Taylor & Francis
Taylor & Francis Group

Spatial and frequency-based super-resolution of ultrasound images

Mon-Ju Wu[a], Joseph Karls[b], Sarah Duenwald-Kuehl[c,d], Ray Vanderby Jr.[a,c,d] and William Sethares[b]*

[a]Materials Science Program, University of Wisconsin-Madison, 2556 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706, USA; [b]Department of Electrical and Computer Engineering, University of Wisconsin-Madison, 2556 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706, USA; [c]Department of Orthopedics and Rehabilitation, University of Wisconsin-Madison, 2556 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706, USA; [d]Department of Biomedical Engineering, University of Wisconsin-Madison, 2556 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706, USA

Modern ultrasound systems can output video images containing more spatial and temporal information than still images. Super-resolution techniques can exploit additional information but face two challenges: image registration and complex motion. In addition, information from multiple available frequencies is unexploited. Herein, we utilised these information sources to create better ultrasound images and videos, extending existing technologies for image capture. Spatial and frequency-based super-resolution processing using multiple motion estimation and frequency combination was applied to ultrasound videos of deforming models. Processed images are larger, have greater clarity and detail, and less variability in intensity between frames. Significantly, strain measurements are more accurate and precise than those from raw videos, and have a higher contrast ratio between 'tumour' and 'surrounding tissue' in a phantom model. We attribute improvements to reduced noise and increased resolution in processed images. Our methods can significantly improve quantitative and qualitative assessments of ultrasound images when compared assessments of standard images.

Keywords: super-resolution; image processing; ultrasound

1. Introduction

Ultrasound images provide a crucial diagnostic tool in both clinical and scientific applications. While individual still images can be viewed and interpreted by trained personnel, modern imaging systems are capable of outputting ultrasound video. While it is evident that a video contains more temporal information than a still image (since it contains information about motion over time), video also contains more spatial information than is immediately apparent. It is not practical to manually examine the information in every frame of the video, so automated methods are becoming crucial. The motivation for this study was to develop a method to create better ultrasound images by exploiting a technique called super-resolution, which combines a collection of small images (number of pixels) into one larger (and more detailed) image.

Super-resolution (Blomgren et al. 2002; Freeman et al. 2002; Park et al. 2003; Farsiu et al. 2004; Shechtman et al. 2005) is perhaps best known as a way of merging multiple camera shots into a larger panorama, but can also be applied when combining multiple (near-identical) video frames into larger pictures. Super-resolution consists of image registration and interpolation. Image registration defines the different locations of the same objects (points) in different frames. It is responsible for aligning the images and retrieving their relative locations. The goal of this study was to create larger, more detailed images from a series of consecutive ultrasound video frames by mimicking this technique. For example, four images that are each 200 by 200 pixels may be combined into one larger image that is 400 by 400 pixels. Applying this procedure to frames {1, 2, 3, 4} of a video, then repeating for frames {2, 3, 4, 5}, then for frames {3, 4, 5, 6} and so on leads to a video that is twice as large (in terms of pixel size) as the original. When the procedure works well, details that were difficult to see in the original, smaller, video become easier to see in the processed, enlarged, video. Image registration is crucial to successful application of super-resolution. Various image registration algorithms may be used in the super-resolution process. Some apply statistical analyses (Elad & Feuer 1997; Borman & Stevenson 1998; Patanavijit & Jitapunkul 2006; Humblot & Mohammad-Djafari 2006; Ho & Zeng 2007; Martins et al. 2007), others use back projection (Haralick et al. 1973; Capel & Zisserman 2000; Ji & Fermuller 2009), and yet others apply frequency domain information (Balci & Foroosh 2006; Vandewalle et al. 2006) for the image registration step. Many image registration algorithms, such as scale invariant feature transform (Lowe 2004) and speeded-up robust features (Bay et al. 2008), perform registration based on certain key points that are usually bright spots, sharp edges and/or well-defined corners. Because ultrasound images do not have sharp

*Corresponding author. Email: sethares@ece.wisc.edu

M.-J. Wu et al.

edges and well-defined corners, keypoint methods do not return useful alignment information.

The second technical challenge was the character of the motion in the ultrasound images. There are many ways of conducting the needed steps within the image registration. Because of the large number of different motions that may occur in the ultrasound images, global methods are inapplicable (Zitová & Flusser 2003). While the individual photographs that are to be combined into a panorama may differ in angle or in spatial placement, the overall mapping that needs to be identified is a single rigid perspective transformation; hence the complete alignment can be accomplished by identifying a single affine mapping. With common ultrasound images (such as tendon images, arterial images and echocardiograms), the range of possible motions is much larger, with some portions of the image moving in one direction and other portions moving in other directions; accordingly, the alignment is needed to consider a broader range of possible mappings.

The ability to gather information from the multiple excitation frequencies available in ultrasound imaging (this might be analogous in the camera application to photos in visible light, infrared and ultraviolet) is a previously unexploited possibility. Different frequencies may lead to different image qualities. Lower frequencies travel further but carry less detail. Higher frequencies do not travel as deeply into the tissue before they damp out, but they contain more texture detail. People have attempted to quantify the attenuation and backscatter of ultrasound in human tissues (D'Astous & Foster 1986; Miller et al. 2008; Sun et al. 2012) and examine the difference of human tissues captured at different frequencies. Fort et al. (1995) proposed to determine texture information by estimating the texture information by singular value decomposition in Fourier domain. Liu et al. (2006) enhanced ultrasound-induced heating by generating the thermal lesions with high frequency and enhancing the bubble activity with low frequency. Ikeda et al. (2006) proposed to use two different frequencies to control cloud cavitation, which refers to structures caused by the collapse or condensation of vapour bubbles formed in a liquid in the body. It is therefore possible to combine

images over a range of frequencies to achieve higher resolution and more accurate imaging. Reflected ultrasound waves at a given frequency amplify and interfere with each other in certain locations, creating a grainy image; these patterns are different for different frequencies. Combining frequencies then offers the possibility of filling in information from the interference gaps for image enhancement. The incorporation of combining these multiple frequency images into the super-resolution scheme presents a third technical challenge.

In order to apply a modified super-resolution technique which successfully resolves these three challenges, images (individual frames in the ultrasound videos) were subjected to (1a) a super-resolution processing using (2a) multiple motion estimation, followed by (3a) combining the multiple frequency images. This method results in a processed video called SuperResolutionVideo 1. Alternatively, it is possible to reverse the order to first (1b) combine the different frequency images, then (2b) super-resolve the results using (3b) the multiple motion estimation; this results in a processed video called SuperResolutionVideo 2. Both SuperResolution videos exploit spatial and frequency diversity to create higher resolution (and more accurate) videos.

2. Materials and methods

2.1 Spatial super-resolution of ultrasound

The typical approach to super-resolution of images is to combine the information from a set of slightly different (i.e. differing in angle or spatial displacement) low-resolution (LR) images of the same scene and use image registration to construct the set into a higher resolution image. In the present case, the set of LR images are consecutive frames in a dynamic ultrasound video. Due to the nature of ultrasound images, it is highly unlikely that complete alignment of the LR images can be achieved by a single affine transformation. To account for this, the images were partitioned into a collection of 36 sub-images, and motion was estimated in each partition separately. The partitioning of the image is shown in Figure 1.
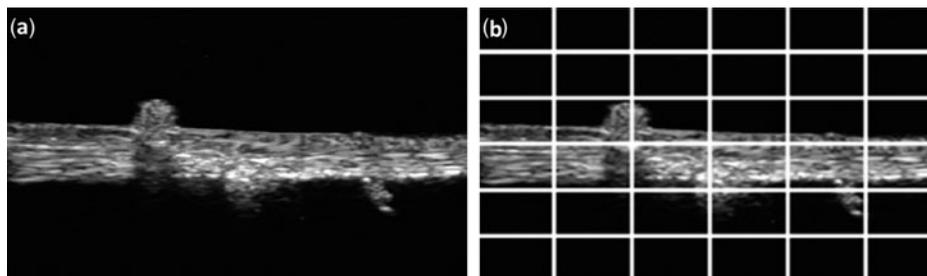

Figure 1. Partitioning of ultrasound images. (a) One frame from an ultrasound video of a small piece of tendon with a hard-edged bead. (b) Ultrasound video frame partitioned into 36 equally sized sub-images.

Each sub-image of every LR image is called $f_{\text{LR},p,m}$ (where $p = 1, \ldots, 36$ partitions and $m = 1, \ldots, M$, where M is the total number of LR images, which will be combined four at a time). These partitions were multiplied by a Tukey window to make them circularly symmetric so that they could be more accurately registered. The windowed sub-images are called $f_{\text{LR},\text{w},p,m}$. The Fourier transform $F_{\text{LR},\text{w},p,m}$ was then computed for each windowed sub-image.

Sub-images from the corresponding partition of consecutive LR images, $f_{\text{LR},\text{w},p,m+1}$, $f_{\text{LR},\text{w},p,m+2}$, $f_{\text{LR},\text{w},p,m+3}$ ($m = 1, \ldots, M-3$ and $p = 1, \ldots, 36$), were used to estimate the rotation of the sub-images with respect to the reference sub-image $f_{\text{LR},\text{w},p,m}$, and the polar coordinates (r, θ) for pixels in each of the sub-images were calculated, with the origin occurring at the centre of each sub-image. To estimate the rotation between sub-images $f_{\text{LR},\text{w},p,m}$ and $f_{\text{LR},\text{w},p,n}$ ($n = m + 1, m + 2, m + 3$) for every angle α, we computed the average value $h_{p,n}(\alpha)$ of the Fourier coefficients for which

$$\alpha - 1 < \theta < \alpha + 1, \quad (1)$$

and

$$0.1\rho < r < \rho_{\max}, \quad (2)$$

where ρ is the image radius (or half the size of the circularly symmetric sub-image). The angle where the maximum of the correlation between $h_{p,m}(\alpha)$ and $h_{p,n}(\alpha)$ is the estimated rotation angle φ_n. To cancel the rotation, the sub-image $f_{\text{LR},\text{w},p,n}$ was rotated $-\varphi_n$ degrees.

The horizontal and vertical shifts between the sub-images $f_{\text{LR},\text{w},p,m+1}$, $f_{\text{LR},\text{w},p,m+2}$, $f_{\text{LR},\text{w},p,m+3}$ and the reference sub-image $f_{\text{LR},\text{w},p,m}$ were estimated. A translation of the sub-image in the space domain can be expressed in the Fourier domain as a linear phase shift. It is well known that the shift parameters can be computed as the slope of the phase difference $\angle(F_{\text{LR},\text{w},p,n}/F_{\text{LR},\text{w},p,m})$. To reduce the sensitivity to noise, a plane was fitted through the phase differences using a least squares method. The shift parameters were the least squares solutions to the linear equations describing a plane through the computed phase differences [following the method used by Vandewalle et al. (2006)].

A high-resolution image was reconstructed using normalised convolution (Pham et al. 2006) from the four registered sub-images. The reconstructed sub-image had twice as many pixels in both dimensions as the LR sub-image. For every sub-image $f_{\text{LR},p,n}$, we computed the coordinates of its pixels in the coordinate frame of $f_{\text{LR},p,m}$ using the estimated rotation angles and estimated shifts. From this, we interpolated the values on a regular high-resolution grid using bicubic interpolation (Keys 1981). Each of these high-resolution sub-images was then pasted back together to form a complete high-resolution ultrasound image.

2.2 Frequency-based super-resolution

The texture of ultrasound images depends on the frequency of the ultrasound; changing frequencies leads to altered image qualities. Ultrasound waves at lower frequencies travel further into the tissue but carry less detail, whereas higher frequencies cannot travel as deeply into the tissue before significant damping, as the shorter wavelengths at higher frequencies cause the energy in the ultrasound waves to dissipate more quickly as they traverse the tissue, but they can carry more detailed texture information (Cross & Jain 1983). Figure 2 demonstrates this concept using two ultrasound images of the same construct (Figure 2(a)), one captured at a frequency of 8 MHz (Figure 2(b)) and one at a frequency of 13 MHz (Figure 2(c)). The 13 MHz image delivers more detailed texture on the 'teeth', whereas the 8 MHz image is brighter but more blurred (note especially the gaps between the teeth). The 13 MHz image attenuates faster than the 8 MHz image, and therefore is darker overall.

As ultrasound waves cannot carry information about the details of an object smaller than the wavelength, shorter wavelengths can carry more information and tend to create a sharper image than longer wavelengths. The relationship between frequency and wavelength is

$$v = f\lambda, \quad (3)$$

where λ is the wavelength, f is the frequency of the wave and v is the speed of the wave.

Four frequencies were used in this study: 8, 10, 12 and 13 MHz. The speed of ultrasound wave is about 1540 m/s
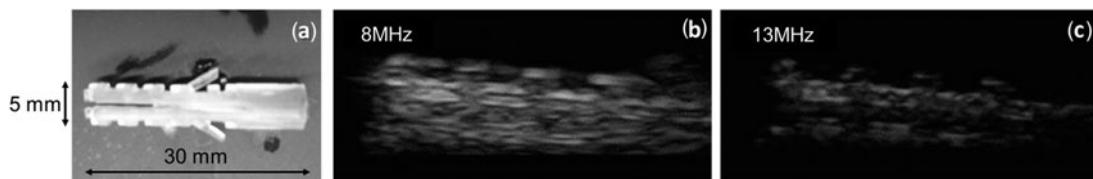

Figure 2. Digital (a) and ultrasound images (b and c) of a plastic device to demonstrate the difference in intensity and detailed texture of a solid object captured at different frequencies; (b) ultrasound image captured at 8 MHz and (c) ultrasound image captured at 13 MHz. The higher frequency image displays more detail but less brightness, while the lower frequency image is brighter but blurred.

4   *M.-J. Wu* et al.

in soft tissue, so the wavelength of frequency 8 and 13 MHz is 0.1925 and 0.1185 mm, respectively. Thus, the wavelength of the 8 MHz signal is close to twice the wavelength of the 13 MHz signal, so the 13 MHz images can carry more detailed information than frequency 8 MHz. This is perceived as a sharper image.

Since different frequencies contain different image information (and have different patterns of attenuation and cancellation), it is advantageous to combine the images from all of the frequencies. The algorithm utilised in this study combined images taken from four frequencies, preserving the detailed texture from the higher frequencies while keep the slow attenuation rate from lower frequencies. It is necessary to weight the images to account for the fact that images from lower frequencies are usually brighter than those from higher frequencies. Suppose the number of frequencies built in the ultrasound machine is R. To eliminate the difference between the overall image brightness captured at different frequencies, the standardisation of overall image brightness captured at different frequencies is essential. Therefore, all pixel values $f_r(i,j)$ in the rth image were scaled by the factor:

$$\frac{\text{sum}(f_R)}{\text{sum}(f_r)}, \tag{4}$$

where $r = 1, 2, \ldots, R$ and represents the index of the frequency in ascending order, and f_r is the image intensity from the corresponding frequency. In our case, $R = 4$, and $r = 1, 2, 3, 4$ represents frequency 8, 10, 12 and 13 MHz, respectively. f_1, f_2, f_3, f_4 are the image intensities from the four frequencies 8, 10, 12, 13, respectively.

This weighting reduced the total intensity of the images at frequencies 8, 10 and 12 MHz and equalised them to the total intensity of the 13 MHz image. This prevented the processed images from being dominated by the intensity of the lower frequency images. The 13 MHz image was used as the base to reduce the overall intensity of the images taken at 8, 10 and 12 MHz; therefore the images taken at 13 MHz itself did not require any intensity adjustment here.

After intensity equalisation, a new image was created by weighting the images of all frequencies differently. The weighting strategy used the attenuation rate of the image from the highest frequency as the reference:

$$P_i = \frac{\sum_{i^*=1}^{i^*=i} f_r(i^*, J)}{\sum_{i^*=1}^{i^*=N} f_r(i^*, J)}, \tag{5}$$

$$A_{iJ} = \frac{i - L}{M - L}, \tag{6}$$

$$f_c(i,j) = f_R(i,j) + \sum_{r=1}^{R-1}\left[A_{iJ} + \left(\frac{r-1}{R-2}\right)(1 - 2A_{iJ})\right]P_i f_r(i,j), \tag{7}$$

where M and N are the number of rows and columns of the image, P_i is the percentage of the accumulative intensity of the image from row 1 to row I in column J, A_{ij} is an adjusting factor and L is the maximum row number where $P_i = 0$ in column J. In our case, $R = 4$, and $r = 1, 2, 3, 4$ represents frequency 8, 10, 12 and 13 MHz, respectively. f_1, f_2, f_3, f_4 are the image intensities from the four frequencies 8, 10, 12, 13 MHz, respectively. The weighting scheme in our case became:

$$f_c(i,j) = f_4(i,j) + (1 - A_{ij})P_i f_3(i,j) + 0.5P_i f_2(i,j)$$

$$+ A_{ij}P_i f_1(i,j). \tag{8}$$

As the ultrasound wave from frequency 13 MHz attenuates and P_i increases, more information from the lower frequencies is contributing to compensate for the decreasing information from the higher frequencies. With the weighting mechanism of this algorithm, the detailed texture from higher frequencies dominated at shallow depths. As the wave travels and attenuates, less information in higher frequencies is present, so more and more information is retrieved from lower frequencies. The frequency combining portion of the algorithm therefore created a composite image that preserved detail from the higher frequencies and also preserved the slow attenuation rate of the lower frequency images.

2.3 Technical validation: strain measurements in tendon

Anatomical lengths and tissue strain are frequently measured in ultrasound images; the accuracy of these measures is dependent upon the resolution and quality of the ultrasound images being analysed. We therefore measured strain in a tendon model in both raw and super-resolved ultrasound images and compared these to actual measured strain in the tendon.

Three porcine digital flexor tendons, completely excised from surrounding tissue but with intact bone-tendon insertion site, were mechanically tested in a servo hydraulic test system (MTS 858, Minneapolis, MN, USA). Original tendon length was recorded for strain calculations. A bead of graphite-impregnated silicone was placed on the surface of the tendons to provide a non-deforming image segment.

Tendons were incrementally stretched, 0.1 mm at a time at the grips, inside a saline-filled bath (which facilitated transmission of ultrasound waves), to a non-damaging physiological strain ($< 6.5\%$ strain). Four ultrasound images, one each at 8, 10, 12 and 13 MHz, were captured at each stretch increment (GE Logiq$e$, Fairfield, CT, USA). Unprocessed images were combined into four videos (one at each frequency); this is the 'raw video'. Images were subjected to (1a) a super-resolution

processing using (2a) multiple motion estimation, followed by (3a) combining the multiple frequency images; this is called 'SuperResolutionVideo 1'. Raw images were then combined in the alternative order, first (1b) combining the different frequency images, (2b) super-resolving the results and using (3b) the multiple motion estimation; this is called 'SuperResolutionVideo 2'.

Each of the videos (raw video, SuperResolutionVideo 1 and SuperResolutionVideo 2) was loaded into a custom DIC-tracking algorithm to calculate average strain in the tissue. Accuracy of the strain values was measured by comparison to the actual strain value, and precision of the strain values was measured by calculating the coefficient of variation (CV).

2.4 Technical validation: strain and contrast ratio in tumour phantom

Several methods of comparing tissues based on relative stiffness (i.e. elastography) require software to compare relative strain changes in adjacent tissues; the repeatability of these measures is dependent on the resolution and quality of the ultrasound images being analysed. We therefore measured strain in a breast tumour phantom in both raw and super-resolved images. Furthermore, the ability of physicians to use imaging modalities such as ultrasound to detect and diagnose lesions (i.e. tumours) is dependent on their ability to visually distinguish the lesion from the surrounding tissues. Therefore, the ability of an algorithm to improve the contrast between different regions in an ultrasound image is highly valuable to this type of ultrasound application. We also measured the contrast ratio between the 'tumour' and 'surrounding tissue' in the breast tumour phantom.

A silicone breast tumour phantom (provided by SuperSonic Imagine, Aix-en-Provence, France) was mechanically compressed in a servo hydraulic test system (MTS 858, Minneapolis, MN, USA). The 'breast' was incrementally compressed, 0.1 mm at a time, with the ultrasound transducer positioned immediately over the 'tumour'. Four ultrasound images, one each at 8, 10, 12 and 13 MHz, were captured at each compression increment. Unprocessed images were combined into four videos (one at each frequency); this is the 'raw video'. Images were subjected to (1a) a super-resolution processing using (2a) multiple motion estimation, followed by (3a) combining multiple frequency images; this is called 'SuperResolutionVideo 1'. Raw images were then combined in the alternative order, first (1b) combining the different frequency images, (2b) super-resolving the results and using (3b) the multiple motion estimation; this is called 'SuperResolutionVideo 2'.

Each of the videos (raw video, SuperResolutionVideo 1 and SuperResolutionVideo 2) was loaded into a custom DIC-tracking algorithm to calculate the average strain in

the tissue. Precision of the strain values was measured by calculating the CV. Videos (raw videos, SuperResolutionVideo 1 and SuperResolutionVideo 2) were also loaded into a custom segmentation algorithm to calculate ultrasound image brightness such that the brightness of the pixels inside the 'tumour' region and in the surrounding 'breast tissue' regions could be quantified and compared, in order to determine the contrast ratio between these two 'tissue' types.

2.5 Statistical analysis

The CV was calculated to compare the accuracy of strain measurements, providing a comparison of the standard deviation to the mean:

$$CV = \frac{\sigma}{\mu} \times 100\%. \qquad (9)$$

ANOVA calculations were performed to compare parameters from groups of videos larger than 3 or more (i.e. between the four frequencies), and Student's *t*-tests were performed to compare pairs of parameters (i.e. between ultra-videos 1 and 2). Significance was set at $p < 0.05$.

3. Results

3.1 Raw ultrasound images

Figure 3 shows the raw ultrasound images of a tendon specimen (with the graphite-impregnated silicone bead in the white box) acquired at 8, 10, 12 and 13 MHz. As expected, images at 8 MHz were brighter than those at higher (i.e. 13 MHz) frequencies due to lower attenuation rate. The lower attenuation rate was also demonstrated in the decreased 'shadowing' effect behind the silicone bead, which was quite marked in the 13 MHz image. Furthermore, the blurring effects were more pronounced in the 8 MHz image, as seen by the amount of artefact present around the silicone bead (lack of a crisp edge, with a halo effect around the edges of the bead).

Similarly, Figure 4 shows the raw ultrasound images of the breast tumour phantom acquired at 8, 10, 12 and 13 MHz. Again, images at 8 MHz were brighter than those at higher (i.e. 13 MHz) frequencies due to the lower attenuation rate.

3.2 Super-resolution image results

Comparison of the raw videos (at each of 8, 10, 12 and 13 MHz) to the processed images shows three major changes to the video quality: first, the new videos were larger (in terms of number of pixels). Second, the images showed greater clarity and detail. Third, there was a

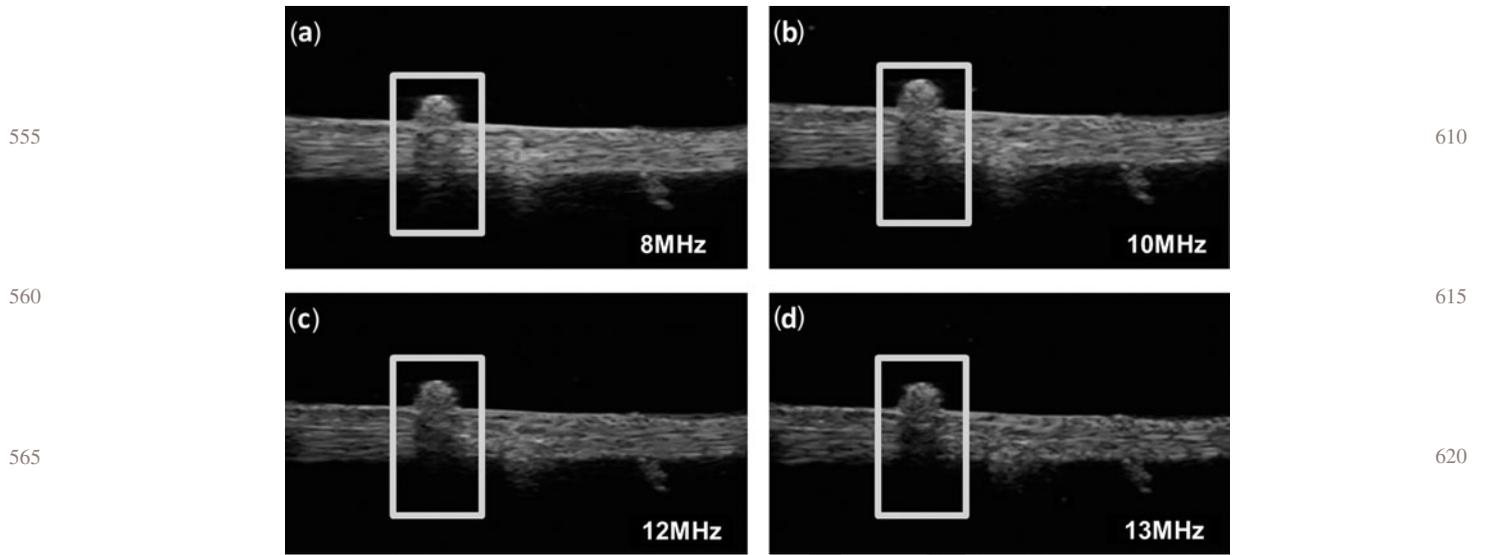6   *M.-J. Wu* et al.

Figure 3.   Raw ultrasound images of a tendon specimen at 8, 10, 12 and 13 MHz. Higher frequency waves carried more detailed texture resulting in sharper edges at the bump in the image (marked by the red box) where there was a halo in the lower frequency image. As the wave travelled farther into the tissue, it attenuated faster in higher frequency, causing the region behind the bump to appear darker.

smooth transition between ultrasound image frames rather than visual 'jittering' between consecutive frames.

The frequency combining portion of the algorithm created a composite image that preserved detail from the higher frequencies and also preserved the slow attenuation rate of the lower frequency images. Figure 5 shows tendon images from lower (8 MHz) and higher (13 MHz) frequencies (Figures 5(a),(b)) along with the image
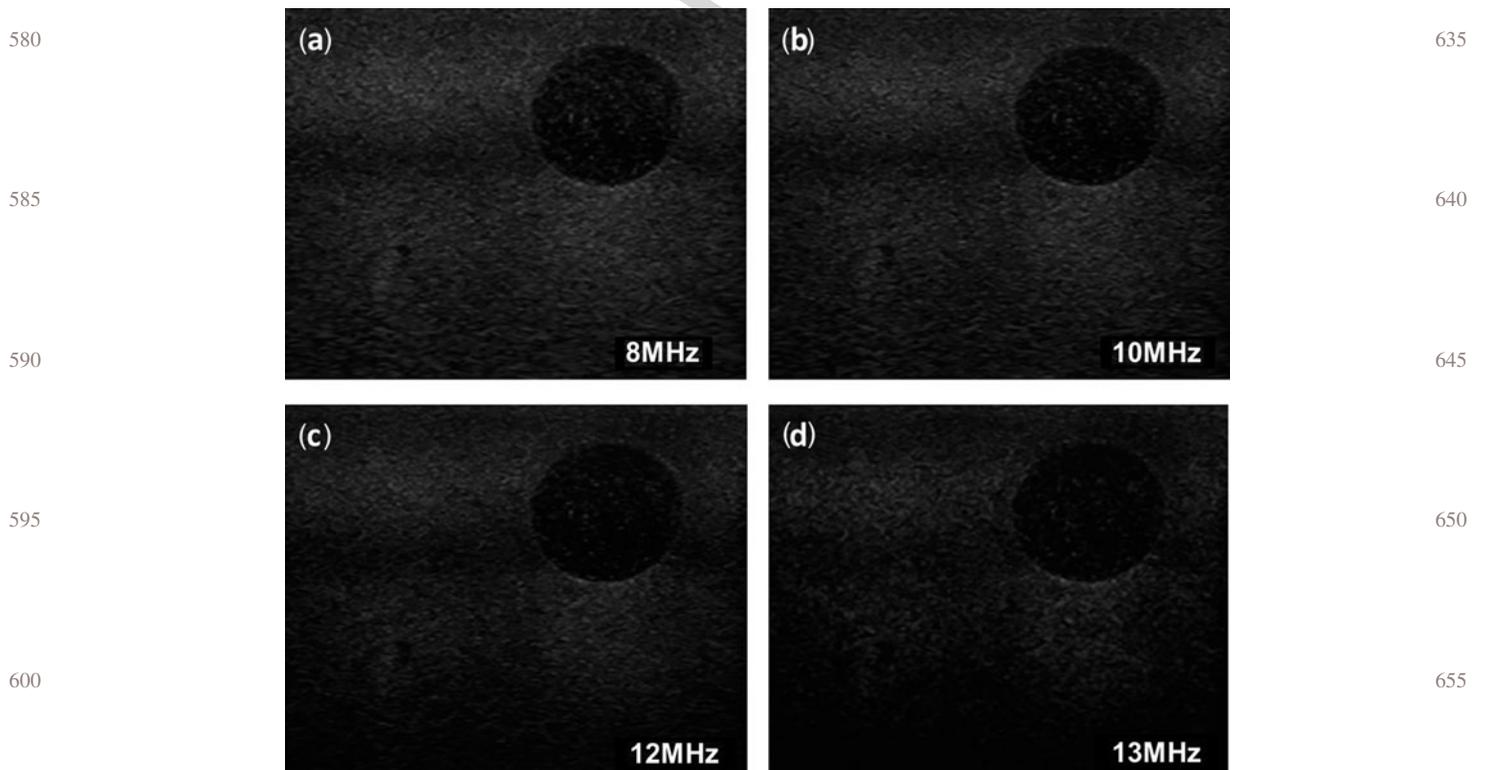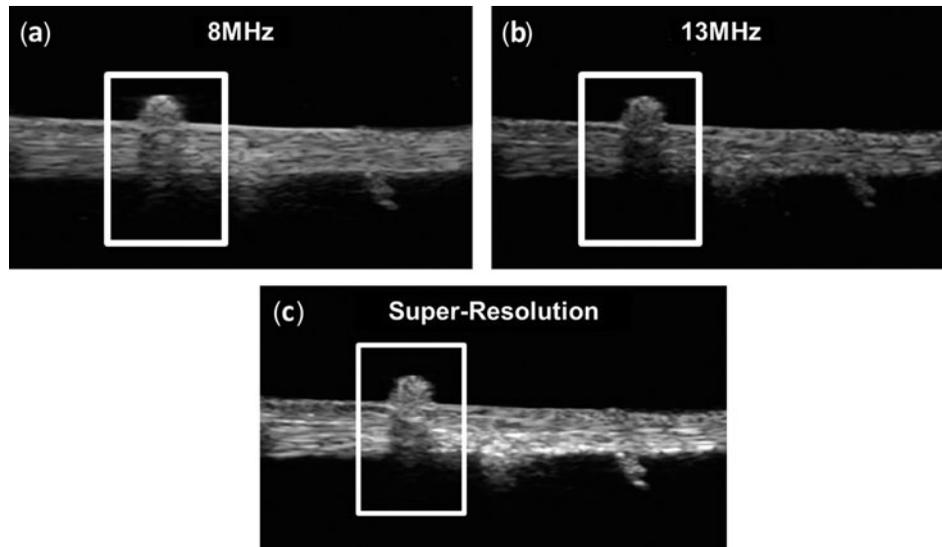

Figure 4.   Raw ultrasound images of a breast tumour phantom at 8, 10, 12 and 13 MHz. Lower frequency images were brighter but more blurry, while higher frequency images had better clarity but faster wave attenuation.

Figure 5.   Raw and processed ultrasound images of tendon: (a) tendon image captured at 8 MHz; (b) tendon image captured at 13 MHz and (c) frequency combined super-resolution image created by the algorithm.

created by our algorithm (Figure 5(c)). The algorithm created an image, which preserved the detailed texture of the higher frequency images while also exploiting extra information at depth from lower frequencies. Therefore, the image showed a sharp image around the bump with no halo, and it also presented information from lower frequencies, so the region behind the bump in our image was brighter than in the higher frequency images.

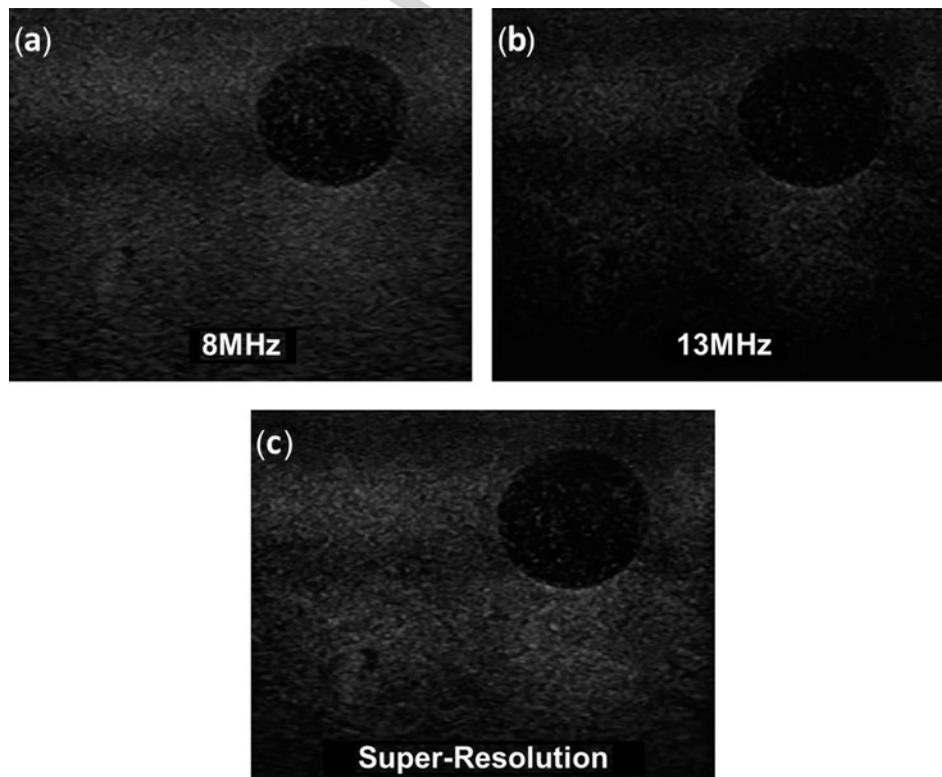Similarly, Figure 6 shows breast phantom images from lower (8 MHz) and higher (13 MHz) frequencies (Figure 6

Figure 6.   Raw and processed ultrasound images of a breast tumour phantom (dark circle): (a) breast phantom image captured at 8 MHz; (b) breast phantom image captured at 13 MHz and (c) frequency combined super-resolution image created by our algorithm.

(a),(b)) along with the image created by the super-resolution algorithm (Figure 6(c)). The algorithm created an image, which preserved the detailed texture of the higher frequency images while also exploiting extra information at depth from lower frequencies.

3.3 *Technical validation: strain measurements in tendon*

Tendon strain measurements in raw and processed videos are summarised in Table 1. Tendons 1, 2 and 3 were stretched to strains of 4.68–6.43%. DIC strain measurements using raw videos ranged from 5.34% to 9.67%, whereas measurements using SuperResolutionVideos ranged from 4.38% to 6.65%.

Measurements made from SuperResolutionVideos 1 and 2 are closer to actual values than those made from the raw videos, and SuperResolutionVideo 1 gives approximately the same values as SuperResolutionVideo 2. Thus, spatial and frequency-based super-resolution of ultrasound images results in improved results regardless of the order of application. Figure 7 demonstrates these results graphically, with the raw videos demonstrating variable and noisy strain tracking results while the super-resolved videos demonstrate strain tracking results which converge on the actual, measured strain (4.68%).

There was no significant difference in strain for raw images at different frequencies ($p = 0.725$; ANOVA calculation), so strain values were averaged across all the four frequency videos. Also, SuperResolutionVideos 1 and 2 were similarly averaged as there was no significant difference in strain values ($p = 0.499$; t-test calculation). Table 2 provides further comparative data.

The results in Table 2 demonstrate improved accuracy (as demonstrated by a reduced strain error; $p = 0.018$, t-test comparison) and precision (as demonstrated by a reduced CV; $p = 0.018$, t-test comparison) in the super-resolved images compared with the raw videos.

3.4 *Technical validation: strain and contrast ratio in tumour phantom*

Variability in strain measurements of the compressed tumour phantom is demonstrated in Table 3. The average

CV for raw videos was 20.62%, compared with an average CV value of 1.18% for super-resolution videos (SRVs). Thus, spatial and frequency-based super-resolution of ultrasound images results in reduced variability in strain measurements in a tumour phantom.

The average contrast ratio in the raw videos was 2.43, while the contrast ratio in the SRVs was 3.38. Thus, spatial and frequency-based super-resolution of ultrasound images results in improved contrast ratio in images of a breast tumour phantom.

4. Discussion

The challenges discussed in Section 1 regarding image registration and complex motion have been successfully resolved in this study. The additional technical challenge of frequency combination has also been addressed. The resulting combination algorithm provides a powerful method for creating better ultrasound images and videos, extending the existing technologies for image capture by post-processing. Comparison of the raw videos (at each of 8, 10, 12 and 13 MHz) with SuperResolutionVideos 1 (in which images undergo spatial super-resolution first and frequency combination second) and 2 (in which images undergo frequency combination first and spatial super-resolution second) shows three major changes to the video quality: (1) videos are larger in terms of number of pixels, (2) the images show greater clarity and detail and (3) images have considerably more smooth transitioning of the ultrasound image from frame to frame (rather than visual 'jittering' between frames). When looking at the videos, the SRVs are visually cleaner with sharper edges and better contrast, and should be easier to interpret in a clinical setting.

In order to validate that the improvements in the SRVs (over the raw videos) are more than cosmetic, we measured the strain in the tendon directly from the various videos. Measurements made from SuperResolutionVideos 1 and 2 are significantly more accurate and precise than those made from the raw videos. We attribute this improvement to the extra stability of the image (i.e. the reduced flicker) and to the increased resolution. Strain and contrast ratio were also measured in a breast tumour phantom with similar improvements in precision and

Table 1. Tendon strain measurements (raw videos vs SuperResolutionVideo).

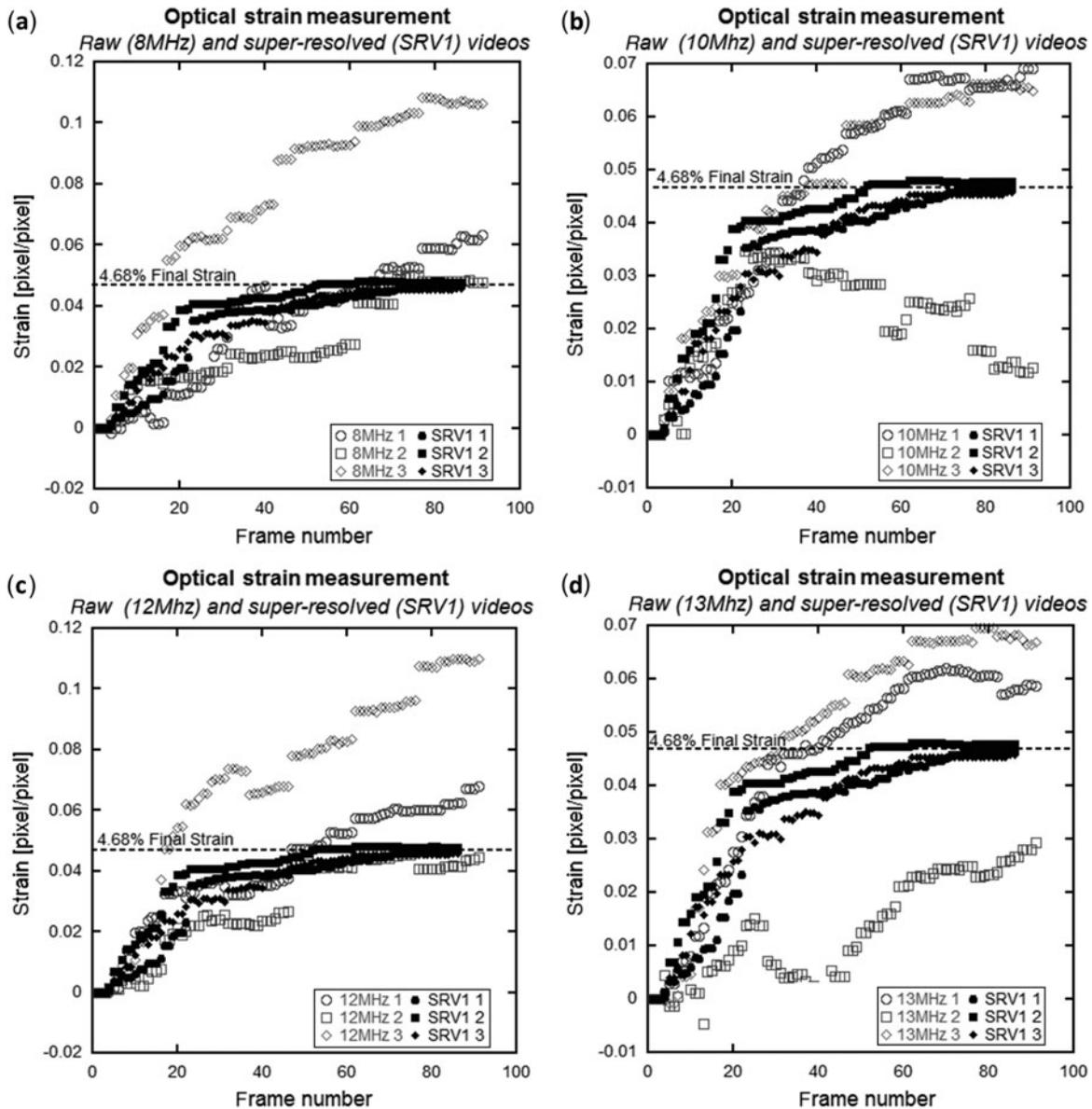|		8 MHz	10 MHz	12 MHz	13 MHz	Actual	SRV 1	SRV 2
Tendon 1	Strain (%)	8.84	8.42	8.90	9.67	6.43	6.65	6.60
	CV (%)	16.84	27.08	10.45	34.50		1.63	1.34
Tendon 2	Strain (%)	6.14	5.69	5.78	6.57	4.83	5.14	4.38
	CV (%)	53.33	33.51	33.98	17.74		9.52	9.68
Tendon 3	Strain (%)	7.39	5.71	7.47	5.34	4.68	4.67	4.68
	CV (%)	41.17	33.89	43.55	39.97		2.38	11.10

Figure 7.　Example tendon strain tracking results. SRV 1 indicates the SuperResolutionVideo in which the spatial super-resolution was performed first followed by frequency combination. Raw videos had variable and strain tracking results, while the super-resolved videos had results which converge on the actual strain (4.68%).

increased contrast ratio. We conclude that SuperResolutionVideo 1 is approximately the same as SuperResolutionVideo 2 and that both are significantly more accurate than the standard B-mode cine images.

Each of the components (frequency combination or spatial super-resolution) of the super-resolution algorithm contributed to the improvements in the images using different approaches; the interpolation used in the spatial super-resolution results in reduced noise and improved overall resolution, while the weighting scheme in the frequency combination creates an image with the increased detail from the higher frequency images and the decreased attenuation from the lower frequency

images, resulting in improved contrast and sharper details within the image. Thus, while the order in which they occur does not affect the outcome, maintaining both components of the algorithm is essential to improve the diagnostic ability of each individual still image while also improving measurements based on the dynamic videos.

While the testing performed thus far focuses on a relatively small number of samples from a tendon model and a tumour phantom, we believe the same general improvement should occur in any processed ultrasound images, such as of pulsing aortas, potentially tumorous lungs or beating hearts. The tendon model was chosen both because it is readily available and because it is a good

10 *M.-J. Wu* et al.

Table 2. Average strain measurement values (raw videos vs SuperResolutionVideos).

		Raw	Actual	SRV
Tendon 1	Strain (%)	8.96	6.43	6.63
	Strain error (%)	35.6		3.3
	CV (%)	22.22		1.49
Tendon 2	Strain (%)	6.05	4.83	4.76
	Strain error (%)	25.10		1.40
	CV (%)	34.64		9.60
Tendon 3	Strain (%)	6.48	4.68	4.68
	Strain error (%)	38.5		0.10
	CV (%)	39.65		6.74
Average	Strain error (%)	33.1		1.6
	CV (%)	31.63		5.85

model of linear tensile motion in the body. The breast tumour model was chosen because it represented an entirely different motion – the compression of a spherical object in a dense medium – while still being conducive to controlled testing. Thus, multiple motions seen in the body – tension versus compression and linear versus radial strain – have been demonstrated with these two models. The success of the algorithm in these different cases suggests that it will be successful in a wide variety of ultrasound images.

The range of frequencies chosen is based on the capabilities of our laboratory ultrasound system. We believe that further enhancements are possible for more advanced systems with broader frequency ranges. However, despite the fact that current ultrasound machines offer multiple frequency selections, they cannot simultaneously (or nearly simultaneously) save the same image at different frequencies. Therefore, the setting (displacement) of the tissue has to remain identical when the ultrasound image is captured by different frequencies. Otherwise, images acquired at different settings would result in different pixel locations between frequencies and would require alignment of the same frame number at different frequencies before being processed. Nearly simultaneous acquisition of different frequencies could be built into future ultrasound systems to obviate this issue.

Due to the nature of movement of biological tissues in ultrasound images, it is highly unlikely that complete alignment of the LR images can be achieved by a single affine transformation. To account for this, the images are partitioned into a collection of 36 sub-images, and motion

Table 3. Breast phantom strain precision results (raw videos vs SuperResolutionVideos).

	8 MHz	10 MHz	12 MHz	13 MHz	Super Resolution Video 1	Super Resolution Video 2
CV (%)	11.95	16.90	12.84	38.64	1.17	1.18

is estimated in each partition separately. This number of sub-images seemed to be adequate for the images in this study; however, if there is greater motion in future images, *n* may need to be increased. Furthermore, if future images have less resolution, *n* may be decreased for enhanced tracking efficiency.

Our algorithm consists of spatial super-resolution and frequency-based super-resolution. Spatial super-resolution improves the image quality by gathering information of the same object from different locations at same frequency. It creates images with higher resolution and removes the noise in the raw data (LR images) by registering and interpolating the consecutive frames at the same frequency. In the meantime, frequency-based super-resolution improves the image quality by collecting information of the same object at the same location but at different frequencies. It preserves the detailed image texture from high-frequency images and the brightness from low-frequency images to create images with higher image contrast. The combination of these two approaches gives images with higher resolution, more details, better quality, less noise, greater contrast and better stability.

The ability of physicians to use imaging modalities such as ultrasound to detect and diagnose the region of interest (e.g. pathological region or tumour) is dependent on their ability to visually distinguish the region of interest from the surrounding tissues. Therefore, the ability of our algorithm to improve the contrast between different regions in an ultrasound image is highly valuable in this type of ultrasound application.

In conclusion, the present spatial and frequency-based super-resolution algorithm successfully addressed the challenges of image registration and complex motion and exploited the ability to utilise information from multiple available ultrasound frequencies. The resulting processed images contain more detail and less noise, which improved contrast within the images and resulted in better strain measurements calculated from the images. Thus, the algorithm could be beneficial to both scientific analyses and clinical diagnoses.

References

Balci M, Foroosh H. 2006. Subpixel registration directly from the phase difference. J Appl Signal Process. 2006:231–231.

Bay H, Ess A, Tuytelaars T, Van Gool L. 2008. Speeded-up robust features (SURF). Comput Vis Image Und. 110:346–359.

Blomgren P, Papanicolaou G, Zhao H. 2002. Super-resolution in time-reversal acoustics. J Acoust Soc Am. 111:230–248.

Borman S, Stevenson R. 1998. Spatial resolution enhancement of low-resolution image sequences – a comprehensive review with directions for future research. Lab. Image and Signal Analysis, University of Notre Dame, Tech. Rep. Research Report.

Capel D, Zisserman A. 2000. Super-resolution enhancement of text image sequences. In: 15th International conference on pattern recognition, 2000. Proceedings. Vol. 1. p. 600–605.

Cross GR, Jain AK. 1983. Markov random field texture models. IEEE Trans Pattern Anal Mach Intell. 1:25–39.

D'Astous FT, Foster FS. 1986. Frequency dependence of ultrasound attenuation and backscatter in breast tissue. Ultrasound Med Biol. 12:795–808.

Elad M, Feuer A. 1997. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. IEEE Trans Image Process. 6:1646–1658.

Farsiu S, Robinson MD, Elad M, Milanfar P. 2004. Fast and robust multiframe super resolution. IEEE Trans Image Process. 13:1327–1344.

Fort A, Manfredi C, Rocchi S. 1995. Adaptive SVD-based AR model order determination for time-frequency analysis of Doppler ultrasound signals. Ultrasound Med Biol. 21:793–805.

Freeman WT, Jones TR, Pasztor EC. 2002. Example-based super-resolution. IEEE Comput Graph Appl. 22:56–65.

Haralick RM, Shanmugam K, Dinstein I. 1973. Textural features for image classification. IEEE Trans Syst Man Cybernet. SMC-3:610–621.

Ho T-C, Zeng B. 2007. Super-resolution images by support vector regression on edge pixels. In: International symposium on intelligent signal processing and communication systems, 2007, ISPACS 2007. p. 674–677.

Humblot F, Mohammad-Djafari A. 2006. Super-resolution using hidden Markov model and Bayesian detection estimation framework. EURASIP J Adv Signal Process. 2006:1–16.

Ikeda T, Yoshizawa S, Tosaki M, Allen JS, Takagi S, Ohta N, Kitamura T, Matsumoto Y. 2006. Cloud cavitation control for lithotripsy using high intensity focused ultrasound. Ultrasound Med Biol. 32:1383–1397.

Ji H, Fermuller C. 2009. Robust wavelet-based super-resolution reconstruction: theory and algorithm. IEEE Trans Pattern Anal Machine Intell. 31:649–660.

Keys R. 1981. Cubic convolution interpolation for digital image processing. IEEE Trans Acoust Speech Signal Process. 29:1153–1160.

Liu H-L, Chen W-S, Chen J-S, Shih T-C, Chen Y-Y, Lin W-L. 2006. Cavitation-enhanced ultrasound thermal therapy by combined low- and high-frequency ultrasound exposure. Ultrasound Med Biol. 32:759–767.

Lowe DG. 2004. Distinctive image features from scale-invariant keypoints. Int J Comput Vis. 60:91–110.

Martins ALD, Homem MRP, Mascarenhas NDA. 2007. Super-resolution image reconstruction using the ICM algorithm. In: IEEE international conference on image processing, ICIP 2007. p. 205–208.

Miller DL, Dou C, Wiggins RC. 2008. Frequency dependence of kidney injury induced by contrast-aided diagnostic ultrasound in rats. Ultrasound Med Biol. 34:1678–1687.

Park SC, Park MK, Kang MG. 2003. Super-resolution image reconstruction: a technical overview. IEEE Signal Process Mag. 20:21–36.

Patanavijit V, Jitapunkul S. 2006. An iterative super-resolution reconstruction of image sequences using affine block-based registration. In: Proceedings of the 2006 international conference on wireless communications and mobile computing. New York, NY: ACM. p. 51–56.

Pham TQ, Van Vliet LJ, Schutte K. 2006. Robust fusion of irregularly sampled data using adaptive normalized convolution. EURASIP J Adv Signal Process. 2006.

Shechtman E, Caspi Y, Irani M. 2005. Space-time super-resolution. IEEE Trans Pattern Anal Mach Intell. 27:531–545.

Sun C, Pye SD, Browne JE, Janeczko A, Ellis B, Butler MB, Sboros V, Thomson AJW, Brewin MP, Earnshaw CH, et al. 2012. The speed of sound and attenuation of an IEC agar-based tissue-mimicking material for high frequency ultrasound applications. Ultrasound Med Biol. 38:1262–1270.

Vandewalle P, Sü S, Vetterli M. 2006. A frequency domain approach to registration of aliased images with application to super-resolution. EURASIP J Adv Signal Process. 2006:1–14.

Zitová B, Flusser J. 2003. Image registration methods: a survey. Image Vis Comput. 21:977–1000.

[Q3]

[Q4]

[Q5]

[Q6]

1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210