

Automatic Normal Positioning of Robotic Ultrasound Probe based only on Confidence Map Optimization and Force Measurement

Zhongliang Jiang¹, Matthias Grimm¹, Mingchuan Zhou¹, Javier Esteban¹, Walter Simson¹, Guillaume Zahnd¹ and Nassir Navab^{1,2}

Abstract—Acquiring good image quality is one of the main challenges for fully-automatic robot-assisted ultrasound systems (RUSS). The presented method aims at overcoming this challenge for orthopaedic applications by optimizing the orientation of the robotic ultrasound (US) probe, i.e. aligning the central axis of the US probe to the tissue’s surface normal at the point of contact in order to improve sound propagation within the tissue. We first optimize the in-plane orientation of the probe by analyzing the confidence map [1] of the US image. We then carry out a fan motion and analyze the resulting forces estimated from joint torques to align the central axis of the probe to the normal within the plane orthogonal to the initial image plane. This results in the final 3D alignment of the probe’s main axis with the normal to the anatomical surface at the point of contact without using external sensors for surface reconstruction or localizing the point of contact in an anatomical atlas. The algorithm is evaluated both on a phantom and on human tissues (forearm, upper arm and lower back). The mean absolute angular difference (\pm STD) between true and estimated normal on stationary phantom, forearm, upper arm and lower back was $3.1 \pm 1.0^\circ$, $3.7 \pm 1.7^\circ$, $5.3 \pm 1.3^\circ$ and $6.9 \pm 3.5^\circ$, respectively. In comparison, six human operators obtained errors of $3.2 \pm 1.7^\circ$ on the phantom. Hence the method is able to automatically position the probe normal to the scanned tissue at the point of contact and thus improve the quality of automatically acquired ultrasound images.

Index Terms—Medical Robots and Systems; Force and Tactile Sensing; Robotic Ultrasound

I. INTRODUCTION

ULTRASOUND (US) is widely used for real-time diagnosis of internal tissues and organs. Unlike X-ray or computed tomography (CT), US imaging is noninvasive, cheap, radiation-free, able to show organ movement in real-time, and therefore widely integrated in clinical practice. However, substantial experience and visuo-tactile skills are required for achieving high-quality US images. Ideal placement of the US probe [2] and suitable contact force between the probe and

Manuscript received: September, 10, 2019; Revised December, 5, 2019; Accepted January, 7, 2020.

This paper was recommended for publication by Editor Pietro Valdastri upon evaluation of the Associate Editor and Reviewers’ comments. (Corresponding author: Mingchuan Zhou.)

¹Zhongliang Jiang, Matthias Grimm, Mingchuan Zhou, Javier Esteban, Walter Simson, Guillaume Zahnd and Nassir Navab are with Computer Aided Medical Procedures, Technical University of Munich, Boltzmannstr. 3, 85748 Garching bei München, Germany. (z1.jiang@tum.de)

²Nassir Navab is with Johns Hopkins University, Baltimore, USA.

The authors would like to thank Dr. Reza Ghobbi from vascular surgery department of Helios Klinikum München West and Mr. Wei Zhang from Klinikum rechts der Isar for their feedback and discussions.

Digital Object Identifier (DOI): see top of this page.

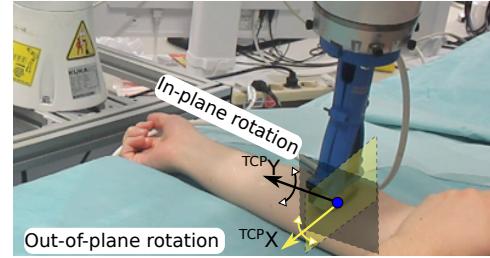


Fig. 1. The two rotations carried out to optimize the probe orientation.

patient [3] are crucial to achieve good image quality. Even for an experienced operator, it is difficult to exactly replicate acquisition parameters (orientation, position and contact force) resulting in large quality variations [4].

To address these challenges, various robotic US systems (RUSSs) have been proposed to assist in US scanning due to their high accuracy, stability and reproducibility. Pierrot *et al.* developed a robotic system with a force control scheme exerting a given force for accurate 3-D reconstruction [5]. Gilbertson *et al.* designed a programmable-force US probe to reduce image variations by providing a known contact force [6]. Conti *et al.* proposed a collaborative RUSS to maintain optimal pressure and provide an interface for 3D volumetric representation [7]. Schneider *et al.* introduced a method to register pre-operative CT to intra-operative camera images with the help of robot [8]. To improve the imaging quality, Burcher *et al.* corrected tissue deformation based on the measured contact force and positional information [3]. More RUSSs have been designed for accurate 3D US imaging [9] and interventional navigation [10].

The previously mentioned works focus on optimizing the contact force F_c in order to achieve better image quality. Besides F_c , the orientation of the US probe also heavily influences the resulting image quality. In order to obtain an optimal image quality, both need to be optimized. However, automatic optimization of the probe orientation has not been fully researched. In a preliminary work, the impact of probe orientation on image quality was investigated [11], and it was concluded that a better quality was obtained when the angle of incidence θ_{AOI} (namely the angle between the US waves and the normal direction of the surface) close to zero. This can be explained by the fact that a zero θ_{AOI} leads to more echoes being reflected back and hence received, rather than scattered

away. Furthermore Chatelain *et al.* discovered that the amount of received sound waves is further increased when the angle between the central axis of the probe and the normal of the tissue in the image plane is zero [12].

Summarizing to achieve optimal image quality the central axis \mathbf{A} of the probe needs to be aligned with the surface normal \mathbf{N} , which requires estimating \mathbf{N} for any contact point. Then the probe can be positioned such that \mathbf{A} and \mathbf{N} align. \mathbf{A} and \mathbf{N} can be decomposed into two orthogonal parts: an in-plane component \mathbf{A}_i and \mathbf{N}_i , aligned with the image plane of the US view and an out-of-plane component \mathbf{A}_o and \mathbf{N}_o orthogonal to the in-plane component, respectively. The location of the two components is depicted in Fig. 2.

To set a suitable probe orientation for autonomous RUSS, Huang *et al.* used the normal direction of a local plane around the scanning path obtained by a depth camera prior to US scanning [9]. Calculating the normal direction \mathbf{N} using a non-deformed scan obtained before the probe reaches the patient is error-prone, as the scan does not include deformations induced by the contact of the probe with the patients skin. To make RUSS adaptable to unknown and/or changing environments, Chatelain *et al.* employed a visual servoing method to optimize the probe orientation in the in-plane direction based on an US confidence map [12]. Briefly, an US confidence map provides a per-pixel measurement of the image quality (signal loss) [1]. However, the method only focused on the in-plane part, while the out-of-plane part still needed to be adjusted via telemanipulation.

Without the ability to finely tune the probe orientation (both from in-plane and out-of-plane view), RUSS would be unable to automatically obtain good quality scans from non-flat surfaces such as human anatomies. To develop a fully automatic RUSS with the ability to provide good image quality, the capability to ascertain the normal direction between probe and surface without requiring external help from human operators is crucial. Due to its independence to calibration accuracy and its sensitivity to changes in contact condition [13], the contact force F_c is a promising candidate for detecting the specified normal direction \mathbf{N} .

The present work aims at introducing a method to identify the normal direction \mathbf{N} of the contact surface for autonomous RUSS for any given point of contact. This is done using estimated forces and the US image. To the best of our knowledge, this is the first paper combining force estimates and real-time US images for estimating the optimal probe orientation. The use of only those two measurements is inspired by the way US sonographers conduct US scans. They do not look at the patient but only at the US images. The main contribution is the ability to fully automatically detect \mathbf{N} and thereby improve imaging quality without requiring an external force sensor or knowledge about the scanned tissue. This is achieved by first computing the in-plane component \mathbf{N}_i based on the confidence map of the real-time US image. Then the out-of-plane component \mathbf{N}_o is computed by first doing a fan motion in the out-of-plane direction and then analyzing the estimated external force from joint torques. The method was validated using both a gel-wax phantom and a volunteer's forearm, upper arm and lower back.

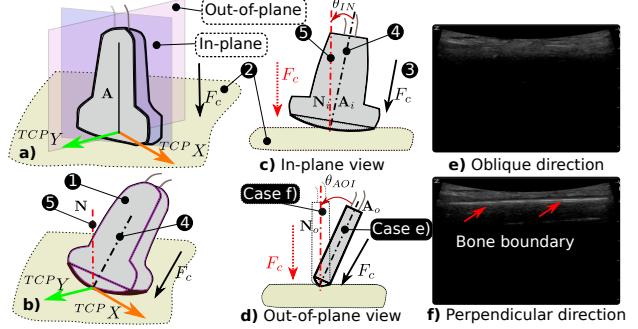


Fig. 2. Impact of the probe orientation on US images. 1: US probe; 2: Surface; 3: Contact force F_c ; 4: probes central axis \mathbf{A} ; 5: Normal direction \mathbf{N} . TCP refers to the tool center point. a) and b) describe ideal and non-ideal probe orientation in 3D, respectively; c) and d) are the in-plane and out-of-plane view, respectively; e) and f) are images from an ideal and non-ideal orientation ($\theta_{AOI} = 10^\circ$). The main difference is the loss of detail (clear boundary of radius bone).

II. ALGORITHM OVERVIEW AND THEORETICAL JUSTIFICATION

This section describes the proposed method for computing the normal direction \mathbf{N} with respect to the contact surface. \mathbf{N} is determined using US images and an estimation of the Cartesian force from joint torques without requiring prior knowledge of the tested tissue or additional sensors. Then an analysis of the estimated force is presented to provide an insight into the dynamic behavior of the estimated force and its impact on the presented application.

A. Strategy for Computing the Normal Direction

In this work, a convex US probe (C5-2/60, Ultrasonix, Richmond, Canada) providing large penetration depth and field of view is attached to the end-effector of a robotic manipulator (KUKA LBR iiwa 7 R800, KUKA Roboter GmbH, Augsburg, Germany).

In order to automatically align the probe axis \mathbf{A} with the normal \mathbf{N} from an arbitrary starting point (see Fig. 2 b)), the in-plane and out-of-plane alignments are carried out separately. Fig. 2 c) and d) describe the two adjustments from the starting orientation (black dashed line) to the desired orientation (red dashed line), respectively. Fig. 2 e) and f) show the corresponding US images of a radial bone obtained with θ_{AOI} equals ten and zero degrees. The latter scan contains a more clear bone outline. This is important, especially for US-CT registration [14], which allows to transfer preoperative CT data into the operating room (e.g. for planning or label transfer). Furthermore the clearer anatomical boundaries aid 3D reconstruction [15]) and carotid stenosis diagnosis [16].

Since the acquired US image is coplanar with the in-plane view, the resulting image quality is highly related to the angular difference between \mathbf{N}_i and \mathbf{A}_i . Methods working directly with pixel intensities are sensitive to artifacts, such as shadows produced by anatomies (see red ellipse in Fig. 3 a)). Building upon the confidence map strategy [1], an original method is used to determine the weighted barycenter of the regions containing tissues (i.e. where the probe is in contact with the imaged object), as opposed to those obscured by

shadows (i.e. out of contact). This allows to compute the correction factor for the orientation in-plane component A_i .

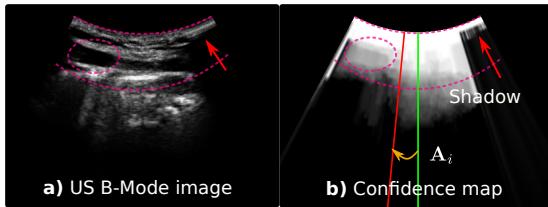


Fig. 3. In-plane orientation adjustment strategy. a) US B-mode scan of forearm; b) corresponding confidence map. The red line depicts N_i .

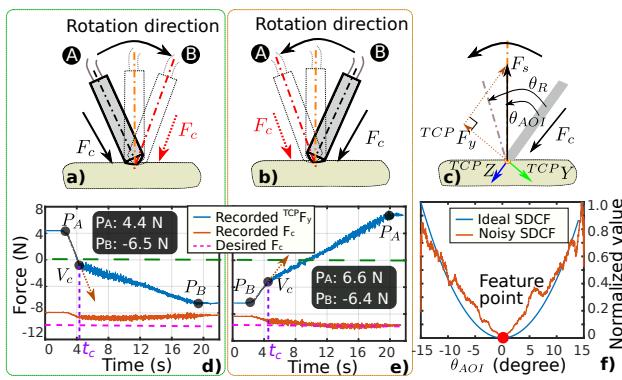


Fig. 4. Cartesian force during fan motion. a) and b) are fan motions of the probe in the out-of-plane direction. Forward from A to B and backward from B to A; c) is the physical force model; d) and e) are the recorded $TCP\hat{F}_y$ (blue line) and contact force (F_c , orange line) corresponding to the cases a) and b) on the gel-wax phantom, where desired $F_c = 10 \text{ N}$ and stiffness = 500 N/m . f) shows the ideal normalized SDCF and normalized first detail coefficient of the noisy SDCF. P_A and P_B marked in d) and e) are the forces recorded at location A and B in a) and b).

To detect the normal component N_o , force information is used, since US images correspond by definition to the in-plane view, thus making the use of image-based metrics to gather information about the out-of-plane dimension challenging. In addition, the force is invariant to registration or calibration errors and it can be estimated by the torque sensors present in all the robot's joints. The contact force F_c between probe and patient is usually set to $3 - 15 \text{ N}$ in clinical practice. Reducing F_c yields a worse contact, whereas a larger F_c reduces the image quality due to deformations.

Then, a fan motion ($\theta_{AOI} = [-15, 15^\circ]$) around the probe long axis (orange arrow in Fig. 2 a)) is executed forward and backward (see Fig. 4 a) and b)) and the resulting force estimated from the joint torques in the direction $TCPY$ (green arrow in Fig. 2 a)) $TCP\hat{F}_y$ is recorded (see Fig. 4 d) and e)). In addition, the theoretical $TCPF_y$ (see Fig. 4 c)) can be calculated as Eq. (1).

$$TCPF_y = -F_c \tan(\theta_{AOI}) \quad (1)$$

Since F_c is constant during the fan motion, $TCPF_y$ only depends on θ_{AOI} . Ideally θ_{AOI} is zero, which corresponds to A_o coinciding with N_o . This leads to a resulting force $TCPF_y$ of zero. TCP refers to the tool center point.

B. Cartesian Force Analysis

The external force $TCPF_y$ at the TCP can be estimated from the joint torques. Since the estimated Cartesian force depends on the joint configuration, singularities should be avoided to get better measurements. The recorded $TCP\hat{F}_y$ during a forward and backward fan motion is shown in Fig. 4 d) and e)). As it can be seen, the same real world point (e.g. A) can yield different recorded forces. The largest observed difference was 5 N without payload and external interaction forces. This difference is mainly caused by the torque sensors inaccuracy and varying static friction in joints. Furthermore, although the forward and backward fan motions are symmetric in Cartesian frame, the joint configurations may be different, because a compliant control scheme is used to guarantee good image quality and patient safety [17].

As shown in Fig. 4 d) and e)), the recorded F_c is initially less than the desired F_c (10 N) because it has been partly compensated by the controller using a virtual force produced by the probe displacement in $TCPZ$ direction after activation of the force control [17]. For each contact, the contact conditions are different and hence the initial recorded \hat{F}_c is varying. However, when the fan motion is started, the absolute value of the recorded \hat{F}_c gradually increases toward the desired force until the tissue deformation creates a reaction force to balance the force \hat{F}_c applied by the robot. The corresponding time stamp is marked as t_c in Fig. 4. Since the initial \hat{F}_c varies, t_c is also varying. Therefore the value of $TCPF_y$ at t_c (marked as V_c) is non-deterministic. Hence, it is not possible to use a fixed threshold for determining the normal N_o . For example, the $TCP\hat{F}_y$ was zero at about 4 s , whereas the normal direction is close to 12 s (middle of time period) in Fig. 4 d).

C. Smooth Derivative of Contact Force (SDCF)

In order to overcome the non-deterministic behaviour of the estimated force, a new feature — smooth derivative of contact force (SDCF) — is defined as in Eq. (2):

$$SDCF(n) = \frac{1}{N} \sum_{i=n-N+1}^n |TCP\hat{F}_y(i+1) - TCP\hat{F}_y(i)| \quad (2)$$

where $n \geq N$, N is the span of the averaging filter and is empirically set to 100.

Since F_c is fixed during the fan motion, the SDCF only depends on the angle θ_{AOI} . The normalized SDCF ($n = 1$) obtained under ideal conditions is shown in Fig. 4 f). Further considering that F_c slightly varies during the fan motion, uniform noise between $[-1, 1] \text{ N}$ is added to the desired force. Then the noisy SDCF is calculated based on the noisy F_c with $n = 100$. The normalized first detail coefficient of discrete wavelet transform (wavelet: sym4, level: 5) result of noisy SDCF is also shown in Fig. 4 f).

As shown in the figure, both ideal and detail coefficient of noisy SDCF are minimal (marked with a red point), when the probe's axis A coincides with the surface normal N . This means that the SDCF is a promising feature to detect the normal direction, even when \hat{F}_c slightly differs from the desired F_c . In addition, since the SDCF represents the change rate of $TCP\hat{F}_y$ rather than the absolute force value, the

effect from the non-deterministic behavior in the beginning is reduced. Furthermore, an approach based on finding the proper minimum of the SDCF is likely to yield better results than a threshold-based approach, especially when non-stationary objects are involved. More details about identifying the desired \mathbf{N}_o are described in Section IV.

III. OPTIMIZATION OF IN-PLANE ORIENTATION

A. Ultrasound Confidence Map

Since the US view is coplanar with the in-plane view, an image-based method can be used to identify the in-plane normal \mathbf{N}_i . Therefore, the US confidence map [1] is exploited. It provides an objective, stable and repeatable per-pixel assessment of imaging quality based on the estimation of the signal loss. An US image is transformed into a probabilistic map $C \in \mathbb{R}^2 \rightarrow [0, 1]$, where the maximum value (1, white) means perfect signal quality and the minimum value (0, black) means no signal. For each pixel, the corresponding value of C can be interpreted as the chance of the signal reaching the transducer, which makes C locally more continuous than the intensity values (see red ellipse in Fig 3). Furthermore, regions where the probe is not in contact with the patient are easily detectable due to their low confidence (see the shadow in Fig. 3). Thus the US confidence map is a good tool to optimize the in-plane orientation, especially in the presence of shadows at the periphery of the US scan.

B. Orientation Computation for In-Plane Component

The present framework is based on the use of convex US probes, however it could straightforwardly be applied to linear transducers as well. For a convex probe, the US image can be defined in polar coordinates as $\Omega = [r_{min}, r_{max}] \times [\theta_{min}, \theta_{max}]$, where Ω represents the US scan; r_{min} and r_{max} denote the radius of the probe and the scan depth, respectively; θ_{min} and θ_{max} are the angular offsets from the two sides of the central line ($\theta_{max} = -\theta_{min}$), respectively. The line from the US focal point and the confidence-weighted barycenter ζ_c is a good approximation of \mathbf{N}_i [12]. Hence the robot is moved such that \mathbf{A}_i intersects ζ_c , which can be calculated as follows:

$$\zeta_c = \frac{1}{\aleph_c} \int_{\theta_{min}}^{\theta_{max}} \int_{r_{min}}^{r'} \frac{\theta}{\theta_{max} - \theta_{min}} C(\theta, r) dr d\theta \quad (3)$$

where $\aleph_c = \int \int_{(\theta, r) \in \Omega} C(\theta, r) dr d\theta$ is the accumulated confidence value over the entire image (Ω), $r' \in (r_{min}, r_{max}]$ defines the region of interest, θ is the angular deviation from central scan-line in the image and θ_{min} and θ_{max} were set to -0.5 and 0.5 respectively.

To reach the desired position, a rotational offset R_c is applied to the current position. R_c is calculated as follows:

$$R_c = -\frac{1}{2} k_c \zeta_c \Psi_p \quad (4)$$

where k_c is the control gain, Ψ_p is the angular field of view of the convex probe. This is done iteratively until no further improvement is reached.

IV. OPTIMIZATION OF OUT-OF-PLANE ORIENTATION

This section proposes a method to compute \mathbf{N}_o using force values recorded during a fan motion of the robot. Due to large noise in the recorded force data, several de-noising steps are applied. An overview of the method is given in Fig. 5.

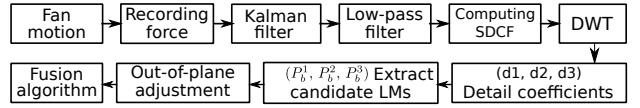


Fig. 5. Overview over the out-of-plane optimization algorithm.

A. Kalman Filter Implementation

Most general-purpose robotic manipulators have the ability to estimate ${}^{TCP}F_y$ based on joint torques. However, the forces involved in RUSS are smaller than 10 N due to safety concerns. Hence, digital noise has severe impact on the estimated Cartesian force for this application. In addition, changes of the contact point between probe and patient due to patient movement further reduce the accuracy of the recorded force (blue line in Fig. 6). The Kalman filter is a standard technique to reduce noise by correcting new measurements based on previous ones. Hence a Kalman filter is employed to denoise the recorded force as in [18].

The fan motion is executed with a low velocity (0.025 rad/s) compared to the sampling frequency of the force (100 Hz). Thus the computed force values differ only slightly for consecutive sampling points. In addition, the Kalman filter operates only on the recorded force, hence the transition matrices for the estimated state (A) and measured state (H) are set to identity. Furthermore, the estimation and measurement noise covariance matrices are empirically set to 10^{-8} and 4×10^{-6} , respectively. The performance of the Kalman filter when applied to the recorded force values during two forward and backward fan-motions with the same motion parameters are depicted in Fig. 6, one on a phantom and one on a volunteer's lower back. The volunteer is breathing, which causes significant fluctuations of the recorded force as shown in Fig. 6 b).

B. Extraction of Frequency-Band Components from SDCF

Due to the low rotational velocity, the important force components are concentrated in the low frequency domain. Hence a low pass filter is employed after the Kalman filter to further reduce noise. In order to determine a suitable stop-band frequency F_s , the Fast Fourier Transform (FFT) is used to investigate the frequency decomposition of the recorded ${}^{TCP}\hat{F}_y$. An example of the FFT result and the corresponding power spectral density (PSD) of the force recorded on the phantom is shown in Fig. 7 a). The high frequency part can be removed because it mainly corresponds to digital noise of the sensors. Based on Fig. 7 a), the PSD is rapidly attenuated after 5 Hz and the energy (cumulative PSD) of the signal components at 15 Hz occupies 85% of the total signal energy. Hence, F_s is set to 15 Hz . The result of the low-pass filter is depicted in Fig. 7 b).

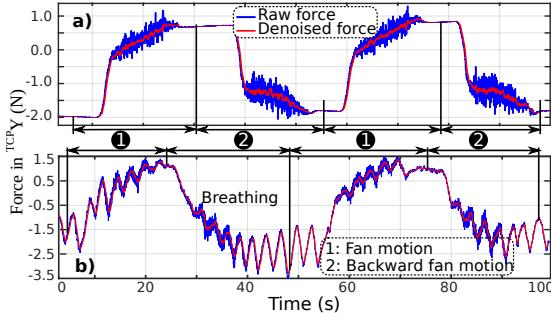


Fig. 6. Result of the Kalman filter. a) and b) show the experiments preformed on the phantom and volunteer's back with breathing. The involved forward and backward fan motion are performed with $F_c = 8 \text{ N}$ for the phantom and 6 N for the volunteer back, stiffness= 250 N/m and Θ_{AOI} was $(-15^\circ, 6^\circ)$ and $(6^\circ, -15^\circ)$, respectively. The parameters of the Kalman filter were set to: estimation noise covariance 10^{-8} , measurement noise covariance 4×10^{-6} , the initial state was set to the first recorded force and the initial error covariance was set to 1.

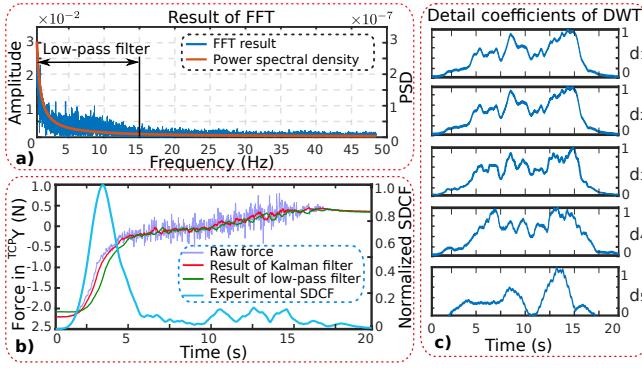


Fig. 7. Frequency decomposition of SDCF. a) result of FFT applied to $TCPY_F_y$; b) result of low-pass filter and the normalized SDCF. The raw force is shown with a purple line, the result of the Kalman filter and the result of the low-pass filter are depicted by red and green lines, and the blue line depicts the SDCF calculated from the result of the low-pass filter. c) Normalized detail coefficients of SDCF's DWT result. The level was set to 5, and the sym4 wavelet was used as basis.

The SDCF is calculated using the result of the low-pass filter and Eq. (2). Its normalized result is shown in Fig. 7 b). Due to the non-periodic nature of the force caused by the fan motion, there is no identifiable main frequency in the FFT result. In order to obtain accurate identification results, the discrete wavelet transformation (DWT) is used to divide the SDCF into its different frequency band components.

The DWT provides high frequency and temporal resolution, which makes it suitable for processing unstable signals. The DWT requires a decomposition level and wavelet function as input. The sym4 wavelet function was empirically chosen as basis function in this paper. To investigate the effect of the different frequency parts, the decomposition level was set to 5. This results in five detail coefficients $d_i, i = 1 \dots 5$ covering the intervals $[\frac{F_s}{2} - F_s], [\frac{F_s}{4} - \frac{F_s}{2}], [\frac{F_s}{8} - \frac{F_s}{4}], [\frac{F_s}{16} - \frac{F_s}{8}]$ and $[\frac{F_s}{32} - \frac{F_s}{16}]$, respectively. The d_i have different frequency interval size, hence their amplitude range also differs. Therefore the amplitudes are normalized between [0, 1] (as depicted in Fig. 7 c)). Since the normalization is linear, it does not affect the locations of local minima.

As shown in Fig. 7 c), the amplitude graphs for d_4 and d_5

are unlike the other three. This is due to them having a low frequency value and small frequency range. Low frequencies are influenced by contact point changes, e.g. due to breathing. Hence, only d_1, d_2 and d_3 are used to detect N_o .

C. Extracting Local Minima from SDCF

In theory, the SDCF is minimal when \mathbf{A} is aligned with \mathbf{N} as shown in Fig. 4 f). Due to noise, this minimum is not necessarily the global minimum for real scenarios. The experimental detail signals ($d_i, i = 1, 2, 3$) from Fig. 7 c) have multiple local minima (LM). The LM can be detected by comparing nearby points using Eq. (5). The detected LM are depicted as red stars in Fig. 8.

$$I_i = \{t | d_i(t - \Delta t) > d_i(t) \wedge d_i(t + \Delta t) > d_i(t)\} \quad (5)$$

where I is the set used to save the corresponding time stamps of LM, $i = 1, 2, 3$ is the detail coefficient index, t is the time stamp and $\Delta t = 10 \text{ ms}$ is the sampling interval. The equation is written in set-builder notation.

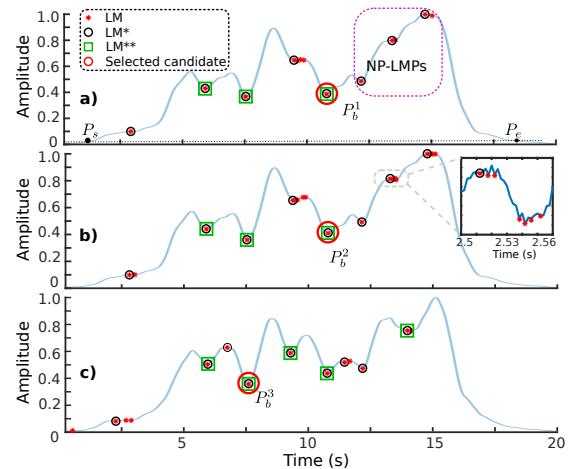


Fig. 8. Suitable local minimum selection. a), b) and c) correspond to d_1, d_2 and d_3 , respectively. The parameters are set as follows: $W_1 = 0.8, L_1 = 0.04, H_1 = 0.08$.

A method is proposed to detect the right LM for each d_i . During this process, two variables (LM^* and LM^{**}) are defined. LM^* are the LM remaining after removing neighboring LM (LM located in close proximity to each other, see the enlarged plot in Fig. 8 b)). LM^{**} are the LM^* remaining after further removing the so-called NP-LMs (noise-perturbed) which are caused by noise perturbing the estimated force (see the pink rectangle in Fig. 8).

To extract the LM^* , a sliding window filter with width W_1 and length L_1 is applied to remove all but the first element for each neighboring LM group. The LM^* are marked with a black circle in Fig. 8 and the corresponding timestamps form the set I_i^* . To further remove NP-LM, the input signal d_i is partitioned into segments along the LM^* . If a LM^* amplitude is close to the maximum amplitude of one of its neighbouring segments, it is discarded. The remaining LM (LM^{**}) are marked with a green square in Fig. 8. A set I_i^{**}

containing their corresponding timestamps is created. This is explained in Eq (6).

$$I_i^{**} = \{t | {}^L M_i^k - d_i(t) > H_1 \wedge {}^R M_i^k - d_i(t) > H_1\} \quad (6)$$

where $i = 1, 2, 3$ is the detail coefficient index, t are the recorded time stamps, $k = 1, 2, \dots, \text{length}(I_i^*) - 2$, H_1 is the amplitude threshold, ${}^L M_i^k$ and ${}^R M_i^k$ are the maximum amplitude values of the left and right segments adjacent to the k th LM. The equation is written in set-builder notation.

In order to identify the desired LM candidate for each d_i , a feature A^* is defined as in Eq. (7).

$$A^*(LM, i) = ({}^L M_i^{LM} - d_i(t)) \cdot ({}^R M_i^{LM} - d_i(t)) \quad (7)$$

where $i = 1, 2, 3$ is the detail coefficient index, LM is the LM, $t = I_i^{**}(k)$ is the timestamp of the LM, ${}^L M_i^j$ and ${}^R M_i^j$ are the maximum amplitude values of the segment left and right of LM .

For each d_i the LM^{**} with the largest A^* is selected as the final candidate (depicted with red circle in Fig. 8).

D. Computation of Normal Direction

The potential normal direction is computed from the final LM (P_b^1, P_b^2 and P_b^3) of the detail coefficients d_i , respectively. To further improve the robustness of the identified result, a fusion algorithm using the three candidate LM is proposed. As it can be seen in Fig. 9, there are three possible cases and five potential results:

Case 1: The maximum A^* of P_b^1, P_b^2 and P_b^3 is larger than the threshold H_A . The timestamp of the LM with maximum A^* is returned.

Case 2: The three selected LM have a similar timestamp. Then the mean timestamp is returned.

Case 3: Two of the selected LM have similar timestamps.

The mean of their timestamps is denoted as t_{mean} , while the third timestamp t_d differs. The two selected LM with similar timestamps are denoted as a and b , the last one is called c . T_1 and T_2 are computed as described in Case 3.1 and Case 3.2, respectively. If T_2 is greater than T_1 , t_d is returned, otherwise t_{mean} .

Case 3.1: A LM corresponding to t_{mean} is searched in the set LM_c^{**} . If there is an element in I_c^{**} with a timestamp not more than H_3 seconds away from t_{mean} , the search is considered a success. T_1 is computed as the sum of the A^* of the searched LM (if available), P_b^a and P_b^b .

Case 3.2: As in case 3.1, the two corresponding LM for t_d are searched among the LM_a^{**} and LM_b^{**} . T_2 is computed as the sum of the A^* of the searched LM (if available) and P_b^c .

Failure: If all three selected LM have different timestamps, then the algorithm returns a failure.

Finally, the returned \mathbf{N}_o is the \mathbf{A}_o corresponding to the returned timestamp.

V. RESULTS

A. Experimental Setup

The overall experimental setup is shown in Fig. 10. A convex US probe is attached to a robot flange using a 3D-printed mount and B-mode US images are acquired using

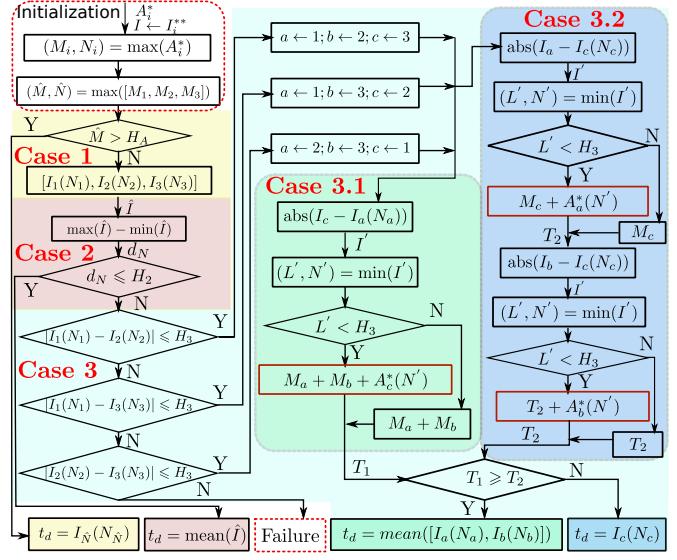


Fig. 9. Fusion algorithm to extract \mathbf{N}_o from the recorded probe orientations based on timestamps. The parameters are set as follows: $H_A = 0.3$, $H_2 = 0.8$ and $H_3 = 1.5$.

an Ultrasonix RP machine (BK Ultrasound, Peabody, MA, USA). B-mode images are transferred to a computer using an Ethernet cable and the OpenIGTLINK protocol. The computer communicates with the robot using ROS (Robot Operating System). Control commands and the robot status are exchanged at 100 Hz. The algorithm was evaluated on a gel-wax phantom with a flat surface and on a human volunteer (lower arm, upper arm, lower back). The phantom is fixed in foam plastic to imitate the behavior of skin. The phantom is placed on a flat surface. Hence the ground truth is represented as $B\vec{V}_g = [0, 0, 1]$ in the robotic base frame. For the human volunteer, the ground truth is computed based on scanning the surface around the desired path. The angular difference between estimated normal $\hat{\mathbf{N}}$ and ground truth normal \mathbf{N} is computed using the equation: $\theta = \cos^{-1}(\hat{\mathbf{N}} \cdot \mathbf{N} / (|\hat{\mathbf{N}}| \cdot |\mathbf{N}|))$.

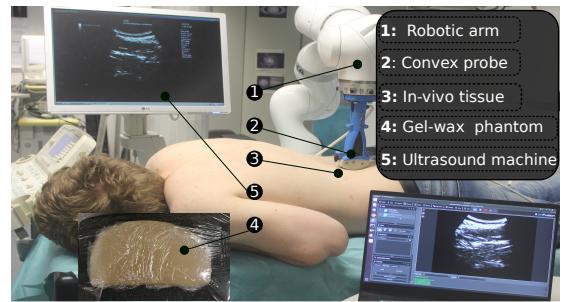


Fig. 10. Experimental setup.

B. Validation of Estimating \mathbf{N}_i using the Gel-Wax Phantom

The computation of \mathbf{N}_i is the first step of the proposed algorithm. It was validated using a series of experiments performed on the gel-wax phantom using three different initial θ_{AOI} and four different initial offset angles θ_{IN} in the in-plane view, where θ_{IN} is the angular difference between

\mathbf{A}_i and \mathbf{N}_i . The US confidence map is calculated from US images, which are highly sensible to deformations. In order to assess the impact of deformation on in-plane adjustment, the robot is controlled in position mode. Each of the twelve experiments is performed four times with different depths ($z = z_0 + 0.5, z_0 + 1, z_0 + 1.5, z_0 + 2 \text{ mm}$), where z_0 is the depth at which the probe starts touching the phantom.

\mathbf{N}_i is estimated using Eq. (3) with $r' = r_{\min} + 0.2(r_{\max} - r_{\min})$. The error (e_{in} , mean \pm STD) for different θ_{IN} and θ_{AOI} is shown in Table-I. The experiments show that the method is able to achieve good estimates for \mathbf{N}_i .

TABLE I
MEAN ABSOLUTE ERROR (\pm STD) OF IN-PLANE IDENTIFIED RESULTS

e_{in}	θ_{AOI}	0	5	10
θ_{IN}				
0		1.11 ± 0.55	0.18 ± 0.09	1.20 ± 0.79
3		0.94 ± 0.66	0.84 ± 0.29	0.56 ± 0.36
6		1.38 ± 0.41	1.61 ± 0.53	1.55 ± 0.39
9		2.71 ± 0.57	2.92 ± 0.20	2.69 ± 0.81

*Units are in degree.

C. Validation of Estimating \mathbf{N}_o using the Gel-Wax Phantom

Five experiments were performed on the gel-wax phantom with flat surface and $F_c = 8 \text{ N}$. To make the tests more realistic and challenging, an asymmetric fan motion ($\theta_{AOI} = [-6, 15^\circ]$) is employed. The calculated angular differences θ_c between \mathbf{A} and \mathbf{N} of the five paired fan motions are shown in Fig. 11. It is calculated by $\sqrt{\theta_{IN}^2 + \theta_{AOI}^2}$. \mathbf{A}_i is not necessarily aligned with \mathbf{N}_i , therefore even if \mathbf{A}_o is aligned with \mathbf{N}_o , an error of zero is not necessarily achieved. Therefore the minimum error for each fan motion is equal to θ_{IN} . For multiple repeated fan motions, it can be seen that θ_{IN} increases with the number of fan motions. This is caused by the probe sliding on the contact surface, due to US gel and deformations of the soft tissue.

The \mathbf{A}_o corresponding to the timestamps of the computed results (red and green dots in Fig. 11) are taken as estimated \mathbf{N}_o . Except for one result (third fan motion, green dot), all results are close to the desired direction ($< 5^\circ$), despite θ_{IN} being non-zero. Furthermore, five of the ten results are very close to the minimum ($< 0.5^\circ$). This shows that the method can detect \mathbf{N}_o purely based on the estimated force provided by a robot without external force sensor.

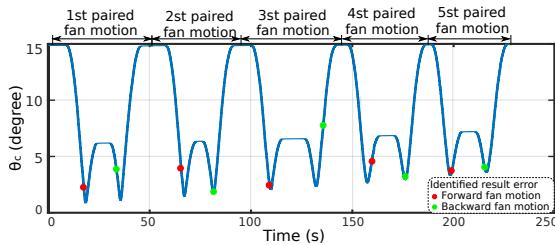


Fig. 11. Computed results for five paired fan motions. The red and green dots depict the results for the forward and backward fan motion, respectively.

To further analyse the proposed method, the normalized amplitude for the d_i for the fifth paired fan motion (largest

initial angle difference with 3.3° and 3.6°) are shown in Fig. 12. The final differences between the ideal \mathbf{N} and detected $\hat{\mathbf{N}}$ are 3.5° and 3.8° for the forward and backward fan motion. However, when the in-plane component is ignored, the errors are 1.0° ($\sqrt{3.47^2 - 3.32^2}$) and 1.3° ($\sqrt{3.84^2 - 3.60^2}$), respectively. This means that \mathbf{N}_o can be estimated reliably, even if the estimation for \mathbf{N}_i is poor.

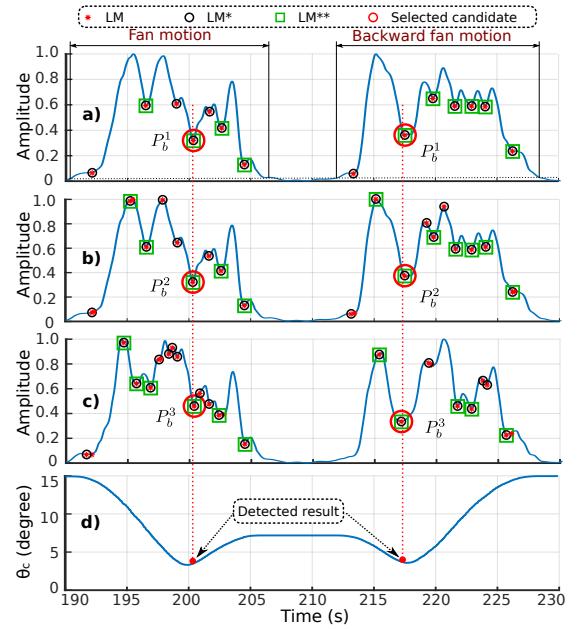


Fig. 12. Feature curve for one paired fan motion. The two red dashed lines represent the selected timestamps based on the detail signals of SDCF (d_i) for the forward and backward fan motion, respectively

D. Comparison with Human Operators

To compare the proposed algorithm with US operators, six volunteers experienced in US scanning were each asked to manually position the US probe along the normal direction of the same phantom ten times (60 times total). The average errors for the human operators are $3.2 \pm 1.7^\circ$ while the error of the proposed method on the same phantom is $3.1 \pm 1.0^\circ$. From Fig. 13, the operators perform better than the proposed method in the best case (error: 0.2° vs 1.7°) while the error of proposed method is more stable and in the worst case (3.6°) performs better than human operators (5.9°). Based on a t-test (probability $p = 0.52 > 0.05$), there is no significant difference between the proposed method and human operators. Hence the proposed method is able to detect \mathbf{N} with a comparable accuracy to the participating human operators on a phantom.

E. In-Vivo Validation

The recorded force is highly related to the properties of the scanned tissue. In order to evaluate the applicability of the proposed method on humans, experiments were performed on the forearm, upper arm and lower back with breathing of a volunteer. Eight scans were performed per anatomy. The computed error is shown in Fig. 13. Since there is variation

in the involved tissue stiffness, the robot stiffness was set to 250 N/m, 350 N/m and 450 N/m during the tests to investigate the impact of the robot stiffness and $F_c = 6$ N.

From Fig. 13, it can be seen that the algorithm performed best on the phantom. This is due to the phantom being stationary whereas humans exhibit motion. For the forearm, the lowest error for each set of experiments is smaller than 2°, and the 25th percentile of all three sets of experiments is close to 2°. The algorithm performed best when the stiffness was 450 N/m, having a maximum error of 5.8°. For the upper arm, the average errors corresponding to the different stiffnesses are distributed around 5.5°. For stiffness 350 N/m, both the variance and the upper error bound (7.6°) were the lowest among the stiffnesses. The error for the lower back is larger than for the other anatomies due to respiratory motion. However, the algorithm still achieves good best case results (1.7°, 2.9° and 1.7°) for the different stiffness, respectively. The results suggest that the robot stiffness should be adapted to the tissues stiffness.

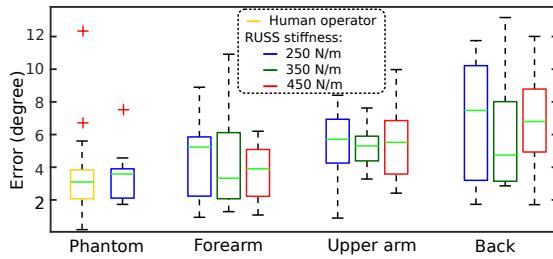


Fig. 13. Identification errors on phantom/volunteer. The first two plots are the experiments performed on the phantom by humans and the proposed method. The other plots show the results of the method on in-vivo tissues.

F. Applicability to Other Anatomies

The proposed method was developed for orthopaedic applications, where deformations are less severe due to the presence of bones possessing large elastic modulus (8.91 GPa) [19]. In order to apply the method to very soft tissues (e.g. breast) the out-of-plane optimization would need to be changed, as the optimal orientation is not necessarily the normal. In that case, one could learn the optimal force for the out-of-plane orientation which may not be systematically normal to the original curvature of the anatomy. The in-plane orientation could be applied without change.

VI. CONCLUSION

A full orientation (both in-plane and out-of-plane) optimization algorithm was introduced to enable RUSS to automatically obtain better quality orthopaedics US images. The focus of the algorithm is on aligning the probe to the (in- and out-of-plane) normal to the surface at the point of contact. The proposed method does not require external force sensors or knowledge about the scanned anatomy. The main drawback of the method is that it requires a full fan motion at each point of contact. The current work only focuses on optimizing the orientation of the US probe. Future work could include jointly optimizing the contact force and the orientation in order to obtain an ideal

image. We hope that the proposed approach moves us a step further towards large-scale use of RUSS in operating theatres by allowing robots to understand the handling of the probe in order to acquire better quality of images.

REFERENCES

- [1] A. Karamalis, W. Wein, T. Klein, and N. Navab, "Ultrasound confidence maps using random walks," *Med. Image Anal.*, vol. 16, no. 6, pp. 1101–1112, 2012.
- [2] A. Scorza, S. Conforto, C. d'Anna, and S. Sciuto, "A comparative study on the influence of probe placement on quality assurance measurements in B-mode ultrasound by means of ultrasound phantoms," *The Open Biomedical Engineering Journal*, vol. 9, p. 164, 2015.
- [3] M. R. Burcher, J. A. Noble, L. Man, and M. Gooding, "A system for simultaneously measuring contact force, ultrasound, and position information for use in force-based correction of freehand scanning," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 52, no. 8, pp. 1330–1342, 2005.
- [4] R. Kojcev, A. Khakzar, B. Fuerst, O. Zettning, C. Fahkry, R. DeJong, J. Richmon, R. Taylor, E. Sinibaldi, and N. Navab, "On the reproducibility of expert-operated and robotic ultrasound acquisitions," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 6, pp. 1003–1011, 2017.
- [5] F. Pierrot, E. Dombre, E. Déglouange, L. Urbain, P. Caron, S. Boudet, J. Gariépy, and J.-L. Ménien, "Hippocrate: A safe robot arm for medical applications with force feedback," *Med. Image Anal.*, vol. 3, no. 3, pp. 285–300, 1999.
- [6] M. W. Gilbertson and B. W. Anthony, "Force and position control system for freehand ultrasound," *IEEE Trans. Robot.*, vol. 31, no. 4, pp. 835–849, 2015.
- [7] F. Conti, J. Park, and O. Khatib, "Interface design and control strategies for a robot assisted ultrasonic examination system," in *Experimental Robotics*. Springer, 2014, pp. 97–113.
- [8] C. Schneider, C. Nguan, R. Rohling, and S. Salcudean, "Tracked pick-up ultrasound for robot-assisted minimally invasive surgery," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 2, pp. 260–268, 2015.
- [9] Q. Huang, J. Lan, and X. Li, "Robotic arm based automatic ultrasound scanning for three-dimensional imaging," *IEEE Trans. Ind. Inform.*, vol. 15, no. 2, pp. 1173–1182, 2018.
- [10] J. Guo, C. Shi, and H. Ren, "Ultrasound-assisted guidance with force cues for intravascular interventions," *IEEE Trans. Autom. Sci. Eng.*, no. 99, pp. 1–8, 2018.
- [11] B. Ihnatsenka and A. P. Boezaart, "Ultrasound: Basic understanding and learning the language," *Int. J. Shoulder Surg.*, vol. 4, no. 3, p. 55, 2010.
- [12] P. Chatelain, A. Krupa, and N. Navab, "Confidence-driven control of an ultrasound probe," *IEEE Trans. Robot.*, vol. 33, no. 6, pp. 1410–1424, 2017.
- [13] Z. Jiang, X. Qi, Y. Sun, Y. Hu, G. Zahnd, and J. Zhang, "Cutting depth monitoring based on milling force for robot-assisted laminectomy," *IEEE Trans. Autom. Sci. Eng.*, pp. 1–13, 2019.
- [14] W. Wein, A. Karamalis, A. Baumgartner, and N. Navab, "Automatic bone detection and soft tissue aware ultrasound-ct registration for computer-aided orthopedic surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 6, pp. 971–979, 2015.
- [15] R. Prevost, M. Salehi, S. Jagoda, N. Kumar, J. Sprung, A. Ladikos, R. Bauer, O. Zettning, and W. Wein, "3D freehand ultrasound without external tracking using deep learning," *Med. Image Anal.*, vol. 48, pp. 187–202, 2018.
- [16] C. Arning, B. Widder, H. Stiegler, M. Görtler *et al.*, "Revision of degum ultrasound criteria for grading internal carotid artery stenoses and transfer to nascent measurement," *Ultraschall in der Medizin*, vol. 31, no. 3, pp. 251–257, 2010.
- [17] O. Zettning, B. Frisch, S. Virga, M. Esposito, A. Riemmüller, B. Meyer, C. Hennersperger, Y.-M. Ryang, and N. Navab, "3D ultrasound registration-based visual servoing for neurosurgical navigation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 9, pp. 1607–1619, 2017.
- [18] G. Welch and G. Bishop, "An introduction to the Kalman filter," *Dept. Comput. Sci. Univ. North Carolina, Chapel Hill, Tech. Rep. TR95041*, 2006.
- [19] K. Choi, J. L. Kuhn, M. J. Ciarelli, and S. A. Goldstein, "The elastic moduli of human subchondral, trabecular, and cortical bone tissue and the size-dependency of cortical bone modulus," *J. Biomech.*, vol. 23, no. 11, pp. 1103–1113, 1990.