

3D ultrasound reconstruction algorithms from analog and digital data

Ole Vegard Solberg^{1,2,3}, Frank Lindseth^{1,2,3}, Lars Eirik Bø^{1,2}, Sébastien Muller^{1,2}, Janne Beate Lervik Bakeng^{1,2}, Geir Arne Tangen^{1,2}, and Toril A. N. Hernes^{1,2,3}

¹SINTEF Technology and Society, Trondheim, Norway

²National Centre for 3D Ultrasound in Surgery, Trondheim, Norway

³Norwegian University of Science and Technology, Trondheim, Norway

November 15, 2010

Abstract

Freehand 3D ultrasound is increasingly being introduced in the clinic for diagnostics and image-assisted interventions. Various algorithms exist for combining 2D images of regular ultrasound probes to 3D volumes, being either voxel-, pixel- or function-based. Previously, the most commonly used input to 3D ultrasound reconstruction has been digitized analog video. However, recent scanners that offer access to digital image frames exist, either as processed or unprocessed data. To our knowledge, no comparison has been performed to determine which data source gives the best reconstruction quality. In the present study we compared both reconstruction algorithms and data sources using novel comparison methods for detecting potential differences in image quality of the reconstructed volumes. The ultrasound scanner used in this study was the Sonix RP from Ultrasonix Medical Corp (Richmond, Canada), a scanner that allow third party access to unprocessed and processed digital data. The ultrasound probe used was the L14-5/38 linear probe. The assessment is based on a number of image criteria: detectability of wire targets, spatial resolution, detectability of small barely visible structures, subjective tissue image quality, and volume geometry. In addition we have also performed the more “traditional” comparison of reconstructed volumes by removing a percentage of the input data. By using these evaluation methods and data from the specific scanner, the results showed that the processed video performed better than the digital scan-line data, digital video being better than analog video. Furthermore, the results showed that the choice of video source was more important than the choice of tested reconstruction algorithms.

*Originally published in *Ultrasonics*, vol. 51, no. 4, pp. 405–419, 2011

1 Introduction

The use of 2D ultrasound for a variety of clinical applications is becoming more common. Compared to other imaging modalities like magnetic resonance imaging (MRI) and computed tomography (CT), ultrasound has the advantages of being cheaper, smaller and more flexible, it has no radiation and it is easier to introduce during surgery. 3D freehand ultrasound offers even more flexibility, and combined with position tracking it is found useful in minimally invasive image-guided surgery (IGS). [1]

The most commonly used input to freehand 3D reconstructions is digitized analog video, either first stored on a video tape and digitalized with an analog video frame grabber [2–4] or obtained instantly by connecting the video frame grabber directly to the video output of the ultrasound scanner [4–18]. Some researchers state that they use digital data from the scanner. In most cases, such data is unavailable to third party users, but exceptions exist like the Sonix RP scanner (Ultrasonix Medical Corp., Richmond, Canada). Some groups also gain access to the digital data by collaborating with the ultrasound scanner manufacturers. [19–23]

Different algorithms for reconstructing 3D volumes from freehand ultrasound exist. In summary, these are [24]:

- Voxel-Based Methods (VBM). Include the Voxel Nearest Neighbor (VNN) where each voxel is assigned the nearest pixel [8] and algorithms where each voxel is assigned a value based on several of the nearest pixels [19, 21, 25–27]. In this group are also algorithms that skip the creation of a voxel volume and reconstruct a 2D slice or surface directly. [28, 29]
- Pixel-Based Methods (PBM). Range from algorithms like the one-step Pixel Nearest Neighbor (PNN) where 2D input images are inserted directly into a target volume [25] to two-step PNN where a second step fills empty voxels afterwards [2, 3, 30–32] to algorithms where input pixels are added with a 3D kernel [6, 17, 33, 34].
- Function-Based Methods (FBM). Algorithms where functions are made based on the input pixels and the target volume is created by evaluating these functions at regular intervals. [35, 36]

For 3D probes [37, 38], the process is digital from acquisition to reconstruction. The DICOM 2008 standard (DICOM 2008, Suppl. 43) also has defined structures for 3D ultrasound volumes. 3D probes still have the disadvantage of poorer resolution compared to 2D probes. 3D probes on the other hand have the advantage of allowing real-time 3D volumes, while 2D probes may only provide real-time 2D images. In IGS, the positions of the data are also necessary in order to navigate in the 3D volume, so a position sensor must be attached to the ultrasound probe and the relation between the

sensors and image data must be defined (probe calibration). In the present study we have evaluated the differences between image volumes originating from different ultrasound data sources and various 3D reconstruction algorithms using a broad range of both quantitative and qualitative comparisons methods.

2 Materials and methods

2.1 Ultrasound data import

The ultrasound scanner used (Figure 1, Sonix RP, Ultrasonix Medical Corp., Richmond, Canada) has a research interface allowing real-time access to digital data from the scanner. A 7.5 MHz linear probe (Figure 1, L14-5/38, Prosonic, Gyeongbuk, South Korea) [39] operating at 10 MHz scanning frequency was used to acquire all images analyzed in this study.

Three different video streams were imported simultaneously:

- Analog video, converted from PAL S-Video with a video-to-FireWire converter (DFG/1394-1e, The Imaging Source, Germany).
- Digital scan converted video, processed by the ultrasound scanner for viewing on a screen.
- Unprocessed digital data delivered as 1D scan lines, only envelope detected and log-compressed.

The digital image sources were imported directly from the ultrasound scanner over a crossed LAN (Local Area Network) cable. The three video streams were imported simultaneously in different threads on a PC with four CPU kernels (Figure 1, Intel Core 2 Quad Processor Q6700 2.66 GHz). Each time an image was received on the computer, a time stamp was created and assigned to the image. A depth setting of 4 cm was used on the scanner for all image acquisitions. This resulted in a pixel size (width \times depth) of 0.147 mm \times 0.147 mm for the analog video, 0.097 mm \times 0.097 mm for the digital video, and 0.150 mm \times 0.077 mm for the unprocessed “video”. The images from the analog and digital video were cropped to only contain the ultrasound data. The difference in pixel sizes between the unprocessed video and digital video is due to the ultrasound scanners internal processing of the video. The pixel size difference between the digital and analog video comes from lesser resolution in the video grabbing hardware. All data sources supplied 8-bit pixels.

2.2 Position tracking and probe calibration

For freehand 3D ultrasound reconstruction, positions and orientations of the 2D images are needed. Several methods for obtaining these positions

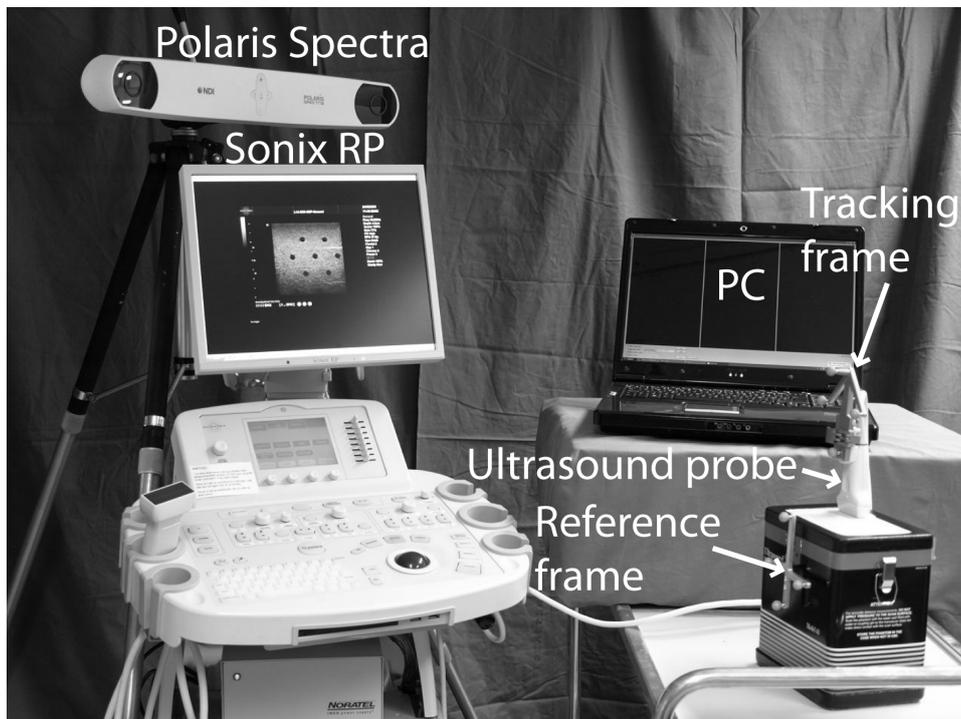


Figure 1: System setup. Polaris Spectra optical positioning system, Sonix RP ultrasound scanner, ultrasound probe with tracking frame, ultrasound phantom with tracking reference frame, and PC for data import.

exist. [24, 40] In our study we used an optical positioning system (Figure 1, Polaris Spectra, Northern Digital Inc., Canada), consisting of a tracking frame attached to the ultrasound probe and a camera unit that were used for calculating the position and orientation of this frame. The positions were obtained using the Image-Guided Surgery Toolkit (IGSTK) [41–43]. As with the video streams, time stamps for the positions were created and assigned by the software at the time the positions were received. The positions were imported with the same application importing the three video streams, but run as a separate thread.

For the 2D ultrasound image to be correctly aligned with the output from the positioning system, a calibration is necessary. We used the spatial calibration method and phantom developed by Chen et al. [18], with threads stretched between the sidewalls in two parallel, ‘N’-shaped configurations. A tracking frame was mounted on the phantom and the positions of the threads relative to this frame were measured. The corresponding structures were identified in the ultrasound images using an automatic segmentation algorithm. The relationship between the image plane and the positioning system was then found using a least-squares minimization method. We also implemented a temporal calibration based on the work by Treece et al. [44] except that we detected a point instead of a line. This temporal calibration was used to synchronize the imported positions with the images. The temporal calibration method matched the vertical movement of one of the segmented points in the 2D images to the vertical movement of the ultrasound probe as reported by the positioning system. Using the temporal calibration, 23.86 ms was subtracted from the analog video time-tags to match the time-tags of the positions, 2.56 ms was added to the digital video time-tags and 0.32 ms to the time-tags of the unprocessed images. The total mean error from the spatial calibration was 1.05 mm, with a root mean square (RMS) value of 1.13 mm and a standard deviation of 0.42 mm. These results were used in the reconstructions and display of data in the following tests.

2.3 Reconstruction algorithms

Two pixel-based 3D reconstruction algorithms were compared in this study. The first is called Pixel Nearest Neighbor [2, 3, 24, 30] and is a relatively fast two-step method. The first step inserts each image pixel in the input 2D images into the target 3D volume based on the position and orientation of the images. The chosen implementation overwrites any existing data in the 3D volume with the most current 2D image. The second step is an interpolation step that traverses the voxels of the target volume and attempt to fill empty voxels with the average value of the nearby voxels. The interpolation first tries to interpolate a voxel with the voxel values from the $3 \times 3 \times 3$ grid around it. If all these voxels are empty a $5 \times 5 \times 5$ grid is used and after that

a $7 \times 7 \times 7$ grid, and if there is still no voxel values within this range the voxel is left empty. The second reconstruction algorithm uses a 3D kernel around the input pixels. Several variations of the input kernel are described [6, 17, 24, 33, 34, 45], and we used an ellipsoid truncated Gaussian-weighted kernel around the input pixels [17, 24]. The size of the kernel is usually set to fill holes in the volume, but we used a novel method of approximately matching the theoretical ultrasound resolution in all three dimensions. For the comparisons we used two slightly different sizes of this kernel, resulting in three different reconstructed volumes:

- Pixel Nearest Neighbor.
- Small 3D ellipsoid kernel around input pixels.
- Large 3D ellipsoid kernel around input pixels.

The reconstruction algorithms were set to create volumes that used the full range of 8 bits to produce volumes that were similar in intensity. The voxel size of all reconstructed volumes were set to 0.2 mm as a compromise between resolution and processing time, the voxel sizes being larger than the input pixel sizes.

2.3.1 Determining kernels for the 3D kernel based reconstruction algorithm

As an approximation of the two-way pulse-echo response of the ultrasound imaging system we have decided to use a 3D Gaussian function. This function is best known as the probability density function for a normal distribution, and in 1D it is given by the formula

$$f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (1)$$

where μ is the mean and σ is the standard deviation. In 3D this becomes

$$f(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(\sqrt{2\pi})^3 \sqrt{|\boldsymbol{\Sigma}|}} e^{-\frac{(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}{2}}, \quad (2)$$

where $\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)$ is the mean vector, and $\boldsymbol{\Sigma}$ is the covariance matrix.

The actual pulse-echo response may be approximated as a product of two sinc functions. The formula for a rectangular aperture in the lateral and elevation direction (Figure 2) may be derived from Angelsen [46, p. 5.54]:

$$H(x, \lambda) = \text{sinc}\left(\frac{x}{\lambda_t f_{\#t}}\right) \text{sinc}\left(\frac{x}{\lambda_r f_{\#r}}\right), \quad (3)$$

where H is the two-way pulse-echo response, x is distance in either the lateral or elevation direction (Figure 2). $\lambda = c/f$ is the wavelength, $f_{\#} = F/D$ is

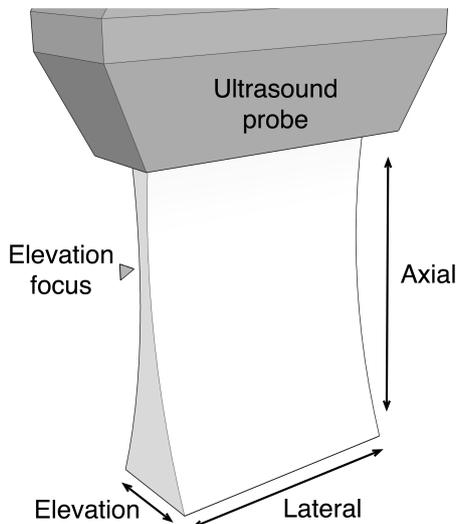


Figure 2: Illustration of the imaging sector of an ultrasound probe. The lateral, axial and elevation direction are marked in addition to the fixed elevation focus.

the f -number, c is the speed of sound, f is the ultrasound center frequency, F is the focal position (depth for the calculation) and D is the effective aperture size. The transmit (t) and receive (r) directions may have different λ and $f_{\#}$. Acoustic absorption reduces the pulse center frequency, resulting in a lower center frequency for the received pulse compared to the transmitted pulse. For a Gaussian pulse envelope the frequency is reduced by the following formula [47]:

$$f_r(z) = f_t - \frac{\alpha B^2 z}{4 \ln 2}, \quad (4)$$

where z is the propagation distance, $\alpha = (a \ln 10)/20$ is the average constant of the absorption coefficient. A commonly used value for a is 0.5 dB/(cm MHz). B is the -6 dB bandwidth of the imaging pulse. In the axial direction (see Figure 2), the pulse-echo response is dependent of the form of the transmitted pulse, but is often approximated as a Gaussian function.

To calculate the theoretical size of the focus of an ultrasound probe in the lateral or elevation direction the function (3) can be evaluated at -6 dB. However, for the Gaussian function we use the width of the main lobe, equal the first zero in the narrower of the two sinc functions in (3). The resulting value was used as the limit for a 99.7% confidence level for a 1D Gaussian distribution (total pulse width obtained from (4) corresponds to 6 standard deviations, so the standard deviation (σ) is estimated simply by dividing the width of the main lobe by 6). The ultrasound scanner uses a Gaussian-like apodization on the elements to dampen the side lobes of the transmitted

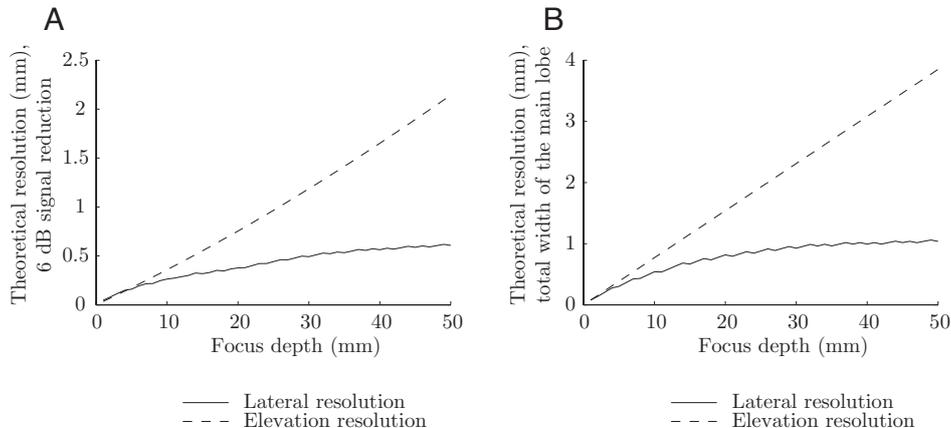


Figure 3: Plot of the theoretical lateral and elevation resolution in focus for a few selected depths for the L14-5/38 probe operating at 10 MHz. (A) The resolutions shown as a 6 dB signal reduction, calculated from (3) and (4). The corresponding axial resolution is 0.154 mm for all depths, calculated from (5). (B) The resolutions illustrated as the total width of the main lobe from (3).

pulse. This apodization will also create a wider main lobe than indicated by the formula (3), which are for a situation without apodization [46, pp. 5.56 – 5.58].

The ultrasound probe had a fixed elevation focus of 16 mm (Ultrasonix Medical Corp., Richmond, Canada) and the ultrasound scanner allows the operator to set several variable focus points in the axial direction. Three focus points were used and at least one was set near the fixed elevation focus. The probe size in the lateral direction is 38 mm, but the effective aperture size varies according to depth (dynamic aperture). The aperture sizes for both transmit and receive were calculated with parameters read from the scanner and code obtained from Ultrasonix. According to information obtained from Ultrasonix the fractional Bandwidth of the probe is minimum 70 % of the center frequency at -6 dB, the center frequency being 7.2 MHz, resulting in a bandwidth of 5.04 MHz. The theoretical lateral and elevation resolutions in focus for the used ultrasound probe are illustrated in Figure 3A by evaluating formula (3) at -6 dB, while the calculations of the total width of the main lobe are shown in Figure 3B. The imaging frequency used is 10 MHz, and the speed of sound in tissue is set to 1540 m/s.

The size in the elevation direction (element height) is 4 mm (Ultrasonix Medical Corp., Richmond, Canada). The theoretical resolution in the axial direction equals half the pulse length [46, p. 1.22], i.e.

$$\Delta a = \frac{cT_p}{2} = \frac{c}{2B}, \quad (5)$$

where T_p is the pulse length in time. With a bandwidth of 5.04 MHz the best theoretical axial resolution is 0.153 mm (5). However, as we wanted to find a value to use for the limit of the 99.7% confidence level of a 1D Gaussian function, the full pulse length should be used: 0.306 mm.

The resolution values in all three dimensions were used to create the covariance matrix Σ for the 3D Gaussian function (2) and thus define the extent of the truncated kernel. For the small kernel we used an imaging depth of 19 mm to calculate the lateral resolution (= 0.78 mm), and for the large kernel we used a depth of 32 mm (= 0.99 mm). To speed up reconstruction computation time we reduced the total kernel size by calculating the elevation resolutions at somewhat shallower depths. The elevation resolutions were calculated for the depth of 18 mm (= 1.39 mm) for the small kernel and for the depth of 24 mm (= 1.85 mm) for the large kernel.

In the 3D reconstruction, the 3D Gaussian function was oriented according to the 2D ultrasound images, and discrete values of the function were used to match the target 3D voxel grid. For the axial resolution we used a resolution of 0.171 mm instead of 0.306 mm, due to initially calculating the axial resolution for 9 MHz instead of 5 MHz. We truncated the 3D Gaussian kernel at 95% confidence level and then increased the kernel size to include the same number of voxels on both sides of the input pixel. Since the axial resolution was so high this resulted in a kernel size of only one voxel in the axial direction. However, the current implementation of the reconstruction requires a larger kernel in this direction, so we increased the kernel size in the axial direction to three voxels.

2.4 Data collection

All acquisitions were performed as freehand translation sweeps with the three different data streams imported simultaneously. To minimize errors from the position-to-image synchronization, and to make sure that the reconstructions got enough data to fill holes, all sweeps were performed with a slow, smooth motion. One scan thus resulted in input data from three different data sources, all showing images from the same structures. The digital video after scan conversion was the same images as presented on the screen of the ultrasound scanner. These images were processed by the scanner for viewing and all user-controlled functions, such as the gain function on the scanner's control panel were applied by the scanner. The analog data is a video signal output from the ultrasound scanner. The signal was converted from digital to analog by the scanner, and again to a digital signal by the video grabber.

Throughout this paper the images used for the 3D reconstructions are called "input images" while images used directly in the comparisons are called "original images". In the comparisons the same images were never used as both input and original images even if the images may be similar

and acquired in approximately the same positions. Both input images and original images came from all three data sources, and while input images were acquired as a stack of several images original images were only acquired as single images.

2.5 Tests for comparing reconstruction algorithms and data sources

1. *Compare the reconstruction algorithms' ability to correctly recreate removed data (test 1):* To compare the quality difference of different reconstructions based on the same input data the method of removing a percentage of input data [31, 32, 35, 48, 49] was used. We scanned a section of the underside of the forearm on two healthy volunteers with two translation scans along the arm of one volunteer and two translation scans across the arm of the other volunteer. The scans along the forearm gave images that only changed slightly from one image to the next, while the images from the scans across the forearm changed more rapidly. Before reconstruction, 0 %, 25 %, 50 %, 75 %, 100 %, 300 %, 500 % and 700 % of the data of one of the input 2D ultrasound images was randomly removed. Random pixels were removed from one slice for the percentages below 100 %, and whole slices were removed for the percentages 100 % to 700 %, e.g.: for 300 % three slices were removed. After reconstruction, all pixel values of this input image were compared with the voxel values from the corresponding positions in the reconstructed volumes and the RMS value of the differences was calculated. The orientation of the reconstructed volume was based on the orientation of the input image from which the data was removed. This procedure was performed on four different positions without holes in each volume, and statistical analyses were used to compare the performance of the reconstruction algorithms.
2. *Compare the reconstructed volumes' ability to display existing structures (test 2):* Measurements of the reconstructed volumes' ability to retain the resolution in the 2D images and of the resolution they manage to obtain in the elevation direction were performed. A comparison was performed by human observers on how well tissue was visualized. The human observers were technical researchers with knowledge of ultrasound imaging ranging from medium to expert. The reconstructed volumes from one scan resulted in a set of volumes with various combinations of the three data sources and three reconstruction algorithms. Original 2D ultrasound images from the ultrasound scanner were acquired simultaneously as a selection of 2D anyplane slices from the reconstructed volumes in the same position and orientation (Figure 4C and D). Statistical comparisons were performed between reconstruc-

tion algorithms/original 2D ultrasound images and video sources. All volunteers were presented with the same set of images on the same computer with brightness and contrast levels unchanged.

- (a) *Visual comparisons of structures placed increasingly closer (test 2a)*: Measuring the volume resolutions was accomplished by scanning an ultrasound resolution phantom, with six small structures (threads) that are placed with distances of 1 mm, 2 mm, 3 mm, 4 mm and 5 mm (Model 040, CIRS Inc., VA, USA). The threads of the CIRS 040 phantom are 0.1 mm in diameter and made of Nylon Monofilament. The threads were scanned with three translation sweeps both along (Figure 4A) and across (Figure 4B) the threads to measure resolutions both in the lateral and elevation direction (Figure 2) of the ultrasound input planes. 2D images were obtained by collecting original ultrasound images with position and orientation (Figure 4C). This position and orientation were used to create 2D anyplanes through the reconstructed 3D volumes (Figure 4D). These images were presented in random order and evaluated by eight volunteers that were given two questions to answer for each image: “Count how many separate bright structures you can see and rate how easy it is to identify those you can see (1–5, where 1 is easy and 5 hard)”.
- (b) *Comparisons based on image measurements of spatial resolution (test 2b)*: The images in test 2a of the CIRS 040 phantom also contained a separate thread, which was used for measurements of resolution: Both the original and a resampled (downsampled) version of the original image were used, and resolution were measured in positions corresponding with the anyplanes through the 3D volumes. Axial and lateral spatial resolutions were measured directly from the 2D ultrasound images by plotting gray level profiles through the center of the scanned wires [50]. The maximum pixel value was used as the center value for each thread. Parabolic curves were matched to these plots and evaluated at -6 dB and -20 dB. Only the values above -6 dB were used for matching the parabolic curves both in the original 2D images and in 2D anyplanes obtained from the 3D volumes. Measurements in the elevation direction were possible to perform on anyplanes through the 3D volumes scanned across the threads (Figure 4B). Each measurement was performed on the same wire at a depth of approximately 27 mm in three different scans.
- (c) *Visual comparisons of small, barely visible structures (test 2c)*: The CIRS 044 (Model 044, CIRS Inc., VA, USA), an ultrasound phantom with small cylinders of varying size, was used to compare

the visibility of small structures after a reconstruction. Three ultrasound translation sweeps were performed on the smallest cylinders on the CIRS 044 phantom, both lengthwise and crosswise. These small cylinders have measurements of 1.5 mm (diameter) and 2.4 mm (length) specified in the fabrication sheet. Original single 2D images were acquired showing as many cylinders as possible, all cylinders appearing as circles in the ultrasound image. Anyplane images (Figure 4D) through the 3D volumes were created from the same position and orientation as the collected original images (Figure 4C). Eight people were presented 2D images, in random order, showing small objects and presented with two questions for each image: “Count how many separate dark structures you can see, and rate how easy it is to identify those you can see (1–5, where 1 is easy and 5 hard).” When presented with several images of varying quality, the test person may “learn” where the structures should be, and this may enable them to identify more structures in a poor image that they would do otherwise. To allow for this the test subjects were also asked to rate (1–5) how easy it was to identify the structures they could identify, and they were also shown an illustration of the corresponding section of phantom beforehand. The 2D images were either an original 2D ultrasound image showing the structures or a 2D anyplane obtained from a 3D volume. The anyplanes from the 3D volumes were either approximately orthogonal or parallel to the 2D images from which the 3D volume was created.

- (d) *Visual comparisons of tissue data orthogonal to the scanning direction (test 2d)*: A section of the underside of the forearm of two healthy volunteers was scanned with freehand translation sweeps. One person was scanned with two sweeps along the arm while the other was scanned with two sweeps across the arm. Original single 2D images were acquired approximately orthogonal to the 3D acquisition sweeps for comparison (Figure 4C) and corresponding anyplane images through the 3D volumes were collected (Figure 4D). A group of eight people were presented with different 2D images in random order, showing tissue data from the same location. They were presented with sets of 3 or 4 images and asked the following question: “Sort the images according to quality and give each image a quality score (1–5), where 1 is best and 5 is worst.” The images were original 2D ultrasound images of the tissue or anyplanes obtained from reconstructed 3D volumes approximately orthogonal to the input 2D images. The sets of 3 images showed the 3 different input sources, being all original 2D images or all anyplane images from a specific reconstruction

algorithm. The sets of 4 images showed an original 2D ultrasound image and anyplane images from different reconstruction algorithms, all images being from the same input source.

3. *Compare the correctness of the reconstructed volume's geometry (test 3)*: A geometry comparison is a measurement on how well a reconstruction manages to recreate a known phantom geometry. Statistical comparisons based on both data sources and reconstruction algorithms were performed.

The CIRS 044 have a set of cylinders with specified measurements of 3 mm (diameter) \times 6 mm (length). Three ultrasound translation sweeps were performed on these structures in both the lengthwise and crosswise direction. Original single 2D ultrasound images with 3D positions and orientations were acquired for each data set both along and across the cylinder. Both height and width of the structure in the image was measured, resulting in measurements of cylinder length in the lateral direction and diameter height in the axial direction for the images along the cylinder, and measurements of diameter width in the lateral direction and diameter height in the axial direction. Three measurements were performed in the original image and three in an image processed with a levels function of an image processing application. The levels function was used to spread out the pixel intensities so that they fill the whole 8 bit range. The reason for this was to try to emulate the processing in the image reconstructions where a similar function was used for the whole volume.

2D anyplane images were created through the reconstructed 3D volumes at positions and orientations matching that of the original 2D images. The same structure measurements were done on these anyplane images and the differences were compared and tested for variations in reconstruction algorithm and data source quality. The anyplane images covered measurements in the elevation direction in addition to measurements in the lateral and axial direction. The elevation measurements were received from cylinder length in the scans along the cylinder and from the cylinder width measurements in the scans across. The same person performed all analyses, repeating each measurement three times. The cylinder measurements performed on the original ultrasound images were used as the gold standard for the comparisons.

2.6 Statistical comparisons

All data were compared using statistical methods using the SPSS Statistics software (SPSS 16 for Mac, SPSS Inc., IL, USA). All statistical tests were performed with a 5% confidence level ($\alpha = 5\%$). To check if all groups came

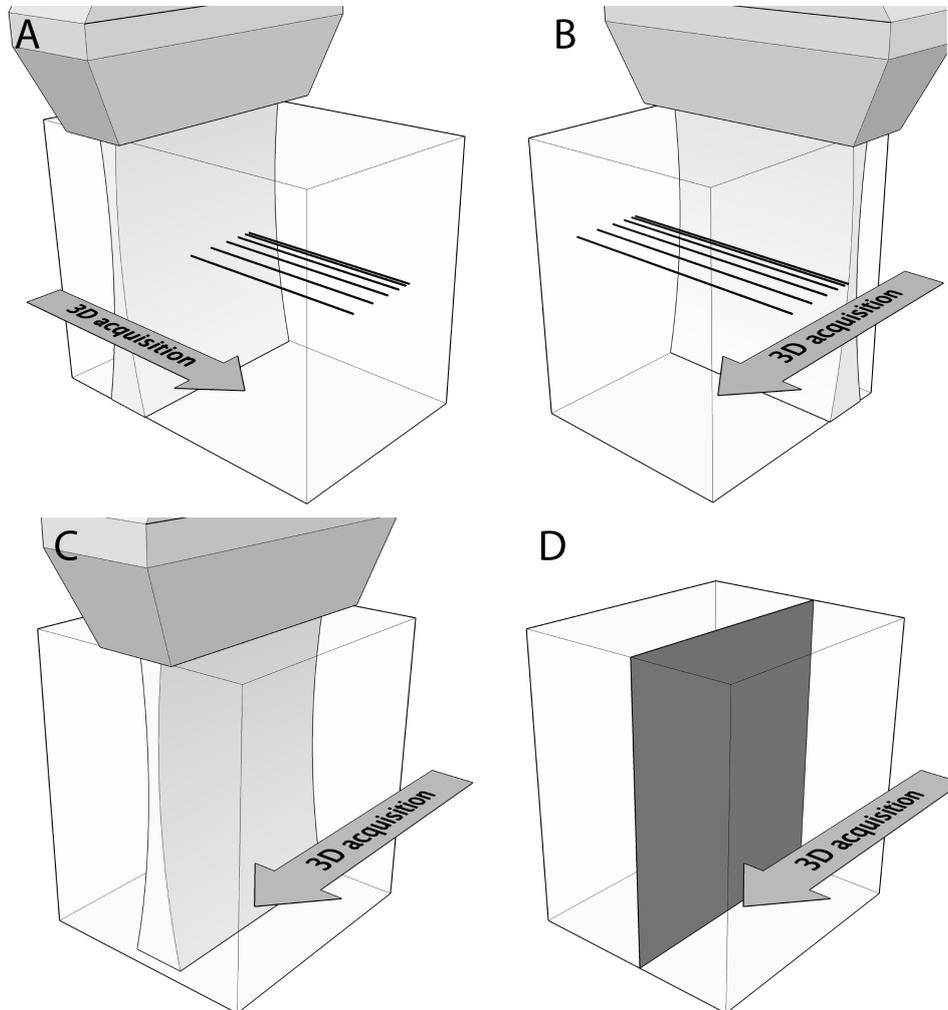


Figure 4: Illustrations of ultrasound sectors. The 3D acquisition scan direction is indicated, and the reconstruction volume is shown as a transparent box. (A) A 3D scan along the threads of a resolution phantom. The threads are seen as points in the ultrasound sector. (B) A 3D scan across the threads of a resolution phantom. The threads are seen as lines in the ultrasound sector. (C) Illustration of a single original ultrasound image acquired orthogonal to the input images in the 3D reconstruction. (D) Illustration of a 2D anyplane through the 3D volume in the same position as the original image in (C).

from the same distribution, an overall test was applied. In case of rejection, each group was tested against the others.

Each group was tested for normal distribution using the Shapiro-Wilk (SW) test to check for the possibility to use parametric statistics. For unrelated samples, if normality was accepted for all groups the Analysis of Variance (ANOVA) test was used for the overall statistics and the Bonferroni multiple-comparisons procedure was used for the pairwise comparisons. If normality was rejected the non-parametric Kruskal-Wallis (KW) test was used for the overall statistics and the Mann-Whitney-Wilcoxon (MWW) test was used for the pairwise comparisons. For test 1 with related samples the Friedman test was used for the overall statistic and the Wilcoxon Signed Rank (WSR) test was used for the pairwise comparisons. All statements in the text regarding differences (e.g.: performed better/poorer than, better/worse result, harder to identify, more accurate detection) are based on statistically significant results, even if this not mentioned explicitly every time.

3 Results

3.1 Compare the reconstruction algorithms abilities to correctly recreate removed data (test 1)

The differences between the slice with the removed data and the data values from the anyplanes in the same positions in the reconstructed volumes were plotted as curves with RMS results for each removed percentage. See Figure 5 as one example showing the RMS values from the different reconstructed volumes using the digital video as input in the scans across the forearm. All RMS values for the different removed percentages were combined for the statistical comparisons (Table 1). Table 1 shows results from the scans taken either across or along the forearm.

Statistically significant differences were found between the following reconstruction algorithms: In the scans across the forearm, representing data with high degree of variation, the large kernel performed poorer than the other two reconstruction algorithms for the digital data, and for the unprocessed video the small kernel performed better than the large. Also when combining the data from all three video sources the small kernel performed better than the large. In the scans along the forearm, representing data with little variation, the PNN reconstruction gave a better result than the others for the analog and digital video while it gave a worse result for the unprocessed data. For the unprocessed data in the scan along the forearm the large kernel reconstruction performed better than the small kernel that performed better than the PNN reconstruction. The differences between the scan across and along the forearm were visible as higher mean values in the scans across the arm (Table 1). When comparing values in Table 1 it was

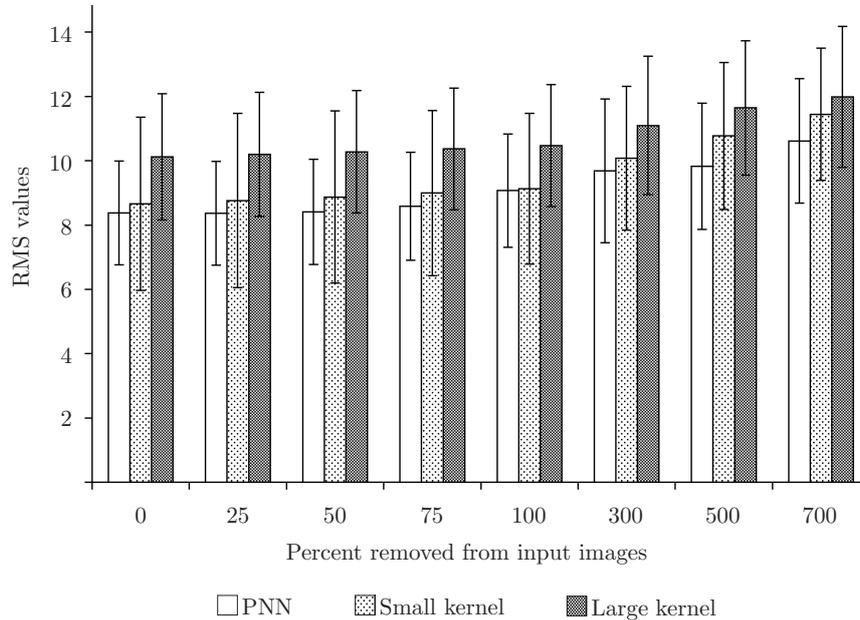


Figure 5: RMS values for the three different reconstruction algorithms for the digital video acquired across the forearm. Mean and standard deviation values are presented for each removed percentage. All RMS measurements for the large kernel are larger than the other two, and the difference is statistically significant.

only possible to compare values from the same data source because of the different characteristics of the data sources.

3.2 Visual comparisons of structures placed increasingly closer (test 2a)

The data from the comparisons can be illustrated as graphs showing the number of threads and the difficulty to identify the threads (Figure 6, Table 2). The compared images came from scans both along (Figure 4A) and across (Figure 4B) the threads (See also Figure 7A and B for examples of the images), the scans along the threads representing a comparison in the lateral direction of the ultrasound images and the scans across the threads representing a comparison in the elevation direction as related to the ultrasound image (Figure 2).

The results showed no statistically significant differences between the reconstruction algorithms, each with a median of 5 identified structures (Table 2). The identification of only 5 structures means that the structures with the distance of 1 mm could not be separated. Comparing the results from the anyplanes through the reconstructed volumes with the original 2D

Table 1: RMS values obtained by removing a percentage of data (0% to 700% removed) from selected input images. The scans across the forearm represent data with large variations while the scans along the forearm represent data with little variation from one image to the next. The values in the table between data sources are not comparable.

	Analog	Digital	Unprocessed
<i>Scan across arm (large variation between images)</i>			
PNN	8.38±2.42	9.12±2.45	16.93±3.43
Small kernel	7.57±1.46	9.59±1.98	16.10±2.96
Large kernel	7.50±1.37	10.77±2.01	17.12±3.19
<i>Scan along arm (small variation between images)</i>			
PNN	5.04±1.01	6.83±1.31	14.07±0.99
Small kernel	5.68±0.96	7.41±1.52	13.21±1.40
Large kernel	5.83±1.04	7.49±1.52	12.86±1.17

ultrasound images, the original images gave statistically significantly better results with a median of 6 identified structures (Table 2), meaning that the structures with the smallest distance of 1 mm could be separated in the majority of the observations. The comparison of the data sources showed that for all tests except the structure count from the scan along the threads, the unprocessed data produced volumes where fewer structures could be identified and they were harder to identify.

3.3 Comparisons based on image measurements of spatial resolution (test 2b)

Parabolic curves were matched to plots through the centers of the scanned wires (Figure 7) and these curves were evaluated at 6 dB and 20 dB levels below the maximum value. The 6 dB values are presented in Table 3.

There were no statistically significant differences between the thread measurements from the different reconstruction algorithms, but they performed better than the original images in some of the tests: In the lateral direction all reconstruction algorithms gave better results than both the original images and the resampled version of these images for both 6 dB and 20 dB. For 6 dB, the measurements in the lateral direction based on the anyplanes through the reconstructed volumes had a mean thread width of 1.17 mm while the original images had a mean width of 1.32 mm. In the axial direction for the 6 dB evaluations, only the small kernel was better than the original images with a mean thread height of 0.67 mm compared to

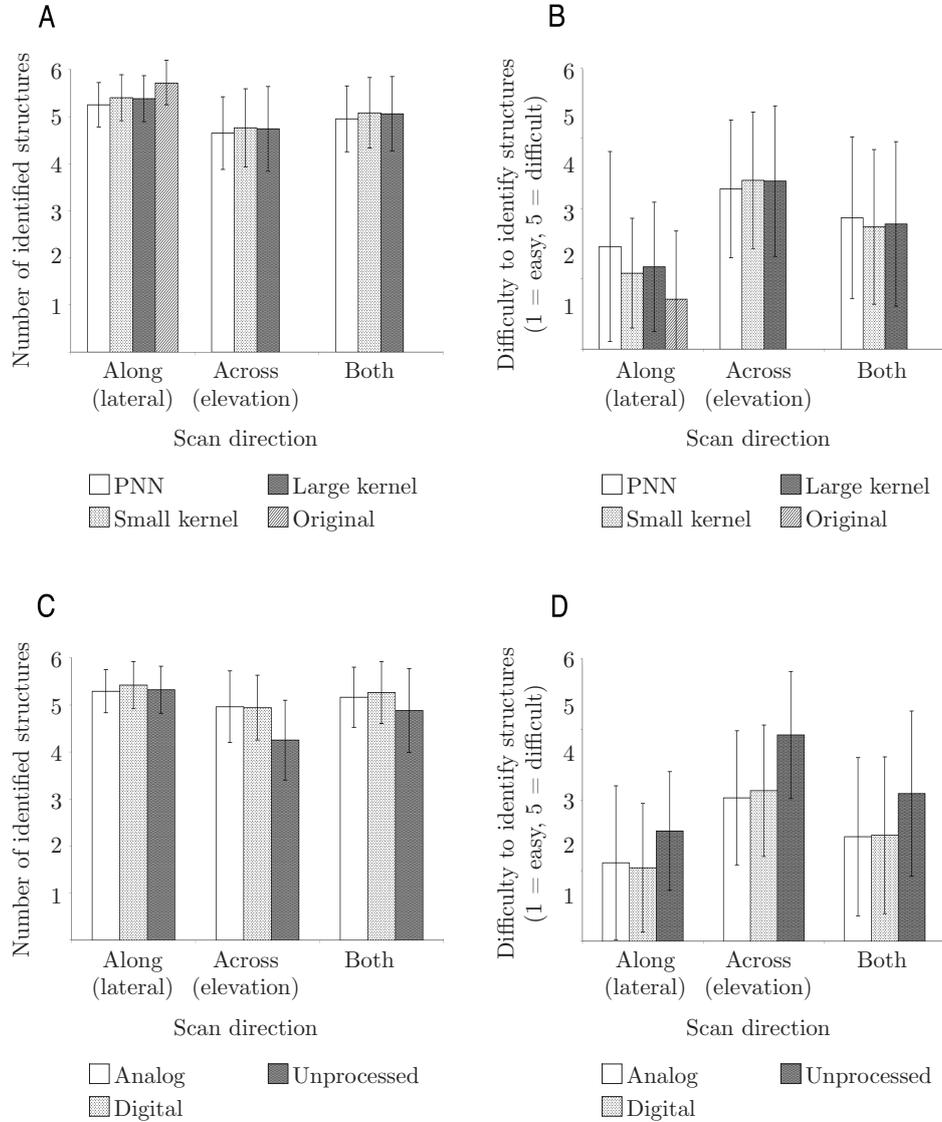


Figure 6: Identification of 6 structures placed increasingly closer. Results from the scans along, across and both combined. Mean and standard deviation values are shown in the graphs. The six threads were identified by eight volunteers. (A) Number of identified threads for the reconstruction algorithms and original images. (B) The difficulty to identify the threads for the reconstruction algorithms and original images. (C) Number of identified threads for the different data sources. (D) The difficulty to identify the threads for the different data sources.

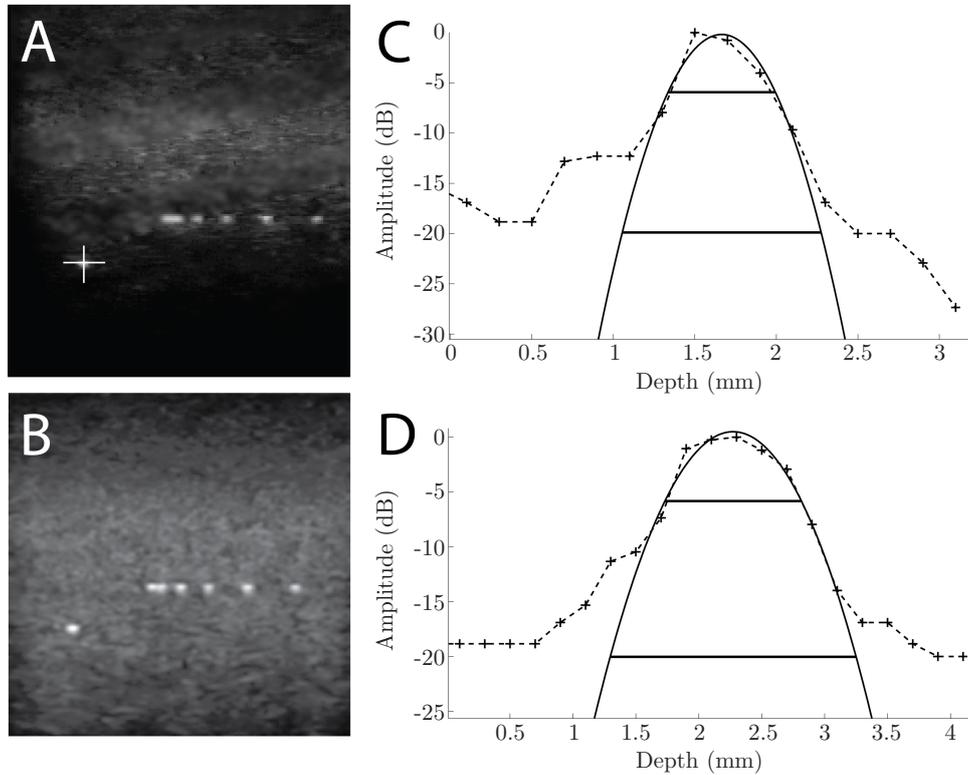


Figure 7: Example of thread size as a measurement of spatial resolution in the axial and lateral direction. The images are based on an anyplane image from a PNN reconstructed volume with analog video as input. (A) The anyplane slice through the ultrasound volume with an illustration of where (crosshairs) the intensity plots in the axial and lateral direction were collected. The horizontal measurement give the elevation resolution when the anyplanes come from the volumes scanned across the threads. The other structures in this image were used in test 2a. (B) The original ultrasound image from the analog video. Pixel values have been modified with the levels function of an image processing application to provide an image with more contrast in the paper. (C) A dB plot of the intensity values in the axial direction from (A) with a fitted parabolic curve. (D) A dB plot of the intensity values in the lateral direction from (A) with a fitted parabolic curve.

Table 2: Number of identified small structures. Mean and standard deviation values for the different combinations of reconstruction algorithm and data source.

	Analog	Digital	Unprocessed
<i>Scan along structures (lateral direction)</i>			
PNN	5.17±0.38	5.42±0.50	5.17±0.48
Small kernel	5.33±0.48	5.46±0.51	5.42±0.50
Large kernel	5.38±0.49	5.38±0.49	5.37±0.49
Original	5.50±0.52	5.94±0.25	5.69±0.48
<i>Scan across structures (elevation direction)</i>			
PNN	4.83±0.82	4.88±0.74	4.25±0.61
Small kernel	5.00±0.72	4.92±0.65	4.38±0.97
Large kernel	5.04±0.76	5.04±0.69	4.12±0.95
<i>Both scan directions combined</i>			
PNN	5.00±0.65	5.15±0.68	4.71±0.71
Small kernel	5.17±0.63	5.19±0.64	4.90±0.93
Large kernel	5.21±0.65	5.21±0.62	4.75±0.98

the mean height of the original images of 0.80 mm. Both the small and the large kernel performed better than the resampled original images. When fewer groups were compared compared without regard to the other groups more differences were detected also for the axial resolution: For 6 dB the PNN algorithm performed better than both the resampled and original images, and also the large kernel performed better than the original images. For 20 dB both the PNN and small kernel reconstruction performed better than the original images. For the 20 dB evaluations in the axial direction, all reconstruction algorithms gave better results than the resampled original images while none showed statistically significant difference from the original images. The results from the data sources showed that the unprocessed ultrasound performed poorer than both analog and digital video in both axial and lateral direction for 6 dB and 20 dB with a measured mean axial thread height for 6 dB of 0.79 mm and a measured mean lateral thread width of 1.49 mm compared to a measured axial thread height of 0.69 mm and measured lateral width of 1.21 mm for the processed video sources. The only difference in the elevation direction was that analog video performed better than unprocessed video for 20 dB.

Table 3: Image measurements of spatial resolutions. The results are presented as mean and standard deviation values in mm, and the measurements are from a fitted parabolic curve evaluated as 6 dB (see Figure 7 for example).

	Axial	Lateral	Elevation
<i>Analog</i>			
PNN	0.65±0.09	1.15±0.04	1.86±0.38
Small kernel	0.69±0.05	1.18±0.05	2.10±0.31
Large kernel	0.69±0.06	1.16±0.07	2.10±0.32
Original	0.74±0.01	1.28±0.01	–
Resampled	0.76±0.01	1.27±0.02	–
<i>Digital</i>			
PNN	0.66±0.15	1.08±0.15	1.58±0.50
Small kernel	0.62±0.07	1.11±0.06	2.31±0.62
Large kernel	0.65±0.08	1.16±0.08	2.24±0.50
Original	0.69±0.03	1.28±0.03	–
Resampled	0.73±0.02	1.25±0.01	–
<i>Unprocessed</i>			
PNN	0.76±0.20	1.22±0.19	1.90±0.43
Small kernel	0.71±0.05	1.23±0.08	2.37±0.51
Large kernel	0.74±0.07	1.28±0.06	2.34±0.47
Original	0.98±0.03	1.80±0.03	–
Resampled	0.78±0.02	1.45±0.02	–

3.4 Visual comparisons of small, barely visible structures (test 2c)

Examples of the images presented to the eight volunteers are shown in Figure 8. The original images allowed more accurate structure detection than the anyplanes through the reconstructed volumes both for the number of structures identified (Figure 9, Table 4) and the difficulty to identify the structures, by deviating from the true number of structures (eleven) by a median of 0. The deviation for the large kernel had a median of 2.5 structures, which was statistically significantly better than the PNN reconstruction with a median deviation of 4 structures from the true number. The small kernel could not be separated from the other reconstructions with a median deviation of 3 structures. For the scans along the cylinders the anyplanes from the PNN reconstruction did not allow the volunteers to identify the same number of structures as the other reconstructions with a median deviation of 3 structures compared to the others with medians of 1 structure deviation. The structures in anyplanes from the PNN reconstructions were more difficult to identify than the small kernel. In the scans across the cylinders there were no statistically significant differences between the reconstruction algorithms with a combined median deviation of 6 structures. The results from comparing only two groups with each other also showed that the structures in the volumes from the PNN reconstruction were statistically more difficult to identify than both the other reconstruction algorithms for the scan direction along the cylinders and for both directions combined. The image sources could not be statistically significantly separated for the scan along the cylinders with a combined median deviation from the true number of 1 structure. In the scan across the cylinders, the unprocessed data performed poorer than the other image sources with a median deviation of 7 structures compared to a median of 6 for the digital video and 5 for the analog video. The digital and analog video could, however, not be separated through statistical significance. When combining both scan directions, the analog video with a median deviation of 3 performed statistically significantly better than the unprocessed data with a median deviation of 4. The digital video with a median deviation of 3 could not be separated from the other algorithms.

3.5 Visual comparisons of tissue data (test 2d)

Figure 10 contains examples of the images used in the comparisons showing both original ultrasound images and anyplanes through reconstructed volumes, all originating from the same position. The differences between the reconstruction algorithms were not statistically significant, but all performed worse than the original images both with respect to ordering and quality. The data sources could not be separated in the scans along the forearm, and

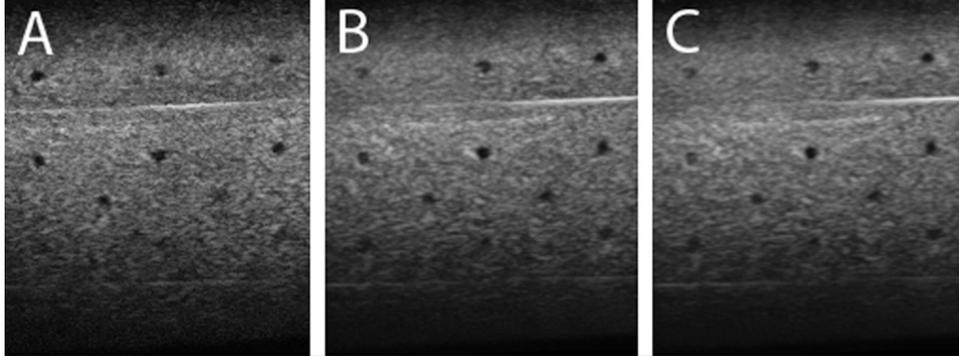


Figure 8: Example of anyplane images of small structures (cylinders). The anyplanes are from reconstructions based on digital video as input. The scan direction of the input images are along the cylinders. (A) PNN reconstruction. (B) Small kernel reconstruction. (C) Large kernel reconstruction.

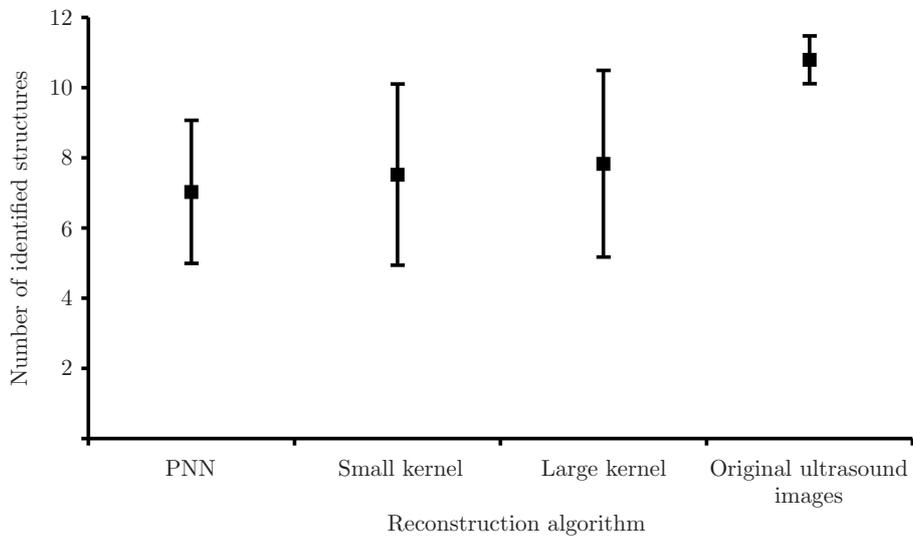


Figure 9: Number of identified structures. The results are shown as the deviation from the true number of structures (=11) presented as mean and standard deviation values. Results from the three different reconstruction algorithms compared with original ultrasound images. The results from both the scans across and along the structures are combined.

Table 4: Comparisons based on the number of identified structures. The values shown are the deviation from the true number of structures (=11) presented as mean and standard deviation values.

	Analog	Digital	Unprocessed
<i>Both scan directions combined</i>			
PNN	3.63±1.47	3.75±2.02	4.52±2.46
Small kernel	3.17±2.30	3.23±2.55	4.04±2.82
Large kernel	2.83±2.30	2.90±2.78	3.77±2.81
Original	0.25±0.77	0.19±0.54	0.19±0.75
<i>Scan across cylinders (elevation direction)</i>			
PNN	4.54±1.28	5.04±1.73	6.62±0.97
Small kernel	5.12±1.36	5.46±1.56	6.58±0.93
Large kernel	4.88±1.39	5.29±1.83	6.42±0.93
<i>Scan along cylinders (lateral direction)</i>			
PNN	2.71±1.00	2.46±1.35	2.42±1.47
Small kernel	1.21±0.98	1.00±0.66	1.50±1.38
Large kernel	0.79±0.41	0.50±0.72	1.12±0.80

the quality score could also not be separated in the scans across the forearm. The results from the ordering in the scans across the forearm showed that the digital video was preferred, followed by the unprocessed data, and the analog video was evaluated as the worst quality. When combining the answers from both scan directions the ordering gives the same result as for scans across the forearm, but now also the quality score shows the analog video to perform poorer than the other sources.

3.6 Compare the correctness of the reconstructed volumes geometry (test 3)

None of the reconstruction algorithms could be separated from the original ultrasound images (used as gold standard) while also being different from the other reconstruction algorithms (Table 5). For the scans across the cylinder the analog and digital video differs from the original images in the length measurements, by a difference in measured mean value of 0.33 mm for the analog video and 0.16 mm for the digital video (Table 6). The unprocessed data differs from the original images in the width measurements, by a difference in measured mean width of 0.28 mm (Table 6).

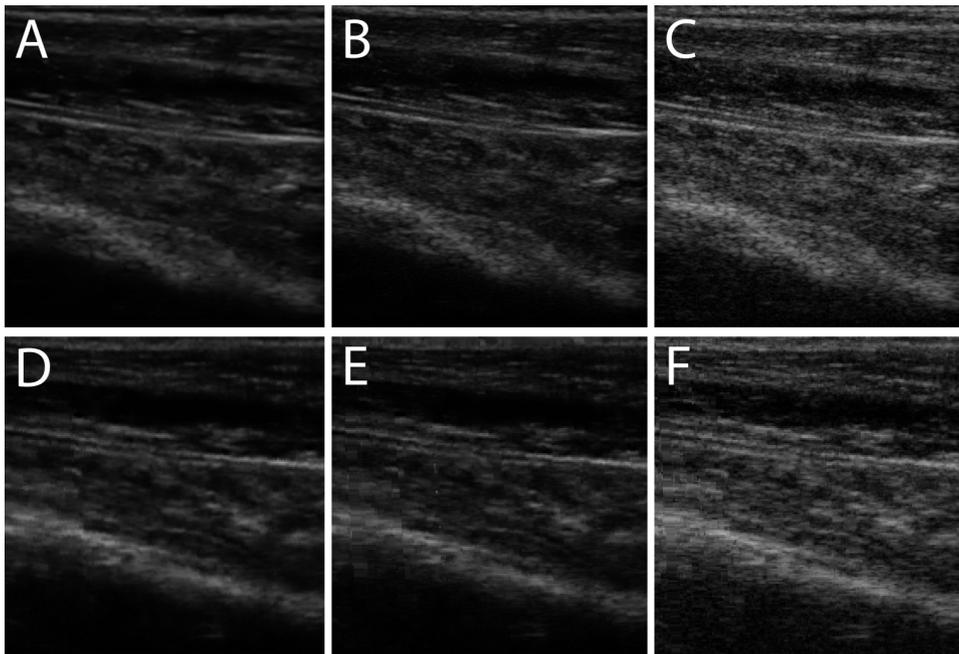


Figure 10: Original ultrasound images and anyplanes of the underside of the forearm. All images are originating from the same position. All anyplane images are reconstructed with a small kernel, and the anyplanes are obtained orthogonal to the reconstruction input images (See Figure 4C and D). (A) Original analog image. (B) Original digital image. (C) Original unprocessed image. (D) Anyplane with analog ultrasound video as reconstruction input. (E) Anyplane with digital ultrasound video as reconstruction input. (F) Anyplane with unprocessed ultrasound data as reconstruction input.

Table 5: Comparisons of reconstruction algorithms based on the measured length (= 6.66 mm), width (= 3.06 mm) and height (= 3.07 mm) of a small cylinder. The absolute of the differences in mm are shown for the results that are statistically significantly different from the original images (“gold standard”). No results where statistically significant when combining both scan directions. NS means not statistically significant.

US Scan direction	Elevation (length)	Lateral (width)	Axial (height)
<i>Along cylinder</i>			
PNN	0.27	NS	NS
Small kernel	0.36		
Large kernel	0.42		
	Elevation (width)	Lateral (length)	Axial (height)
<i>Across cylinder</i>			
PNN	NS	0.13	NS
Small kernel		0.22	
Large kernel		0.25	

Table 6: Comparisons of input ultrasound sources based on the measured length (= 6.66 mm), width (= 3.06 mm) and height (= 3.07 mm) of a small cylinder. The absolute of the differences in mm are shown for the results that are statistically significantly different from the original images (“gold standard”). No results where statistically significant when combining both scan directions. NS means not statistically significant.

US Scan direction	Elevation (length)	Lateral (width)	Axial (height)
<i>Along cylinder</i>			
Analog	0.29	NS	NS
Digital	0.32		
Unprocessed	0.44		
	Elevation (width)	Lateral (length)	Axial (height)
<i>Across cylinder</i>			
Analog	NS	0.33	NS
Digital	NS	0.16	
Unprocessed	0.28	NS	

Table 7: Short summary of the most important results from the different comparison tests.

	Reconstruction algorithms	Data sources
Test 1a: Remove data	Large k. worst (for varying images)	–
Test 2a: Identification of close structures	Equal performance	Unprocessed worst
Test 2b: Image resolution measurements	Equal performance	Unprocessed worst
Test 2c: Identification of small structures	PNN worst	Unprocessed worst
Test 2d: Tissue data comparisons	Equal performance	Digital best, Unprocessed middle, Analog worst
Test 3: Geometry measurements	PNN barely best, Large k. middle, Small k. barely worst	Indecisive (Digital slightly better)

4 Discussion

In image-guided surgery it is important to have reliable images available from all directions, and the quality of the reconstructed 3D ultrasound data to be used for guidance and therapy is important. In this paper we have studied if the choice of 3D reconstruction algorithm and input data source may affect image quality and resolution in a 3D volume as evaluated by various methods. A quick summary of the results is presented in Table 7.

4.1 Comparing the performance of the reconstruction algorithms

Our study showed that the performance of the 3D ultrasound reconstruction algorithms varied from test to test. The small kernel reconstruction algorithm performed slightly better overall than the two other tested algorithm implementations, but a conclusion depends on how the tests are

being weighted due to importance. If all tests are to be equally weighted, the choice of reconstruction algorithm doesn't matter, and a good choice may be the fastest algorithm, PNN. All quality measures may, however, not be equally important, and the preferred algorithm may be selected based on intended application.

The method of removing input data and measuring the reconstruction algorithms success at recreating removed data is the "classical" comparison test of reconstruction algorithms [35]. For the scan along the forearm, the PNN algorithm surprisingly has the best performance for the analog and digital video. However, this scan direction creates images that change only slightly from one image to the next. For the unprocessed data, however, the PNN reconstruction has the worst performance, probably because of the increased level of information in these images. The scans across the forearm were more interesting, since they gave more variation from one image to the next, and the ability to perform well under such situations is more important in a practical situation, so these results were given more weight in Table 7 and our conclusions. For the scans across the forearm the reconstruction with the largest ellipsoid kernel around input pixels performed worse than the smallest kernel for the digital video and the unprocessed data. The reconstruction with the largest kernel performed also worse than the PNN reconstruction for the digital video. The reason for the poorer performance of the large kernel may be that we used a lateral resolution for the reconstruction based on a depth of 32 mm when the image depth was 40 mm, and thus blurring the data too much in a large part of the reconstructed volume.

In addition to test 1, only test 2c and test 3 showed any statistical differences in reconstruction algorithms. All the comparisons of counted small structures in test 2c showed that the PNN reconstruction performed poorer than the small kernel in two cases and also poorer than the large kernel in two cases (Table 4). When only comparing two groups at a time in test 3, the PNN algorithm performed better than the small kernel in one comparison and better than the large kernel in three. However, the differences in measured mean values were very small: less than the size of the volume cells (< 0.2 mm), so even if the differences were statistically significant they may not result in much practical difference.

In image-guided surgery, it is most important to identify very small structures in the correct location, so the small or large kernel may be a better choice than the current implementation of the PNN algorithm. However, it should be noted that the small structures in test 2c were so small that a minor position change made the structures disappear. In the test we used the position of a single ultrasound image showing all the structures. The nature of the PNN reconstruction algorithm implementation just replacing existing data in the 3D volume with the latest data might have changed the position of the small structures just enough that they could not be identified. This does not mean that they could not have been identified in a neighboring

position, just that the PNN reconstruction algorithm introduced a small position bias. It is possible to remove this bias by changing the bin-filling [24] of the PNN implementation to an averaging [30, 34] or by keeping the maximum value [30] instead of using only the most recent value.

4.2 Comparing reconstructions with directly acquired original ultrasound images

All tests, except test 1, may be used to test for differences between original, directly acquired 2D ultrasound images and anyplanes through the reconstructed volumes. Most tests showed the original ultrasound images to perform better than the anyplanes, but test 2b showed the opposite.

An interesting result from test 2b was that the means acquired for resolution measurements on anyplanes from the reconstructions showed better resolutions than those measured on the original images (Table 3). The down-sampling of the original ultrasound images seemed to lead to slightly better results, but this difference was not enough to explain why the anyplanes in the lateral direction from the reconstructions performed better than original ultrasound images for this test. Also, in the axial direction for the 6 dB reduction, the small kernel obtained better results than both sets of original data, and when fewer groups were compared also the other algorithms performed better than both sets of original data. However, the explanation to the differences may be that the reconstruction algorithms processed the input data to make full use of the 8-bit range, while the original images were unprocessed. This processing increased the distance between the pixel values and the 6 dB and 20 dB measurements were not at the same level for the anyplanes and the original images.

Another interesting observation was obtained when comparing results for 6 dB reductions in Table 3 with the theoretical results shown in Figure 3. The differences between measured and theoretical resolutions were quite large: The measured axial resolution was about 4.7 times larger than the theoretical, the measured lateral resolution about 2.7 times larger and the measured elevation resolution about 2.1 times larger. For the axial direction, the assumption about a very short transmitted pulse for the theoretical calculation was clearly too optimistic, and the transmitted pulse was probably several wavelengths long. The thickness of the thread may also lead to a slightly increased measurement, even if the thread diameter was only 0.1 mm. When imaging the thread we made sure that a focus point was at the approximate depth of the thread so that the lateral resolution should be comparable to the theoretical values. However, the formula for the lateral resolution (3) did not take into account apodization, and this may explain some of the difference. The available data of the ultrasound probe may also provide additional information: The probe center frequency is 7.2 MHz, and even if the imaging frequency was set to 10 MHz, the real center frequency of

the submitted pulse may be closer to 7.2 MHz. In addition the absorption in the imaged area of the ultrasound phantom was 0.7 dB/(cm MHz). Using $f_t = 7.2$ MHz in (3) and $a = 0.7$ dB/(cm MHz) in (4) resulted in the measured lateral resolution being only 1.8 times larger than the theoretical resolution, and measured elevation resolution being only 1.4 times larger than the theoretical. The elevation focus is fixed at depth 16 mm, and compared to the wire depth of 27 mm the difference for the elevation direction was quite understandable. To explore the differences between the theory and practical results further it may be necessary to perform hydrophone measurements of the transmitted pulse, especially to determine the differences in axial resolution.

Test 2c showed that directly acquired ultrasound images performed better than the reconstruction algorithms for the purpose of identifying small structures in phantoms. Test 2d showed that the human observers preferred directly acquired ultrasound images to anyplanes through reconstructions, also for tissue data. Test 3 measuring all axes of a cylinder showed that in most cases the distances measured on anyplanes was not different from directly acquired ultrasound images. Still, Tables 5 and 6 showed that images related to the lateral and elevation reconstruction directions gave significant differences from the original ultrasound data for larger distances (lengths measurements) but not for smaller measurements (width measurements), except for the unprocessed data. Another interesting result was that all the length measurements in anyplanes in the lateral direction (scans across cylinder) were larger than the measurements in the original images while the length measurements in the elevation direction (scans along cylinder) were smaller. The reason for this difference may be inaccuracies in the probe calibration, as this comparison test is dependant of accurate orientation. The original images for the length measurements were obtained orthogonal to the input images in the scan along the cylinder, and a small difference in orientation may have resulted in a relatively large difference in length measurement.

4.3 Difference between data sources

Both processed video sources performed generally better than the unprocessed data. The digital video was the data source with the best results throughout all the tests. However, it could not be separated from the analog video in tests 2a, 2b, 2c and 3. On the other hand, it performed better than the analog video in the tissue comparison test 2d. The analog video showed good results in several tests and was often better than the unprocessed data, but not statistically significantly better than the digital video in any of the comparisons. However, in the volunteers evaluation of the tissue data, the analog video performed poorer than both the digital and unprocessed video.

When observing the input video sources, it was obvious that the unprocessed data had more information with the higher resolution, while the processed video sources represented a resampling of the sampled data, which could lead to a loss of information. So far not much work is published comparing the quality of the unprocessed with processed data. Use of the unprocessed data could probably lead to better results in comparisons; however, our results indicate the opposite conclusion. The reason for this may be that the reconstruction algorithms do not process the data as thoroughly as the ultrasound scanner does. This result may be seen as a proof that the image processing done in the ultrasound scanner does improve the image quality as seen from the user perspective even if the resolution may be somewhat reduced. Another important aspect of this is that if unprocessed data are used directly there may be a reduction of quality and the images may be harder to interpret even if the data contain more information. To prevent this, a processing similar to that of the ultrasound scanner could be performed. It should however be noted that all comparisons are based on data from a single scanner using one ultrasound probe, and the quality of the unprocessed and processed data may vary between different probes and scanners, and the results of our study can hence not be generalized.

4.4 Comparing input sources and reconstruction algorithms

Several studies try to compare the performance and quality of the reconstruction algorithms [24, 31, 32, 35, 36, 48, 49]. The most commonly used comparison method is the method of removing a percentage of the input data and then determining the algorithms ability to recreate this data. An RMS error value or an equivalent value is usually used to disclose the quality difference. A specific slice is selected, usually in the middle of the ultrasound volume, and a set of different percentages are selected for removal of data [35]. Coupé et al. [48] uses a variation over this and removes input slices (1, 2, 3, 4, 5 and 6) over the whole input data set and calculates a mean and standard deviation for the MSE (Mean Squared Error) values of the different slices. The advantage of this method is that it is an objective method that may be done automatically and is usable with most kinds of ultrasound data.

Drawbacks of this method are that it only compares one aspect of reconstruction quality and the results are dependant of the imaged tissue. In addition, this comparison method is not suitable to compare different ultrasound data sources as we have done in this paper. We have still used this existing method, but in addition we have devised several additional tests for the purpose of comparing both reconstruction algorithms and data sources for the reconstructions.

It is important to note that in our use of the test we performed the statistical comparisons over the full range of removed data (0 %, 25 %, 50 %, 75 %, 100 %).

75 %, 100 %, 300 %, 500 % and 700 %) without regard to the data having more samples in the range from 0 % to 100 %, leading to a higher weight given to this range. Also, a situation may arise where the RMS values of one reconstruction algorithm are better for one range of percentages but worse for another range when compared to another reconstruction algorithm. In this situation, the statistical tests of the whole range may be indecisive, while a test of a smaller range might give a result. The described situation happened for the combination of all data sources in the scan along the forearm, and as may be seen in Table 1 this test could not differentiate between the data sources. We decided not to perform any additional tests of specific smaller ranges as we wanted to focus on the overall score.

Several of the tests we have used in this paper deal with the practical resolution of the ultrasound images and volumes, especially tests 2a, 2b, 2c and 3. However, test 2c was also highly dependent on accurate positions of the small structures or the ultrasound probe positions related to the structures. Some of the differences between the reconstruction algorithms and original 2D ultrasound images may be due to small errors in the probe calibration having a total mean error of 1.05 mm. The structures being only 1.5 mm in diameter and 2.4 mm long may easily be missed with this calibration accuracy since the comparisons were based on a position and orientation with the structures visible in the original images. A small error in either position or orientation may lead to several missing structures. One of the important features of ultrasound is how good the image quality appear to a human and how easy it is to interpret the image. Test 2d try to test these aspects, but the test is very unspecific. To detect more important differences a set of specific tests could be created asking volunteers (preferably clinical ultrasound users) to evaluate sets of images from different clinical cases.

We have tried to create tests that may represent different uses of the ultrasound volumes, but the range of tests is not exhaustive. Our comparison tests also have a varying degree of manual interaction and may not be practical in all situations. Several other tests could be devised in addition to the tests we have performed in this paper. One idea is to perform an automatic segmentation of a known structure and compare volume sizes [33]. However, our initial experiments showed that the segmentation algorithms that we had access to created quite different segmented volumes from the same reconstructed ultrasound volume with only small changes in parameters, and it was difficult to get comparable segmentations from volumes originating from different input sources. Another possible test is the comparison of measurements of relatively large distances. An automated method like the one described in Lindseth et al. [51] might be possible, and this test may detect more and larger differences than the test we have performed in this paper, which only compares measurements of small distances.

The general problem with comparison tests for reconstructed ultrasound volumes is that there are several error sources that may affect the results,

and these error sources may be larger than the factors we want to compare. Our method of comparing several aspects of the reconstructed volumes is a way to limit the error sources when looking at all tests combined. Still the error sources should be minimized before the comparisons. We have also applied statistical comparisons of all the data collected in the tests since this gives a better way of knowing if one value really is better than another or if it only is natural variation.

4.5 Relevance and further work

3D probes may be used for acquisition of 3D data instead of 2D probes. 3D probes will allow both easy acquisition of near real-time 3D data and the possibilities of 3D data with time as a fourth dimension. However, in IGS, tracked 2D probes have a few advantages over 3D probes: The data stream from the ultrasound probe is limited, so still it is possible to get a higher resolution volume from a 2D probe. In IGS, positioned intraoperative data is easier to correlate to preoperative data, so a tracking sensor is usually needed. 3D probes are shown to interfere more with electromagnetic tracking systems than 2D probes [52]. The volume covered by a 3D probe is much smaller than is practical in most IGS applications, and to overcome this limitation, an application must be created that combines data from the 3D probe. The data from the 3D probe are not easily available to third party ISG applications, needed to integrate the data with preoperative data, or for combining the data into larger volumes.

Most new ultrasound scanners support the DICOM standard so that it should be possible to get access to digital data for 3D reconstructions. Even if only a stack of 2D images is stored, a 3D reconstruction is still possible with this data if the 2D images are tagged with accurate global time-tags. The positions may then be recorded at the same time with their own time-tags, and a time calibration [44, 53] may be performed to match the positions with the images. Most applications for 3D ultrasound reconstructions require a near real-time implementation, and currently the video grabber approach is the fastest solution in most cases for freehand 3D reconstructions, as DICOM don't support real-time images yet. However, DICOM working group 24 focuses on developing DICOM standards and services for Image Guided Surgery. [54] This work may result in real-time protocols that open up for easily accessible real-time digital ultrasound images suitable for 3D reconstructions. Another option is the method we have used in this paper by using an ultrasound scanner that allows real-time access to digital data.

All comparisons of data in this paper are performed with data from a single scanner using one ultrasound probe. To get more general results it would be interesting to collect data from more scanners and probes. Only a few ultrasound scanners supply unprocessed data (e.g. RF) to third parties without special collaboration agreements. However, a study using data from

various scanners and probes could make it easier to arrive to general conclusions regarding choice of input data. Examples of scanner producers with scanners that may supply unprocessed data are: Winprobe (North Palm Beach, FL, US), VeraSonics (Redmond, WA, US), and Terason (Burlington, MA, US), in addition to the Sonix RP scanner from Ultrasonix used in our study.

5 Conclusion

The present study shows that the choice of data source may be more important than the choice of reconstruction algorithm in order to achieve high quality image volumes from tracked freehand 2D ultrasound data. Furthermore, scan converted digital and analog data gave better results than unprocessed ultrasound data in our comparison tests performed with one scanner. Overall, digital video performs slightly better than analog video, but in most cases the two video sources were difficult to separate by the comparison methods, indicating that the quality loss of using flexible video-grabbing solutions over scanners with a digital interface may not be significant. It must be taken into account that the conclusions on data source quality were based on comparisons performed on data from a single scanner using one ultrasound probe, and that quality and processing may differ between probes and scanners. More work including comparison tests on several scanners and probes should therefore be performed in order to obtain more general conclusions. By giving all the comparisons the same weight, no reconstruction algorithm of those tested performs significantly better than the others in terms of image quality indicating that the fastest reconstruction method should be chosen, e.g. PNN. However, in general, the reconstruction algorithm must be selected according to the application and the intended usage of the 3D volume.

Acknowledgements

This work was supported by the Research Council of Norway, through the FIFOS Programme Project 152831/530 and Medical Imaging Laboratory (MI Lab); the Ministry of Health and Social Affairs of Norway, through the National Centre of 3D Ultrasound in Surgery; and by SINTEF Technology and Society. We want to thank Thomas Kuiran Chen and Purang Abolmaesumi at Queen's University, Kingston, Canada, for supplying us with both a phantom and C++ code for ultrasound probe calibration. We also want to thank Reidar Brekken (SINTEF, NTNU), Rune Hansen (SINTEF, NTNU) and Hans Torp (NTNU) for discussions regarding ultrasound theory, Sjur Gjerard (NTNU) for aid with laboratory work, Thor Andreas Tangen (NTNU) and Jochen Deibele (NTNU) for help with the Ultrasonix scanner,

and we like to thank the rest of the people at SINTEF Technology and Society, Medical Technology for volunteering for some of the comparison tests. The authors are thankful to the reviewers, who provided comprehensive and constructive suggestions and indicated that in addition to Ultrasonix there are two other systems on the market that allow access to the RF data: Winprobe (and possibly Verasonics).

References

1. Unsgård, G, Rygh, OM, Selbekk, T, Müller, TB, Kolstad, F, Lindseth, F, and Hernes, TAN. Intra-operative 3D ultrasound in neurosurgery. *Acta Neurochirurgica* 2006;148:235–253.
2. McCann, HA, Sharp, JC, Kinter, TM, McEwan, CN, Barillot, C, and Greenleaf, JF. Multidimensional ultrasonic imaging for cardiology. *Proceedings of the IEEE* 1988;76:1063–1073.
3. Hottier, F and Billon, AC, eds. 3D echography: status and perspective. Vol. F 60. *3D Imaging in Medicine*. Springer-Verlag Berlin Heidelberg, 1990.
4. Nadkarni, SK, Boughner, D, and Fenster, A. Image-based cardiac gating for three-dimensional intravascular ultrasound imaging. *Ultrasound in Medicine & Biology* 2005;31:53–63.
5. Prager, RW, Gee, A, Treece, G, and Berman, L. Freehand 3D ultrasound without voxels: volume measurement and visualisation using the Stradx system. *Ultrasonics* 2002;40:109–115.
6. Meairs, S, Beyer, J, and Hennerici, M. Reconstruction and Visualization of Irregularly Sampled Three- and Four-Dimensional Ultrasound Data for Cerebrovascular Applications. *Ultrasound in Medicine & Biology* 2000;26:263–272.
7. Welch, JN, Johnson, JA, Bax, MR, Badr, R, and Shahidi, R. A real-time freehand 3D ultrasound system for image-guided surgery. In: *Ultrasonics Symposium*. Ed. by Schneider, SC, Levy, M, and McAvoy, BR. Vol. 2. IEEE, 2000.
8. Sherebrin, S, Fenster, A, Rankin, RN, and Spence, D. Freehand three-dimensional ultrasound: implementation and applications. In: *Medical Imaging 1996: Physics of Medical Imaging*. Ed. by Van Metter, RL and Beutel, J. Vol. 2708. SPIE, 1996:296–303.
9. Ogawa, S, Itoh, K, Omoto, K, Cheng, X, Ohya, A, and Akiyama, I. Three dimensional ultrasonic imaging for diagnosis of breast tumor. In: *1998 IEEE Ultrasonics Symposium*. Ed. by Schneider, SC, Levy, M, and McAvoy, BR. Vol. 2. IEEE, 1998:1677–1680.

10. Amin, VR, Wang, B, Sonka, M, and Lauer, RM. Three-dimensional ultrasound imaging of vessel wall for evaluating atherosclerosis risk and disease. In: *Medical Imaging 2002: Ultrasonic Imaging and Signal Processing*. Ed. by Insana, MF and Walker, WF. Vol. 4687. SPIE, 2002:255–263.
11. Rohling, RN, Gee, AH, and Berman, L. Automatic registration of 3-D ultrasound images. *Ultrasound in Medicine & Biology* 1998;24:841–854.
12. Gilja, OH, Hausken, T, Olafsson, S, Matre, K, and Odegaard, S. In vitro evaluation of three-dimensional ultrasonography based on magnetic scanhead tracking. *Ultrasound in Medicine & Biology* 1998;24:1161–1167.
13. Detmer, PR, Bashein, G, Hodges, T, Beach, KW, Filer, EP, Burns, DH, and Jr., DS. 3D ultrasonic image feature localization based on magnetic scanhead tracking: In vitro calibration and validation. *Ultrasound in Medicine & Biology* 1994;20:923–936.
14. Moskalik, AP, Carson, PL, Meyer, CR, Fowlkes, JB, Rubin, JM, and Roubidoux, MA. Registration of three-dimensional compound ultrasound scans of the breast for refraction and motion correction. *Ultrasound in Medicine & Biology* 1995;21:769–778.
15. Laporte, C and Arbel, T. Combinatorial and Probabilistic Fusion of Noisy Correlation Measurements for Untracked Freehand 3-D Ultrasound. *IEEE Transactions on Medical Imaging* 2008;27:984–994.
16. Solberg, OV, Tangen, GA, Lindseth, F, Sandnes, T, Enquobahrie, AA, Ibanez, L, Cheng, P, Gobbi, D, and Cleary, K. Integration of a real-time video grabber component with the open source image-guided surgery toolkit IGSTK. In: *Medical Imaging 2008: PACS and Imaging Informatics*. Ed. by Andriole, KP and Siddiqui, KM. Vol. 6919. SPIE, 2008:69190Z–69199.
17. Ohbuchi, R, Chen, D, and Fuchs, H. Incremental volume reconstruction and rendering for 3-D ultrasound imaging. In: *Visualization in Biomedical Computing '92*. Ed. by Robb, RA. Vol. 1808. SPIE, 1992:312–323.
18. Chen, TK, Thurston, AD, Ellis, RE, and Abolmaesumi, P. A Real-Time Freehand Ultrasound Calibration System with Automatic Accuracy Feedback and Control. *Ultrasound in Medicine & Biology* 2009;35:79–93.
19. Berg, S, Torp, H, Martens, D, Steen, E, Samstad, S, Hoivik, I, and Olstad, B. Dynamic three-dimensional freehand echocardiography using raw digital ultrasound data. *Ultrasound in Medicine & Biology* 1999;25:745–753.

20. Martens, D and Gilja, OH. The EchoPAC-3D software for image analysis. In: *Basic and New Aspects of Gastrointestinal Ultrasonography*. Ed. by Ødegård, S, Gilja, OH, and Gregersen, H. Advanced series in Biomechanics. World Scientific Publishing, 2005. Chap. 10:305–329.
21. Thune, N, Gilja, OH, Hausken, T, and Matre, K. A practical method for estimating enclosed volumes using 3D ultrasound. *European journal of ultrasound* 1996;3:83–92.
22. Grønningsæter, Å, Kleven, A, Ommedal, S, Aarseth, TE, Lie, T, Lindseth, F, Langø, T, and Unsgård, G. SonoWand, an ultrasound-based neuronavigation system. *Neurosurgery* 2000;47:1373–1739.
23. Lindseth, F, Bang, J, and Langø, T. A robust and automatic method for evaluating accuracy in 3-D ultrasound-based navigation. *Ultrasound in Medicine & Biology* 2003;29:1439–1452.
24. Solberg, OV, Lindseth, F, Torp, H, Blake, RE, and Hernes, TAN. Freehand 3D ultrasound reconstruction algorithms – a review. *Ultrasound in Medicine & Biology* 2007;33:991–1009.
25. Trobaugh, JW, Trobaugh, DJ, and Richard, WD. Three-dimensional imaging with stereotactic ultrasonography. *Computerized Medical Imaging and Graphics* 1994;18:315–323.
26. Wein, W, Pache, F, Röper, B, and Navab, N. Backward-Warping Ultrasound Reconstruction for Improving Diagnostic Value and Registration. In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2006*. Ed. by Larsen, R, Nielsen, M, and Sporring, J. Springer, 2006:750–757.
27. Karamalis, A, Wein, W, Kutter, O, and Navab, N. Fast hybrid freehand ultrasound volume reconstruction. In: *Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling*. Ed. by Miga, MI and Wong, KH. Vol. 7261. SPIE, 2009:726114–726118.
28. Treece, GM, Prager, RW, Gee, AH, and Berman, LH. Correction of probe pressure artifacts in freehand 3D ultrasound. *Medical Image Analysis* 2002;6:199–214.
29. Zhang, Y, Rohling, R, and Pai, DK. Direct surface extraction from 3D freehand ultrasound images. In: *IEEE Visualization 2002 — VIS 2002*. Ed. by Moorhead, R, Gross, M, and Joy, KI. IEEE Computer Society, 2002:45–52.
30. Nelson, TR and Pretorius, DH. Interactive acquisition, analysis, and visualization of sonographic volume data. *International Journal of Imaging Systems and Technology* 1997;8:26–37.

31. Estépar, RSJ, Martín-Fernández, M, Caballero-Martínez, PP, Alberola-López, C, and Ruiz-Alzola, J. A theoretical framework to three-dimensional ultrasound reconstruction from irregularly sampled data. *Ultrasound in Medicine & Biology* 2003;29:255–269.
32. Estépar, RSJ, Martín-Fernández, M, Alberola-López, C, Ellsmere, J, Kikinis, R, and Westin, CF. Freehand ultrasound reconstruction based on ROI prior modeling and normalized convolution. In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2003*. Ed. by Ellis, RE and Peters, TM. Vol. 2879. Springer, 2003:382–390.
33. Barry, CD, Allott, CP, John, NW, Mellor, PM, Arundel, PA, Thomson, DS, and Waterton, JC. Three-dimensional freehand ultrasound: image reconstruction and volume analysis. *Ultrasound in Medicine & Biology* 1997;23:1209–1224.
34. Gobbi, DG and Peters, TM. Interactive intra-operative 3D ultrasound reconstruction and visualization. In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002*. Ed. by Dohi, T and Kikinis, R. Vol. 2489. Lecture Notes in Computer Science. Springer, 2002:156–163.
35. Rohling, R, Gee, A, and Berman, L. A comparison of freehand three-dimensional ultrasound reconstruction techniques. *Medical Image Analysis* 1999;3:339–359.
36. Sanches, JM and Marques, JS. A multiscale algorithm for three-dimensional free-hand ultrasound. *Ultrasound in Medicine & Biology* 2002;28:1029–1040.
37. Nikolov, SI, González, JPG, and Jensen, JA. Real time 3D visualization of ultrasonic data using a standard PC. *Ultrasonics* 2003;41:421–426.
38. Eulenstein, S, Lange, T, Hünerbein, M, Schlag, PM, and Lamecker, H. Ultrasound-based navigation system incorporating preoperative planning for liver surgery. *International Congress Series* 2004;1268:758–763.
39. Huang, Sw, Kim, K, Witte, RS, Olafsson, R, and O’Donnell, M. Inducing and Imaging Thermal Strain Using a Single Ultrasound Linear Array. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 2007;54:1718–1720.
40. Cinquin, P, Bainville, E, Barbe, C, et al. Computer assisted medical interventions. *IEEE Engineering in Medicine and Biology Magazine* 1995;14:254–263.
41. Cleary, K, Ibáñez, L, Gobbi, D, Cheng, P, Gary, K, Aylward, S, Jomier, J, Enquobahrie, A, Zhang, H, Kim, Hs, and Blake, MB. *IGSTK: The Book*. 2.0. Gaithersburg, Maryland, USA: Signature Book Printing, 2007.

42. Enquobahrie, A, Cheng, P, Gary, K, Ibáñez, L, Gobbi, D, Lindseth, F, Yaniv, Z, Aylward, S, Jomier, J, and Cleary, K. The Image- Guided Surgery Toolkit IGSTK: An Open Source C++ Software Toolkit. *Journal of Digital Imaging* 2007;20:21–33.
43. Gary, K, Ibáñez, L, Aylward, S, Gobbi, D, Blake, MB, and Cleary, K. IGSTK: An open source software platform for image-guided surgery. *Computer* 2006;39:46–53.
44. Treece, GM, Gee, AH, Prager, RW, Cash, CJC, and Berman, LH. High-definition freehand 3-D ultrasound. *Ultrasound in Medicine & Biology* 2003;29:529–546.
45. Gobbi, DG, Lee, BKH, and Peters, TM. Correlation of preoperative MRI and intraoperative 3D ultrasound to measure brain tissue shift. In: *Medical Imaging 2001: Visualization, Display, and Image-Guided Procedures*. Ed. by Mun, SK. Vol. 4319. SPIE, 2001:264–271.
46. Angelsen, BAJ. *Ultrasound Imaging. Waves, signals and signal processing*. Trondheim, Norway: Emantec AS, 2000.
47. Wear, KA. The effects of frequency-dependent attenuation and dispersion on sound speed measurements: applications in human trabecular bone. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 2000;47:265–273.
48. Coupé, P, Hellier, P, Azzabou, N, and Barillot, C. 3D freehand ultrasound reconstruction based on probe trajectory. In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2005*. Ed. by Duncan, J and Gerig, G. Vol. 3749. Springer, 2005:597–604.
49. Sanches, JM and Marques, JS. A Rayleigh reconstruction/interpolation algorithm for 3D ultrasound. *Pattern Recognition Letters* 2000;21:917–926.
50. Thijssen, JM, Weijers, G, and Korte, CL de. Objective Performance Testing and Quality Assurance of Medical Ultrasound Equipment. *Ultrasound in Medicine & Biology* 2007;33:460–471.
51. Lindseth, F, Langø, T, Bang, J, and Hernes, TAN. Accuracy evaluation of a 3D ultrasound-based neuronavigation system. *Computer aided surgery* 2002;7:197–222.
52. Hastenteufel, M, Vetter, M, Meinzer, HP, and Wolf, I. Effect of 3D ultrasound probes on the accuracy of electromagnetic tracking systems. *Ultrasound in Medicine & Biology* 2006;32:1359–1368.
53. Rousseau, F, Hellier, P, and Barillot, C. A novel temporal calibration method for 3-D ultrasound. *IEEE Transactions on Medical Imaging* 2006;25:1108–1112.

54. Lemke, HU. Summary of the white paper of DICOM WG24 'DICOM in Surgery'. In: *Medical Imaging 2007: PACS and Imaging Informatics*. Ed. by Horii, SC and Andriole, KP. Vol. 6516. SPIE, 2007:651603–651613.