

# EKF-based In-hand Object Localization from Joint Position and Torque Measurements

Martin Pfanne, Maxime Chalon

*Institute of Robotics and Mechatronics, German Aerospace Center (DLR), Wessling, Germany*

*E-mails: {Martin.Pfanne, Maxime.Chalon}@dlr.de*

**Abstract**—The ability to manipulate objects is the primary purpose of any robotic hand. However, when executing a grasp, the object and fingers rarely move exactly as planned. These unobserved deviations in the pose of the object or the contact configuration can make it impossible to solve a given task.

In this paper, we present a new approach to estimate the state of the grasp using only position and torque measurements from the joints of the hand. Based on the popular extended Kalman filter framework, the algorithm estimates the pose of the manipulated object, as well as the contact forces and positions on the surface of the object. It is able to reliably detect new and the loss of contacts and to incorporate this information in the estimation filter. The validity of this new method is shown in different grasp and manipulation tasks using David, a humanoid platform of DLR.

## I. INTRODUCTION

Technological advancements in the last few decades allowed the design and development of robotic manipulators to mature from simple grippers to anthropomorphic hands, like the one of DLR robot David (Fig. 1, left) [1]. While the mechanical capabilities of these hands have almost caught up with those of humans, the intelligence when it comes to grasping and manipulation still leaves much to be desired.

One of the main problems, which makes complex tasks like in-hand manipulation so challenging, is the lack of reliable information about the state of the object after the grasp acquisition. In real world scenarios, the actual grasp of an object differs from the planned one at least to some degree for many reasons. Errors in the position of the object or the hand, inaccuracies in the planning model or deviations in the motion of the fingers may all lead to a different equilibrium pose of the object, than was anticipated. Even the real contact configuration of the grasp may be unknown at this point, if one of the fingers slid on the object in an unforeseen way and lost contact. Although the object may still be grasped in a stable configuration, subsequent tasks may fail because of wrong assumptions about the pose of the object and the configuration of the grasp. Imagine the difficulty of writing with a pen or tightening a screw with a screwdriver without knowing the exact position of these tools or how they are being held by your hand. A bottle that has been picked up may tip over when placed again because it moved during the grasp.

Usually, robotic applications rely on the use of visual localization techniques to determine the pose of an object [2]. However, after grasping the object, these approaches

may be hindered by occlusions of the object by the fingers. In [3], an integrated grasping simulator and real-time visual tracking system was proposed that, after planning a grasp, visually monitors its execution. In [4], a Monte Carlo filter was applied to solve 3D global localization problems with force controlled robots. However, their approach suffered from high computational requirements in the context of the Markov localization. Tactile data was the basis of another Bayesian estimation technique in [5].

The problem of localizing both the pose and the shape of an object is considered in [6], [7], using particle filtering and tactile sensing. [8] presented an algorithm that locally estimates the pose of a manipulated object from contact information and a geometric description of the object. After pre-computing characteristic relations between the object and the fingers offline, tactile measurements are used online to find the best match in the database. [9], [10] addressed the problem of learning to predict the interactions of rigid bodies in a probabilistic framework. [11] evaluated different contact models in their ability to predict object motion during simple manipulation tasks. In comparison, contact modeling based on Coulombs friction and maximum energy principle performed best, but was still unable to make accurate predictions. In contrast to this and other recent work on dynamic state estimation [12]–[14], our method is

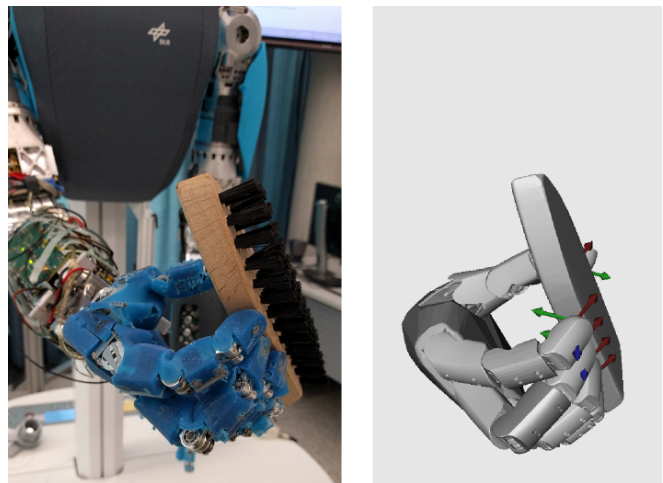


Fig. 1. Left: Inhand manipulation of a brush with the DLR robot David. Right: 3D representation of the grasp environment.

based primarily on kinematic contact models. Previously, we proposed a localization method based on a particle filter that used kinematic data and tactile sensing to online estimate the object pose [15].

The method that is proposed in this paper aims to solve the localization problem of a manipulated object without requiring additional specialized sensing hardware like artificial skin or contact sensors. Using only position and torque measurement of the fingers, this approach is directly applicable to a wide range of robotic hand systems, for example any of the torque controlled manipulators of the German Aerospace Center (David, DLR Hand II, HIT Hand, DEXHAND). Additionally, by utilizing an extended Kalman filter (EKF), the presented algorithm is able to explicitly account for any uncertainties in the measurements and the state of the localization. Using this approach, our method not only works for fingertip grasps, but also for other constraint grasps. Although extremely common in many robotic applications, these grasps often pose problems for other approaches because they are highly over-constraint (i.e. power grasps).

While this paper illustrates the core concept of a state estimation from joint measurement, it also introduces a flexible design that allows the integration of additional sensor information or the estimation of other quantities, like measurement biases. Future work will focus on extending the capabilities of the localization and making use of more measurements (i.e. visual data), whenever available.

Section II of the paper, will give a brief description of the problem and define any assumptions the method is based on. Section III focuses on the design of the estimation filter that tracks the state of the grasp. In section IV, we explain how new and the loss of contacts is detected and incorporated into the filter. Section V describes experiments that were conducted to validate the approach and presents the results. Finally, section VI summarizes the findings and gives concluding remarks.

## II. PROBLEM DESCRIPTION

The problem of localizing a grasped object is not well defined in and of itself. It greatly depends on the available data and information. But also on the intended use of the localization results. Therefore, this first section aims to define the problem, as it is approached in this paper.

Generally speaking, the purpose of the presented algorithm is to provide the necessary information about the state of the object and the grasp in order to enable subsequent tasks. Specifically, this means that in a pick and place scenario, the object has to be localized w.r.t the palm, in order to properly plan the motion to the desired target position. In an in-hand manipulation scenario, in addition to the pose of the object, the contact configuration between the object and the hand has to be determined to plan the following motions. Moreover, the algorithm is required to be able to incorporate new measurements online and provide a current best estimate of the grasp state and configuration at all times.

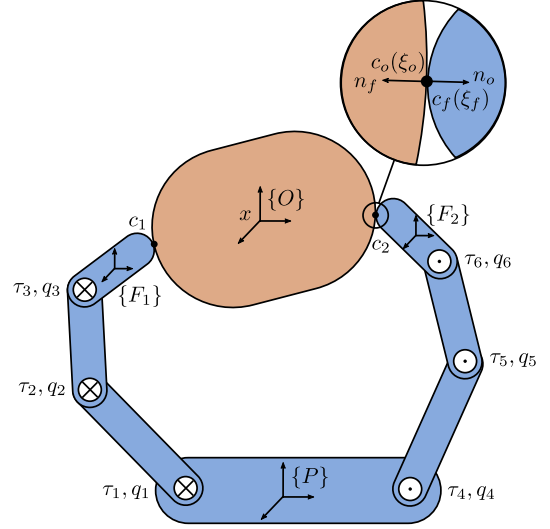


Fig. 2. Illustration of the grasp of an object with two fingers.

The proposed algorithm is dividing and solving the problem in two steps, each individually described in the subsequent two sections:

- 1) Estimation of the state of the grasp, i.e. the pose of the object w.r.t. the palm and the contact positions and forces
- 2) Detection of new or loss of contacts between the object and the fingers

To solve these problems, the algorithm relies on the assumption that the following information and measurements are available:

- (Inaccurate) measurements of the joints positions and torques of the fingers
- (Inaccurate) initial pose of the object at the start of the grasp acquisition (e.g. from vision system)
- The object is rigid and of known geometry (3D mesh available)
- All parts of the hand are rigid and of known geometry (3D meshes available)

Fig. 1 shows the grasp of an object and the respective 3D representation as an example of a possible localization problem and the information that is available to solve it.

## III. ESTIMATION OF THE GRASP STATE

This section describes the EKF as it is used for this problem. The first part defines the most important quantities and vectors. The second part focuses on the design of the filter.

### Definitions

Fig. 2 illustrates the main quantities of an object that is manipulated by a robotic hand. The pose  $x \in \mathbb{R}^6$  of the object is described by the translation and rotation of an object fixed frame  $\{O\}$  w.r.t. a palm fixed frame  $\{P\}$ . The primary purpose of any object localization algorithm is the determination of  $x$ . However, the grasp of the object is also described by the contacts between the object and

the fingers. A contact of index  $i$  is defined by its position  $c^{[i]}(\xi^{[i]}) \in \mathbb{R}^3$ , where  $\xi^{[i]} \in \mathbb{R}^2$  is the position on the surface of the object [16], as well as a scalar force  $f^{[i]}$  in the direction of the surface normal  $n^{[i]}$ . Therefore, the current state of the estimation  $y \in \mathbb{R}^{6+3n}$  is described by the pose of the object and a set of contact positions and forces at time  $t \in \mathbb{R}$ :

$$y_t = \{x_t, \xi_t^{[1]}, \dots, \xi_t^{[n]}, f_t^{[1]}, \dots, f_t^{[n]}\} \quad (1)$$

where  $n \in \mathbb{N}$  is the number of contacts. The measurement vector  $z \in \mathbb{R}^{2m}$  contains the set of all joint positions  $q \in \mathbb{R}^m$  and torques  $\tau \in \mathbb{R}^m$ :

$$z_t = \{q_t^{[1]}, \dots, q_t^{[m]}, \tau_t^{[1]}, \dots, \tau_t^{[m]}\} \quad (2)$$

where  $m \in \mathbb{N}$  is the number of joints. Lastly, the control vector  $u \in \mathbb{R}^m$  shall be given by the joint velocities of the fingers  $\dot{q} \in \mathbb{R}^m$ :

$$u_t = \{\dot{q}_t^{[1]}, \dots, \dot{q}_t^{[m]}\} \quad (3)$$

### EKF design

The state, measurement and control vectors are utilized in the EKF framework, which assumes that the state at time  $t$  can be described as [17]:

$$y_t = f(y_{t-1}, u_t) + w_t \quad (4)$$

with  $f$  being the motion model and  $w_t$  being additive, zero mean Gaussian motion disturbances. Similarly, measurements are described as:

$$z_t = h(y_t) + v_t \quad (5)$$

with  $h$  being the measurement model and  $v_t$  being additive, zero mean Gaussian measurement disturbances. In order to estimate the state and uncertainty of the object, the motion model and measurement model of the system are required. These models are based on digital representation of the kinematics and geometry of the hand and the object, which are then locally linearized at each step for the EKF. The main contribution of the proposed algorithm lies in the formulation of these differentiable models that allow for an application of the EKF to this highly nonlinear problem.

The motion model  $f$  is utilizing the grasp matrix  $G \in \mathbb{R}^{6 \times 3n}$  and hand Jacobian  $J \in \mathbb{R}^{3n \times m}$  for hard-finger contacts as defined in [18] to predict the motion of the object from the motion of the joints:

$$f(y_{t-1}, u_t) = y_{t-1} + \begin{pmatrix} G^+ J u_t \Delta t \\ 0_{3n \times 1} \end{pmatrix} \quad (6)$$

where  $\Delta t$  is the time between two steps and  $G^+$  the pseudo-inverse of  $G$ . The rest of the state vector remains unchanged.

Under this model, the prediction assumes that the contact state, i.e. the contact positions and forces, does not change. The rolling or slipping of contacts are considered as the noise of the prediction. Using a more complex motion model would be possible, i.e. to predict rolling contacts. However, practically this more computationally expansive calculation only marginally improves the prediction. In the context of

the complete filter, its effect is negligible and sufficiently accounted for as noise.

The proposed measurement model  $h$  can be divided in a joint position model  $h_q$  and a joint torque model  $h_\tau$ :

$$h(y_t) = \begin{pmatrix} h_q(y_t) \\ h_\tau(y_t) \end{pmatrix} \quad (7)$$

$h_q$  is estimating the current joint positions by using the Jacobian to minimize the displacement between the contacts on the object  $c_o$  and the contacts on the fingers  $c_f$ :

$$h_q(y_t) = h_q(y_{t-1}) + J^+(c_{o,t} - c_{f,t-1}) \quad (8)$$

where  $c_{o,t}^{[i]}$  is the position of contact  $i$  on the object. Knowing the geometry of the object, it can be calculated from  $x_t$  and  $\xi_t^{[i]}$ , which are part of state  $y_t$ . Similarly,  $c_{f,t-1}^{[i]}$ , the position of contact  $i$  on the finger, can be calculated from the previous estimate of the joint positions,  $h_q(y_{t-1})$ , and the estimated position of the contact on the surface of the finger link,  $\xi_{f,t-1}^{[i]} \in \mathbb{R}^2$ .

$$c_{o,t}^{[i]} = c_o^{[i]}(x_t, \xi_t^{[i]}) \quad (9)$$

$$c_{f,t-1}^{[i]} = c_f^{[i]}(h_q(y_{t-1}), \xi_{f,t-1}^{[i]}) \quad (10)$$

$\xi_{f,t}^{[i]}$  is recalculated in each step to minimize the difference between the current contact normal on the object,  $n_{o,t}^{[i]}$ , and the previous one on the finger,  $n_{f,t-1}^{[i]}$ :

$$\xi_{f,t}^{[i]} = \xi_{f,t-1}^{[i]} + \frac{\partial n_f^{[i]}}{\partial \xi_f^{[i]}} \bigg|_{\xi_{f,t-1}^{[i]}} ((-n_{o,t}^{[i]}) - n_{f,t-1}^{[i]}) \quad (11)$$

Similarly to the contact positions, the contact normal  $n_{o,t}^{[i]}$  is calculated from  $x_t$  and  $\xi_t^{[i]}$ , and  $n_{f,t-1}^{[i]}$  is calculated from  $h_q(y_{t-1})$  and  $\xi_{f,t-1}^{[i]}$ .

$h_\tau$  is estimating the current joint torques:

$$h_\tau(y_t) = J^T \lambda_t \quad (12)$$

where  $\lambda_t$  is the vector of the  $n$  contact forces  $\lambda_t^{[i]} \in \mathbb{R}^3$ . Using the point-contact-without-friction model [18],  $\lambda_t^{[i]}$  is a force normal to the surface of the object:

$$\lambda_t^{[i]} = n_o^{[i]} f_t^{[i]} \quad (13)$$

where  $f_t^{[i]}$  is the estimated contact force in  $y_t$ .

Using these definitions, the mean of state  $y_t$  and its covariance  $P_t$  can be recursively computed in two steps with the EKF equations [19]. First, the prediction:

$$\bar{y}_t = f(y_{t-1}, u_t) \quad (14)$$

$$\bar{P}_t = F_{t-1} P_{t-1} F_{t-1}^T + Q_t \quad (15)$$

where  $Q_t$  is the covariance of the motion disturbance and:

$$F_t = \frac{\partial f}{\partial y} \bigg|_{y_{t-1}, u_t} \quad (16)$$

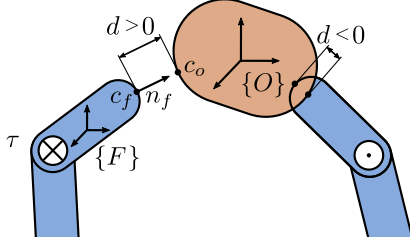


Fig. 3. Important quantities for the detection of new contacts:  $d_j$  is the positive smallest distance or negative deepest penetration between a finger link and the object.  $f_j$  is the magnitude of the normal contact force.

And second, the update:

$$y_t = \bar{y}_t + K_t(z_t - h(\bar{y}_t)) \quad (17)$$

$$P_t = (I - K_t H_t) \bar{P}_t \quad (18)$$

with:

$$H_t = \left. \frac{\partial h}{\partial y} \right|_{y_t} \quad (19)$$

$$K_t = \bar{P}_t H_t^T (H_t \bar{P}_t H_t^T + R_t)^{-1} \quad (20)$$

where  $R_t$  is the covariance of the measurement disturbance.

The calculation of the first EKF step requires initial values for the mean,  $y_0$ , and covariance,  $P_0$ , of the state. The initial pose of the object was assumed to be available, e.g. provided by a vision system.  $P_0$  is initialized to a diagonal matrix and set to the assumed uncertainty in the initial pose.

Similarly,  $Q_t$  and  $R_t$  are diagonal matrices representing uncertainties in the motion and measurement model. While  $Q_t$  has to account for errors from rolling or slipping contacts,  $R_t$  should be set to the expected inaccuracies of the measurements of the joint positions and torques.

#### IV. DETERMINATION OF THE CONTACT CONFIGURATION

The previous chapter described how joint measurements are incorporated to track the pose of the object, as well as the contact positions and forces. So far, the algorithm assumed a static grasp configuration, i.e. a constant number of contacts. However, it is apparent that during the execution of a grasp the number of contacts between the fingers and the object is not constant at all. Starting from zero contacts, as the hand approaches the object, one finger after another will come in contact. Later, when the hand is opened and the object is placed again, the fingers lose contact to the object, eventually not touching it at all anymore. Therefore, any in-hand object localization algorithm has to be able to detect new or the loss of contacts, as well as incorporate this information accordingly into the estimation. While the first part of this section focuses on detecting new contacts and adding them to the EKF, the second part describes the loss and removal of them.

##### *Detecting and adding new contacts*

Before the execution of a grasp begins, no parts of the hand are assumed to be in contact with the object. As the fingers

start moving, the distance between their links and the object becomes smaller. As they touch, the fingers begin to apply a force to the object, resulting in a measurable torque on the joints of the fingers. Using the available measurements of the joint torques and positions, as well as knowledge about the initial pose of the object, new contacts can be detected.

Without inaccuracies, both a zero distance between a finger and the object, and non-zero torques applied to its joints would indicate a contact. However, both properties have to be inferred from non-accurate measurement. Therefore, this paper proposes a heuristic based on the estimated distance between a finger link and the object, and the scalar contact force, to evaluate the state of the contact:

$$w_j = W_j(d_j, f_j) \quad (21)$$

where  $d_j$  is the scalar distance between link  $j$  and the object,  $f_j$  is the scalar contact force and  $W_j$  is a heuristic function to calculate a unit-less scalar weight  $w_j$  to represent the probability of having a contact between link  $j$  and the object. If  $w_j$  is bigger than a set threshold  $\rho_+$ , the link is assumed to be in contact. The value of  $\rho_+$  and the function for  $W_j$  were empirically chosen after evaluating the results of a number of varying scenarios to work well for different objects and and grasps:

$$w_j = \frac{f_j}{\sqrt{d_j}} \quad (22)$$

Since  $f_j$  and  $d_j$  can not be measured directly, they have to be calculated from the available measurements and data. Fig. 3 illustrates the most important quantities for the calculation.

To determine the smallest distance between two geometric objects, a wide range of methods have been proposed in the past. For this algorithm, a modified version of the Gilbert-Johnson-Keerthi (GJK) distance algorithm was used [20], [21]. Given the pose and geometry of two bodies, this version of the GJK algorithm is able to return the smallest distance between them, if they are not colliding. If the two bodies are indeed in collision, the algorithm returns the negative penetration depth. In addition to the distance or penetration depth, the algorithm also returns the respective positions on the surface of the two bodies.

The scalar contact force has to be inferred from the torque measurement of the finger joints. However, this calculation also requires the position of the contact on the finger. This information is taken from the GJK algorithm, i.e. the point of smallest distance or deepest penetration to the object, respectively. The relation between the joint torques and the contact forces is given by the hand Jacobian. Since a scalar force value is required, the point-contact-without-friction model is used for this calculation. That means, only the normal component of the force vector is transmitted through the contact. The scalar force can then be calculated as follows:

$$f_j = (J_j^T n_j)^+ \tau \quad (23)$$



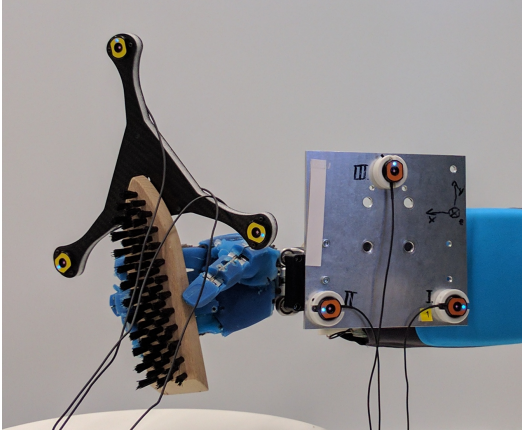


Fig. 4. Experimental setup for the evaluation of the estimation algorithm. The pose of the object and the palm are tracked with the vision system K610 from Nikon.

where  $J_j$  is the partial hand Jacobian for a hard-finger contact on link  $j$ ,  $n_j$  is the normal vector on the surface of the link and  $\tau$  is the measured torque vector.

Once a new contact between a link and the object has been detected, it has to be added to the estimation filter. Specifically, the state vector  $y_t$ , covariance matrix  $P_t$  and covariance of the motion disturbance  $Q_t$  have to be modified to account for the additional contact:

$$y_t = \{x_t, \xi_t^{[1]}, \dots, \xi_t^{[n]}, \xi_t^*, f_t^{[1]}, \dots, f_t^{[n]}, f_t^*\} \quad (24)$$

where  $\xi_t^*$  is the position of the new contact on the surface of the object, as provided by the GJK algorithm and  $f_t^*$  is the scalar contact force  $f_j$ .  $P_t$  is extended with the assumed initial uncertainty of the contact position and force as diagonal elements of the matrix. Similarly,  $Q_t$  has to be enlarged to account for the motion disturbances effecting the new contact.

#### Detecting loss of and removing contacts

During the manipulation and use of an object, contact to some or even all fingers may be lost. This could be planned, e.g. when putting down an object, or be accident, e.g. when one finger slips off the object. In any case, to allow for the correct estimation of the grasp configuration, these events have to be detected correctly and the contacts have to be removed from the filter. Physically, the loss of contact means that there is a non-zero distance between the respective link and the object, as well as a zero contact force.

The detection of the loss of contact between one finger link and the object is realized similarly to that of new contacts, in that the same heuristic is used to evaluate the contact state:

$$w_j = \frac{f_j}{\sqrt{d_j}} \quad (25)$$

However, instead of using the GJK algorithm and the raw torque measurement to obtain  $f_j$  and  $d_j$ , the necessary values are calculate from the state of the estimation filter,  $y_t$ . While  $f_j$  can be taken directly from  $y_t$ ,  $d_j$  can be calculate from the estimated position of the contact on the object,  $c_o$ ,

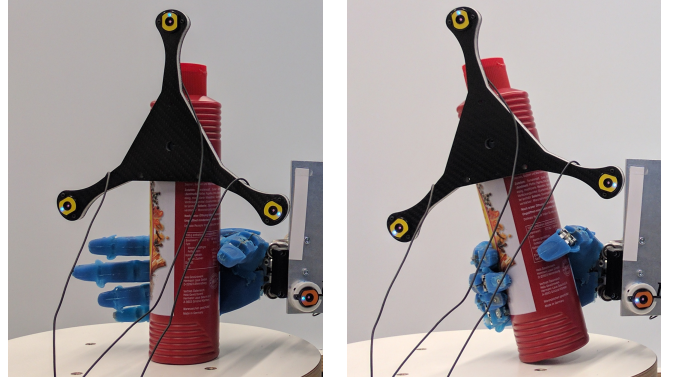


Fig. 5. Power grasp of a ketchup bottle. The bottle tilts during the grasp, potentially causing it to fall over when set down again.

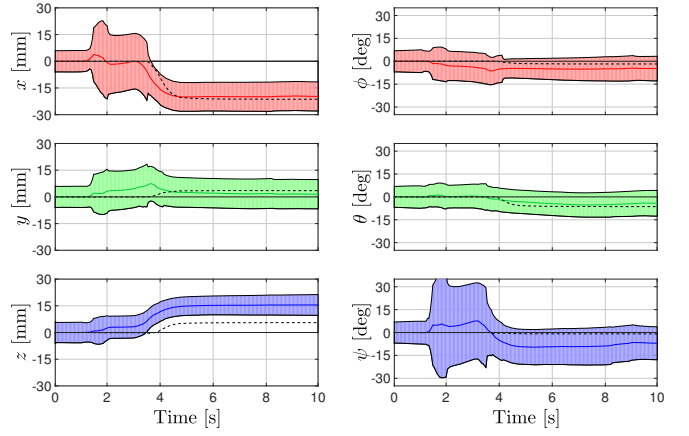


Fig. 6. Displacement in position (left) and orientation (right) of the grasped ketchup bottle from ground truth (dashed, black line) and EKF estimation (colored, solid line), including  $3\sigma$  uncertainty range (colored area). No tracking would have assumed that there is 0 displacement.

and on the finger,  $c_f$ . If  $w_j$  gets smaller than an empirically determined threshold  $\rho_-$  the contact to link  $j$  is removed from the estimation.

Similar to adding contacts, the removal of one involves adjusting  $y_t$ ,  $P_t$  and  $Q_t$ . In  $y_t$ , the position and force of the respective contact have to be removed from the vector. In  $P_t$  and  $Q_t$  the lines and rows that correspond to the uncertainties of this contact have to be eliminated.

#### V. EXPERIMENTAL VALIDATION

The proposed algorithm was evaluated using the DLR robot David. Fig. 4 illustrates the experimental setup of one of the experiments. In order to evaluate the estimation results, a ground truth of the object pose w.r.t. the palm of the hand was required. For this purpose, the K610 visual tracking system from Nikon was used. Marker plates, each containing three tracking LEDs, were rigidly attached to the object and the wrist of the hand. Measuring the 3D position of all six LEDs allowed the calculation of the pose ground truth. However, this system did not provide any reference for the estimated contact positions or forces. A more elaborate evaluation of the proposed method would therefore require additional sensors. While that was not part of the experiments

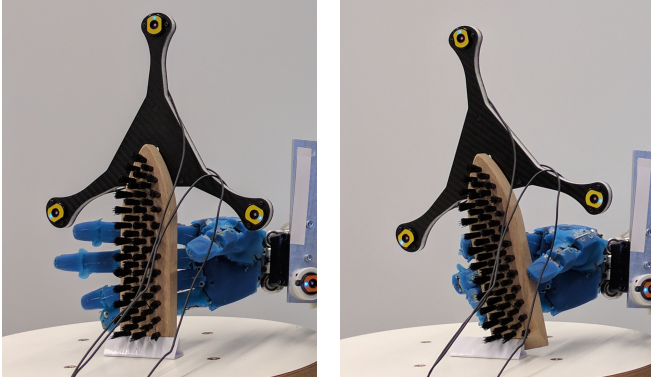


Fig. 7. Fingertip grasp of a brush. The brush turns in the hand as the contact forces are physically balanced between all five fingers.

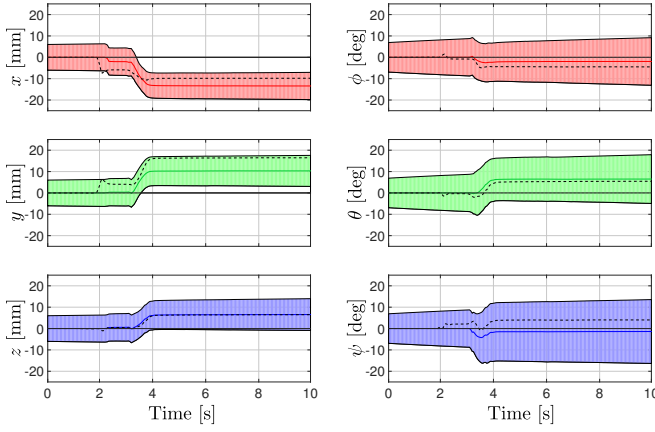


Fig. 8. Displacement in position (left) and orientation (right) of the grasped brush from ground truth (dashed, black line) and EKF estimation (colored, solid line), including  $3\sigma$  uncertainty range (colored area). No tracking would have assumed that there is 0 displacement.

for this publication, future work could expand on these shortcomings with a more complete validation setup.

The measurements from the visual tracking system were also used to provide an initial pose of the object to the localization algorithm. During manipulation, only the torque and position measurements of the fingers were accessible to the estimation and processed with a  $\Delta t$  of 50 ms.

The main purpose of the presented experiments was to validate the algorithm, demonstrating its ability to estimate a feasible object pose that satisfies all constraints given by the measurements. This section describes the three scenarios that were evaluated and illustrates their results.

#### Exp. 1: Power grasp

The first experiment was a power grasp of a ketchup bottle. In practice, executing a grasp like that rarely fails. That means, even without perfect knowledge of the scene or precise planning of the finger motions, the bottle usually ends up tightly grasped inside the hand. However, these inaccuracies may cause the object to move during the grasp, for example tilt w.r.t. to the table. This motion, if not accounted for, may then cause the bottle to fall over, when it is later placed on the table and released again. Fig. 5

illustrates this scenario, showing the ketchup bottle before and after grasping.

Using the position and torque measurements from the joints, the motion of the bottle is estimated. Fig. 6 shows the estimated pose of the object compared to the ground truth from the tracking system. These results show both the capabilities and limitations of the algorithm. The motion along  $x$  and  $y$ , as well as the rotation around these axes was estimated very well. The deviation in and around  $z$  is rather big. However, this was to be expected. This direction corresponds to the symmetry axis of the ketchup bottle. Given the available measurements, poses along or around this axis can not be distinguished. A better estimation of these degrees of freedom therefore requires additional sensor information. After grasping, all uncertainties slowly increase as the noise of the motion model makes the object pose more and more uncertain over time.

If the motion of the object was not tracked during the grasp, the assumed position on the table (in the  $x$ - $y$ -plane) would have been wrong by 22 mm. By estimating the object motion with the proposed method this error was reduced to 3 mm. Similarly, the error in the assumed tilt of the object w.r.t. the table plane could be reduced from  $7^\circ$  to  $4^\circ$ . This improvement already made it possible to move the bottle to a pose, where it could be properly placed on the table again, where otherwise it would have fallen because of the unknown tilt of  $7^\circ$ .

Furthermore, this experiment demonstrates that the application of the proposed algorithm is not limited to grasps, where only the fingertips of the hand manipulate the object. Power grasping the ketchup bottle leads to several links of the same finger to be in contact. Although this greatly over-constrains the object pose, the estimation filter is able to robustly track the object motion.

#### Exp. 2: Fingertip grasp

The second experiment comprised of a fingertip grasp of a brush. As shown in Fig. 7, grasping an object like that may cause it to move until it reaches an equilibrium where all outside forces are balanced. To pre-compute or avoid this motion is often challenging or not possible at all because of inaccuracies or limitations to the planning model.

More so than the previous power grasp, estimating this grasp configuration greatly depends on the torque measurements from the finger joints. The joint positions make sure that the estimated object stays in contact with the fingers, but this does not constraint it the correct pose. However, the joint torques infer the direction of the contact forces and therefore the normal direction of the contacts on the surface of the object.

Fig. 8 illustrates the results of the experiment. While generally not as precise as the grasp of the ketchup bottle, the knowledge about the pose of the object could still be improved. Without tracking, the assumed position after grasping would be wrong by 20 mm. This error could be reduced to 7 mm. Additionally, the error in the rotation of the brush was reduced from  $8^\circ$  to  $6^\circ$ .

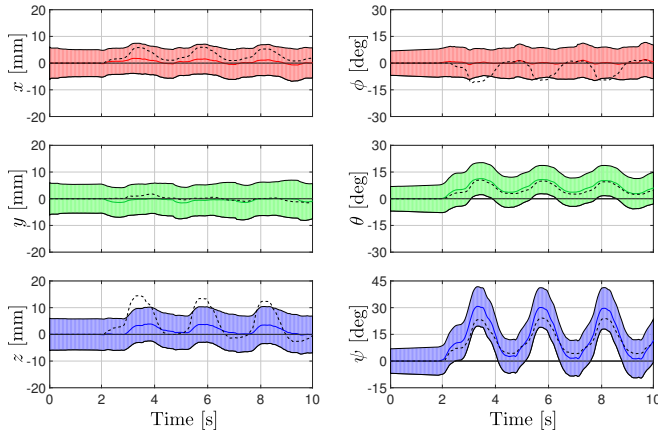


Fig. 9. Displacement in position (left) and orientation (right) of the in-hand manipulated brush from ground truth (dashed, black line) and EKF estimation (colored, solid line), including  $3\sigma$  uncertainty range (colored area).

### Exp. 3: In-hand manipulation

The final experiment was a simple in-hand manipulation task. The grasped brush, as shown in Fig. 4, was rotated back and forth inside the hand, while tracking its pose. The contact configuration remained the same during this motion, i.e. the same finger links stayed in contact with the object at all times.

The results of the experiment are illustrated in 9. The main components of the manipulation, namely the rotation around  $\theta$  and  $\psi$  were estimated correctly. However, more subtle motions, i.e. around  $\phi$  and along  $x$  and  $z$ , were generally underestimated. Additionally, at times the ground truth lies even outside the  $3\sigma$  uncertainty range of the estimation, which means that the filter is too confident about the estimated values. The pose error at the end of the experiment was 2 mm in position, and  $2^\circ$  in rotation, respectively.

## VI. CONCLUSION

In this paper we proposed a new solution to the problem of determining the pose of a manipulated object, as well as the contact configuration of the grasp. Based on the popular extended Kalman filter framework, this novel algorithm relies only on position and torque measurements from the finger joints in order to estimate the pose of the object, as well as the contact state. The method was successfully validated with real measurements from three different grasping and manipulation tasks using the DLR robot David.

While the proposed algorithm is able to estimate the object pose with a limited set of measurements, when available, additional information about the state of object should be used to improve the localization. The flexible design of the presented system has the potential to be expanded to make use of additional sensor data. For example, contact sensors could replace the heuristic function for the contact detection, or visual data could be fused in the EKF to help estimate degrees of freedom that are not kinematically constraint. Further work will focus on incorporating these additional sources of information.

## REFERENCES

- [1] M. Grebenstein, A. Albu-Schäffer, T. Bähls, M. Chalon, O. Eiberger, W. Friedl, R. Gruber, S. Haddadin, U. Hagn, R. Haslinger *et al.*, “The DLR hand arm system,” in *Proceedings. IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 3175–3182.
- [2] P. Allen, A. Timcenko, B. Yoshimi, and P. Michelman, “Automated tracking and grasping of a moving object with a robotic hand-eye system,” *IEEE Transactions on Robotics and Automation*, vol. 9, no. 2, pp. 152–165, 1993.
- [3] D. Kragic, A. Miller, and P. Allen, “Real-time tracking meets online grasp planning,” in *Proceedings. IEEE International Conference on Robotics and Automation (ICRA)*, 2001, pp. 2460–2465.
- [4] K. Gadeyne and H. Bruyninckx, “Markov techniques for object localization with force-controlled robots,” in *Proceedings. 10th International Conference on Advanced Robotics (ICAR)*, 2001, pp. 91–96.
- [5] A. Petrovskaya, O. Khatib, S. Thrun, and A. Ng, “Bayesian estimation for autonomous object manipulation based on tactile sensors,” in *Proceedings. IEEE International Conference on Robotics and Automation (ICRA)*, 2006, pp. 707–714.
- [6] C. Corcoran and R. Platt, “A measurement model for tracking hand-object state during dexterous manipulation,” in *Proceedings. IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 4302–4308.
- [7] R. Platt, F. Permenter, and J. Pfeiffer, “Inferring hand-object configuration directly from tactile data,” in *Electronically published proceeding of the Mobile Manipulation Workshop, ICRA*, 2010.
- [8] S. Haidacher and G. Hirzinger, “Estimating finger contact location and object pose from contact measurements in 3D grasping,” in *Proceedings. IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2. IEEE, 2003, pp. 1805–1810.
- [9] M. Kopicki, R. Stolkin, S. Zurek, T. Morwald, and J. Wyatt, “Predicting workpiece motions under pushing manipulations using the principle of minimum energy,” in *Robotics: Science and Systems conference, Proceedings of the Workshop: Representations for object grasping and manipulation in single and dual arm tasks*, 2010.
- [10] M. Kopicki, R. Stolkin, S. Zurek, and T. Morwald, “Learning to predict how rigid objects behave under simple manipulation,” in *Proceedings. IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1. IEEE, 2011, pp. 5722 – 5729.
- [11] R. Kolbert, N. Chavan-Daffe, and A. Rodriguez, “Experimental validation of contact dynamics for in-hand manipulation,” in *International Symposium on Experimental Robotics*. Springer, 2016, pp. 633–645.
- [12] L. Zhang, S. Lyu, and J. Trinkle, “A dynamic bayesian approach to real-time estimation and filtering in grasp acquisition,” in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 85–92.
- [13] M. Posa, C. Cantu, and R. Tedrake, “A direct method for trajectory optimization of rigid bodies through contact,” *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 69–81, 2014.
- [14] S. Li, S. Lyu, and J. Trinkle, “State estimation for dynamic systems with intermittent contact,” in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3709–3715.
- [15] M. Chalon, J. Reinecke, and M. Pfanne, “Online in-hand object localization,” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 2977–2984.
- [16] S. Rusinkiewicz, “Estimating curvatures and their derivatives on triangle meshes,” in *3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on*. IEEE, 2004, pp. 486–493.
- [17] H. Durrant-Whyte and T. Bailey, “Simultaneous localization and mapping: part i,” *IEEE robotics & automation magazine*, vol. 13, no. 2, pp. 99–110, 2006.
- [18] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Springer, 2008, ch. 28. Grasping, pp. 671–700.
- [19] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005, ch. 2. Recursive State Estimation, pp. 13–38.
- [20] E. G. Gilbert, D. W. Johnson, and S. S. Keerthi, “A fast procedure for computing the distance between complex objects in three-dimensional space,” *IEEE Journal on Robotics and Automation*, vol. 4, no. 2, pp. 193–203, 1988.
- [21] S. Cameron, “Enhancing GJK: Computing minimum and penetration distances between convex polyhedra,” in *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, vol. 4. IEEE, 1997, pp. 3112–3117.