
**SENSORLESS RECONSTRUCTION
OF UNCONSTRAINED FREEHAND
3D ULTRASOUND DATA**

R. J. Housden, A. H. Gee,
G. M. Treece and R. W. Prager

CUED/F-INFENG/TR 553

May 2006

University of Cambridge
Department of Engineering
Trumpington Street
Cambridge CB2 1PZ
United Kingdom

Email: rjh80/ahg/gmt11/rwp@eng.cam.ac.uk

Sensorless reconstruction of unconstrained freehand 3D ultrasound data

R. James Housden, Andrew H. Gee, Graham M. Treece and Richard W. Prager

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

Abstract

Freehand 3D ultrasound can be acquired without a position sensor by finding the separations of pairs of frames using information in the images themselves. Previous work has not considered how to reconstruct entirely freehand data, which can exhibit irregularly spaced frames, intersecting frames, non-monotonic out-of-plane probe motion and significant in-plane motion. This paper presents reconstruction methods that overcome these limitations and are able to robustly reconstruct unconstrained freehand data. The methods are assessed on freehand data sets and compared to reconstructions obtained with a position sensor.

1 Introduction

3D ultrasound [3, 13] is an emerging medical imaging modality with many potential applications [6]. The data can be acquired using dedicated 3D probes incorporating an oscillating head which sweeps the scan plane over a fixed volume of interest. The alternative, freehand approach involves the clinician manually sweeping a conventional probe over the target: by attaching a position sensor to the probe, each B-scan can be labelled with its position and orientation. The B-scans thus form a 3D data set which can be visualised and processed in a number of ways to extract clinically useful information. The freehand approach offers the advantages of arbitrary acquisition volumes, with translation as well as rotation of the scan head, low cost, and compatibility with the full palette of existing 2D probes. There are also disadvantages, including slow acquisition (freehand is not suitable for 4D scanning) and image distortion caused by varying probe contact pressure [16].

The oscillating head approach is the current focus of commercial activity. Perhaps the greatest barrier to more widespread uptake of freehand scanning is the need for the add-on position sensor. This requires careful end-user calibration [12] and imposes cumbersome constraints on the scanning protocol. Typically, the operator must maintain a clear line of sight between the probe and the position sensor’s fixed base station, and must be careful not to stray outside the sensor’s operating region. Also, small-scale jitter in the position sensor readings can lead to visible artefacts in images derived from the reconstructed data. It is therefore not surprising that researchers have been investigating alternative, freehand acquisition techniques that dispense with the position sensor.

One such approach exploits image-based registration to find the offsets between adjacent frames using information in the images themselves. The in-plane offsets (i.e. in the axial and lateral directions) can be found using standard image registration and motion tracking techniques [7, 18]. In the out-of-plane, elevational direction, speckle decorrelation [1, 11, 17] can be used to estimate the absolute separation between two patches of data, as shown in Figure 1. Since the ultrasound beam is not perfectly focused, the backscattered signal at any point depends on the scatterers in a certain volume, the so-called “resolution cell”, around that point. If two points on neighbouring B-scans have overlapping resolution cells, the two signals will be correlated. The correlation depends on the amount of overlap and therefore on the separation between the two points. Therefore, by calculating the correlation between the two patches of data, it is possible to infer their elevational separation using a precalibrated decorrelation curve. Given at least three non-collinear patch pairs, the elevational separation, tilt and yaw can be estimated for the frame pair.

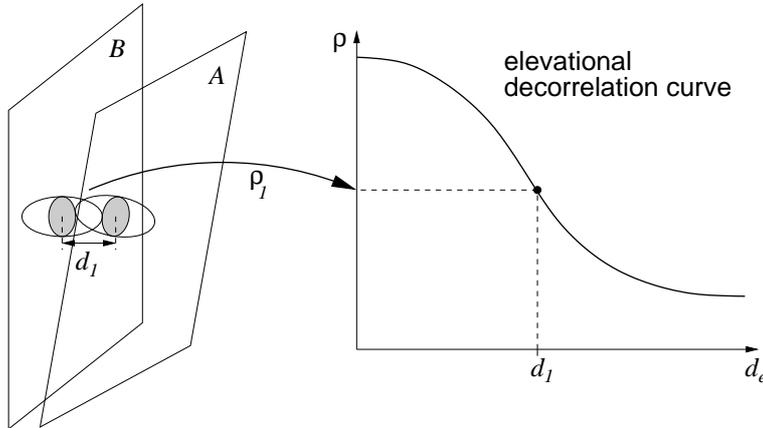


Figure 1: **Speckle decorrelation.** Consider two patches of data on nearby image frames A and B . The correlation between the data varies with the resolution cell overlap and therefore with their separation. The precalibrated decorrelation curve can be used to look up a distance d_1 , given the correlation ρ_1 .

There are several difficulties with this approach that impede its direct application to in vivo freehand data. First, the theory holds only for fully developed speckle. Real tissue contains regions of coherent scattering which decorrelate at a slower rate than regions of fully developed speckle. In recent work, we have shown how to allow for this in an adaptive speckle decorrelation framework [5]. Secondly, there are other causes of speckle decorrelation (noise, in-plane motion, transducer rotation, tissue compression, physiological motion) which might be confused for elevational motion. While most of these remain a matter for future research, we have studied the effects of in-plane motion in detail [10]. We showed how to estimate in-plane motion to sub-sample accuracy using novel Gaussian interpolation techniques. We also showed how to correct the elevational correlation values to allow for the in-plane motion.

Another difficulty is that speckle decorrelation can only provide the absolute value of the separation between two patches. There is a direction ambiguity: is the patch on B-scan A to the left or right of the patch on B-scan B ? All previous work has ignored this problem and instead constrained the probe motion to be monotonic with no intersecting frames, in which case all patch separations can safely be assumed to be in the same direction. In this paper, we present for the first time sensorless reconstructions of truly unconstrained freehand data. Building on preliminary work [9] with regularly spaced frames, we present new algorithms to allow for arbitrary hand-held probe motion, including cases where the probe reverses direction mid-way through the scan, and axial rotation scans where the frames intersect one another.

The paper is organised as follows. In Section 2, we present the robust reconstruction technique, which is conveniently broken into three modules: frame selection, deducing intersections and non-monotonic reconstruction. In Section 3, we describe experiments with a hand-held probe designed to test the capabilities and limitations of the new algorithms. Finally, in Section 4 we draw some conclusions and suggest avenues for further work.

2 Robust reconstruction

Here is a broad overview of the sensorless freehand 3D reconstruction technique.

1. Each image frame is divided into an 8 column \times 12 row grid of patches. For the radio frequency (RF) acquisition platform described in Section 3, this gives more than 5000 RF samples per patch, sufficient to calculate meaningful second order statistics.

2. For each frame A , we calculate the set of frames B_i within range of A . By “within range”, we mean that the majority of the patches on B_i are offset by an amount within the search range of the in-plane alignment [10] and are close enough elevationally to correlate to some extent with a corresponding patch on A . Frames that are elevationally too distant from A will be totally decorrelated and not allow any meaningful estimation of elevational or in-plane separation.
3. For all frame pairs within range of each other, we estimate the spatial separation between corresponding patches as three degrees of freedom: axial, lateral and elevational translation. This is performed as accurately as possible, making use of the technique in [5] to compensate for coherent scattering and the methods in [10] to calculate the lateral shift and subsequent correlation to sub-sample precision.
4. Using the patch-wise separation estimates calculated in the previous stage, we select several interleaved sets of coarsely spaced frames, using the method described in Section 2.1. Each set of frames spans the entire data set, but each will be reconstructed independently. This allows a robust consensus approach for the final reconstruction, with any outlier sets rejected.
5. For each frame pair for which we have separation estimates, we next check to see whether the frames intersect one another, using the algorithm described in Section 2.2. If there are any intersections, the signs of the elevational separation estimates are adjusted appropriately.
6. Each set of frames is reconstructed three-at-a-time, instead of the usual pair-wise approach, using the algorithm in Section 2.3. This allows any non-monotonic probe motion to be detected. The independent reconstructions are combined into a final result using a consensus check, as described in Section 2.3

2.1 Frame selection

There are two requirements to consider when selecting interleaved sets of frames for independent reconstruction. First, we would like consecutive frames in each set to be quite widely spaced, though still within range. This is so that we can operate on the steep part of the decorrelation curve in Figure 1, where small errors in the correlation ρ do not produce large errors in the elevational separation estimates d_e [8, 14]. Secondly, we would like the sets to be mutually independent, with no single frame belonging to more than one set. When these two requirements conflict, we should accept sub-optimal spacing in order to preserve the independence of the sets.

Consider initially the problem of choosing only one set of frames from the sequence, so that independence is not an issue. We have a numbered sequence of irregularly spaced frames from which we need to select a coarsely spaced sequence with approximately even spacing, as shown in Figure 2. The task is therefore to determine the coarse sequence using the available patch-wise elevational separation estimates.

In order to detect non-monotonic motion, each frame in the coarse sequence must be within range of at least the next two frames. The reason for this is explained in detail in Section 2.3. We therefore calculate two pieces of range information for each frame, as shown in Figure 2. Firstly, from each frame f , we can move forward through the sequence by a maximum number of steps r_{\max} while remaining within range. This value has already been determined when the patch separation estimates were calculated. We also need another value, r_{half} , which is less than r_{\max} and determines another frame, f_h , in the sequence between f and f_m . f_h divides the offset between frames f and f_m into two parts. It is important for tracking non-monotonic motion that both of these offsets are as large as possible. f_h is therefore chosen by maximising the smaller of the two halves of the offset. Figure 3 shows three examples of the ideal position of f_h between f and f_m . For a simple, linear monotonic frame pair, this is at half the distance between them. The same is true of intersecting frames. For a non-monotonic sequence, however, f_m could be elevationally closer to f than some of the frames between them in the sequence. Half the distance between f and f_m would then not give the largest possible pair of offsets.

(a) single coarse sequence

```
choose a starting frame f_n;
coarse sequence length, l = 0;
N = number of frames in data set;

for frame, f = 1 to N {

    if f = f_n {

        add frame f_n to coarse sequence;
        l = l + 1;

        if l = 1
            f_n = f_n + r_half(f_n);

        else
            b_1 = the last frame added to the sequence;
            b_2 = the frame one coarse step back from b_1;
            max_limit = b_2 + r_max(b_2);
            half_limit = b_1 + r_half(b_1);
            f_n = min (max_limit, half_limit);

    }
}
```

(b) multiple coarse sequences

```
choose M = number of coarse sets;
choose a different starting frame, f_n,s, for each coarse set, s;

for frame, f = 1 to N {

    for set, s = 1 to M {

        if f = f_n,s
            update set s as above;

    }

    if f has been used by more than one set
        adjust coarse sequences to avoid repetition (see text);

}
```

Figure 4: Pseudo-code for choosing coarse sequences of frames.

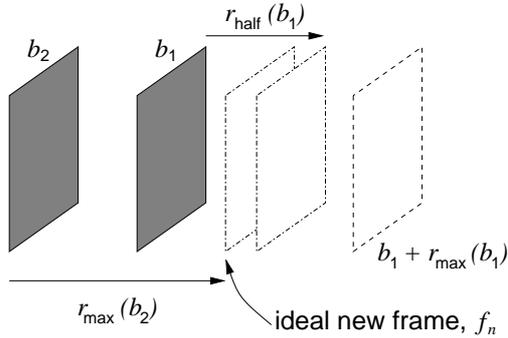


Figure 5: **Choosing the next frame in the sequence.** The ideal next frame depends on two maximum limits: it must be within range of both the preceding frames, b_1 and b_2 , and leave room for another frame within range of b_1 . We therefore choose the smaller of $b_2 + r_{\max}(b_2)$ and $b_1 + r_{\text{half}}(b_1)$.

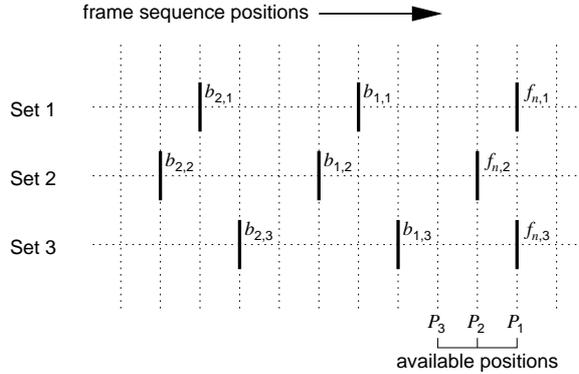


Figure 6: **Repeated frames.** Sets 1 and 3 both expect the same frame as their next frame. We can avoid repeating this frame by moving both $f_{n,1}$ and $f_{n,2}$ back by one position. Note that the frame positions in this diagram represent their position in the numbered frame sequence and not necessarily their actual locations in space.

slots. In more extreme cases, there might not be enough slots to play with. We must then accept a repeated frame as unavoidable, but continue to adjust any further repetitions. Having established how many frames to adjust, we distribute the frames evenly into the available slots, minimising the number of times each position is used. We order the frames according to the order of the preceding frames in the sets. In our example, this means that both frames $f_{n,2}$ and $f_{n,1}$ are moved back one position, while $f_{n,3}$ is unchanged.

2.2 Intersecting frames

Consider a pair of frames which happen to intersect along a line parallel to the x axis, as shown in Figure 7. Speckle decorrelation provides an 8×12 grid of estimates of the absolute elevational z offsets. If we were to look at one y column of the grid, we would expect to see a y - z distribution something like the one in Figure 7(a). All previous work on sensorless reconstruction has assumed no frame intersections: a least squares fit to the z values would then position frame B roughly parallel to frame A , as shown by the dotted line in Figure 7(a). To allow for intersections, we need to deduce the signed elevational z offsets. If we assume that the 8×12 grid of offsets should define

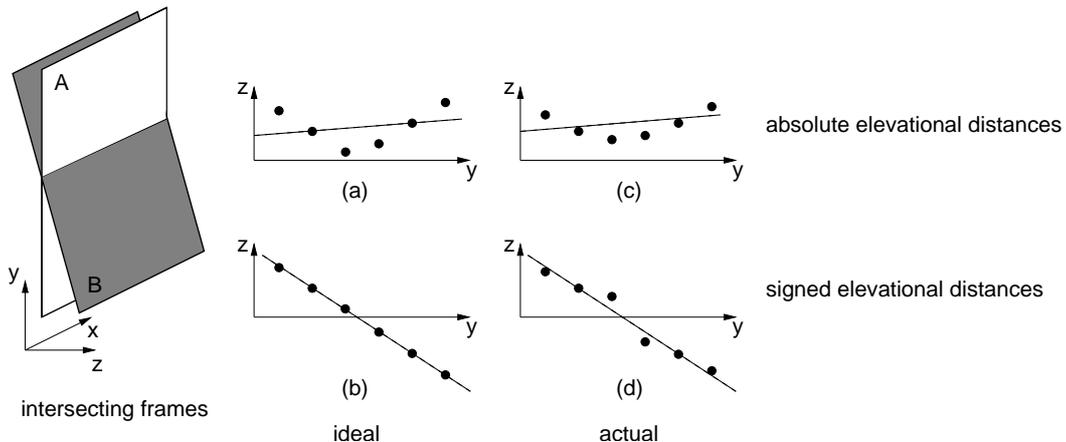


Figure 7: **Intersecting frames.** In this example, frames A and B intersect along a line parallel to the x axis. Looking at the absolute elevational z offsets along any column y of the grid, we'd expect to see a y - z distribution like the one in (a), though various sources of error mean that we actually see something more like (c). Speckle decorrelation can only give the absolute values of the elevational offsets. To correctly reconstruct intersecting frames, we need to deduce the signs, as in (b) and (d).

a flat plane, then we could consider possible intersection points in the y direction and choose the one which gives the smallest least squares residual for the fitted plane. In the example in Figure 7, the smallest residual is given by changing the signs of the distance estimates for $y > 3$, as shown in Figure 7(b). We would therefore deduce that frames A and B intersect between grid rows 3 and 4, and position them accordingly.

The task is complicated by a significant amount of uncertainty in the elevational separation estimates. As mentioned in Section 1, there are sources of error which we have not allowed for, including electrical noise, tissue compression and physiological motion. For intersecting frames, the errors are particularly severe around the line of intersection, where there is speckle decorrelation which is not due to elevational motion, but due to rotation of one frame relative to the other. Since the decorrelation curve is fairly flat at its left hand end (see Figure 1), even a small underestimate of the correlation can result in a large overestimate of the elevational separation [8, 14]. Consequently, the estimated z offsets do not approach zero at the line of intersection, but are instead somewhat larger, as shown in Figure 7(c) and (d). The residual error for the no-intersection solution (c) may not be very different from that of the correct solution (d).

So there is considerable uncertainty in the individual z values but, equally, considerable redundancy to exploit. We have 8×12 offsets, but we only need three to position frame A relative to frame B . We therefore adopt a robust regression approach, the aim being to ignore any outliers in the z values.

We start by considering individual rows and columns of the offset grid independently: see Figure 8(a). For a row or column comprising N offsets, there are N configurations of signed offsets to consider, including the no intersection case, as shown in Figure 8(b). We fit a straight line to each configuration and assume that the one with the smallest residual error reflects the correct configuration. If this configuration indicates an intersection, then that row or column contributes one putative point on the line of intersection of the two frames, as shown in Figure 8(c). Other rows and columns suggest other points. It is tempting to find the line of intersection by a simple least squares fit to the ensemble of points, though the aforementioned sources of error render such an approach inappropriate.

Instead, we use a variation of the random sample consensus (RANSAC) algorithm [4]. Each pair of intersection points defines a candidate line of intersection. We exhaustively evaluate all

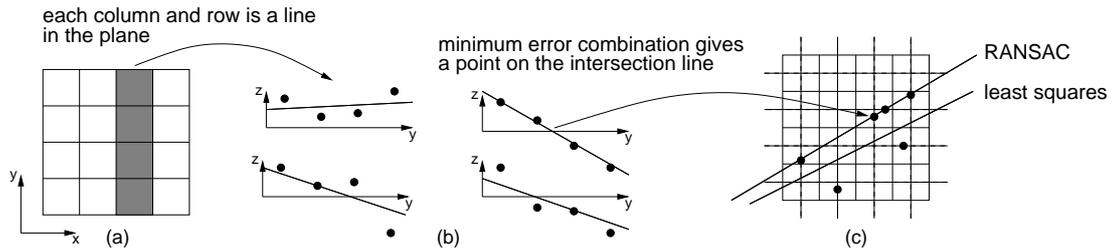


Figure 8: **The intersection algorithm.** Each row and column (a) of the grid of elevational offsets should define a straight line. We try all permutations of the signed offsets (b) and select the one which gives the best straight line fit. This provides one point (c) on the line of intersection of the two frames. Other rows and columns provide other points, though inevitable errors mean that a least squares fit to the ensemble of points is not appropriate.

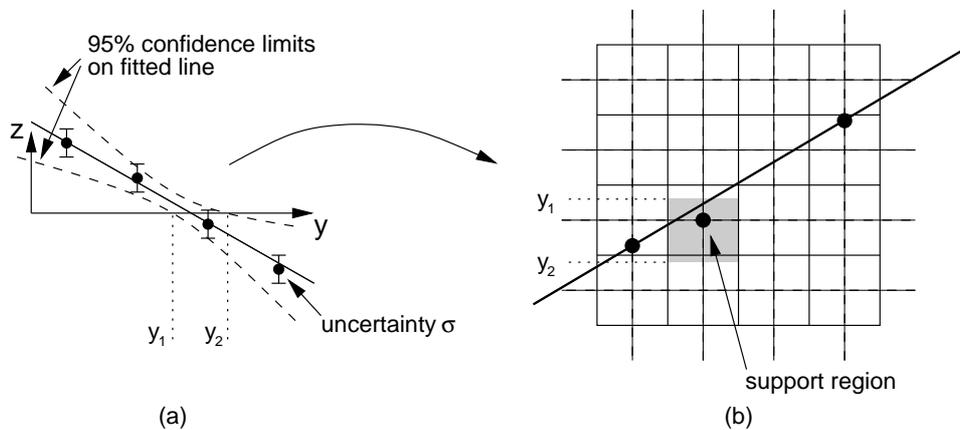


Figure 9: **The uncertainty model.** Uncertainty in the z_i values leads to uncertainty in the fitted line. The 95% confidence limits on the intersection point are given by y_1 and y_2 , where the uncertainty hyperbolae intersect the y -axis. This defines one dimension of the support region in the offset grid. The other dimension is simply the width of the corresponding row or column.

such lines of intersection and count how many of the other points support the hypothesis that the line is correct. To qualify as a “supporter”, a point has to be sufficiently close to the line such that any discrepancy can be explained by the uncertainty in the original elevational z offsets.

To deduce precisely how close a point should be to lend its support to a candidate line, we return to the least squares fitting process which we used to obtain the putative intersection points from the z values. For a standard linear regression on the points (y_i, z_i) , $i \in \{1 \dots N\}$, a y value predicts a particular z value according to the regression result $z = my + c$. The 95% confidence limits on this z value are [2]

$$z \pm 1.96 \sigma \sqrt{\left(\frac{1}{N} + \frac{(y_0 - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \right)} \quad (1)$$

where σ is one standard deviation of the uncertainty in the z_i values and \bar{y} is the mean of the y_i values. The loci of the points in (1) define two hyperbolae around the fitted line, as shown in Figure 9(a). The two points y_1 and y_2 where the hyperbolae cross the y axis give the 95% confidence interval on the intersection point. This defines one dimension of the support region,

as shown in Figure 9(b). The other dimension is simply the width of the corresponding row or column. If the candidate line of intersection passes through a point’s support region, then that point supports the hypothesis that the line is correct.

The value of σ can be estimated from the patch-wise elevational separations themselves. If these define a flat plane, the difference between any two adjacent z offsets in the x direction should be constant; likewise in the y direction. Any variation in these pair-wise differences is indicative of uncertainty σ in the z offsets. It is straightforward to show that the variance of the z offsets is half the variance of the pair-wise differences: this is how we estimate σ from the data itself¹.

The output of the RANSAC algorithm is the largest support set found amongst the candidate lines of intersection. The next step is to decide whether or not to accept this intersection. Where frames do not actually intersect, it only takes two incorrect intersection points, from any two rows or columns, to suggest a line of intersection. We therefore compare the size of the support set s to the maximum possible support s_{\max} , and accept the intersection only if $s > ks_{\max}$. Note that s_{\max} is larger for a line of intersection crossing a good portion of the grid than for one which just clips one of the grid’s corners. k is the intersection algorithm’s only tunable parameter. As illustrated in Figure 7(c) and (d), errors in the elevational offsets can cause many intersection points to be missed: we should therefore set k to a rather low value (0.2 was used for the experiments in Section 3). Finally, if the intersection is accepted, the line of intersection is refined by an orthogonal least squares fit to all the points in the support set.

2.3 Non-monotonic reconstruction

Consider a sequence of frames to be reconstructed. The objective is to determine the six degrees of freedom defining the offset between each adjacent frame pair. These offsets can then be concatenated to give the position and orientation of each frame relative to the first frame in the sequence. Rather than finding all six values together, we simplify the problem by dividing it into in-plane and out-of-plane components. In the plane of the images, the offset is defined by the axial and lateral translation at the frame centre and a roll angle about the elevational axis. These values are found by a least squares fit to the axial and lateral separation estimates at each patch. Out of the image plane, there are two rotations (tilt and yaw) and an elevational translation at the frame centre. We again determine these values with a least squares fit to the patch-wise elevational separation estimates.

As mentioned in Section 1, there is a direction ambiguity in the elevational separation estimates. At this stage, the intersection algorithm (Section 2.2) would have resolved this ambiguity within each frame pair, but there is still an ambiguity between different pairs. Figure 10 illustrates the problem. Given three frames A , B and C , the AB and BC offsets can be found independently, but their relative directions are not known. Frame C has two possible positions relative to A and we need to determine which one is correct.

We can distinguish the two possibilities by looking at the distances between frames A and C and solving an expanded least squares problem for the three frames together. This is in contrast to all previous work, which has only considered pairs of frames. For a single frame pair, the least squares problem can be written as $\mathbf{d} = \mathbf{X}\mathbf{w}$, where \mathbf{d} is a vector of the patch-wise elevational separation estimates, \mathbf{w} is a vector of the three out-of-plane frame-wise offset parameters and \mathbf{X} is a coefficient matrix. We can expand the problem to three frames as follows:

$$\begin{pmatrix} \mathbf{d}_{AB} \\ \mathbf{d}_{AC} \\ \mathbf{d}_{BC} \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{X} & \mathbf{X} \\ \mathbf{0} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \mathbf{w}_{AB} \\ \mathbf{w}_{BC} \end{pmatrix} \quad (2)$$

The additional information of the separations between frames A and C disambiguates the two possible directions of BC relative to AB : we consider both combinations and accept the solution with the smallest residual error. Note how the frame selection algorithm in Section 2.1 facilitates

¹The signs of the pair-wise differences become problematic when the two frames intersect. Ignoring the sign errors results in an overestimate of σ , though the error is small for approximately parallel (but intersecting) frames.

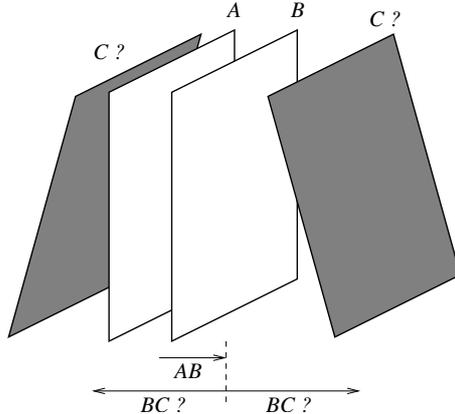


Figure 10: **Direction ambiguity between frame pairs.** There is a direction ambiguity between offsets AB and BC . Frame C cannot be placed relative to A using just the offsets between adjacent frame pairs.

this reconstruction technique, by ensuring that three consecutive frames in a coarse sequence are always within range of each other.

Figure 11(a) shows how we use the multiframe approach to reconstruct the coarsely spaced sequences of frames. We consider three consecutive frames in a sequence as an independent block and position them using equation (2). We then consider another block of three frames that overlaps the first, as shown in the figure. These independently positioned blocks are joined together by averaging the offset in the overlapping range. By repeating this process, we can build up an entire sequence of frames.

Correctly reconstructing a sequence relies on correctly reconstructing and joining each individual block of three frames. It only takes one spurious change of direction to make the entire sequence useless. The important next step is therefore to exploit the redundancy afforded by the multiple coarse sets (Figure 11(b)). We again use a variation of the RANSAC algorithm. We consider each coarse reconstruction in turn and count how many of the others support the hypothesis that it is correct. The largest set of supporters is taken as the set of correct reconstructions and the remaining coarse sequences are rejected.

A remaining detail is how we determine whether one reconstruction supports another. Consider the example in Figure 12. Of the five reconstructions, four are correctly monotonic but one is not. This is clear when we look at the accumulated offsets over the length of the sequences. We locate the maximum difference between the accumulated offsets and compare it to a threshold (0.5 mm was used for the experiments in Section 3). If the difference is less than the threshold, then the sequences are deemed to support one another.

Having rejected the outliers, we are left with several similar coarse scale reconstructions. By averaging the maximum offset of each of these, we determine the length of the final reconstruction. By “maximum offset”, we mean the maximum value in the accumulated offsets. For a monotonic sequence, this would be at the end of the sequence. For a non-monotonic sequence, it could be at any of the turning points. The averaging is intended to reduce any non-systematic drift error. We then arbitrarily choose one of the sequences to form a skeleton for the final reconstruction, and apply a uniform stretch (or shrink) so that it conforms to the average length (Figure 11(c)).

A complete reconstruction requires that the remaining frames are added to the coarse structure. We do this by considering blocks of frames between the coarsely spaced frames, as shown in Figure 11(c). These are reconstructed as an independent block, using an expanded version of equation (2), and then scaled to fit into the available gap. The positions of the coarsely spaced frames are unchanged, so the fine detail is added without affecting the established coarse scale reconstruction.

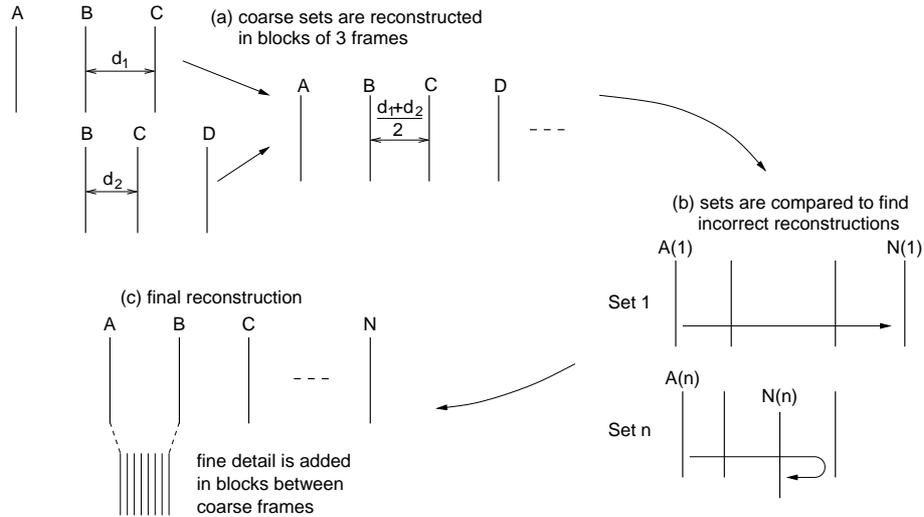


Figure 11: **The reconstruction algorithm.** The algorithm can be summarised in three stages. The multiple sets of coarsely spaced frames are reconstructed independently. They are then compared to find a consensus for the correct coarse scale reconstruction, resulting in a single coarsely spaced sequence. Finally, the remaining frames are added to the coarse structure.

3 Experiments and results

Ultrasound data was recorded with a 5–10 MHz linear array probe connected to a Dynamic Imaging Diasus ultrasound machine (<http://www.dynamicimaging.co.uk>). The depth setting was 4 cm with a single focus at 2 cm. Analogue RF echo signals were digitised after receive focusing and time-gain compensation, but before log-compression and envelope detection, using a Gage Compuscope CS14200 14-bit digitiser (<http://www.gage-applied.com>). The system works in real time, with acquisition rates of about 30 frames per second. The signal was sampled at 66.67 MHz, synchronous with the ultrasound machine’s internal clock: this synchronisation minimises phase jitter between adjacent A-lines. The RF vectors were filtered with a 5–10 MHz filter, then envelope detected using a Hilbert transform. Each B-scan’s position and orientation was recorded using a Northern Digital Polaris optical tracking system (<http://www.ndigital.com>), with spatial and temporal calibration performed according to the techniques in [15].

The scanning subject was a joint of beef. Three types of scan were performed:

- A hand-held linear sweep over the beef, with no intentional change of direction. Five of these data sets were recorded (data sets 1–5).
- A hand-held linear sweep over the beef, with the direction of motion intentionally reversed approximately half way through the scan². Five of these data sets were recorded (data sets 6–10).
- A rotational scan of the beef, with the probe twisting around an axis aligned roughly with the centre A-line, and no deliberate change of direction. The probe was mounted in a mechanism to constrain the motion appropriately, though it was still turned by hand, resulting in irregularly spaced frames as expected in a freehand data set. Three of these data sets were recorded (data sets 11–13).

²This is admittedly not a realistic scanning protocol. In practice, direction changes are likely to be accidental and transient, as opposed to intentional and sustained. However, having a definite turn at a known position is a good way to test the algorithm’s ability to reconstruct non-monotonic motion.

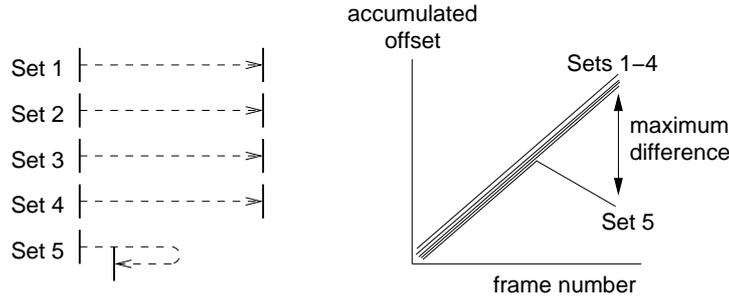


Figure 12: **Finding incorrect reconstructions.** In this example, four of the five reconstructions are correct. The incorrect one is found by looking at the maximum difference between the accumulated offsets in each set.

A meaningful comparison between the sensorless and position sensor reconstructions would require the beef to remain absolutely stationary with respect to the Polaris system’s reference frame. To this end, the probe was separated from the beef by a thick layer of coupling gel, to minimise the chance of any contact force displacing the beef. Note, however, that this precaution is solely for the benefit of the sensor-based reconstructions. Sensorless reconstructions are not sensitive to rigid movement of the scanning subject.

The probe was moved sufficiently slowly to yield five interleaved sets of frames in each of the thirteen experiments. Faster motion would be possible with internal processing of the RF signal within the ultrasound machine: our acquisition rate is limited by the speed of data transfer between the Gage card and main memory across the computer’s PCI bus. Note, however, that moving the probe too fast is not immediately catastrophic, but results in a graceful degradation of the reconstruction performance. Instead of five independent reconstructions to exploit, there would be fewer.

Figure 13 shows the linear reconstructions in terms of the cumulative elevational offset at the centre of the frames. Figure 14 shows the axial rotation reconstructions in terms of the cumulative yaw angle. All five of the coarse sets are shown separately, along with the equivalent values from the position sensor. The final reconstructions are not shown, to avoid cluttering the graphs. Adding the fine detail to the reconstruction is the least important stage, it is the coarse sets that determine whether or not the reconstruction is successful. Note that the coarse scale reconstructions each start at different frames in the sequence. In Figures 13 and 14, each line has therefore been offset at the start by the appropriate amount both horizontally and vertically. The required vertical offsets were determined by solving an expanded version of equation (2) for a block containing the five frames at the start of each set.

The results demonstrate robust reconstruction in all thirteen cases. In the worst cases, data sets 10 and 11, the final reconstruction relied on a consensus of two from five. Data set 10 featured two points with uncertain monotonicity, whereas most data sets have none. Considering the two points individually, there is a consensus of three out of five for the correct solution. In data set 11, there is clear evidence of missed intersections where the graphs flatten out around frame 50. This is undoubtedly a result of the phenomenon illustrated in Figure 7(c) and (d): unwanted speckle decorrelation around the line of intersection makes it easy to miss the intersection altogether. In general, the most brittle part of the reconstruction process is the intersection check, and the principal mode of failure is missed intersections, not false positives. To verify this last point, we repeated the reconstructions of the linear data sets with the intersection step skipped: we simply assumed no intersections at the coarse scale. The results were almost identical, so we can state with confidence that the errors in Figure 13 are not caused by false positives in the intersection test.

The sensorless reconstructions differ from the position sensor reconstructions in two ways: small scale jitter and large scale drift. The former is a known limitation of the position sensor

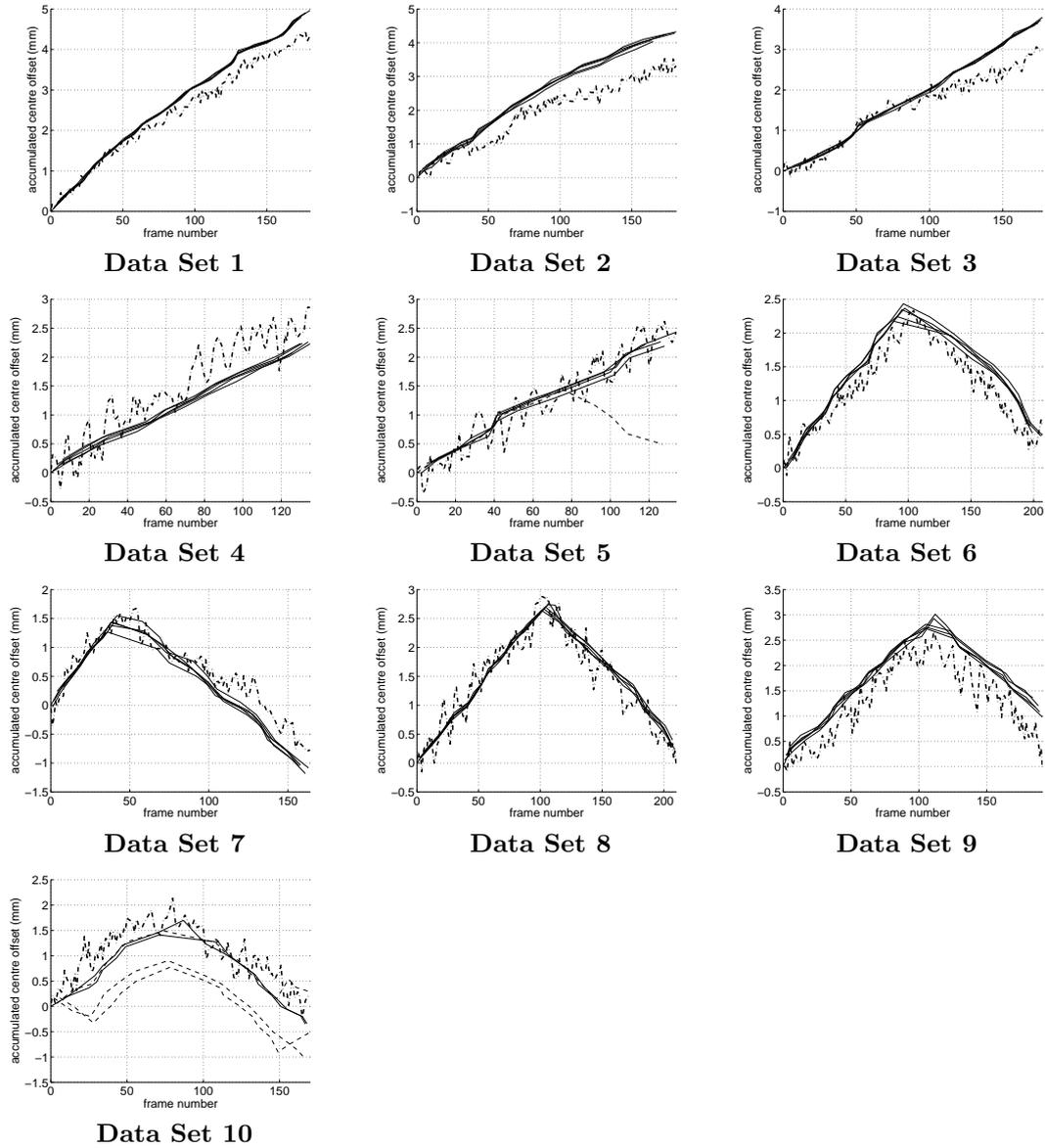


Figure 13: **Reconstruction of freehand linear data sets.** The graphs show the cumulative elevational offsets at the centre of the frames. For each data set, the five interleaved reconstructions are plotted separately. Reconstructions that contribute to the final consensus are shown as solid lines (—), those that are rejected are shown as dashed lines (- -). For comparison, the centre offsets given by the position sensor are also shown (- -).

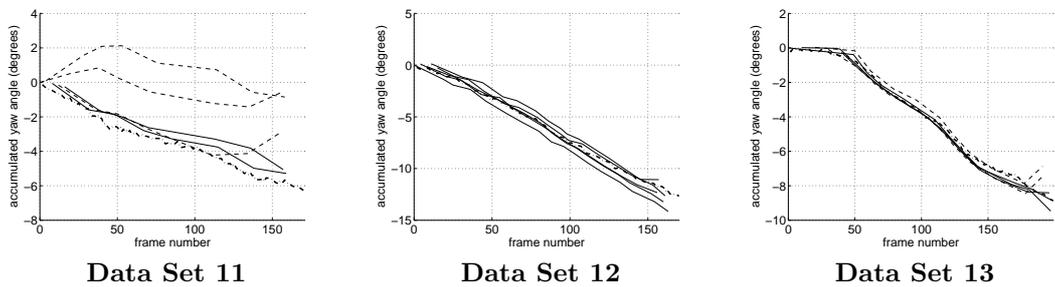


Figure 14: **Reconstruction of axial rotation data sets.** The graphs show the cumulative yaw angle. For each data set, the five interleaved reconstructions are plotted separately. Reconstructions that contribute to the final consensus are shown as solid lines (—), those that are rejected are shown as dashed lines (- -). For comparison, the rotations given by the position sensor are also shown (-·-).

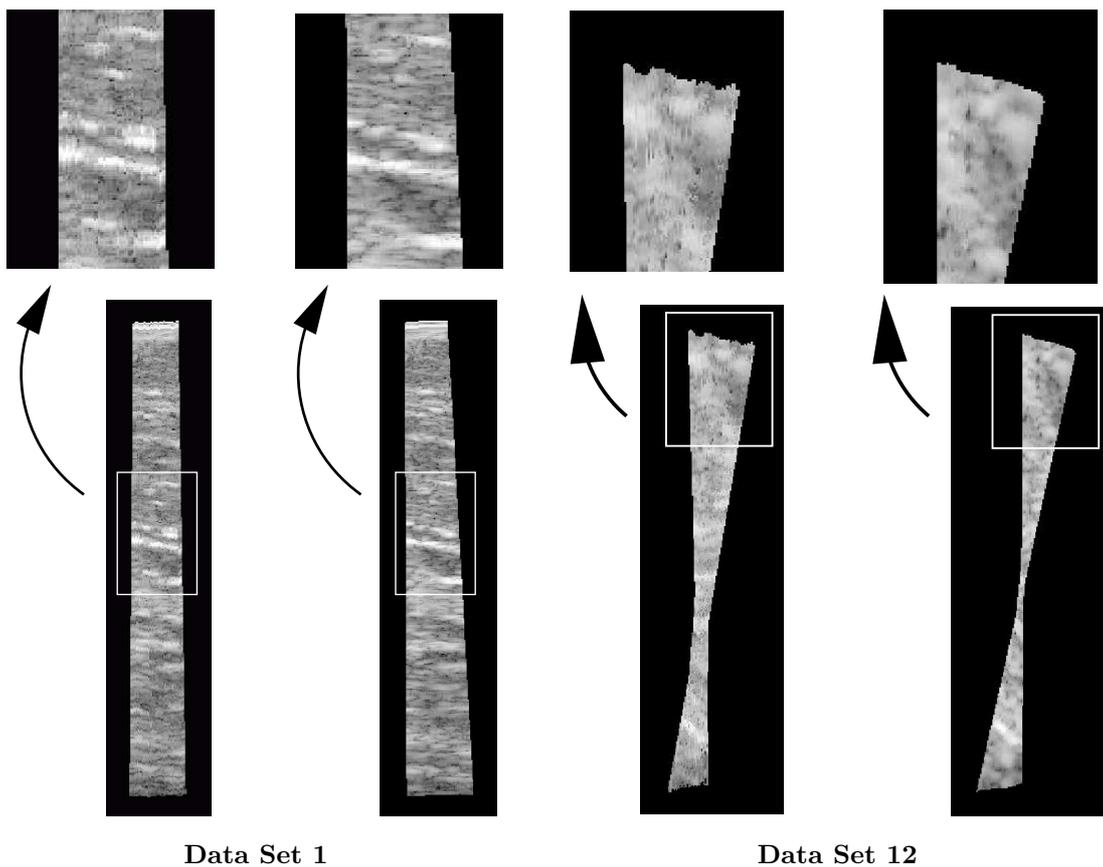


Figure 15: **Reslices.** On the left are reslices along the length of data set 1, a monotonic linear scan. On the right are reslices from data set 12, an axial rotation scan: in this case, the reslices are parallel to the probe face. In each case, the left images are derived from the position sensor reconstruction and the right ones from the image-based reconstruction.

which affects reslices of the reconstructed data sets. In Figure 15, small scale mispositioning of the B-scans results in discontinuities in the sensor-based reslices. In contrast, the sensorless reslices are far smoother.

On the other hand, large scale drift is a known limitation of sensorless reconstruction techniques. It arises from small biases in the decorrelation-based elevational separation estimates. The principal sources of the bias are lateral interpolation between A-lines [10], compensation for coherent scattering [5] and speckle decorrelation caused by probe rotation instead of elevational motion. The last of these would result in systematic overestimation of elevational separation — as is evident in many of the results in Figure 13 — and helps explain the somewhat larger drift for, say, data set 2 compared with previous results obtained with purely translational probe motion [5, 10]. There is also some drift in the elevational yaw and tilt estimates (see, for example, the reslice of data set 1 in Figure 15), though we have not cluttered the results by presenting quantitative details here. This is because drift has its roots in the low-level, patch-wise separation estimates, which are not the subject of this paper. The high-level frame selection, intersection and three-at-a-time reconstruction algorithms have, as expected, neither exacerbated nor ameliorated the drift.

4 Conclusions

We have demonstrated robust sensorless reconstruction of unconstrained freehand 3D ultrasound data. The most brittle part of the process is the intersection test, though false positives are rare so this is not a problem with the usual linear scanning protocols. Sensorless reconstructions are not sensitive to rigid movement of the scanning subject. Moreover, small scale detail is not obscured by position sensor registration jitter. However, unavoidable drift limits the accuracy of any large scale measurements based on the sensorless reconstructions.

In future work, we intend to build a hybrid system combining image-based reconstruction with an unobtrusive, optical fibre-based position sensor (ShapeTape, <http://www.measurand.com/>) that runs down the probe cable. Unlike magnetic and optical position sensors, there is no line of sight requirement and the clinician need not be aware of the extra hardware. Sparse readings from the ShapeTape could be used to correct the drift in the image-based reconstructions, and bridge any gaps caused by excessively rapid probe motion.

Acknowledgements

This work was carried out under EPSRC grant number GR/S34366/01. Dynamic Imaging provided a modified Diasus US machine with direct access to the analogue RF signals.

References

- [1] J-F Chen, J. B. Fowlkes, P. L. Carson, and J. M. Rubin. Determination of scan plane motion using speckle decorrelation: theoretical considerations and initial test. *International Journal of Imaging Systems Technology*, 8:38–44, 1997.
- [2] N. R. Draper and H. Smith. *Applied regression analysis*. Wiley, New York, 1966.
- [3] A. Fenster, D. B. Downey, and H. N. Cardinal. Three-dimensional ultrasound imaging. *Physics in Medicine and Biology*, 46:R67–R99, 2001.
- [4] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

- [5] A. H. Gee, R. J. Housden, P. Hassenpflug, G. M. Treece, and R. W. Prager. Sensorless freehand 3D ultrasound in real tissue: Speckle decorrelation without fully developed speckle. *Medical Image Analysis*, 10(2):137–149, 2006.
- [6] A. H. Gee, R. W. Prager, G. M. Treece, and L. H. Berman. Engineering a freehand 3D ultrasound system. *Pattern Recognition Letters*, 24:757–777, 2003.
- [7] B. J. Geiman, L. N. Bohs, M. E. Anderson, S. M. Breit, and G. E. Trahey. A novel interpolation strategy for estimating subsample speckle motion. *Physics in Medicine and Biology*, 45:1541–1552, 2002.
- [8] P. Hassenpflug, R. Prager, G. Treece, and A. Gee. Distance measurement for sensorless 3D US. In *Medical Image Computing and Computer-Assisted Intervention — MICCAI'04*, volume 2, pages 1087–1088, Saint-Malo, France, September 2004. LNCS 3217, Springer.
- [9] R. J. Housden, A. H. Gee, R. W. Prager, and G. M. Treece. Sensorless freehand 3D ultrasound for non-monotonic and intersecting frames. In *Proceedings of the 9th Annual Conference on Medical Image Understanding and Analysis*, pages 127–130, Bristol, UK, 2005.
- [10] R. J. Housden, A. H. Gee, G. M. Treece, and R. W. Prager. Sub-sample interpolation strategies for sensorless freehand 3D ultrasound. *Technical Report CUED/F-INFENG/TR 545*, Cambridge University Engineering Department, 2006.
- [11] M. Li. System and method for 3-D medical imaging using 2-D scan data. *United States Patent*, 5,582,173, 1995.
- [12] F. Lindseth, G. A. Tangen, T. Lango, and J. Bang. Probe calibration for freehand 3-D ultrasound. *Ultrasound in Medicine and Biology*, 29(11):1607–1623, November 2003.
- [13] T. R. Nelson and D. H. Pretorius. Three-dimensional ultrasound imaging. *Ultrasound in Medicine and Biology*, 24(9):1243–1270, 1998.
- [14] W. Smith and A. Fenster. Statistical analysis of decorrelation-based transducer tracking for three-dimensional ultrasound. *Medical Physics*, 30(7):1580–1591, 2003.
- [15] 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*, 29(4):529–546, April 2003.
- [16] 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.
- [17] T. A. Tuthill, J. F. Krücker, J. B. Fowlkes, and P. L. Carson. Automated three-dimensional US frame positioning computed from elevational speckle decorrelation. *Radiology*, 209:575–582, 1998.
- [18] L. Weng, A. P. Tirumalai, C. M. Lowery, L. F. Nock, D. E. Gustafson, P. L. V. Behren, and J. H. Kim. US extended-field-of-view imaging technology. *Radiology*, 203(3):427–441, 1997.