Kinematic Synergies in Human Bimanual Manipulation

Julia Starke, Lukas Ruf, Andre Meixner and Tamim Asfour

***Abstract*— Bimanual manipulation tasks present significant challenges for robots due to the high dimensionality and complex coordination required between two arms. While humans perform such tasks effortlessly, transferring this capability to robots remains challenging. In this paper, we introduce a novel synergy-based representation for human bimanual manipulation that captures the characteristics of coordinated movements in a low-dimensional space, the synergy space. We train a variational autoencoder network on human motion data to learn synergies in bimanual manipulation tasks. We compare a Euclidean representation with a Riemannian representation in the latent space. Experimental results demonstrate that our synergy-based representation reduces dimensionality while improving both representation accuracy and motion smoothness. Specifically, our method achieves a $38\,\%$ reduction in representation error compared to linear dimensionality reduction techniques, while simultaneously improving the smoothness of generated movements by $42\,\%$. We validate our approach through qualitative and quantitative analysis of reconstructed and newly generated bimanual movements, showing that the resulting motions preserve the essential characteristics of human manipulation.**

I. Introduction

Bimanual manipulation of objects is a fundamental capability for both humans and humanoid robots. While humans perform bimanual tasks effortlessly in everyday activities, humanoid robots still face significant challenges when attempting to replicate this functionality. Human bimanual manipulation demonstrates remarkable efficiency despite these inherent challenges. By studying and modeling how humans coordinate their arms during manipulation tasks, we can develop more efficient control strategies for humanoid robots. The concept of kinematic synergies – correlated joint movements that reduce the dimensionality of complex motor tasks – has been widely applied in robotic grasping, where it has significantly simplified hand design and control. These synergies transform high-dimensional joint spaces into lower-dimensional representations that capture essential grasp characteristics [1], [2]. By these means, synergies can simplify the adaptation and transfer of human motion demonstrations for use in robotic control algorithms. However, extending this concept to bimanual manipulation poses unique challenges due to the larger workspace, increased degrees of freedom, and the coordination patterns required between both arms.

The research leading to these results has received funding from the European Union's Horizon 2021 Research and Innovation programme under grant agreement No 101070292 (HARIA), the Carl Zeiss Foundation under the project JuBot and the German Federal Ministry of Research, Technology and Space (BMFTR) under the German Robotics Institute (RIG).

The authors are with the Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany. Julia Starke is with the Institute of Robotics and Cognitive Systems, University of Lübeck, Lübeck, Germany. E-mails: julia.starke@uni-luebeck.de, asfour@kit.edu

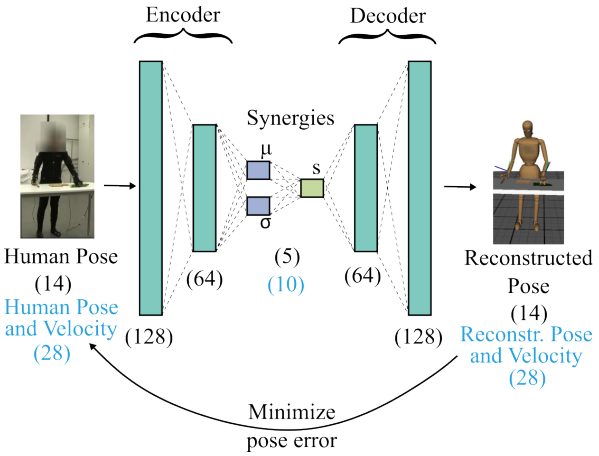


Fig. 1: Autoencoder network used to learn a bimanual synergy space; the network is trained on human demonstrations, which are mapped to a simulated human model

In this work, we develop a novel synergy representation for bimanual manipulation motions, that captures the inherent structure of the demonstrated human motions. The goal is the generation of bimanual motion trajectories directly from the bimanual synergy space. We train a variational autoencoder network, shown in Fig. 1, to learn a lower-dimensional synergy representation of human bimanual upper body postures and joint velocities. We compare a traditional Euclidean representation with a Riemannian representation in the latent space. Finally, the representation and generation of human-like bimanual manipulation motions are shown in the learned synergy space.

The main contributions of the paper are (i) a bimanual synergy space for representing a wide variety of household manipulation tasks, and (ii) a comparison between Euclidean and Riemannian latent representations of bimanual human motion synergies.

II. Related Work

The concept of synergies comes from neuroscience and describes correlations between human joints in grasping [1], [3], [4], [5]. It has been adopted widely in robotic grasping to simplify hand design [6], [2], [7], [8] and grasp control [9], [10], [11], [12]. However, it has been shown that synergies do not only exist within hand motions in grasping, but also within upper- and whole-body motions [13], [14]. The whole-body motion generation [15] and control [16] are challenging recent research questions due to the high dimensionality of the joint space. Especially in bimanual

manipulation, direct transfer of unimanual motions is only partially possible [17], as usually both arms work together towards a single goal and need to be coordinated [18]. Therefore, it is beneficial to describe upper body motions in bimanual manipulation tasks in a synergistic representation in order to understand the underlying motion characteristics and simplify motion control.

Bimanual manipulation synergies have been defined by Suárez et al. [19]. They learn a synergy space spanning the postures of two robot arms during teleoperation by performing *Principal Component Analysis* (PCA). The synergy space is learned for each task individually. Hence, the subsequent motion planning needs to take the task into account and pick the synergy space accordingly.

This procedure has later been improved by García et al. [20] by learning zero- and first-order synergies. Zero-order synergies describe the body posture and are thereby equal to the postural synergies commonly used. First-order synergies are learned on joint velocities and thereby capture correlations in motion direction and speed [21]. Synergies are still learned for each task, but the segmentation of the overall space into different synergy cells is performed automatically based on zero- and first-order synergies. These synergies can be used to evaluate and maximize the human likeness in robotic motion generation [22], [23].

For holding, moving and rotating a ball held with two robot arms, local synergies have been found by estimating a Riemannian Gaussian distribution using *Maximum Likelihood Estimation* (MLE) [24]. The covariance of the estimated Gaussian serves as the local synergy description. These can be used to perform a task-specific optimal control.

Based on teleoperated, unimanual robot motions, motion skills have been learned using a *variational autoencoder* (VAE) both in task- and joint-space [25]. In both cases, a Riemannian metric based on the robot's task space is induced in the latent space by training *radial basis functions* (RBF) following the approach presented by Arvanitidis et al. [26]. The learned robotic motion skills are used for motion generation under consideration of obstacle avoidance. In further work, human motions have been split temporally into geodesics, which can interpreted as motion synergies [27].

In this paper, we present bimanual manipulation synergies that are learned from natural human motions. We apply a variational autoencoder to learn a lower-dimensional synergy space from bimanual human motions in different kitchen and household actions. We compare a classical Euclidean synergy representation with a synergy space based on a Riemannian metric for distance measurement. Thereby we aim to allow for functional and human-like motion generation considering the non-linear nature of human motions.

III. Problem Statement

To describe human bimanual manipulation motions and transfer them to the control of bimanual humanoid robots, a simple yet understandable representation of such motions is needed. This work aims to generate bimanual manipulation motions on a kinematic human model. In a second step, these can be mapped to a humanoid robot. To this end, we consider human upper-body motions, including both arms, for everyday activities. These can be described by the arm joint angles. The challenges in learning bimanual arm motions lie in the high-dimensional nature of the data, the nonlinearity of the human configuration space and the need to consider both arms within a single, coordinated motion. To simplify the representation of such motions, we rely on the concept of kinematic synergies, that reduce the complexity of the motion representation by transferring the joint angles into a lower-dimensional synergy space. This synergy space should be shaped to provide an accurate and descriptive representation of the complex joint angle trajectories and should thereby allow for simplified, goal-driven motion planning and control. To achieve this, the synergy space not only needs to provide a good representation of the original joint angles, but also needs to be smooth and should reflect the motion strategies that are naturally applied by humans. In the following, we will discuss these different aspects required for a motion representation suitable for the control of simple, yet human-like manipulation actions.

A. Human Bimanual Manipulation Motions

We search for a simplified representation of human joint angle data for bimanual manipulation tasks. To evaluate the developed representation, we consider motion data from the KIT Bimanual Manipulation Dataset [28]. It consists of 588 motion recordings each demonstrating one complex bimanual manipulation task executed by one of two subjects. The motions are picked from a household context and include cooking and cleaning tasks. Each motion consists of a sequence of basic manipulation actions (e. g. hold, lift, move, pour) that are composed into longer bimanual actions (e. g. serving a drink or mixing a salad).

The recordings have been made with a multi-modal sensor setup, however, we only consider the kinematic data recorded by a VICON motion capture system. The data is mapped to the *Master Motor Map* (MMM) model [29], which is a reference model of the human body that includes kinematics and dynamics based on biomechanical studies. A motion is represented by the angles of the joints within the MMM model. We concentrate on the upper body motion, taking into account both arms with a three *degrees of freedom* (DoF) shoulder joint and two DoF elbow and wrist joints, respectively. Overall, the motion can thereby be described by 14 joint angles representing both human arms. In order to consider the direction of the motion, we further use the joint angular velocity. The velocity is derived from the joint angles using a third-order Savitzky-Golay-Filter with a window size of 21.

The recordings of the KIT Bimanual Manipulation Dataset provide a comprehensive database for manipulations performed in daily activities around the household. However, additional data might be needed to learn specialized motions performed by skilled workers in manufacturing tasks.

B. Kinematic Synergies

To simplify the representation of these bimanual manipulation motions, the concept of kinematic synergies should be applied. Kinematic synergies are originally used in grasp representation [10], [30] and exploit correlations between different joint angles. They open a lower-dimensional space $\mathcal{S} \in \mathbb{R}^m$ that is able to represent the high-dimensional motion $f(t)$ in the joint angle space $\mathcal{F} \in \mathbb{R}^n$ as a synergy motion $s(t)$, with $m < n$. Typically, the motion is described as a timed trajectory $s(t) = (s_{t=0}, ..., s_{t=T})$ in a static, linear synergy space. Different from the widely used hand synergies, we aim to represent motions of both arms during human manipulation actions as series of static body postures in a lower-dimensional subspace. By these means, we strive for a simplified description of bimanual manipulation actions, that captures the essential characteristics of a demonstrated human motion. This lower-dimensional representation shall be general and independent of specific human demonstrations in order to allow for the generation of new robotic manipulation motions that are human-like. The synergy representation shall be general and shall represent any motion seen in the human demonstrations of longer bimanual household actions.

C. Motion Generation

In order to use the derived human motion representation for the control of robotic manipulation actions, it does not only need to be able to flexibly encode the entire range of human manipulation motions, it should also inherently capture the non-linearity of the human configuration space. It has been shown that humans tend to optimize their arm motion with respect to the kinetic energy needed [31]. Such trajectories requiring minimal muscular effort are called geodesics. These geodesics, which represent the shortest paths in a Riemannian space [32], are more complex to describe in the classical Euclidean space representation. Typically, synergy spaces are learned with a Euclidean structure. In order to improve the representation of geodesics, we aim to evaluate the synergy space from a Riemannian perspective. Thereby we aim to properly describe the natural human geodesics and ensure the generation of meaningful motions from the resulting representation space.

IV. Bimanual Synergies

We aim to represent human bimanual manipulation motions as trajectories in a lower-dimensional synergy space. To learn these bimanual posture synergies, we train a variational autoencoder with the joint angles of both human arms during the manipulation motion. The synergies are trained and analyzed according to two assumptions: (i) due to the inherent non-linearity of human motions, a Riemannian space is better suited to describe the motion synergies than a classical Euclidean space, and (ii) in accordance with [20] joint positions and velocities should be considered in learning motion synergies. Therefore we learn from 14 joint angles and 14 velocities for the seven joint rotations in shoulder, elbow and wrist of each arm. Thereby the input data is 28-dimensional for posture and velocity or 14-dimensional when only considering the posture. The human demonstrations of household manipulation motions, as described in Section III-A, are split into training, validation and test set by a ratio of 2:1:1. The bimanual tasks are trained from the long action sequences present in the database. Later, these are split into short actions to evaluate the behavior of the synergy space with respect to different motion characteristics. However, it is important to note that the training data represents a variety of complex bimanual manipulation actions from a household context and is not confined to short, isolated action snippets. The autoencoder architecture and training parameters are defined based on a hyperparameter search. In order to impose a Riemannian metric on the synergy space, coherent with assumption 1, training is performed in a two-step procedure with the standard network training and the training of the Riemannian characteristics of the latent space.

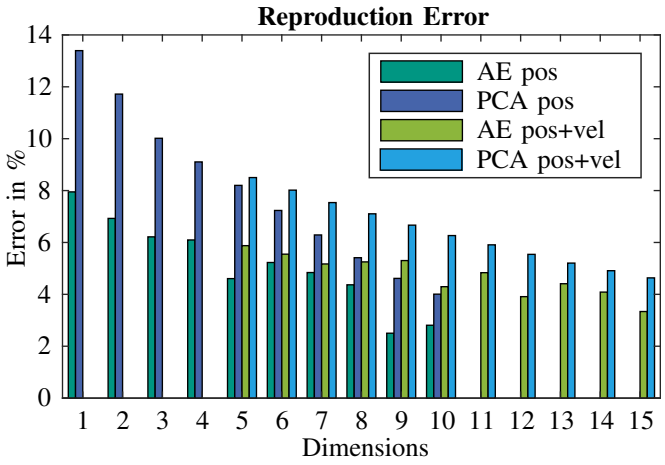


Fig. 2: Reproduction error in percent of the maximum joint angle for synergy spaces with varying dimensions measured on the validation data set; shown is the representation error of synergies learned only on postures (AE pos) and on both postures and velocities (AE pos+vel) as well as the error of a linear dimensionality reduction on both datasets (PCA pos and PCA pos+vel)

A. Autoencoder Architecture

The variational autoencoder network consists of three layers each in the encoder and decoder structure. All layers are fully connected and the *tangens hyperbolicus* is used as activation function. A schematic view of the autoencoder network structure is shown in Fig. 1. The latent synergy space has ten dimensions when trained on posture and velocity and five dimensions when trained on posture only. A posture is comprised of the joint angles of shoulder (3 DoF), elbow (2 DoF) and wrist (2 DoF) in both arms. As shown in Fig. 2, the reproduction error within the validation data set is optimal for these dimensions. When more synergy variables are provided, overfitting to the training data can be observed, while less synergy variables cause a higher reproduction error. For modeling the output of the decoder, we opt for a Gaussian representation in order to account for the probabilistic nature of the resulting bimanual posture. Therefore, both for the latent and for the output layer, a mean and a standard deviation are learned in parallel.

TABLE I: Autoencoder network and training parameters

Parameter	Value
Training epochs	300
RBF training iterations	100
L1 regularization λ	$1e^{-7}$
Gradient clipping c	1

B. Training and Optimization

The network is trained in a two-step procedure. First, the latent mean and the reconstructed bimanual postures are trained in a classical way. In a second step, the latent variance is trained in order to induce a Riemannian representation of the probabilistic latent space.

For the first autoencoder training step, the loss function consists of the representation quality RQ and the Kullback-Leibler divergence KLD:

$$\mathcal{L} = \alpha\,\mathrm{KLD} - \mathrm{RQ} \tag{1}$$

with a weighting parameter α that is gradually increased during learning. Thereby, the general representation is promoted first and over the training iterations, more emphasis is also given to the latent distribution enforced by the KLD loss. We start training with the representation loss only ($\alpha = 0$) for the first half of the training and then increase α up to one during the second half of the training.

Training is performed using an Adam optimizer with a learning rate of $1e^{-3}$. The latent variance is kept constant throughout this part of the training. An overview of the network and training parameters is given in Table I.

C. Latent Variance Representation

The variance of the latent synergy space is learned separately, taking into account the distribution of the given training data. By these means, a Riemannian metric can be imposed on the synergy space to allow geodesic motion generation, as described by Arvanitidis et al. [26]. The variance σ (corresponding to the σ-layer in Fig. 1) is represented by modeling the precision β with a *radial basis function (RBF)* neural network [33]. The network has one hidden layer with 100 neurons with an RBF activation function. The RBF centers are initialized by a K-Means clustering. During training, a least squares objective function is applied to fit the RBF model. The precision is defined as $\beta = \sigma^{-2}$. By ensuring the extrapolation of β towards zero, it can be guaranteed that the variance increases with the lack of training data. The RBF kernels are fitted in a second training step, after the mean of the latent representation is already trained with the autoencoder network. By these means, we separately train a well-structured variance of the latent synergy space also at the borders of the latent area covered by the training data.

This structured synergy variance allows us to impose not only a Euclidean, but also a Riemannian metric on the synergy space. During motion planning, traditionally a Euclidean metric is applied to find the shortest path between two postures in the cartesian coordinate system spanned by the synergies. In addition, we study a Riemannian metric $\mathbf{M}_z = \mathbf{J}_z^T \mathbf{J}_z$ motivated by the nonlinearity of human motion patterns. Due to the stochastic nature of the autoencoder, the resulting metric is the sum of the metrics applied individually to the mean and variance functions of the latent space $\mathbf{M}_z = \mathbf{M}_z^{(\mu)} + \mathbf{M}_z^{(\sigma)}$. The calculation of this metric is straightforward within the area covered by the training data. Since the latent variance is explicitly trained to increase at the boundaries of the training data, $\mathbf{M}_z^{(\sigma)}$ becomes high in unseen regions and ensures a smooth representation of the latent Riemannian metric.

V. Evaluation

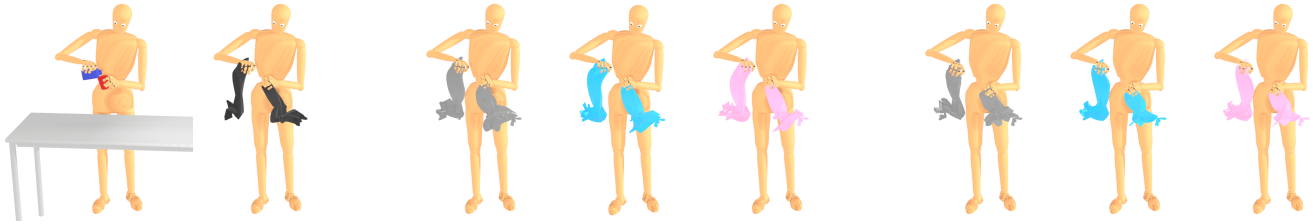
In order to assess the validity of the synergy space, we analyze the reproduction error for human motion demonstrations. The representation quality is also compared to a classical, linear synergy space. In addition, the smoothness and structure of the synergy space is evaluated and the merit of applying the Riemannian metric is analyzed. Finally, we assess the capability of the synergy representation to generate new, human-like motions.

A. Autoencoder Validation

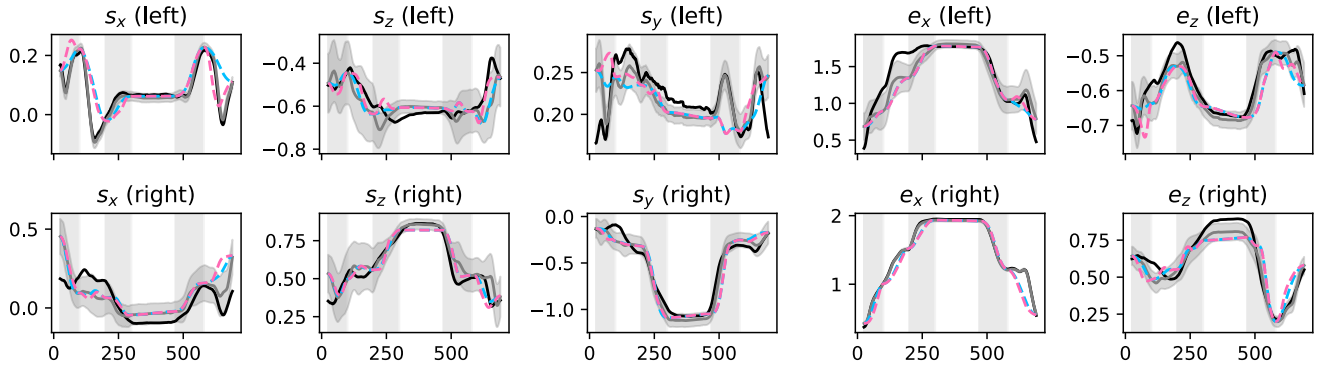
The representation quality of the latent bimanual synergy space is evaluated based on the *mean squared error* (MSE) between the demonstrated human motions and their reproduction from synergy space. The error is calculated for entire bimanual manipulation task sequences and is given as the squared mean of the joint angle deviation in radian. By these means, the autoencoder achieves a reproduction error on the test dataset of 0.0022 for the synergy space trained with posture only and 0.0024 for synergies trained on posture and velocity. Two exemplary reconstructed bimanual motion for both synergy spaces are shown for *pouring* into a cup in Fig. 3 a) and for *opening* a bottle in Fig. 4 a) with the original and deconstructed motions depicted in black and gray respectively. In comparison, a PCA is performed on the same data and a linear synergy space is constructed with the five (for posture only) or ten (for posture and velocity) most informative principal components as synergies. These linear synergies achieve a reconstruction error of 0.0067 learned only on postures or 0.0039 learned on posture and velocity. Hence, the trained bimanual synergy space improves reconstruction by at least $38\,\%$ compared to the classical approach.

In addition to the reproduction error, the smoothness of the latent space is of great importance, especially when the latent space shall be used for data generation. To validate the smoothness of the latent space, we follow the approach by Dimou et al. [34]. Postures are uniformly sampled from the area of latent space, that is covered by the training data. By measuring the distance between postures that are adjacent in the latent space, a relative measure of the smoothness of transitions in latent space is obtained.

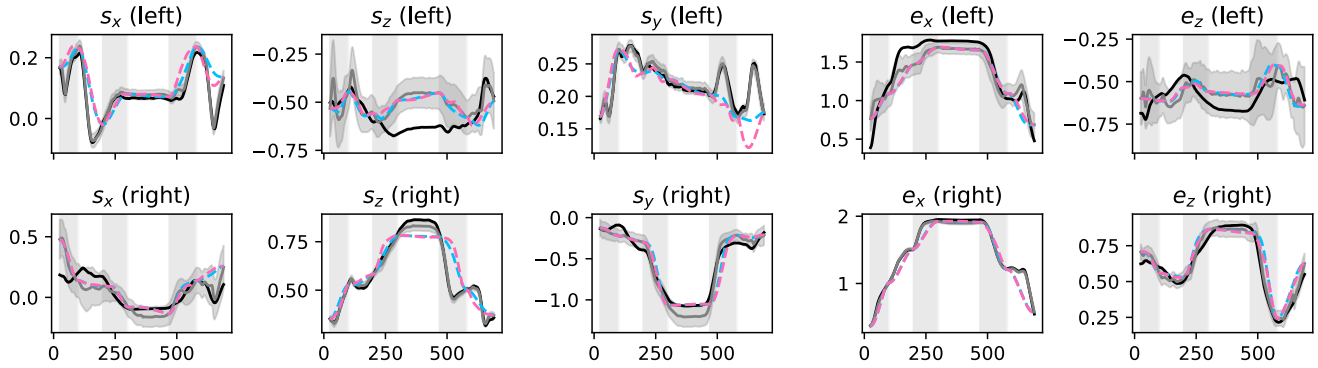
For sampling the latent space with a step size of 0.1, a mean posture variation of $4.1 \cdot 10^{-4}$ is measured. For the linear reference synergy space obtained by performing a PCA



(a) End-effector trajectories based on synergies; a static posture from the motion (frame 400) is shown on the MMM model of the human body



(b) Joint angle trajectories based on synergies learned only on bimanual posture



(c) Joint angle trajectories based on synergies learned on bimanual posture and velocity

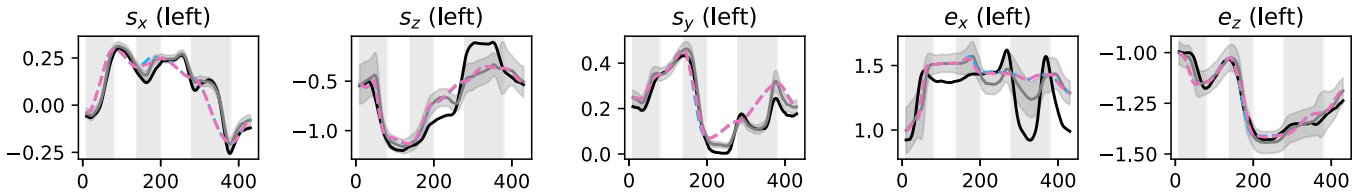
Fig. 3: End effector and joint trajectories of the MMM reference model [29] from synergies learned on *(a, b) only posture* or *(a, c) posture and velocity* for a pouring motion (Pour1_cs_ms_0_20l_02 of subject 1480 from test data). *(b, c)* - Joint angle trajectories in radian (*y-axis*) over time in frames (*x-axis*) of the three shoulder (s_x, s_z, s_y) and two elbow (e_x, e_z) joints for left (*top*) and right (*bottom*) arm. The order of the joints corresponds to the kinematic chain of the MMM arm. The human demonstration (—) and the reconstructed (—) trajectory including variance (of the decoder) are visualized as well as the generated and decoded shortest paths with respect to the Euclidean (- -) and the Riemannian metric (- -) in the latent space between segmented configurations ( ). *(a) - left to right:* Original recording, demonstrated human motion, reconstructed (*posture*), Euclidean (*posture*), Riemannian (*posture*), reconstructed (*posture and velocity*), Euclidean (*posture and velocity*), and Riemannian (*posture and velocity*).

and using the same number of synergy parameters, the mean posture variation amounts to $7.1 \cdot 10^{-4}$. Since the applied measure of smoothness only allows for a relative comparison, the latent synergy space learned by the autoencoder is proven to be at least as smooth as the linear PCA synergies, which are continuous by design.

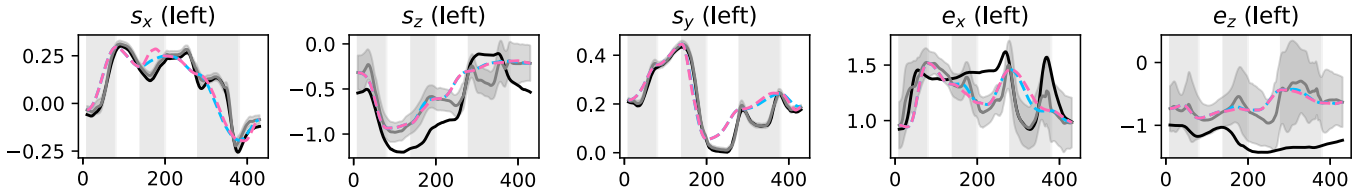
B. Synergy Space

The Riemannian metric ensures that motions observed in the human demonstrations are represented within the structure of the latent synergy space. Fig. 5 visualizes the geometric measure based on the Riemannian metric in the first three variables of the latent synergy space learned only from the body postures. It can be seen that the variance increases in areas of high uncertainty, specifically at the border encasing the area covered by the training data. The high variance in turn increases the Riemannian metric via $\mathrm{M}_z^{(\sigma)}$. For motion planning in the synergy space, these areas of high uncertainty are therefore avoided due to the high metric cost and planned motions follow the inherent pattern of the demonstrated human motions. This synergy space has been trained on a large variety of human bimanual tasks, as described in Section III-A. Overall, the borders of the learned latent synergy space are neatly defined by the metric. However, due to the high variety of motions represented

(a) End-effector trajectories based on synergies; a static posture from the motion (frame 185) is shown on the MMM model of the human body



(b) Joint angle trajectories of the left arm based on synergies learned only on bimanual posture



(c) Joint angle trajectories of the left arm based on synergies learned on bimanual posture and velocity

Fig. 4: End effector and joint trajectories from synergies learned on *(a, b) only posture* or *(a, c) posture and velocity* for opening a bottle (Open3_ms_180_30h_05 of subject 1723 from training data), the naming and color scheme of the plots is explained in Fig. 3, only the trajectories for the left arm are shown that unscrews the lid of the bottle

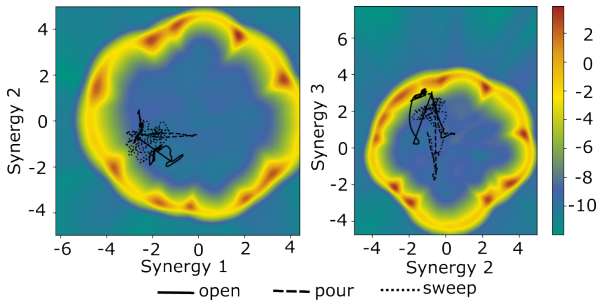


Fig. 5: Structure of the geometric measure in the first three variables of the bimanual synergy latent space trained only on postures with three exemplary trajectories from the test set; the area of the training data is encased by a barrier of high uncertainty and thereby results in a high Riemannian metric

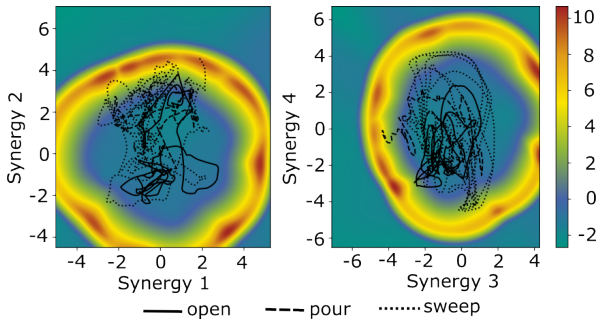


Fig. 6: Structure of the geometric measure in the first four variables of the bimanual synergy latent space trained on postures and motion velocities with three exemplary trajectories from the test set; including joint velocities in synergy training results in a smoother boundary of the synergy space and more expansive synergy trajectories

by the synergy space, the synergy representation is evenly distributed in a smooth synergy space with no structural separation. This also means that there are no strictly separated clusters inherent in the human demonstrations, but instead it is generally possible to perform very different motions from the same starting posture.

Fig. 6 shows the synergy space learned from both posture and joint velocity. The additional input of joint velocities provides the autoencoder with a notion of the motion in progress and therefore allows to distinguish different motions that pass through the same posture, but fulfill very different tasks. It can be seen that the resulting synergy space is smoother at the boundaries. In addition, each individual motion covers a large part of the synergy space, since the same posture needs to be represented by an area within the synergy space in order to distinguish between differences in motion speed and direction.

Overall, different bimanual tasks are mapped to different areas within the synergy space. All considered tasks except for sweeping the floor are separated. Since sweeping motions are larger and include the entire body, this task spreads over the entire synergy space. Due to the local separation of tasks in the synergy space, new synergy trajectories can be associated to tasks.

C. Motion Generation

We evaluate the learned synergy space and motion generation capabilities on a bimanual demonstration of a *pouring* motion and a motion for *opening* a bottle for both synergy

spaces learned only on posture and both on posture and velocity. These actions are part of larger motion sequences, namely grasping a bottle and a glass or cup from the table, pouring from the bottle into the container and placing all objects back (pouring) and grasping a bottle from the table, unscrewing the lid, and placing the lid and the bottle back on the table (opening). For motion generation, these action sequences need to be split into individual actions in order to account for boundary conditions (e. g. the location of the objects on the table). Therefore, we focus on the central pouring and opening actions in these sequences. It is important to note that the synergy representation allows to generate a trajectory that ensures the coordination of both arms. The temporal progress of the motion execution needs to be defined manually.

We consider the following three approaches: (i) Reconstruction of the joint trajectories using the forward path of the trained VAE (see validation in Section V-A), (ii) straight lines in the latent space between key configurations, i.e., assuming a Euclidean metric, and (iii) shortest paths (geodesics) based on the induced Riemannian metric. The geodesics are computed by solving the boundary value problem (given the initial and final configuration mapped in the latent space) using the fixed-point method [35]. Given that the velocity of a movement along a geodesic remains constant, it is possible to reparameterize the geodesic by modifying the acceleration at which the path is traversed [31], [27], e.g., in order to move from a stationary position. For comparability, this reparameterization is applied to both the geodesic and the straight line.

The resulting trajectories of all three methods applied for *pouring* into a cup are presented in Fig. 3. For *opening* a bottle, the trajectories of the left arm that unscrews the lid are shown in Fig. 4. Both the Euclidean and Riemannian motion generation provide human-like manipulation motions. This can be seen in Fig. 3 b) and c) and Fig. 4 b) and c) in comparison to a motion trajectory demonstrated by a human subject. Motions generated using the Euclidean and Riemannian metric are very similar. In general, the generated motions are very similar to the reconstructed human demonstration, which matches the overall representation error. A small deviation is mainly visible in the y-axis of the left shoulder for pouring and the z-axis of the left elbow for opening. The learned uncertainty of the decoder reflects the quality of the synergy-generated motion data. Whenever the reconstructed data deviates from the original human demonstration, the synergy uncertainty also increases.

The visualization of demonstrated human motions together with the motions generated from the bimanual synergy space in Fig. 3 a) and Fig. 4 a) shows that the synergy-generated motions are human-like and generally functional.

This qualitative assessment is supported by a quantitative analysis of the postural deviation of generated motions from a human ground truth trajectory. The results of this evaluation are shown in Table II. The *mean squared error* (MSE) of the joint angles (in radians) between the generated trajectory and the human motion data as ground truth or the reconstructed trajectory is used as a metric for comparison. The approaches based on Euclidean and Riemannian metrics in the latent space yield generally similar results. This similarity is likely due to the extensive distribution of the training data, yielding a more even metric space and the relatively small distances between the key joint configurations in the segmented original data. However, the synergy space is practicable for the generation of new, human-like bimanual manipulation motions.

TABLE II: Quality of synergy-generated motions (MSE)

	Pour Example			
	Only Posture		**Posture and Velocity**	
	Ground truth	Reconstr.	Ground truth	Reconstr.
Reconstr.	0.008	-	0.008	-
Euclidean	0.012	0.004	0.011	0.005
Riemannian	0.013	0.005	0.011	0.005
	Open Example			
	Only Posture		**Posture and Velocity**	
	Ground truth	Reconstr.	Ground truth	Reconstr.
Reconstr.	0.012	-	0.021	-
Euclidean	0.014	0.003	0.025	0.007
Riemannian	0.014	0.003	0.023	0.007

VI. Conclusion

This paper presents a novel bimanual synergy space that fulfills the objective to generate human-like bimanual motion trajectories, which have not been directly demonstrated by a human. The synergy space is learned by training a variational autoencoder with demonstrated human manipulation motions. In addition to the classical autoencoder training, a Riemannian metric is learned on the variance of the latent synergy space. The learned synergy space is able to represent bimanual human manipulation motions as trajectories of the joint posture and velocity and is significantly smoother than a classical linear synergy space derived by a PCA. Training the synergy space on both posture and joint velocity provides an inherent representation of the dynamic motion and yields a smoother boundary of the synergy space. On the other hand, a synergy space trained only on postures requires less synergy variables and achieves similar quality in the representation and generation of motions. We compare a Riemannian latent embedding with the classical Euclidean synergy representation with respect to the generation of human-like bimanual manipulation motions. Both classical Euclidean synergy paths as well as geodesics leveraging the Riemannian metric of the synergy space allow for the generation of human-like bimanual manipulation motions.

Thereby, the bimanual synergy space provides a simple yet comprehensive representation to describe and generate human-like manipulation motions. Our synergy space is able to represent a wide range of daily household activities. The Riemannian latent embedding ensures a correct variance metric to guide the generation of motions.

In the future, we plan to evaluate the motion generation based on geodesics with respect to interpolation capabilities [26] and consider training and evaluating multiple VAEs

on a smaller subset of the demonstrations for specific bimanual actions or categories [18] to highlight the advantages of the Riemannian metric. So far, the bimanual synergies are purely kinematic in nature, hence we plan to use them as a feedforward trajectory generator for a humanoid robot that needs to be complemented with a feedback controller for stability. We believe that this representation can simplify and improve the design of bimanual manipulation control for humanoid robots.

REFERENCES

[1] M. Santello, M. Flanders, and J. F. Soechting, "Postural Hand Synergies for Tool Use," *The Journal of Neuroscience*, vol. 18, no. 23, pp. 10 105–10 115, 1998.

[2] A. Bicchi, M. Gabiccini, and M. Santello, "Modelling natural and artificial hands with synergies," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, no. 1581, pp. 3153–3161, 2011.

[3] E. J. Weiss and M. Flanders, "Muscular and Postural Synergies of the Human Hand," *Journal of Neurophysiology*, vol. 92, no. 1, pp. 523–535, 2004.

[4] J. K. Shim and J. Park, "Prehension synergies : principle of superposition and hierarchical organization in circular object prehension," *Experimental Brain Research*, vol. 180, pp. 541–556, 2007.

[5] M. Santello, M. Bianchi, M. Gabiccini, E. Ricciardi, G. Salvietti, D. Prattichizzo, M. Ernst, A. Moscatelli, H. Jörntell, A. M. Kappers, K. Kyriakopoulos, A. Albu-Schäffer, C. Castellini, and A. Bicchi, "Hand synergies: Integration of robotics and neuroscience for understanding the control of biological and artificial hands," *Physics of Life Reviews*, vol. 17, pp. 1–23, 2016.

[6] C. Brown and H. Asada, "Inter-Finger Coordination and Postural Synergies in Robot Hands Via Mechanical Implementation of Principal Components Analysis," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2007, pp. 2877–2882.

[7] S. B. Godfrey, A. Ajoudani, M. Catalano, G. Grioli, and A. Bicchi, "A synergy-driven approach to a myoelectric hand," in *IEEE Int. Conf. on Rehabilitation Robotics*, 2013.

[8] M. Catalano, G. Grioli, E. Farnioli, A. Serio, C. Piazza, and A. Bicchi, "Adaptive Synergies for the Design and Control of the Pisa/IIT SoftHand," *Int. Journal of Robotics Research*, vol. 33, no. 5, pp. 768–782, 2014.

[9] M. T. Ciocarlie and P. K. Allen, "Hand Posture Subspaces for Dexterous Robotic Grasping," *Int. Journal of Robotics Research*, vol. 28, no. 7, pp. 851–867, 2009.

[10] J. Romero, T. Feix, C. H. Ek, H. Kjellström, and D. Kragic, "Extracting Postural Synergies for Robotic Grasping," *IEEE Transactions on Robotics*, vol. 29, no. 6, pp. 1342–1352, 2013.

[11] B. A. Kent, J. Lavery, and E. D. Engeberg, "Anthropomorphic Control of a Dexterous Artificial Hand via Task Dependent Temporally Synchronized Synergies," *Journal of Bionic Engineering*, vol. 11, no. 2, pp. 236–248, 2014.

[12] J. Starke, C. Eichmann, S. Ottenhaus, and T. Asfour, "Human-inspired representation of object-specific grasps for anthropomorphic hands," *International Journal of Humanoid Robotics (IJHR)*, vol. 17, no. 2, p. 2050008, 2020.

[13] T. R. Kaminski, "The coupling between upper and lower extremity synergies during whole body reaching," *Gait & Posture*, vol. 26, no. 2, pp. 256–262, 2007.

[14] N. García, R. Suárez, and J. Rosell, "Task-Dependent Synergies for Motion Planning of an Anthropomorphic Dual-Arm System," *IEEE Transactions on Robotics*, vol. 33, no. 3, pp. 756–764, 2017.

[15] X. Chen, B. Jiang, W. Liu, Z. Huang, B. Fu, T. Chen, and G. Yu, "Executing your Commands via Motion Diffusion in Latent Space," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 18 000–18 010.

[16] T. He, W. Xiao, T. Lin, Z. Luo, Z. Xu, Z. Jiang, C. Liu, G. Shi, X. Wang, L. Fan, and Y. Zhu, "HOVER: Versatile Neural Whole-Body Controller for Humanoid Robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 9989–9996.

[17] G. Lu, T. Yu, H. Deng, S. S. Chen, Y. Tang, and Z. Wang, "AnyBimanual: Transferring Unimanual Policy for General Bimanual Manipulation," 2024, arXiv:2412.06779 [cs:RO].

[18] F. Krebs and T. Asfour, "A bimanual manipulation taxonomy," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 4, pp. 11 031–11 038, 2022.

[19] R. Suárez, J. Rosell, and N. García, "Using Synergies in Dual-Arm Manipulation Tasks," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 5655–5661.

[20] N. García, R. Suárez, and J. Rosell, "First-Order Synergies for Motion Planning of Anthropomorphic Dual-Arm Robots," in *World Congress of the Federation of Automatic Control*, 2017, pp. 2247–2254.

[21] N. García, J. Rosell, and R. Suárez, "Motion planning using first-order synergies," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2015, pp. 2058–2063.

[22] ——, "Motion Planning by Demonstration With Human-Likeness Evaluation for Dual-Arm Robots," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 49, no. 11, pp. 2298–2307, 2019.

[23] A. Meixner, M. Carl, F. Krebs, N. Jaquier, and T. Asfour, "Towards unifying human likeness: Evaluating metrics for human-like motion retargeting on bimanual manipulation tasks," in *IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, May 2024, pp. 13 015–13 022.

[24] M. J. A. Zeestraten, I. Havoutis, S. Calinon, and D. G. Caldwell, "Learning task-space synergies using Riemannian geometry," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 73–78.

[25] H. Beik-Mohammadi, S. Hauberg, G. Arvanitidis, G. Neumann, and L. Rozo, "Reactive motion generation on learned Riemannian manifolds," *The International Journal of Robotics Research*, vol. 42, no. 10, pp. 729–754, 2023.

[26] G. Arvanitidis, L. K. Hansen, and S. Hauberg, "Latent Space Oddity: on the Curvature of Deep Generative Models," in *International Conference on Learning Representations*, 2018, pp. 1–15.

[27] H. Klein, N. Jaquier, A. Meixner, and T. Asfour, "A Riemannian take on human motion analysis and retargeting," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Kyoto, Japan, October 2022, pp. 5210–5217.

[28] F. Krebs, A. Meixner, I. Patzer, and T. Asfour, "The KIT bimanual manipulation dataset," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Munich, Germany, July 2021, pp. 499–506.

[29] C. Mandery, O. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "Unifying representations and large-scale whole-body motion databases for studying human motion," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 796–809, August 2016.

[30] J. Starke and T. Asfour, "Kinematic synergy primitives for human-like grasp motion generation," in *IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, May 2024, accepted.

[31] A. Biess, D. G. Liebermann, and T. Flash, "A Computational Model for Redundant Human Three-Dimensional Pointing Movements: Integration of Independent Spatial and Temporal Motor Plans Simplifies Movement Dynamics," *Journal of Neuroscience*, vol. 27, no. 48, pp. 13 045–13 064, 2007.

[32] F. Bullo and A. D. Lewis, "Geometric Control of Mechanical Systems," in *Texts in Applied Mathematics*. Springer, 2005, vol. 49.

[33] Q. Que and M. Belkin, "Back to the Future: Radial Basis Function Network Revisited," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 1856–1867, 2020.

[34] D. Dimou, J. Santos-Victor, and P. Moreno, "Learning Conditional Postural Synergies for Dexterous Hands: A Generative Approach Based on Variational Auto-Encoders and Conditioned on Object Size and Category," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 4710–4716.

[35] G. Arvanitidis, S. Hauberg, P. Hennig, and M. Schober, "Fast and robust shortest paths on manifolds learned from data," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2019, pp. 1506–1515.