Please cite the Published Version

Real-Time Ultrasound Segmentation, Analysis and Visualisation of Deep Cervical Muscle Structure

Ryan J. Cunningham, *Member, IEEE*, Peter J. Harding, and Ian D. Loram, *Member, IEEE*

Abstract—Despite widespread availability of ultrasound and a need for personalised muscle diagnosis (neck/back pain-injury, work related disorder, myopathies, neuropathies), robust, online segmentation of muscles within complex groups remains unsolved by existing methods. For example, Cervical Dystonia (CD) is a prevalent neurological condition causing painful spasticity in one or multiple muscles in the cervical muscle system. Clinicians currently have no method for targeting/monitoring treatment of deep muscles. Automated methods of muscle segmentation would enable clinicians to study, target, and monitor the deep cervical muscles via ultrasound. We have developed a method for segmenting five bilateral cervical muscles and the spine via ultrasound alone, in real-time. Magnetic Resonance Imaging (MRI) and ultrasound data were collected from 22 participants (age: 29.0 ± 6.6, male: 12). To acquire ultrasound muscle segment labels, a novel multimodal registration method was developed, involving MRI image annotation, and shape registration to MRI-matched ultrasound images, via approximation of the tissue deformation. We then applied polynomial regression to transform our annotations and textures into a mean space, before using shape statistics to generate a texture-to-shape dictionary. For segmentation, test images were compared to dictionary textures giving an initial segmentation, and then we used a customized Active Shape Model to refine the fit. Using ultrasound alone, on unseen participants, our technique currently segments a single image in ≈0.45s to over 86% accuracy (Jaccard index). We propose this approach is applicable generally to segment, extrapolate and visualise deep muscle structure, and analyse statistical features online.

Index Terms—ultrasound, cervical dystonia, segmentation, MRI, shape model, skeletal muscle, trapezius, splenius, semispinalis, multifidus, rotatores, generative shape model, electomyography, pattern recognition.

I. INTRODUCTION

THIS paper concerns segmentation of structures (skeletal muscles) which are dynamic, homogeneous in tissue type and speckle structure, and imaged by a low quality modality, ultrasound (US). There is currently negligible literature on segmentation of skeletal muscle within ultrasound

[1], [2], and no literature on the segmentation of cervical muscles. This is despite the high medical and economic impact of chronic neck/back pain and injury, the current need for personalised musculoskeletal diagnosis, and the widespread availability of ultrasound. These facts testify to the challenging nature of these medical imaging problems and their inadequate solution by existing methods. The solutions we provide here open this entire domain (i.e. low cost, personalised muscle diagnosis using ultrasound) for further development.

Ultrasound is the most challenging modality for image segmentation [2]. Compared with Computerised Tomography (CT), and (functional) Magnetic Resonance Imaging (f)MRI, the spatial resolution is low, the probe-tissue target relationship is variable, probe contact is variable, features, shape and texture within the target are inconsistent through time, the signal to noise ratio is low and in this domain the target (muscle) moves. Current methods are not suited to this domain of medical image analysis (see section II). Even if one assumes "perfect" ultrasound images (i) it is not currently established how the muscle boundaries map to the ambiguous shape and texture manifestation of the ultrasound image, and (ii) the complete muscle boundaries lie outside of the viewing region. These reasons rule out direct labelling of ultrasound images, and prompt the use of a multiple imaging modalities. MR images provide complete and visible cross-section boundaries; however, registration of these between modalities must address the compression of tissue which occurs only in US, and not MRI.

This study contributes a novel method for segmenting five bilateral cervical muscles and the spine, in real-time from ultrasound *alone*, providing real-time *in vivo* analysis and prediction/visualisation of whole muscle structure beyond the visible boundaries of the ultrasound plane. Further, this methodology is applicable generally to any set of muscles accessible with US, for example is immediately applicable to the entire spine, including the back. To deliver our contributions while addressing the challenges named above, we have developed a novel multimodal MRI-US shape annotation and registration method which facilitates the construction of physiologically accurate labelled ultrasound datasets, which is currently not possible. We have also developed a novel segmentation method, in which we construct a texture-to-shape dictionary from minimal amounts of labelled data, which we use to initialise and inform a local statistically-constrained heuristic pixel intensity search based

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2016.2623819, IEEE Transactions on Medical Imaging

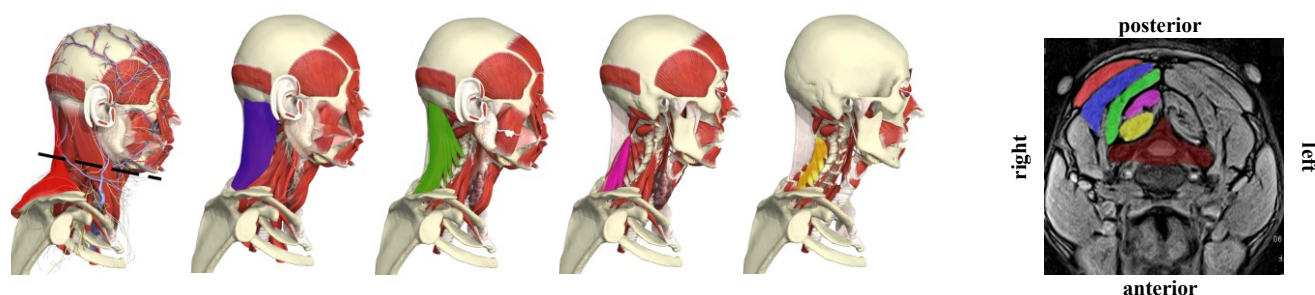> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <   2

Fig. 1. Cervical anatomy and cervical MR image. This figure shows a colour-coded sagittal view of the cervical anatomy (left) and a colour-coded axial MRI of a single participant's cervical muscles (right); the colour segments are respectively (MRI: posterior to anterior, anatomical: left to right), **upper trapezius**, **splenius**, **semispinalis capitis**, **semispinalis cervicis**, **multifidus**, and the **spine**. The dashed black line across the neck on the anatomical image of the trapezius (far left) represents approximately where the image on the right was taken from (i.e. an axial scan at that level and inclination).

on the well-established Active Shape Model (ASM; [3]). Delivery of our contributions provide novel insight into a new domain and fundamentally test the hypothesis that ultrasound contains the information required to locate an annotation defined in the MRI domain. This paper might be regarded as an experiment in which the fundamental question is whether or not ultrasound provides sufficient information to facilitate an anatomically accurate segmentation of the neck muscles.

As this domain is new and the literature is relatively sparse, we have written a more detailed background and rationale (section II), which includes basic cervical muscle anatomy, current alternative technologies for measuring the parameters of the cervical muscle system, its application to neck/back pain and injury with a specific application to Cervical Dystonia (CD), review of the relevant state-of-the-art segmentation literature, and the justification for our proposed method. Following the background and rationale we describe our methods (section III) which include the data collection protocol (*A*), our novel image annotation protocol (*B*), the construction of our novel dictionary-to-shape template matching segmentation algorithm (*C-E*), and finally our novel heuristic image search algorithm based on the ASM (*F*). We present results (section IV) on the accuracy of our algorithm using the popular Dice similarity coefficient (DSC), and the more representative Jaccard index, which collectively reveal the extent to which whole muscle regions can be automatically identified. We also present accuracy results using the Hausdorff distance, which reveals the extent to which the muscle boundaries (shape) can be accurately traced. To facilitate comparison with state-of-the-art, we also compute and compare results of the widely used standard Active Appearance Model (AAM; [4]); an industry standard medical image segmentation algorithm. Finally we compare the segmentations of 3 experts following our annotation process. After presentation of results, we give a detailed discussion (section V) of our contributions and the derived results, detailing limitations and successes. We conclude (section VI) that the application is important and that real-time segmentation of the cervical muscles *in vivo* directly from ultrasound *alone*, is possible, where current standard techniques are inadequate to serve this purpose.

II. BACKGROUND AND RATIONALE

The morphometry of human cervical muscles is architecturally complex [5]; the posterior neck contains over five bilateral muscle layers and those muscles can cross joints and attach to multiple bones [5]. The size and shape of cervical muscles across a population exhibit wide variability, and cross-sectional areas do not scale proportionally with body height and/or weight [5]. Furthermore, there are significant differences in muscle shape between genders [6]. The variability and complexity of the cervical muscles present a challenge in defining a generalized model of muscle shape/architecture. The human cervical muscles are also functionally complex, exhibiting functional redundancy across multiple muscles [7].

There are a multitude of conditions (e.g. neck and upper limb pain and injury) that would benefit from the ability to measure the properties of the cervical muscle system *in vivo*. Neck pain is a highly prevalent condition that causes substantial disability [8]. Out of 291 conditions studied in the Global Burden of Disease Study (2010), neck pain ranked 4th highest in terms of disability, as measured by years lived with disability, and 21st in terms of overall burden [8]. Work-related upper limb and neck musculoskeletal disorders (MSDs) are one of the most common occupational disorders around the world [9]. Work-related upper limb disorder (WRULD), repetitive strain injury (RSI), cumulative trauma disorder, occupational overuse syndrome and work-related complaints of the arm, neck or shoulder are the most frequently used umbrella terms for disorders that develop as a result of repetitive movements, awkward postures and the impact of external forces [10]. The cost of WRULDs in the EU has been estimated to be between 0.5% and 2% of gross national product [9].

CD is a complex neurological disorder which causes involuntary muscle spasticity which can severely impair quality of life. One successful treatment of CD is to inject botulinum toxin into the affected muscle(s), which reduces the contraction(s) and relieves the person. However, there is difficulty in identifying which muscles need to be injected. There is also currently no non-invasive method for quantifying response to treatment, or severity of dystonia in deep cervical muscles – in fact the deep muscles often go untreated. A non-invasive method for identifying contractions in deep cervical muscles would be beneficial for targeting and monitoring treatment of CD. Although methods have been developed which would differentiate CD patients from healthy subjects

[11], there is currently no clinically accepted method for localizing and quantifying severity of CD in deep cervical muscles.

Kinematic analysis of head orientation can reveal force loads in the neck via mechanical inference [12], [13] however, due to muscle groupings (functional redundancy), complex muscle structure [5], and the counterintuitive nature of muscle activation [7], information extraction at the level of the muscle is currently not possible. Surface electromyography can measure muscle activation during contraction in superficial and partially superficial muscles in the neck [14], although it has been argued that sEMG is not appropriate for measurement of partially superficial muscles such as the splenius capitis layer [15]. Needle EMG is required to measure contraction in deep muscle layers. However, this approach is impractically invasive, time-consuming, and has a relatively small measurement volume (typically $\approx 1\text{mm}^3$) [16]. Furthermore, there can be no certainty about the clarity of the positioning of intramuscular electrodes [16]. The numerous methodological problems with EMG are well documented; this includes cross-talk, electrical interference, mechanical artefacts, and activity of adjacent muscles in contact with the same electrode [17].

Imaging technology such as MRI and ultrasound allow non-invasive imaging of cross-sectional areas of every bilateral layer of muscle in the neck [7], [18]–[20]. MRI gives high spatial, low temporal resolution images of internal body structures. Studies have shown that functional MRI (fMRI) has sufficient temporal resolution (0.1-0.5Hz) to measure contractions in deep cervical flexors [20], however, image quality is severely impaired. In comparison to MRI, ultrasound has higher temporal frequency (20-100Hz+) – allowing observation of muscle function – and has lower image quality. While, in comparison to fMRI, ultrasound has higher temporal frequency and higher image quality. Studies have shown that ultrasound is comparatively as accurate as MRI and computerized tomography (CT) for measuring thickness of deep cervical muscles [21]; the authors of that study note that consistency between comparisons of the different modalities was improved by the use of image-plane markers.

Previous work has considered shape parameter measurement consistency in the deep multifidus via US, concluding that it was a reliable method for measuring muscle dimensions, both while at rest, and under contraction [18], [19], [22]. A survey on ultrasonography of the cervical muscles concludes that there is insufficient literature on assessment of the cervical muscles via rehabilitative ultrasound imaging, and that there is a need for proper identification of muscle boundaries, using landmarks and knowledge of functional anatomy [1]. The authors further argue that standardized positions of subjects and ultrasound transducers are important for statistical analysis of shape parameter measurements.

A broad review of the literature reveals that there is negligible published work on the segmentation of skeletal muscle within ultrasound. The cervical muscles represent one of the most challenging groups to segment within one of the most challenging imaging modalities, ultrasound [2]. There are no previously published works on segmentation of the cervical muscles; a morphologically and physiologically complex muscle group. There is however a body of work on cardiac muscle segmentation within ultrasound [23]–[27]. Cardiac muscle is a multi-segment muscle system, which is rhythmical and continually active. This differs from skeletal muscle groups, which can activate independently, and appear differently depending on the configuration of the connecting bones and the active state of the muscles, which is particularly prominent with a disease such as cervical dystonia. A further distinction can be made between cardiac muscle and skeletal muscle in the sense that cardiac muscle is not subject to compression resulting from probe contact, whereas skeletal muscle is.

ASMs provide a powerful paradigm for combining shape statistics to regularize heuristic segmentation algorithms, although they require shape statistics built up from many annotations of object boundaries over a population. Previous work has used ASMs to achieve fully automatic segmentation of a two-layer skeletal muscle system in the human triceps surae [28], of which this is the only example of automated muscle segmentation/analysis via ultrasound. However, annotating images of the cervical muscles – directly from ultrasound – is extremely uncertain due to the complexity of the architecture of the cervical muscles [5], and the challenging nature of ultrasound images [2], therefore there is currently no standard approach to obtain truth data for cervical ultrasound, which is required for an ASM, and more generally for testing of any arbitrary technique. Further, this method did not present a workable solution for initialising the segmentation, and as such required user intervention in many cases. There is evidence that AAMs outperform ASMs at matching shapes to patches of texture (and ASMs outperform AAMs at matching shapes to edges) [29]. However, AAMs require comparatively more training examples due to the large increase of data dimensions, and are notoriously slower.

Here we consider a study which proposes the use of an ASM with a Gamma Mixture Model (GMM) to segment heart cavities in 3D ultrasound [23]. The authors propose use of a GMM to classify pixels from the histogram of pixel distributions. The GMM is a 2-class one, which aims to differentiate hyper-echoic regions (class 1: muscle) from hypo-echoic regions (class 2: blood). This information is used to initialize a standard ASM of the whole heart, which then iteratively searches the volume for a refined segmentation. Then they use the standard ASM of individual heart cavities to optimize individual cavities and complete the segmentation. The authors evaluate their technique on just 20 cases and report an accuracy of 71%-90% (Dice similarity coefficient), and one failed segmentation. Two independent experts manually annotated the individual 2D images of the 3D volumes, identifying 5 distinct segments. Finally, they report a runtime of over 2 minutes for a single volume, so this is not a real-time solution. Their study further reinforces the case for applying ASMs; they are robust and accurate and they

maintain population-defined statistical anatomical shapes and relationships between shapes, which is important in a clinical sense. However, one of the main problems with the ASM – particularly for real-time applications – is its initialization; their approach is to use the pixel intensity histogram to initialize the ASM, yet the contrast between a heart cavity and heart muscle is very high, whereas the application to skeletal muscle concerns the muscle itself, and the internal structure (fascicles) is of a similar composition to the boundary (aponeurosis), with a similar level of acoustic impedance. The cervical muscles are densely packed together, separated almost always only by muscle boundaries which are not always visible; this rules out histogram-based methods for initializing an ASM.

Skeletal muscle is homogenous in texture, making differentiation of individual muscle segments within a group via histogram-based methods – particularly where there is large muscle-boundary dropout – inappropriate and unlikely. An example of failed muscle segmentation and differentiation is given here [30], where the authors present a convoluted multi-stage method based on speckle patch similarity (histograms); their technique 'lumps' the whole muscle group of only 3 muscles together (even the aponeuroses), as does the benchmark technique which they compare against [31]. Upon inspection of the example images they present, one can see that the entire muscle region (but not individual muscles in the region) is distinctly differentiable from the other segments by pixel intensity alone. Furthermore, their method is highly parametric, where parameters are chosen empirically with minimal justification and evidence of generalization. Their work has further reinforced the idea that histogram-based methods are inappropriate for segmentation of skeletal muscle within ultrasound. In contrast to the shortcomings of this work, the authors have demonstrated that in principal the texture of different structures within ultrasound can be informative for segmentation of distinct regions. The authors note that a supervised shape-regulated model might provide a better solution if the deformation of shapes due to probe pressure can be successfully modeled – they further add that this is a daunting task which is subject to inter-observer variations.

There are a number of papers on ultrasound segmentation based on the level-set methods [27], [32], [33]. The popular level-set can be a powerful segmentation technique, particularly when combined with shape/texture priors [26]. However, level-set methods are parametric and are difficult to tune, sensitive to initial conditions (initialization), can suffer from contour leakage due to segment boundary dropout, and computational complexity (runtime) increases with the number of objects (the N-level set paradigm) [34]. By far the most popular approach to initialization of a level-set region is to have an expert manually select the initial region, as in [26]. In contrast, and by far the most popular automatic initialization method is the use of classification techniques based on the pixel intensity histogram [27], [32] or local patches of texture [33]. Within the domain we are investigating, manual initialization is not possible for two

reasons: the first is the lack of expert certainty (subjectivity) on the location of each of the 10 muscle segments, and the second is the usefulness of a system which requires online manual labeling (i.e. not real-time analysis and visualization). The problems with use of histogram information have already been addressed in this paper, namely: lack of contrast between homogenous regions.

Machine learning, and in particular, deep learning methods [25], [35]–[37] are rapidly gaining a reputation as the state of the art in vision systems. We consider work on object detection within ultrasound, which demonstrates a remarkably robust solution to detecting features of fetal brains and computing measurements of fetal head structures, using a novel technique known as the Integrated Detection Network (IDN) [38]. This work proposes the use of an IDN for extracting features from training data, which are used in a sequential probabilistic object detection framework for hierarchical detection of object locations. The authors do not attempt to use this information for region extraction, but in principal their method could be used to initialize an ASM. They use a very large dataset of 13,000 annotations, which we note is not always feasible. The runtime of their technique is approximately 14.7s on a CPU, which we note is not suitable for real-time visualization. In their results they disregard the top 5% "large" errors, claiming that large errors were correlated with low probability from the IDN and can therefore be informative to an operator. The main drawbacks of this technique are the required amount of labeled data, and the runtime of the algorithm; for these reasons this technique has been ruled out, but in principal they have demonstrated that if one can construct a feature dictionary, one may be able to create an object detection system with which one can initialize an ASM.

Machine learning methods in general suffer from the problem of obtaining large volumes of annotated data, and in cases where obtaining annotated data is difficult – such as the cervical muscles within ultrasound – machine learning is not a preferred method. Difficulty obtaining labeled data is not uncommon within a research lab or a clinic. We can therefore rule out all techniques which use machine learning and are dependent on large volumes of data. One such example study uses a powerful machine learning method known as random forests (lots of bagged decision trees) for liver segmentation within ultrasound [25]. That study used over 940 annotated images, and since liver segmentation is a single segment problem, they therefore do not demonstrate application to multi-segment problems such as the cervical muscles. Further, they do not regulate their solution with anatomically correct boundary models, and we have therefore ruled out this technique.

In this paper, we propose a solution which is fast and effective for comparatively small datasets. We use the principals of the AAM to create a texture-to-shape dictionary, in which annotations (shapes) from a training set are warped into a mean shape space using nonlinear polynomial transformations. Those transformations are then applied to the associated texture. After texture/shape transformations we
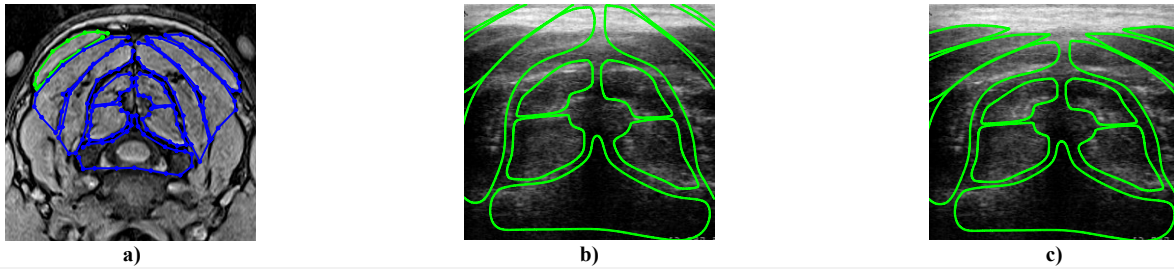
Fig. 2. Image annotation/registration process. In *a* an expert annotates the boundaries of 10 muscles and the spine within the MRI image which contained the image plane markers (green highlight informs the expert which muscle is currently being annotated). After interpolation and smoothing of the annotated points, in *b* the boundaries are manually registered – by the same expert – to the ultrasound image by rotation and translation (no scaling was permitted deliberately). Notice that the superficial boundaries do not agree with the texture. Finally, in *c* the nonlinear squashing function was then applied – by the same expert – to approximate the soft tissue deformation in the superficial muscles and complete the manual registration.

create a mean shape and texture. Then we create a principal component model of the shape alone, and use the first n component axes to symmetrically generate thousands of shapes. Then we warp the mean images from the mean shape space to every shape in the components space, resulting in thousands of texture templates to be used as a texture-to-shape dictionary. To segment a new image we propose an initial segmentation using this dictionary, and then we extract the profiles of the selected shape from the associated texture, and this is used to guide a heuristic search and refine the fitting according to the principals of the ASM. We predict that the initial segmentation using the dictionary is compatible with parallel processing and therefore has the potential to execute fast at runtime, and that it will provide a stable and accurate initial segmentation from which we can gauge the fit by taking some measure of the discrepancy between test image and dictionary texture. We also predict that since the initial segmentation will be close to the final segmentation, then only few iterations of the heuristic search need to be executed at full image resolution, which combined will allow accurate real-time segmentation. The following section describes our proposed method of segmentation.

III. METHODS

A. Data collection

These experiments were approved by the Research Ethics Committee of the Faculty of Science and Engineering, Manchester Metropolitan University (MMU). Participants gave (written) informed consent to these experiments, which conformed to the standards set by the latest revision of the Declaration of Helsinki. Experiments were performed at the Cognitive Motor Function laboratory, in the School of Healthcare Science, Healthcare Science Research Institute, MMU, Manchester, England.

Data were collected from 22 participants (age: 29.0 ± 6.6, male: 12, female: 10). Firstly, images of participants' posterior cervical muscles were acquired via transverse ultrasound. A T-shaped ultrasound probe was held to the back of the neck, just above C7 in the vertebra (see supplementary video). We chose a T-shaped probe for its suitability for taping to the skin and "wearing" during natural movement while imaging the cervical muscles, and not for its imaging quality. We emphasize that it is possible to choose alternative probes

which would provide superior image quality. Participants stood upright, while a single image was recorded. Then, 2 cod liver oil capsules were taped to the neck either side of the probe (using Transpore medical tape) to mark the image plane. The probe was removed, leaving the capsules in place, and an MRI scan (0.3T open MRI scanner, G-Scan, Esaote, Italy) was then obtained with participants lying supine on the scanning bed and their neck positioned central within a cervical imaging coil. Axial scans (Spin T1-weighted HF, matrix 512×512) were performed in a range from the upper jaw line to the clavicle, orthogonal to the spine, in 19 equidistant sections.

B. Image annotation

Following data collection, a human expert then annotated the MRI images, identifying boundaries of trapezius, splenius, semispinalis capitis, semispinalis cervices, multifidus, and the spine, bilaterally for all participants. The expert was permitted to annotate with as many points as necessary to capture the essential detail of each muscle. The expert was also instructed to begin annotation at the medial apex of each muscle, proceeding anti-clockwise around the muscle boundary. The annotation instructions were to facilitate point interpolation and statistical shape modelling. Following annotation, the expert then manually registered the MRI annotations to their corresponding ultrasound image via translations and rotations only (deliberately) – no scaling was permitted. In addition to translation and rotation, the expert was allowed to optimize the parameters of a contour squashing function, to approximate the soft tissue deformation and subsequently match the annotations to the affected muscles (see figure 2). This novel contribution was a necessary step, due to the pressure on the muscles from the probe during ultrasound imaging, which resulted in the muscles directly beneath the probe being squashed and deformed and a mismatch between the boundaries of MRI and US.

To approximate the soft tissue deformation, we first normalize the **x** component of a given shape,

$$\mathbf{x}' = c\left(\frac{\mathbf{x}}{w} - \frac{1}{2}\right) + \frac{1}{2}, \tag{1}$$

and then we proceed to apply an exponential squashing function parameterized by curvature and pressure coefficients,

$$\mathbf{y}' = f(\mathbf{x}, \mathbf{y}, c, s) = \left(\left(1 - \left(1 \oslash (1 + e^{-\mathbf{y}})\right)\right)\right) s \odot$$
$$\left(\mathbf{x}' \odot (1 - \mathbf{x}')\right), \tag{2}$$

where $\odot$ and $\oslash$ are Hadamard element-wise product and division operators respectively, $\mathbf{x}$ is a vector of x (horizontal) coordinates of some annotation, $\mathbf{y}$ is a vector of y coordinates of some annotation, w is the image (probe) width, c is a curvature parameter, and s is a pressure parameter. The result of equations 1 and 2 is a function which allows the expert to control the depth magnitude of contour displacement by tuning the s coefficient which simulates applied pressure, and by tuning the extent to which displacement causes increased curvature in the medial portion of the probe by tuning the c coefficient, which simulates the curvature of the back of a person's neck prior to probe contact. Using this approach we keep the metric anatomical information intact and express annotations in terms of a non-linear parametric function (see figure 3). We emphasize that modelling the soft tissue deformation may be a better approach to consider, though that is not the immediate objective of this study.

For the majority of cases all muscles were visible, permitting accurate annotation and registration in both MRI and ultrasound images. For a few cases the superficial muscles, particularly trapezius, were unidentifiable in the ultrasound image, and this was particularly the case for female participants. In the cases with invisible muscles, the expert aligned all other muscles as accurately as possible, inferring the positions of the trapezius and splenius muscles. To validate this process we performed inter-expert comparisons of annotations/registrations with an expert population of 3. Inter-expert registered annotations were compared in the ultrasound image plane using the performance metrics defined in the results section. The 3 experts annotated and registered muscle contours for all 22 participants in isolation of the other experts. Then we cross-validated the registered annotations by comparing experts 1 and 2, 1 and 3, and finally 2 and 3. Their ultrasound registrations were compared in the ultrasound plane only, using the Jaccard and Hausdorff metrics (i.e. no extrapolation, since the difficulty is in matching ultrasound gradients to contours, and not annotating the MR images).

C. Shape model

We follow the paradigm of the ASM to build a statistical model of shape and contour variance. Firstly, the entire dataset of 22 ultrasound images and their registered annotations was mirrored (image-central laterally). This was done to account for intrinsic asymmetries in the data, and it resulted in a total dataset of 44 annotated images. Then, all annotation points were interpolated per muscle such that every muscle consisted of 50 point annotations with the same anatomical origin and end point. Then, we built a principal components model of the annotation points. With this model we generated thousands of shapes equidistant in the component space. We use the "elbow method" to select only the most descriptive components, which resulted in retention of only 9 components, collectively encoding over 80% of the total variance. Later (see next
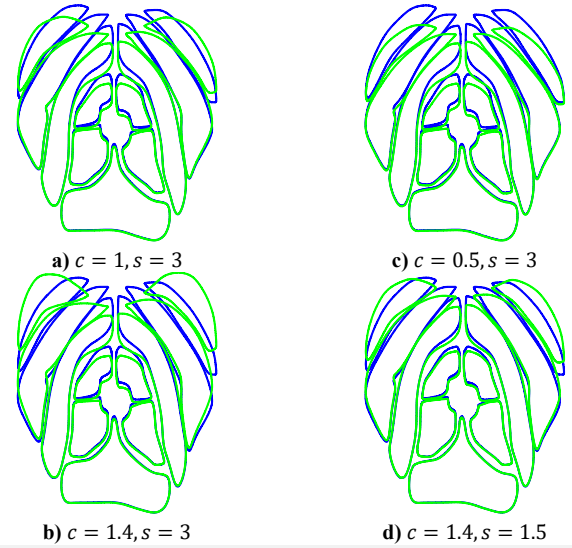

a) $c = 1, s = 3$ **c) $c = 0.5, s = 3$**

b) $c = 1.4, s = 3$ **d) $c = 1.4, s = 1.5$**

Fig. 3. **Parametric model of the soft tissue deformation.** Each graphic shows the original MRI annotation in blue, and the annotation in green after adjusting the parameters of the squashing function. The parameter values are shown below each graphic. Notice that decreasing s (**c** compared with **d**), decreases the amount of simulated pressure, and increasing c (**a** compared with **b**), accentuates curvature of the affected muscles. We found that the mean compression of the superficial muscles was 6.54mm $\pm$ 3.1mm, as characterised by mean function parameters c: 0.38 ± 0.21 and s: 5.4 ± 2.04. The coeficcient of variation (CV $= \frac{\sigma}{\mu}$) of the two parameters c: 0.55, and s: 0.37, is consistent with a low standard deviation (CV < 1), which indicates consistent behaviour of the function over the group. These findings are consistent with previous studies [16], where the authors empiracally show that compression can be anywhere between 7mm and 25mm.

section), we use 33 components (99% of the variance) to optimise the segmentation.

To generate shapes from the model we explore all combinations of the 9-dimensional component space by creating a 9-dimensional grid of n equidistant steps, within ± 2 standard deviations of each component. These points define component weightings which were used to generate distinct shapes. We chose number of bins per component $\mathbf{n} =$ [35 35 25 9 5 5 5 3 3 3] for each component [1: 9] respectively, such that n was as large as possible while the dictionary remained feasible in size. To expand on that point, the vector $\mathbf{n}$ represents how fine the mesh is in component space for each component; e.g. the first element of $\mathbf{n}$ means that we generate 35 points ± 2 standard deviations of the first component, and the second element means that we generate all combinations of first and second components in a 35×35 grid, and finally if we take the last element of $\mathbf{n}$ we generate all combinations of all 9 components in a 3^9 grid. At each point in all of the grids a hypothetical shape was generated by adding weighted components to the mean shape,

$$\mathbf{x} = \bar{\mathbf{x}} + P(\mathbf{b} \odot \sqrt{\boldsymbol{\lambda}}), \tag{3}$$

where $\mathbf{x}$ is a generated shape, $\bar{\mathbf{x}}$ is the mean shape vector, P is an len($\mathbf{x}$)×9 matrix of shape components, $\mathbf{b}$ is a vector of length len($\mathbf{x}$) which contains the component weightings at a single point in the 9-dimensional components space, and $\boldsymbol{\lambda}$ is a vector of eigenvalues. This was done for all weighting
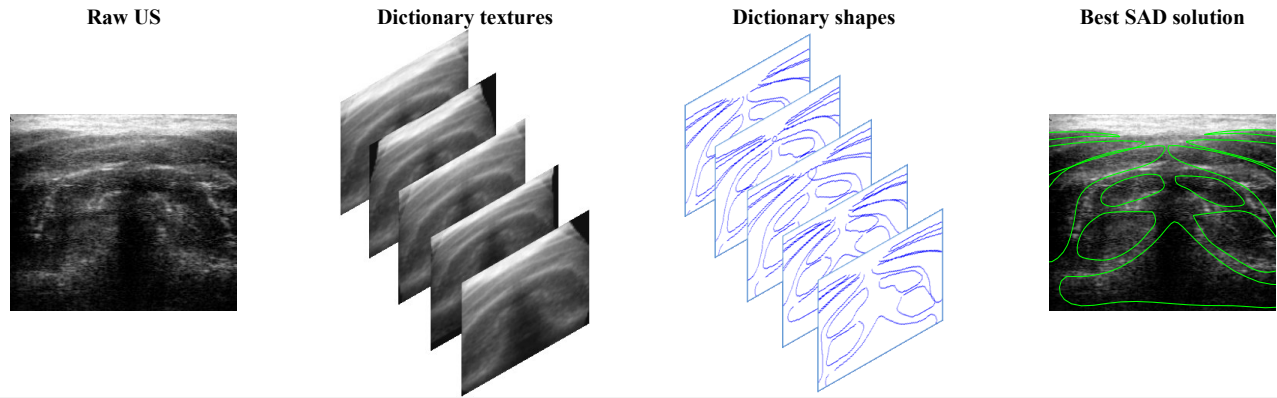
| Raw US | Dictionary textures | Dictionary shapes | Best SAD solution |

Fig. 4. Dictionary segmentation pipeline. From let to right, the raw ultrasound image is acquired, then the dictionary of textures is compared one by one on the CPU (or all in parallel on multiple cores, or a GPU) by taking the average of the sum of absolute pixel differences (SAD). The index of the best matching texture (lowest SAD score) is used to extract the contours which were used to generate that texture, which are subsequently used to segment the raw image.

combinations, which resulted in a database of 70,619 generated shapes.

D. Texture model

To create a texture model, the mean texture was required. Intuitively the textures of each image in the set are not in a comparable space (i.e. the muscles are in different locations for each person). In order to create a mean texture, we optimized fourth order polynomials to map each shape to the mean shape,

$$\mathbf{s} \approx \mathbf{\mu} = [\mathbf{x_1 y_1 x_2 y_2 \dots x_m y_m}] \approx \mathbf{p_0} + \sum_{j=1}^{j=n} \mathbf{p}_j \mathbf{\mu}_i^{n-(j-1)}, \quad (4)$$

where $\mathbf{s}$ is a shape vector of length $2m$, $\mathbf{\mu}$ is the mean shape vector of length $2m$, $\mathbf{p}$ is a vector of polynomial terms, and n is the order of the polynomial. Equation 3 provides the mapping of $\mathbf{\mu}$ to $\mathbf{s}$ via the polynomial coefficients $\mathbf{p}$. We can then use the function to map pixel locations of a shape to the mean texture space. We can solve for the polynomial coefficients analytically by redefining mean shape in fourth order polynomial form,

$$X = [\mathbf{\mu^4}\ \mathbf{\mu^3}\ \mathbf{\mu^2}\ \mathbf{\mu}\ 1]. \quad (5)$$

In this form a linear solver can find the vector of polynomial coefficients used to transform X to $\mathbf{s}$,

$$\mathbf{p} = (X^T X)^{-1} X \mathbf{s}^T. \quad (6)$$

After solving for the coefficients, we then transformed all pixels in the individual textures to a comparable space, using bilinear interpolation to fill gaps in the transformed texture, and then we computed the mean texture from all of the transformed images.

Following construction of the mean texture, we repeat the same process by optimizing fourth order polynomials to map the mean shape to every shape in the shape database. This resulted in a texture-to-shape dictionary which describes where the muscle boundaries are in those textures. Each texture in the dictionary was interpolated to an 88×88 image matrix (approximately one fifth of the original image size) using bilinear interpolation. We retain the full resolution image for the refined fitting process (section III. F). The

reason for down-sampling was to reduce computation time during template matching (see next subsection), and to add in some feature-translation invariance.

E. Dictionary segmentation

After construction of a texture-to-shape dictionary, segmentation of a new *out of sample* image can now be achieved by searching the dictionary for the texture which minimizes the sum of absolute differences (SAD) between a texture and a new image, as is standard for lighting-invariant template matching. Once the best matching texture is known, the shape associated with that texture can be extracted and used as an initialisation for the refined fitting procedure. We parallelised the search (partitioning the data) using all 4 cores (8 threads with hyper-threading) of a CPU to compute the SAD measure in parallel, storing the results in a list, where the indexes of the list correspond to the index in the texture-to-shape dictionary. We acknowledge that further gains in computation speed could be achieved by using multiple CPUs, and Graphics Processing Units (GPU), however in that regard this study only aims to demonstrate the increase in computation speed as a result of parallelization. After the SAD metric is computed for all textures in the dictionary, we then search the list for the minimum, which gives the best texture and consequently the best segmentation, from which we can use the component model to refine the segmentation (see figure 4).

F. Refined fitting segmentation

The initial segmentation not only provides a reliable estimate with regards to the shape and location of each muscle, it also gives the warped mean texture. We extract contour profile intensities from the warped texture, and then we use the statistical model to refine the segmentation with additional components and a heuristic search. We slightly modify the ASM search routine which is described in detail here [3]. In brief, we define a set of image intensity search profiles from the test image, perpendicular to the muscle contours at each of the 50 points in each muscle, 20 pixels (which is larger than the standard ASM) above and below the contour. We also define a set of intensity profiles for the mean shape using the warped mean texture from the initial segmentation, half the length of the search profiles. The reason for using the warped texture and larger profiles is to use as
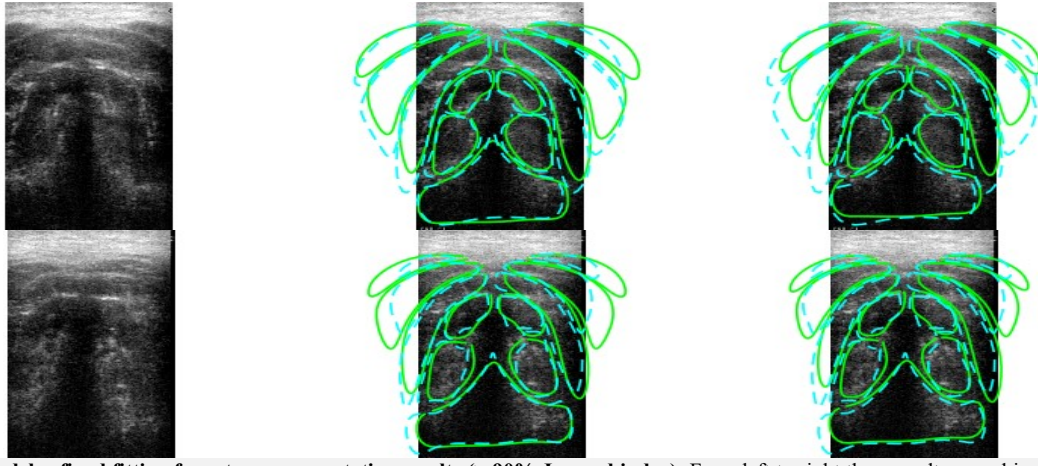
Fig. 5. Shape model refined fitting for a two representative results (> 90% Jaccard index). From left to right the raw ultrasound is shown, followed by the dictionary segmentation result, followed by the refined shape model fitting result. Solid green colour shows the automatic segmentation, and dashed cyan colour shows the expert annotation/segmentation. These results empirically demonstrate the accuracy of the proposed tool. Importantly, even without the refined fitting step, the live segmentation is accurate, even giving a good approximation of the muscle shapes and boundaries beyond the lateral ultrasound plane. The refined fitting increases the accuracy and subsequently improves the extrapolation of the muscles beyond the image, which can be used to visualise the muscle structure of a patient, beyond the visible features in the ultrasound plane.

much of the texture as possible, since edges and features are not typically high contrast in ultrasound images. Then, for each contour point we move the mean profile along the search profile (from end to end) and compute the sum of squared differences (SSD) between intensities of the normalised test image and the intensities of the corresponding static profile in the normalised dictionary image. We take the moved contour point at the minimum SSD and store it in a new shape vector. When we have computed the coordinates which minimise SSD for all contour points we have a new shape vector,

$$\mathbf{s'} = [\mathbf{x_1 y_1 x_2 y_2 \dots x_m x_m}]. \tag{7}$$

We then project this shape into components space,

$$\mathbf{c} = \mathrm{P}\omega(\mathbf{s'} - \boldsymbol{\mu}) \odot \sqrt{\boldsymbol{\lambda}}, \tag{8}$$

where ω is a gradient term (which we empirically set to 0.1), and finally we re-project the component vector $\mathbf{c}$, back into shape space,

$$\mathbf{s} = \boldsymbol{\mu} + \mathbf{c}\mathrm{P}^{\mathrm{T}}. \tag{9}$$

The rationale for using SSD rather than SAD here, is that during dictionary segmentation, the whole texture is considered and thus all pixels should be considered with equal weight (which empirically was the case), however the for the refined fit, we consider the boundary (edge) information, therefore normalisation adds weight to these components for focussed comparison.

For this procedure we retained 33 components (99% of variance), therefore the process of projection into components space and re-projection into shape space the model will discard spurious contour movements (i.e. contour shapes that are not statistically representative). We do this iteratively for some maximum number of iterations (200). To prevent new shape vectors wandering away from the statistical norm defined by the shape model, we regularised the component projection such that each vector lies within 3 standard

deviations of the model. All parameters were chosen empirically. For each iteration, after projection of the new shape vector into components space, we compute the following sum,

$$v = \Sigma \big(|\mathbf{c}| \oslash 3\sqrt{\boldsymbol{\lambda}}\big)^2, \tag{10}$$

and where $v > 1$ we regularise the vector $\mathbf{c}$ with,

$$\mathbf{c'} = \mathbf{c}\frac{1}{v}. \tag{11}$$

This regularisation forces the norm of any generated component to lie within a hyper-ellipse, 3 standard deviations from the model, and therefore will limit the optimisation to explore only valid points in the manifold. After 200 iterations the refined segmentation is complete (see figure 5).

G. Visualisation and real-time analysis of muscle features

After segmentation of the muscles, we present live on-screen visualisation of the whole cervical muscle structure, and quantitative and visual analysis of the components of the segmented muscles. On the screen we show the raw ultrasound with the segmented muscle boundaries overlaid, showing the extrapolated muscle boundaries beyond the ultrasound plane (see figure 6). Extrapolation is an intrinsic property of the ASM, no explicit procedure is needed for this – the component model simply re-projects the hypothetical boundaries based on the gradients present in the ultrasound image (i.e. its best approximation at what the whole muscle structure looks like, given the image).

Further to visualisation of the whole muscle structure, since the segmentation is inherently composed of re-projected components, and since each component uniquely captures a visual feature (e.g. asymmetry, relative muscle size/orientation, whole muscle group size) we present normalised statistics in the form of a bar chart on the screen. We first normalise each shape component,
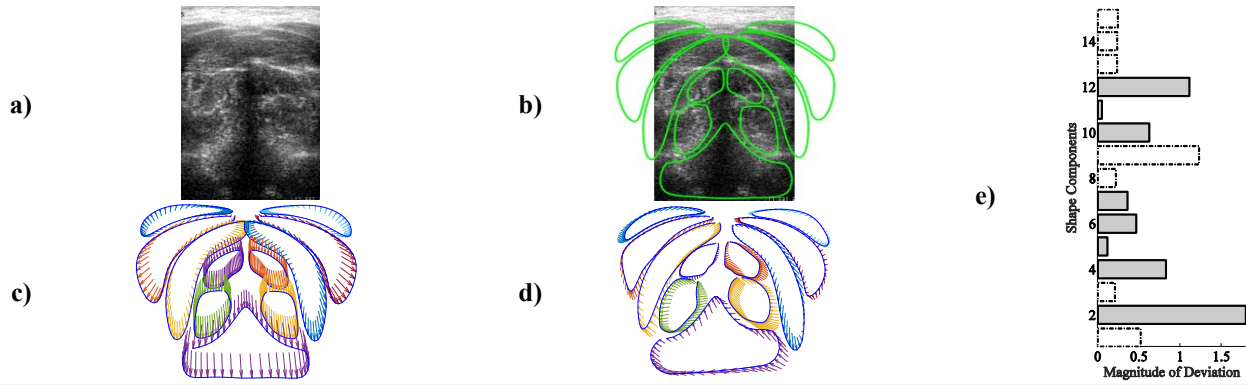
Fig. 6. Muscle feature segmentation visualisation. This figure shows how segmentation can provide visualisation of identified component features, and also how those components can be represented quantitatively. The raw ultrasound is given in *a*, and in *b* the refined (post) segmentation is given. *c* and *d* are a vector visualisation of the two main components of variance (2 and 9, respectively) – where the solid blue contours represent the magnitude of deviation from the mean shape, and the arrows represent angle of deviation (i.e. each arrow originates at the mean shape, and ends at the component-projected shape – highlighting the major differences of a person's neck from a population). The bar plot (*e*) shows the magnitude of the first 15 components, where the colour represents the sign (white dashed = negative). With this visualisation a clinician and/or a patient can visually interpret the features detected by the segmentation, which can recorded for group and longitudinal studies.

$$\mathbf{c}' = \frac{\mathbf{c}}{\sqrt{\lambda}}, \tag{12}$$

where $\mathbf{c}'$ is a vector of standard deviations representative of the proportional magnitude of each component. We can now express $\mathbf{c}'$ as a bar chart visualisation, and record/interpret its meaning in real time. Further to the quantitative measure, we provide an interface for visually interpreting each component. We take the top n proportional components and display them on the screen as a shape with arrows ending at each contour point, originating at each corresponding mean point. This allows visual identification, recording and characterisation of features present in a person with CD or some other condition.

IV. RESULTS

To consider the performance of our technique we consider the two features which would be most useful in a clinical setting, muscle contour detection and muscle cross-sectional area approximation. While these two features are intrinsically linked, it is important to know the extent to which they can be measured. To measure the accuracy of muscle contour detection, we used the Hausdorff distance, which is defined as,

$$H(\mathbf{s}, \mathbf{h}) = max \left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right), \tag{13}$$

where,

$$\mathbf{a}_i = min(\mathbf{s}_i - \mathbf{h}),$$

$$\mathbf{b}_i = min(\mathbf{h}_i - \mathbf{s}),$$

and $\mathbf{s}$ is the ground truth (expert annotation), and $\mathbf{h}$ is the model segmentation. The Husdorff distance gives the furthest distance (i.e. the greatest deviation) of all points between the ground truth and the segmentation, which we express in millimetres.

To measure the accuracy of cross-sectional muscle area measurement we use the Jaccard index, which is defined as the intersection between ground truth and model segmentation,

divided by the union of the ground truth and model segmentation. This Jaccard index is computed for individual muscle segments to ensure that the correct regions are being classified by the model,

$$J(\mathbf{s}, \mathbf{h}) = \frac{|\mathbf{s} \cap \mathbf{h}|}{|\mathbf{s} \cup \mathbf{h}|}. \tag{14}$$

The Jaccard index gives the proportion of the combined areas of 2 arbitrary polygons (a ground truth muscle, and a model segmentation muscle) that is overlapping, which was computed for all 11 segments, averaged, and then expressed as a percentage. We also give the Dice similarity coefficient, is less discriminating than the Jaccard index but due to its popularity is given to allow comparisons with other methods,

$$D(\mathbf{s}, \mathbf{h}) = \frac{2|\mathbf{s} \cap \mathbf{h}|}{|\mathbf{s}| + |\mathbf{h}|}, \tag{15}$$

In addition to measuring segmentation accuracy, we also measured the time taken to segment a single image using 4 cores on a CPU (i7-4720HQ), where all code is original and was written in C++. Leave one out (LOO) cross validation was used for all participants' data, where any given test case contained 2 ultrasound images (original and mirrored), and 42 (21 originals and 21 mirrored) training images for the construction of the shape and texture models.

Results showed high Jaccard indices and Dice coefficients, and a low Hausdorff distance for the majority of test cases, demonstrating consistent and accurate performance for both measures. In the best cases, over 99% of the muscle areas were in agreement, and the largest deviation of all points in the annotation and the segmentation was less than 4mm (meaning that all points deviated at most by 4mm). In the worst cases we see agreements as low as 33% and 50% for Jaccard index and Dice coefficient respectively, and a maximum contour deviation of 20mm according to the Hausdorff distance. Positively, these cases represent outliers in results terms, and their failure can be predicted from their SAD values. The trickier cases (x > 60% < 80% Jaccard)

Table. 1. Table of results. This table summarises the accuracy and timing metrics for 200 iterations of the segmentation algorithm we have proposed and developed. It also summarises the results and timing for 25 iterations at 2 scales (50 total) of the standard **AAM** [39]. Results are given for the dictionary segmentation (**pre**) and the refined fitting (**post**), for both: segmentation within the ultrasound image (**US**) and extrapolation of whole muscle structure beyond the ultrasound image (**extrap**). Results show that our method is accurate to over 86%, with the best result recording over 99% accuracy according to the Jaccard index and the Dice similarity coefficient, and maximum contour discrepancy of just 3.8mm according to the Hausdorff distance. *Notice that our method is an improvement over the standard AAM, particularly when predicting shape* (**extrap**) *beyond the image texture. We also demonstrate the vast difference in segmentation runtime between our method and the AAM*.* Additionally we give inter-**expert** comparisons, which empirically demonstrate within the ultrasound: subjectivity, increased variability, and less agreement than the automated methods. Further to analysis of performance metrics, the timing measurements show that our tool can segment a new image approximately 2-3 fps.

		μ ± σ	min	max
Jaccard index	**pre - US**	85.52% ± 17.36%	33.77%	99.45%
	post - US	86.67% ± 15.87%	32.43%	99.55%
	AAM - US	84.09% ± 14.67%	40.11%	96.26%
	expert - US	79.16% ± 19.70%	37.68%	99.32%
	expert - MRI	96.84% ± 0.02%	88.34%	99.99%
	pre - extrap	86.64% ± 13.76%	46.48%	98.21%
	post - extrap	87.19% ± 13.04%	43.85%	98.77%
	AAM - extrap	82.98% ± 13.07%	41.36%	95.35%
Dice similarity coefficient	**pre - US**	91.08% ± 12.24%	50.49%	99.72%
	post - US	91.91% ± 11.51%	48.98%	99.78%
	AAM - US	90.55% ± 10.43%	57.25%	98.10%
	expert - US	86.87% ± 13.88%	54.73%	99.66%
	expert - MRI	98.38% ± 0.01%	93.81%	99.99%
	pre - extrap	92.19% ± 9.06%	63.47%	99.10%
	post - extrap	92.56% ± 8.80%	60.96%	99.38%
	AAM - extrap	90.06% ± 9.20%	58.52%	97.62%
Hausdorff distance	**pre - US**	7.00mm ± 2.76mm	4.24mm	15.01mm
	post - US	6.47mm ± 3.03mm	3.84mm	16.06mm
	AAM - US	7.65mm ± 2.73mm	4.21mm	13.71mm
	expert - US	7.69mm ± 3.61mm	3.43mm	20.98mm
	expert - MRI	2.45mm ± 0.81mm	0.81mm	4.77mm
	pre - extrap	8.75mm ± 2.79mm	4.03mm	16.95mm
	post - extrap	8.32mm ± 2.89mm	4.72mm	17.33mm
	AAM - extrap	11.21mm ± 3.08mm	6.11mm	17.27mm
Timing	**pre**	0.203s ± 0.002s	0.203s	0.218s
	post	0.130s ± 0.007s	0.125s	0.141s
	AAM*	72.87s ± 2.14s	69.98s	81.63s
SAD	**pre - US**	11.47px ± 0.97px	10.11px	13.34px

*The AAM was downloaded for use with the MATLAB interface, which we acknowledge would not be as efficient as the same algorithm written in c/c++.

represent cases where either the ultrasound image was sub-optimal, or not all muscles were not visible, or the physiology of the muscles as appeared in the ultrasound image was unusual; for example one of the participants' spine was barely visible above the lower boundary of the image (i.e. a comparatively large neck outlier). To address these cases one might suggest increasing the number of standard deviations within which the ASM can search the image, however, we found that 3 standard deviations was a good regularizer, and increasing this value empirically produced some over-fitting, even overlap in the muscle segments. One other suggestion may be to use a larger population and develop multiple models, which the operator could rapidly switch between at runtime. This is not dissimilar to (f)MRI brain segmentation, or facial segmentation, in which different models are
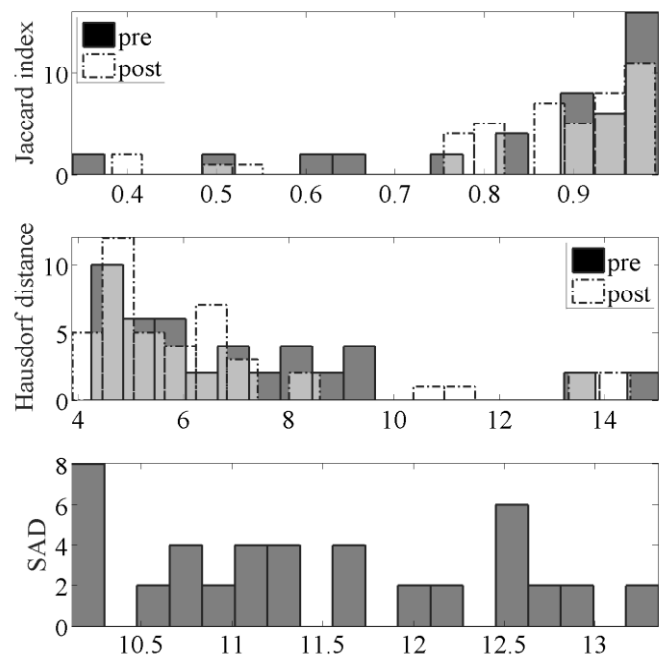

Fig. 7. Histogram of results. These histograms summarise the 3 performance metrics, Jaccard, Hausdorff and SAD, for our method. The majority of results recorded over 86% accuracy (Jaccard index). Similarly, the majority of results recorded below 7mm (Hausdorff distance). Correlation of SAD and Jaccard index ($r = -0.4397, p = 0.0028$), and the Hausdorff distance ($r = 0.3885, p = 0.0092$), revealed a strong relationship between the 3 metrics.

employed to segment different ethnical groups.

Finally, we evaluated the annotation process using the above named metrics to assess the agreement between an expert population of 3. Comparisons empirically demonstrate that our image annotation protocol was robust and consistent, showing high accuracy and agreement between operators, although there were disagreements occurring in challenging ultrasound images. Results also showed that the agreement between operators was slightly lower than our automated technique for which we give probable reasons in the discussion section.

V. DISCUSSION

A real-time cervical muscle segmentation and analysis tool was successfully developed for use with standard B-mode medical ultrasound machines. This study has opened up a new domain in medical imaging research (real-time muscle segmentation and analysis via ultrasound), which has particular relevance to many applications including cervical dystonia, concussion, whiplash, and many other neck/back pain medical problems. Furthermore, segmentation is a precursor step enabling normalisation of individual muscles for statistical and machine learning approaches. Hence we anticipate the results achieved by this approach will stimulate widespread development within this domain.

To deliver our tool, we have developed a new methodological approach to modelling and segmenting any arbitrary semi-rigid structure via ultrasound. Our methods presented here allow operators to confidently annotate ultrasound images using a novel multimodal (MRI-ultrasound)
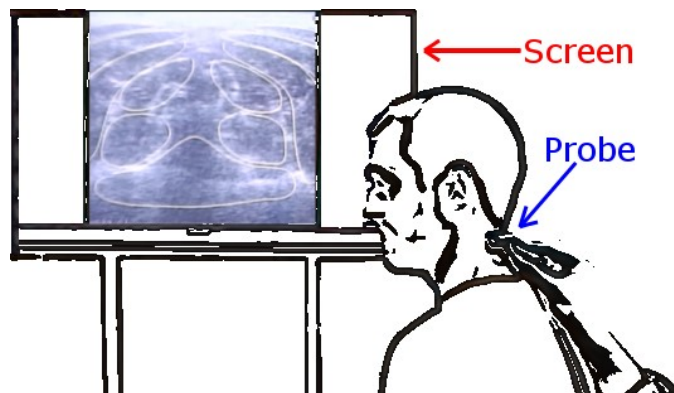
Fig. 8. Lab setup. In the figure a participant stands upright, with the ultrasound probe held against the posterior neck, and a screen opposite. The screen shows segmentation overlaid on the raw ultrasound in real time. **See supplementary video**; all participants in the video were not included in any of the methodological/modelling stages presented in this study; i.e. no MRI or ultrasound were acquired or labelled from these participants, and thus the video is representative of real-world performance. The video shows some ability to handle changes in head orientation/position, although of course the accuracy of this currently cannot be quantified; i.e. no ground truth labels.

manual registration method. We have also developed a novel method for approximating the soft tissue deformation due to probe contact. We acknowledge that there are techniques for imaging the muscle system without applying pressure to the structure being imaged; however the tool presented here was developed in such a way that it could be easily deployed and used with minimal complications in a clinical setting, and is generally applicable without restriction – there is nothing preventing the same methods being used where there is no tissue deformation.

On the polynomial mapping of textures from shapes (section III. D.), the authors would like to note that various other methods were attempted among various levels of polynomials. We note that spline-based and triangulation methods warped the textures to unrecognisable forms with huge holes in the texture (over-fitting), which we attribute to the vast differences between matched shapes. In contrast, we tried polynomial orders 2 through 4 which all produced desirable results, where greater than 4 started to give too much over-fitting again. We settled on order 4 polynomials to maximise the fitting, while simultaneously preserving regions through minimal warping.

The tool was rigorously evaluated against a gold standard, popular and powerful segmentation method (AAM), using LOO cross validation under two conditions; its ability to extract detailed contour (shape) information (Hausdorff), and its ability to localize and identify the area of each muscle within the texture (Jaccard/Dice). Results showed that the tool is accurate and consistent at over 86% accuracy on average, with the best cases recording > 99% accuracy (see table 1, and figure 7), compared with the AAM which reported consistent accuracy over 84%, with the best cases only recording > 96%. We draw your attention to figure 7, which shows how the initial segmentation (pre) performs well overall at identifying the correct muscle cross-sectional areas within the texture (Jaccard), and how the refined fitting (post) worsens slightly for the upper quartile of cases according to the same metric. We also draw your attention to the Hausdorff

results in the same figure, which conversely show an improvement after refined fitting for the upper quartile of cases. This was predicted in our introduction and supported by the literature [29], where we expected AAM-type techniques to perform best at matching texture areas than visible contours. The initial dictionary segmentation performs well at matching muscle areas, but then the refined fitting allows capture of the essential detail of the muscle boundary, while degrading slightly the quality of the overall texture match. This phenomenon can be explained by annotation error (see table 1 for inter-expert disagreement), where the whole shape of the muscle boundaries is anatomically correct and validated due to the rigid registration process, yet somehow the general approximation of the location of muscle areas is slightly mismatched. This would mean that after refined fitting, the overall Hausdorff distance should decrease, and the Jaccard index should increase as the muscle shape is corrected to fit, and the area has some constant translational/rotational error. It may also be that our compression function is not entirely physiological, which would also bring about some translational error. These findings validate our method and justification for our annotation protocol.

Inter-operator comparisons of the annotation process proved comparatively robust, while showing slightly increased variability and slightly lower agreement between experts than the automated methods. We note that the largest disagreements between experts occurred generally in the poorest quality ultrasound image data, and were also generally the same images which the tool performed poorest on. However, our tool provides an interface for regulating quality of data via operator feedback, since we have shown that it is more consistent and can self-report the SAD metric during live segmentation – which was related to the accuracy of the segmentation (see figure 7). We can also apply a threshold to the SAD metric and discard segmentations below that threshold (see supplementary video). With respect to the poor agreements between experts, we note that there are fewer degrees of freedom during the annotation/registration process; by definition the method has more flexibility to transform and optimise muscle boundaries using a population-generated model, where the experts were not permitted (deliberately, to retain the anatomical muscle boundary relationships of individual participants). A combination of feedback of muscle boundary graphical overlays and a self-reporting error metric may lead to regulation of data quality and improved segmentation results.

We also demonstrated how our tool would be used to extrapolate and visualise the wider cervical muscle structure beyond the ultrasound image plane, and quantitatively analyse whole muscle features detected by the segmentation procedure, in real time. Live visualisation of the whole cervical muscle structure can help inform people with CD about their condition and how it affects their anatomy, while also giving the clinician a detailed view of the location of any features which are characteristic of dystonia. Live visualisation of analysis allows interpretation of a person's condition, which may be recorded and studied longitudinally.

The proposed application of this tool is to facilitate treatment monitoring, and diagnosis of CD, although there are many conditions which would benefit; whiplash, burners,

cervical radiculopathy, spondylosis, stenosis, and repetitive motion disorders, to name a few. With respect to CD, the current gold standard treatment involves inserting needles directly through the muscle layers into the target muscle, and injecting botulinum toxin. However, there are numerous problems with this procedure:

1) There is often difficulty in identifying the affected muscle(s), and often the only solution is to feel around the neck for tight (active) muscles, and consequently the deep muscles cannot be explored or treated because they cannot be identified as symptomatic, and they are too close to the spine to risk injection.

2) There is also difficulty in validating whether the correct muscles have been targeted by the botulinum needle. Botulinum takes approximately a fortnight to take effect; therefore feedback about the efficacy of any single injection cannot be obtained in real-time. The use of needle EMG or feeling for active muscles is not applicable here.

3) There is currently no quantitative way to monitor the progression of CD, or the effect of any treatment in individual muscles or muscle groups, whether that is Deep Brain Stimulation (DBS), Parkinson's drugs, muscle relaxants, or botulinum toxin. EMG can provide session-subjective information about muscle activation, but it is invasive and time consuming, and cannot measure muscle wasting, regeneration, or asymmetry about the spine.

Our tool avoids the need for MR imaging, which would otherwise prevent people who meet exclusion criteria from benefiting from its use. Our tool can provide clinicians with real-time identification of every layer of muscle, allowing image optimization by on-screen feedback. With an optimised image, there are existing techniques [40] which would allow tracking of the needle end point in the ultrasound image plane, promoting confident placement of the needle in even the deepest muscles. Our tool provides quantifiable data on cross-sectional muscle area and shape, and asymmetries. Data can be recorded with confidence and can be used to inform future treatments. We propose that the methods presented here allow potential modelling of CD and other neuromuscular disorders using pattern recognition methods (e.g. [41]), which may be incorporated into the real-time feedback to help identify suspected symptomatic muscles, informing decisions to target identified muscles. We highlight this as an important research area which needs to be explored, which would contribute to the development of new treatments for CD.

We propose this tool has more general application. The structure of muscles along the spine is consistent between cervical, thoracic and lumber regions. Hence all the methods are immediately applicable to segmentation of deep lumber muscles in application to chronic low back pain. In principle the methods are applicable to complex muscles structures within the upper and lower limb. With respect to image analysis, deep learning shows much promise for the extraction of features characterising muscle structure and muscle action for medical diagnosis. However, deep learning is applied to individual muscles requires normalisation to a standard form and we propose this tool is a first stage in that process [42].

VI. Conclusions

In summary, a tool has been developed which allows real-time segmentation and identification of the posterior cervical muscles. This study has validated the efficacy of the tool for quantifying muscle area and shape parameters with minimal error, as validated by high-resolution expert annotations of MR images. We have demonstrated that our technique is effective with only a small amount of training data, which is common in many medical imaging problems. This study also presents a generalized methodology for reproducing these results using an annotation and our novel multimodal manual registration method, to allow modelling and segmentation of generally any muscle group in the body that is visible via ultrasound. We have tested and validated the hypothesis set out in our introduction, that "ultrasound provides sufficient information to facilitate an automatically accurate segmentation of the neck muscles". We have also tested and validated the hypothesis set out in our introduction, that "ultrasound contains the information required to locate an annotation defined in the MRI domain". Since our tool is non-invasive and has no exclusion criteria, it is suitable for clinical environments for the monitoring and guidance of CD treatment. Further work is required to clinically validate this tool on a large CD population, and to develop the tool to allow automatic identification of, and quantification of severity, of symptomatic patients and individual patients' muscles. Importantly, we have given a methodology for applying this technique to any arbitrary muscle group and/or disease. There is currently a lack of literature on modelling CD in terms of measurable muscle parameters; therefore we conclude that the tool presented in this study would promote inexpensive acquisition of large volumes of quantitative data on symptomatic muscles, expanding the scope of understanding of the pathophysiology of CD.

References

[1] K. Javanshir, M. Amiri, M. A. Mohseni-Bandpei, A. Rezasoltani, and C. Fernández-De-Las-Peñas, "Ultrasonography of the cervical muscles: A critical review of the literature," *Journal of Manipulative and Physiological Therapeutics*, vol. 33, no. 8. pp. 630–637, 2010.

[2] J. a. Noble and D. Boukerroui, "Ultrasound image segmentation: a survey," *IEEE Trans. Med. Imaging*, vol. 25, no. 8, pp. 987–1010, 2006.

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, no. 1. pp. 38–59, 1995.

[4] T. Cootes, "Active Appearance Models," *Pattern Anal. ...*, vol. 23, no. 6, pp. 681–685, 2001.

[5] L. K. Kamibayashi and F. J. Richmond, "Morphometry of human neck muscles.," *Spine (Phila. Pa. 1976).*, vol. 23, no. 12, pp. 1314–1323, 1998.

[6] G. Rankin, M. Stokes, and D. J. Newham, "Size and shape of the posterior neck muscles measured by ultrasound imaging: Normal values in males and females of different ages," *Man. Ther.*, vol. 10, no. 2, pp. 108–115, 2005.

[7] A. N. Vasavada, B. W. Peterson, and S. L. Delp, "Three-dimensional spatial tuning of neck muscle activation in humans," *Exp. Brain Res.*, vol. 147, no. 4, pp. 437–448, 2002.

[8] D. Hoy, L. March, P. Brooks, F. Blyth, A. Woolf, C. Bain, G. Williams, E. Smith, T. Vos, J. Barendregt, C. Murray, R. Burstein, and R. Buchbinder, "The global burden of low back pain: estimates from the Global Burden of Disease 2010 study," *Ann. Rheum. Dis.*,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2016.2623819, IEEE Transactions on Medical Imaging

> REPLACE THIS LINE WITH YOUR PAPER IDENTIFICATION NUMBER (DOUBLE-CLICK HERE TO EDIT) <   13

vol. 73, no. 6, pp. 968–974, 2014.

[9]   V. C. W. Hoe, D. M. Urquhart, H. L. Kelsall, and M. R. Sim, "Ergonomic design and training for preventing work-related musculoskeletal disorders of the upper limb and neck in adults.," *Cochrane Database Syst Rev*, vol. 8, p. CD008570, 2012.

[10]   V. Ap, B. Sma, A. Burdorf, S. Sm, D. V. Hcw, and K. Bw, "Conservative interventions for treating work-related complaints of the arm , neck or shoulder in adults (Review)," no. 12, 2013.

[11]   C. Lungu, A. W. Tarulli, D. Tarsy, P. Mongiovi, V. G. Vanderhorst, and S. B. Rutkove, "Quantifying muscle asymmetries in cervical dystonia with electrical impedance: A preliminary assessment," *Clin. Neurophysiol.*, vol. 122, no. 5, pp. 1027–1031, 2011.

[12]   B. L. Suderman, B. Krishnamoorthy, and A. N. Vasavada, "Neck muscle paths and moment arms are significantly affected by wrapping surface parameters.," *Comput. Methods Biomech. Biomed. Engin.*, vol. 15, no. 7, pp. 735–44, 2012.

[13]   V. Esat and M. Acar, "Viscoelastic finite element analysis of the cervical intervertebral discs in conjunction with a multi-body dynamic model of the human head and neck," *Proc. Inst. Mech. Eng. Part H J. Eng. Med.*, vol. 223, no. 2, pp. 249–262, 2009.

[14]   C. M. Sommerich, S. M. B. Joines, V. Hermans, and S. D. Moon, "Use of surface electromyography to estimate neck muscle activity," *J. Electromyogr. Kinesiol.*, vol. 10, no. 6, pp. 377–398, 2000.

[15]   M. A. M. Benhamou, M. Revel, and C. Vallee, "Surface electrodes are not appropriate to record selective myoelectric activity of splenius capitis muscle in humans," *Exp. Brain Res.*, vol. 105, no. 3, pp. 432–438, 1995.

[16]   M. Kramer, I. Schmid, S. Sander, J. Högel, R. Eisele, L. Kinzl, and E. Hartwig, "Guidelines for the intramuscular positioning of EMG electrodes in the semispinalis capitis and cervicis muscles," *J. Electromyogr. Kinesiol.*, vol. 13, no. 3, pp. 289–295, 2003.

[17]   K. Türker, "Electromyography: some methodological problems and issues," *Phys. Ther.*, vol. 73, no. 10, pp. 698–710, 1993.

[18]   J. P. Lee, C. L. Wang, Y. W. Shau, and S. F. Wang, "Measurement of cervical multifidus contraction pattern with ultrasound imaging," *J. Electromyogr. Kinesiol.*, vol. 19, no. 3, pp. 391–397, 2009.

[19]   J. P. Lee, W. Y. I. Tseng, Y. W. Shau, C. L. Wang, H. K. Wang, and S. F. Wang, "Measurement of segmental cervical multifidus contraction by ultrasonography in asymptomatic adults," *Man. Ther.*, vol. 12, no. 3, pp. 286–294, 2007.

[20]   B. Cagnie, N. Dickx, I. Peeters, J. Tuytens, E. Achten, D. Cambier, and L. Danneels, "The use of functional MRI to evaluate cervical flexor activity during different cervical flexion exercises.," *J. Appl. Physiol.*, vol. 104, no. 1, pp. 230–235, 2008.

[21]   A. C. Dupont, E. E. Sauerbrei, P. V. Fenton, P. C. Shragge, G. E. Loeb, and F. J. R. Richmond, "Real-time sonography to estimate muscle thickness: Comparison with MRI and CT," *J. Clin. Ultrasound*, vol. 29, no. 4, pp. 230–236, 2001.

[22]   E. Kristjansson, "Reliability of ultrasonography for the cervical multifidus muscle in asymptomatic and symptomatic subjects," *Man. Ther.*, vol. 9, no. 2, pp. 83–88, 2004.

[23]   A. Haak, G. Vegas-Sanchez-Ferrero, H. W. Mulder, B. Ren, H. A. Kirišli, C. Metz, G. Van Burken, M. Van Stralen, J. P. W. Pluim, A. F. W. Van Der Steen, T. Van Walsum, and J. G. Bosch, "Segmentation of multiple heart cavities in 3-D transesophageal ultrasound images," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 6, pp. 1179–1189, 2015.

[24]   M. Rousson and D. Cremers, "Efficient kernel density estimation of shape and intensity priors for level set segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2005, vol. 3750 LNCS, pp. 757–764.

[25]   G. Carneiro, J. C. Nascimento, and A. Freitas, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 968–982, 2012.

[26]   C. Corsi, G. Saracino, A. Sarti, and C. Lamberti, "Left ventricular volume estimation for real-time three-dimensional echocardiography," *IEEE Trans. Med. Imaging*, vol. 21, no. 9, pp. 1202–1208, 2002.

[27]   N. Lin, W. Yu, and J. S. Duncan, "Combinative multi-scale level set framework for echocardiographic image segmentation," *Med. Image Anal.*, vol. 7, no. 4, pp. 529–537, 2003.

[28]   J. Darby, E. F. Hodson-Tole, N. Costen, and I. D. Loram,

"Automated regional analysis of B-mode ultrasound images of skeletal muscle movement.," *J. Appl. Physiol.*, vol. 112, no. 2, pp. 313–327, 2012.

[29]   T. F. Cootes, G. Edwards, and C. J. Taylor, "Comparing Active Shape Models with Active Appearance Models," *Br. Mach. Vis. Conf. BMVC*, pp. 173–182, 1999.

[30]   N. Narayan, P. Marziliano, J. Kanagalingam, and C. Hobbs, "Speckle Patch Similarity for Echogenicity based Multi-Organ Segmentation in Ultrasound Images of the Thyroid Gland," *IEEE J. Biomed. Heal. Informatics*, vol. 2194, no. c, pp. 1–1, 2015.

[31]   B. Glocker, O. Pauly, E. Konukoglu, and A. Criminisi, "Joint classification-regression forests for spatially structured multi-object segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7575 LNCS, no. PART 4, pp. 870–881.

[32]   N. Paragios, "A level set approach for shape-driven segmentation and tracking of the left ventricle," *IEEE Trans. Med. Imaging*, vol. 22, no. 6, pp. 773–776, 2003.

[33]   J. Xie, Y. Jiang, and H. Tsui, "Segmentation of kidney from ultrasound images based on texture and shape priors.," *IEEE Trans. Med. Imaging*, vol. 24, no. 1, pp. 45–57, 2005.

[34]   E. Angelini, Y. Jin, and A. Laine, "State of the art of level set methods in segmentation and registration of medical imaging modalities," *Handb. Biomed. Image Anal.*, pp. 47–101, 2005.

[35]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst. 25*, pp. 1–9, 2012.

[36]   C. Farabet, C. Couprie, L. Najman, and Y. Lecun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, 2013.

[37]   H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2016, vol. 11–18–Dece, pp. 1520–1528.

[38]   M. Sofka, J. Zhang, S. Good, S. K. Zhou, and D. Comaniciu, "Automatic detection and measurement of structures in fetal head ultrasound volumes using sequential estimation and integrated detection network (IDN)," *IEEE Trans. Med. Imaging*, vol. 33, no. 5, pp. 1054–1070, 2014.

[39]   G. Tzimiropoulos and M. Pantic, "Optimization problems for fast AAM fitting in-the-wild," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 593–600.

[40]   S. Brinkmann, H. Vaghadia, A. Sawka, and R. Tang, "Methodological considerations of ultrasound-guided spinal anesthesia using the Ultrasonix GPS??? needle tracking system," *Canadian Journal of Anesthesia*, vol. 60, no. 4. pp. 407–408, 2013.

[41]   S. Quijano-Roy, D. Avila-Smirnow, R. Y. Carlier, V. Allamand, A. Barois, C. Barnerias, R. Ben Yaou, V. Biancalana, G. Bonne, P. Carlier, F. Colas, O. Dubourg, N. Essid, B. Estournet, A. Ferreiro, P. Guicheney, M. Hamida, K. Hankiewicz, C. Ioos, M. Jarraya, C. LeBreton, F. Leturcq, N. Monnier, D. Monpoint, D. Orlikowski, N. Pellegrini, P. Richard, N. B. Romero, L. Viollet, and S. Wehbi, "Whole body muscle MRI protocol: Pattern recognition in early onset NM disorders," *Neuromuscul. Disord.*, vol. 22, no. SUPPL. 2, 2012.

[42]   R. Cunningham, "The application of b-mode ultrasonography for analysis of human skeletal muscle," Manchester Metropolitan University, Manchester, England, UK, 2015.