Using Ultrasound Images of the Forearm to Predict Finger Positions

Claudio Castellini, *Member, IEEE*, Georg Passig, and Emanuel Zarka

Abstract—Medical ultrasound imaging is a well-known technique to gather live views of the interior of the human body. It is totally safe, it provides high spatial and temporal resolution, and it is nowadays available at any hospital. This suggests that it could be used as a human–computer interface. In this paper, we use ultrasound images of the human forearm to predict the finger positions, including thumb adduction and thumb rotation. Our experimental results show that there is a clear linear relationship between the features we extract from the images, and finger positions, expressed as angles at the metacarpo-phalangeal joints. The method is uniformly valid for all subjects considered. The unavoidable movements of the ultrasound probe with respect to the skin and of the skin with respect to the inner musculoskeletal structure are compensated for using the optical flow. Typical applications of this system range from teleoperated fine manipulation to finger stiffness estimation to ergonomy. If successfully applied to transradial amputees, it could be also used to reconstruct the imaginary limb, paving the way to, e.g., fine control of hand prostheses, treatment of neuropathic/phantom limb pain and visualization of the imaginary limb as a tool for the neuroscientist.

Index Terms—Learning and adaptive systems, rehabilitation, ultrasound imaging.

I. INTRODUCTION

DEVELOPED soon after the second World War as a diagnostic device, ultrasound imaging, also known as medical ultrasonography (US) is a noninvasive technique to visualize structures inside the human body. (A comprehensive reference to medical ultrasound is the classical textbook [1].) The general principle is that of wave reflection/refraction: in modern ultrasound medical devices, an array of piezoelectric transducers is used to generate a focused wave of ultrasound in the range of 2–20 MHz which penetrates the body part of interest; partial reflection of the wave at the interfaces between tissues with different acoustic impedance (density) is then gathered and converted, in the so-called B-mode, to a gray-scale 2-D image. High-gray-valued ridges in the image denote therefore tissue interfaces. Modern US machines can achieve sub-millimeter spatial resolution and/or real-time temporal resolution, penetrating several centimeters below the subject's skin. The technique is

easy to set up and it has no known side effects, to the extent that one of its best known applications is the imaging of the fetus with prebirth diagnostic purposes. Additionally, although ultrasound machines are not cheap, they are nowadays found essentially in any hospital, making the technique easily available.

Such an accurate, safe, and widespread technique is likely to have an immediate application as a human–computer interface. In particular, here we focus on one such use, namely, determining finger positions using features extracted from US images of the human forearm. For a long time, US imaging has successfully been used as a diagnostic tool for hand musculoskeletal disorders such as, e.g., synovitis and rheumatoid arthritis [2]–[4], so US images should contain enough information to reconstruct the position, velocity and/or force exerted by the fingers. If this happens to be the case, a system enforcing this idea would have potential applications in, e.g., ergonomy and precise teleoperation and manipulation when operated by intact subjects, and could be a breakthrough if used by amputees, letting them control hand prostheses to a so-far unknown degree of precision. (Of course this claim is subject to many assumptions—see Section IV for more about it.)

The feeling that the idea is viable stems, as a start, from plain observation of the US imaging of the human wrist as the fingers move. Consider the movie "fingers.avi" included in the supplemental material. The movie is recorded from a healthy subject using a standard portable US machine (see Section II-A2 for more details). The transducer lies on the ventral side of the right wrist orthogonal to the axis of the forearm (see also Fig. 2, right panel). As the subject's fingers move, a clear relationship between the changes in the image and the hand configuration is apparent; although not elementary, the changes appear repeatable and related to the flexion of single fingers, even in the case of the thumb adduction and rotation. Flexion of the pinkie finger for example, results in a "hole" opening and closing near the left-upper corner of the image. Comparison with an anatomical representation of the very same section (see Fig. 1) suggests that we are looking at the motion of one of the tendons of the *M. Flexor Digitorum Superficialis*.

A. Overview of the Methodology and Results

In this paper, we show a positive result along this line. Six healthy human subjects were instructed to mimic with their right hand the movements performed by an animated human hand model on a computer screen. The hand configuration and US images of the corresponding wrist would be gathered and synchronized during the experiment, using a dataglove and the above mentioned US machine. The movements consisted of repeated flexion of the fingers plus adduction and rotation of the
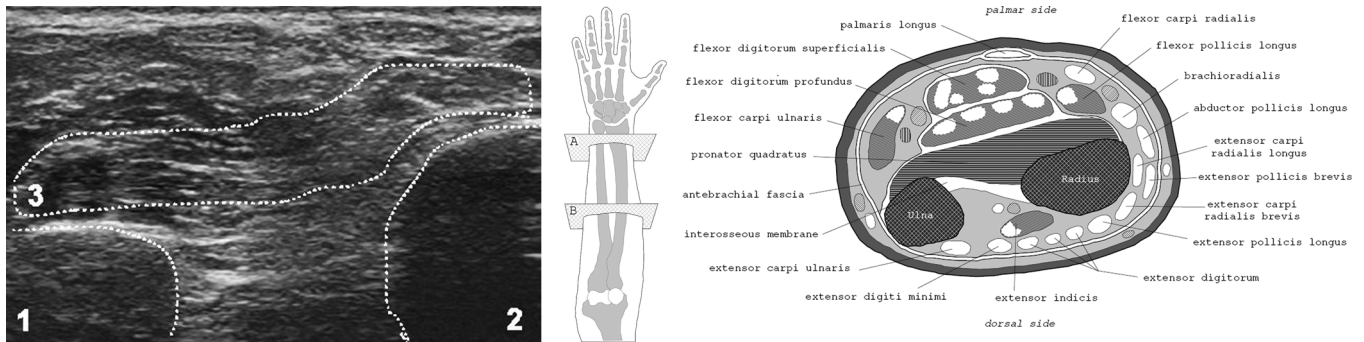
Fig. 1. Comparison of a typical US image of the human wrist (left) and a pictorial representation of its anatomy (right, representing section "A" in the middle panel, both figures reproduced from [26]). Annotations on the left panel highlight the ulna (1), the radius (2), and the *M. Pronator Quadratus* (3) on the ultrasound image. Above the area of the *Quadratus*, circular sections of the flexor tendons are clearly visible.

thumb, in several different configurations. Offline, local spatial features were extracted from each frame. The features encode a linear approximation of the gray-level distribution around a set of interest points. The points are chosen to belong to a uniformly spaced grid, irrespective of any anatomical detail of the subject.

Regression analysis reveals that these features are almost perfectly correlated (in the sense of the standard Pearson correlation) to finger positions, leading to a prediction error of as little as about 1% of the position range. The correlation is higher where the sections of anatomically relevant muscles appear; for example, pinkie movement is highly correlated with features extracted near the section of the *F.D.Superficialis*, i.e., from the upper-left corner of the images seen in the movie—where the "hole" grows and shrinks. A regression matrix K can be estimated via, e.g., least-squares, so that $\mathbf{p} = k\mathbf{v}$, where $\mathbf{p}$ represents the position of the fingers (six angles measured at the metacarpo-phalangeal joints) and $\mathbf{v}$ encodes the visual features extracted from the US frames. In order for the approach to work fine, the transducer must be as still as possible with respect to the skin and skeleton of the subject; unwanted motion is compensated for by evaluating the optical flow at key frames and using it to shift the position of the interest points in order for the extracted features to always represent the same information. This method is shown to significantly prevent drift errors during the experiment and improve the prediction accuracy.

The paper is organized as follows. After reviewing the related work, we describe the experimental setup and the processing techniques employed (Section II); we then show the experimental results (Section III) and lastly, conclusions and future work are presented (Section IV).

B. Related Work

Extensive work on the use of live ultrasound imaging to control a one-degree-of-freedom hand prosthesis appears since 2006 in a series of joint Chinese/British studies by, among others, Zheng [5]–[7]. The authors focus on the large extensor muscle of the forearm, *M. Extensor Carpi Radialis*, and show high correlation between the wrist extension angle (detected with a goniometer) and the change in size of the projection of the muscle itself in the image. This is a computationally easy visual feature (unsurprisingly, the authors claim it can be evaluated in real-time) which happens to be linearly related to a particular movement. The technique of determining muscle

contraction using ultrasound imaging is therein named *sonomyography*, a term that we will not use here as we do not necessarily target muscle contraction only.

As far as we know this method has not yet been applied to the whole hand/finger system, and the authors have never considered more than one feature at the same time. This restricted focus is probably motivated by the diversity and complexity of the changes in US images as joint positions change: the single identified feature is related to a precise anatomical change, a relation which would be quite hard to assess in the general case. It is likely that a more general treatment in that case would require a detailed model of the kinematics of the human forearm, *plus* a detailed model of the changes in the projected US image as the hand joints move—a task which seems overtly complex.

As opposed to that, in this work we take a more image-processing-/machine-learning-oriented approach: we employ a uniformly-spaced set of visual features from each image, extract uniform features from each point and then use data gathered from a dataglove to relate joint positions and images, without the need of a detailed anatomical model. The approach is therefore dramatically extended, and since the relationship between image features and joint positions turns out to be linear, we can keep the computational burden within such limits that it can be used online.

The only attempt so far at modelling finger positions appears in [8], where significant differences among optical flow computations for finger flexion movements are reported, but not analyzed in detail. Optical flow [9] does not really seem the best feature choice in this case, since it is a derivative operator, hard to compute and prone to accumulating integral errors when applied to position recognition.

This paper can be seen as the follow-up to [10], in which the same approach was applied to the limited case of one human subject in extremely controlled conditions (namely, taking care that no probe/skin/skeleton relative movement would happen).

II. EXPERIMENT DESCRIPTION

A. Data Gathering Setup

1) Hand Motion: An 18-sensor right-handed Cyberglove (Cyberglove Systems[1], see also Fig. 2, left panel) is used to gather the finger positions. The Cyberglove is a light fabric,
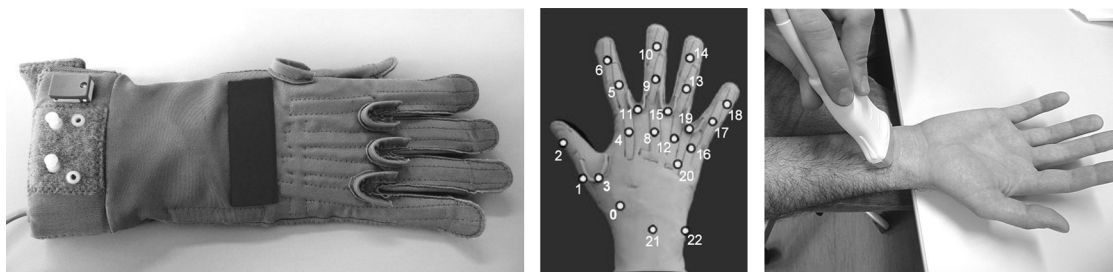
[1]http://www.cyberglovesystems.com

Fig. 2. Data capturing devices: (left to right) the Cyberglove; the location of its sensors (sensors 16, 12, 8, 4, 0, and 3 are used); the ultrasound transducer placed onto the subject's wrist. Moisture due to ultrasound conductive gel is clearly visible.

rather elastic glove, onto which 18 strain gauges are sewn; the sewing sheaths are chosen carefully by the manufacturer, so that the gauges exhibit a resistance which is proportionally related to the angles between pairs of hand joints of interest. The device can then return 18 8-bit values, proportional to these angles, for an average resolution of less than one degree, depending on the size of the subject's hand, a careful wearing of the glove and the rotation range of the considered joint. (For practical reasons, the subject must wear a cotton glove below the Cyberglove; we verified that this would not limit the precision of the device.)

We consider six hand motions, namely flexion/extension of the five fingers and thumb adduction/abduction. Thumb flexion/extension is roughly equivalent to thumb rotation, indeed a very important motion, characteristic of the high primates and paramount for most activities of daily living. The choice of these six motions is also motivated by the fact that they are enforced by the most advanced hand prosthesis of the world at the time of writing, namely the Vincent Hand (Vincent Systems GmbH[2]).

The above motions are captured by considering the five metacarpo-phalangeal glove sensors, placed where the proximal phalanxes of the fingers meet the palm, plus the thumb/index abduction sensor for the thumb abduction/adduction. For each subject we performed a simple calibration consisting of asking to stretch/bend the subject's joints to extreme positions, therefore recording each motion's extreme values. We then used these extreme values to normalize the sensor values between 0 and 1, so that 0 corresponds to the relaxed stance and 1 to the maximum voluntary contraction. (Values are nevertheless sometimes outside this range since nothing prevents the subjects to go beyond the limits while performing the required movements.)

According to the placement of the sensors on the Cyberglove (see Fig. 2, central panel), we choose sensors 16, 12, 8, 4, and 0 for the pinkie, ring, middle, index, and thumb flexion/extension, and sensor 3 for the thumb rotation. A careful hardware calibration enables us to obtain a resolution of 7–7.5 bits over the considered ranges, actually way below one degree in all cases. Approximate angle ranges and resolutions are reported in Table I.

The glove values are captured at 88 Hz, the maximum rate allowed by the serial port communication on the setup machine under Windows XP.

2) Ultrasound Imaging: US images are gathered using a General Electric *Logiq-e* portable ultrasound machine (see web-

TABLE I
APPROXIMATE RANGES (IN DEGREES AND IN CYBERGLOVE UNITS) AND RESOLUTIONS (IN DEGREES) OF THE SIX CONSIDERED FINGER MOTIONS

motion	rng[deg]	rng[cyb]	res[deg]
pinkie flexion	93	170	0.55
ring flexion	82	123	0.67
middle flexion	84	149	0.56
index flexion	88	183	0.48
thumb adduction	60	125	0.48
thumb rotation	51	153	0.33

site[3]) equipped with a 12L-RS linear transducer; the ultrasound "B"-mode is selected, resulting in a gray-valued image representing a section of what lies directly under the transducer. After an initial round of examinations, we chose the following settings: ultrasound frequency of 12 MHz, minimal onboard image preprocessing (i.e., noise rejection/edge enhancement), focus point at a depth of about 1.3 cm, minimum depth of field ("focus number" set at 1). This results in a frame rate of 28 Hz. These settings match those already used in [10].

The transducer is lightly but firmly fixed on the subject's wrist using a commercially available vise; this gives to the setup some flexibility, accounting for smaller and larger forearms. The transducer lies near the distal radioulnar articulation at the level of the *M. Pronator Quadratus* as depicted in Fig. 2, right panel. The output image (consider Fig. 1 again) is therefore a section of the wrist along the transverse plane, containing the ulna and radius, the *M. Pronator Quadratus* itself, and the tendons related to almost all flexor muscles.

Images generated by the US machine are captured from the VGA video output using a commercial VGA frame grabber, then sent via Ethernet to a standard PC.

B. Experimental Protocol

1) Subject Pool and Stimulus: Six subjects, all able-bodied, 22–39 years old, joined the experiment. Each subject would sit in front of a large PC screen, wear the glove and then lie her/his hand and part of the forearm relaxed on an orthopaedic support. The chair, the screen, the support, and the glove were adjusted prior to the experiment in order to obtain maximum comfort. The US transducer was then fixed just above and onto the wrist, tightly but comfortably. Standard ultrasound gel was applied between the transducer's head and the skin to allow the correct functionality of the US machine. The subjects were given no knowledge of what the experiment was about; they were simply

---

[2]http://www.handprothese.de/vincent-hand

[3]http://www.gehealthcare.com/euen/ultrasound/products/portable/logiq-e

Fig. 3. The experimental setup: the subject would mimic the hand-model movements, as seen on the computer screen; meanwhile, the glove and ultrasound machine would gather hand motions and US images.

instructed to perform with the right hand what the hand model on the screen would do, trying to mimic both the movement and its speed. The model is controlled using exactly the same six motion values at a real-time rate of 25 Hz. Fig. 3 shows the situation.

The stimulus (that is, what the hand model would actually do) consists of a sequence of basic movements, either single- or multi-finger. Single-finger movements are: pinkie, ring, middle, index, and thumb full flexion and back, and thumb full adduction and back. Multi-finger movements are: (a) simultaneous flexion of the pinkie and ring, (b) simultaneous flexion of the middle and index, (c) simultaneous flexion of the pinkie, ring, middle, and index, and (d) like (c) but also adducting the thumb, as in a typical "flat grasp," used to grasp credit cards or DVDs. Each movement is performed at three different speeds (1, 3, and 5 s for full flexion and back) and repeated two times (single-finger movements) or three times (multi-finger movements); in between movements, 1.5 s of rest are allowed. All in all, there are 72 movements (36 single- and 36 multi-finger ones); appropriate labels are applied to all samples in order to understand what movement and what speed is associated to each US frame and hand position. Each experiment lasted about 6 min (3 min single-, 3 min multi-finger movements); no fatigue or discomfort were reported by any of the subjects.

2) Synchronization and Preprocessing: Data synchronization is enforced on a Windows PC equipped with a multi-core processor, by gathering data from each device (i.e., the hand model, the glove, and the US machine) asynchronously and accurately time stamping each received datum. Time stamping is enforced by the HRT library [11], giving an experimentally-determined precision of 1.9 μs on the machine used. Linear interpolation is used to find the glove motion and stimulus values best corresponding to the time at which each image is received on the PC. All data are then low-pass filtered with a Butterworth fifth-order filter, cutoff frequency at 1 Hz. This relatively strong filtering was chosen after an initial round of experiments, and found to yield the best results. Actually, since visual features are linearly related to position features [10] and position features are

linearly related to finger joint angles (due to the linearity of the cyberglove), the bandwidth of both signals is directly limited by the speed of the movements (we have verified this for a relevant number of sample signals). As per the stimulus, movements are performed in cycles lasting 1, 3, 5 s, therefore the signal bandwidth is uniformly limited below 2 Hz, so that 1 Hz seems to be a reasonable value to cut off all high-frequency noise without introducing too much delay in the real-time prediction.

3) Image Grabbing and Validation: As the US machine we employed has no way of streaming images directly to a PC, images are grabbed from an external VGA connector using a commercial frame grabber as stated above. The US machine generates images at a rate which depends on the chosen B-mode settings; in our case, that was 28 Hz. These images are streamed through the VGA port at a resolution of 1024×768 at 60 Hz, as is customary. The frame grabber grabs the images at an unsynchronized rate of about 56 Hz (the precise rate cannot be fixed *a priori* as it depends on on-board software compression which cannot be disabled) and streams them in turn through an Ethernet connection.

Since the image stream is asynchronous, images must be checked for validity before being used as they are received by the system; i.e., one must verify that frame $t + 1$ is a complete, different frame with respect to frame t. This is done by checking the difference between two consecutively grabbed frames, evaluated as the sum of absolute differences of the gray levels. Three clearly increasing, repeatable noise levels were identified and associated to the following.

a) *Framegrabber noise:* the same ultrasound image is grabbed twice by the framegrabber. The frame is considered invalid and discarded, except in case c (see below).

b) *Ultrasound noise plus framegrabber noise* an update of the ultrasound image occurred; the frame has changed on the US since the last grab. The frame is considered valid.

c) *Tearing* due to the unsynchronized grabbing the top half of the image has been updated (type b above) whereas the bottom half has not (type a); the frame is then discarded and the next frame of type a is then scheduled for usage.

At the end, only the valid frames are retained. Inspection of the length of the recorded sequence has proved the effectiveness of the approach. Additionally, each frame is cropped to the portion of interest; that is, irrelevant information such as, e.g., the windows border, menus, etc., are discarded.

C. Data Processing and Analysis

1) Choice of the Image Features: Visual inspection of the US images while moving the fingers reveals that image deformations are highly localized according to anatomy and the position of the transducer. For instance, flexion of the pinkie finger appears as a local modification at the upper-left corner of the image, as that is the location where one would expect the section of the related tendon to appear. The choice of features is therefore that of *local* features—numbers extracted from a region of interest (ROI) centered around one of N interest points $p_i = (x_i, y_i)$, $i = 1, \ldots, N$.

Potentially, there are interest points in each and every image area where finger-related modifications appear. In this case, one must first identify the areas of interest for each finger and for
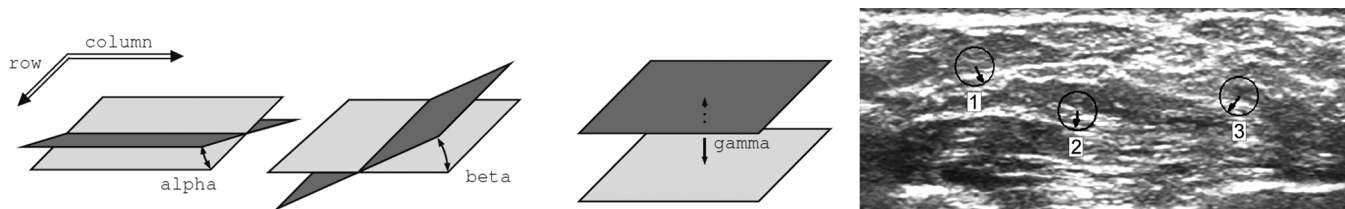
Fig. 4. (Left) A graphical representation of the three features α, β, γ extracted from each interest point. (right) Three ROIs located near the zones of the image where most changes are seen for pinkie flexion and thumb adduction (1,3) and where almost no movement is observed (2). The features are represented as normalized vectors where α and β are the vector components. Consider also the movie "features.avi," included in the supplemental material.

each subject, then estimate a potentially different shape for each ROI. We rather employed a simpler alternative: we chose a uniformly-spaced grid of interest points, and a circular area around each point. The center coordinates of each ROI, and the radius r, are chosen such that there is minimum overlapping among pairs of ROIs. Formally

$$\mathrm{ROI}_i = \left\{ (x, y) : (x - x_i)^2 + (y - y_i)^2 \leq r^2 \right\}.$$

Each ROI is the local axial section of a very complex 3-D object, namely the complete musculoskeletal structure of the wrist, which is made of soft tissue structures, adjacent and shifting along each other as muscles and tendons move. A full model of the motion seen in each ROI would require modelling carefully this 3-D motion and its view from across. This is clearly an extremely hard task, and anyway not in the focus of this work. We rather make some abstract considerations.

Firstly, local changes manifest themselves as alterations in the gray values inside the ROI, in particular as rotations, enlargements/shrinkings, shifts, combinations of these, and even less evident deformations due to the details of the musculoskeletal structure entering and exiting the section. This hints at some form of approximation of the gray value contents of the ROI; that is, for each p_i we extract a feature vector $\mathbf{f}(\mathrm{G}(\mathrm{ROI}_i))$ where $G(x, y)$ is the image gray level found at pixel (x, y) and, with a slight abuse of notation, $\mathrm{G}(\mathrm{ROI}_i)$ is the set of gray values found in the ith ROI.

Secondly, image changes seem related by visual inspection to positions of the fingers (and not, e.g., to velocities); this is intuitively sensible since all changes in the body structure due to voluntary contractions must be reversible when contraction is released. This hints at the use of temporal zeroth-order features, rather than first-order features such as, e.g., the optical flow (see, e.g., [9]), also since in this framework, this technique would mostly detect movement components in the x- and y-direction of the image plane, but not along the axis perpendicular to the section, which would be the main movement direction of muscles and tendons. As a differential measure, it can yield information on absolute position after integration, but indeed will be prone to integration errors (i.e., a random offset every time the fingers go back to the resting position).

For all these reasons, the features used here are, for each ROI, the local linear approximation of its gray-value contents. More in detail $\mathbf{f} = (\alpha_i, \beta_i, \gamma_i)$, where

$$\mathrm{G}(\mathrm{ROI}_i) \approx \alpha_i (x_i - x) + \beta_i (y_i - y) + \gamma_i$$

for all $(x, y) \in \mathrm{ROI}_i$. Intuitively, α_i denotes the mean image gradient along the x direction (rows of the image), β_i is the same value along the y (columns) direction, and γ_i is an offset. Fig. 4 (left panel) graphically represents the coefficients. In order to extract these features (and for all other image-related computations and evaluations) we used the HALCON v10.0 library by MVTec.[4]

2) A Toy Example: In order to test the effectiveness of the chosen features, they are first applied to a toy example. Namely, flexion of the pinkie and adduction of the thumb are selected as two very different movements, and repeatedly performed by one of the experimenters, in controlled conditions. Two corresponding ROIs are then selected, near the regions of the images in which most changes appear during either movement; a third ROI is chosen where almost no movement is noticed. Fig. 4 (right panel) shows a graphical representation of the α and β evaluated at the three ROIs, superimposed to a typical US image. The features are represented as rotating normalized vectors where α and β are the vector components.

Visual inspection of the three points, changing as the related motions are performed, reveals that the chosen features are well related to the finger positions (as recorded by the glove). Consider the movie "features.avi" included in the supplemental material; as one can see, ROI 1 (represented in red in the movie), clearly correlates with the flexion of the pinkie finger; ROI 3 (in green in the movie) correlates with the thumb position; and ROI 2 (in magenta in the movie) shows no apparent correlation with either movement.

This toy example confirms strong local correlation between finger positions and features extracted from each ROI, and proves in hindsight that the choice of local features is a good choice. In fact, an initial experiment revealed that high "correlation zones" exist between finger motions and features, as one can see from Fig. 5, showing the correlation coefficients between each ROIs and the six degrees of motion. (For the purpose of illustration, the correlation coefficients of the three features α_i, β_i, and γ_i have been averaged out for each feature point.) As one can see, specific areas of high correlation for the first four degrees of motion exist (flexion of the pinkie, ring, middle, and index finger) with highest correlation coefficients of about 0.7. In contrast to this, for the movements of the thumb, no feature points show very high correlation (for instance, the highest correlation coefficient for thumb adduction is about 0.4), but in such a case we can still hope that the *combination*
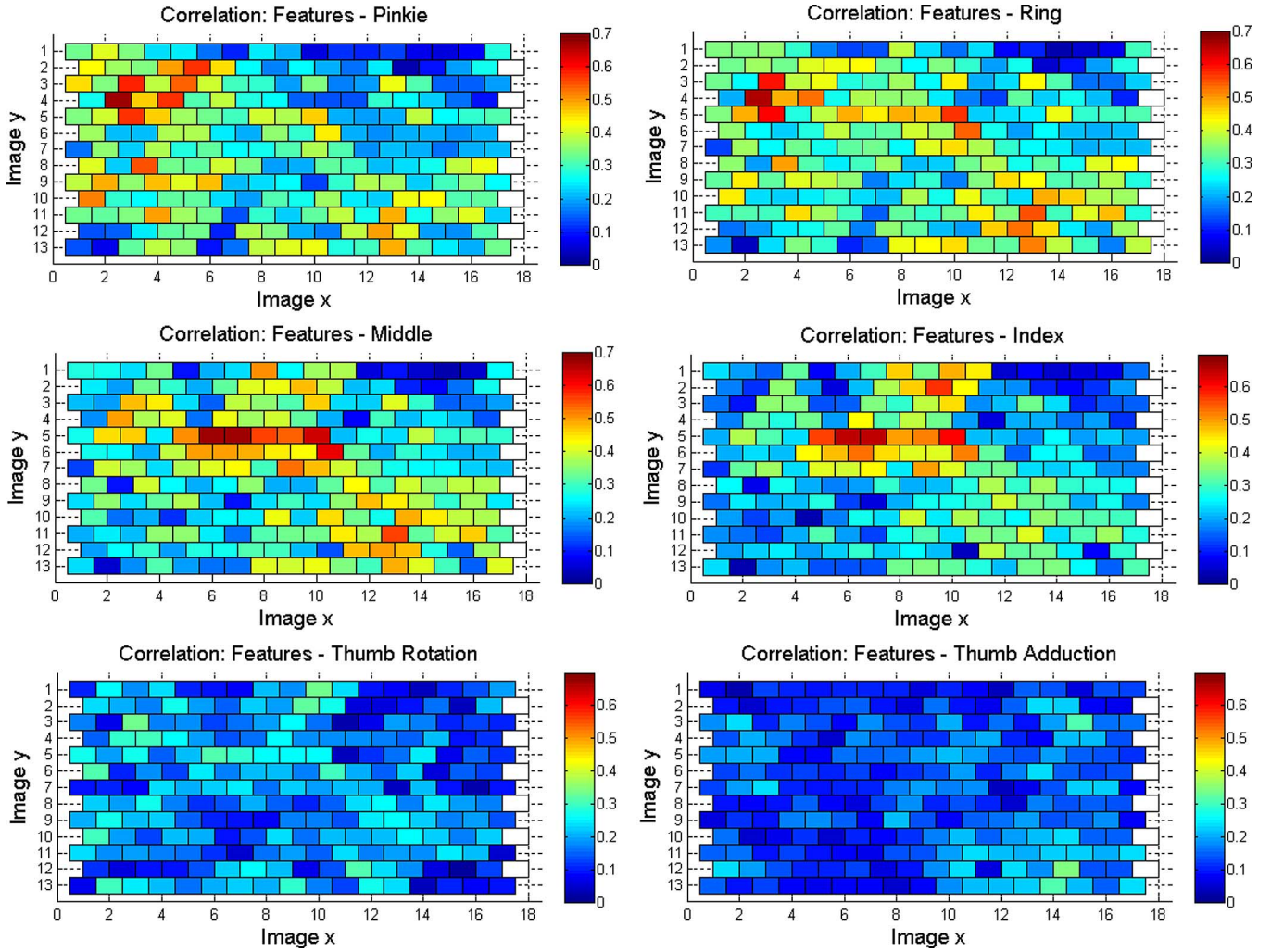
[4] http://www.mvtec.com/halcon

Fig. 5. Correlation between image features (average of α_i, β_i, γ_i for each ROI) and finger motions. Each colored rectangle matches the position of a ROI in the image, and the color denotes the correlation existing between the features evaluated at the related ROI and the finger motion of the subfigure title.

of all local features is effective to predict the thumb position. (See Section III for a full analysis.)

After an initial round of experiments, it was verified that the optimal setting is to have $N = 215$ ROIs, resulting in $215 \cdot 3 = 645$ visual features; the space between pairs of ROIs was chosen to be 50 pixels and the radius of each ROI, 20 pixels. As we have verified, the computation of all these features is extremely fast and can be done in cinema-like real-time, that is, 25 Hz, on the machine we used. As well, and somewhat surprisingly, we have verified that the number and distribution of the ROIs, and their radiuses, are not crucial to the prediction accuracy, as long as the related values are chosen reasonably, that is, in order for them to duly cover the surface of the US image and not to overlap too much.

3) Linear Regression: In [10], we showed that a linear relationship exists between the above-described gray level approximations and the finger positions (angles at the hand joints). So, for each frame an image feature vector $\mathbf{v} \in \mathbb{R}^{645}$ is associated to a finger position vector $\mathbf{p} \in \mathbb{R}^6$. Least-squares regression is applied to each dimension of the output space in order to obtain linear coefficients for the input space values. In other words,

for each degree of motion p_j with $j = 1, \ldots, 6$, we evaluate $k_1, \ldots, k_{645}$ with $k_i \in \mathbb{R}$ such that

$$p_j \approx \sum_{i=1}^{645} k_i v_i.$$

This procedure ends up in a 6×645 matrix K, which can further on be used to estimate new image feature vectors: $\mathbf{p} = K\mathbf{v}$. We employ the Matlab standard linear regression function.

In order to have an idea of the generality of this procedure, i.e., of how applicable this procedure is to features extracted from so-far-unseen images, we perform a form of cross-validation. For each subject, the collected data are first randomly permuted; then a subset of the data set (*training set*) is chosen to perform the estimation of K; the prediction error is evaluated on the rest of the data set (*testing set*). Different training set sizes are chosen, namely containing $1/n$ of the whole data set, where $n = 2, \ldots, 8$; it is expected that the prediction error would rise as n increases. No sample normalization is performed. This training/prediction cycle is repeated for 20 times, each time with a different permutation, then mean and standard deviation of

the obtained error rate are reported. As an error measure, we evaluate the square-root mean-square error normalized over the range of the target values (NRMSE).

4) Compensation of Transducer Motion: As is intuitively clear—and so it turned out after an initial round of experiments—the features $(\alpha_i, \beta_i, \gamma_i)$ are highly sensitive to the values inside the ROI: even a small shift in (x_i, y_i) can sensibly alter the features-to-position map and make K, as previously evaluated, much less effective. Small movements of the subject's forearm with respect to the transducer are inevitable, due to the gel, the normal motion of human limbs and fatigue. We compensate for this by evaluating the optical flow between reference frames and then using it as a map to locally shift the center of each ROI.

We consider as reference frames the images taken in the resting stance. During one session, composed of 72 movements, rest is reached 73 times. Related frames are identified by comparison with the stimulus plus a visually-determined time shift. The first resting frame is taken as reference; then for each further resting posture, the frame in the middle is considered for comparison. Optical flow is then evaluated and the feature points are shifted accordingly to the optical flow vector values.

One problem with the probe motion compensation is that, due to the shift of the feature points, some of them move out of scope during the session. In that case, these feature points transfer wrong information and are therefore not considered when estimating K. Hence, the algorithm has to be applied twice: the first time to evaluate all the features that transfer correct information through out the whole session, and the second time to extract the feature values from these feature points. This also means that when applying the probe motion compensation, some information is lost. In other words, the marginal feature points (on average 45 feature points; 135 features) are not taken into account when estimating K. The motion compensation can be switched on and off, enabling a full analysis of its effectiveness.

III. EXPERIMENTAL RESULTS

A. Regression Accuracy

Consider first Fig. 6, showing the NRMSE obtained by linear regression on each finger position, according to the size of the training set and for each subject. It is apparent from the figure that linear regression is highly effective in all cases. When the training set size is one-half of the total data set ($n = 2$ in the figure), the NRMSE ranges from 1.08% (middle finger, subject 3) to 2.65% (pinkie finger, subject 1). As expected, as the training set size is reduced, the NRMSE increases to a maximal value of 9.67% (pinkie finger, subject 1 again) when $n = 8$. The increase looks superlinear. Notice that there is no "best" finger overall. Fig. 7 shows comparisons between typical real and predicted finger positions.

B. Probe Motion Compensation

Consider now Fig. 8, showing the effect of the probe motion compensation mechanism on the NRMSE. We choose here to show some of the worst cases of Fig. 6, that is Subjects 1 and 2 when $n = 8$, and then to repeat the experiment with an even smaller training set size, namely $n = 12$.
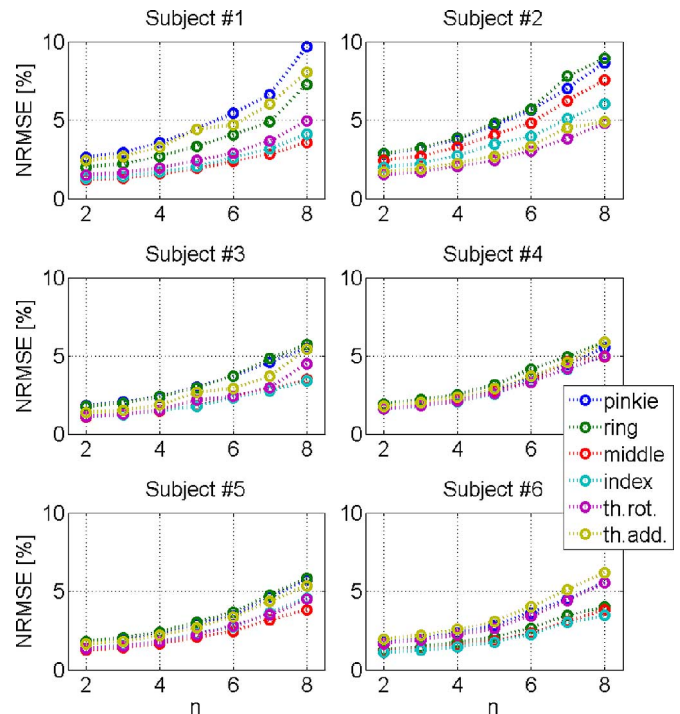

Fig. 6. Per subject accuracy of linear regression. Mean values obtained over 20 random permutations of the training set.

As is apparent, the use of probe motion compensation is highly effective, reducing the NRMSE by a statistically significant amount, especially in the cases when the error is high. For example, considering the pinkie flexion when $n = 12$, the NRMSE of subject 1 drops from 35.82% ± 9.94% to 13.54% ± 3.76%; for subject 2 the drop is from 30.82% ± 7.26% to 13.65% ± 4.22%. Student's t-test yields $p < 0.01$ in both cases.

C. Implementation

The above described system has been demonstrated using a three dimensional, 22 DoFs graphical model of a human hand, both to show the stimulus to the subject and to show the prediction. The model bone structure is configured using the six estimated finger positions and a simple nonlinear coupling among the internal DoFs; it is then rendered using the free utility Blender (The Blender Foundation[4]). The six values are streamed to the model by the prediction system using a local host-based UDP stream at 25 Hz.

The whole sequence of movements was administered to a single, intact subject while the system would gather synchronized joint angles and US images. Subsequently, the matrix K was estimated and then used to predict the hand configuration from the images alone. The result is visible in the movie "demo.avi," included in the supplemental material. At the end of the demo the subject reported a remarkable feeling of ownership of the model, while no fatigue or discomfort was detected. Notice that, in this case, the probe is not placed on the wrist but on the ventral side of the forearm, at a medial distance between the wrist and the elbow. (See the Section IV for more details on the probe positioning.)
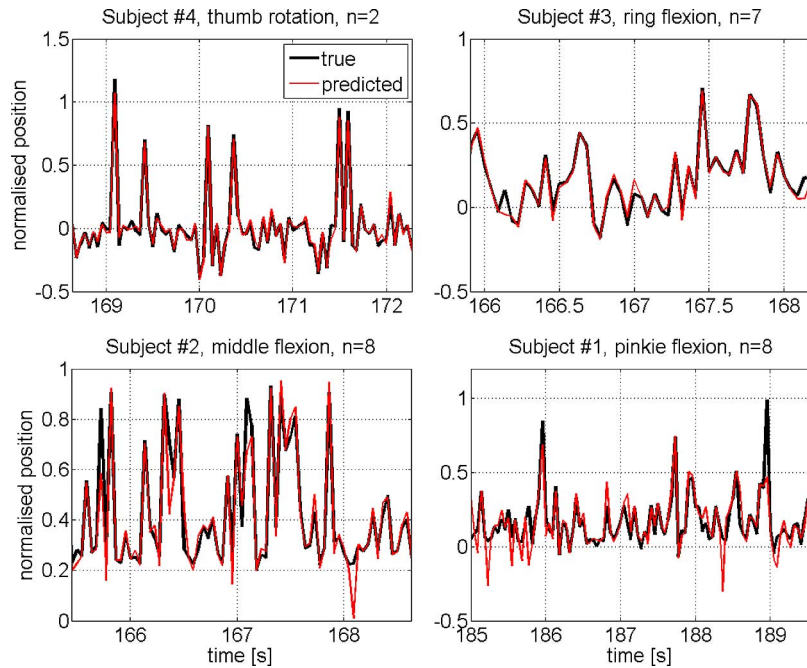
[5]http://www.blender.org

Fig. 7. Typical example comparisons between true and predicted finger positions. Notice the higher errors (for instance for Subject 1 around $t = 4080$ samples) as n is increased.
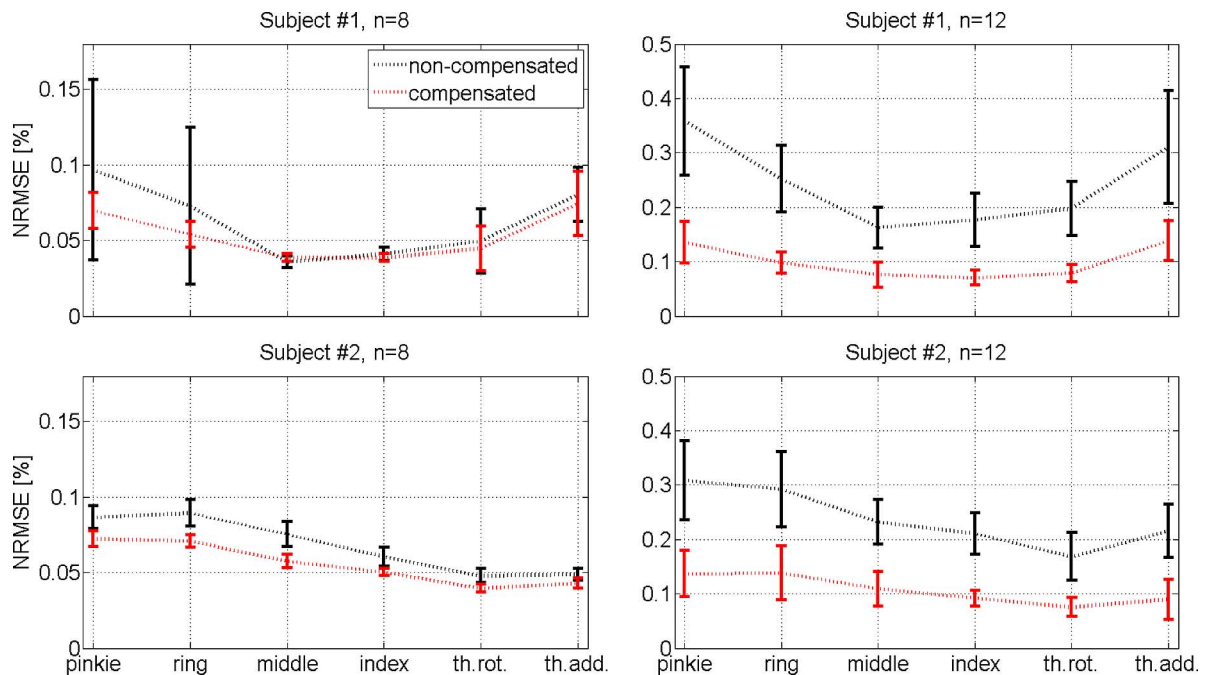

Fig. 8. Effect of the motion compensation on typically high errors. Error bars denote the mean and standard deviation NRMSE obtained on the 20 permutations.

IV. CONCLUSION

A. Discussion

The experimental results shown above clearly indicate that US imaging of the human wrist can be effectively used as a means to reconstruct the hand posture, up to a remarkable precision. This is the main finding of this work. More precisely: it is no surprise that the information required to reconstruct the hand configuration is in the images, since US is a widespread tool for medical diagnosis of hand conditions (and, in fact, for a

number of other conditions). What remained to be demonstrated is that the information can be effectively extracted and used to this aim. We claim that the answer to this question if affirmative.

A further, more surprising outcome of this experiment is that, if the type of features we use is chosen, then there is a *linear* relationship between them and the hand posture (finger joint angles). This result was already published in [10], but here we show that it holds uniformly for all subjects considered; it holds when subsampling is applied; and the relationship is robust with respect to relative motion between the subject's forearm and the

US transducer—in that case we demonstrate the usefulness of a simple compensation technique based upon optical flow computation. (Of course, too much motion will disrupt the approach.)

In our opinion, this is not trivial at all even though the features are themselves first-order approximations of the gray levels in the ROIs. Indeed, such features encode local structural motion as seen in a section of the forearm, but it is not clear why this motion is linearly related to finger position. Actually, in order to fully understand this phenomenon, a model of the musculoskeletal structure, its motion and the view across a section of it would be required, but this is not the focus of this paper. Notice that our approach is totally blind to anatomy: features are the same for each ROI, and ROIs are chosen to be uniformly distributed across the image.

The most interesting consequence of the linearity of this relationship is that the whole system can be implemented online (ongoing research). The only problematic point seems that of optical flow calculation, but that is required only whenever a new reference frame is detected, that is, not in real-time.

B. Future Work

The main problem so far encountered is the forearm/transducer motion, which eludes the compensation mechanism in extreme cases; but this problem should be largely solved by improving the setup—especially, a more apt transducer is required. For example, the transducer might be steadily fixed on the forearm via a belt; or it might be worn as a glove (silicone liner) with single transducers sewn atop, placed according to optimal muscular activity. We actually have hints that B-mode sonography might not be required at all, and this would greatly simplify the design of a new generation of transducers. Research in this direction is already on the way.

A more interesting point is that of finding features which are less sensitive to small changes in the US image. So far we use Least-Squares regression to estimate K, but this could be made more robust by employing some for of regularized regression such as, e.g., regularized least squares [12], [13]. An orthogonal direction to be explored is that of *allowing* for more transducer motion, therefore probably breaking the linearity of the map, and then using more involved machine learning methods rooted in the probabilistic framework (e.g., support vector machines [14]). A further very desirable characteristic is that of enforcing online learning, which implies in the linear case the periodic recomputation of K as the map changes, due to, e.g., transducer motion or new movements enforced by the subject. In that case a sensible way ahead is represented, e.g., by incremental regularized least squares, an approach which does not depend on the number of acquired samples (i.e., it does not grow indefinitely), but only on the dimension of the input space.

Further experiments have revealed that within reasonable limits, the radius of each ROI does not affect the accuracy of the system; on the other hand, there is a clear trade-off between the *number* of ROIs and the error. Finding a minimum set of ROIs is of course an interesting path ahead; ROIs could also be determined by local correlation with single-finger movements, and then some form of source separation could be used to enforce composition of these movements into multi-finger motions.

C. Applications

The application of the results shown in this paper are multifarious and potentially quite wide. Ultrasound imaging could be used as a means of interaction in virtual reality as well as in real environments, to control a mechanical hand or a model, leading to very accurate teleoperation, performed without any sensor on the subject's hand (such as, e.g., an instrumented glove).

An even more interesting application, though, is in rehabilitation of neuropathic pain patients, mainly upper-limb amputees and complex regional pain syndrome patients. Neuropathic pain is felt in the missing or impaired limb and has a so-far unknown origin; according to neurological studies performed, among others, by Flor [17], [18] and Maihöfner [19], this kind of pain is related to the degree of cortical reorganization happening in the brain as a consequence of the sensorial feedback loop breakdown implied by such diseases. As first shown by Ramachandran [20] and confirmed more recently [21], one promising path ahead is represented by *mirror therapy*, in which a mirror is used to give the patient the illusion of a restored limb, therefore closing back the sensorimotor loop, albeit in a very incomplete fashion.

During mirror therapy the patient must necessarily perform with the imaginary limb what (s)he performs with the intact limb. Apparently this simple exercise can ease the pain and even, in some cases, awaken the phantom limb. Our system could then be used to show to a patient the imaginary limb, this time moving freely and voluntarily. Employed in a virtual reality scenario, or with a real robotic hand, the system could give the patient an almost perfect illusion and feeling of immersion, and constitute a better treatment of neuropathic pain.

Of course this idea relies on the assumption that enough residual activity can be found in the stump/impaired limb. Hints at a positive answer to this question exist [22]–[25] although so far only the electromyographic signal has been studied to this end. Given the enormously richer amount of information carried by the ultrasound images with respect to electromyography, we conjecture that this would be the case with our system, too.

REFERENCES

[1] R. S. C. Cobbold, *Foundations of Biomedical Ultrasound*, ser. Biomedical Engineering. New York: Oxford Univ. Press, 2007.

[2] P. L. Cooperberg, I. Tsang, L. Truelove, and W. J. Knickerbocker, "Gray scale ultrasound in the evaluation of rheumatoid arthritis of the knee," *Radiology*, vol. 126, pp. 759–763, 1978.

[3] L. De Flaviis, P. Scaglione, R. Nessi, R. Ventura, and G. Calori, "Ultrasonography of the hand in rheumatoid arthritis," *Acta Radiol.*, vol. 29, pp. 457–460, 1988.

[4] G. A. W. Bruyn and W. A. Schmidt, *Introductory Guide to Musculoskeletal Ultrasound for the Rheumatologist*. Houten, The Netherlands: Bohn Stafleu & Van Loghum, 2006.

[5] Y. Zheng, M. Chan, J. Shi, X. Chen, and Q. Huang, "Sonomyography: Monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis," *Med. Eng. Phys.*, vol. 28, pp. 405–415, 2006.

[6] X. Chen, Y.-P. Zheng, J. Y. Guo, and J. Shi, "Sonomyography (SMG) control for powered prosthetic hand: A study with normal subjects," *Ultrasound Med. Biol.*, vol. 36, no. 7, pp. 1076–1088, 2010.

[7] J.-Y. Guo, Y.-P. Zheng, L. P. J. Kenney, A. Bowen, D. Howard, and J. J. Canderle, "A comparative evaluation of sonomyography, electromyography, force, and wrist angle in a discrete tracking task," *Ultrasound Med. Biol.*, vol. 37, no. 6, pp. 884–891, 2011.

[8] J. Shi, S. Hu, Z. Liu, J. Guo, Y. Zhou, and Y. Zheng, "Recognition of finger flexion from ultrasound image with optical flow: A preliminary study," in *Proc. Int. Conf. Biomed. Eng. Comput. Sci. (EMBC)*, Wuhan, China, 2010, pp. 1–4.

[9] B. K. P. Horn and B. G. Schunk, "Determining the optical flow," *Artif. Intell.*, vol. 17, pp. 185–203, 1981.

[10] C. Castellini and G. Passig, "Ultrasound image features of the wrist are linearly related to finger positions," in *Proc. IROS—Int. Conf. Intell. Robots Syst.*, 2011, pp. 2108–2114.

[11] J. Nilsson, "Implementing a continuously updating, high-resolution time provider for windows," *MSDN Mag.* 2004 [Online]. Available: http://msdn.microsoft.com/en-us/magazine/cc163996.aspx

[12] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models*, ser. Monographs Stat. Appl. Probabil.. London, U.K.: Chapman & Hall, 1994, ch. 58.

[13] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," *Adv. Computat. Math.*, vol. 13, no. 1, pp. 1–50, 2000.

[14] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. ACM Workshop Computat. Learn. Theory*, D. Haussler, Ed., 1992, pp. 144–152.

[15] V. Vovk, "Competitive on-line statistics," *Int. Stat. Rev.*, vol. 69, pp. 213–248, 2001.

[16] K. S. Azoury and M. Warmuth, "Relative loss bounds for on-line density estimation with the exponential family of distributions," *Mach. Learn.*, vol. 43, no. 3, pp. 211–246, 2001.

[17] H. Flor, T. Elbert, S. Knecht, C. Wienbruch, C. Pantev, N. Birbaumers, W. Larbig, and E. Taub, "Phantom-limb pain as a perceptual correlate of cortical reorganization following arm amputation," *Nature*, vol. 375, no. 6531, pp. 482–484, 1995.

[18] H. Flor and N. Birbaumer, "Phantom limb pain: Cortical plasticity and novel therapeutic approaches," *Current Opin. Anaesthesiol.*, vol. 13, pp. 561–564, 2000.

[19] C. Maihöfner, R. Baron, R. DeCol, A. Binder, F. Birklein, G. Deuschl, H. O. Handwerker, and J. Schattschneider, "The motor system shows adaptive changes in complex regional pain syndrome," *Brain*, vol. 130, no. 10, pp. 2671–2687, 2007.

[20] V. S. Ramachandran, D. Rogers-Ramachandran, and S. Cobb, "Touching the phantom limb," *Nature*, vol. 377, no. 6549, pp. 489–490, 1995.

[21] B. L. Chan, R. Witt, A. P. Charrow, A. Magee, R. Howard, P. F. Pasquina, K. M. Heilman, and J. W. Tsao, "Mirror therapy for phantom limb pain," *N. Eng. J. Med.* vol. 357, no. 21, pp. 2206–2207, 2007.

[22] F. C. P. Sebelius, B. N. Rosén, and G. N. Lundborg, "Refined myoelectric control in below-elbow amputees using artificial neural networks and a data glove," *J. Hand Surg.*, vol. 30A, no. 4, pp. 780–789, 2005.

[23] F. V. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, and N. V. Thakor, "Decoding of individuated finger movements using surface electromyography," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1427–1434, May 2009.

[24] C. Castellini and P. van der Smagt, "Surface EMG in advanced hand prosthetics," *Biol. Cybern.*, vol. 100, no. 1, pp. 35–47, 2009.

[25] C. Cipriani, C. Antfolk, M. Controzzi, G. Lundborg, B. Rosen, M. Carrozza, and F. Sebelius, "Online myoelectric control of a dexterous hand prosthesis by transradial amputees," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 3, pp. 260–270, Jun. 2011.

[26] F. H. Netter, *Atlas der Anatomie des Menschen*, 3rd ed. Stuttgart, Germany: Thieme, 2006.

Claudio Castellini (M'XX) received the Laurea in biomedical engineering from the University of Genova, Genova, Italy, in 1998, and the Ph.D. degree in artificial intelligence from the University of Edinburgh, Edinburgh, U.K., in 2005.

He then spent 4.5 years as a postdoctoral researcher in the Advanced Robotics Laboratory of the University of Genova, Genova, Italy, working on machine learning applied to human sensorimotor data. Since 2009 he is a researcher at the DLR (German Aerospace Center) in the Bionics group. He is (co)author of more than 30 peer-reviewed papers.

Georg Passig received an electrical engineering diploma at the Technische Universität München, München, Germany, in 1999, and continued as a researcher at the Institute of Realtime-Computersystems finishing with the Ph.D. degree in robotics and image processing, in 2005.

He spent six years as a postdoctoral researcher at the DLR (German Aerospace Center) developing the Mirosurge surgical robot and has started in 2011 as a Professor for Mechatronics and Image Processing at the University of Applied Sciences Ingolstadt.

Emanuel Zarka received the B.Sc. degree in mechatronics and robotics and the M.Sc. degree in mechatronics and robotics from the University of Applied Sciences FH Technikum Wien, Wien, Austria, in 2009 and 2011, respectively. This work is part of his M.Sc. thesis, completed at the DLR (German Aerospace Center) in the Bionics Group.