

Extracting Postural Synergies for Robotic Grasping

Javier Romero, Thomas Feix, *Member, IEEE*, Carl Henrik Ek, *Member, IEEE*, Hedvig Kjellström, and Danica Kragic, *Senior Member, IEEE*

Abstract—We address the problem of representing and encoding human hand motion data using nonlinear dimensionality reduction methods. We build our work on the notion of postural synergies being typically based on a linear embedding of the data. In addition to addressing the encoding of postural synergies using nonlinear methods, we relate our work to control strategies of combined reaching and grasping movements. We show the drawbacks of the (commonly made) causality assumption and propose methods that model the data as being generated from an inferred latent manifold to cope with the problem. Another important contribution is a thorough analysis of the parameters used in the employed dimensionality reduction techniques. Finally, we provide an experimental evaluation that shows how the proposed methods outperform the standard techniques, both in terms of recognition and generation of motion patterns.

Index Terms—Grasping, humanoid robots, motion analysis, multifingered hand.

I. INTRODUCTION

ENCODING and representation of grasping movements in robotic applications such as mapping and planning are often inspired by the analysis of human motion data [1]–[5]; see Fig. 1. Human grasping has been studied widely in neurophysiology and psychology, where the goal is to understand the processes behind the control of movement [6]–[9]. A central result of such studies is that there exist strong correlations between the spatial configurations of fingers. The correlations allow the extraction of compact representations, often referred as *synergies*, e.g., a lower dimensional space \mathbf{X} in Fig. 1, which can concisely describe complex motions in the high-dimensional space \mathbf{Y}^h . Most of the work in neurophysiology use linear models to encode postures and movements [7]–[9]. Even comparative studies of techniques used for synergy extraction like [10] employ linear methods only. The first contribution of this paper is to show that, apart from being high dimensional, finger movements are

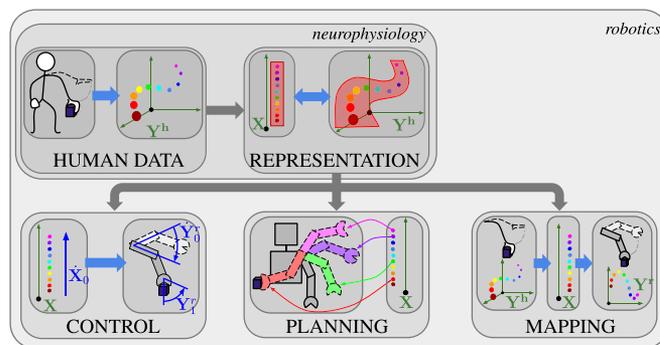


Fig. 1. Human motion data have been used in neurophysiology to study how the brain and the nervous system command motor actions. The results of these studies have been used for control of dynamic movement in robots (Control), search for suitable poses for a specific task (Planning), and transferring kinematic actions between different embodiments (Mapping).

also highly nonlinear. A more natural approach is thus to encode them using techniques that deal with both high-dimensionality and nonlinearity, as also suggested in [3].

In neurophysiology, it is argued that synergies parameterize the modulation of muscle forces for control of human posture [6]. Thus, synergies provide a natural modeling paradigm for robot control, where control laws in low-dimensional space \mathbf{X} can drive the forces applied on the higher dimensional robot space \mathbf{Y}^r ; see Fig. 1. Synergies have been used to design reference robot hand movements capable of adapting to external forces [11]. Reference movements become lower dimensional by using synergies, with the obvious advantages of requiring fewer degrees-of-freedom (DoF) for control. In motion planning, the computational cost of searching for adequate kinematic configuration increases exponentially with the dimensionality of the structure. Ciocarlie and Allen [12] reduced the complexity by searching for grasping postures in a postural synergy space.

Synergies have also been used as a representation that transcends differences in embodiments—the “correspondence problem” [13]. This problem, also referred to as *Mapping* in Fig. 1, deals with transferring postures or movements from one agent to another, e.g., mapping human postural space \mathbf{Y}^h to robot space \mathbf{Y}^r . While in [12] the correspondence is designed manually (i.e., the eigengrasp-robot correspondences $\mathbf{X} \rightarrow \mathbf{Y}^r$ are manually assigned), it is found automatically in [14] by employing the concept of Virtual Fingers. In summary, the efficient parameterization provided by postural synergies have been used to address inherent problems related to the high-dimensional representation. In this paper, we evaluate the use of nonparametric latent variable models for extracting postural synergies. These models are capable of handling both the high-dimensionality and the nonlinear structure of human motion data. We conduct a thorough analysis of the employed model and evaluate

Manuscript received July 20, 2012; revised March 1, 2013; accepted June 24, 2013. Date of publication July 19, 2013; date of current version December 2, 2013. This work was supported by the European Union under Grant FLEXBOT (FP7-ERC-279933) and Grant TOMSY (IST-FP7-270436) and by the Swedish Foundation for Strategic Research. This paper was recommended for publication by Associate Editor J. Dai and Editor W. K. Chung upon evaluation of the reviewers’ comments.

J. Romero is with the Perceiving Systems Department, Max Planck Institute for Intelligent Systems, Tübingen 72076, Germany (e-mail: jromero@tuebingen.mpg.de).

T. Feix is with the Department of Mechanical Engineering and Materials Science, Yale University, New Haven, CT 06511 USA (e-mail: thomas.feix@yale.edu).

C. H. Ek, H. Kjellström, and D. Kragic are with Computer Vision and Active Perception Lab, Centre for Autonomous Systems, CSC, The Royal Institute of Technology KTH, Stockholm 100 44, Sweden (e-mail: chek@kth.se; hedvig@kth.se; dani@kth.se).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TRO.2013.2272249

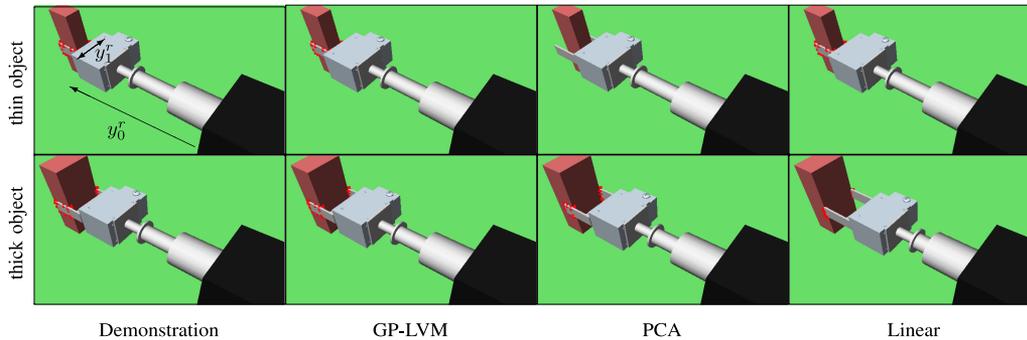


Fig. 2. Simple grasp execution: The columns represent, from left to right, one of the original examples, i.e., GP-LVM, PCA, and Linear models. The GP-LVM models mimics almost perfectly the training example. PCA fails to make contact for the thin object and grasps the thick object slightly prematurely. The Linear model grasps properly the thin object but collides with the thick object. Graphs are generated with OpenRave [23].

experimentally its advantages with respect to the standard techniques both in terms of recognition and generation of motion patterns.

We start by reviewing related work in Section II. Section III outlines dimensionality reduction methods and Section IV presents the dataset used for the evaluation. We analyze the extracted grasping models in Section V, evaluate their performance in Section VI, and conclude the work in Section VII.

II. MOTIVATION AND RELATED WORK

Grinyagin *et al.* [9] classify synergies into three separate types. First, static postural synergies, that refer to models of the correlation between single kinematic poses, e.g., [7]. Second, kinematic synergies that consider time dependent correlation of postures during an action, e.g., [8]. Finally, muscle synergies address the covariation of lower level representations of movement such as electromyographic activity [6], [15]. While muscle synergies are specifically bound to a human embodiment, postural synergies, and kinematic synergies have inspired a large body of work in robotics [7], [11]. Synergies are commonly extracted using linear dimensionality reduction methods such as principal component analysis (PCA) [16]. One of the earliest approaches is presented in [7]. The authors showed that a substantial part of the variations of grasping hand poses (80% of the data *variance*) can be expressed as a 2-D linear combination of hand joints.

While Santello *et al.* [7] showed the correlation of joints for different static hand poses, later research focused on the temporal correlation (i.e., kinematic synergies) of the hand pose while executing specific grasps. In [9], multiple executions of precision grips are analyzed with PCA to conclude that a 1-D direction can explain more than 97% of the data variance. The data from six different subjects and three different grasping conditions were analyzed separately, generating a different 1-D manifold for each of these series of 20 trials. Mason *et al.* [8] studied the correlation in the position of different parts of the hand for specific grasps applied to different objects, again using similar techniques as Santello *et al.* They concluded that for each subject and grasp, more than 96% of the variance could be explained by a 1-D manifold.

Ciocarlie *et al.* [12] introduced the concept of Eigengrasps based on the grasp synergies defined in [7]. Grasps are planned

using Eigengrasps, making the optimization of the hand pose computationally tractable. Eigengrasps were also used in [4] to control the 12-DoF of a robotic hand. In [17], the grasping control of a 17-DoF hand was performed by moving on a 2-D manifold extracted with Isomap. In [18], correlations between wrist and finger movements were modeled, validated, and applied to solve control of redundant DoF. There is also some work focused on imposing certain desired characteristics on the synergies. In [19], the low-dimensional space variables are constrained to represent task evolution (time) and relevant parameters for the action (such as object size). Another constraint is used in [20], where the method interpolates between high-dimensional actions by imposing certain spatial relations between the training actions in a low-dimensional representation which resembles the task space. Another area of robotics influenced by the concept of synergies is robotic hand design [2], [21], [22].

The majority of the aforementioned approaches rely on linear dimensionality reduction methods. It is clear that the linear approximation of the mapping can result in inaccuracies. For motivational purposes, let us depict a concise example of how a linear model that represents most of the variance still fails to represent the motion correlation in an action such as reaching and grasping.

Our example consists of a set of robot grasping executions for which we wish to acquire a low-dimensional representation. For illustrative purposes, our robot is fairly simple. It has two components: a telescopic arm and a gripper, each with 1-DoF (see Fig. 2). Our goal is to obtain a 1-D representation of a grasping task. Despite its simplicity, this example shows that nonlinearities in high-dimensional space can be critical. Grasping an object can be roughly divided into two components: transport phase and grasping phase. During the transport phase, the hand slowly reaches its maximum aperture (considering the hand being initially closed) and in the grasping phase, the hand closes very fast, while the distance to the object decreases only slightly. The following equation formalizes this idea:

$$y_{t,1}^r \propto (1 - y_{t,0}^r)^{\frac{1}{3} + \epsilon_0} + \epsilon_1 \quad (1)$$

where y_t^r represents the 2-D configuration of the robot at every time, $y_{t,1}^r$ represents the gripper aperture at time t , $y_{t,0}^r$ represents the arm extension at time t , and ϵ_i represent different noise

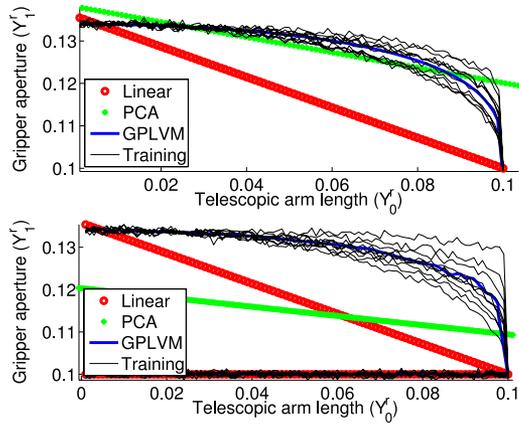


Fig. 3. Grasp trajectories generated from different models computed from the ten demonstrations to be imitated [Training, generated from (1)]. *Linear* is a linear interpolation between the beginning and end gripper aperture. *PCA* and *GP-LVM* extract a 1-D representation of data and generate the trajectory by linearly sampling that 1-D manifold. The top figure shows only approach and grasp, and the bottom figure includes retreat.

terms. Ten grasping sequences $\{y^r\}$ were generated (labeled as “Training” in Fig. 3), and three models were extracted from them. The first one, which is marked as “Linear” in top Fig. 3, is a linear interpolation between the maximum and minimum aperture with linear decrease of the arm length. The other two are the trajectories generated by sampling a 1-D space extracted dimensionality reduction methods PCA and GP-LVM (Gaussian Process Latent Variable Model).

The nonlinear method GP-LVM best resembles the training trajectories which were highly nonlinear. Let us consider the consequences of the linearization. In Fig. 3, when the linear trajectories are below the example trajectories (all the Linear model, and part of the PCA one), the gripper’s aperture is smaller than in the original examples, which means that the gripper can potentially collide with the object; see Linear column, thick object row, in Fig. 2. When the line is above the examples (last stage of PCA model), the gripper is more open than in the examples, and therefore, the gripper might fail to enter contact with the object (PCA column, thin object row in Fig. 2).

The relation between the two joints in this action set was functional, and the PCA reconstruction is relatively good, even though it fails in some cases. If the actions also includes the retreat movement, the relation becomes more nonlinear and nonfunctional and, thus, is not possible to model as a causal relation between master and slave [3]; see the bottom part of Fig. 3. The linear manifold computed by PCA averages the approach and retreat actions, setting a gripper aperture which is too small to approach the object. However more importantly, it cannot retreat with a constant aperture.

The relation between the transport and grasp components in a grasping action is inherently nonlinear. Trying to model it as a linear manifold can result in early collisions or failure to contact the object. This can happen even when a large part of the variance is kept in the linear manifold. PCA managed to represent 98.6% of the variance of the grasp-approach action with its linear representation and still fails to properly grasp

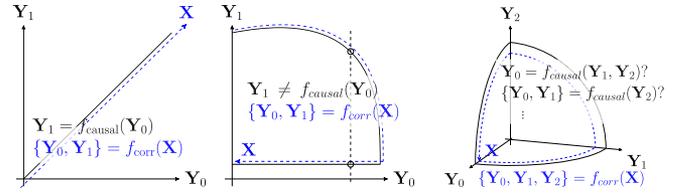


Fig. 4. Applicability of causal relation between data dimensions. Y_i represent robot dimensions such as finger joints or gripper aperture versus distance to the object, while X represents a (nonlinear) synergy that can model the correlations on Y_i . (Left) The master-slave relation between different Y_i is only applicable in simple cases; (middle) such relation is not possible when the relation between variables is not functional, because points with the same coordinate Y_0 (e.g., arm extension) have different coordinate Y_1 (e.g., hand aperture).

the object. The amount of variance represented by PCA for the grasp-approach-retreat action was also high, i.e., 78.6%.

We now relate our work to recent approach of [3] that improves the robustness to perturbations in grasping actions by exploiting coupling between reaching and grasping. The coupling acts as an “attractor” that allows the task to be perturbed in terms of goal position or grasp type. The approach is based on the assumption that there is a causal relationship between two processes (reach and grasp). This implies that the system is *a priori* divided into two parts and the role of the master and slave are manually assigned. While this may seem reasonable in the case of reaching and grasping, causal assumptions easily break down when the task becomes more complex. For example, extending the reaching/grasping with a retreat movement breaks the causal relationship between reaching and grasping, because closeness to the body does not *cause* large hand aperture any more. The relation between reach and grasp is no longer functional, i.e., $y = f(x)$ is unique; see the middle part of Fig. 4. Another problem of the causal relation between master and slave is that it only relates two variables or sets of variables. Imagine that the dataset to model is a path embedded on a sphere sector in 3-D space Y_0, Y_1, Y_2 ; see the rightmost part of Fig. 4. According to [3], multiple variables can be grouped into master or slave. This approach implies that one of the dimensions generate the other two, or *vice versa*. However, that model misses important correlations within the 2-D set.

Latent variable methods tackle the correlations problem not by considering causal relationships between dimensions but by considering them as generated from a shared underlying variable. Often it is assumed that this latent variable is parameterized on, or near to, a low-dimensional manifold (X in Fig. 4) that can define the synergy space. The following sections explain the underlying principles behind PCA and GP-LVM, focusing on the implications of the assumptions made about the data.

III. METHODS FOR EXTRACTING POSTURAL SYNERGIES

We first formulate the problem of dimensionality reduction, and then, introduce the approaches evaluated in the paper. We refer to data in three separate representations: *observed*, which indicates the “raw” data, *intrinsic*, which is the representation aligned with the DoF of the data, and *latent*, which represents the parameterization of the data learned by the dimensionality

reduction approach. We strive for a scenario in which the latent variable representation is aligned with the intrinsic representation, which we argue is the synergic representation of the data.

For many scenarios, observed data are commonly different from its intrinsic representation, and the task of dimensionality reduction is to, given data in a specific representation, recover the intrinsic representation. The problem is formalized as follows. Given a set of N data points $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ where $\mathbf{y}_i \in \mathbb{R}^D$ (e.g., $N = 10$ time-instances of $D = 2$ dimensional joints of the robot in Fig. 2), we assume this to have been generated from an intrinsic representation $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$, $\mathbf{x}_i \in \mathbb{R}^q$ through the generative mapping f , $\mathbf{y}_i = f(\mathbf{x}_i)$. Further, we will assume the observed representation to be an over parameterization (i.e., $q < D$). The objective of dimensionality reduction is to recover \mathbf{X} from \mathbf{Y} .

The problem is severely ill constrained since an infinite combination of input representations \mathbf{X} and mappings f may be generated \mathbf{Y} . There are two main branches of work in dimensionality reduction, *spectral* and *generative*. Spectral approaches assume the generative mapping f to have a smooth inverse. This is different compared with the generative class of models which directly tries to model the generative mapping. The spectral assumption is stronger and does constrain the solution space further compared with the generative. This implies that while the generative models are applicable to a larger range of data, recovering the solution is likely to be much more challenging. There are both linear and nonlinear formulations of the methods.

In this paper, we apply four different algorithms: PCA, GP-LVM, locally linear embedding (LLE), and Isomap. PCA is a linear model, while the other three are nonlinear. We will not evaluate LLE and Isomap in depth, as our initial results showed that they are not suitable for this type of data. Consequently, we will focus on PCA being the dominant algorithm for extracting synergies and GP-LVM, being a flexible nonlinear algorithm.

A. Principal Component Analysis

The objective of PCA is to find a low-dimensional hyperplane which maximizes the variance of the data projected onto it. Mathematically, this implies finding a low-rank approximation of the covariance matrix. The best rank solution (in terms of the Frobenius norm) can be found in closed form and corresponds to a linear projection. PCA relies on the global statistics of the data and optimizes linearly the reconstruction of the data covariance. Thus, it applies well to data with noise of low variance and when the intrinsic signal occupies a linear subspace. However, these assumptions do not apply when a significant portion of the observations corresponds to noise and/or the correlations in the data are nonlinear.

B. Gaussian Process Latent Variable Model

Spectral and generative approaches share the assumption that the mapping from the latent space to observed space takes functional form. We will refer to the observed data \mathbf{Y} as the representation of the latent space \mathbf{X} generated through the mapping f . By assuming that the observations we have gathered have been corrupted by additive Gaussian noise, i.e.,

$\mathbf{y}_i = f(\mathbf{x}_i) + \epsilon$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^{-2}\mathbf{I})$, we can formulate the likelihood of the data under the model as

$$p(\mathbf{Y}|\mathbf{f}, \sigma^{-2}) = \prod_i \mathcal{N}(\mathbf{y}_i|\mathbf{f}_i, \sigma^{-2}). \quad (2)$$

This formulation requires to find both the generative mapping f and the latent representation \mathbf{X} that maximize the likelihood. This is a severely ill-constrained problem as an infinite number of combinations of mappings and latent structures could have generated the data. To make the problem tractable, priors (i.e., preferences over particular types of solutions) have to be introduced. The methodology used in this paper will introduce a prior over the possible mappings f in the form of a distribution over functions, a Gaussian Process priors (\mathcal{GP}) [24]. This approach is referred to as a GP-LVM [25]. The definition of our preference over mappings as a \mathcal{GP} allows us to obtain the marginal likelihood by averaging over all possible smooth mappings

$$p(\mathbf{y}|\mathbf{X}, \sigma) = \int p(\mathbf{y}|\mathbf{f}, \sigma)p(\mathbf{f}|\mathbf{X})d\mathbf{f}. \quad (3)$$

Thus, we still need to specify the \mathcal{GP} prior that encodes a relevant preference over generative mappings.

A \mathcal{GP} is specified by a mean $\mu(\cdot)$ and a covariance function $k(\cdot, \cdot)$. We are generally not interested in the mean function and remove this DoF by assuming it to be a the constant zero function. More interesting is the covariance function $k(\cdot, \cdot)$, which encodes how different points in the observed space $\mathbf{y}_{i,j}$ covary as a function of their latent representation $\mathbf{x}_{i,j}$. We wish to encode a preference toward generating mappings which are smooth. This is done by specifying a covariance function where points that are close to each other have a large covariance. As such, we choose a radial-basis function where the covariance between two points is a function of their interdistance

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_r e^{-\frac{\gamma}{2}(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)}. \quad (4)$$

The covariance matrix can also encode prior knowledge about the structure of the data, e.g., the grasp type of each data point (so that same grasps are close to each other) or the neighbors in a time series of a particular data point. These priors can result in better manifolds but have not been used in this paper to show that simple interpoint distances can already reveal the underlying grasping synergies. In the GP-LVM, each of the dimensions in the observed data is modeled using an independent \mathcal{GP} , which leads to the marginal likelihood of the model

$$P(\mathbf{Y}|\mathbf{X}, \theta) = \prod_{j=1}^D \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{K}|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{y}_j^T \mathbf{K}^{-1} \mathbf{y}_j} \quad (5)$$

where \mathbf{K} is the matrix resulting from evaluating the covariance function on the latent representation \mathbf{X} . The solution is found by minimizing the negative log of (5) with respect to \mathbf{X} and the parameters of the covariance function θ . This is done using gradient-based methods which require an initialization of the parameters. Throughout the experiments presented in this paper, we initialize the GP-LVM from the solution provided by PCA.

While the smooth covariance in (4) implies that points close in latent space remain close in the observed space, the opposite

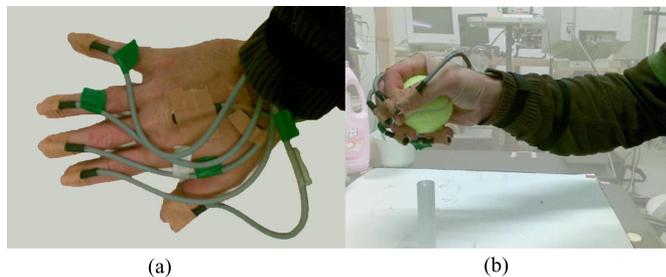


Fig. 5. Magnetic sensors setup. (a) Five sensors are placed on the fingertips and one on the wrist. (b) Grasp posture for grasp 11.

is not guaranteed. Spectral methods make an assumption that the inverse of the generative mapping is also a function. This is not true in the generative case where it is quite possible for two different latent locations to map to the same point in the observed space. This type of constraint can be encoded into the model by representing the latent locations in (5) as a function of the observed data \mathbf{Y} . The latent representations can then be learned indirectly by finding the parameters of this mapping that maximizes the marginal likelihood of the observed data. This approach was introduced in [26] as the back-constrained GP-LVM. In this paper, we will use a back-constraint where the latent coordinates are parameterized by the observed data using kernel regression.

C. Spectral Nonlinear Models

There exist several nonlinear spectral dimensionality reduction models [27] such as Isomap [28], maximum variance unfolding [29], and LLE [30]. Therefore, it may seem unfair to compare the nonlinear GP-LVM with the linear PCA. However, our reasons for focusing on the comparison with PCA are three fold. First, PCA is the most commonly applied method in the field of synergy extraction, in particular for grasping data. Therefore, it is natural that we compare our method with the state-of-the-art in the field. Second, LLE and Isomap do not provide mechanisms to map unseen data into an existing manifold. Third, our experiments show that the performance of LLE and Isomap on our data are worse than PCA (see Figs. 8 and 13), which has also been confirmed in [31].

IV. DATASETS

The dataset was generated from five subjects (three male, two female), and it is available from <http://grasp.xief.net/>. All subjects were right handed and have not reported any hand disabilities. The hand lengths were 185.2 ± 13.3 mm, and the hand width were 81.1 ± 7.4 mm. A Polhemus Liberty system with six magnetic sensors was used for recording the data. Each sensor provided its orientation and position with respect to a base point as a 4-D quaternion and a 3-D vector. The spatial and angular resolution of each sensor is 0.8 mm and 0.15° , respectively. One sensor was applied to each fingertip, positioned on the fingernail and one was placed on the dorsum of the hand; see Fig. 5(a). The fingertip sensors were calibrated so that fingers resting on the table and pointing forward had zero rotation. Because of the

limited articulability of finger phalanges, the measured positions and orientations provide a good approximation of the full-hand configuration. Measuring additional phalanx orientations would improve accuracy in grasps that do not involve fingertip contact, at the cost of further constraining the mobility of the subject.

The sensors provide position and orientation of the fingertips. This data are converted relative to the reference sensor, placed on the back of the hand, representing our task-space data. Some studies [7], [9] prefer the use of joint angle data. We argue that task space is preferable for our task. First, data in task space are easier to compare because, in joint-space data, proximal joints have a higher impact in the pose than distal joints. Second, task-space data are more portable across embodiments, because it directly encodes the relation of the hand with the object (the contact points and normals).

The subjects were asked to perform 31 different grasp types [32] with their right hand on an object typical for the specific grasp. They were shown a picture of each grasp and a demonstration of the grasp was performed if the subject had problems mimicking the grasp on the picture. Initially, they placed the hand in front of them on the table in a flat-hand posture. Upon a starting signal, they grasped an object with the desired grasp type, lifted the object [this moment is shown in Fig. 5(b)], put it down again, and retreated the hand to the starting position. The data recording started when the hand began to move and ended when the hand was put back to the initial position. Each grasp was discretized into 30 uniformly distributed time instances for which we recorded the fingertip poses. In summary, this means that we have acquired a database consisting of five subjects performing 31 different grasps resulting in 4650 data points. Furthermore, subjects were asked to perform each grasp twice. The first instance was used for testing and the second for training.

V. SYNERGETIC REPRESENTATIONS

Fig. 6 shows a schematics of the evaluation framework. For representing temporal information, as well as multiple subject variance in low-dimensional space, we use Gaussian mixture regression (GMR) [31], [33]; see the bottom left module in Fig. 6. We first briefly explain the use of GMR, and then, examine qualitatively the distribution of the low-dimensional representation extracted with different dimensionality reduction techniques.

A. Gaussian Mixture Regression of Grasps

Our modeling of low-dimensional action data has two main parts. First, a mixture of Gaussians is fitted to the data after extending it with a time dimension; see second column of Fig. 7. This Gaussian Mixture Model (GMM) is computed by initializing the mixture of Gaussians with k -means and optimizing it through expectation maximization. Empirically, we found that using more than three Gaussians did not improve the generalization capabilities of the model; see Section VI-B2. Second, a hand posture is inferred for each time step by using GMR. This creates a continuous path through the latent space that describes the grasp (see the third column of Fig. 7 and bottom left module of Fig. 6). That path has a mean and a variance. The

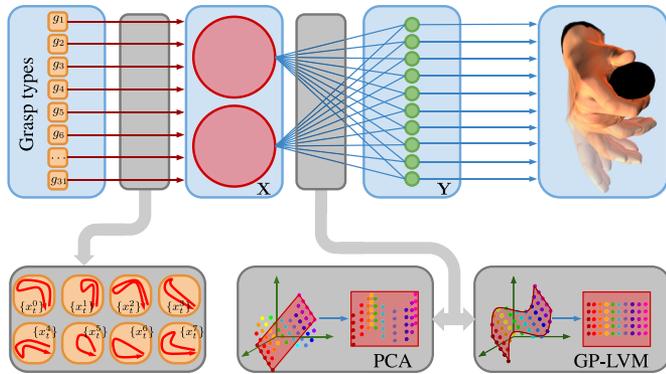


Fig. 6. Schematic overview of our system. We model different grasp types g_i (top left box) as time series of poses $\{x_t^i\}$ in low-dimensional space \mathbf{X} . We show eight examples of those models in the lower left box. The low-dimensional (2-D in this example) space \mathbf{X} is extracted from the high-dimensional space \mathbf{Y} using unsupervised dimensionality reduction methods, such as (bottom middle) PCA and (bottom right) GP-LVM. The hand poses (right) are fully described by the space \mathbf{Y} . The arrows follow the generative meaning of our model: grasp models describing low-dimensional data that can be mapped into high-dimensional space and be executed. However, the extraction process runs from right to left: high-dimensional executions mapped into low-dimensional space that can be analyzed to create different grasp models.

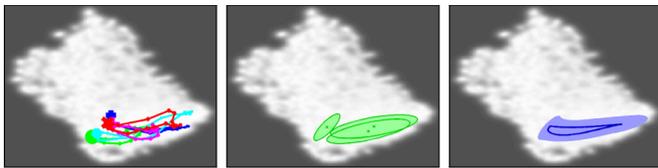


Fig. 7. Projection of grasp number 1 into latent space, GMM fitting, and GMR regression. Each color in the leftmost plot represents the projected path of grasp 1 for a single subject. These paths are encoded using GMR to create a compact representation (rightmost plot).

paths corresponding to four representative grasps can be found in Fig. 10.

The GMM/GMR representation of the grasps is a tool that can be used for several purposes. One is the generation of new actions under some constraints [34]. In our case, this could help to generate an action composed of two grasps without coming back to the rest position between them. The second grasp can be constrained to start in a specific pose or after a specific time frame of the first grasp. The GMR can be optimized taking into account this constraint, thereby providing a smooth transition between those grasps. In the remainder of this section, we present the representations extracted with PCA, GP-LVM, Isomap, and LLE.

B. Principal Component Analysis

For PCA, we decided to use 2- and 3-D manifolds for three reasons: first, because their accuracy is similar to the accuracy reported in [7]; second, because the accuracy increases very slowly by adding further dimensions; and third, because the visualization of the manifolds is difficult for dimensionalities beyond three.

PCA 2D [see Fig. 8(a)]: To visualize the shape of the space, the data points are plotted as white dots over a dark background. The points cover a rather narrow arc in space, on which all

movements are situated. The initial starting posture is on the right side of the space. During the approach movement, the trajectory progresses leftwards. The overall flexure of the fingers determine how far the trajectory moves away from the start point. The reason for this is that the starting posture is a flat hand, and therefore, increasing the finger flexion increases the difference to the starting posture.

C. Spectral Nonlinear Methods

For the sake of conciseness, we have not explained in depth the spectral nonlinear methods. The reason is that their results are clearly worse than PCA and GP-LVM for our problem. Fig. 8(c) and (d) shows that executions of the same grasp by different subjects are located in very different positions in the manifold and have very diverse directions (some trials in Fig. 8(d) are almost perpendicular to the rest). Moreover, since different trials of the same grasp are scattered around the manifold, different grasps will be hardly separable.

D. Gaussian Process Latent Variable Model

Each point in latent space is connected via a Gaussian process mapping to a point in high-dimensional space. The mapping predicts the mean and the variance of a point in high-dimensional space given the latent location. The mean can be directly used as the reconstruction of a latent point in high-dimensional space. The variance, which is connected to the prediction, can be used to quantify the confidence the model has while generating the point in high-dimensional space. How fast the variance increases while moving away from data points gives a hint on the belief of the model to generalize to previously unseen points. For simplicity and coherence with the rest of the methods, Fig. 8 only shows the data points themselves. However, Fig. 7 shows the variance of the GP-LVM as the brightness of the background (white corresponding to low variance and dark corresponding to high variance).

The points in the GP-LVM 2-D space [see Fig. 8(b)] cover a relatively larger area compared with the PCA 2-D space [see Fig. 8(a)]. The nonlinear character of GP-LVM allows it to spread the grasp types better and, therefore, having a finer differentiation between them, as we will further explore in Section VI-B. This comparison is valid as well for GP-LVM 3-D and PCA 3-D. As we have seen in Section III-B, the kernel (or covariance) matrix is a core part of the GP-LVM methodology. In the following section, we inspect the computed GP-LVM representation in terms of its kernel matrix.

1) Interpreting the Kernel Matrix: The mapping between latent space \mathbf{X} and observed space \mathbf{Y} is commanded by the kernel matrix $\mathbf{K}_{i,j}$. This matrix has as many rows and columns as the number of data points (e.g., number of grasps times, number of subjects times, and number of samples per grasp) and will determine the closeness of points in latent space. It is computed according to (4). Different properties, such as grasp similarity and grasp action structure can be observed in this matrix; see Fig. 9. We observe a block of the kernel matrix around the diagonal, corresponding to the points belonging to three different grasp types (right side of Fig. 9). We can quantify how similar

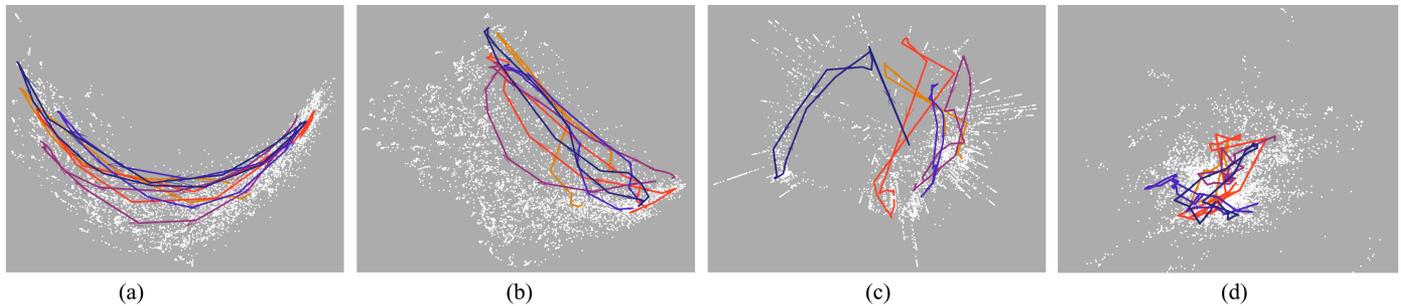


Fig. 8. Comparison of the synergistic representations of grasping data. In all figures, the trajectories of the five subjects performing grasp number 4 are plotted in different colors. The white dots represent the projection of all the grasping points. Each axis represents one dimension in the low-dimensional space. (a) PCA 2-D space. (b) GP-LVM 2-D space. (c) Isomap 2-D space. (d) LLE 2-D space.

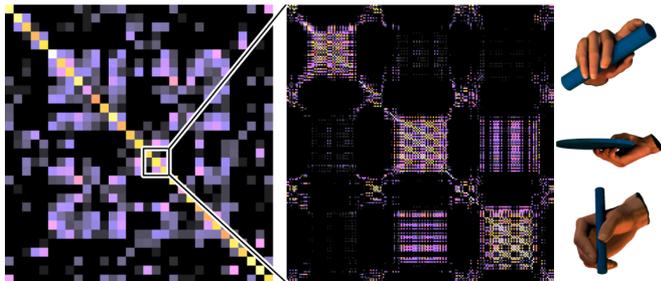


Fig. 9. (Left) Mean of the elements of the kernel matrix corresponding to the grasping frames, averaged for each grasp block. Dark tones represent low values [i.e., low correlations according to (4)], whereas bright ones denote highly correlated points. Bright colors in off-diagonal denote high grasp similarity. (Middle) Correlation of data points within three grasps, showing the action structure (approach, grasp, retreat, where approach and retreat are highly similar for all grasps). (Right) The three compared grasps. Second and third are very similar apart from global rotation, which is not modeled, while the first one clearly has a different index pose.

or dissimilar two grasps are based on the correlation. Another observation is that the initial and final points of a specific grasp (at the extremes of each grasp block) are not very informative in terms of grasp distinction, since there is always a large correlation between them. This is natural since those points correspond to the approach and retreat phase of the grasp, which is very similar across grasps.

In summary, GP-LVM-based encoding enables a finer differentiation between grasp types while keeping different user instances of the same grasp close to each other. Both GP-LVM and PCA are able to generalize across subjects. The presented analysis is applicable to any time-series data with multiple classes.

VI. EVALUATION

In this section, we evaluate the grasp synergies proposed by the GP-LVM and compare them with the ones extracted using PCA. We first evaluate how well poses can be represented in the low-dimensional space. Then, we evaluate the semantics of the synergies, i.e., how compact are the models of each grasp, and how well can we discriminate between them.

A. Evaluation of the Reconstruction Error

One important requirement on the extracted synergies is that they accurately represent the observed space of the data. There-

fore, we evaluate the quality of the learned mapping in terms of the reconstruction error. The reconstruction error shows how much the mapping connecting the observed and the latent representation distorts the data. It is computed by pushing a point from the high-dimensional space through the latent-space and back to the original space; the difference between the original and back-projected point is the reconstruction error. This is performed for both the training points, which tests how well the model adapts to those points, and the test set, where it allows us to assess the performance for points which are new but similar to the training data. Since no information about the classes of the grasp types is included, it does not test the generalization ability of the model. The reconstruction error only tests how much information is lost in the mapping from high-dimensional space to low-dimensional space and back; it does not provide information about the semantics of the space, e.g., how similar the executions of a particular grasp by different subjects are. For all four datasets, the positional [see Fig. 11(a)] and rotational [see Fig. 11(b)] errors were calculated.

Any model created with GP-LVM outperforms all models created by PCA in terms of reconstruction error. It is worth consideration the difference in performance between the training and test data. In both GP-LVM models, the training data have lower errors (both positional and rotational) compared with the test data. Interestingly, such a trend is not visible for PCA where the error on the training and test data are very similar. This is due to the fact that the synergies from PCA tend to be oversmoothed, average trajectories (see Fig. 12). Such average trajectories are “equally wrong” for training and testing. GP-LVM adapts better to the trajectories at the cost of a slight overfit. Nevertheless, the reconstruction error of GP-LVM is around 20% better than the error from PCA of the same dimensionality.

Increasing the dimensionality of the latent space allows better fitting of the training data onto the manifold, decreasing the reconstruction error for the training set. Similar effects are observed for the test set, suggesting that the higher dimensional models are not overfitting the data. Importantly, although the decrease is significantly larger for the nonlinear GP-LVM indicating that the correlations in the observed space are nonlinear and cannot be modeled using PCA.

Both algorithms seem to reconstruct positions and rotations with equal importance and in a similar fashion. If one compares the figures of positional [see Fig. 11(a)] and rotational errors

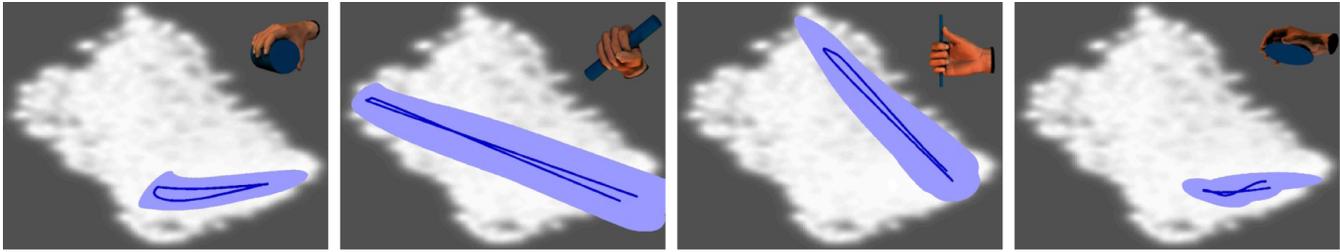


Fig. 10. GMR regression on four representative grasps. The dark line indicates the mean trajectory and the light area correspond to the uncertainty. It can be observed that thumb adduction/abduction varies strongly along the vertical axis (the grasp with the most abducted thumb, represented in the third figure, has the uppermost ending point). Similarly, hand aperture of the hand varies along the horizontal axis.

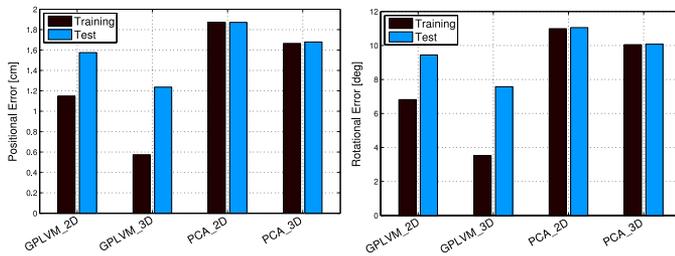


Fig. 11. Reconstruction errors regarding the position and orientation of the training and test datasets.

[see Fig. 11(b)], the relative differences between models and test/training set are very similar.

The overall performance of the extracted synergies by GP-LVM are better compared with the PCA synergies of the same dimensionality, as it has better results for the training set, as well as for the test set. Human hand motion in general is nonlinear, and therefore, an algorithm that can cope with nonlinearities (such as GP-LVM) in the data will be superior. Additionally, PCA focuses on reconstructing the data variance, which might be dominated by noise. GP-LVM, being a probabilistic approach, also has the ability to explicitly model the noise which increases the performance of the algorithm.

B. Semantic Evaluation

Low reconstruction error is not the only desirable characteristic. The representation should also capture and reflect semantic details of the data, like similarity between different users executing the same grasp or dissimilarity across different grasps.

1) *Visualization of the Mean Grasp Model:* To assess how well the GMM/GMR models fit to the original data, we project the latent trajectory of the GMM/GMR model back to the high-dimensional space. The comparison with the original data gives insight into how good the created GMM/GMR grasp models are. We reduced the amount of data displayed for the sake of clarity. The movements are projected to the plane spanned by the palmar-distal directions. Fig. 12 shows the fingertip movements of the index finger and the thumb projected onto that plane. The top image shows the corresponding grasp type, as well as the plane onto which the movements are projected. In the background of the other images, the original movements of the five subjects are shown in a lighter color.

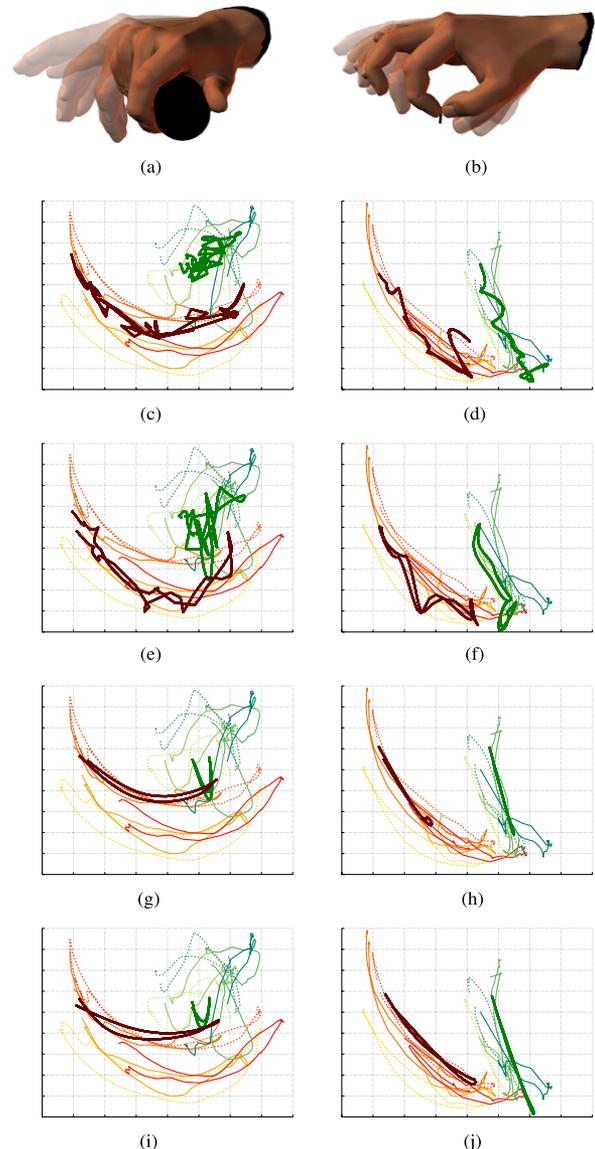


Fig. 12. Reconstruction of the thumb and index trajectory for two grasp types according to PCA and GP-LVM, projected into 2-D for visualization purposes. The bold trajectories show the mean trajectory in high-dimensional space according to different models, while the lighter trajectories correspond to the original demonstrations executed by the five subjects. The right-top trajectories correspond to the thumb finger, and the left-bottom trajectories correspond to the index. The figures show that PCA oversmooths the finger trajectories. (a) Grasp 4, adducted thumb. (b) Grasp 22, tip pinch. (c) GP-LVM 2-D. (d) GP-LVM 2-D. (e) GP-LVM 3-D. (f) GP-LVM 3-D. (g) PCA 2-D. (h) PCA 2-D. (i) PCA 3-D. (j) PCA 3-D.

The grasp on the left side is a special variation of the power grasp, where the thumb is aligned with the axis of the cylinder. In this grasp type, the thumb is relatively static, as it only abducts for the grasp. Since abduction/adduction is a movement largely perpendicular to the plane, most of the movement is lost by the projection. This makes the appearance of the thumb relatively random, but the 3-D trajectory shows a distinct pattern.

The grasp on the right side is a precision grasp, where the index and the thumb are used to pick up a small object. Therefore, it is important that those two digits are close in the actual grasping phase. The background trajectories in Fig. 12 clearly show this.

GP-LVM produces relatively rugged trajectories, but they follow the subjects trajectories quite well. They have roughly the same range of motion as the subjects. In the precision grasp (right column), they reach a position where thumb and index finger are very close, which is a functional requirement of the grasp type. The 3-D GP-LVM is smoother and its trajectories fit the original ones even better.

PCA produces very smooth curves, but it cannot create the curved path of the subjects trajectories. There is an offset and the trajectory cannot follow the full motion amplitude of the subjects. When the dimension is increased to three, the shape of the trajectory improves—the curvature gets a little bit larger, and the length of the trajectory better fits the subject's one.

Overall GP-LVM outperforms PCA, since it is able to follow the path of the human fingertips much better for a given dimension. That comes at the cost of having more rugged trajectories which is likely to be problematic for planning. In most applications, this is more desirable than having smooth trajectories, which follow the wrong path.

The rotational component of the fingertip cannot be easily visualized, so a comparable analysis on the rotations was not performed. Nevertheless, it can be assumed that they will behave in a very similar fashion, as the reconstruction error is very similar in positions and orientation (see Section VI-A).

2) *Grasp Similarity*: In this section, we classify grasping actions according to the GMM/GMR models; see Section V-A. In Fig. 10, the evaluation of the trajectories from the dynamical model applied to the GP-LVM 2-D representation is shown. Following the process depicted by Fig. 7, each grasp model is based on five trajectories, as performed by the subjects. The dark line corresponds to the mean trajectory and the light area shows the variance the model has on certain points of the trajectory. It can be observed that the vertical axis is closely related to thumb abduction/adduction, while distance to the lower left corner relates to how much is the hand open. One limitation of the data is that contact between object and hand is not explicitly observed, which makes some grasps look more similar than they really are (for example power and precision grasps with similar hand configuration).

Computing the similarity between grasps (pose sequences) is not straightforward. We will use the probabilistic description of grasps in terms of GMMs for this purpose. Based on these probabilistic models, we can compute how likely it is that each

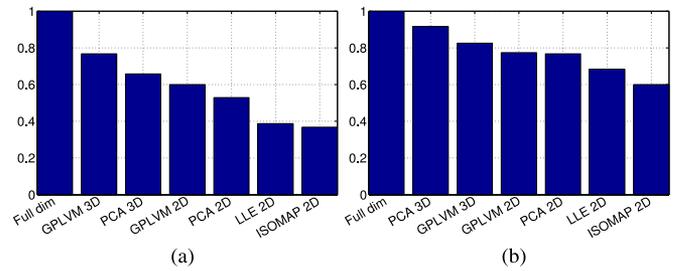


Fig. 13. Classification rate for GMM/GMR models tested with the training data itself. Although the grasp training set is more separable in PCA 3-D and full dimensionality than in GP-LVM 3-D for six Gaussians, the latter generalizes better over new data and, therefore, outperforms the rest when classifying previously unseen data [see Fig. 14(b)]. Isomap and LLE clearly perform worse than the rest of the methods. (a) Three Gaussians. (b) Six Gaussians.

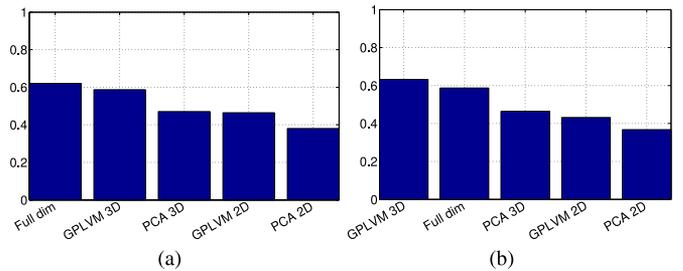


Fig. 14. Classification rate for GMM/GMR models on the test set. GP-LVM outperforms PCA for a given dimensionality and performs similarly to the full-dimensional model, which uses more than ten times more dimensions. (a) Three Gaussians. (b) Six Gaussians.

point x in the space is generated by a grasp g_i

$$p(x|g_i) = \sum_{k=1}^K \pi_k^{g_i} \mathcal{N}(x|\mu_k^{g_i} \sigma_k^{g_i}) \quad (6)$$

$$p(g_j|g_i) = \prod_{\forall x \in g_i} p(x|g_i) \quad (7)$$

$$s(g_j, g_i) = (p(g_j|g_i) + p(g_i|g_j))/2. \quad (8)$$

Equation (6) states that the probability of a point x belonging to a grasp g_i is modeled as a weighted mixture of Gaussians, as explained in Section V-A, once we make the simplifying assumption of independence between the poses, (7). This probability is not symmetric: $p(g_j|g_i) \neq p(g_i|g_j)$. Thus, we define the similarity between two grasps $s(g_j, g_i)$ as the average of those two quantities, (8). Other methods like hidden Markov models matching of sequences could be applied here instead [35].

Following this, we can compute the probability of a new grasp sequence having been generated by a particular GMR model. By comparing those probabilities, we can estimate which is the most likely grasp class that generated that sequence, and compare it with the real grasp that was actually executed. We performed this classification task for the grasp actions in the test set (not used for training). In Fig. 14, we can see that GP-LVM 3-D manages to generalize the model over new sequences equally well or better than the full-dimensional representation, while having a dimensionality more than ten times smaller than the original representation. GP-LVM is consistently better than PCA for a

given manifold dimensionality. The amount of Gaussians used in the GMM/GMR model does not make large differences in performance.

It is not possible to perform this classification task with Isomap and LLE manifolds, since the standard version of these methods do not provide a way of projecting new data (i.e., not existing in the training set) onto the lower dimensional manifold. Therefore, for LLE and Isomap, we can only show how well they separate the training data in lower dimensional space by classifying the training set based on the models extracted from the same set. Fig. 13 shows the classification performance for the methods in Fig. 14 plus Isomap 2-D [see Fig. 8(c)] and LLE 2-D [see Fig. 8(d)].

The first observation we can make is that the models computed in full-dimensional space perfectly separate the data. However, we should remember that those models generalize over new data similarly or worse than the GP-LVM 3-D models. Second, we can observe that PCA performs well in this classification task; this tells us that it clearly overfits the training data, since its performance is much worse in Fig. 14. The fact that the classification capabilities of mixtures of six Gaussians is only better for training data indicates that mixtures of more than three Gaussians produce overfitting. LLE and Isomap perform clearly worse than the 2-D versions of PCA and GP-LVM, as we expected from the shape of their manifolds in Fig. 8. We should remember, however, that this classification is much more sensitive to the particularities of the training set than the classification used in Fig. 14.

VII. CONCLUSION

The study that has been presented in this paper relates to two important areas of robotics: 1) human observation and motion analysis and 2) representations that enable successful action planning and control. In applications that consider hand activities, a common way of controlling grasping actions for robotic hands is to use high-dimensional human grasping data. Different representations based on dimensionality reduction techniques have been used to enable viable planning and control solutions. The correspondence problem between human and robot hand activities has also benefited from the usage of postural synergies. As such, the original work on postural synergies was based on linear dimensionality reduction methods which, as we have shown in this paper, do not represent the human hand activity in an appropriate manner due to the inherent nonlinearities in the data. We have argued that this significantly limits the usefulness of (linear) postural synergies as a modeling paradigm and that nonlinear dimensionality techniques should be exploited to represent the data in a more appropriate manner.

The work presented in this paper addressed the nonlinear dimensionality reduction methods and their application to encoding highly nonlinear human grasping data. Apart from encoding of postural synergies, our work relates closely to recent work in control of combined reaching and grasping movements in robots. However, this study is built on assumptions of a causal relationship between reaching and grasping, which is something that may not hold. An illustrative example in the beginning of

this paper and detailed discussion of related work serve as a motivating example of the applicability of the proposed approach.

As the first contribution, we have shown that nonlinear dimensionality reduction methods can be used to tackle the correlations problem without considering causal relations between dimensions but by assuming them to have been generated from an external manifold which we infer from data. Our second contribution is a thorough analysis of the internal parameters used in dimensionality reduction techniques, shedding light onto algorithms which have been traditionally used as a “black-box.” Finally, we have provided an extensive experimental evaluation that showed how the proposed methods outperform the standard techniques in the field both in terms of recognition and generation of motion patterns. To this end, we have presented both qualitative and quantitative results of applying two different approaches for learning low-dimensional representations of hand pose data.

REFERENCES

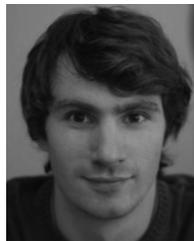
- [1] Y. Demiris and G. Hayes, *Imitation as a Dual-Route Process Featuring Predictive and Learning Components: A Biologically-Plausible Computational Model*. Cambridge, MA, USA: MIT Press, 2002, pp. 327–362.
- [2] M. Ciocarlie and P. Allen, “Data-driven optimization for underactuated robotic hands,” in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 1292–1299.
- [3] A. Shukla and A. Billard, “Coupled dynamical system based arm hand grasping model for learning fast adaptation strategies,” *Robot. Auton. Syst.*, vol. 60, no. 3, pp. 424–440, 2012.
- [4] C. Granville, D. Southerland, J. Platt, and A. H. Fagg, “Grasping affordances: Learning to connect vision to hand action,” in *The Path to Autonomous Robots: Essays in Honor of George A. Bekey*, G. S. Sukhatme, Ed. New York, NY, USA: Springer-Verlag, 2009, pp. 59–80.
- [5] J. Peters and S. Schaal, “Reinforcement learning of motor skills with policy gradients,” *Neural Netw.*, vol. 21, no. 4, pp. 682–697, 2008.
- [6] L. H. Ting, “Dimensional reduction in sensorimotor systems: A framework for understanding muscle coordination of posture,” *Progr. Brain Res.*, vol. 165, pp. 299–321, 2007.
- [7] M. Santello, M. Flanders, and J. Soechting, “Postural hand synergies for tool use,” *J. Neurosci.*, vol. 18, no. 23, pp. 10 105–10 115, 1998.
- [8] C. R. Mason, J. E. Gomez, and T. J. Ebner, “Hand synergies during reach-to-grasp,” *J. Neurophysiol.*, vol. 86, no. 6, pp. 2896–2910, Dec. 2001.
- [9] I. V. Grinyagin, E. V. Biryukova, and M. A. Maier, “Kinematic and dynamic synergies of human precision-grip movements,” *J. Neurophysiol.*, vol. 94, no. 4, pp. 2284–2294, 2005.
- [10] M. C. Tresch, V. C. K. Cheung, and A. d’Avella, “Matrix factorization algorithms for the identification of muscle synergies: Evaluation on simulated and experimental data sets,” *J. Neurophysiol.*, vol. 95, no. 4, pp. 2199–2212, 2006.
- [11] M. Gabbicini, A. Bicchi, D. Prattichizzo, and M. Malvezzi, “On the role of hand synergies in the optimal choice of grasping forces,” *Auton. Robots*, vol. 31, no. 2–3, pp. 235–252, 2011.
- [12] M. T. Ciocarlie and P. K. Allen, “Hand posture subspaces for dexterous robotic grasping,” *Int. J. Robot. Res.*, vol. 28, no. 7, pp. 851–867, 2009.
- [13] C. L. Nehaniv and K. Dautenhahn, *The Correspondence Problem*. Cambridge, MA, USA: MIT Press, 2002, pp. 41–61.
- [14] S. B. Kang and K. Ikeuchi, “Toward automatic robot instruction from perception: Mapping human grasps to manipulator grasps,” *IEEE Trans. Robot. Autom.*, vol. 13, no. 1, pp. 81–95, Feb. 1997.
- [15] F. J. Valero-Cuevas, F. E. Zajac, and C. G. Burgar, “Large index-fingertip forces are produced by subject-independent patterns of muscle excitation,” *J. Biomech.*, vol. 31, no. 8, pp. 693–703, 1998.
- [16] I. T. Jolliffe, *Principal Components Analysis*. New York, NY, USA: Springer-Verlag, 1986.
- [17] A. Tsoli and O. C. Jenkins, “Neighborhood denoising for learning high-dimensional grasping manifolds,” in *Proc. Intell. Robots Syst.*, 2008, pp. 3680–3685.
- [18] K. Nguyen and V. Perdereau, “Arm-hand movement: Imitation of human natural gestures with tenodesis effect,” in *Proc. Intell. Robots Syst.*, Sep. 2011, pp. 1459–1464.

- [19] J. Steffen, M. Pardowitz, and H. Ritter, "A manifold representation as common basis for action production and recognition," in *Proc. German Conf. Artif. Intell.*, 2009, pp. 607–614.
- [20] S. Bitzer and S. Vijayakumar, "Latent spaces for dynamic movement primitives," in *Proc. IEEE-RAS Int. Conf. Human. Robots*, 2009, pp. 574–581.
- [21] D. Praticchizzo, M. Malvezzi, and A. Bicchi, "On motion and force controllability of grasping hands with postural synergies," presented at the Robotics: Sci. Syst. Conf., Zaragoza, Spain, Jun. 2010.
- [22] M. Malhotra and Y. Matsuoka, "The relationship between actuator reduction and controllability for a robotic hand," in *Proc. IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatron.*, Sep. 2010, pp. 331–336.
- [23] R. Diankov, "Automated construction of robotic manipulation programs" Ph.D. dissertation, Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Aug. 2010.
- [24] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced Lectures on Machine Learning* (Lecture Notes in Computer Science, vol. 3176). New York, NY, USA: Springer-Verlag, 2003, pp. 63–71.
- [25] N. D. Lawrence, "Probabilistic non-linear principal component analysis with gaussian process latent variable models," *J. Mach. Learn. Res.*, vol. 6, pp. 1783–1816, 2005.
- [26] N. D. Lawrence and J. Quinero-Candela, "Local distance preservation in the gp-lvm through back constraints," in *Proc. Int. Conf. Mach. Learn.*, 2006, pp. 513–520.
- [27] L. K. Saul, K. Q. Weinberger, F. Sha, J. Ham, and D. D. Lee, "Spectral methods for dimensionality reduction," in *Semi-Supervised Learning*, O. Chapelle, B. Schölkopf, and A. Zien, Eds. Cambridge, MA, USA: MIT Press, 2006, ch. 1.
- [28] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
- [29] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 77–90, Oct. 2006.
- [30] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [31] J. Romero, T. Feix, H. Kjellström, and D. Kragic, "Spatio-temporal modeling of grasping actions," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2010, pp. 2103–2108.
- [32] T. Feix, R. Pawlik, H. Schmiedmayer, J. Romero, and D. Kragic, "A comprehensive grasp taxonomy," in *Proc. Robot., Sci. Syst.: Worksh. Understand. Human Hand Adv. Robot. Manipulat.*, Jun. 2009.
- [33] S. Calinon, *Robot Programming by Demonstration: A Probabilistic Approach*. Boca Raton, FL, USA: EPFL/CRC, 2009.
- [34] S. Calinon, F. Guenter, and A. Billard, "On learning, representing and generalizing a task in a humanoid robot," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 286–298, Apr. 2007.
- [35] S. Calinon and A. Billard, "Recognition and reproduction of gestures using a probabilistic framework combining PCA, ICA and HMM," in *International Conference on Machine Learning*. New York, NY, USA: ACM, 2005, pp. 105–112.



Javier Romero received the M.Sc. degree in telecommunication engineering from the Technical University of Madrid, Madrid, Spain, in 2007 and the Ph.D. degree in 2011 from the Royal Institute of Technology, Stockholm, Sweden.

He is currently a Research Scientist with the Perceiving Systems group, Max-Planck-Institute for Intelligent Systems, Tübingen, Germany. His research interest includes robotic perception of humans and their environment, focusing on generative models of human appearance and human body tracking.



Thomas Feix (M'12) received the M.Sc. degree in sports equipment technology from the University of Applied Sciences Technikum Wien, Vienna, Austria, and the Ph.D. degree in 2011 from the Vienna University of Technology.

He is currently a Postdoctoral Associate with the GRAB Lab Department of Mechanical Engineering, Yale University, New Haven, CT, USA. He is working on human hand kinematics for manipulation. His research interests include human grasping and manipulation and its application to robotics and prosthetics.



Carl Henrik Ek (M'10) received the M.Sc. degree in vehicle engineering from The Royal Institute of Technology KTH, Stockholm, Sweden, in 2005 and the Ph.D. degree in computer science from the Oxford Brookes University, Oxford, U.K., in 2009.

He is currently a Postdoctoral Researcher with the School of Computer Science and Communication, The Royal Institute of Technology KTH. His research interests include machine learning, robotics, and computer vision with a focus on probabilistic methods for learning efficient representations of data.



Hedvig Kjellström received the M.Sc. degree in engineering physics and the Ph.D. degree in computer science from The Royal Institute of Technology KTH, Stockholm, Sweden, in 1997 and 2001, respectively.

He is currently an Associate Professor with The Royal Institute of Technology KTH. She has written more than 50 papers in the fields of computer vision, information fusion, robotics, speech, and human-computer interaction.

Dr. Kjellström received the Koenderink Prize for her European Conference on Computer Vision (ECCV) 2000 article on human motion reconstruction, with M. Black and D. Fleet in 2010.



Danica Kragic (SM'12) received the M.Sc. degree in mechanical engineering from the TU Rijeka, Rijeka, Croatia, in 1995 and the Ph.D. degree in computer science from The Royal Institute of Technology KTH, Stockholm, Sweden, in 2001.

He is currently a Professor with the School of Computer Science and Communication, The Royal Institute of Technology KTH. Her research interests include computer vision and robotics.

Dr. Kragic received the 2007 IEEE Robotics and Automation Society Early Academic Career Award.

She is a Member of the Swedish Royal Academy of Sciences and the Swedish Young Academy.