3D US reconstruction results are very similar to those obtained via traditional phantom-based calibration.

1   Introduction

Freehand ultrasound is a cheap, safe and portable imaging modality for intra-operative guidance and visualization. Although 3D ultrasound sensors exist, 2D probes remain widely used in clinical practice due to their relatively high image quality and low cost [1]. Intraoperative guidance and visualization frequently require relating 2D US slices acquired from a hand-held probe to the 3D patient anatomy. Most published techniques rely heavily on external tracking systems, e.g. optical or electromagnetic, which provide the rigid location and orientation of a tracking target fixed to the US probe for each US image frame [2, 1, 3, 4].

Although useful, probe tracking systems have important limitations. Calibration procedures are typically required to transform the US image geometry to that of the tracked probe [2], which involve scanning a specialized phantom every time the image-to-probe geometry may have changed, e.g. following instrument sterilization. Even relatively simple calibration methods [3] require additional equipment, procedures and expertise, complicating the clinical work flow. Loss of calibration or tracking (e.g. due changing US probe depth settings, physical accidents [4], loss of line-of-sight) during a medical procedure result in image-guidance failure. Tracker-free matching of US data [5] offers a potential online solution, however current systems do not address out-of-plane alignment. Speckle decorrelation techniques can potentially measure out-of-plane motion [6–8], however practical applications based on manually-acquired data in a clinical setting remain challenging.

This paper proposes a new computational methodology in order to reduce the reliance of US-guidance on probe tracking systems: inter-slice correspondence (ISC). ISC identifies local image correspondences between 2D US slices acquired from different probe positions, robustly and in near real-time, despite out-of-plane rotations, in-plane translation, orientation and scale changes. Experiments demonstrate that ISCs can be accurately computed between manually acquired US slices during neurosurgery without probe tracking. A novel application of ISC is proposed: US volume reconstruction without calibration, where reconstruction results are similar to those obtained via phantom-based calibration.

2   Inter-slice Correspondence

Inter-slice correspondence (ISC) aims to map image structure from a US slice to homologous structure in a sequence of previously acquired 2D US slices. Although ISC can be used in conjunction with probe tracking information, for example in the 3D reconstruction method proposed in the following section, it is achieved without probe tracking information. To be effective in an arbitrary interventional setting, ISC must account for a significant degree of deformed or missing tissue, e.g. in the case of resection, in addition to in-plane geometrical variations such as such as translation, rotation and scaling, e.g. in case of variation in probe position or depth settings. Most importantly, ISC must account for out-of-plane geometrical variations, as US data are generally acquired from different 3D slices through the patient anatomy. To this end, ISC operates by identifying correspondences between informative local patterns or features that can be reliably localized in multiple, approximately intersecting US planes.

ISC makes use of local scale-invariant image features, distinctive image patterns that can be automatically extracted in each US slice in a manner independent of in-plane translation, orientation and scale changes [9, 10]. A feature $f = \{\bar{u}, \theta, \sigma, \bar{a}\}$ is an oriented image patch described geometrically by 2D pixel location $\bar{u} = \{u, v\}$, orientation θ and scale σ, along with a local intensity descriptor $\bar{a}$ used for computing correspondence. A number of scale-invariant feature extractors exist, we adopt the computationally efficient 'SURF' algorithm [10]. Briefly, salient image regions $(\bar{u}, \theta, \sigma)$ are identified based on the local Hessian determinant, and local intensity is encoded via a 64-element descriptor of $\bar{a}$ of Haar wavelet responses.

Robust feature correspondence has been the focus of extensive research in the context of computer vision, where techniques such RANSAC and the Hough transform are commonly used to identify correct feature matches between pairs of projective images, for example [9–11]. The context of US data bears distinct challenges, however, and these techniques cannot be applied directly. To illustrate, in the context of projective image data, the same 3D point is generally observable in images acquired from different sensor geometrical configurations, i.e. camera viewpoints, and relatively dense correspondences can be identified via pair-wise image matching. In contrast, US data are not formed via projection but rather as slices through 3D space, and the same 3D point cannot

generally be observed as slices generally do not pass through precisely the same point. Thus whereas relatively dense correspondences can be identified between an arbitrary pair of projective images, e.g. photographs of the same scene, valid correspondences generally do not exist between an arbitrary pair of US images.

ISC thus adopts a novel robust one-to-many correspondence technique, that aims to identify approximately correct matches between a new US image and an entire sequence of previously acquired US slices, as illustrated in Figure 1. Let I and $\bar{I}'$ represent sets of features extracted in a new US image and in a sequence of previously acquired (i.e. prior or training) US slices, respectively. ISC identifies correspondences (f, f') between all features $f \in I$ and some feature $f' \in \bar{I}'$, by pairing each f with a nearest neighbor (NN) f' based on the Euclidean distance of local feature intensity descriptors $\|\bar{a}, \bar{a}'\|$. Fast approximate NN techniques operate in $O(N \log N)$ computation complexity in the number of training features $N = |\bar{I}'|$, e.g. using the k-d tree data structure [12]. An additional US-related consideration is the case of mirrored image correspondence due to lateral probe flipping, this can be accounted for efficiently by mirroring the geometry and appearance descriptor elements of $f \in I$ prior to correspondence.

Many ISCs are spurious and noisy, and robust estimation is required to identify a reduced set of valid correspondences. The geometrical mapping between ISC arising from multiple US slices is modeled as a global in-plane similarity transform followed by local feature-specific variations. Intuitively, the similarity transform coarsely approximates global image pattern variation between nearby slices in the 3D world, about which variations due to out-of-plane rotation, non-linear tissue deformation, occlusion, etc., are approximated as locally linear via individual correspondences. The Hough transform is used to identify a globally optimal similarity transform, as follows. Let $\Psi = \{d\bar{u}, d\theta, d\sigma\}$ represent the 2D translation, rotation and scaling parameters of an in-plane similarity transform mapping the geometry of f' to that of f. The set $C(\Psi) = \{\ (f, f') :$

$$\mathrm{NN}(a, a') \ \wedge \ \|\bar{u}, \bar{u}' + d\bar{u}\|/\sigma \leq \epsilon_u \ \wedge \ |\log \sigma' d\sigma/\sigma| \leq \epsilon_\sigma \ \wedge \ |\theta - \theta' + d\theta| \leq \epsilon_\theta$$

$\}$ contains the inliers of Ψ, i.e. all nearest neighbor correspondences $\mathrm{NN}(f, f')$ such that the differences in scale-normalized location, scale and orientation of feature f and f' transformed by Ψ are within thresholds $(\epsilon_u, \epsilon_\sigma, \epsilon_\theta)$. An optimal transform $\Psi^* = \underset{\Psi}{\operatorname{argmax}} |C(\Psi)|$ maximizing the cardinality or inlier count of $C(\Psi)$ is identified by testing candidate transforms generated by each correspondence, and the final set of ISC inliers is $C(\Psi^*)$. Thresholds are set empirically, higher values allow a high degree of ISC deformation while increasing the incorrect ISC rate. Values of $(\epsilon_u = 3/4, \epsilon_\sigma = \log 1.5, \epsilon_\theta = 20°)$ are used here for all experiments. A final constraint requires correspondences to occur within sequential frames of $\bar{I}'$, under the assumption of a smoothly varying US probe acquisition trajectory.

The accuracy of ISC with respect to ground truth is evaluated in the context of neurosurgery, with two freehand US sweeps of the same human brain acquired prior to major resection. The BITE data set is used [4], where a calibrated probe tracking system provides a ground truth mapping between US pixel coordinates
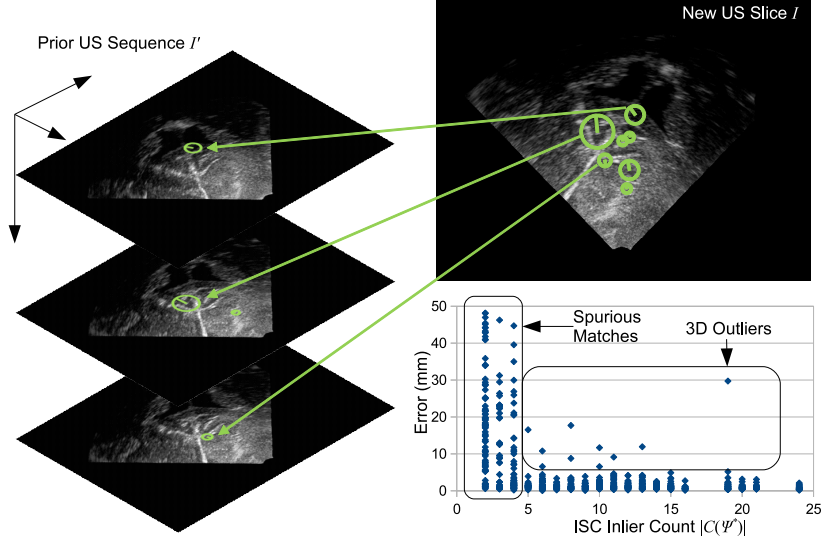
Fig. 1. Illustrating inter-slice correspondences via scale-invariant feature (green circles) from a new US slice (upper right) to multiple US slices in a prior sequence (left). A graph of 3D ISC error vs. ISC inlier count (lower right) reveals low error for inlier counts greater than 4, with the exception of a small number of 3D outliers.

to 3D anatomical locations. Images are 640×480 with 0.2mm resolution, each producing ≈ 800 features. One sweep is arbitrarily chosen as the prior sequence $\bar{I}'$ (240 slices), and ISC is performed to each slice I of a second sweep (144 slices). Figure 1 graphs the error (Euclidean distance) for ground truth 3D locations of ISC correspondences. Error is high for slices where $|C(\Psi^*)| < 5|$, these are generally incorrectly aligned due to insufficient overlap of US content. This is expected, since the volumetric overlap of sweep 2 with sweep 1 here is $\approx 78\%$. Error is low for slices where $|C(\Psi^*)| \geq 5$, with the exception of a small number (2.2%) of *3D Outliers*, i.e. coincidental correspondences that are in-plane inliers but incorrect outliers in 3D. In total, 41% of testing slices I result $|C(\Psi^*)| \geq 5$, with an average ISC error of 1.3mm. Note that the number of ISC inliers can be used to identify US frames with no valid correspondence. 3D outliers have a negligible impact on volume reconstruction in the following section.

3 US Volume Reconstruction Without Calibration

US Volume reconstruction is commonly used for visualizing 3D patient anatomy from a set of tracked, freehand 2D US frames. Reconstruction typically requires estimating the US image-to-probe transform via a specialized calibration procedure and apparatus [2]. Calibration-free methods for computing the image-to-probe transform have been proposed, however they come with limitations (e.g.

no recovery of out-of-plane rotation [13] or scaling parameters [14]), or require a specific US acquisition protocol (e.g. two image sequences separated by a 90° out-of-plane rotation about a fixed 3D point [14]).

ISC allows calibration-free reconstruction from relatively unconstrained US sweeps through arbitrary, textured tissues. Let $\bar{x}_i = [x_i, y_i, z_i, 1]^T$ represent the location of a 3D point in the world in homogenous coordinates, and let $\bar{u}_i = [u_i, v_i, 0, 1]^T$ represent the pixel location of the same point within a 2D US image. Points $\bar{u}_i$ and $\bar{x}_i$ are related by the following equation:

$$\bar{x}_i = T_w T_p T_s \bar{u}_i, \tag{1}$$

where 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, typically 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 $T_{ps} = T_p T_s$ represent the unknown image-to-probe transform matrix.

The goal of 3D reconstruction here is to determine the unknown T_{ps} from ISCs identified between US slices. Let $(\bar{u}_i, \bar{u}_i{}')$ represent corresponding US image points, i.e. the locations of the same 3D point x_i observed in two different US images. From Equation (1), corresponding points $\bar{u}_i$ and $\bar{u}_i{}'$ are related via the following equation:

$$T_w' T_{ps} \bar{u}_i{}' = T_w T_{ps} \bar{u}_i + \eta_i, \tag{2}$$

where η_i is assumed to be zero-mean Gaussian noise. Thus given a set of N correspondences $\{(\bar{u}_1, \bar{u}_1{}'), \ldots, (\bar{u}_i, \bar{u}_i{}'), \ldots, (\bar{u_N}, \bar{u_N}')\}$, the optimal matrix T_{ps}^* minimizing the squared reconstruction error of corresponding points in the 3D world becomes:

$$T_{ps}^* = \underset{T_{ps}}{\mathrm{argmin}} \left\{ \sum_{i=1}^{N} ||T_w' T_{ps} \bar{u}_i{}' - T_w T_{ps} \bar{u}_i||^2 \right\}, \tag{3}$$

T_{ps} is a scaled rigid transform defined by eight intrinsic parameters: a 3D displacement vector $\bar{d} = \{d_1, d_2, d_3\}$, two positive scaling factors (s_u, s_v) 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 (3), 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 the norm $||\bar{\omega}||$ defines the rotation angle.

Optimization thus seeks to identify a parameter vector $\{\log(s_u), \log(s_v), \bar{d}, \bar{\omega}\}$ minimizing Equation (3). A variety of non-linear optimization methods could be brought to bear, here we adopt the Nelder-Mead simplex method [15] which does not require explicit gradient computation and converges reliably in optimization scenarios involving small numbers of parameters. Additionally, rather than minimize the squared error over all correspondences, optimization considers the squared error of the 75% of correspondences with minimum error, in order to reduce the influence of potential 3D outlier ISCs.

Reconstruction is tested using two 900-slice tracked US sequences of the brain, acquired prior to major resection in the context of neurosurgery. The sequences are acquired via arbitrary sweeps along similar trajectories, consisting of translation and minor 3D rotations. ISCs are identified between slices in different sequences as in Section 2, then used to estimate T_{ps} via optimization of Equation (3). An important note is that ISCs must be identified over a degree of out-of-plane rotation in order avoid degeneracy in estimating T_{ps}. For comparison purposes, T_{ps} is also estimated via a standard wire phantom-based calibration procedure (see PLUS perk.cs.queensu.ca). The discrepancy between T_{ps} parameters for the two methods is low: scale parameters (isotropic here) differ by $0.1045 - 0.1064$ or $\approx 1.8\%$, rotation direction cosines differ by $2.8°, 2.4°, 1.6°$. The reconstruction error for 3D correspondences following estimation of T_{ps} is 0.35mm. Volumes are reconstructed from one tracked 900-frame sequence using a simple trilinear interpolation method, using T_{ps} obtained via the calibration-free ISC method and standard calibration. These are shown in Figure 2, note the high degree of visual similarity.

4 Discussion

Inter-slice correspondence (ISC) is presented as a novel method for aligning 2D ultrasound data without external probe tracking. Correspondences can be identified across a wide range of variations, including out-of-plane rotation, and exhibit relatively low error (1.3mm) with respect to ground truth. Correspondence failure can be detected, e.g. in the case where no valid correspondence exists, by the number of ISC inliers. ISC is efficient; 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 via parallelized GPU processing.

The primary potential for ISC is in increasing the robustness of 2D US-guidance by reducing the dependency on external tracking. This is demonstrated in a novel calibration-free US volume reconstruction formulation based on ICSs, where results are similar to those obtained by a phantom-based calibration procedure with an average reconstruction error of 0.35mm. Other potential applications could include automatic verification of calibration during a procedure or navigation and visualization in the case of tracking or calibration failure, these are left for future work.
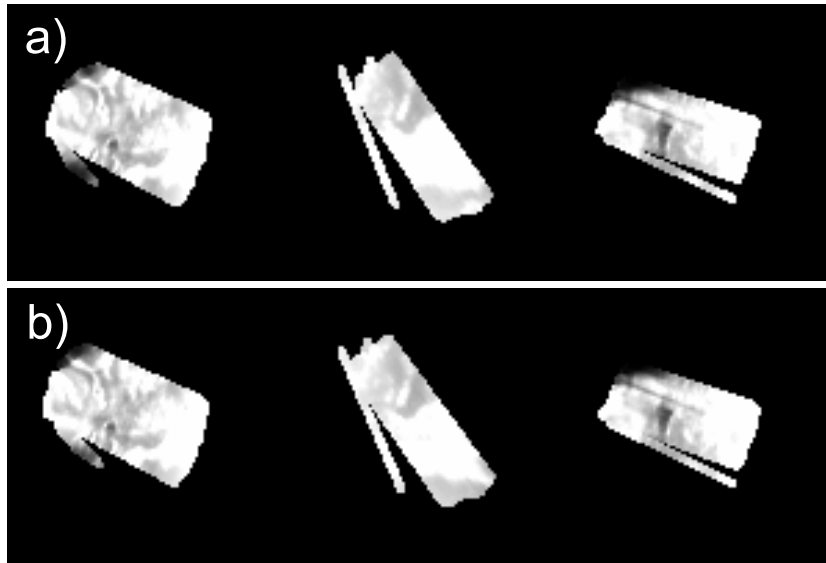
Fig. 2. Axial, sagittal and coronal cross sections of reconstructed US volumes: (a) calibration-free ISC and (b) standard calibration-based. Note the visual similarity.

Several considerations regarding ISC should be noted. The anatomy of interest must contain distinctive, localizable image structure from which distinctive features can be extracted. Qualitatively, we have noted ISC to be effective in US of various organs, including muscle. ISCs cannot be computed if the degree out-of-plane rotation between slices is too great, this breakdown point will be investigated. Future work will include quantifying ground-truth ISC accuracy in intraoperative data, investigating ISC for the purpose of online navigation for image guidance in neurosurgery and in other domains (e.g. abdomen), improvements to tracker-less visualization (e.g. deformable tumor contour alignment) and developing a probabilistic model of ISC.

5 Acknowledgements

References

1. Prager, R.W., Ijaz, U.Z., Gee, A.H., Treece, G.M.: Three-dimensional ultrasound imaging. Journal of Engineering in Medicine **224** (2010) 192–223
2. Mercier, L., Lang, T., Lindsesth, F., Collins, D.L.: A review of calibration techniques for freehand 3-d ultrasound systems. Ultrasound in Medicine & Biology **31**(4) (2005) 449–471

3. Khamene, A., Sauer, F.: A novel phantom-less spatial and temporal ultrasound calibration method. In: MICCAI. (2005) 65–72
4. Mercier, L., Del Maestro, R.F., Petrecca, K., Araujo, D., Haegelen, C., Collins, D.L.: Online database of clinical mr and ultrasound images of brain tumors. Medical Physics **39** (2012) 3253
5. Kwitt, R., Vasconcelos, N., Razzaque, S., Aylward, S.: Recognition in ultrasound videos: Where am i? In: MICCAI. (2012) 83–90
6. Housden, R., Gee, A., Treece, G., Prager, R.: Sensorless reconstruction of freehand 3d ultrasound data. MICCAI 2006 (2006) 356–363
7. Laporte, C., Arbel, T.: Learning to estimate out-of-plane motion in ultrasound imagery of real tissue. MIA **15** (2011) 202–213
8. Afsham, N., Najafi, M., Abolmaesumi, P., Rohling, R.: 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. Volume 8320. (2012) 832017–1
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV **60**(2) (2004) 91–110
10. Bay, H., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. CVIU **110**(3) (2008) 346–359
11. Mirota, D., Uneri, A., Schafer, S., Nithiananthan, S., Reh, D., Ishii, M., Gallia, G., Taylor, R., Hager, G., Siewerdsen, J.: 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 **32**(7) (2013) 1215–1226
12. Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. In: International Conference on Computer Vision. (2009) 331–340
13. Boctor, E., Iordachita, I., Fichtinger, G., Hager, G.: Ultrasound self-calibration. In: SPIE Medical Imaging. (2006) 61412N–61412N
14. Wein, W., Khamene, A.: Image-based method for in-vivo freehand ultrasound calibration. In: SPIE Medical Imaging. Volume 6920. (2008)
15. Nelder, J.A., Mead, R.: A simplex method for function minimization. Computer Journal **7** (1965) 308–313