2011 IEEE International Conference on Robotics and Automation
Shanghai International Conference Center
May 9-13, 2011, Shanghai, China

Towards a Unifying Grasp Representation for Imitation Learning on Humanoid Robots

Martin Do, Tamim Asfour and Rüdiger Dillmann

Abstract— **In this paper, we present a grasp representation in task space exploiting position information of the fingertips. We propose a new way for grasp representation in the task space, which provides a suitable basis for grasp imitation learning. Inspired by neuroscientific findings, finger movement synergies in the task space together with fingertip positions are used to derive a parametric low-dimensional grasp representation. Taking into account correlating finger movements, we describe grasps using a system of virtual springs to connect the fingers, where different grasp types are defined by parameterizing the spring constants. Based on such continuous parameterization, all instantiation of grasp types and all hand preshapes during a grasping action (reach, preshape, enclose, open) can be represented. We present experimental results, in which the spring constants are merely estimated from fingertip motion tracking using a stereo camera setup of a humanoid robot. The results show that the generated grasps based on the proposed representation are similar to the observed grasps.**

I. Introduction

The acquisition of novel grasping skills plays an essential role in enabling humanoid robots to fully interact with the environment and the human. Considering the variety of objects and the different ways that an object can be dealt with, grasping strategies have to be developed which go beyond simple closure grasps towards a complete projection of the numerous possible grasps associated with each object. In order to bootstrap this process, imitation learning provides a fast alternative in terms of acquisition of new skills. Learning from human demonstration features the possibility of generating a representation of a demonstrated action which encodes human-likeness and subtle characteristics such as constraints which are satisfied during the execution of a specific task by a human.

Therefore, an essential issue in imitation learning that has to be addressed, is the question of what features have to be stored and processed and how, in order to obtain a generalized representation, which can be adapted and applied to new objects and situations. This issue becomes even more evident, when we look at the grasp problem for a robot hand where the motion of a highly complex system with several degrees of freedom (DoF) has to be controlled.

In [1], an early attempt of a generalized grasp representation in joint space is given by the concept of the grasp taxonomy which describes the assignment of grasp hand postures to a finite number of classes. Based on this grasp taxonomy [2] provides an extension which additionally incorporates a number of hand posture patterns common in a manufacturing environment. This concept suggests that given a class, represented by an exemplary posture, for a specific object, an instance in the vicinity of the posture can be found which forms a good grasp. In previous works such as [3] and [4] grasp taxonomies are applied to grasp synthesis. However, due to large number of DoF being involved, the configuration space remains huge.

Neuroscientific studies (see [5]) demonstrated that a lesser number of DoF needs to be actively controlled to cover the range of possible human hand postures during daily grasping activities, due to consistent covariations between the finger joints, known as postural hand synergies. In previous works ([6],[7],[8]), this concept has been applied to control the entire hand by a lower dimensional set of base postures commonly extracted by applying dimensionality reduction algorithms such as Principal Component Analysis. A similar approach is proposed in [9] where a grasp representation is generated in the form of manipulation manifolds consisting of hand postures and a mapping from joint space onto manipulation parameter in task space.

Nevertheless, these approaches operate in joint angle space, respectively in its projection to a lower dimensional subspace, which features unfavorable characteristics in terms of imitation learning through the observation by a humanoid robot. One major issue lies in the high complexity of observing and tracking human hands in joint space. Vision-based tracking algorithms which can be used with a stereo camera setup of a humanoid head do not provide the necessary performance and accuracy, while highly accurate motion capture systems involve high costs and time-consuming operation. Furthermore, the question arises how and whether the continuous reach movement in task space can be properly aligned with discrete hand postures in joint space. As stated in [10] and [11], the preshape and the enclose phase, respectively the final grasp phase, are accompanied by the reaching movement. Therefore, from the kinematic point of view, it seems to be reasonable to treat the whole grasping process as a single unit in which the three phases persist in permanent correlation to each other. Hence, attaining human-likeness is contradictory to the decoupled processing of the preshape, reach, and enclose phase. Ordinary task space representations (see [12], [13]) considering merely contact points and fingertip positions do not address this issue, while representations in the form of separate finger trajectories neglect correlating finger movements.

Hence, this work proposes a task space representation which incorporates synergies between the fingers by means

M. Do, T. Asfour and R. Dillmann are with the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, as members of the Institute for Anthropomatics, e-mail: martin.do,asfour,dillmann@kit.edu

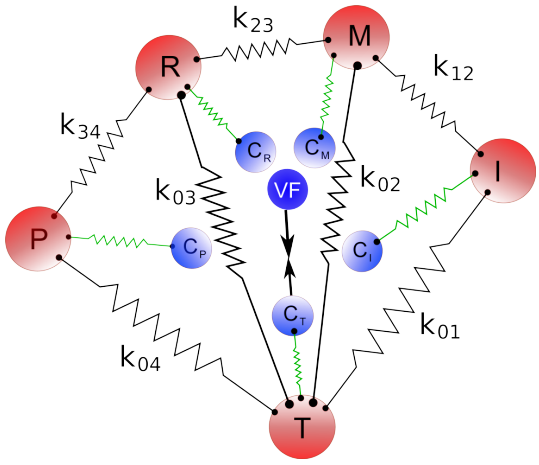

Fig. 1. Visualization of the bodies representing the fingertips. The virtual springs are visualized in black and denoted by the corresponding spring constant k_{ij}. VF describes the virtual finger while thumb (T), index (I), middle (M), ring (R) and pinkie (P) are indexed with 0, 1, 2, 3, 4.

of mass-spring-damper systems whose parameters form the representation of a grasp. Furthermore, we will show that this representation can be an initial building block for a grasp imitation learning framework which allows the parameter estimation from human observation, mapping to a humanoid platform, and the execution by reaching and grasping.

The paper is organized as follows. Section II describes the proposed representation of a grasp consisting of the model of the fingertip motion and the reach movement. In Section III, the experimental setup is explained containing a description of the humanoid platform, the observation mechanism and mapping to the robot platform. Finally, experimental results are given in IV. In conclusions, the work is summarized and notes to future works are given.

II. Representation of a Grasp

A. Concept

In this work, we suggest a representation which exploits synergies in task space by exploring and modulating fingertip movements during the grasp process. To establish synergies, one has to ensure that the trajectory of each fingertip is influenced by the motion of the remaining fingers, especially the neighboring ones. To model these relationships, ordinary mass-spring-damper systems are introduced as virtual springs between the fingertips as depicted in Fig. 1. The motion at time t of finger i at position $\mathbf{p_i} \in \mathbb{R}^{dim}$ connected to a finger j at position $\mathbf{p_j} \in \mathbb{R}^{dim}$ via a virtual spring can be inferred from following second order system of differential equations:

$$\ddot{\mathbf{x}}_{\mathbf{ij}}(t) = -\frac{k_{ij}}{m_i}\mathbf{x_{ij}}(t) - \frac{d}{m_i}\dot{\mathbf{x}}_{\mathbf{ij}}(t), \tag{1}$$

with

$$\mathbf{x_{ij}}(t) = \frac{\mathbf{p_i} - \mathbf{p_j}}{\|\mathbf{p_i} - \mathbf{p_j}\|}(\|\mathbf{p_i} - \mathbf{p_j}\| - l_{ij}), \tag{2}$$

describing the displacement concerning the equilibrium length l_{ij} along the spring direction and $\dot{\mathbf{x}}_{\mathbf{ij}}(t)$ and $\ddot{\mathbf{x}}_{\mathbf{ij}}(t)$ representing the corresponding velocity and acceleration. m_i denotes the mass of the physical body which represents the fingertip i while k_{ij} denotes the spring constant and d the damping constant. Auxiliary springs, which link each finger to its supposed contact position $\mathbf{c_i}$ are added to the system reducing oscillations in order to maintain stability and to retain the system centered around the contact points. The forces of each auxiliary spring can be determined by following equation:

$$\ddot{\mathbf{x}}_{\mathbf{c_i}}(t) = -k_c\mathbf{x}_{\mathbf{c_i}}(t) - d\dot{\mathbf{x}}_{\mathbf{c_i}}(t), \tag{3}$$

where $\mathbf{x}_{\mathbf{c_i}}(t)$ is obtained by replacing $\mathbf{p_j}$ with $\mathbf{c_i}$ in Eq. 2. k_c is constant and holds the same value for all auxiliary springs. In order to grasp an object, we desire a smooth, simultaneous movement of all finger towards their corresponding contact points. According to neuroscientific studies (see [14]), during the grasp process humans tend to focus on a mainly fixed spot on the object surface which corresponds to the thumb contact position. Therefore, the thumb is assumed to lead the reaching movement of the end effector towards the object. Following a concept introduced in [15], the remaining fingers form the virtual finger whose forces are supposed to build up an opposition force to the force exerted by the thumb in order to achieve a stable grasp. Based on these findings, to attain balanced, simultaneous finger movements, a central force $\mathbf{f_{cen}}(t)$ is applied on the entire system, which exerts a force $\mathbf{f_{i,ext}}(t)$ on each auxiliary spring resulting in:

$$\mathbf{f_{i,ext}}(t) = \begin{cases} \mathbf{f_{cen}}(t) & ,i = 1 \\ \left(\frac{\sum_{i=2}^{N}\|\mathbf{x_{c_i}}(\mathbf{t})\|}{(N-1)\|\mathbf{x_{c_1}}(\mathbf{t})\|}\right)^2 \mathbf{f_{cen}}(t) & ,else, \end{cases} \tag{4}$$

with finger $i = 1$ indexing the thumb. For the description of the entire system, we introduce a connection matrix $\mathbf{K}$ with $\mathbf{K}(i,j) = k_{ij}$ and a damping matrix $\mathbf{D}$ with $\mathbf{D}(i,j) = d_{ij}$ where $k_{ij} > 0$, $d_{ij} = d$ if two fingers are connected by virtual spring, and $k_{ij} = 0$, $d_{ij} = 0$ otherwise. Furthermore, a vector $\mathbf{c}$ is introduced indicating whether a finger is involved in a grasp by setting $c(i) = 1$, and $c(i) = 0$ otherwise. To obtain the complete equation for the motion of the body i, Eq. 3 and Eq. 4 are added to Eq. 1 which leads to following equation:

$$\ddot{\mathbf{x}}_{\mathbf{i}}(t) = -\mathbf{K}\mathbf{x_{ij}}(t) - \mathbf{D}\dot{\mathbf{x}}_{\mathbf{ij}}(t) + \mathbf{c}^{\mathbf{T}}(\ddot{\mathbf{x}}_{\mathbf{c_i}}(t) + \mathbf{f_{i,ext}}(t)). \tag{5}$$

By solving Eq. 5 one obtains the displacements $\mathbf{x_i}(t)$ by which the position $\mathbf{p_i}$ is updated. Due to stability reasons the implicit fourth-order Runge-Kutta method is used as a solver.

In case contact points are not available, the proposed system can be applied on simple shaped objects by modifying the $\mathbf{x}_{\mathbf{c_i}}(t)$. Replacing each contact point with the object center $\mathbf{c_o}$ and introducing the mean distance $\bar{d}(\{\mathbf{p_s}\}, \mathbf{c_o})$ between relevant surface points $\{\mathbf{p_s}\}$ and $\mathbf{c_o}$ as equilibrium length one obtains:

$$\bar{\mathbf{x}}_{\mathbf{c_i}}(t) = \frac{\mathbf{p_i} - \mathbf{c_o}}{\|\mathbf{p_i} - \mathbf{c_o}\|}(\|\mathbf{p_i} - \mathbf{c_o}\| - \bar{d}(\{\mathbf{p_s}\}, \mathbf{c_o})). \tag{6}$$

Applying Eq. 6 to Eq. 5 instead of Eq. 3 leads to the desired system equations. Relevant surface points can be extracted e. g. from a silhouette which one obtains when intersecting

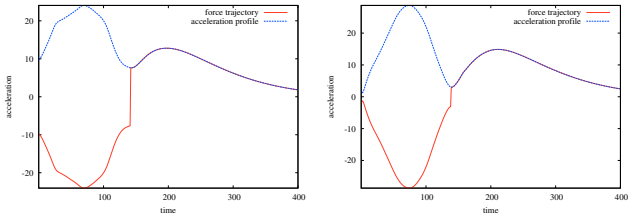

Fig. 2. Blue line, starting from positive, describes the acceleration for a reaching movement towards the object. The red line is the derived force trajectory. Left: Grasp from above the object. Right: Grasp from the side of the object.

a plane orthogonal to the reach direction of the end effector with a volumetric object representation.

Most of the parameters within the system are assumed to be constant or can be calculated except for the central force $\mathbf{f_{cen}}(t)$ whereas $\mathbf{f_{cen}}(t)$ plays a crucial role in the modulation of the system. For $\mathbf{f_{cen}}(t) < 0$ finger movements are generated which lead the fingertips away from the object whereas this process can be considered as the preshaping of the hand. The enclosing movement is initiated when $\mathbf{f_{cen}}(t) > 0$, causing the system to progress towards the final grasp pose. Therefore, $\mathbf{f_{cen}} = \{\mathbf{f_{cen}}(0) \dots \mathbf{f_{cen}}(t)\}$ can be interpreted as a force trajectory which controls the execution and transition of the different grasp phases. The trajectory can be inferred from the acceleration profile of the reach movement as follows:

$$\mathbf{f_{cen}}(t) = \begin{cases} -\|\mathbf{a}(t)\| & ,t < t_{a_{min}} \\ \|\mathbf{a}(t)\| & ,else, \end{cases} \tag{7}$$

where $t_{a_{min}}$ denotes the moment where the acceleration magnitude becomes minimal. The force trajectories for two different reach movements are depicted in Fig. 2.

B. Parameterization

The shape of the the finger trajectories emerging from the modulation of the system mainly depends on the spring constants of the virtual springs. For M springs the constants are collected in a vector $\mathbf{k} = (k_1, \dots, k_M)^T$. A major advantage of the proposed representation is that $\mathbf{k}$ can be estimated from the observation of the fingertip motion. For N fingers given their observed trajectory $\{\mathbf{p_i} = \{\mathbf{p_i}(0) \dots \mathbf{p_i}(t)\} | i = 1 \dots N\}$ and the force trajectory $\mathbf{f_{cen}}$ of the human hand, in order to estimate the springs constants, one has to rewrite Eq. 5 resulting in:

$$\mathbf{X_t k} = -\ddot{\mathbf{x}}_\mathbf{i}(t) - \mathbf{D}\dot{\mathbf{x}}_\mathbf{ij}(t) + \mathbf{c}^\mathbf{T}(\ddot{\mathbf{x}}_\mathbf{c_i}(t) + \mathbf{f}_{\mathbf{i,ext}}), \tag{8}$$

where matrix $\mathbf{X_t} \in R^{(N \cdot dim) \times M}$ describes the displacements of each fingertip along the spring m:

$$\begin{aligned} X_t(i * dim + 1, m) &= x_{m_i}(t) \\ &\vdots \\ X_t(i * dim + dim, m) &= x_{m_i}(t), \end{aligned} \tag{9}$$

with $n = 1, ..., dim \cdot N$ and $x_{m_i}(t) = x_{ij}(t)$ if body i is connected to j via spring m, and $x_{m_i}(t) = 0$ otherwise. $\ddot{\mathbf{x}}_\mathbf{i}$ represents the accelerations of the fingers in task space calculated from the observed finger trajectories. Since Eq. 8 forms a linear regression problem, we apply Singular Value Decomposition to produce $\hat{\mathbf{X}}_\mathbf{t}^{-1}$, the generalized inverse matrix of $\mathbf{X_t}$. Subsequently, by multiplying $\hat{\mathbf{X}}_\mathbf{t}^{-1}$ on both sides an estimation for $\mathbf{k}$ is obtained.

C. Representation of the reach movement

For a complete grasp representation, the reach movement extracted from human observation has to be represented in a way, which allows the adaptation of learned action to new situations. We investigated different approaches for the representations of movement primitives based on splines [16], Hidden Markov Models [17] and applied dynamic motor primitives (DMP) as proposed in [18]. A DMP provides a representation of a movement segment by shaping an attractor landscape described by a second order dynamical system. In [19], a motion representation based on DMPs is applied to represent pick-and-place actions. Similar to a linear spring system, using second order dynamics the basic point attractive system can be written as follows:

$$\tau \dot{\mathbf{v}} = k(\mathbf{g} - \mathbf{x}) - d\mathbf{v} - k(\mathbf{g} - \mathbf{x_0})s + kf(s) \tag{10}$$

$$\tau \dot{\mathbf{x}} = \mathbf{v}, \tag{11}$$

where $\mathbf{x}$ and $\mathbf{v}$ are position and velocity of the system; $\mathbf{x_0}$ and $\mathbf{g}$ are the start and goal position; τ is a temporal scaling factor; k acts like a spring constant; the damping term d is chosen such that the system is critically damped. To enable the encoding of arbitrarily complex movements, the non-linear function f is introduced which is defined as follows:

$$f(s) = \frac{\sum_i w_i \psi_i(s) s}{\sum_i \psi_i(s)}, \tag{12}$$

where $\psi_i(s) = \exp(-h_i(s - c_i)^2)$ are Gaussian basis functions, with center c_i and width h_i, and w_i are adjustable weights. The function f depends on a phase variable s, which monotonically changes from 1 towards 0 during a movement and is obtained by following equation:

$$\tau \dot{s} = -\alpha s \ , \tag{13}$$

where α is a pre-defined constant. Eq. 13 is referred to as canonical system. Based on a demonstrated movement $\mathbf{x}(t)$ with time steps $t = 0, \dots, T$ and its corresponding velocity and acceleration profile $\mathbf{v}(t)$ and $\dot{\mathbf{v}}(t)$, a DMP can be adapted to a movement by adjusting the weights w_i within f. Since f can be computed from Eq. 11 with the demonstration parameters and s can be obtained through integrating the canonical system, the adjustment of w_i is reduced to a linear regression problem. The DMP formulation features several advantageous properties such as guaranteed convergence towards the goal, spatial and temporal invariance and robustness against perturbations. However, the most important property lies in the simple adaption towards a new situation, which is mainly accomplished by specifying new start and goal positions. Once specified, the execution of the movement is attained through integration and evaluation of $s(t)$. The obtained

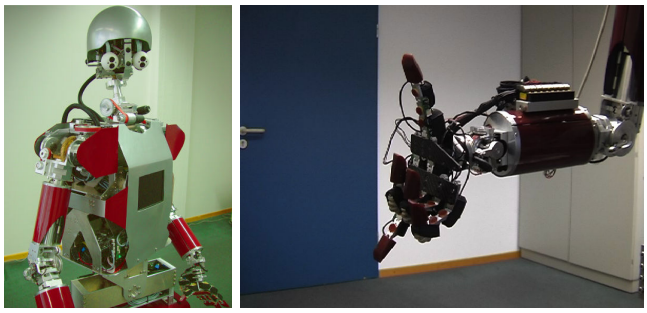

Fig. 3. Left: The humanoid robot ARMAR-IIIb. Right: Position-controlled right hand with 8 DoF.

phase variable then drives the non-linear function f which in turn perturbs the linear spring-damper system to compute the desired attractor landscape. Regarding the concept introduced in Section II-A, the acceleration profile $\dot{\mathbf{v}}(t)$ of the DMP can be used to derive the force trajectory $\mathbf{f}_{\mathbf{cen}}$ as defined in Eq. 7. For each specific grasp type a DMP is generated and stored in motion library along with the parameters of the grasp representation.

III. Experiments

A. Experimental Platform

The humanoid robot ARMAR-IIIb, which serves as the experimental platform in this work, is a copy of the humanoid robot ARMAR-IIIa [20]. From the kinematics point of view, the robot consists of seven subsystems: head, left arm, right arm, left hand, right hand, torso, and a mobile platform. The head has seven DoF and is equipped with two eyes, which have a common tilt and can pan independently. Each eye is equipped with two digital color cameras, one with a wide-angle lens for peripheral vision and one with a narrow-angle lens for foveal vision. The upper body of the robot provides 33 DoF: 2·7 DoF for the arms and three DoF for the torso. The arms are designed in an anthropomorphic way: three DoF for each shoulder, two DoF in each elbow and two DoF in each wrist. Each arm is equipped with a pneumatic-actuated five-fingered hand with eight DoF. The locomotion of the robot is realized using a wheel-based holonomic platform.

B. Observation

In the following, a method for capturing the human fingertip motion is presented. Based on [21] and [22], we implemented a real-time tracking algorithm combining particle filter and mean shift based on color information. The input to the system is a stereo color image sequence, captured with the built-in wide-angle stereo pair of the humanoid robot ARMAR-IIIb. To obtain accurate and robust position estimations of the fingertips, markers in the form of green caps are attached to the fingers. In the first frame, the color information of the markers is exploited to segment the images in order to determine the regions of interest surrounding the N fingertips. These regions are labeled and a color histogram model in HSV space is calculated. A single particle filter instance is applied to obtain an estimation for all fingertip positions based on the previous observation and the weighted particles. Each particle represents a set of N candidate regions whereas the corresponding weight is calculated by comparing the regions color histograms to the histogram model and the posture of these candidates to the one in the previous frame. The estimation is refined using an ordinary mean shift algorithm driving each region towards the maxima of the density distribution within the color histogram. Since the markers are of the same color, overlaps and false labeling might occur. For grasp observation, the assumption is made that the palm is facing towards the camera where in most cases the finger order Thumb → Index → Middle → Ring → Pinkie is valid. By representing the coordinates in polar space, it can be checked easily if this order is violated. If so, a search for candidate regions for the false estimated fingers is initiated in the vicinity of the previous configuration. Since this algorithm operates on monocular images, for each view a tracking instance is created whereas the 3D positions of the fingers are calculated by exploiting epipolar geometry. The presented framework is capable of online tracking of fingertip motion with a frame rate of 23 Hz on a 2 GHz dual core CPU. Sample images during the tracking process are depicted in Fig. 4.

C. Mapping and Execution

The grasp reproduction on ARMAR-IIIb is performed in several stages. In the first stage, the DMP for the reaching movement is adapted to position and the orientation of the object to be grasped. Subsequently, the force trajectory is derived to modulate the grasp representation resulting in the fingertip trajectories. The trajectories are mapped and scaled to fit the coordinate system of an intermediate hand model. For this purpose, the Master Motor Map (MMM), introduced in [23] and extended in [24], is used. The core feature of this framework is a reference kinematic model which facilitates the mapping from a human motion capture system to the kinematic structure of a robot. The model incorporates a biomechanical hand model with 21 DoF. By solving the inverse kinematics problem for the MMM hand model, one obtains the joint angle configuration respective to the given fingertip positions. Due to different measurements and less DoF of the robot hand, in order to attain a goal-directed reproduction, which additionally features a high similarity to the demonstrated human hand movement, joint

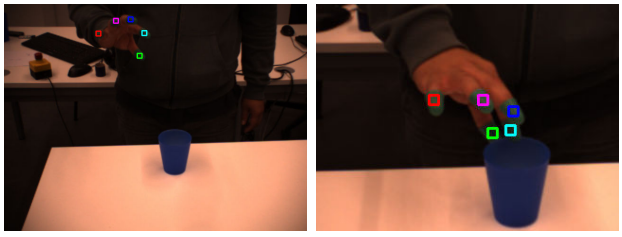

Fig. 4. Left camera views of the tracking method. Red denotes thumb region, light blue the index, blue the middle, pink the ring finger, and red pinkie region

angles as well as the desired fingertip positions have to be considered during execution. In [25], we developed an approach, which supports reproduction of observed human motion on the robot using non-linear optimization methods. To formulate an optimization problem for each finger which comprises displacements in Cartesian space regarding the fingertip position as well as the finger joints, a similarity measure is defined as follows:

$$S(\sigma) = 2 - \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\sigma}_i{}^t - \sigma_i\right)^2}{\pi^2} - \frac{\frac{1}{3}\sum_{k=1}^{3}\left(\hat{p}_k{}^t - p_k\right)^2}{\left(2\cdot l_{finger}\right)^2} \qquad (14)$$

with n representing the number of finger joints, $\sigma_i, \hat{\sigma}_i{}^t \in [0,\pi]$ and $p_k, \hat{p}_k{}^t \in \left[-l_{finger}, l_{finger}\right]$, whereas l_{finger} describes the considered finger length. The reference joint angle configuration is denoted by $\hat{\sigma} \in \mathbb{R}^n$, while $\hat{\mathbf{p}} \in \mathbb{R}^3$ stands for the desired fingertip position. The current fingertip position $\mathbf{p}$ can be determined by applying the forward kinematics of the robot to the joint angle configuration σ. Based on Eq. 14 and the joint constraints $\{(C_{min}, C_{max})\}$ of a robot with n joints, one obtains following constrained optimization problem:

$$\min S'(\sigma) = 2 - S(\sigma) \qquad (15)$$

$$\text{subject to} \quad C_{i_{min}} \leq \hat{\sigma}_i \leq C_{i_{max}} \qquad (16)$$

For solving Eq. 15, we apply the Levenberg-Marquardt algorithm. Following this optimization approach a trade-off is attained, which on the one hand results in an accurate finger positioning with small displacement error while it provides on the other hand a feasible robot joint angle configuration resembling the observed human configuration.

D. Results

The N-body system of the grasp representation is implemented in 2D. Therefore, currently only planar grasps which only require fingertip contact can be represented. For the reproduction of grasp, the force trajectory defined in Eq. 7 is applied to modulate the systems. The trajectories emerging from this modulation describe a fingertip posture in x,y direction in task space of the hand. The pose of the hand needed for grasping is obtained from the DMP movement and represented in the robot's platform coordinates. The proposed grasp representation is evaluated for a pinch, tripod, power, and lateral grasp. Based on image sequences captured by the humanoid robot, in addition to the fingertip trajectories the hand movement was determined by segmenting and tracking the hand by means of skin color information. From the resulting trajectory, a DMP is generated which, due its properties, allows the adaptation to new targets and the reproduction of smooth trajectories. The results of the motion reproduction of are depicted in Fig. 5.

Based on the fingertip trajectories, it is possible to estimate the spring constants of each virtual spring. The constants of the remaining springs are fixed independently of the considered grasp type. To maintain stability and avoid oscillation during the modulation, the system is assumed to be over-damped. For our experiments, the trajectories emerging

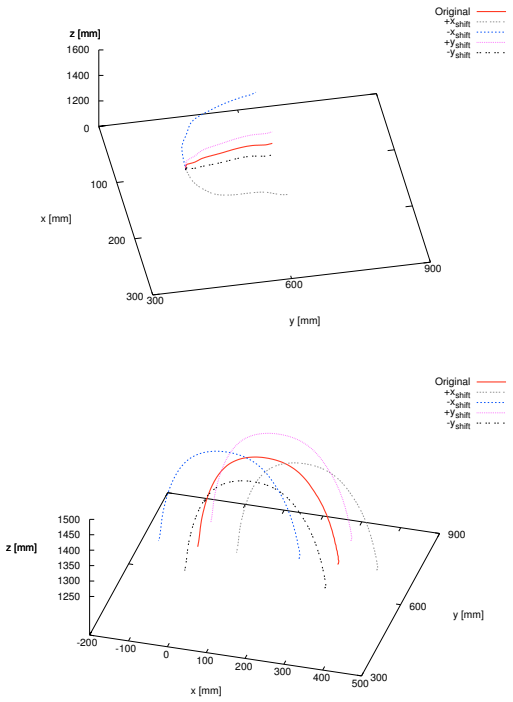

Fig. 5. Top: From a DMP reproduced reach trajectories towards different goals. Bottom: From a DMP reproduced place trajectories starting from different start points towards different goals.

from a grasp instantiations are compared to the observed movements. Due to the small number of contact points, the fingertip movements during a pinch grasp reproduction are highly informative in terms similarity to the human demonstration. As depicted in Fig. 6, using the virtual spring grasp representation, for thumb, index, middle, and ring finger, fingertip movements could be generated which are similarly shaped as the observed trajectories. Due to noisy motion data regarding the pinkie movement, the spring constants linked to the pinkie could not be estimated accurately enough leading to a slightly diverse trajectory. Furthermore, it could be observed that due to the integration of springs, we were able to produce smooth finger trajectories. To complete the specification of the grasp representation, the equilibrium lengths between the fingers were measured at a human subject whereas the hand is to be maintained in a very relaxing posture. On the platform, we were able to reproduce grasping movements where the hand preshapes and contact with the object is made at the end of the reaching movement. However, due to small number of DoF of the ARMAR-IIIb hand and its unstable pneumatic control mechanism, the optimized joint angle configuration could not be accurately reproduced. Results of the grasp reproduction for a pinch and power grasp are depicted in Fig. 8 and Fig. 7.

IV. Conclusion

In this work, we have presented a grasp representation which exploits finger movement synergies in task space and, hence, allows the formulation of grasps in a goal-directed and

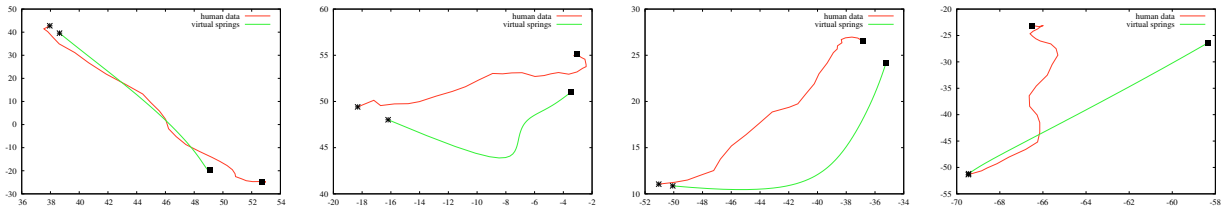

Fig. 6. Plots showing the trajectory of each finger during observation (red) and reproduction (green) of a pinch grasp. The star symbol denotes the start of each trajectory, whereas the box symbol represents the end. The measurements are given in mm. From left to right: index finger; middle finger; ring finger; pinkie.

low dimensional fashion facilitating several processes such as the observation of the human hand, which is a cumbersome task in joint space. Along with the parameter estimation procedure, a grasp can be learned and represented from human demonstration, even online. In order to reduce the dimensions of the control variables for trajectory generation, synergies on task space level were successfully established by means of the virtual springs. The result is a continuous grasp representation, which unifies the different grasp stages (preshape, reach, and enclose) leading to a smooth, human-like movement reproduction. In the near future, we focus on extending our implementation of the dynamical system from 2D to 3D in order to represent grasps which require additional contact areas besides the fingertips. Furthermore, we will extend our library of represented grasps and investigate how complex object representations can be integrated.

V. Acknowledgments

The work described in this paper was conducted within the EU Cognitive Systems project GRASP (FP7- 215821) funded by the European Commission.

References

[1] J. Napier, "The prehensile movements of the human hand," *Journal of Bone and Joint Surgery*, vol. 38B, no. 4, p. 902913, 19.

[2] M. R. Cutkosky and P. K. Wright, "Modeling manufacturing grips and correlation with the design of robotic hands."

[3] A. T. Miller, S. K. H. T. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *Proc. IEEE International Conference on Robotics and Automation (ICRA'03)*, Taipei, Taiwan, September 2003, pp. 1824–1829.

[4] M. Do, J. Romero, H. Kjellström, P. Azad, T. Asfour, D. Kragic, and R. Dillmann, "Grasp recognition and mapping on humanoid robots," in *IEEE International Conference on Humanoid Robots*, Paris, France, December 2009.

[5] M. Santello, M. Flanders, and J. F. Soechting, "Postural hand synergies for tool use," *The Journal of Neuroscience*, vol. 18, no. 23, pp. 10 105–10 115, 1998.

[6] M. T. Ciocarlie and P. K. Allen, "Hand posture subspaces for dexterous robotic grasping," *The International Journal of Robotics Research*, vol. 28, no. 7, pp. 851–867, 2009.

[7] M. Gabiccini and A. Bicchi, "On the role of hand synergies in the optimal choice of grasping forces," in *Proceedings of Robotics Science and Systems*, Zaragoza, Spain, June 2010.

[8] D. Pratichizzo, M. Malvezzi, and A. Bicchi, "On motion and force controllability of grasping hands with postural synergies," in *Proceedings of Robotics Science and Systems*, Zaragoza, Spain, June 2010.

[9] J. Steffen, R. Haschke, and H. Ritter, "Towards dextrous manipulation using manipulation manifolds," in *Proc. IEEE International Conference on Intelligent Robots and Systems (IROS'08)*, Nice, France, September 2008, pp. 2738–2743.

[10] B. Hoff and M. A. Arbib, "Models of Trajectory Formation and Temporal Interaction of Reach and Grasp," *Journal of Motor Behaviour*, vol. 25, pp. 175–192, 1993.

[11] M. Jeannerod, M. A. Arbib, G. Rizzolatti, and H. Sakata, "Grasping objects: the cortical mechanisms of visuomotor transformation," *Trends in Neurosciences*, vol. 18, no. 7, pp. 314–320, 1995.

[12] N. Vahrenkamp, M. Do, T. Asfour, and R. Dillmann, "Integrated grasp and motion planning," in *Proc. IEEE International Conference on Robotics and Automation (ICRA'10)*, Taipei, Taiwan, 2010, pp. 2883–2888.

[13] S. Ekvall and D. Kragic, "Interactive grasp learning based on human demonstration," in *Proc. IEEE International Conference on Robotics and Automation (ICRA'04)*, New Orleans, USA, April 2004, pp. 3519–3524.

[14] A. Schiegg, H. Deubel, and W. X. Schneider, "Attentional selection during preparation of prehension movements," *Visual Cognition*, vol. 10, pp. 409–431, 2003.

[15] M. Arbib, T. Iberall, and D. Lyons, "Coordinated control programs for movements of the hand," *Experimental Brain Research Supplement*, vol. 10, pp. 111–129, 1985.

[16] A. Ude, C. G. Atkeson, and M. Riley, "Programming Full-Body Movements for Humanoid Robots by Observation," *Robotics and Autonomous Systems*, vol. 47, no. 2-3, pp. 93–108, June 2004.

[17] T. Asfour, P. Azad, F. Gyarfas, and R. Dillmann, "Imitation Learning of Dual-Arm Manipulation Tasks in Humanoid Robots," *International Journal of Humanoid Robotics*, vol. 5, no. 2, pp. 183–202, December 2008.

[18] A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Movement imitation with nonlinear dynamical systems in humanoid robots," in *Proceedings of the IEEE International Conference on Robotics and Automation)*, 2002.

[19] P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal, "Learning and generalization of motor skills by learning from demonstration," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Kobe, Japan, 2009.

[20] T. Asfour, K. Regenstein, P. Azad, J. Schröder, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control," in *IEEE/RAS International Conference on Humanoid Robots*, 2006.

[21] E. Maggio and A. Cavallaro, "Hybrid particle filter and mean shift tracker with adaptive transition model," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2005.

[22] S. Yonemoto and M. Sato, "Multitarget Tracking using Mean-Shift with Particle filter based intialization," in *12th International Conference on Information Visualisation*, 2008, pp. 314–320.

[23] P. Azad, T. Asfour, and R. Dillmann, "Toward an Unified Representation for Imitation of Human Motion on Humanoids," in *IEEE International Conference on Robotics and Automation*, Rome, Italy, April 2007.

[24] S. Gärtner, M. D. Azad, T. Asfour, R. Dillmann, C. Simonidis, and W. Seemann, "Generation of Human-like Motion for Humanoid Robots Based on Marker-based Motion Capture Data," in *ISR 2010 (41st Internationel Symposium on Robotics*, Munich, Germany, June 2010.

[25] M. Do, P. Azad, T. Asfour, and R. Dillmann, "Imitation of Human Motion on a Humanoid Robot using Non-Linear Optimization," in *IEEE International Conference on Humanoid Robots*, Daejeon, Korea, December 2008, pp. 545–552.

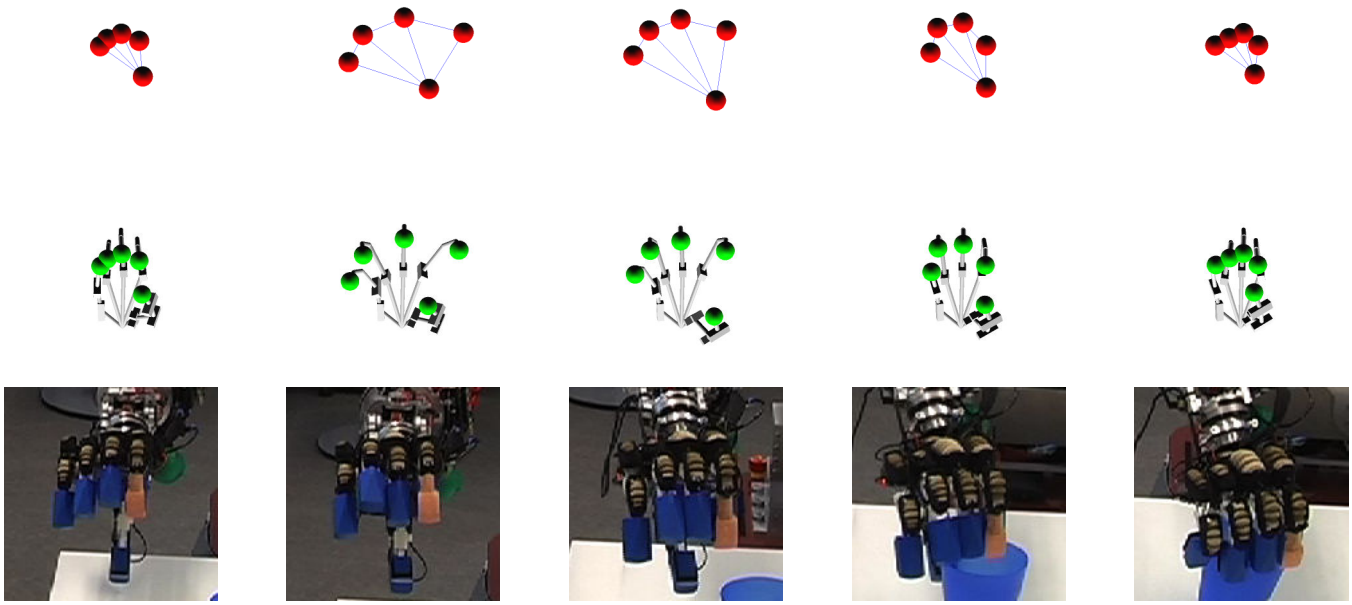

Fig. 7. Top: Image sequence depicting the capture human fingertip motion for a power grasp. Middle: The MMM hand model maintaining a hand posture that matches the fingertip motion. Bottom: Reproduction of the represented power grasp.

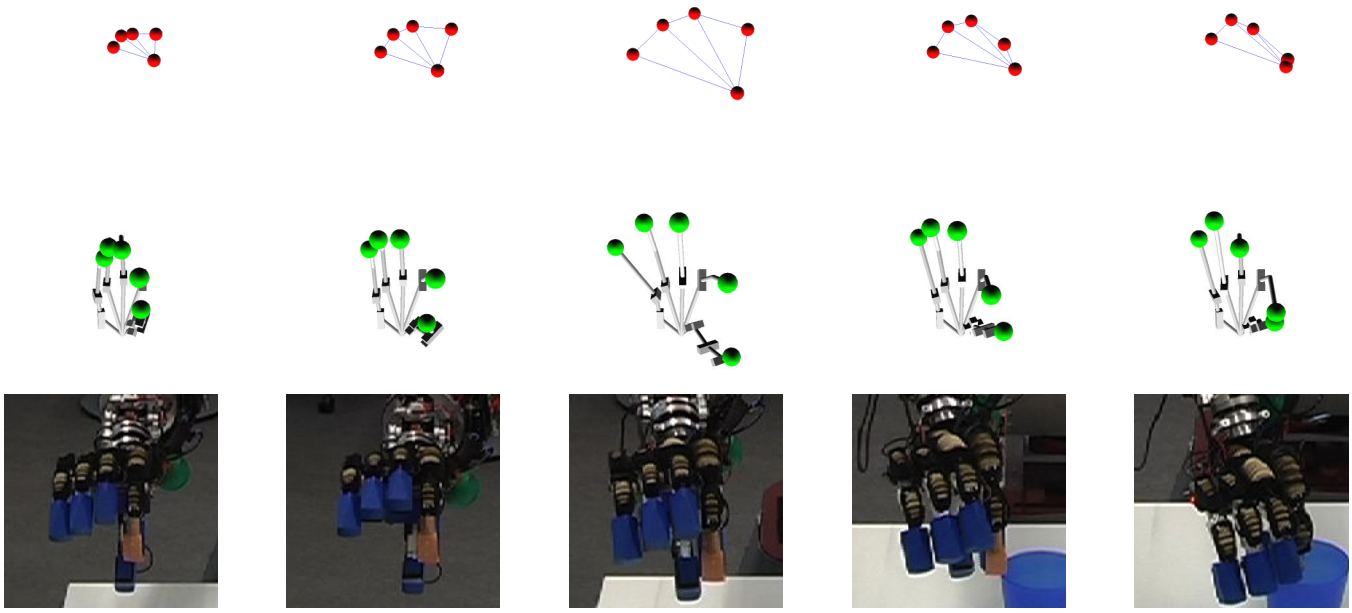

Fig. 8. Top: Image sequence depicting the capture human fingertip motion for a pinch grasp. Middle: The MMM hand model maintaining a hand posture that matches the fingertip motion. Bottom: Reproduction of the represented pinch grasp.