1. INTRODUCTION

1.1. Motivation

The term ‘elastography’, first suggested by Ophir et al. [1], refers to methods of imaging the mechanical properties of tissue, specifically those related to the elastic (Young’s) modulus. The idea that mechanical properties are related to pathology is not new: clinicians have employed manual palpation to feel for stiff lumps as long ago as 400 BC [2]. However, recent improvements in imaging technology have allowed accurate, high resolution, in some cases quantitative visualization of the stiffness of relatively deep-lying structures.

The rapid development of this technique has spawned clinical trials in many anatomical areas. These indicate that elastograms may differentiate between benign and malignant lesions, and distinguish between different types of malignancy [3–8], in some cases better than conventional B-mode ultrasound images [9]. Possible applications include discrimination without biopsy between complex cysts and malignant breast lesions [10], monitoring of atherosclerosis [11,12], detection and grading of deep vein thrombosis [13], assessment of skin pathologies [14] and evaluation of myocardial fitness [15]. Sample images are shown in figure 1 from some of our own clinical trials. However, according to a recent evidence-based review [16], only transient ultrasound elastography (a quantitative technique for making point stiffness measurements, not images) was considered a clinically proven technique for assessing liver fibrosis: in all other cases, the evidence was inconclusive. Nevertheless, most manufacturers now offer or are developing some form of elastography imaging on their flagship ultrasound systems.

1.2. Elastographic techniques

Elastography involves the measurement of local tissue deformation in response to some sort of applied mechanical stress. The way in which the tissue deforms gives us information about the tissue’s mechanical properties. The many elastographic (or related) techniques can be loosely categorized according to how the stress is applied and how the deformation is measured. For example, stress application can be external by the probe [1] or internal via ultrasonic radiation force [17], while deformation measurement can be achieved using magnetic resonance imaging (MRI) [18] or ultrasound [19]. In quasi-static ultrasound elastography, the deformation is induced by manually pressing on the anatomy with the transducer, and measured using ultrasound. Such techniques are known by various names commercially, including eSie...
1.3. Tissue mechanics

The mechanics of tissue are highly complex and many assumptions have to be made in order to relate measurable values to a mechanical material property. These assumptions have the effect of introducing artefacts into elastograms, which are often incorrectly interpreted as a direct visualization of the material stiffness. The linear viscoelastic model [21–23] relates stress to strain, describing anisotropic material properties with a dependency over time, which might be related to biochemical changes [24]. Owing to the quasi-static nature of the scanning process, we can ignore viscous effects, giving rise to the linear elastic model [21,22]; however, this still contains far more parameters than can be constrained by practical empirical measurement. Hence, tissue is generally assumed to be isotropic, in which case stress and strain can be modelled as a simple function of the Young modulus $E$ and the Poisson ratio $v$. It is usually assumed that $v$ is close to 0.5 in most tissue, leaving $E$ alone to characterize an incompressible, isotropic material’s stiffness. It has been suggested that this parameter accounts almost entirely for the useful information accessible by manual palpation [25,26].

The time dependency is sometimes reintroduced to this simplified model by considering the material to be suffused by an incompressible, near-inviscid fluid [27]. Resulting measurements of porosity might be useful for assessing oedematous tissue [28,29].

Even with the simple, single parameter model, estimating $E$ is still problematic, since this relies on both stress and strain, and the stress field is not known. Figure 3 highlights this point with a finite-element simulation (ABAQUS v. 6.7, Simulia, RI, USA). We can measure the strain in figure 3c, but we really need to divide this into the unknown stress in figure 3b in order to estimate the stiffness in figure 3a. Normalization and inversion techniques will be discussed in §2.4: their aim is to recover something similar to $E$ from the strain measurements. However, even though the strain image does not directly relate to stiffness, it is apparent that the strain in the hard and soft lesions has good contrast and high spatial accuracy. Nearly all elastograms are actually versions of this strain image, rather than true stiffness images, and hence the term strain imaging or pseudo-strain imaging is sometimes used as a surrogate for elastography.

1.4. Review structure

The quasi-static elastography technique can be split into several distinct stages, which are all necessary for producing high quality, clinically useful elastograms. Section 2.1 reviews the methods for tracking the relative local deformation of ultrasound data between images, which lies at the heart of the technique. Section 2.2 discusses the measures used to calculate the quality of the displacement estimates and their use in limiting major tracking errors. Section 2.3 considers the process of creating strain data from displacement data, and methods of filtering these data. Section 2.4 looks at ways of creating images that...
are closer to stiffness than strain. Finally, §2.5 covers the process of creating dynamic, meaningful images from the raw image data. In §3, we also take a brief look at the related areas of three-dimensional imaging, and imaging of shear strain for detecting slip.

We summarize the various techniques that have been used for each stage, drawing examples from our own work in this area. The organization of the material in the following sections naturally follows our own understanding and emphasis in elastographic imaging. We neither claim to include everything which is good within the field, nor exclude everything which is not. Our aim is to give a broad overview of the field from which the technique can be explored.

2. QUASI-STATIC ELASTOGRAPHY

2.1. Tracking tissue movement

A key part of an elastographic imaging system is the algorithm used to estimate tissue deformation between two frames of ultrasound data. The probe movement in quasi-static elastography is seldom purely in the axial direction, and in any case even axial movement will deform the tissue in all dimensions. However, the axial deformation contains most of the useful mechanical information, and it can be measured more accurately than motion in other directions [31]. Hence lateral or elevational tracking, if employed at all, is mainly of use to improve the estimation of the axial deformation.

There are many different types of displacement estimators [32,33], but they can be grouped into certain categories. Most estimators are concerned with the matching of blocks of RF ultrasound data by rigid relative movement of the blocks until some similarity metric is optimized. Correlation maximization was the first such metric to be proposed [1,19,34] and remains the most widely used. Others include sum of absolute differences [35,36], sum of squared differences [37], correlation phase [38–40] and weighted phase separation [41]. Displacement must be estimated to sub-sample precision, either by interpolation of the similarity metric before finding the peak [37,39,42–53] or by interpolation of the RF data during the matching process [41,54]. Phase-based similarity metrics are particularly efficient at providing sub-sample accuracy. Optimal alignment is indicated by zero correlation phase [38,55], or zero phase separation [41], but a non-zero result, coupled with knowledge of the transducer’s centre frequency, allows rapid iteration towards the correct answer provided the initial alignment is within half a wavelength of the zero-phase shift [39]. An inaccurate estimate of the centre frequency does not affect the accuracy of the displacement estimate, just the speed of convergence [39].

Correlation maximization can be achieved by independent exhaustive searches at each window [19], or using uphill search from a guessed initial displacement, which must be within half a wavelength of the correct displacement, i.e. exactly the same constraint as in phase-based estimation. The initial guess can be achieved by imposing inter-window continuity, either by dynamic programming [56,57], or by minimizing a global cost function that penalizes discontinuous displacements [37,48,58], or by tracking the displacements from one window to the next [41,45,49,50,54,59,60]. The search at each window might be fully two-dimensional [44], or there may be separate, one-dimensional axial and lateral searches, perhaps with some iteration between the two [46,54]. There could be a single, high-resolution matching process [39,41,43,47,54], or a multi-level approach with a coarse matching stage initializing subsequent stages at progressively higher resolutions [36,37,48,61–63]. This matching process affects both the speed and robustness of displacement estimation. For example, exhaustive search is neither fast nor robust: it requires a large search range, and false-positive matches are common when the window size is small (as required for fine-scale displacement estimation) and the noise level significant [64].

A potential problem with all such approaches is that pre- and post-deformation windows can match poorly, because the data in the windows themselves may be deformed to different extents. Several techniques have been developed that involve warping the data to increase the correlation between windows. These are variously referred to as ‘temporal stretching’ [65], ‘adaptive stretching’ [42] or ‘companding’ [66], and involve searching over a stretch as well as a
displacement parameter when matching each window. Companding increases the precision of deformation data, but at huge computational cost. However, it has been noted that the main advantage of this approach is not so much the increase in correlation that it affords, but rather the ability to match each window at all points simultaneously [67].

The precision of displacement estimates from rigidly displaced window matching can be increased by noting the position within the window at which the pre- and post-displacement data match best. For most similarity measures, particularly phase-based methods, the estimation of this location is a simple calculation based on the signal amplitude, as in figure 4a [67], and has hence been termed amplitude modulation correction (AMC). Taking note of the location at which each displacement estimate is most valid, rather than just assuming that the estimate applies to the centre of the fixed window from which it is derived, provides a considerable increase in precision, as shown in the phantom data of figure 4d (elastography phantom, CIRS Inc., Norfolk, VA, USA).

The performance of all displacement estimators varies with the amount of relative strain between the pre- and post-deformation data. Elastographic signal-to-noise ratio (SNR$_e$), used as a measure of the performance of such algorithms, is shown in figure 5 for the best phase- and correlation-based algorithms with AMC applied, and for adaptive companding. This is based on simulated ultrasound data for which the ground truth deformation is known [67]. In practice, typical strains are less than 2 per cent, often much less than 1 per cent, and in this region the AMC methods are both quicker and potentially more precise.

**2.2. Quality metrics**

Since displacement estimation techniques are all based on optimizing similarity measures between pre- and post-deformation data, the value of this measure at the optimal displacement can also be used to give an indication of displacement estimation quality. In fact, the precision $W$ (the reciprocal of the variance or the mean-squared error) of each displacement estimate is approximately given by [68]

$$W = \frac{\rho}{1 - \rho}, \quad (2.1)$$

where $\rho$ is the normalized correlation coefficient between the RF data in the matched windows. An equivalent expression can be derived for phase-based methods, which makes use of residual phase error rather than normalized correlation [69].

Such quality metrics can be put to good use in many of the algorithms within quasi-static elastography. For example, all displacement tracking approaches that use previous estimates to initialize the following search can fail catastrophically if an incorrect displacement is found at any point and then propagated into other parts of the image. For simple forward tracking, such errors can be called drop-outs [60], since they generate a vertical line of errors in an elastogram. Typically, tracking algorithms assume that tissue exhibits continuous displacement in order to eliminate such
Figure 5. Comparison of SNR, for different displacement estimators. The best phase-based and correlation-based displacement estimators are shown together with a much slower adaptive estimate which does not use fixed windows. The window-based estimates outperform the adaptive estimate at typical strains of 1–2%, although adaptive estimation is increasingly worthwhile for larger strains. Derived from data used in Lindop et al. [67]. Bold line, phase-based using AMC; thin line, correlation-based using AMC; dashed line, adaptive.

errors. However, in in vivo scanning, apparent continuity is often violated either by slip planes or regions of complete ultrasound signal loss.

It is inevitable that such regions will result in localized errors in displacement estimation. However, propagation of these errors can be prevented if the tracking path is allowed to evolve dynamically according to the quality measure: so windows with high-quality matches are preferentially used to initialize the searches at their neighbours [41,54,59,60]. Displacement can even be tracked in disjoint regions separated by slip planes, provided multiple seeds are employed, at least one seed is planted in each region, and the tracking wavefronts evolve in a quality-guided manner [59,70].

A good tracking algorithm can therefore be seen as a means to increase the robustness of the displacement estimation process, whereas it is the fundamental displacement estimation algorithm that determines the precision. Figure 6 provides an example of this increase in robustness. Here, varying levels of noise are added to simulated data, and displacement estimation attempted using various fixed window lengths [59]. At each noise level, there are clear upper and lower bounds on the window lengths that allow successful recovery of the displacement field. Since varying the window length changes the trade-off between the precision and the resolution of the elastogram, an increased range allows much greater flexibility in optimizing the system parameters. The quality-based tracking strategies, particularly the multiple-seed [59] and two-pass [60] methods, which allow for multiple initialization points in the image, are more robust than the simpler tracking strategies.

2.3. Gradient estimation and filtering

Most quasi-static elastography systems follow displacement estimation with axial gradient estimation in order to generate axial strain. Filtering of some kind is nearly universal at this stage, since differencing of consecutive samples [1] (i.e. the digital equivalent of differentiation) amplifies the high-frequency components of the measurement noise. For instance, moving-average filtering (MAF) can be applied either before [48] or after [38] differentiation. Staggered strain estimation (SSE) [71], where displacements are differenced over a wider interval, has also been applied in a wide range of practical strain imaging investigations [29,72–78]. MAF and SSE have been shown to be mathematically identical [79].

Another popular technique for gradient estimation is piecewise-linear least-squares regression (PLLSR) [80], where the gradient is calculated from fitting a linear model to a window of data. This too can be interpreted as simple differencing followed by a linear filter, and it has been shown that even simple Gaussian-based filters can achieve lower estimation noise than these methods at the same resolution [79]. Many other linear filters are possible, such as cubic polynomial fitting [81].

This linear filtering operation has a significant impact on the ultimate resolution and accuracy of strain images. For the case of possibly the most popular, PLLSR, the precision of the strain estimates is then [68]

\[
W_A(x, y) = \frac{(\sum_j \hat{y}_j)^2}{\sum_j \hat{y}_j^2(1/W_j)},
\]

where the summations are over displacement windows in the regression kernel, \(\hat{y}_j\) denotes axial distance from the kernel centre, and \(W\) is defined in equation (2.1). The precision of strain estimates calculated through simple differencing can be derived from this by taking \(\hat{y}_j = \{-1, 1\}\).
Since most displacement tracking and filtering techniques make use of windows with fixed size, subsequent strain images have fixed resolution but variable quality. It is also possible to use matrix inversion schemes that act on the entire image and produce results that are globally optimal in some sense. For instance, non-parametric regression (NPR), using the strain quality as a per-pixel weighting, can generate strain images in real time of fixed quality but variable resolution [69].

Figure 7 contains some sample filtered images for three different types of filter, and three different extents of filtering. Global quality-based algorithms compare well with PLLSR or even quality-weighted Gaussian filtering, with well-defined borders between different strain regions, good suppression of noise and better preservation of the size of stiff objects.

2.4. Normalization and inversion of strain images

Thus far we have created a precise, filtered, version of the relative axial strain between two frames in an ultrasound scanning sequence, i.e. equivalent to the simulated data of figure 3c. However, ideally we want to display tissue stiffness, \( E \), in figure 3a, rather than axial strain. To do this, we need to solve an inverse problem for which we also need either the stress, figure 3b, or the stiffness \( E \) everywhere on the boundary of the data [82]. Although stress can be measured to some extent on the tissue surface using load cells [83], some assumptions (e.g. of uniformity [84]) are necessary to satisfy these conditions, even slight violations of which can lead to incorrect estimates of \( E \) [82].

The more popular, less quantitatively motivated, though much faster and better constrained approach is to display an image based directly on the axial strain. Such images always need some form of normalization to convert the strain into a displayable range, and to reduce variation that is simply a result of varying applied stress [68]. Hence if \( s_A(x,y) \) is the measured strain, the normalized value \( s_B \) is

\[
s_B(x, y) = \frac{s_A(x, y)}{\hat{s}(x, y)},
\]

where \( \hat{s}(x, y) \) is a normalization value that may simply be a constant representing the maximum strain in each image. The quality of \( s_B \) is then

\[
W_B(x, y) = W_A(x, y) \times \hat{s}(x, y)^2,
\]

where \( W_A \) was defined in equation (2.2). It is also possible to use more complex definitions of the normalizing function \( \hat{s}(x, y) \), for instance, to allow for uneven pressure across the probe face, or stress decay with depth in tissue [30, 68]. Such functions can be fitted to the strain data by making broad assumptions about the typical stiffness distribution in clinical data.

Figure 8a,b shows an example of axial strain-based display (breast biopsy phantom, CIRS Inc., Norfolk, VA, USA). Even the ‘raw’ strain image in figure 8a requires some form of normalization for display: this is really a normalized image with \( \hat{s}(x, y) \) as a constant value. The stiff inclusion is visible even in this image.
The normalized ‘pseudo-strain’ image in figure 8b has been created using the flexible scheme outlined in Lindop et al. [68]. The effect is to remove distracting variations in the background that did not represent variations in material stiffness. Normalization has also accentuated the noisy measurements at the bottom-left, which are the result of very little axial probe movement on that side owing to rotation of the probe about the elevational axis. In order to remove such noisy regions, we need to persist data from more than one pair of ultrasound frames, leading to the result in figure 8d: this is the subject of §2.5.

2.5. Persistence and visualization

The formation of an elastogram only requires two frames of ultrasound data: one before and one after an axial deformation of the tissue. However, there are several reasons for using many more frames of data to form a stiffness image. Firstly, if the probe is moved by hand, there will be some frame pairs that do not exhibit any relative deformation at all, or for which the deformation is inappropriate for measuring axial strain. Secondly, figure 5 shows that displacement estimators tend to perform best at fairly low strain, whereas it can be seen from equation (2.4) that higher strain increases the precision of the displayed, normalized value. The easiest way to satisfy both these constraints is to measure displacement between multiple low-strain frames, which accumulate to give higher strain deformations. Thirdly, ultrasound data are available at rates up to 200 frames per second, depending on the focusing technique and the imaging depth, generally much higher than the rates required for good visual responsiveness in the images.

One common approach to this problem is to create elastograms between all pairs of frames and then select the ‘best’ pair, either manually or with the use of a metric based on the quality discussed in §2.2 [85,86]. In fact, such a metric is often displayed dynamically as the scan progresses, to give users some feedback on the quality of their scanning technique. Such schemes do not make optimal use of the large amount of data available. To make use of all the data, sequences of images can be averaged in order to reduce noise [87], but this does not take into account the variations in quality within each elastogram.

It is possible to generate better sequences of elastograms by physically constraining the types of freehand probe motion [88]. Alternatively, the known quality $W_p(x,y)$ of the data can be used either as a per-image [89] or per-pixel weighting [68] in persisting the stream of images. An example of per-pixel quality-based persistence is shown in figure 8. The single elastogram in figure 8b contains some very good data, but also some areas of very poor data, either owing to inappropriate tissue deformation, or difficulty in tracking displacements owing to low or decorrelated ultrasound signals. The persisted elastogram in figure 8d has been created from a sequence containing the frame in figure 8b, and some other frames which contained good data in other regions. Per-pixel quality weighting has the effect of combining the best regions from each frame to build up the final displayed image. Poor-quality data in the scan sequence are effectively ignored, since in this case, the quality $W_p(x,y)$ is low [30].

There is currently little consensus on how elastograms should be displayed. Images in this paper have used greyscale for strain, with black indicating stiff and white indicating soft regions, but this scale is sometimes inverted. It is also popular to use a colourmap to display strain values, often based on a blue–yellow–red progression, though blue has been used to represent either stiff or soft regions. Such a colourmap can be overlaid on top of a conventional greyscale ultrasound image, with some transparency, so that, at least to some extent, both modalities are simultaneously visible. This display is familiar to clinicians in that it is similar to Doppler ultrasound, used for visualization of fluid flow. However, unlike Doppler, where the blue–red colour transition marks a change in flow direction, in quasi-static elastography, there are no clear pseudo-strain boundaries that could appropriately be aligned with the marked changes in colour. Hence it is potentially dangerous to associate regions which are ‘red’, ‘blue’ or any other colour with any particular pathology.

Quality values are generally also displayed alongside some form of pseudo-strain, at least as an independent visualization of the overall quality of the image. However, since the quality is available at a per-pixel level,
much better use can be made of it either by fading out strain data according to quality, or displaying quality as an overlay on top of the strain data [30]. The images in this paper have used a red overlay to indicate poor quality, gradually fading to a solid colour as the quality degrades, generated by our own freely available Stradwin software (http://mi.eng.cam.ac.uk/rwp/stradwin) [90,91]. Quality display acts as an important indication of poor scanning performance. Per-pixel quality overlays can also be used to obscure low-quality strain data, which might otherwise appear to represent actual stiffness variations, and hence confuse the interpretation of an elastogram, as in figure 8b,c.

3. RELATED TOPICS

3.1. Three-dimensional elastographic techniques

Many of the algorithms in the previous sections extend naturally to three dimensions. Indeed, since the stress and strain fields induced during deformation of tissue are inherently three-dimensional, it might therefore be assumed that much better stiffness data, particularly from inversion techniques, should be available by acquiring and processing in three dimensions. Many three-dimensional techniques are already well developed for conventional ultrasound imaging. However, the lateral and elevational components of displacement can generally be estimated with less precision [93], and hence much of the focus of three-dimensional elastography has been in increasing the precision of axial displacement, and hence axial strain.

There are two main approaches for three-dimensional elastography. Firstly, a sequence of two-dimensional elastograms can be created as the probe is moved elevationally [94], for instance, for intra-vascular [95] or prostate [96] imaging. Alternatively, a mechanically swept two-dimensional probe or two-dimensional phased-array probe can be used to acquire a volume of data, then a controlled or freehand compression applied [54,76,93,97–99] before further volumes are acquired. Some mixed approaches have also been suggested, either using a three-dimensional pre-compression volume with a two-dimensional post-compression scan [100], or a mechanically swept two-dimensional probe but with two-dimensional elastogram processing [101]. Such approaches generate three-dimensional datasets of axial strain. Figure 9 shows an example with an in-house and commercial phantom using the system presented in Treece et al. [54]. These data may be useful in accurately quantifying the volume of stiff lesions.

The difference between using probes that acquire fundamentally three-dimensional data, and sequences of data from a two-dimensional probe moved across the anatomy, is more than just in the techniques for processing the data. Mechanically swept two-dimensional probes and two-dimensional phased arrays have a much larger probe face than two-dimensional probes, and since it is the probe face that provides the pressure contact with the anatomy, this affects the nature of the generated stress and strain fields [1,102]. In theory, a larger probe acts more like an infinite compressor plate and should generate more uniform stresses. In practice, better images are sometimes generated by techniques which use two-dimensional probes [101], or one-dimensional probes with compression plates [88].

3.2. Slip and shear strain imaging

Elastograms based on axial strain show relative anatomical stiffness; however, it has been suggested that how lesions are attached to surrounding tissue may also correlate with malignancy [103]. The slip patterns generated by un-attached lesions are to some extent visible on elastograms, since slip results in discontinuity in the axial strain. A more direct way to visualize these patterns is by the axial-shear strain, i.e. the lateral gradient of axial displacement [78,104–107], or indeed some combination of axial- and lateral-shear strain [108,109].

Most of the discussion in this paper regarding axial strain can also be applied to the generation of an axial-shear strain image. Aside from the obvious difference in the direction in which the gradient is taken, there are two further differences. Both axial and axial-shear strain change sign dependent on whether they are induced by a compression or relaxation of the anatomy. In an elastogram, this sign change is removed at the normalization or inversion stage, but in
techniques to enable its development to date. Many of these techniques have been briefly outlined and contrasted in this paper. It is sufficiently mature to allow successful commercialization, even though the clinical case has yet to be conclusively made. However, it is clear that many assumptions are required in the formation of an elastogram, particularly in the mechanics of the tissue deformation. Relaxing some of these assumptions continues to provide a fruitful area of ongoing research.

We thank Ellen Neale for acquiring the clinical data in figure 1. Full ethical approval was obtained from the Cambridgeshire 3 Research Ethics Committee (reference no. 07/H0306/90). We also thank Chris Uff for acquiring the clinical data in figure 10. Full ethical approval was obtained from the National Hospital for Neurology and Neurosurgery and the Institute of Neurology Joint Regional Ethics Committee (reference 18/H0716/92).

REFERENCES


10 Barr, R. G. 2006 Clinical applications of a real time elastography technique in breast imaging. In Proc. 5th Int. Conf. on Ultrasonic Measurement and Imaging of Tissue Elasticity, Snowbird, UT, 8–11 October 2006, p. 112. Austin, TX: ITEC.


4. CONCLUSIONS

Quasi-static ultrasound elastography is an apparently simple idea, which has required surprisingly complex axial-shear strain images, the sign change is useful and needs to be preserved. Secondly, whereas the average magnitude of axial strain is non-zero, typically axial-shear strain images do have a mean value close to zero.

These changes mean a new normalization process is required for imaging axial-shear strain [110], though all other processing steps can remain the same. Figure 10 contains some examples of phantom (breast biopsy phantom, CIRS Inc., Norfolk, VA, USA) and in vivo axial-shear strain images produced using this normalization technique. The alternating bright and dark quadrants surrounding the inclusions are typical of axial-shear strain images, the location of the bright quadrants indicating whether the inclusions are stiff or soft. Relative shear strain levels inside and outside the inclusions may indicate how well they are bonded to the surrounding material [111,112].

Figure 10. Imaging of axial-shear strain. (a) B-mode image, (b) normalized axial strain and (c) normalized shear strain of a hard inclusion in a phantom. (d–f) Corresponding images acquired during neurosurgery. The level of axial shear within the inclusion, and the variation in axial shear surrounding the inclusion, provides an indication of how well it is bonded to the surrounding tissue. Derived from data used in Chen et al. [110].
10 Review. Quasi-static ultrasound elastography G. Trecce et al.


