Automated analysis of 3D echocardiography

Stralen, M. van

Interpolation of irregularly distributed sparse 4D ultrasound data using normalized convolution

4

WE DEVELOPED A NOVEL MULTI-BEAT image fusion technique using a special spatiotemporal interpolation for sparse, irregularly sampled data (ISIS). It is applied to irregularly distributed 3D cardiac ultrasound data acquired with a fast rotating ultrasound (FRU) transducer. ISIS is based on normalized convolution with Gaussian kernels tuned to irregular beam data spacing over cardiac phase (τ), and beam rotation (θ) and elevation angles (ϕ).

Images are acquired with the FRU transducer, a linear phased array rotating mechanically continuously at very high speed (240-480rpm). High-quality 2D images are acquired at ±100 frames/s over 5-10 seconds. ECG is recorded simultaneously. Images are irregularly distributed over τ and θ, because rotation is not synchronized to heart rate. ISIS was compared quantitatively to spatiotemporal nearest neighbor interpolation (NNI) on synthetic (distance function) data of a pulsating ellipsoid.ISIS was also tested qualitatively on 20 in-vivo cardiac image sets and compared to classical temporal binning with trilinear voxel interpolation, at resolutions of $256 \times 256 \times 400$ for 16 phases.

From the synthetic data simulations, ISIS showed absolute distance errors (mean ± standard deviation) of 1.23 ± 1.52 mm; considerably lower than for NNI (3.45 ± 3.03 mm). For in-vivo images, ISIS voxel sets showed reduced motion artifacts, suppression of noise and interpolation artifacts and better delineation of endocardium. ISIS improves the quality of 3D+T images acquired with a FRU transducer in simulated and in-vivo data.

This chapter has been derived from (© 2006 SPIE):
Novel spatiotemporal voxel interpolation with multibeat fusion for 3D echocardiography with irregular data distribution. J.G. Bosch, M. van Stralen, M.M. Voormolen, B.J. Krenning, C.T. Lancé, J.H.C. Reiber, A.F.W. van der Steen, N. de Jong. Proc SPIE Med Imaging 2006; 6147; 61470Q.

4.1 | Introduction

3D ultrasound techniques

3D ultrasound imaging is rapidly gaining clinical importance. For acquisition of 3D images, several techniques are possible: freehand probe manipulation with position sensing; motorized probe translation/rotation; or matrix array transducers. For acquisition of 3D sets over the full cardiac cycle (4D echocardiography), this is generally combined with ECG triggering. The speed and quality of a freehand or stepper motor approach are insufficient for such applications. As an alternative to the costly and complex matrix array solution, we have developed a linear phased array transducer that rotates continuously at high speed. Such a fast rotating linear phased array can be used to acquire images of the left ventricle covering 3D space over the full cardiac cycle in near real time. Image quality within the 2D image planes is superior to matrix array transducers, but these planes are spread sparsely and irregularly over space and time, since rotation and cardiac phase are not synchronized. We developed a novel spatiotemporal interpolation technique to tackle the complex problem of generation of optimal voxel sets and standardized cross sections from this data.

the challenge

4.1.1 | Fast rotating ultrasound transducer

transducer design

The fast rotating ultrasound (FRU) transducer developed in our laboratory [Djoa et al. 2000; Voormolen et al. 2002; Voormolen et al. 2003] is a special device for near real-time 3D cardiac image acquisition (fig. 4.1). It consists of three major parts: a phased array, a DC motor that drives the array, and a slip-ring device establishing signal transfer to and from the continuously rotating array. The DC motor drives the array at a high rotation speed ranging from 4 to 8 Hz and is connected to an external control system with a manual setting for the rotation speed. The array of the transducer, custom made by Delft Instruments (Delft, The Netherlands), contains 64 elements with a pitch of 0.21 mm and is tapered into an octagonal shape, approximating a circle with a radius of 7 mm. It has a fractional bandwidth of 86% with a center frequency of 3.2 MHz.

4.1.2 | Image acquisition

Cardiac apical images are acquired at about 100 2D frames per second (fps) with the FRU transducer rotating at 4 – 8 Hz (240 – 480 rpm) and recorded with a GE/VingMed Vivid 5 ultrasound system (VingMed, Horten, Norway). The typical acquisition

Figure 4.1: The fast rotating ultrasound (FRU) transducer

time is approximately 10 seconds, which has proven to be convenient for clinical application and allows a complete acquisition within one breath hold.

2D images are acquired and stored digitally in polar format (ultrasound beams). Each image consists of approximately 80-90 beams at about 1° spacing; each beam consists of 400-500 samples. The patient's ECG is recorded simultaneously. Due to the fast rotation, individual 2D images have a curved 3D shape (fig. 4.2) because the image plane rotates over 10° – 30° during the collection of the consecutive beams of the image.

curved image planes

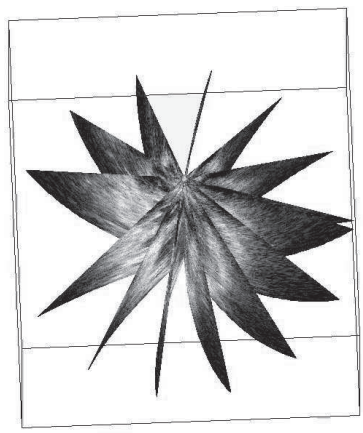

Figure 4.2: A sequence of seven consecutive FRU images with curved image planes

Because of variable cardiac cycle duration and lack of synchronization between rotation and cardiac phase, images are irregularly distributed over rotation angle and cardiac phase. After acquisition, exact rotation angle (θ) and cardiac phase (τ, time offset from R-peak) is calculated for each 2D image frame and its beams. From this information, multi-beat fusion can be applied: around 10 consecutive cardiac cycles are merged into one to reach a denser coverage of the 4D space.

4.1.3 | Sparse data and multi-beat fusion

In fig. 4.3a, an example of the coverage of 3D space is shown for an interval of about $1/12^{\text{th}}$ of a cardiac cycle. It can be seen that in that period, the volume is swept almost two times. At 6 rotations per second and 100 fps, a full volume is swept in about 80 ms and contains about 8 images (16 half sectors). The 3D distances between images amount to 10 – 20 mm, which is too large for appropriate voxel space generation; therefore, corresponding cardiac phases from several cardiac cycles are merged. This multi-beat fusion results in a denser coverage of 3D space per cardiac interval (fig. 4.3b).

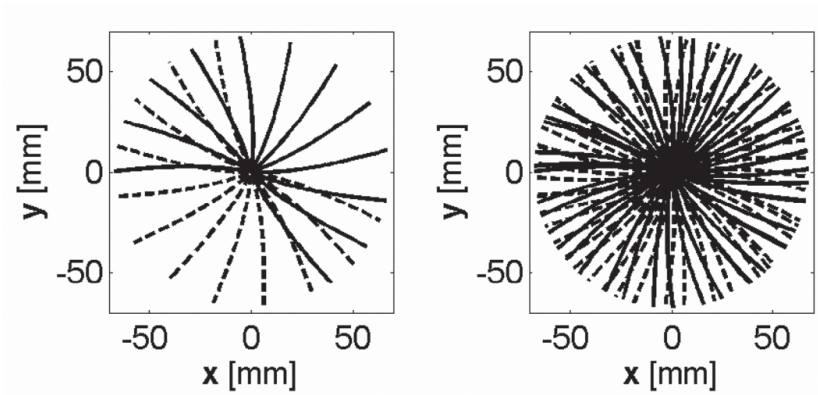

Figure 4.3: *left*) Example of the sparse sampling of the time-volume space resulting from one time interval (1/12) of a cardiac cycle. The individual recorded images are depicted as lines at a certain C-plane level (the z-axis is the transducer's axial direction). The left half of each individual sector image is shown as a solid line and the right half as a dashed line. *right*) The denser sampling of the volume-time space when multi-beat fusion is used

Irregular data distribution | 4.1.4

The distribution of the curved image planes over the 4D data space (3D plus time) is determined by three factors: the rotation speed, the frame rate and the duration of each cardiac cycle. The frame rate is determined by the settings of the ultrasound machine, remains fixed over an acquisition and is exactly known. Rotation speed is set at the motor controller of the FRU transducer and measured precisely after the acquisition by a special image analysis procedure [Voormolen et al. 2003]. The beat lengths, however, are variable and unpredictable, since they are determined by the patient's spontaneous sinus rhythm, which exhibits a natural short-term variability. Therefore, the start of the next beat (marked by the R-peak in the ECG) comes at an unpredictable moment, and thus at a highly unpredictable rotation angle. Consecutive beats may easily vary in beat length by 50 ms, which corresponds to a rotation angle difference of about 100°. An actual example of the irregular distribution of images over rotation angle θ and cardiac phase τ is given in fig. 4.4.

unpredictable beat length

Interpolation of sparse, irregular samples | 4.1.5

Interpolating data from sparse and irregularly distributed samples is a classical problem. In the image processing domain, it has been addressed for the filling of missing parts in images with "plausible" data (image inpainting) [Drori et al. 2003; Pham and van Vliet 2003], resolution improvement by combining multiple low-resolution images into one high-resolution images (superresolution) [Farsiu et al. 2004; Pham and van Vliet 2003] and converting irregularly sampled data into regular images [Loke and du Buf 2004]. In the (ultra)sonic imaging domain, such issues have been addressed for sonar imaging [Loke and du Buf 2004] and interpolation of freehand 3D ultrasound [Estépar et al. 2003; Meairs et al. 2000]. For inpainting and superresolution applications, the so-called normalized convolution [Knutsson and Westin 1993; Westin 1994] approach (also known as normalized averaging) has been applied with success [Drori et al. 2003; Pham and van Vliet 2003].

normalized convolution

The technique has been applied to freehand 3D ultrasound images of cerebrovascular structures [Meairs et al. 2000] and kidney and thyroid [Estépar et al. 2003] as well, but these images were relatively densely sampled and there was no temporal interpolation involved. In our case, we deal with three spatial dimensions and one temporal dimension. The default technique in this 4D domain is temporal binning (assigning all data to some discrete time points) in combination with nearest neighbor interpolation or trilinear spatial averaging. To our knowledge, normalized convolution has not been applied to this 4D problem domain.

application to 4D domain

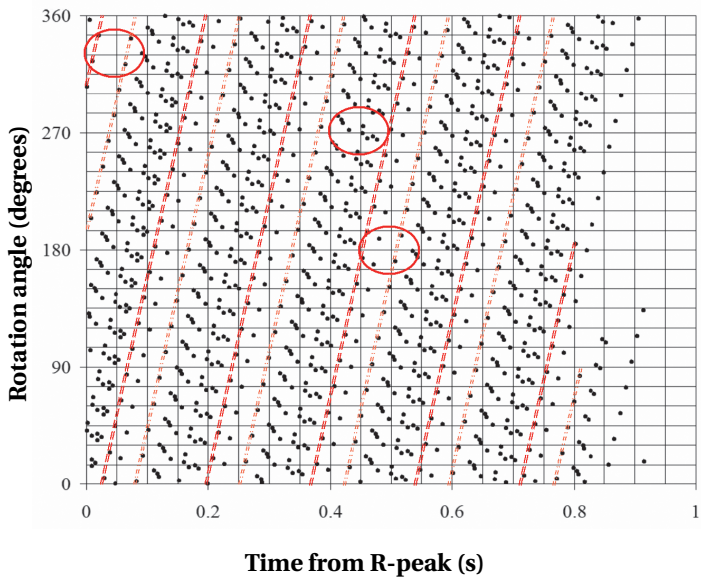

Figure 4.4: Distribution of 1012 consecutive image frames (14 beats) over rotation angle and cardiac phase. Beat durations from 0.8 to 0.92 s, rotation speed 5.82 Hz, frame rate 106.1 Hz. Each dot represents one frame. The redlines highlight the consecutive frames from the first beat (from left to right); the light redline marks those of the second beat. The redcircles show areas of dense and sparse distribution

4.2 | Methods

4.2.1 | Introduction

We developed a novel multi-beat fusion technique using a special spatiotemporal interpolation for sparse, irregularly sampled data (ISIS). The technique is based on normalized convolution with Gaussian kernels tuned to the (irregular) spacing of the data over cardiac phase (τ) and over the rotation (θ) and elevation angle (ϕ) of the spherical beam coordinates. Normalized convolution is a technique known from superresolution approaches and irregular data interpolation, but has not been applied to this area. The interpolation of the sparse irregular beam data takes place in the (θ,ϕ,τ) domain; results are then converted into Cartesian (x, y, z) voxel space for each τ.

Image fusion/interpolation | 4.2.2

For constructing a 4D image set (a time sequence of 3D voxel sets), this irregularly and locally sparsely populated dataset must be interpolated into a regular, fully populated set. In fig. 4.4, this corresponds to estimating the data at each of the grid line crossings (or even on a finer grid). Various techniques can be applied to fulfill this task (fig. 4.5).

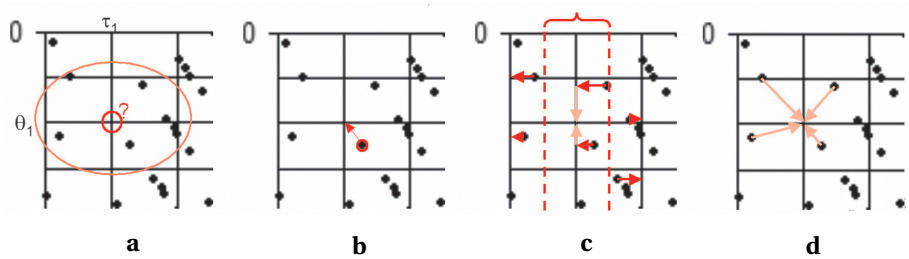

Figure 4.5: Different approaches to sparse data interpolation, demonstrated on a detail of fig. 4.4. *a*) grid position to be filled. *b*) Spatiotemporal nearest neighbor interpolation. *c*) temporal binning and spatial linear interpolation. *d*) spatiotemporal weighted average

The simplest approach is to find the closest sample for each grid point. This means we bluntly neglect the error in time and angle, which can result in considerable motion and fusion artifacts. Moreover, many of the available samples are simply neglected, since only one sample per grid point is used for the estimation. Furthermore, several grid points may use the same single sample. In terms of interpolation, this is a spatiotemporal "nearest neighbor" approach.

nearest neighbor

The interpolation in our case involves both temporal and spatial differences. A sensible classical solution is to apply temporal binning followed by interpolation. All samples within a time interval are assigned to the midpoint of the interval, but their angular information is preserved. Interpolation is then applied between consecutive angles. Although no samples are wasted and angular information is preserved, this may still result in considerable motion artifacts, since information from different time points is combined without taking object motion into account.

temporal binning

An alternative solution for spatiotemporal interpolation is to use a weighted combination of all neighboring samples in space and time. For sparse, irregularly distributed data, finding the neighboring samples for each grid point is not an easy task. Therefore, most approaches use the inverse approach and distribute the sample data to neighboring grid points.

our approach

We have developed a novel technique using a dedicated spatiotemporal interpolation for sparse, irregularly sampled data (ISIS). The technique is based on normalized convolution (NC) [Knutsson and Westin 1993; Pham and van Vliet 2003; Westin 1994]. Normalized convolution applies the signal/certainty principle: a sparse data sampling can be represented as a combination of a signal image S, containing the available samples, and a certainty image C, containing value '1' for the sample positions and value '0' elsewhere.

The empty spaces can be filled by diffusing both signal strength and certainty around the available samples. By normalizing the accumulated samples with accumulated certainty, properly weighted values result. NC is a technique known from superresolution approaches [Knutsson and Westin 1993; Pham and van Vliet 2003] and irregular data interpolation [Knutsson and Westin 1993], but has not been applied to this area of spatiotemporal ultrasound interpolation.

2D example

In the example of fig. 4.6, we will use the distribution shown in fig. 4.4. Each sample point in this distribution represents one complete image (all values of r and ϕ). Therefore, with this example we can simulate the interpolation over the (θ,τ) domain which is sparsely and irregularly sampled. By applying the distribution of the 1012 samples on the 128×128 test image (fig. 4.6a), we keep only 6% of the original data (fig. 4.6b and c). By applying a convolution ($\otimes$) with some diffusion kernel G on both the signals (fig. 4.6d) and their certainty (fig. 4.6e) and calculating their ratio (eqn. 4.1, fig. 4.6f), we get the normalized convolution result R.

$$R = \frac{(S \cdot C) \otimes G}{C \otimes G} \tag{4.1}$$

In fig. 4.6, a 2D Gaussian kernel was applied with a width $\sigma = 3$ in both directions. As can be seen in fig. 4.6f, the structure of the image is well preserved, although only 6% of the original samples were used.

4.2.3 | Implementation in 4D

Instead of the 2D example described above, we are addressing a 4D problem: one continuous temporal dimension and three spatial dimensions. In our case, the goal of the interpolation is the generation of a time sequence of regular Cartesian (x, y, z) voxel sets. The input data, however, consists of discrete ultrasound beams; their native domain is the spherical coordinate system, where every beam is fully characterized by its rotation angle θ, elevation angle ϕ and cardiac phase τ. Each sample within the beam is addressed by its depth r (fig. 4.7a).

We propose to apply spatiotemporal interpolation between complete beams, so the variable r can be left out of the interpolation task. The interpolation of the

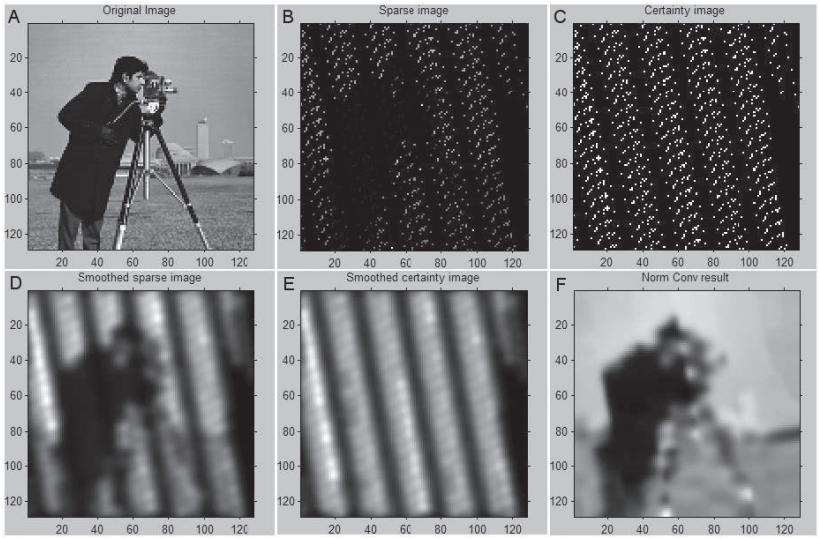

Figure 4.6: Example of normalized convolution. *a*) original image 128*128. *b*) sparse sampling S of image following distribution of fig. 4.4. *c*) Certainty image C. *d*) image *S* diffused by Gaussian kernel. *e*) image C diffused by Gaussian kernel. *f*) Result image, D normalized by E

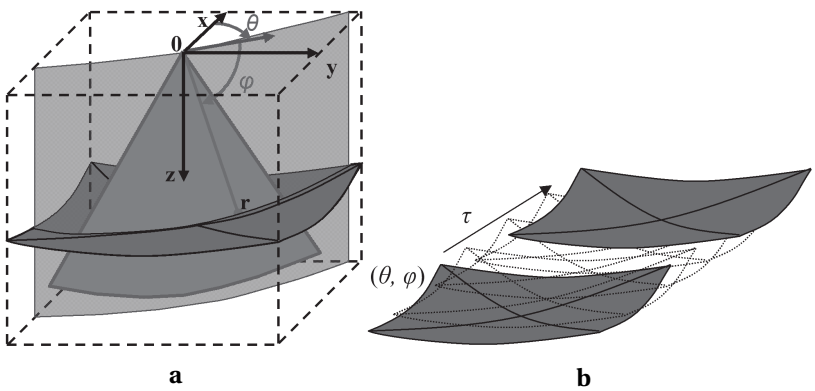

Figure 4.7: *a*) Spherical and Cartesian coordinate system. The z-axis corresponds to the rotation axis of the transducer. r : depth, ϕ : elevation angle, θ : rotation angle. A single curved image plane with a scan sector and a single beam is shown in gray. A spherical plane $r = C$ is shown in dark gray. *b*) Domain of ISIS interpolation: for planes of constant r, ISIS is applied over (θ, ϕ, τ) space

sparse beam data takes place in the (θ,ϕ,τ) domain (fig. 4.7b), by 3D normalized convolution using a 3D Gaussian kernel. This results in a dense, oversampled, regular (θ,ϕ,τ) space, or a regular, dense (θ,ϕ) plane for each desired value of τ. Results are then converted into Cartesian (x,y,z) voxel space for each τ by trilinear interpolation in a regular grid between the eight closest points in (θ,ϕ,r) for every voxel (fig. 4.8).

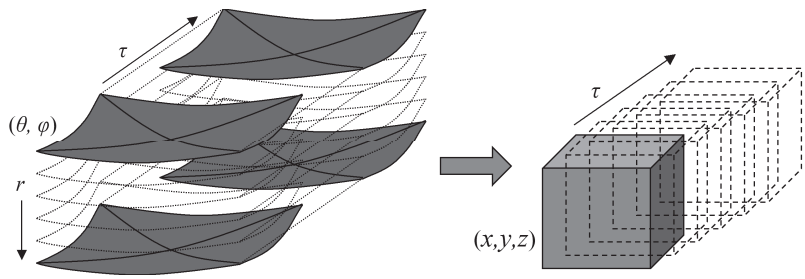

Figure 4.8: Conversion from spherical to Cartesian domain: ISIS interpolated sets for all r are combined and converted into regular voxel sets for each desired value of τ

4.3 | Experiments

ISIS was tested both quantitatively (by simulations) and qualitatively on in-vivo data. It was tested quantitatively by simulations on synthetic data representing the beating left ventricle and compared to spatiotemporal nearest neighbor interpolation (NNI). From the synthetic data, spatiotemporal sampling errors D of both methods were calculated by sampling a distance function from the beating ellipsoid's surface and subtracting the true (known) distance at the exact desired time and spatial position. ISIS voxel space generation was also tested qualitatively on 20 in-vivo cardiac image sets of 15 patients, compared to a classical temporal binning with spatial trilinear voxel interpolation.

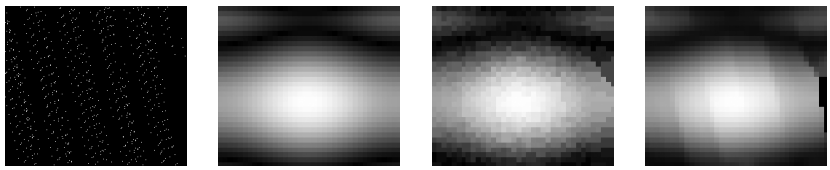

Figure 4.9: Results of simulations on a distance function for a pulsating ellipsoid over (θ, τ) space. *a*) Distribution of samples (beams) over phase τ and rotation angle θ (300×256). *b*) Distance function sampled at ideal positions for θ and τ (37×32). *c*) NNI results (37 × 32) from samples in *a*. *d*) ISIS results (37 × 32) from samples in *a*

Quantitative evaluation on synthetic data | 4.3.1

ISIS was tested quantitatively by simulations on synthetic data of a pulsating ellipsoid representing the beating left ventricle. It was compared to spatiotemporal nearest neighbor interpolation (NNI). For every point $\boldsymbol{p}$ in space at time t, an analytical distance function $D(\boldsymbol{p}, t)$ from the ellipsoid's surface can be calculated,

error measure

$$D(\boldsymbol{p}, t) = \frac{\left(\|\boldsymbol{p} - \boldsymbol{a}\| + \|\boldsymbol{p} - \boldsymbol{b}\|\right)}{2} - D_0(t) \tag{4.2}$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are the 3D positions of the ellipsoid's foci and $D_0(t)$ is the ellipsoid's long-axis radius as a function of time.

From the synthetic data, spatiotemporal sampling errors ΔD of both interpolation methods were calculated by sparsely sampling the distance function from the beating ellipsoid's surface (eqn. 4.2), interpolating the sparse distance data and comparing the interpolated distance values $D'_{NNI}(\boldsymbol{p}, t)$ resp. $D'_{ISIS}(\boldsymbol{p}, t)$ with the true (known) distance $D(\boldsymbol{p}, t)$ at all exact desired spatiotemporal positions $(\boldsymbol{p}, t)$,

$$\Delta D = D'(\boldsymbol{p}, t) - D(\boldsymbol{p}, t) \tag{4.3}$$

ΔD expresses both temporal and spatial errors in a distance in mm, which is meaningful since the beating ellipsoid's dimensions model the actual application.

Target resolutions for the rotational image sets were 32 angles (θ) and 37 phases (τ) within the cardiac cycle (which corresponds to resolutions of about 11°, 25 ms) (fig. 4.9). Object displacements per time step were in a similar range as per angle step. For simplicity, this evaluation was only done in two dimensions (one temporal, one spatial); the images can be interpreted as an M-mode of a set of 32 samples on a coaxial circle in a C-plane (horizontal; ϕ, r fixed; θ varying).

4.3.2 | Qualitative evaluation on in-vivo data

ISIS voxel space generation was also tested qualitatively on 20 in-vivo cardiac image sets of 15 patients, compared to a classical temporal binning with spatial trilinear voxel interpolation. Target spherical resolutions for normalized convolution interpolation over (θ, ϕ, τ) were $(256 \times 100 \times 64)$, in order to achieve (x, y, z) voxel sets of $(256 \times 256 \times 400)$ at 16 cardiac phases. Typical dimensions of the input data were 800 – 1100 polar frames containing 80 – 90 beams with 400 – 500 samples, covering a sector of about 90° and 10 cm depth. The applied 3D Gaussian diffusion kernel $G(\sigma_\theta, \sigma_\phi, \sigma_\tau)$ had widths of (1.0, 1.0, 3.0) respectively. A qualitative evaluation was performed by side-by-side comparison of equivalent cross sections (fig. 4.10) of voxel sets generated by both techniques.

parameter settings

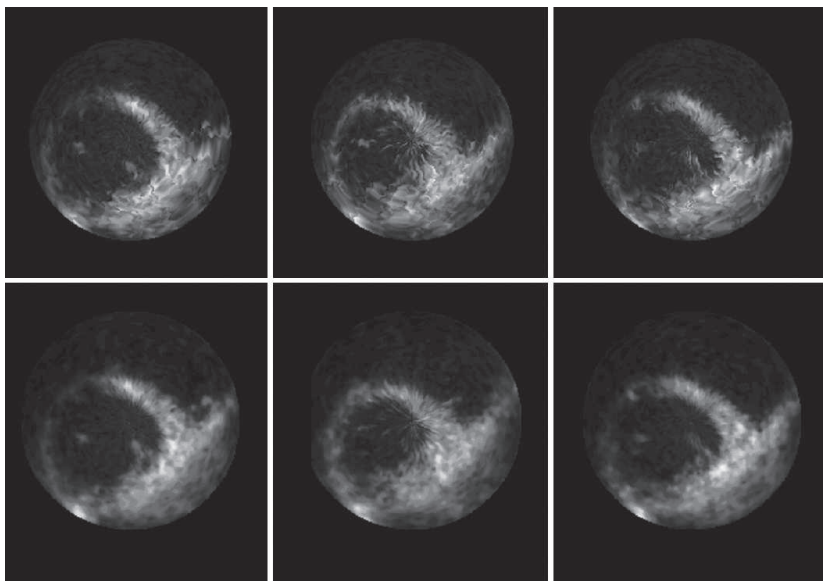

Figure 4.10: *top*) Reconstructed planes using classical trilinear interpolation in 3D space, assuming constant cardiac phase within each time bin. *bottom*) same planes using ISIS. Depth ± 10 cm, just above mitral valve at (left to right) phase ED, ES and iso-volumetric contraction. ISIS gives higher tissue-to-blood ratio and a decrease in motion artifacts

Table 4.1: Results of interpolations on synthetic data of pulsating ellipsoid

$n = 17$	**Distance errors** (mm)	**Absolute distance errors** (mm)
NNI	0.31 ± 4.58	3.45 ± 3.03
ISIS	0.16 ± 1.95	1.23 ± 1.52

All the results are expressed as mean ± standard deviation.

Results | 4.4

Simulations on synthetic data | 4.4.1

In the simulations, nearest neighbor interpolation (NNI) showed temporal errors of -1.1 ± 13.8 ms (mean ± standard deviation) and angular errors of -0.2 ± 4.8 degrees. Mean and SD of absolute errors were 10.2 ± 9.4 ms and 3.8 ± 3.0 degrees, respectively. These values are directly calculated from the nearest neighbor sparse sample positions. From the synthetic data simulations, ISIS showed distance errors ΔD_{ISIS} of 0.16 ± 1.95 mm; this was considerably lower than for NNI ($\Delta D_{NNI} = 0.31 \pm 4.58$ mm). Absolute distance errors were 1.23 ± 1.52 mm for ISIS vs. 3.45 ± 3.03 mm for NNI. Results are shown in fig. 4.9 and listed in table 4.1.

In-vivo images | 4.4.2

In the in-vivo images, the resulting voxel sets showed reduced motion artifacts, suppression of noise and interpolation artifacts and better delineation of structures and tissue-to-blood contrast (fig. 4.10). Especially the typical zigzag artifacts (resulting from interpolations between frames that are spatially very close but temporally different) are strongly suppressed. However, the choice of the kernel size parameters seemed suboptimal in some cases. In areas of very sparse sampling, some artifacts remain visible. Sometimes, interpolation seems irregular in these areas and spurious edges may result. In very densely sampled areas, sometimes some anatomical details are smoothed out.

4.5 | Discussion

As the most important result from this study, we can conclude that the ISIS method gives better interpolation results than a classical approach using temporal binning. However, several limitations apply and we assume further improvements are possible; these are discussed below.

4.5.1 | Limitations

The setup of this study was limited in several aspects.

simulation only in 2D

First of all, the simulations have so far been applied only in the 2D case. Although we have chosen this simulation in a way that matches the real problem well, we would like to extend it to the actual 4D situation .

qualitative in-vivo evaluation

Secondly, the in-vivo interpolation results have so far only been judged qualitatively by visual comparison to other interpolations. A more elaborate evaluation needs to be performed, e.g. on a moving phantom, to objectively investigate artifacts and to examine the effects of the method on volume estimations after segmentation.

4.5.2 | Alternatives and further development

There are several issues that should be further investigated.

size of the diffusion kernel

As described in section 4.4.2, the sizes of the Gaussian diffusion kernels seem suboptimal in some situations. The size of the diffusion kernel should match the local sample distribution; if the kernel size is much smaller than the spatiotemporal sample distances, the diffusion process can result in an irregular interpolation, possibly introducing spurious edges. If the kernel size is much larger than these local sample distances, data from many points will contribute to fill the holes and results will look overly smoothed. This can be understood if we consider the weighted sum of two overlapping Gaussian kernels at different distances. Where sample points are relatively close, the Gaussian kernels will result in an equal weighting and therefore in an unweighted averaging; at the tails of the distributions, the nearest sample will fully dominate the weighting and the result will behave like a nearest neighbor interpolation.

locally adaptive kernel

This issue can be addressed by applying alternative convolution kernels, such as proposed by Knutsson and Westin [1993] and Westin [1994], or by locally adjusting the kernel sizes to the distribution densities, e.g. by using a distance transform to

determine the widths of the anisotropic diffusion kernels. Preliminary efforts in this direction showed promising results.

The ISIS method fulfills its tasks in the sense that it supplies plausible spatiotemporal interpolation of sparse, irregular data. It effectively reduces noise and suppresses motion artifacts. However, a technique like this will not always preserve sharp edges. Especially in regions of sparse sampling and/or considerable motion, edges will inevitably be smoothed. Spatiotemporal interpolation of a moving, sharp edge will result in a smoothed or double edge. This could be handled with specific assumptions about objects and edges and a certain motion model, but this is subject of further research. Also, registration could be combined with the interpolation (e.g. as proposed by Penney et al. [2004]) to address this limitation. However, such an approach is iterative, computation intensive and known to be extremely slow. Therefore, this will result in considerable further increase in complexity and processing time.

Conclusions | 4.6

The proposed interpolation technique (ISIS) improves the quality of 3D + T images acquired with a fast rotating transducer in simulated and in-vivo data. Several further improvements seem feasible. The technique may also prove useful in similar cases where multi-beat fusion, irregular sampling and/or phasic motion artifacts induce image errors, e.g. freehand 3D ultrasound.