The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Spatio-Temporal Modeling of Grasping Actions

Javier Romero Thomas Feix Hedvig Kjellström Danica Kragic

Abstract— **Understanding the spatial dimensionality and temporal context of human hand actions can provide representations for programming grasping actions in robots and inspire design of new robotic and prosthetic hands. The natural representation of human hand motion has high dimensionality. For specific activities such as handling and grasping of objects, the commonly observed hand motions lie on a lower-dimensional non-linear manifold in hand posture space. Although full body human motion is well studied within Computer Vision and Biomechanics, there is very little work on the analysis of hand motion with nonlinear dimensionality reduction techniques. In this paper we use Gaussian Process Latent Variable Models (GPLVMs) to model the lower dimensional manifold of human hand motions during object grasping. We show how the technique can be used to embed high-dimensional grasping actions in a lower-dimensional space suitable for modeling, recognition and mapping.**

I. Introduction

Modeling of human hand motion is receiving an increasing interest in areas such as computer vision, graphics, robotics and psychology. The goal of the work presented here is to study and model the spatial dimensionality and temporal context of human hand actions to i) provide representations for programming grasping actions in robots, and ii) use these for designing new robotic and prosthetic hands. In robotics, it has been argued that continuous motion mapping from human to robot requires suitable spatio-temporal representations of human and robot motions, [1], [2]. However, most of the work on grasp mapping is based on hand-designed grasping taxonomies considering a *discrete* set of hand postures, [3], [4].

The main contribution of the work presented here is the study of spatial and temporal context of human grasping actions. Recently, significant advances have been achieved in full body human motion modelling by exploiting the low-dimensional nature of the data [5]. Despite the wide use of subspace representations in human body motion analysis, the work on human hand motion is limited. The work in [6] indicates that, similarly to full body modelling, hand data would also benefit from a similar methodology.

However, compared to the state-of-the-art on the human body modelling, the methods applied in [6] are significantly less flexible since they assume the manifold is linear. In this work we apply a recently proposed generative model (GPLVM) to shape the manifolds of high-dimensional human grasping motions. The adopted technique has been used in several recent studies of human body motion but has not been applied to human hand motion. One of the motivations for this model is its capability to regenerate grasping actions when, for example, task constraints need to be taken into account, [7].

Despite the wide use of subspace representations in human body motion analysis [5], the work on human hand motion is limited. An analysis of low-dimensional embeddings of human hand grasping was performed in [6]. However, the data was recorded from subjects imagining grasp actions instead of applying them. Furthermore, the low-dimensional space was created with PCA, which is limited by its linear nature. The benefits of non-linear dimensionality reduction schemes are shown in [8], where a 2D space is used for the control of robotic grasps. In a similar way, we compare the 2D latent space generated by our model and other dimensionality reduction techniques such as Principal Component Analysis (PCA), Isomap or Locally Linear Embedding (LLE), showing the advantages of our approach.

II. Contributions and Related Work

Research on human grasps concentrates on the design of grasping taxonomies based on the observation of human grasping actions. Grasp classes are often based on heuristics motivated by intuition and application. In general, there is very little consensus between the different taxonomies. Our recent study, [4] analyzes several taxonomies proposed in the areas of robotics, biomechanics and medicine. An important observation is that the taxonomies have not been contrasted against the actual data extracted from subjects performing the grasps. Work in that direction was presented in [6]. Subjects were asked to shape the hand as if they were grasping different objects. A CyberGlove was used to record 15 joint angles of the grasping hand. This data was projected onto a low-dimensional space with PCA. The main conclusion was that the first two components of the projected data accounts for 80% of the variance of the data.

In our work, we further develop these ideas in several directions. First, we consider the whole grasping sequence instead of just a single grasp posture. This facilitates the spatial and temporal reconstruction of a grasping action. Second, the latent space is reconstructed from end-effector data (fingertip position and orientation relative to the palm) instead of joint angles. Thus, we avoid the problem of proximal joints having a higher impact on the position of

This work was supported by EU IST-FP7-IP GRASP and Swedish Foundation for Strategic Research.

J. Romero,H. Kjellström and D.Kragic are with Computational Vision and Active Perception Lab, Centre for Autonomous Systems, School of Computer Science and Communication, KTH, SE-100 44 Stockholm, Sweden jrgn,hedvig,dani@kth.se

T. Feix is with Otto Bock Healthcare Gmeinbh, 1070 Vienna, Austria Thomas.Feix@ottobock.com

The authors would like to thank Prof. Heiner Deubel and his team for help with data collection and Carl Henrik Ek for insights into using GPLVM

the fingertip. End-effector data is also easier to translate to other embodiments than the joint angle data. Third, instead of studying how different objects are grasped, we study how different grasps are performed. The motivation for this is that some objects can be grasped in different ways depending on the goal of the grasp (pick a pen or write with a pen). Fourth, due to the non-linearity of the human hand motion, we use non-linear methods to construct the low-dimensional representation space.

The work by Ciocarlie et al. [9], [10] focuses on reducing the complexity of robotic grasping through the use of PCA. The low-dimensional space extracted in [6] is used both for reducing the complexity of grasp space exploration [9] and for mapping between an operator and a simulated hand [10]. Since the space contains only hand postures where the final grasp has been achieved, the approach phase of the grasp is not taken into account. In [11] whole grasp sequences (including the approach phase) are used for optimizing the final grasp pose in a simulation environment. However, the optimization does not distinguish between approach and grasping phase; therefore the optimized grasp pose might be based only on approaching poses. In [8] data from a Vicon optical motion capture system is used to create a latent space for "Interactive Control of a Robot Hand" using Isomap. The data is a concatenation of different grasps and tapping demonstrations. Contrary to our approach, the authors do not provide any study of the similarity between the demonstrated grasps. In Section IV-C, we will also discuss the performance of Isomap for our purposes.

In this paper, we use GPLVMs for creating a low-dimensional grasp space in which we can reason about the similarities and differences between a set of predefined grasps [4]. GPLVM places a Gaussian Process (GP) prior over the generative mapping from latent space to data space. Through marginalization of this mapping, the marginal likelihood of observed data given the latent locations can be found. The latent locations are then found by maximizing this likelihood. Due to the flexibility of GPs, the generative mapping is not constrained to be linear as is the case of PCA. Moreover, it has been shown that it is more efficient than other techniques like Isomap when dealing with noisy and incomplete training data, [12].

In the context of full-body human motion, GPLVMs have been employed both for visual tracking of full-body motion [5] and for classification of full-body actions [13]. Modeling the dynamics in embedded spaces of lower dimensionality decreases the amount of training data needed [5] and facilitates the generation of natural and physically plausible motion [14].

Similarly to full-body motion analysis, an immediate application of this latent space is a non-parametric dynamic model of grasping actions for tracking and classification; however, this is out of the scope of the work presented here. For our purpose, we do not model dynamics explicitly as in [5], but include back-constraints (Section III) that indirectly enforce temporal continuity in the latent space. This avoids the unimodal nature of the GPDM dynamics. The created GPLVM model allows the generation of concatenated grasping actions with natural transitions. This can be done by applying constraints in the latent space in a similar way as constraints are applied in [15].

III. Theoretical Formulation

Let D denote the dimension of the data space and q the dimension of the latent space. Given N observations, the matrix containing the data points is denoted $Y \in \mathbb{R}^{N\times D}$ and the matrix of the corresponding points in the latent space is $X \in \mathbb{R}^{N\times q}$. The marginal likelihood P of the datapoints y, given the latent positions x and the hyper-parameters θ, is a product of D independent GPs [16]:

$$P(Y|X,\theta) = \prod_{j=1}^{D} \frac{1}{(2\pi)^{\frac{N}{2}}|K|^{\frac{1}{2}}} e^{-\frac{1}{2}y_j^T K^{-1} y_j} \tag{1}$$

where $y_j \in \mathbb{R}^{N\times 1}$ is the jth column of the data matrix and $K \in \mathbb{R}^{N\times N}$ is the kernel or covariance matrix specified by θ. In the GP-LVM framework the latent location $\mathbf{X}$ and the hyper-parameters θ are found using Maximum Likelihood. In general, this optimization has many solutions since the function is not convex [16]. To remove additional degrees of freedom non-informative priors are placed over the latent locations and hyper-parameters.

A. Covariance Functions

The covariance matrix K in Eq. (1) is determined by the covariance or kernel function k:

$$K_{i,j} = k(x_i, x_j) \tag{2}$$

The choice of the covariance function is critical, since characterizes the functions most prominent in the prior. This is also an advantage of the method, since it allows adaptation to the specific needs of the task and the dataset at hand. Most commonly, K is determined by a sum of several different kernels, like the Radial Basis Function (RBF), bias and noise kernels. The RBF kernel is defined as follows:

$$k(x_i, x_j) = \alpha\ e^{-\frac{\gamma}{2}(x_i-x_j)^T(x_i-x_j)} \tag{3}$$

where α defines the output variance and the inverse kernel width γ controls the smoothness of the function. By using a smooth covariance function like the RBF kernel, we encode a preference towards smooth generative mappings in the GP prior. This implies that points close in the latent space will remain close in the observed space (when projected using the mean prediction of the GP). However, it is not guaranteed that the inverse is true, i.e. points close in the observed space remain close in the latent space (see Subsection III-B). In addition to the RBF kernel we also include a bias term which accounts for translations in the data and a white noise term.

B. Back Constraints

As stated above, a GPLVM in its basic form does not guarantee that a smooth inverse exists to the generative mapping [17]. However, this can be incorporated into the model by representing the latent locations x_i in terms of a smooth parametric mapping g_j from the observed data y_i.

$$x_{ij} = g_j(y_i, a) = \sum_{n=1}^{N} a_{jn} k_{bc}(y_i - y_n) \tag{4}$$

where k_{bc} is the back constraint kernel. This means that the maximum likelihood solution of these parameters a rather than the latent locations are sought. This is referred to as a back-constrained GPLVM. In addition to preserve the local smoothness of the observed data, previously unseen data can be projected onto the latent space in an efficient manner by pushing them through this back-mapping.

IV. Evaluation on Real Data

The proposed technique was evaluated on data generated by 5 subjects (3 male, 2 female). All subjects are right handed and have not reported any hand disabilities. A Polhemus Liberty system with six magnetic sensors was used for recording the data. The spatial and angular resolution of each sensor is 0.8 mm and 0.15 degrees respectively. One sensor was applied to each fingertip, positioned on the fingernail and one was placed on the dorsum of the hand. See Figure 1 for an image of the markers applied to the hand.

The subjects were asked to perform 31 grasp types from the 33 in [4]. We excluded "Distal Type" and "Tripod Variation" due to their very specific nature. A picture of each grasp was shown to the subjects and a demonstration of the grasp was performed for the most difficult ones. Then the subjects were instructed to grasp with that specific grasp type. The data was then further processed as follows:

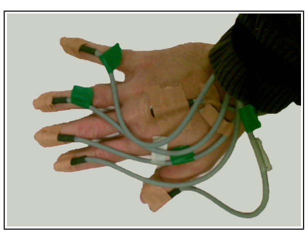

Fig. 1. Placement of the sensors. Five sensors are placed on the fingertips and one is positioned on the wrist.

1) Calibration that aligns the coordinate systems of the sensors with the actual anatomical direction.
2) Transformation of the fingertip data into the wrist coordinate system. To provide some invariance to different approach movements, the hand pose is defined in terms of the relative position and orientation of the fingertip sensors with respect to the wrist.
3) Translation of the position of the fingertip origin to the center of the distal finger segment and normalization of the dimensions to a standard range.

The transportation component of a grasp movement varies significantly depending on the orientation and distance of the object to the hand. Therefore, hand pose is here defined as the pose of the fingers relative to the palm. The sensors create a space of dimensionality 35 where each of the 5 sensors has

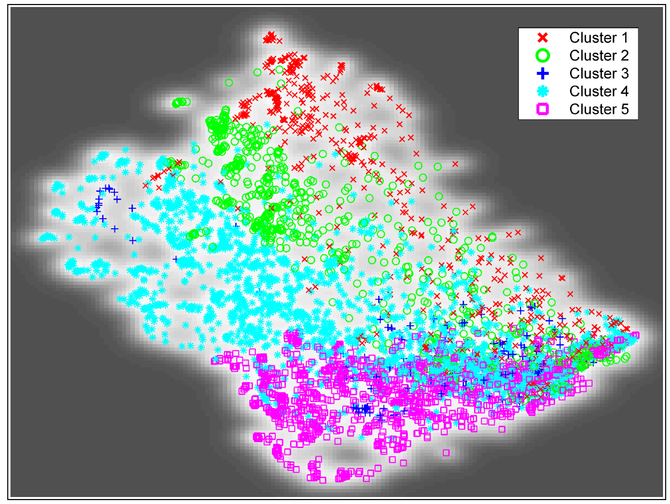

Fig. 2. Grasp space spanned by the execution 31 grasp types by five subjects. Each color represents one of the clusters represented in Figure 5.

7 dimensions: 3 for position and 4 for orientation (we used quaternions to represent rotations). From each trial we took 30 equally distributed samples. Overall this resulted in a data matrix of size 4650×35.

A. Low dimensional representation of the grasp movements

We created the GPLVM latent space spanned by this data with the Matlab FGPLVM toolbox [17]. We should note that the datapoints were not tagged with any information about grasp class, subject or timestamp in the creation of the GPLVM. Several different configurations of the GPLVM parameters (with and without back constraints, different back constraint types, variation of parameters) were analyzed. Scaled conjugate gradient optimization was used in order to obtain maxima of Eq.(1). The best results were achieved with a kernel composed of RBF, bias and noise, and kernel based regression back constraints with an RBF kernel. The inverse width parameter was set to 0.001 by inspection. Following [6], [8], we selected a dimensionality of 2 for the latent space, simplifying the visualization of the results. Although higher dimensional latent spaces could improve the separability of grasps, the advantages of GPLVM over other dimensionality reduction techniques can be shown already in 2D. The model was initialized with different dimensionality reduction methods (PPCA, Isomap, LLE) and the one with lowest reconstruction error was kept. In our case this was an initialization with PPCA and the result of the optimization can be seen in Figure 2. We observe that the space has a common starting point in the lower right corner (corresponding to the initial "flat hand" position), then all the grasps follow the same path for few timesteps (transition from flat hand to relaxed hand) and finally different grasp classes diverge.

B. Gaussian Mixture Regression (GMR) of Grasps

Since the data contains multiple subject demonstrations over time, the representation of each grasp in latent space

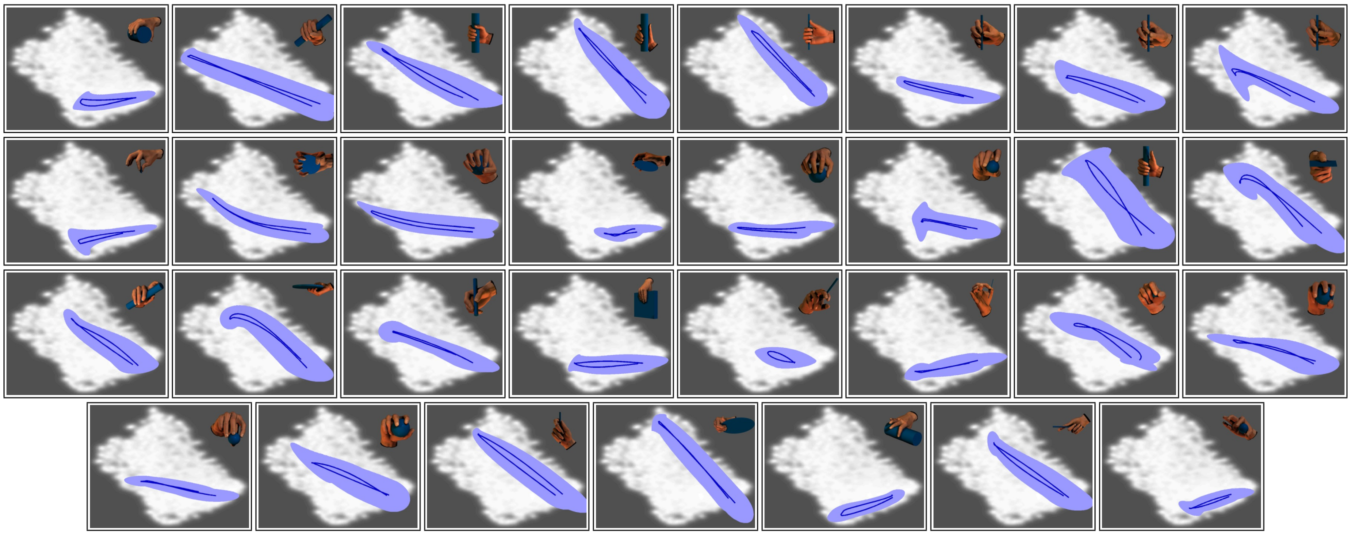

Fig. 3. GMR regression on the 31 grasp movements of all subjects. The dark line indicates the mean trajectory and the light area correspond to the uncertainty. The grasp are sorted, so the first row contains grasps 1 to 8 and so on.

should encompass temporal information as well as multiple subject variance. We have used GMR [15] for getting a unique dynamic model for each grasp type. We will briefly introduce this representation (check [15] for more information). First, for each grasp the datapoints in latent space (2D data, see first row of Figure 4) are extended with the time dimension. Then this data (3D) is fitted into a Gaussian Mixture Model (GMM)(second row of Figure 4) by an expectation-maximization procedure initialized with K-means. Empirically, we found that using more than 3 gaussians did not improve the quality of the fitting. Based on that mixture of gaussians 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 (third row of Figure 4). That path has a mean and a variance. The paths corresponding to each of the 31 grasps can be found in Figure 3. The GMM/GMR representation of the grasps is a powerful tool that can be used for several purposes. One is the generation of new actions under some constraints [15]. 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.

C. Comparison of Dimensionality Reduction Algorithms

For comparison, other dimensionality reduction algorithms were applied to the same dataset of real human grasps. The latent space dimensionality for all the algorithms was set to 2. Algorithms used were Principal Components Analysis (PCA), Isomap (from [18]) and Locally Linear Embedding (LLE) (from [19]). Figure 4 shows the low dimensional trajectories of all subjects performing grasp 1 and as background the corresponding latent space. This grasp is a typical example and the other grasp types show a similar pattern throughout all dimensionality reduction algorithms. The points of the PCA solution lie on an "arc" and the starting position is on the right side. This shape seems to be due to PCA being a linear method. It can only unravel the global motion in the data. Since this arc is rather narrow there is little distinction between different grasp trajectories and fine details of the manifold cannot be extracted.

Isomap shows some sort of star-like structure, but one branch does not represent one grasp type as would be expected. Also the ability to generalize between subjects is not present: the trajectory of each subject is different without showing common trends. Modifying the numbers of neighbors did not improve the result, so either the neighborhood size is too small or the locally linear assumption is violated.

LLE fails to discover any meaningful structure. All datapoints are centered in a certain location without any inner structure or common trajectories for grasp types.

For a latent space dimensionality of 2, GPLVM (Figure 2 and first column of 4) obtains a higher inter-grasp separability, lower intra-grasp variance (comparable to PCA) and preserves time continuity in the trajectories in the latent space in a better way than the other methods. PCA is limited since it is a linear method; Isomap and LEE fail since they are based on local distance measurements which are very sensitive to noise. Of course these problems also alter the GMM/GMR algorithm, so that the output is nearly a point (Isomap) or the trajectory has a very high variance (LLE). The ability to generalize between subjects is also visible in PCA, but the whole space is very packed and the trajectories of all grasp types are within a very small area. This comparison show us the advantages of GPLVM for representing high dimensional noisy data in very low dimensional spaces. The study of which dimensionality is optimal is left for future work.

D. Similarity Measure and Clustering of Grasps

We used GMM/GMR to measure similarity between human grasps. Since we have a probabilistic model for each

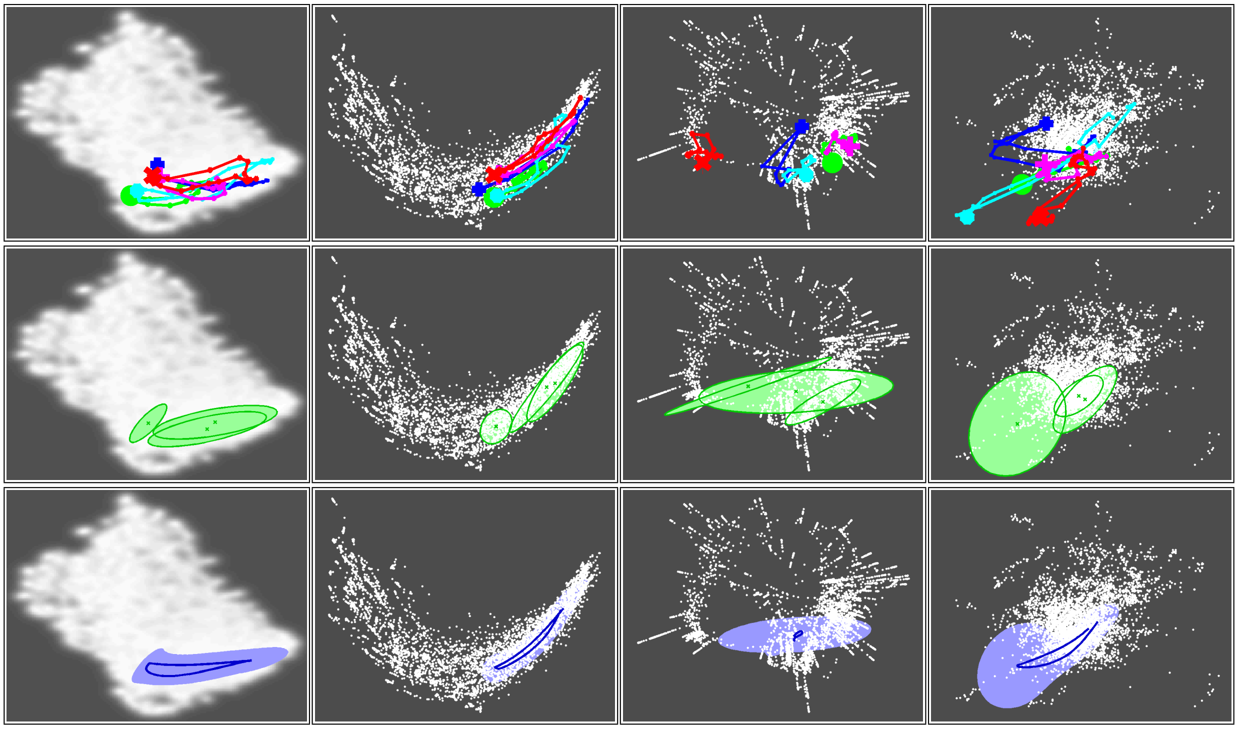

Fig. 4. From left to right: GPLVM, PCA, ISOMAP, LLE. From top to bottom : projection of grasp number 1 into latent space(different colors correspond to different users), GMM fitting, GMR regression. The other grasp types show similar patterns.

grasp in the latent space (through their GMM representation), we can compute how likely it is that each point x in the space is generated by a grasp g_i.

$$p(x|g_i) = \sum_{k=1}^{3} \pi_k^{g_i} \mathcal{N}(x|\mu_k^{g_i} \sigma_k^{g_i})$$

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

$$s(g_j, g_i) = (p(g_j|g_i) + p(g_i|g_j))/2$$

The product of the likelihoods of points in grasp g_j being generated by grasp g_i give us a measure of how well is g_j supported by the g_i model. Note that this measure is not symmetric. We can define the similarity between two grasps $s(g_j, g_i)$ as the average of those two quantities.

We performed average linkage clustering (UPGMA from the Matlab Statistical Toolbox) based on this similarity measure. The result of the algorithm can be seen in Figure 5. The number of clusters was chosen to be 5 since further subdividing the clusters overfits the data, i.e. *cluster four* was split into two groups with similar characteristics. Reducing the number of clusters resulted in large, too general clusters.

The grasps in *cluster one* resemble each other quite well. They all are power grasps with all four fingers in contact with the object. In addition the thumb is in a very adducted and extended position. The fingers are all in a very similar position, the MCP joint is rather extended, but the PIP and DIP joints are strongly flexed.

Cluster two is constructed by grasps that have a “straight” (extended MCP and IP joint) and mostly adducted thumb. Side opposition (see [20] for a description of the concept) is dominant in grasps 16, 27, 30 or at least there are some aspects that side opposition is involved as in grasps 17 and 18. None of those grasps is a precision grasp.

Only one grasp belongs to *Cluster three*. This grasp type does not impose many constraints in the hand pose of the subject. Therefore the variability of the grasp was high; most subjects formed this grasp with all fingers extended, but one subject flexed the ring and the middle finger. Also the index and the middle finger, which are in contact with the object, can be bent to a certain degree without affecting the stability of the grasp. Overall it seems that this grasp is formed in a rather extended position; this explains why the center of that grasp is close to the starting position unlike the rest of the grasps which involve much more flexion of the digits.

The biggest group of grasps is in *cluster four*. This group is quite diverse and it offers less distinct properties than the other groups. Yet all four fingers are all in a mid-flexed positions and the flexion increases towards the little finger. This is a clear difference to cluster five, where the little finger is in an extended position. In addition the thumb is mostly abducted, except grasp 23 where it is adducted.

Cluster five has a distinct inner structure. The horizontal direction in latent space modulates the overall extension/flexion of the fingers, whereas the vertical direction changes the individual index finger flexion.

In addition to those clusters properties, there are some general trends of the latent space. First, the further away a grasp is from the starting position (right side of the latent space) the more flexed the fingers will be. This is due to the fact that the starting position is with fingers and thumb totally extended and the transition between grasp types is smooth. The clusters seem to be elongated in the start-end posture direction. This makes sense, since the whole movement was taken into account when clustering the grasp types.

In the grasp taxonomy of [4] the thumb plays a crucial role in classifying the grasp types. The clusters which were created here tend to go in accordance with this thumb classification, but there are some conflicts. The reason for this could be that the clustering algorithm gives each finger equal importance, while in [4] the thumb plays a prominent role. Some grasp types do not employ all fingers, which means that potentially some fingers are not relevant for the grasp definition. Currently those fingers are taken into account with the same importance as fingers in contact with the object.

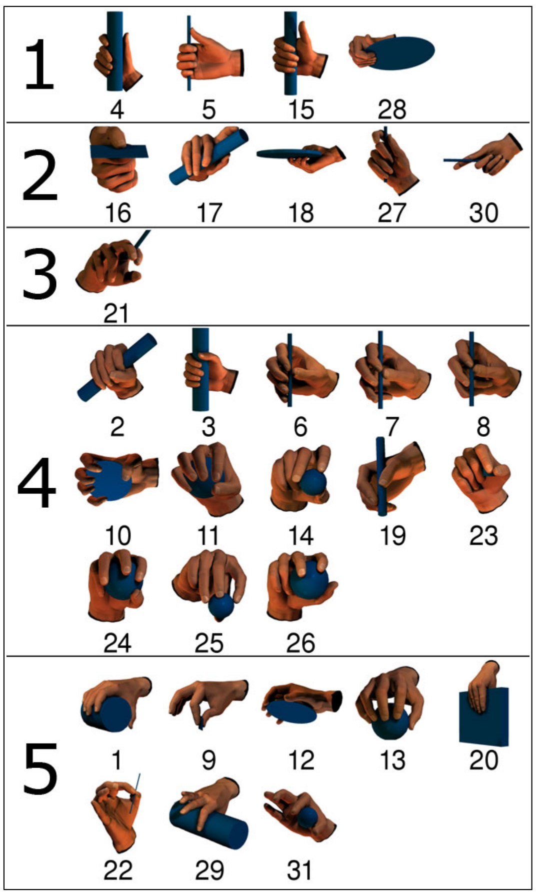

Fig. 5. Clusters of the human grasps.

V. Conclusions

The goal of the work presented here, differently from all the existing grasp taxonomies, was to model the spatial dimensionality and temporal context of hand actions. Instead of studying how different objects are grasped, we study how different grasps are performed. Apart from the important insights of human hand motion, the developed technique has also been used to evaluate the state-of-the-art taxonomies. We have shown how the technique can be used to embed high dimensional grasping actions in a lower-dimensional space suitable for modeling, recognition and mapping. Considering the whole grasping sequence instead of just a single grasp posture facilitates the spatial and temporal reconstruction of a grasping action. The method is evaluated on real data.

An immediate application of the extracted latent space is a non-parametric dynamic model of grasping actions for tracking and classification [21]. We do not model dynamics explicitly but include back-constraints that indirectly enforce temporal continuity in the latent space. This avoids the unimodal nature of using an auto-regressive dynamic model. The created GPLVM model potentially allows the generation of concatenated grasping actions with natural transitions. Thus, one idea is to apply constraints in the latent space in a similar way as in [15]. Together with the evaluation in terms of grasp classification this remains our future work.

References

[1] S. Ekvall and D. Kragic, "Grasp recognition for programming by demonstration tasks," in *ICRA*, 2005, pp. 748–753.

[2] S. Bitzer and S. Vijayakumar, "Latent spaces for dynamic movement primitives," in *Humanoids 2009*, 2009.

[3] S. Ekvall and D. Kragic, "Interactive grasp learning based on human demonstration," in *ICRA*, 2004, pp. 3519–3524.

[4] T. Feix, R. Pawlik, H. Schmiedmayer, J. Romero, and D. Kragic, "A comprehensive grasp taxonomy," in *Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation*, June 2009.

[5] R. Urtasun, D. J. Fleet, and P. Fua, "3D people tracking with gaussian process dynamical models," in *CVPR*, 2006, pp. I: 238–245.

[6] J. S. M. Santello, M. Flanders, "Postural hand synergies for tool use," in *The Journal of Neuroscience*, 1998.

[7] D. Song, K. Huebner, V. Kyrki, and D. Kragic, "Learning task constraints for robot grasping using graphical models," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010.

[8] A. Tsoli and O. C. Jenkins, "Neighborhood denoising for learning high-dimensional grasping manifolds," in *IROS*, 2008, pp. 3680–3685.

[9] M. T. Ciocarlie, C. Goldfeder, and P. K.Allen, "Dimensionality reduction for hand-independent dexterous robotic grasping," in *IROS*. IEEE, 2007, pp. 3270–3275.

[10] M. T. Ciocarlie, S. T. Clanton, M. C. Spalding, and P. K. Allen, "Biomimetic grasp planning for cortical control of a robotic hand," in *IROS*. IEEE, 2008, pp. 2271–2276.

[11] H. B. Amor, G. Heumer, B. Jung, and A. Vitzthum, "Grasp synthesis from low-dimensional probabilistic grasp models," *Journal of Visualization and Computer Animation*, vol. 19, no. 3-4, pp. 445–454, 2008.

[12] R. Urtasun, D. J. Fleet, A. Geiger, J. Popović, T. J. Darrell, and N. D. Lawrence, "Topologically-constrained latent variable models," in *ICML08*, vol. 307. ACM, 2008, pp. 1080–1087.

[13] R. Urtasun and T. Darrell, "Discriminative gaussian process latent variable model for classification," in *ICML*, 2007, pp. 927–934.

[14] S. Bitzer, S. Klanke, and S. Vijayakumar, "Does dimensionality reduction improve the quality of motion interpolation?" in *ESANN09*.

[15] S. Calinon, F. Guenter, and A. Billard, "On learning, representing and generalizing a task in a humanoid robot," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 37, pp. 286–298, 2007.

[16] N. D. Lawrence, "The gaussian process latent variable model," The University of Sheffield, Department of Computer Science., Tech. Rep.

[17] N. D. Lawrence and J. Quinonero-Candela, "Local distance preservation in the gp-lvm through back constraints," in *ICML06*, pp. 513–520.

[18] J. B. Tenenbaum, V. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, December 2000.

[19] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, no. 5500, pp. 2323–2326, 2000.

[20] T. Iberall, G. Bingham, and M. A. Arbib, "Opposition space as a structuring concept for the analysis of skilled hand movements," *Experimental Brain Research Series*, vol. 15, pp. 158–173, 1986.

[21] J. Romero, H. Kjellström, and D. Kragic, "Hands in action: Real-time 3d reconstruction of hands in interaction with objects," in *ICRA*, 2010.