# RAPID REGISTRATION FOR WIDE FIELD-OF-VIEW FREEHAND 3D ULTRASOUND 

A. H. Gee, G. M. Treece,

R. W. Prager, C. J. C. Cash
and L. Berman
CUED/F-INFENG/TR 447
January 2003

University of Cambridge
Department of Engineering
Trumpington Street
Cambridge CB2 1PZ England

# Rapid Registration for Wide Field-of-View Freehand 3D Ultrasound 

Andrew Gee, Graham Treece, Richard Prager, Charlotte Cash* and Laurence<br>Berman*<br>University of Cambridge<br>Department of Engineering<br>Trumpington Street<br>Cambridge CB2 1PZ<br>*Department of Radiology<br>Addenbrooke's Hospital<br>Cambridge CB2 2QQ


#### Abstract

A freehand scanning protocol is the only way to acquire arbitrary large volumes of 3D ultrasound data. For some applications, multiple freehand sweeps are required to cover the area of interest. Aligning these multiple sweeps is difficult, typically requiring nonrigid imagebased registration as well as the readings from the spatial locator attached to the ultrasound probe. Conventionally, nonrigid warps are achieved through general elastic spline deformations, which are expensive to compute and difficult to constrain. This paper presents an alternative registration technique, where the warp's degrees of freedom are carefully linked to the mechanics of the freehand scanning process. The technique is assessed through an extensive series of in vivo experiments, which reveal a registration precision of a few pixels with comparatively little computational load.


## 1 Introduction

3 D ultrasound $[2,3,10]$ is being applied in an ever growing number of clinical scenarios, often replacing more costly CT and MRI studies. There are essentially two ways to acquire 3D ultrasound data: using a dedicated 3D probe, which is placed on the skin and automatically scans a small, fixed volume beneath; or using a freehand system [3], in which an add-on spatial locator records the position and orientation of a conventional 2 D probe as it is swept over the region of interest, and a 3 D volume is constructed from the resulting B-scans and their relative positions. While most commercial systems employ volume probes for good reasons of ergonomics and practicality, the freehand protocol allows the acquisition of arbitrarily sized volumes and consequently has more potential applications.

For the largest volumes, use of a freehand system is essential, and it might even take several sweeps of the probe to cover the entire region of interest. For instance, three sweeps might be needed for a human liver: one for the left lobe, one for the centre and another for the right lobe. But it is not only large organs, like the liver, which might require multi-sweep acquisition: it is really an issue of organ size with respect to the B-scan resolution. Later in this paper we present high resolution scans of the human breast, used to aid radiotherapy planning following lumpectomy. Use of a small parts probe at a 6 cm depth setting was essential to achieve the necessary axial resolution, but the probe's small lateral footprint ( 3.5 cm ) meant that two sweeps were required to cover the resectioned area.

Multi-sweep freehand acquisition brings with it a problem of registration. Even though the add-on spatial locator, if properly calibrated [13], is able to locate each B-scan within a single sweep fairly accurately, the relative locations of B-scans from distinct, overlapping sweeps are seldom as good. The principal reason for this is displacement and deformation of the underlying anatomy by the contact pressure of the probe. One sweep pushes the anatomy one way, the next sweep pushes it in a different direction. Combining the two sweeps into a single 3D data set results in artifacts akin to motion blur in conventional 2D photography. The subject of this paper is to show how an image-based technique can correct these registration errors, using the spatial locator readings only as a starting point for the alignment process. To be of any clinical relevance, the technique needs to be both fast and accurate.

There is a lot of literature on medical image registration (see $[4,8]$ for recent reviews), but very little of it concerns 3D ultrasound, which differs from the other modalities in three important ways: the often irregular sampling of space by the acquisition protocol, the way the anatomy's appearance depends on the direction of insonification, and the low signal to noise ratio of the images themselves. The following brief literature survey is restricted to the relevant works on nonrigid, intramodal 3D ultrasound registration. Meyer [9] presents an algorithm to register multiple volumes for compound imaging, wide field-of-view mosaicing and sequential scanning studies. A nonrigid, arbitrary, thinplate spline deformation is used for the registration. Krücker [5, 6] later refined the algorithm, using a sub-volume approach to speed up the computation. Pennec [11] uses an alternative radial basis parameterisation to register 3D ultrasound volumes in the course of neurosurgery, with the aim of solving the so-called 'brain shift' problem. As before, there are a large number of arbitrary degrees of freedom to cope with the nonrigid brain deformations. With the same application in mind, Pratikakis [14] introduces a multi-scale registration algorithm set in a many degree-offreedom motion estimation framework. Aiger [1] uses a general elastic transformation to register ultrasound subvolumes for visual simulation applications.

What all these works have in common is their reliance on general elastic deformations to align the ultrasound volumes. This leads to long computation times (invariably of the order of minutes) and also a poorer registration precision, since the problem is typically underconstrained: several different elastic warps of one volume might give a good apparent match to the other volume. Nevertheless, general elastic warps are essential for many of the above applications. By restricting our attention to the wide field-of-view mosaicing application, we show how a deformation with physically motivated degrees of freedom can produce reliable registrations in a fraction of the time required for a general elastic warp. Section 2 describes the registration technique in detail, while Section 3 presents a rigorous in vivo assessment of the technique's accuracy and precision. Conclusions and suggestions for further work can be found in Section 4.

## 2 Aligning multiple freehand sweeps

## Overview

Figure 1 illustrates the difficulty in constructing large field-of-view ultrasound data sets from multiple freehand sweeps. A reslice through the uncorrected data set (top left) shows up serious artifacts. There is a clear misregistration between the two sweeps, and also registration errors between the individual B-scans that make up each sweep. The inter-sweep registration error is caused mostly by the probe pushing the anatomy in different directions in the two sweeps. The inter-B-scan registration errors are mainly due to patient respiration and varying probe contact pressure during the sweep.

Registering sweep B to sweep A in the conventional manner (a spline-based, nonrigid warp) would be both difficult and pointless: difficult because a relatively fine spline grid would be required to align the high frequency inter-B-scan oscillations, leading to an underconstrained registration problem, and pointless because the registered volume would still exhibit the inter-Bscan artifacts. Instead, we propose an approach which abandons free-form spline deformations in favour of constrained, physically motivated corrections.

The first step is to correct the inter-B-scan errors in the two sweeps independently. This is achieved through pair-wise correlation of neighbouring B-scans. B-scan $n+1$ is translated within its plane until it most closely matches B-scan $n$. A further, nonrigid warp is applied in the axial direction to compensate for varying probe contact pressure: B-scans which have been compressed by the probe are retrospectively expanded to match the B-scan acquired with the minimum probe pressure. The intention is to recapture how the anatomy would have appeared if scanned with a non-contact acquisition protocol. Full details of this process can be found in [19]. The effects of the inter-B-scan registration are illustrated at the right of Figure 1. The probe pressure and respiration artifacts are no longer evident, though there remains the clear discontinuity between the two sweeps.


Figure 1: Constrained registration of overlapping sweeps. Instead of opting for an underconstrained, spline-based deformation of sweep B to align it with sweep A, we propose a two stage, physically motivated process. First, the B-scans in each sweep are translated within their planes and stretched in the axial direction to remove probe pressure and respiration artifacts. The two sweeps are then aligned using just a rigid body translation.

To correct the inter-sweep registration, we no longer need so many degrees of freedom. In fact, assuming both sweeps are compressed the same amount by the probe, a rigid body transformation should suffice. Furthermore, considering the mechanics of acquiring two sweeps, it is unlikely that rotation need play a significant part in their registration: unless the anatomy is articulated around some sort of axle, pressing from a different direction is going to translate the anatomy, not rotate it. With this justification (which will be verified in Section 3), we propose just a three degree of freedom translation for the inter-sweep registration, as illustrated at the bottom left of Figure 1.

The success of the inter-sweep registration is contingent on the aforementioned proviso that "both sweeps are compressed the same amount by the probe". To achieve this, each sweep need contain only one B-scan acquired with minimal probe pressure: recall that, in the course of the inter-B-scan registration, the other B-scans in the sweep are expanded to match this one. This is the only constraint imposed on what is otherwise a completely freehand scanning protocol. Even though the focus of this work is on freehand scanning, we should mention that the inter-sweep registration algorithm, as presented in the rest of this section, would also work with multiple volumes acquired using dedicated 3D probes, provided the volumes were acquired with approximately the same contact pressure.

## Inter-sweep registration

A conventional approach to registration would utilise the entire overlap region when estimating the optimal translation between sweeps A and B. However, calculating a similarity measure across the overlap region is computationally expensive and probably unnecessary, given that we are only considering inter-sweep translations and not more general nonrigid transformations. An alternative is to compare sweeps A and B on a single plane within the overlap region, as in Figure 2. This is certainly faster, and conceivably as accurate as using the entire overlap region, which in any case is often narrow. The accuracy assertion will be tested in Section 3. The plane on which sweeps are


Figure 2: Dividing planes. Instead of comparing the sweeps across the entire overlap region, the comparison can be performed more efficiently on a single plane, a dividing plane, which approximately bisects the overlap region. The position and orientation of the plane can be determined automatically from the relative locations of the two sweeps.
compared is termed the dividing plane. Its position and orientation can be computed automatically from the relative locations of the two sweeps [18]: the intention is that it should bisect the overlap region. Dividing planes have other uses in multi-sweep freehand 3D ultrasound [18]. For example, when the entire volume is resliced for visualisation purposes, only data from sweep A is used to the left of the dividing plane, while data from sweep B is used to the right. This avoids having to blend data from both sweeps in the overlap region. This was the approach used to generate the reslices in Figure 1 and subsequent figures.

Hence, the inter-sweep alignment involves finding the $(x, y, z)$ translation of sweep B relative to sweep A which results in the two sweeps appearing most similar when resliced on the dividing plane. There are thus two key ingredients to the alignment process: the selection of a suitable criterion to assess the similarity of the two reslice images, and the design of a search algorithm to find the optimal $(x, y, z)$ as quickly as possible. We use well established solutions for both aspects of the problem, namely a multiresolution search and a choice of three conventional similarity measures. The multiresolution search is described in detail in Figure 3 and its caption. It allows good alignments to be found in the order of a few seconds ( 2 GHz Pentium 4), whereas a single level, high resolution search would take many minutes. Optimal efficiency is achieved by performing the alignment in dividing plane coordinates, with $x$ and $y$ spanning the dividing plane, and $z$ perpendicular to it. Thus, only $z$ offsets involve calculating a new reslice of sweep B: $x$ and $y$ offsets can be evaluated by sliding the same sweep B reslice over the sweep A reslice. The reslices are calculated directly from the B-scans (there is no intermediate voxel representation) using a more efficient version of the algorithm described in [12].

Note that the search is undirected: no use is made of the gradient of the similarity measure to decide where to look next. This is because the similarity function is often multimodal in the search range (see Section 3), so relying on local gradients to find the global optimum is risky. Also, the sampling of sweep B in the $z$ direction is sparse at level 1: a more rigorous approach would be to sample the entire volume, replacing the sparse reslices with volume renderings of the slabs around each reslice. Unfortunately, the computational burden of constructing the volume renderings is unacceptable for a clinically practical tool.

For each potential $(x, y, z)$ alignment of sweep A relative to sweep B, one of the following three functions is calculated to measure the similarity of the dividing plane reslices:

$$
\text { sum of absolute difference } \quad S A D=\frac{\sum|A-B|}{n_{\text {overlap }}}
$$

$$
\begin{aligned}
\text { correlation ratio } & C R=\frac{\operatorname{Var}[E(B \mid A)]}{\operatorname{Var}(B)} \\
\text { mutual information } & M I=H(A)+H(B)-H(A, B)
\end{aligned}
$$

The sum of absolute difference per overlapping pixel is the simplest criterion. The differences in grey levels between corresponding pixels are summed across the overlapping portions of the reslice images, and normalised by the number of overlapping pixels. A low $S A D$ value suggests a good alignment. Note that $S A D$ is unable to recognise image similarities in the event that the grey levels in sweep B are transformed with respect to those in sweep A (as might happen, for instance, if sweep B were acquired from a significantly different direction to sweep A, and suffer different attenuation in the ultrasound propagation path). Also, the absolute difference criterion often favours alignments which minimize the overlap between the two sweeps, since it is relatively easy to find (and align) small, featureless regions at the extremities of the sweeps which happen to look the same. To avoid this, an optional feature of the multiresolution search disregards those alignments which significantly reduce the inter-sweep overlap, as illustrated in Figure 3.

The correlation ratio [15] is a more general criterion which allows for a functional mapping between the intensities in sweeps A and B . It is computed only in the overlap region of the two reslice images, using simple formulae presented in [15]. The correlation ratio copes implicitly with the sweep overlap issue: it does not favour disconnected alignments. Values for the correlation ratio lie in the range 0 to 1 , with 0 implying no functional dependence between A and B , while 1 implies purely deterministic dependence. Thus, high values of $C R$ are indicative of a good alignment.

Mutual information [7, 20] is the most general criterion, allowing for an arbitrary mapping between the intensities in sweeps A and B. The mapping does not have to be linear or even functional, but only predictable [15]. In the $M I$ formulation, $H(X)$ denotes the entropy of image $X$, while $H(X, Y)$ denotes the joint entropy of the two images $X$ and $Y$. Again, the criterion is calculated only in the overlapping portions of the two reslice images. The entropies can be estimated from the joint and marginal intensity histograms of the images [7] (the approach adopted in this paper), or more rigorously using Parzen density estimates [20].

The computation times for the $C R$ and $M I$ criteria are dependent on the number of grey levels. For this reason, the reslice images are quantised to 32 grey levels (from the original 256, by shifting the 8 -bit intensities down to 5 bits) before computation of the similarity function. This not only speeds things up, but also acts as a somewhat arbitrary noise filter, since there is little significance in ultrasound intensity values differing by a few grey levels in 256 . For consistency, the 5-bit quantisation is also used with the $S A D$ criterion, even though this makes no difference to the computation speed.

Figure 4 contains a concise pseudo-code description of the overall search algorithm. Enough detail is provided to allow readers to implement the algorithm should they so wish.

## 3 In vivo assessment of the alignment technique

## Inter-sweep assessment protocol

The rigid, inter-sweep alignment technique described in Section 2 was assessed through a series of in vivo experiments. While in vitro experiments might allow a more precise assessment of registration accuracy, such experiments are of limited value since B-scans of phantoms are totally unlike those of the human body, typically posing far less of a challenge to an image-based registration algorithm.

Two very different freehand data sets were studied. The first was a high resolution scan of the human forearm, with two partially overlapping parallel sweeps (see Figure 5(a)). The Bscans were acquired using an $5-10 \mathrm{MHz}$ linear array probe on a Diasus ${ }^{1}$ ultrasound machine. The depth setting was 4 cm , giving a B-scan resolution of $0.1 \mathrm{~mm} /$ pixel. Every effort was made to ensure that the sweeps were correctly registered at the time of acquisition: the arm was immersed in a warm water bath and scanned from a distance to eliminate contact pressure artifacts, the

[^0]

Figure 3: The multiresolution search algorithm. At level 1, both sweeps are resliced along the dividing plane. The reslice images are smoothed by convolution with a Gaussian of standard deviation 6 pixels and downsampled by a factor of 8 . The resulting images are aligned in the $x-y$ plane by sliding the B reslice over the A reslice and recording the $x-y$ offset which optimizes the similarity criterion: this is illustrated towards the top right of the figure. $z$ alignment is achieved by repeating this process for parallel reslices of sweep B, spaced at 8 pixel increments as shown in the figure. The limits of the $x, y$ and $z$ search ranges are user-defined parameters. The optimal $(x, y, z)$ level 1 alignment is used as the starting point for the level 2 search. This time, the Gaussian has standard deviation 3 pixels and the images are downsampled by a factor of 4 . The $z$ increment for the sweep B reslices is 4 pixels, and the $(x, y, z)$ search ranges are now $\pm 2$ steps on either side of the level 1 optimum. At level 3, the Gaussian has standard deviation 1.5 pixels and the downsample factor is 2. Finally, level 4 employs no downsampling and a Gaussian of standard deviation 0.75 pixels. An optional feature of the search algorithm forcibly preserves the overlap between the two sweeps, by not considering potential alignments which significantly reduce this overlap (eg. the two dashed $z$ alignments at level 1 in the above example). Use of this feature is essential with simple comparison criteria, like the absolute difference per overlapping pixel, which tend to favour disconnected alignments.

```
set centre_x, centre_y & centre_z to 0;
set x_range, y_range & z_range to user-defined values (in pixels);
set Gaussian_standard_deviation to 6;
set downsample_factor to 8;
divide x_range, y_range & z_range by downsample_factor;
for level = 1 to 3 {
    let IM_A = reslice of sweep A on dividing plane;
    smooth IM_A with Gaussian and downsample, then reduce grey levels to 5 bits;
    for z = (centre_z - z_range) to (centre_z + z_range) {
        let IM_B = reslice of sweep B on dividing plane shifted by (z*downsample_factor);
        [optional] if size of IM_B significantly reduced, continue onto next z iteration;
        smooth IM_B with Gaussian and downsample, then reduce grey levels to 5 bits;
        for x = (centre_x - x_range) to (centre_x + x_range)
            for y = (centre_y - y_range) to (centre_y + y_range) {
                compute similarity measure of IM_A and IM_B at relative shift of (x,y) pixels;
                keep track of (optimal_x, optimal_y, optimal_z);
            }
    }
    set (centre_x, centre_y, centre_z) to 2*(optimal_x, optimal_y, optimal_z);
    set x_range, y_range & z_range to 2;
    divide downsample_factor & Gaussian_standard_deviation by 2;
}
```

Figure 4: Pseudo-code for the multiresolution search algorithm.

(a) arm

(b) liver

Figure 5: Scanning patterns for the arm and liver experiments. Each B-scan is shown as an inverted ' $U$ ' shape, with the probe face at the closed end of the ' $U$ '. Typical dividing planes are shown as disks. The arm was scanned in two approximately parallel, linear sweeps. Sweep A comprised 159 B-scans, each of dimension $412 \times 438$ pixels. Sweep B comprised 177 B-scans of the same size. The liver was scanned with one approximately linear sweep, and a second fanshaped sweep which intersected the first at right angles. The linear sweep comprised 188 B-scans of dimension $496 \times 412$ pixels, the fan-shaped sweep comprised 147 B-scans of the same size. See Figures $6(\mathrm{a})$ and $7(\mathrm{a})$ for typical B-scans from the two data sets.

B-scans were transferred from the Diasus to the 3D workstation digitally (eliminating analogue video artifacts), and the probe was tracked using a Polaris ${ }^{2}$ optical spatial locator. With careful calibration of all components, this acquisition system has measurable in vitro precision of less than $0.65 \mathrm{~mm}[17]$. While we would not expect the in vivo sweeps to be registered quite as well as this - the difference between the speed of sound in the water and in the tissue introduces further errors $^{3}$ - any image-based registration correction of less than 1 mm would be strong evidence of the algorithm's accuracy.

The second data set was a low resolution scan of the human liver, acquired using two sweeps, one approximately linear, the other fan-shaped, the two sweeps intersecting at right angles (see Figure $5(\mathrm{~b})$ ). The B-scans were acquired using a 3.5 MHz convex array probe on a Powervision $7000^{4}$ ultrasound machine. The depth setting was 16 cm , giving a B-scan resolution of $0.4 \mathrm{~mm} /$ pixel. Although the subject was not immersed in a water bath, we would not expect significant probe pressure artifacts with such deep, low resolution scans. The B-scans were transferred from the Powervision to the 3D workstation via an analogue video link, and the probe was tracked using a Fastrak ${ }^{5}$ magnetic spatial locator. The precision of this system is far lower than that of the high resolution system used for the arm experiments, and there is little we can say a priori about the correct registration of the two sweeps.

The accuracy and precision of the alignment algorithm were assessed as follows. For each data set, the B-scans belonging to one of the sweeps were retrospectively displaced from their recorded positions by a known, random offset $\boldsymbol{p}$. Each component of $\boldsymbol{p}$ (in spatial locator coordinates) was in the range $\pm 50$ pixels, giving a maximum magnitude displacement of 86.6 pixels. The registration algorithm was then applied to the distorted data set, with a search range of $\pm 87$ pixels in each dividing plane coordinate, and the calculated correction $\boldsymbol{q}$ recorded for each of the three similarity measures described in Section 2. By repeating this process for 100 different random displacements, a set of offsets $\boldsymbol{\Delta}_{i}=\boldsymbol{q}_{i}-\boldsymbol{p}_{i}, i \in\{1 \ldots 100\}$, was obtained. For the arm data set, the mean value

$$
\overline{\boldsymbol{\Delta}}=\frac{1}{100} \sum_{i} \boldsymbol{\Delta}_{i}
$$

is indicative of the registration accuracy: as discussed earlier, we would expect $\|\overline{\boldsymbol{\Delta}}\|$ to be less than 1 mm . For the liver, a visual inspection of the registered data set gives a more qualitative indication of the algorithm's accuracy. The expected deviation

$$
\bar{\delta}=\frac{1}{100} \sum_{i}\left\|\boldsymbol{\Delta}_{i}-\overline{\boldsymbol{\Delta}}\right\|
$$

reflects the precision of the registration technique: ideally, all the $\boldsymbol{\Delta}$ values should be the same, giving $\bar{\delta}=0$.

To prevent the absolute difference criterion from favouring disconnected sweeps, the search range was restricted to consider only those corrections $\boldsymbol{q}$ which maintain the overlap between the two sweeps. To be precise, any potential correction $\boldsymbol{q}$ which reduced the overlap by more than $5 \%$ (compared with the overlap at $\boldsymbol{q}=\mathbf{0}$ ) was disregarded. For consistency, the overlap restriction was also enforced for the correlation ratio and mutual information experiments, even though it makes very little difference with these criteria.

Three sets of experiments were performed on each data set with each similarity criterion. First, the 100 registrations were performed with a fixed dividing plane, the task therefore being to find the rigid displacement which best aligns sweep B with the same, fixed reslice from sweep A. Given a correctly functioning multiresolution search, we would expect identical $\boldsymbol{\Delta}$ values from each of the 100 trials, at least up to the interpolation precision of the reslicing algorithm [12] used to generate the candidate matches from sweep B. The results of this set of experiments are given in Table 1.

[^1]|  | mutual information |  | correlation ratio |  | absolute difference |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  | $\overline{\boldsymbol{\Delta}}$ | $\bar{\delta}$ | $\overline{\boldsymbol{\Delta}}$ | $\bar{\delta}$ |  | $\bar{\delta}$ |
| arm | $(3.8,-3.9,3.3)$ | 0.9 | $(3.5,-4.1,3.0)$ | 0.7 | $(3.4,-4.3,2.6)$ | 0.8 |
| liver | $(-2.2,4.6,-14.0)$ | 0.7 | $(-2.5,5.6,-16.4)$ | 1.4 | $(-1.4,3.2,-17.4)$ | 1.2 |

Table 1: Alignment results for a fixed dividing plane. All $\overline{\boldsymbol{\Delta}}$ and $\bar{\delta}$ values are given in pixels. The expected deviation $\bar{\delta}$ is sub-pixel for the arm and within 1.5 pixels for the liver, irrespective of the similarity criterion. Across all 600 trials, the maximum deviation (relative to the expected deviation) was $4.1 \bar{\delta}$, this being observed for the liver data set with the absolute difference criterion.

|  | mutual information |  | correlation ratio |  | absolute difference |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  | $\overline{\boldsymbol{\Delta}}$ | $\bar{\delta}$ | $\overline{\boldsymbol{\Delta}}$ | $\bar{\delta}$ | $\overline{\boldsymbol{\Delta}}$ | $\bar{\delta}$ |
| arm | $(3.7,-5.1,2.8)$ | 2.6 | $(3.0,-4.6,1.8)$ | 3.5 | $(3.8,-5.6,2.8)$ | 2.4 |
| liver | $(-2.4,4.8,-12.9)$ | 2.9 | $(-2.6,5.2,-14.3)$ | 3.3 | $(-1.8,3.3,-14.4)$ | 3.6 |

Table 2: Alignment results for an automatically positioned dividing plane. All $\overline{\boldsymbol{\Delta}}$ and $\bar{\delta}$ values are given in pixels. Allowing the dividing plane to roam within the overlap region increases the variation of the calculated corrections, but only slightly: the expected registration precision is in all cases less than 4 pixels, and all the $\delta$ values are less than $5 \bar{\delta}$.

The second set of experiments used a different dividing plane for each random displacement, positioned automatically relative to the two sweeps using the procedure described in [18]. Compared with the previous set of experiments, any increased variation in the registration results can be largely attributed to different reslices from sweep A suggesting different alignments with sweep B. This set of experiments therefore tests the assertion that aligning the sweeps based on a single reslice in the overlap region does not significantly prejudice the result compared with aligning the sweeps based on the entire overlap region. The results are given in Table 2.

Finally, the third set of experiments tested the alignment algorithm's resilience to uniform compression of sweep A relative to sweep B, as would be the case were the minimum probe pressure for sweep A significantly different to that for sweep B. While the algorithm is not designed to correct this sort of misregistration, we would hope its performance degrades gracefully, and not catastrophically, with increased pressure mismatch. The B-scans in sweep A were retrospectively compressed in the axial direction to imitate increased probe pressure. The compression was performed as follows: the bottom row of pixels remained in situ, other rows were displaced downwards by an amount depending linearly on the row's distance from the bottom of the B-scan, up to a maximum of $m$ pixels for the top row. Experiments were run with $m=10,20,30$ and 40 pixels, and an automatically positioned dividing plane as in the previous set of experiments. Typical compressed (b) and uncompressed (a) B-scans are shown in Figures 6 (arm) and Figures 7 (liver). The results of the experiments are presented in Table 3.

## Discussion

Table 1 shows that the multiresolution search algorithm is good, though not perfect, at finding the same registration irrespective of where this lies within the search range. The mean offsets found by the three similarity criteria were consistent to within a few pixels, and the 100 offsets for each similarity criterion formed tight clusters around their means, with an expected deviation of less than one pixel from the mean for the arm data set, and less than 1.5 pixels for the liver data set. Some of this variation can be attributed to the nearest neighbour interpolation [16] used to generate the candidate reslices from sweep B which are compared with the fixed reslice from sweep A: these reslices will vary very slightly as sweep B is shifted to different positions. Most of the variation, though, arises from failure of the multiresolution search algorithm. Consider, for example, trial 37 of the liver experiments, which, for the absolute difference criterion, yielded


Figure 6: Arm B-scans. The small tick marks along the edges of the images are millimetres, the larger ticks are centimetres. (a) is an original B-scan, (b) has been compressed by 40 pixels ( 4 mm ) in the axial direction, to simulate the effects of increased probe contact pressure.


Figure 7: Liver B-scans. The tick marks along the edges of the images are centimetres. (a) is an original B-scan, (b) has been compressed by 40 pixels ( 16 mm ) in the axial direction, to simulate the effects of increased probe contact pressure.

|  | mutual information |  |  | correlation ratio |  |  |  | absolute difference |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  | $m$ | $n$ | $\overline{\boldsymbol{\Delta}}$ | $\bar{\delta}$ | $n$ | $\overline{\boldsymbol{\Delta}}$ | $\bar{\delta}$ | $n$ | $\overline{\boldsymbol{\Delta}}$ |  |
| arm | 10 | 100 | $(11.4,-5.4,0.5)$ | 2.7 | 100 | $(11.0,-5.5,-0.2)$ | 2.9 | 100 | $(11.2,-6.4,0.2)$ | 2.8 |
| liver | 10 | 100 | $(2.2,4.2,-12.9)$ | 2.9 | 100 | $(1.9,4.6,-14.4)$ | 3.2 | 100 | $(2.8,2.6,-14.6)$ | 3.8 |
| arm | 20 | 100 | $(17.6,-6.0,-4.1)$ | 3.8 | 100 | $(17.6,-6.1,-4.8)$ | 3.6 | 100 | $(17.4,-6.9,-5.0)$ | 3.6 |
| liver | 20 | 97 | $(6.8,3.5,-13.9)$ | 3.4 | 100 | $(6.5,4.6,-15.3)$ | 3.4 | 100 | $(7.7,2.1,-15.3)$ | 4.0 |
| arm | 30 | 100 | $(24.3,-5.1,-9.3)$ | 5.1 | 100 | $(24.6,-6.0,-9.6)$ | 5.5 | 100 | $(24.0,-6.2,-10.3)$ | 4.6 |
| liver | 30 | 93 | $(10.8,3.4,-16.3)$ | 5.1 | 100 | $(10.3,4.9,-17.3)$ | 4.6 | 100 | $(12.0,2.2,-16.0)$ | 4.5 |
| arm | 40 | 100 | $(31.2,-4.3,-14.3)$ | 6.8 | 100 | $(30.4,-3.5,-16.0)$ | 7.2 | 100 | $(31.1,-5.2,-14.6)$ | 5.4 |
| liver | 40 | 89 | $(14.7,3.8,-16.8)$ | 5.7 | 95 | $(14.4,5.3,-18.4)$ | 4.9 | 100 | $(15.9,2.0,-17.8)$ | 5.6 |

Table 3: Alignment results for $m=10,20,30$ and 40 pixels compression. All $m, \bar{\Delta}$ and $\bar{\delta}$ values are given in pixels. To allow for outliers (gross failures of the alignment algorithm), the results are presented for the $n$ inliers which give relatively tight $\boldsymbol{\Delta}$ clusters with the maximum $\delta$ less than $5 \bar{\delta}$. The frequency of outliers increases with the degree of compression and the generality of the similarity criterion.

|  | full search range |  | restricted search range |  |
| :--- | :--- | :--- | :--- | :--- |
|  | offset | $S A D$ | offset | SAD |
| level 1 | $(8,48,-24)$ | 1.41 | $(16,48,-8)$ | 1.52 |
| level 2 | $(12,48,-20)$ | 1.55 | $(16,48,-12)$ | 1.58 |
| level 3 | $(12,50,-20)$ | 1.73 | $(14,48,-12)$ | 1.72 |
| level 4 | $(13,49,-20)$ | 1.95 | $(13,48,-12)$ | 1.94 |

Table 4: The failure mode of trial 37 of the liver experiments. With the full search range, the low resolution search finds a false minimum absolute difference at offset $(8,48,-24)$ pixels (in dividing plane coordinates). From this poor start, the search algorithm is unable to find the high resolution minimum at $(13,48,-12)$ pixels. Narrowing the search range, however, allows the algorithm to converge to the true minimum.
the most deviant result, $4.1 \bar{\delta}$ from the mean. The failure mode is illustrated in Table 4 , which shows the corrections (in dividing plane pixel coordinates) and $S A D$ measures at each stage of the multiresolution search. With the full search range of $\pm 35 \mathrm{~mm}$ in each direction, the algorithm heads off in the wrong direction and cannot recover at lower levels of the search. By limiting the search range, the algorithm can be persuaded to find the optimal high resolution solution.

This failure is a consequence of the multimodal similarity function within the search range: there would be no problem were the function locally convex. But in this case there is a global minimum of 1.94 at $(13,48,-12)$ pixels, and a local minimum of 1.95 at $(13,49,-20)$ pixels. Unfortunately, the search is easily drawn towards the wrong solution at the low resolution stages of the search. Figure 8 illustrates this phenomenon with a simple, one dimensional example.

Despite the occasional failure of this nature, the multiresolution search algorithm works to sub-pixel accuracy in the vast majority of cases, and is easily corrected (by manually selecting a different similarity criterion or narrowing the search range) when it fails. Moreover, the search ranges for these experiments are unusually large to allow for the artificial displacement of sweep B by up to 86 pixels: we would not expect to require such a large correction for untampered in vivo data, so we could restrict the search range to reduce the chances of the algorithm getting attracted by a local minimum. In summary, the algorithm would appear to offer a sensible compromise between accuracy, speed and autonomy.

Table 2 justifies the use of a single dividing plane on which to align the two sweeps, instead of the entire overlap region. Allowing the dividing plane to roam within the overlap region does increase the variation of the calculated registrations, but only slightly. The table indicates that we can expect to arrive at the same alignment, irrespective of the initial registration error and the consequential positioning of the dividing plane, to within a few pixels every time. The most extreme variations from the mean (as before, caused by failure of the multiresolution search) were of the order $5 \bar{\delta}$. The accuracy of the alignment technique is confirmed by the arm experiments, with $\|\bar{\Delta}\|=0.7 \mathrm{~mm}$ for the mutual information experiments, 0.6 mm for the correlation ratio and 0.7 mm for the absolute difference. These are all well within the 1 mm limit expected for an accurate registration.

Figures 9 and 10 illustrate, for each similarity criterion, the effects of the mean and most deviant corrections on typical reslices through the two data sets. They also show the same reslices for no correction and for the most extreme random offset applied during the 100 trials. For all three similarity criteria, the mean correction is certainly an improvement over no correction for the liver data set, and perhaps an improvement for the arm data set. The most deviant corrections appear almost identical to the mean corrections, even though they are up to 18 pixels away (though not necessarily in the plane of the reslice) in the case of the arm. This illustrates a difficulty in registering data with strong continuous features aligned in the same direction, like the tissue layers in the arm: the registration quality is not very sensitive to inter-sweep translations perpendicular to the dividing plane, giving rise to local minima in the similarity functions which can easily trap a multiresolution search algorithm.


Figure 8: Limitations of multiresolution search algorithms. (a) shows the $S A D$ measure for two sine waves of amplitude one and period 125 pixels, for offsets in the range -58 to +58 pixels. The sine waves are identical, except one has a spike of amplitude 30 at pixel 20 , while the other has a spike of the same amplitude at pixel 40 . Both sine waves were smoothed by convolution with a Gaussian of standard deviation 0.75 pixels before calculation of the absolute difference. According to this criterion, the best registration is achieved through a shift of zero, which aligns the two sine waves. In a multiresolution search, however, we would start by registering more heavily smoothed, subsampled versions of the two signals. (b) shows the absolute difference per overlapping pixel between the two signals smoothed by a Gaussian of standard deviation 3 pixels and subsampled by a factor of four. The best registration is now the one which aligns the two spikes, a shift of 5 downsampled pixels, or 20 true pixels. From this false start, a multiresolution search algorithm cannot find the high resolution optimal shift of zero pixels.


Figure 9: Reslices through the arm data set. Of the 100 trials with a roaming dividing plane, (a) shows the maximum applied displacement of $(-42.8,-49.6,42.3)$ pixels, resulting in a gross misregistration of the two sweeps. (b) shows the same reslice with no inter-sweep correction: as explained in the text, great care was taken to ensure that the sweeps were accurately registered at the time of acquisition, and this reslice confirms this to be the case. (c) shows the mean correction $\bar{\Delta}$ found using the mutual information criterion, (d) shows the mean correlation ratio correction and (e) shows the mean absolute difference correction. Each shows perhaps a slight improvement in the alignment compared with (b). (f) shows the most deviant mutual information correction (10.9 pixels away from the mean), (g) the most deviant correlation ratio correction (17.6 pixels away) and (h) the most deviant absolute difference correction (10.4 pixels away).


Figure 10: Reslices through the liver data set. Of the 100 trials with a roaming dividing plane, (a) shows the maximum applied displacement of $(-42.8,-49.6,42.3)$ pixels, resulting in a gross misregistration of the two sweeps. (b) shows the same reslice with no inter-sweep correction: some misregistration is apparent as discontinuities in the dark vessels and the bright diaphragm. (c) shows the mean correction $\overline{\boldsymbol{\Delta}}$ found using the mutual information criterion, (d) shows the mean correlation ratio correction and (e) shows the mean absolute difference correction. Each shows a significant improvement in the alignment compared with (b). (f) shows the most deviant mutual information correction (8.5 pixels away from the mean), (g) the most deviant correlation ratio correction ( 7.2 pixels away) and (h) the most deviant absolute difference correction (8.4 pixels away).


Figure 11: Reslices through the 40 pixel squashed arm data set. Sweep A (on the right of the reslices) was synthetically squashed by 40 pixels ( 4 mm ). (a) shows the uncorrected data set: the extent of the squash is evident by comparison with Figure 9 (b). (b) shows the mean correction found using the mutual information criterion. Even though a rigid registration is unable to correct the relative compression between the two sweeps, the alignment procedure has found a good compromise, under-correcting at the top and over-correcting at the bottom.


Figure 12: Reslices through the 40 pixel squashed liver data set. Sweep A (on the right of the reslices) was synthetically squashed by 40 pixels ( 16 mm ). (a) shows the uncorrected data set: the extent of the squash is evident by comparison with Figure 10 (b). (b) shows the mean correction found using the correlation ratio criterion. Even though a rigid registration is unable to correct the relative compression between the two sweeps, the alignment procedure has found a good compromise, aligning the principal distinguished features of the data set.

Finally, Table 3 confirms the graceful degradation of the alignment procedure in the presence of the principal misregistration which it was not designed to correct: uniform compression of sweep A relative to sweep B. The mean corrections $\overline{\boldsymbol{\Delta}}$ vary as the algorithm attempts to find a compromise alignment between two sweeps which don't match. Examples of these mean corrections, for the most severely deformed data sets, can be found in Figures 11 and 12. The alignment precision worsens as the sweeps become less similar: this is to be expected, since we can no longer expect a dominant global optimum in the similarity criterion. Indeed, we may find significant local optima some distance away from each other, especially with the mutual information and correlation ratio criteria, which are able (by design) to find similarities between images even when the grey levels are very different. This was indeed the case, with both of these similarity criteria occasionally finding totally inappropriate corrections, more than $5 \bar{\delta}$ away from the mean $\bar{\Delta}$. As before, this was a fault of the multiresolution search algorithm: the false optima were dominant at the low resolutions, but not at the high resolutions. The difference now is that these optima may be a long way from the 'best' solution. It is reassuring to note that the mutual information criterion, the most general of the three, is most susceptible to outliers of this sort, followed by the correlation ratio, which is the second most general criterion. The absolute difference criterion did not arrive at any grossly incorrect registrations, since it is unable to find similarities between the two sweeps except in the vicinity of the mean correction: its similarity function is more convex than the others'.

While the absolute difference criterion proved reliable and robust for the liver and arm experiments, one can envisage scenarios where a more general criterion is required to achieve a proper alignment, especially when the two sweeps are from very different insonification directions, so that the same features look different in each sweep. Also, as noted in Section 2, the absolute difference criterion often favours alignments which minimize the overlap between the two sweeps. Even though this effect was suppressed in the liver and arm experiments by disregarding registration hypotheses which significantly reduced the sweep overlap, this is an inelegant complication which is not required when using the other two criteria.

The results of the arm and liver experiments point to the need for a graphical user interface (GUI) to verify and control the alignment process. Even though the registration works well, without user interaction, in the vast majority of cases, it is not infallible and there is a need for user verification and, where necessary, correction (usually by selecting an alternative alignment criterion, or narrowing the search range to avoid a local minimum). A suitable GUI is shown in Figure 13: this has been fully implemented in the Stradx freehand 3D ultrasound system ${ }^{6}$, and will be made available for free download with the next release. The GUI allows concatenation of inter-sweep alignments when there are more than two sweeps. An example of this can be found in Figure 14.

## Combined inter-sweep and inter-B-scan alignment

Finally, we present a qualitative assessment of the combined inter-sweep and inter-B-scan registration protocol in the context of a clinical application. The uncorrected images in Figure 15 show reslices through freehand data sets of the human breast, used to aid radiotherapy planning following lumpectomy. Use of a small parts probe at a 6 cm depth setting was essential to achieve the necessary axial resolution, but the probe's small lateral footprint ( 3.5 cm ) meant that two sweeps were required to cover the resectioned area. The uncorrected reslices exhibit significant probe pressure and respiration artifacts (inter-B-scan), and also a clear misregistration between the two sweeps.

The figure shows successful removal of both types of registration error, using the inter-Bscan alignment technique described in [19] followed by inter-sweep alignment as described in this paper. Significantly, the scans were acquired before the overall registration protocol was devised, so the sonographer was not making any effort to match the minimum contact pressures in each sweep. This lends credence to the clinical practicality of the technique. Registration times were

[^2]

Figure 13: The Stradx inter-sweep alignment GUI. The GUI allows the user to select the pair of sweeps to be aligned (for very large data sets, as in Figure 14, there may be more than two sweeps), the alignment criterion to use, the search ranges in the three principal directions and whether to forcibly preserve the overlap between the two sweeps. The two reslice images are shown superimposed in red and green in the main display area, both pre- and post-alignment. The effects of the alignment can be observed in all of Stradx's other display windows by pressing the 'Apply' button. Should the registration be incorrect, the same button allows the registration to be removed before another is calculated.


Figure 14: Concatenation of alignments for a three-sweep scan. (a) shows a reslice through a freehand scan of a human liver. Three sweeps were required: one for the left lobe, one for the centre and another for the right lobe. Minor misregistration is evident at the two inter-sweep boundaries. (b) shows the same reslice after inter-sweep alignment. The GUI in Figure 13 was used to calculate a registration between sweeps 1 and 0 , apply this registration to the data set, then calculate a second registration between sweeps 2 and 1 . Mutual information was used for the first alignment, the correlation ratio for the second.


Figure 15: Combined inter-B-scan and inter-sweep registration. The figure shows reslices through two data sets of the human breast, part of an experimental radiotherapy planning protocol following lumpectomy. The uncorrected reslices exhibit probe pressure and respiratory artifacts, as well as a clear misregistration between the two sweeps. Both types of artifact are effectively removed by the combined registration process.
around one minute for the nonrigid inter-B-scan registration, and a couple of seconds for the rigid inter-sweep registration ( 2 GHz Pentium 4).

The lumpectomy application allows us to examine the hypothesis that the overall registration reveals the undeformed anatomy, as would be observed with a non-contact imaging modality. Figure 16 shows a reslice through the 3D ultrasound breast data superimposed on a corresponding reslice through a CT scan acquired at the same time. The two data sets were aligned using external fiducial points, not any sort of image-based registration. The excellent alignment is testament to the careful calibration of the freehand ultrasound system [17] as well as the effectiveness of the registration algorithms. It must be stressed, however, that this is a preliminary, one-off 3DUSCT registration: further work is required to rigorously validate the combined inter-B-scan and inter-sweep registration process, and its implications for the veracity of the 3 D ultrasound data set.

## 4 Conclusions

We have presented and assessed an algorithm for the alignment of multiple freehand 3D ultrasound sweeps. Used in conjunction with an earlier inter-B-scan registration technique, and with minimal impact on the freehand scanning protocol, it allows faithful visualisation of arbitrarily large data blocks. Speed and precision are acceptable for the vast majority of clinical applications, with a couple of seconds of computation achieving alignments to within a few pixels' precision.


Figure 16: Comparison of corrected 3D ultrasound data with CT data. A reslice through the breast data is shown superimposed on a corresponding reslice through CT data of the same breast, acquired at approximately the same time. The two modalities were aligned using external fiducial points, not any sort of image-based registration. The figure reveals excellent correspondence between the corrected ultrasound data and the non-contact CT image. The area targeted for radiotherapy is outlined in white.

Further work is required to fully automate the process, since at the moment some human interaction is required to visually verify the alignment and re-run the algorithm, with different parameters, in the rare cases of failure. Despite this limitation, the technique is already being used, with considerable success, in an experimental radiotherapy planning application.

## Acknowledgements

The authors would like to thank Charlotte Coles and Andrew Hoole for providing the CT data in Figure 16.

## References

[1] D. Aiger and D. Cohen-Or. Mosaicing ultrasound volumes for visual simulation. IEEE Computer Graphics and Applications, 20(2):53-61, March/April 2000.
[2] A. Fenster, D. B. Downey, and H. N. Cardinal. Three-dimensional ultrasound imaging. Physics in Medicine and Biology, 46:R67-R99, 2001.
[3] A. Gee, R. Prager, G. Treece, and L. Berman. Engineering a freehand 3D ultrasound system. Pattern Recognition Letters, 24(4-5):757-777, February 2003.
[4] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. Physics in Medicine and Biology, 46:R1-R45, 2001.
[5] J. F. Krücker, G. L. LeCarpentier, J. B. Fowlkes, and P. L. Carson. Rapid elastic image registration for 3D ultrasound. IEEE Transactions on Medical Imaging, 21(11):1384-1394, November 2002.
[6] J. F. Krücker, C. R. Meyer, G. L. LeCarpentier, J. B. Fowlkes, and P. L. Carson. 3D spatial compounding of ultrasound images using image-based nonrigid registration. Ultrasound in Medicine and Biology, 26(9):1475-1488, 2000.
[7] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodaility image registration by maximization of mutual information. IEEE Transactions on Medical Imaging, 16(2):187-198, April 1997.
[8] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. Medical Image Analysis, 2(1):1-36, March 1998.
[9] C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, G. L. LeCarpentier, J. B. Fowlkes, M. A. Roubidoux, and P. L. Carson. Semiautomatic registration of volumetric ultrasound scans. Ultrasound in Medicine and Biology, 25(3):339-347, 1999.
[10] T. R. Nelson and D. H. Pretorius. Three-dimensional ultrasound imaging. Ultrasound in Medicine and Biology, 24(9):1243-1270, 1998.
[11] X. Pennec, P. Cachier, and N. Ayache. Tracking brain deformations in time-sequences of 3D US images. Pattern Recognition Letters, 24(4-5):801-813, February 2003.
[12] R. W. Prager, A. H. Gee, and L. Berman. Stradx: real-time acquisition and visualization of freehand three-dimensional ultrasound. Medical Image Analysis, 3(2):129-140, 1999.
[13] R. W. Prager, R. N. Rohling, A. H. Gee, and L. Berman. Rapid calibration for 3-D freehand ultrasound. Ultrasound in Medicine and Biology, 24(6):855-869, 1998.
[14] I. Pratikakis, C. Barillot, P. Hellier, and E. Mémin. Robust multiscale deformable registration of 3D ultrasound images. International Journal of Image and Graphics, in press.
[15] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In Medical Image Computing and ComputerAssisted Intervention - MICCAI'98, pages 1115-1124, Cambridge MA, October 1998. LNCS 1496, Springer.
[16] R. N. Rohling, A. H. Gee, and L. Berman. A comparison of freehand three-dimensional ultrasound reconstruction techniques. Medical Image Analysis, 3(4):339-359, 1999.
[17] G. M. Treece, A. H. Gee, R. W. Prager, C. J. C. Cash, and L. Berman. High definition freehand 3D ultrasound. Ultrasound in Medicine and Biology, in press.
[18] G. M. Treece, R. W. Prager, A. H. Gee, and L. Berman. 3D ultrasound examination of large organs. Medical Image Analysis, 5(1):41-54, 2001.
[19] G. M. Treece, R. W. Prager, A. H. Gee, and L. Berman. Correction of probe pressure artifacts in freehand 3D ultrasound. Medical Image Analysis, 6(3):199-215, 2002.
[20] P. Viola and W. M. Wells. Alignment by maximization of mutual information. International Journal of Computer Vision, 24(2):137-154, 1997.


[^0]:    ${ }^{1}$ Dynamic Imaging Ltd., http://www.dynamicimaging.co.uk/

[^1]:    ${ }^{2}$ Northern Digital Inc., http://www.ndigital.com/
    ${ }^{3}$ For the comfort of the person being scanned, the temperature of the water bath was kept below $50^{\circ} \mathrm{C}$, the temperature at which the speed of sound in water is the same as in average human tissue.
    ${ }^{4}$ Toshiba Corporation, http://www.toshiba.com/
    ${ }^{5}$ Polhemus Inc., http://www.polhemus.com/

[^2]:    ${ }^{6}$ http://svr-www.eng.cam.ac.uk/~rwp/stradx/.

