2014 Joint IEEE International Conferences on Development and Learning
and Epigenetic Robotics (ICDL-Epirob)
October 13-16, 2014. Palazzo Ducale, Genoa, Italy

Correcting Pose Estimates during Tactile Exploration of Object Shape: a Neuro-robotic Study

Claudius Strub*†, Florentin Wörgötter†, Helge Ritter‡ and Yulia Sandamirskaya*
*Institut für Neuroinformatik, Ruhr-Universität Bochum, 44780 Bochum, Germany
Email: claudius.strub@ini.rub.de, sandayci@rub.de
†Department of Computational Neuroscience, III Physics Institute, Georg-August-Universität, 37077 Göttingen, Germany
Email: worgott@gwdg.de
‡Center of Excellence Cognitive Interaction Technology, Bielefeld University, 33619 Bielefeld, Germany
Email: helge@cit-ec.uni-bielefeld.de

***Abstract*—Robots are expected to operate autonomously in unconstrained, real-world environments. Therefore, they cannot rely on access to models of all objects in their environment, in order to parameterize object-directed actions. The robot must estimate the shape of objects in such environments, based on their perception. How to estimate an object's shape based on distal sensors, such as color- or depth cameras, has been extensively studied. Using haptic sensors for this purpose, however, has not been considered in a comparable depth. Humans, to the contrary, are able to improve object manipulation capabilities by using tactile stimuli, acquired from an active haptic exploration of an object. In this paper we introduce a neural-dynamic model which allows to build an object shape representation based on haptic exploration. Acquiring this representation during object manipulation requires the robot to autonomously detect and correct errors in the localization of tactile features with respect to the object. We have implemented an architecture for haptic exploration of an object's shape on a physical robotic hand in a simple exemplary scenario, in which the geometrical models of two different n-gons are learned from tactile data while rotating them with the robotic hand.**

I. Introduction

Object manipulation capabilities are critical for robotic systems which are designed to assist humans in their everyday environment. Manipulation of common objects is a complex sensorimotor task, requiring coordination between the motor system of the hand and the sensory feedback, both proprioceptive and tactile. These coordination patterns between sensory inputs and motor control variables depend on the object manipulated by the robot and its particular shape. In many real-world cases, it is infeasible to pre-compute these coordination patterns or object shape models in advance. This would require to consider every possible object which a robot may encounter in an unconstrained human environment. Therefore, an autonomous robot needs a capability to acquire such models based on its own sensory information. Typically, vision is used to estimate the object shape for grasping and object manipulation (e.g., [16]). This estimation is often prone to errors inherent in the process of visual object segmentation (e.g. occlusions, distractors, complex background). Tactile exploration of object shape could provide an additional source of information about the geometric parameters of the object, relevant for object manipulation.

The tactile feedback is often used in robotic systems to determine contact with the object and to control the force of the grasp [15]. Estimation of object shape based on haptic feedback has been considered in robotics only in constrained settings, using one of two simplifications in the haptic learning process: (1) a rigidly mounted object is used for learning its shape and geometry, e.g. [5], [12]; (2) haptics is used to localize objects, the shape of which is assumed to be known [3], [14].

In this paper, to the contrary, we study haptic learning of object shape in a less constrained setting. Here, both shape and pose of the object are initially unknown and have to be estimated while the object is autonomously manipulated by the robot. During haptic exploration in such an unconstrained setting, there is a need to estimate and correct for the errors in the localization of detected features with respect to the object. The pose of the object changes during learning due to the physical interaction between the hand and the object. These changes in the object pose are only partly intentional, as grasping and releasing of the object as well as slippage induce unintended and uncontrolled object movements. Small errors in the pose estimate accumulate over time, which calls for an error correction mechanism to prevent drift in the subsequent object shape estimate. Thus, the haptic learning problem becomes equivalent to the well-known simultaneous localization and mapping (SLAM) problem in robotic navigation [2], [6].

There have been several biologically motivated approaches to the SLAM problem, e.g. [4], [11], [13]. Several challenges make most of the approaches to SLAM in navigation not applicable to the equivalent problem in haptic learning, however. First, in tactile SLAM, the sensor information is only present during periods of object contact in contrast to the typical continuously available information of distance sensors (e.g. sonar, infra-red, or laser). This leads to only very sparsely distributed information in space and time, comparable to solving SLAM in navigation by only using the bumpers of a robot. Second, the tactile sensor data typically does not have spatially distinct features. Surface curvature, edges, and texture are often repetitive in space. Hence, the capabilities of computing, detecting, matching and tracking unique or salient features, which serve as landmarks, are very restricted [8].

978-1-4799-7540-2/14/$31.00 ©2014 IEEE

In this paper, we explore haptic learning in a simple robotic scenario and suggest a neural-dynamic model, which realizes a solution to the haptic SLAM problem. We derive inspiration for our model from what is known about how humans and primates use haptics to determine the pose and shape of objects [10], [21]. In particular, in our model we use biologically plausible tactile features and Dynamic Neural Fields – a framework proposed for both, modeling human cognition [22] and dynamics of neuronal populations [24]. This provides an opportunity to use and refine the same model to account for human behavioral data on haptic learning, when such data will be available.

In the experimental setting, we use two fingers of a robotic hand to rotate different n-gon objects and build the representation of the object shape online, in an incremental fashion. The robot autonomously learns the shape of the object, while simultaneously estimating errors in the moment-to-moment localization of tactile features on the object surface. In particular, the system builds on a recently introduced model for haptic learning which is able to correct errors in the orientation of the localized object pose [23]. Here, this model is extended to additionally enable the correction of shifts in the object position, directed lateral to the grip.

II. Overview of the Architecture

This section gives a brief overview of the developed architecture for learning object shape representations based on tactile inputs. The architecture (depicted in Fig. 1) has to realize a combined mapping and localizing of the object from self-generated motions. To this end, an *Object Manipulation Behavior* drives an interactive exploration loop, which constitutes a sequence of reactive behaviors controlling the movements of the robot's fingers. A **Estimation Path (EP)** processes the generated tactile and proprioceptive data to estimate the change of the object pose. Allocentric features are extracted for the **Mapping Path (MP)** that maintains the object model, and for the **Localizing Path (LP)** that maintains the object pose. They both are controlled by the *Matching Module* which constitutes the core element of our architecture: it uses Dynamic Neural Fields (DNF, see Sec. III below) to compare the sensed features with the current object representation in order to decide whether an update of the pose (LP) or the object shape (MP) is appropriate.

In the **EP** the location of the contact blob on the tactile sensor is transformed to an allocentric frame of reference using the *Forward Kinematics* and proprioception. Two consecutive contact locations are then used in order to estimate the change in the object pose, induced by the robot's action. The estimated pose changes are integrated into an *Object Pose Estimate* in Fig. 1. The first grasp initializes the estimate of the object position, assumed to be in the middle of the two contact points of the fingers.

The **MP** extracts allocentric *3D Features* from the tactile pressure patterns, computed via the *Forward Kinematics* and proprioception. These features are then *Transformed* into an object-centered coordinate system using the current *Object Pose Estimate*. The *Matching Module* compares the current features with previous ones, stored in the *Object Representation* in order to detect and compensate for drifts in the representation. Finally, the corrected features are fused into the accumulating *Object Representation*.

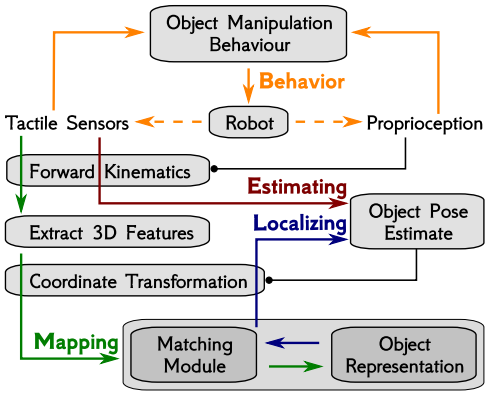

Fig. 1. Overview of the general architecture: black lines with a dot at the end indicate a parametrization relationship. Arrows pointing to boxes indicate inputs for processing in the box, and arrows passing through boxes indicate a parametrized transformation of the information through the box.

The **LP** uses the detected drifts of the the *Matching Module* to correct the *Object Pose Estimate* in 3D external (robot-centered) space, i.e. tracking the object.

The central part of this architecture is the *Matching Module* in Fig. 1, which implements the error estimation based on current tactile features and the object representation. Estimating the error requires to decide whether the current feature corresponds to a new haptic "landmark" and should lead to an adaptation of the object shape representation (i.e. mapping). Alternatively, the input may correspond to a previously mapped, but incorrectly localized object feature, which would require an adaptation of the pose estimate in order to reduce the feature localization error. This is essential for ensuring a consistent mapping and localization of the object, which is the core problem to be addressed by SLAM, and can only be solved to the degree of object ambiguity. The objects used in the experiments, as well as many everyday objects, are symmetric and repetitive in their appearance. Hence, the localization of the object is necessarily ambiguous, i.e. the same features may be detected for multiple object poses.

III. The Model

In this Section, we describe the neural-dynamic model implementing the combined matching module and object representation (bottom of Fig. 1), which is crucial for the localization and mapping of the object. First the extraction of features serving as input to the model is described, followed by a detailed description of the model itself. The model is evaluated in robotic experiments, described in Section IV.

A. Computing Features from Tactile Inputs

The proposed model is inspired by neural processing of shape by haptic pathways in monkeys and humans, which are

highly interleaved with visual processing pathways [9], [21]. Prominent features in both systems are zero-, 1st, and 2nd order moments, which correspond to position, orientation, and curvature of the tactile contact, respectively. These features are therefore used as inputs to our model. Note, that the normal of the object surface does not necessarily coincide with the normal of the sensor surface, due to the rigid fingers. Curvature is modeled by the eigenvectors and eigenvalues of the covariance of the tactile "pressure blobs", along with the angle of the first eigenvector. All these features are associated with their location in a 3D external coordinate system, computed using the forward kinematics. In the current setup, the induced object motion is restricted to one degree of freedom: rotation along it's z-axis. Nevertheless, there is an initial error in the estimate of the object's position which the model needs to compensate for.

Further, the perceived tactile features are transformed from external into an object-centered reference frame by shifting and rotating the features' spatial representation according to the current estimation of the object's pose. These features with their locations in an object-centered reference frame are used as inputs to the model, implemented using DNFs.

B. Dynamic Neural Fields (DNFs)

The basic dynamical components in our model are Dynamic Neural Fields (DNFs), see [1], [24] for a neural derivation and the analysis of dynamics. DNFs have been used in cognitive science to model dynamics and development of cognitive processes, such as, e.g., memory formation, decision making, or categorization [20], as well as to integrate the low-level sensory inputs and motor dynamics into cognitive architectures, e.g., for scene representation, sequence generation, and grounded language [18].

The DNFs used in our model are described by Eq. (1), which defines the rate of change in activation $u(x,t)$ of the field:

$$\tau \dot{u}(x,t) = -u(x,t) + h + S(x,t) + \int f\big(u(x',t)\big)\omega(|x-x'|)dx'. \quad (1)$$

In Eq. (1), $u(x,t)$ is the activation of the DNF at time step t and position x. The position x describes a behavioral variable, such as a perceptual feature, location in space, or motor control variable and may be multi-dimensional: $\vec{x} \in \mathbb{R}^n$. The activation $u(x,t)$ can be interpreted as the confidence of value x for this behavioral variable in the current state.

The term $-u(x,t)$ stabilizes an attractor for the activation function at values, defined by the last three terms in the equation. The time constant τ determines how fast activation $u(x,t)$ relaxes to the attractor. The negative resting level h ensures that the DNF produces no output in a deactivated state and $S(x,t)$ is an external input, driving the DNF. The convolution term models lateral interactions between sites of an active DNF, shaped by the interaction kernel, $\omega(|x-x'|) = c_{exc}\exp\left[-\frac{(x-x')^2}{2\sigma_{exc}^2}\right] - c_{inh}\exp\left[-\frac{(x-x')^2}{2\sigma_{inh}^2}\right]$, with a short-range excitation (strength c_{exc}, width σ_{exc}) and a long-range inhibition (strength c_{inh}, width σ_{inh}). A sigmoidal non-linearity, $f\big(u(x,t)\big) = \frac{1}{1+\exp[-\beta u(x,t)]}$ defines the output of the DNF with which the DNF impacts on other dynamics in the model, as well as on its own dynamics through the lateral interactions.

The lateral interactions of DNFs stabilize a localized peak-attractor for the activation function, i.e. even for a noisy and varying input, the DNF "stabilizes a decision" for the most active peak position, leading to discretization of continuous sensory and motor spaces.

To build a long-term memory of the object's shape, we use memory trace dynamics, Eq. (2), [17]:

$$\tau_l \dot{P}(x,t) = \lambda_{build}\big(-P(x,t) + f(u(x,t))\big)f(u(x,t)) - \lambda_{decay}P(x,t)\big(1-f(u(x,t))\big). \quad (2)$$

Here, $P(x,t)$ is the strength of the memory trace at site x of the DNF with activity $u(x,t)$ and output $f\big(u(x,t)\big)$, λ_{build} and λ_{decay} are the rates of build-up and decay of the memory trace. The build-up of the memory trace is active on sites with a high positive output $f\big(u(x,t)\big)$, the decay is active on the sites with a low output.

DNFs are used in our model for the following functions: to represent the currently perceived haptic features; to match these features to the accumulated long-term memory of the object shape (implemented by memory traces); and to compute errors in the pose estimation, thus stabilizing the object-centered shape representation.

C. Stabilization of the shape representation

Here the networks of DNFs in the proposed model are described, which implement the matching module combined with the object representation for mapping and localizing of the object. To stabilize the object shape representation against drift during the object manipulation, three dynamic networks process in parallel the input features described in III A, shown in Fig. 2. One network stabilizes the rotation estimate using the contact normal information of both fingers, shown on the left side of the figure. Additionally, for each finger a separate network stabilizes the translation by using the contact position information of the corresponding finger. The networks for the translation stabilization of each finger are identical and therefore only one of them is shown on the right side of Fig. 2.

In all networks, tactile features are processed in separate mapping and localization pathways: the localization pathway holds current information of the perceived haptic features and the mapping pathway with slower dynamics accumulates a long-term memory of past inputs. The latter serves as the object shape representation and is used for matching with the current features of the localization pathway, in order to correct for errors in the pose estimate (i.e. object localization).

a) The mapping pathways: The mapping pathways, depicted by the gray-shaded area in Fig. 2, operate with slow time constants. In the **rotational network**, a short time-window of recent contact normals of both fingers serves as

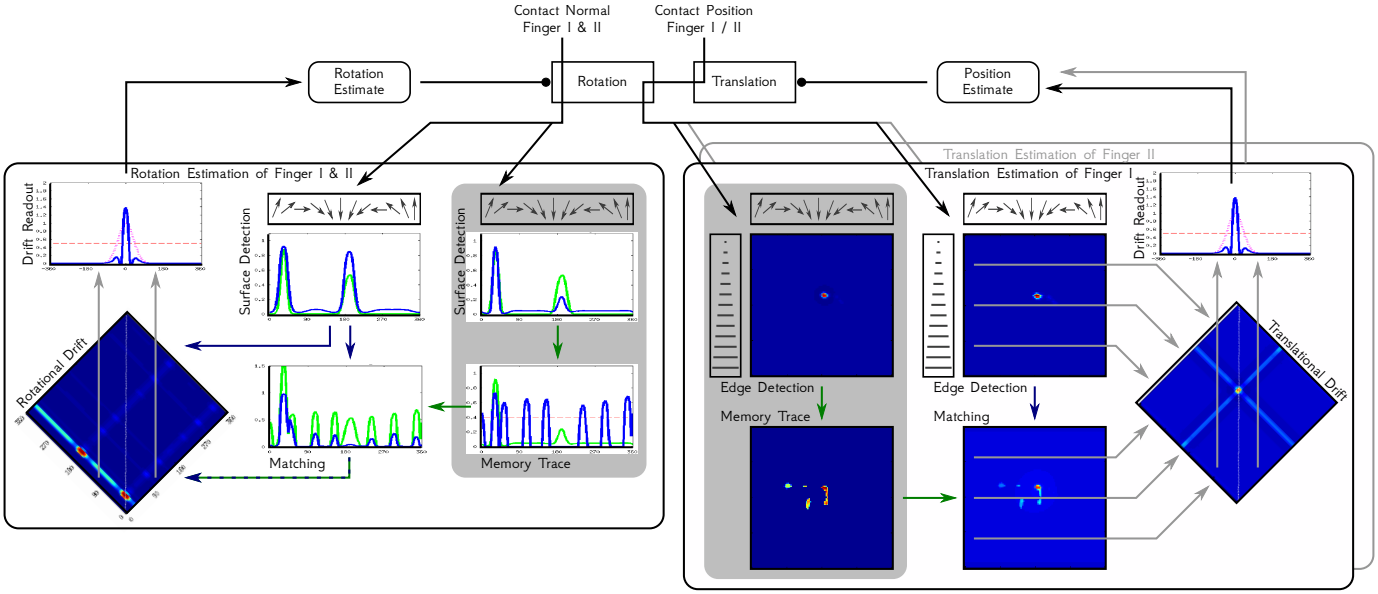

Fig. 2. Overview of the Model - Left: Orientation of the surface normal is the input for a mapping (dark-green arrows) and a localization (dark-blue arrows) pathway, consisting of multiple DNFs. Mapping in the gray underlaid DNF and MT corresponds to the object representation and operates with a slower time constant. Localization utilizes the memory and outputs a correction term for the estimated object rotation. 1D plots show input (green) and activation (blue) of the DNFs, thresholds are marked by dashed red lines. 2D plots only show the DNF output. Right: contact position is represented in 2D DNFs with polar coordinate encoding on the axis. Just like in the rotational case (left), there is a localization and a mapping pathway with memory, which are compared to determine the drift in translation.

input to a 1D orientation selective field. As the finger surface is curved, flat object surfaces lead to more normals with the same orientation, while object edges lead to a continuous change in the orientation of the normals. Therefore, the distribution of contact normal orientations is broad and weak for contacts with an object edge and is sharp and strong for contacts with flat surfaces. Hence, the orientation selective *Surface Detection* DNF generates a stabilized peak for a detected surface, which is used as input to the *Memory Trace* dynamics. The memory trace (Eq. (2)) is an exponentially fading memory of past peaks, i.e. stores the orientations of previously detected surfaces. The exponential fading of the memory is crucial for overcoming incorrectly matched features which have previously been fused into the object shape memory.

Analogously, in the **translational network** a short time-window of recent contact positions of the corresponding finger is transformed into polar coordinates. These are then weighted with the ratio of the larger to the smaller eigenvalue of the covariance of the contact pressure pattern. Object edges generate long and thin pressure patterns in contrast to surfaces, which generate more round "blobs". Contacts, which correspond to edges are thus boosted by such weighting and induce peaks in the 2D *Edge Detection* field, which represent tactile features for translation estimation. These peaks are then passed into the *Memory Trace* for storing the positions of the recently detected edges in an object-centered reference frame.

b) The localization pathways: In the localization pathways, the input processing is identical with the mapping pathways, except for the faster timescale. The detected edges / surfaces are passed as inputs to a *Matching* field, which also receives the corresponding (edge or surface) *Memory Trace* as input. The lateral interactions in the matching DNF have two main effects: first, that peaks of the localization pathway are "pulled" towards previously detected peaks stored in the memory trace, if they are sufficiently close. The radius of this correction-effect can be tuned by the kernel-width of the matching field and determines the maximal spatial resolution of detectable features.

With a small temporal delay a second effect emerges: as the current feature position in the mapping pathway is propagated to the memory trace, the peak in the matching field gets stabilized in its position. Thus if the input features have a drift over time, the low-pass filtered memory trace counteracts it.

c) Correcting the pose estimate: The dynamics in the localization pathways result in a corrected feature representation, requiring a mechanism to extract the according correction term for the pose estimate. For correction of the **object rotation** estimate, the matched activation peak in the *Matching* DNF (which represents the corrected orientation) is compared to the original activation peak in the *Surface Detection* DNF. This comparison is performed in a 2D DNF (*Rotational Drift* in Fig. 2): each 1D field is projected along one of the dimensions of the 2D field, where these two projections sum up. A diagonal readout of this 2D DNF, i.e. a projection to a 1D *Drift Readout* DNF (Fig. 2), creates a representation of the two inputs in relation to each other. This mechanism is inspired by a neural implementation of reference frame transformations described in [19] and provides information of the peak shift due to the matching. The activation in the readout DNF is suppressed at locations far from the center to give close by

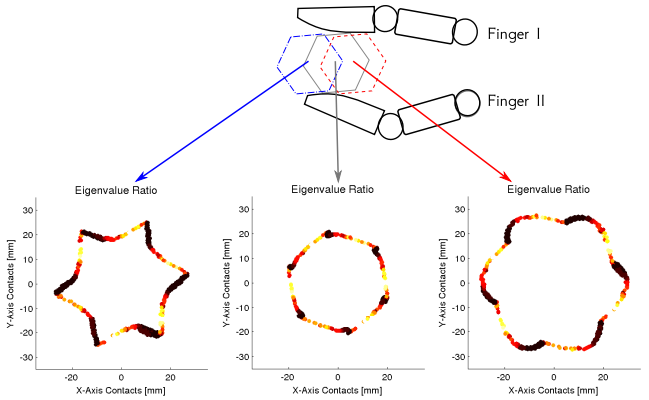

Fig. 3. Effects of errors in the position estimate. The gray object in the middle is the true position of the object. The plots in the bottom row show the object shape representation in object-centered reference frame from finger II when the estimated object position is either to the left (blue), coincides with the estimated (gray) or is to the right (red). The color-coding of the object shape representation corresponds to the computed eigenvalue-ratio, low values (black) indicate an edge and high values (yellow) a surface.

matches a higher weighting (magenta dots in *Drift Readout* plots, Fig. 2).

The deviation of the peak position from the center of the rotational *Drift Readout* DNF is the estimated error and is subsequently used to correct the current object rotation estimate. Thus, for consecutive time steps, the perceived feature is mapped to the corrected location in the object-centered reference frame.

While a wrong estimate of the object orientation leads to a corresponding rotational error in the object shape representation (see Fig. 5), a wrong estimate of the **object position** leads to distortions of the object shape representation, as shown in Fig. 3.

For a translation along this proximal-distal axis, the features are rotated with respect to a wrong rotation center, since the object is fixed such that rotations are only possible along the axis at the object center. For an object which is more distal than estimated, this leads to an inward drift of the feature (edges) location in the object shape representation for one finger and an outward drift of the edges in the other finger. Accordingly, the edges in the object shape representation drift into the opposite directions for a more proximal object than estimated, as visualized in Fig. 3.

In the current setup, only detection and correction of translations along the distal-proximal axis is implemented. Translations towards one of the fingers lead to a different pattern of distortions and will be subject of future research.

In order to compare the location of the currently perceived feature (*Edge Detection*) with the location of the matched feature (*Matching*), a coordinate frame transformation is performed. The outputs of the two 2D DNFs, which represent these locations, are summed over the orientation dimension to receive 1D activation functions. This may be done, as the translation correction is only dependent on the feature drift in the amplitude (radius of the polar representation) and not its orientation. Finally, the current and the matched amplitude

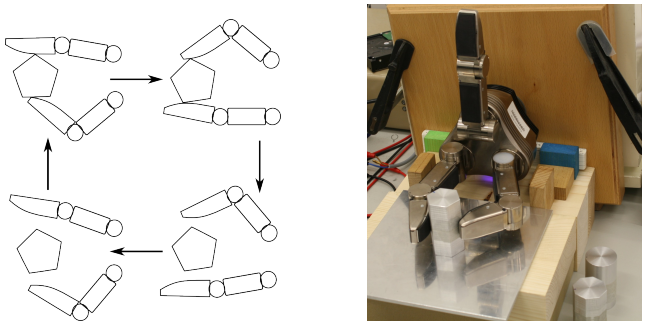

Fig. 4. Left: sketch of the rotation behavior; Right: picture of our experimental setup

of the estimated drift are projected into the two dimensions of the 2D *Translation Drift* field. The distance of the peak from the diagonal of this DNF is proportional to the estimated correction, i.e. how much the peak, induced by the new input, is "pulled back" by the memory representation. This is computed by the *Drift Readout* field, which is then used to correct the estimated object position accordingly.

IV. EXPERIMENTS

In our experimental setup a Shunk Dexterous Hand 2 is used and configured such that only two of the three fingers are used, each having two degrees of freedom (i.e. controlled joints). The two phalanges of the fingers are each equipped with a tactile sensor. These each consist of an array of 6×13 tactile elements (texels) on the distal phalanges, whereas the width decreases to 4 texels at the fingertips. Figure 4 shows the robotic setup, used for evaluation of the model, as well as the manipulation behavior used in our experiments.

A. The Setup

Rotation experiments were performed with two different aluminum objects (n-gons): an 8-sided and a 6-sided cylinder which had the same medium diameter (4.0 cm) and object height (7.0 cm). The objects had a hole in the bottom, by which they were attached to a steel axis to prevent translations of the object.

The robot hand performed an object rotation behavior as sketched in Fig. 4-left while recording the haptic data. First, the fingers are moved towards each other until tactile feedback signals sufficient contact with the object. This event triggers a movement parallel to the object's surface (i.e. rotating the object) while controlling the contact force with the object. When the contact blob reaches the edge of a tactile sensor array, the fingers are moved apart from each other, the parallel movement is reversed and the cycle begins anew.

Due to only having two joints per finger, there are not only forces orthogonal to the object surface, but also tangential components. These lead to an uncontrolled movement when the object is released, which cannot be detected nor prevented in the proposed setup. In general, this unintended movement should be systematic and indeed, there is a strong tendency of systematically underestimating the object rotation. Errors

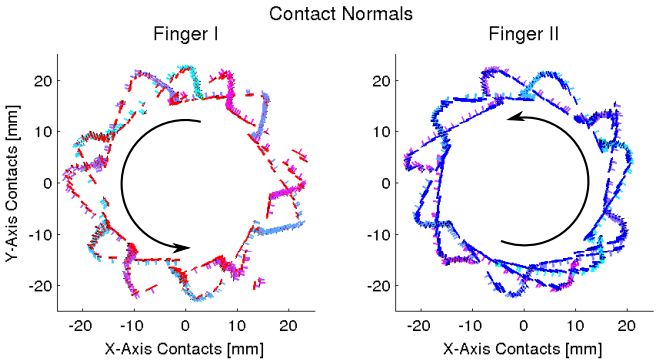

Fig. 5. Raw, uncorrected data of the last two full rotations for a 6-Sided object. The position and orientation of the tactile contacts in the object coordinate system is shown. The color-code shows the temporal order from start (teal) to end (pink) of the two rotations, the object rotation direction is indicated by the arrows.

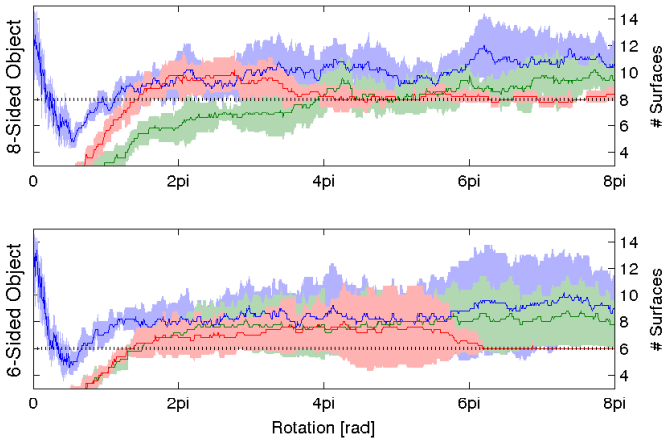

Fig. 6. The mean and standard deviation of the number of estimated object surfaces, evolving over time (i.e. object rotations). For each of the two objects (subplots) five datasets were used for computing the graphs with three different methods, respectively. Blue: based on an accumulated histogram. Green: the proposed model with deactivated error feedback. Red: the proposed model with error correction.

in the translation estimation are due to not perfectly centered objects during the initial grasps. This leads to a static offset between the true position and the initial estimate, as the object has only minor translations during the manipulation.

B. Generated Datasets

With each of the objects, five datasets were recorded, each consisting of an estimated rotation of four times 360 degrees. Hence, 10 datasets were collected, in which the tactile patterns and joint angles were sampled with approximately 2-3 Hz and the according features were computed and stored.

An exemplary subset of two rotations is visualized in Fig. 5 from the first dataset of the six-sided object, where finger I corresponds to the upper finger of the sketch in Fig. 4.

It is clearly visible that there is a drift in the object rotation estimate as the data points do not align for consecutive full rotations of the object (note the temporal color-coding in Fig. 5). Furthermore the impact of the error in the position estimate is visible in the "bump-like" shifts between two consecutive surfaces.

V. RESULTS

The benefits of the proposed neural-dynamic model are evaluated first for the rotation and then for the translation correction, as these rely on independent features and mechanisms. To evaluate the rotation correction, three different estimators of object shape, i.e. the number of detected surfaces during manipulation, are used:

First, the accumulated histogram of all past contact normal orientations is computed. This approach operates directly on the accumulated raw inputs. *Second*, the memory trace of the proposed model is used for evaluation, however without any correction in the pose estimates. In this case, the model performs a discretization with a fading memory. *Third*, the memory trace of the model with error correction is used for correction, as depicted in Fig. 2.

Each estimator was analyzed to determine the number of detected surfaces in order to demonstrate how the process of memory formation and correction of drifts improves performance. For this the activation of the estimator is smoothed and the number of peaks above a threshold is determined.

Figure 6 shows the mean and standard deviation of the number of estimated surfaces using the three estimators. The number of surfaces detected is shown for each measurement step, computed by averaging all datasets for each object.

In general, the simple accumulated histogram (first estimator) and the uncorrected model (second estimator) are incapable of building a consistent representation of the object shape, as the errors in the pose estimate accumulate and lead to a constant drift over time (see Fig. 6, blue and green lines). Only the memory trace with error correction (full model) converges to a correct estimate of the number of surfaces of the objects (red line in the two plots in Fig. 6).

As the memory is empty in the beginning and only incrementally builds up, every new surface is "corrected" into the direction of the previously seen surface and has no counterpart on the other side yet. This leads to a systematic overestimation of the number of surfaces during the first 360 degrees of the rotation. However, during further exploration of the object the model shifts and merges the orientations of surfaces in the memory and finally converges to a stable representation, as can be seen in Fig. 6, red lines. Hence the model is capable of coping with errors in the object shape memory due to previously incorrect matches.

In Fig. 7, the time-courses of the histogram, the uncorrected memory, and the corrected memory are shown as the rotation behavior is performed for the eight-sided object. In the histogram approach, each column (i.e. every rotation step) is normalized for an increased visibility. The memory is implicitly normalized, as given by Eq. 2. Note the clear increase in alignment of the detected surfaces during the rotation when the model performs corrections in the pose estimate ("Corrected MT" in Fig. 7). Initially incorrectly mapped features are fused

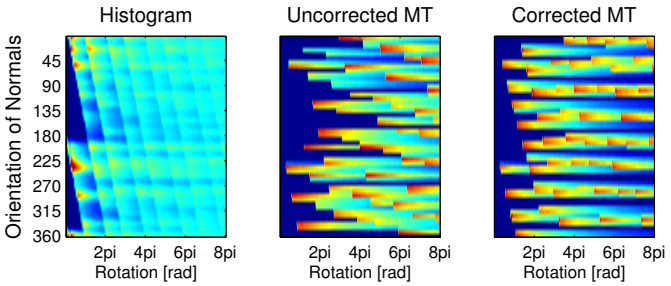

Fig. 7. The histogram, uncorrected memory (i.e. memory trace) and corrected memory for dataset 5 of the 8-Sided object. Note that the y-axis is circular and should have eight equally spaced peaks aligned for all rotation steps.

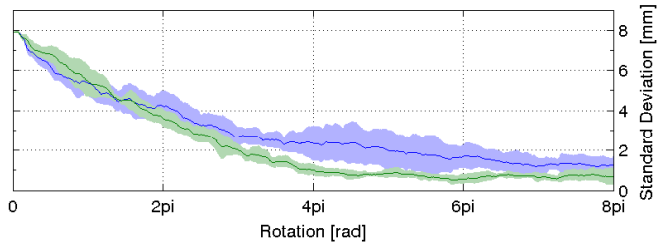

Fig. 8. Temporal development of the mean and standard deviation (across all datasets of an object) of the position estimate standard deviation (across five different initial translation estimates). Blue: the 8-Sided object; Green: the 6-Sided object.

with others or washed out as the memory trace converges to a stable representation.

Corresponding to finding the right number of surfaces, the model also finds an appropriate correction for the orientation estimate of the object. The shown data in Fig. 9 are the last two full rotations of each dataset in the object centered coordinate system for the raw data (top row), the rotation corrected data (middle row) and the combined rotation and translation corrected data (bottom row). Note that the amount of drift in rotation of the raw data was not controlled for and varies from hardly any (edge on successive edge) to the maximum possible (edge on successive surface).

The plots in Fig. 9 show for different datasets, that the neural-dynamic model for haptic learning is capable to compensate drifts in the rotation estimate as well as wrong initial estimates of the object position.

In order to evaluate the translation correction, the initial estimate of the object position was systematically varied from -1.0 cm to +1.0 cm for each of the 10 datasets. This corresponds to an translational error of up to $\pm 25\%$ of the object diameter (4.0 cm). Figure 8 shows that for all datasets the object position estimate converges (although not always to the same value, which is not visible in the figure). Hence, the system is able to create a self-consistent representation of the object, despite the false estimation of its initial pose and a drift of the pose estimate during manipulation (bottom row in Fig. 9). However, the absolute location of the object in external, allocentric coordinates, cannot be extracted from this representation.

VI. DISCUSSION

In this paper, we proposed a neural-dynamic model capable of autonomously learning a representation of the shape of an object from purely haptic data during manipulations in a closed-loop. This study revealed a principled problem in autonomous learning of shape from haptics – the haptic SLAM problem: simultaneously tracking the object pose (localization) by using a representation of the object, while building this object representation and maintaining its consistency as perceived features are continuously fused into it (mapping). We have evaluated the model in a simple, proof-of-concept robotic set-up, using two n-gon objects, which were rotated by two fingers of a robotic hand. We have demonstrated that the proposed model leads to a self-consistent object representation by correcting drifts in rotations and errors in the translation estimates. This includes mechanisms implementing a (1) working memory representation of the latest tactile stimuli; (2) long-term memory, which integrates tactile features over the whole interaction with the object; (3) correction of the location of a current tactile feature on the object towards a matched feature, stored in the long-term memory; (4) comparison of the corrected location representation with the working memory of the initial location for derivation of the according corrective term for the pose estimate of the object; (5) correction of the pose estimate. Using neural-dynamics allows to extract and correct the shape representation in an online fashion, since the lateral interaction in DNFs match transient representations and operate on multiple time scales, thereby implementing memory.

Surely, our setup was simplified and we envision several directions of extensions of the model. First, in order to build representations of more complex, every-day objects, combinations of features to higher order features will be investigated in future research. These could be a composition of surfaces, curvature and edges as well as the angle of an edge, i.e. distance between the surface-peaks in the memory trace of the model. For this composition into higher order features, an investigation of the dynamics of tactile patterns during interaction with the object surface will be necessary. Here, a higher resolution of the sensing array (both in space and in response amplitude) would be of benefit.

Second, for demonstrating that object shape can be learned purely from haptics during manipulations, it is necessary to go towards fully-fledged 3D haptic SLAM, i.e. localizing the rotation and position in 2D space. To accomplish this, the translation estimate in the remaining dimension (along the line between the fingers) needs to be corrected. A promising solution may be to use asymmetries between the data of the two fingers, which are currently neglected.

Finally, linking the haptics-derived representations with vision is another possible direction of future research. Haptic learning as implemented in our model, leads to representations compatible with vision-based representations previously developed in the DNF framework [7].

Eventually, the object shape representation is meant to be

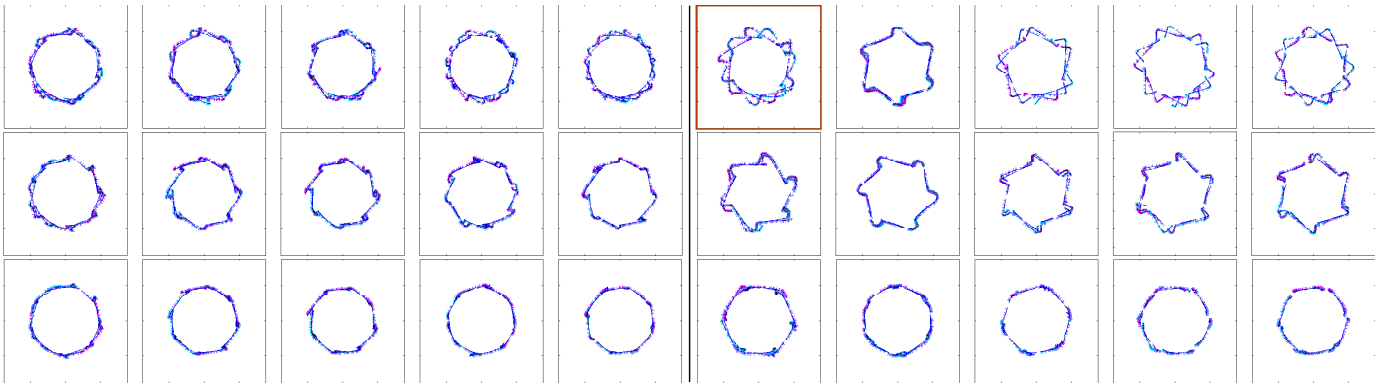

Fig. 9. Tactile measurements in the object coordinate system for the last two full rotations of all datasets of the 8-Sided (left) and 6-Sided (right) objects. The top row shows the uncorrected data of finger II; the second row shows the corresponding dataset with rotation correction and the third row with the combined rotation and translation correction. The dataset with the red surrounding box is shown in Fig. 5.

used for predicting tactile input while manipulating the object. This prediction could be used to detect changes in the interaction with an object, caused by unpredictable perturbations, which are inherent in real-world manipulation behavior. The tactile input prediction can also be used to optimize the manipulation behavior with respect to a grasp dependent cost function, e.g. falling on flat surfaces only. The system then could learn to perform faster and with higher precision in such tasks and produce behavioral predictions for equivalent human experiments. Learning in the closed behavioral loop of both the object representation and the control of motor actions, based on this representation, is the final goal of this project.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of the German Japan Collaborative Research Program on Computational Neuroscience (WO 388/11-1) and the DFG SPP *Autonomous Learning* within Priority Program 1567.

REFERENCES

[1] Shun-ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[2] Tim Bailey and Hugh Durrant-Whyte. Simultaneous localization and mapping (slam): Part ii. *Robotics & Automation Magazine, IEEE*, 13(3):108–117, 2006.

[3] Maxime Chalon, Jens Reinecke, and Martin Pfanne. Online in-hand object localization. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 2977–2984. IEEE, 2013.

[4] Nicolas Cuperlier, Mathias Quoy, and Philippe Gaussier. Neurobiologically inspired mobile robot navigation and planning. *Frontiers in neurorobotics*, 1, 2007.

[5] Stanimir Dragiev, Marc Toussaint, and Michael Gienger. Gaussian process implicit surfaces for shape estimation and grasping. In *Robotics and Automation (ICRA), International Conference on*, pages 2845–2850. IEEE, 2011.

[6] Hugh Durrant-Whyte and Tim Bailey. Simultaneous localization and mapping: part i. *Robotics & Automation Magazine, IEEE*, 13(2):99–110, 2006.

[7] Christian Faubel and Gregor Schöner. Learning to recognize objects on the fly: a neurally based dynamic field approach. *Neural Networks*, 21(4):562–576, 2008.

[8] Charles Fox, Mat Evans, Martin Pearson, and Tony Prescott. Tactile slam with a biomimetic whiskered robot. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 4925–4930. IEEE, 2012.

[9] Steven Hsiao and M. Gomez-Ramirez. *Neurobiology of Sensation and Reward - Chap. 7: Touch*. Frontiers in Neuroscience. CRC Press, 2011.

[10] Roland S Johansson and J Randall Flanagan. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience*, 10(5):345–359, 2009.

[11] Yangming Li, Shuai Li, and Yunjian Ge. A biologically inspired solution to simultaneous localization and consistent mapping in dynamic environments. *Neurocomputing*, 2012.

[12] Martin Meier, Matthias Schopfer, Robert Haschke, and Helge Ritter. A probabilistic approach to tactile shape reconstruction. *Robotics, IEEE Transactions on*, 27(3):630–635, 2011.

[13] Michael J Milford, Gordon F Wyeth, and David Prasser. Ratslam: a hippocampal model for simultaneous localization and mapping. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 1, pages 403–408. IEEE, 2004.

[14] Zachary Pezzementi, Caitlin Reyda, and Gregory D Hager. Object mapping, recognition, and localization from tactile geometry. In *Robotics and Automation (ICRA), International Conference on*, pages 5942–5948. IEEE, 2011.

[15] Robert Platt, Leslie Kaelbling, Tomas Lozano-Perez, and Russ Tedrake. Simultaneous localization and grasping using belief space planning. In *Robotics and Automation (ICRA), IEEE International Conference on, Workshop on Manipulation Under Uncertainty*, 2011.

[16] Mila Popovic, Gert Kootstra, Jimmy Alison Jorgensen, Danica Kragic, and Norbert Kruger. Grasping unknown objects using an early cognitive vision system for general scene understanding. In *Intelligent Robots and Systems (IROS), International Conference on*, pages 987–994. IEEE, 2011.

[17] Yulia Sandamirskaya. Dynamic neural fields as a step towards cognitive neuromorphic architectures. *Frontiers in Neuroscience*, 7:276, 2013.

[18] Yulia Sandamirskaya, Stephan K.U. Zibner, Sebastian Schneegans, and Gregor Schöner. Using dynamic field theory to extend the embodiment stance toward higher cognition. *New Ideas in Psychology*, 31(3):322 – 339, 2013.

[19] Sebastian Schneegans and Gregor Schöner. A neural mechanism for coordinate transformation predicts pre-saccadic remapping. *Biological cybernetics*, 106(2):89–109, February 2012.

[20] Gregor Schöner. Dynamical Systems Approaches to Cognition. In *Cambridge Handbook of Computational Cognitive Modeling*, pages 101–126. Cambridge University Press, 2008.

[21] Jacqueline C Snow, Lars Strother, and Glyn W Humphreys. Haptic shape processing in visual cortex. *Journal of Cognitive Neuroscience*, 2013.

[22] John P. Spencer and Gregor Schöner. Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, 6(4):392–412, 2003.

[23] Claudius Strub, Florentin Wörgötter, Helge Ritter, and Yulia Sandamirskaya. Using haptics to extract object shape from rotational manipulations. In *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*. IEEE, 2014.

[24] H R Wilson and J D Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13:55–80, 1973.