

International Journal of Humanoid Robotics
© World Scientific Publishing Company

HUMAN-INSPIRED REPRESENTATION OF OBJECT-SPECIFIC GRASPS FOR ANTHROPOMORPHIC HANDS

JULIA STARKE, CHRISTIAN EICHMANN, SIMON OTTENHAUS and TAMIM ASFOUR

*High Performance Humanoid Technologies Lab
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT)
Adenauerring 2, 76131 Karlsruhe, Germany
{julia.starke, asfour}@kit.edu*

Received Day Month Year

Revised Day Month Year

Accepted Day Month Year

The human hand is a complex, highly-articulated system, which has been the source of inspiration in designing humanoid robotic and prosthetic hands. Understanding the functionality of the human hand is crucial for the design, efficient control and transfer of human versatility and dexterity to such anthropomorphic robotic hands. Although research in this area has made significant advances, the synthesis of grasp configurations, based on observed human grasping data, is still an unsolved and challenging task. In this work we derive a novel, constrained autoencoder model, that encodes human grasping data in a compact representation. This representation encodes both the grasp type in a three-dimensional latent space and the object size as an explicit parameter constraint allowing the direct synthesis of object-specific grasps. We train the model on 2250 grasps generated by 15 subjects using 35 diverse objects from the KIT and YCB object sets. In the evaluation we show that the synthesized grasp configurations are human-like and have a high probability of success under pose uncertainty.

Keywords: Grasping; Anthropomorphic Hands; Hand Synergies.

1. Introduction and Related Work

The human hand is a versatile and complex system containing 23 *degrees of freedom* (DOF)^{1,2,3}. It provides robust and sensitive grasping capabilities and allows dexterous manipulation of diverse objects. Improving the understanding of human grasp control is a necessary condition to foster the transfer of such grasping abilities onto anthropomorphic robotic hands. To reduce the complexity in design and control of anthropomorphic hands, researchers addressed the question of finding a low-dimensional representation of grasp postures. To that end, couplings within the grasp posture configuration space have been observed during natural human grasping motions. These couplings have been referred to in the literature as grasping synergies. The synergies allow the description of grasping postures in a significantly lower dimensional space while preserving their functional characteristics. However,

a direct, intuitive understanding of a grasp based on its synergy is usually not possible. On the one hand, individual dimensions of synergies cannot be mapped directly to descriptive parameters such as aperture or object size. On the other hand, a dimension in the synergy space often encodes several properties of a grasp simultaneously.

The existence of *postural hand synergies* in static grasp poses was first described by Santello et al. in 1998⁴. They could show that only two dimensions extracted by a *Principal Component Analysis* (PCA) are sufficient to account for more than 80 % of the postural information contained in a grasp configuration, being an inspiration for many further studies on synergy representation and implementation. However, the authors also showed that descriptive grasp type classifications or object properties were not directly visible in the low dimensional synergies. While these observed postural synergies can be partially explained by muscular joint couplings, Weiss and Flanders showed that muscular and postural synergies are not identical⁵. They found that roughly a third of all postural synergies was significantly involved in the execution of one muscular synergy based on *electromyographic* (EMG) activation. This was confirmed by Castellini and van der Smagt⁶ performing a PCA on 18 joint angle values and ten EMG readings. While the first two muscular synergies and the first three postural synergies accounted for more than 70 % of the total variance, both sets of synergies did not correlate for the same grasps.

Regarding static postural synergies, Bicchi et al. introduced the model of *soft synergies* to increase the flexibility of synergy-based grasping regarding the shape of the object⁷. By temporally sampling from a space of postural synergies, Romero et al. introduced a fundamental concept for *kinematic synergies* to describe the entire pregrasp motion⁸. A comparison of several linear and nonlinear dimensionality reduction methods was carried out, showing that the nonlinear nature of human grasping data cannot be entirely described by linear reduction techniques like the mainly applied PCA. Other approaches regarding kinematic synergies include the interpolation of joint trajectories⁹ and the description as a combination of short and basic *eigenmotions* extracted from human grasping data. With a non-linear representation, the reproduction error of hand postures was notably decreased, but the proposed model still did not contain intuitively understandable information describing the resulting grasp.

In control, highly articulated robotic hands offer a large amount of individually adjustable degrees of freedom, thereby enabling versatile motion capabilities. Nevertheless the subspace of sensible posture configurations resulting in feasible grasps or gestures is considerably smaller than the configuration space of such hands. Ciocarlie and Allen first applied the synergy findings from neuroscience to the control of several robotic hands ranging from three fingered grippers to a model of the fully actuated human hand¹⁰. Two *eigengrasps* dividing the available joint space into two main sets of correlating joints were defined for each robotic hand individually. Thereby, a low-dimensional subspace for computational grasp optimization was created. An application of the eigengrasp control scheme on a fully actuated humanoid

robotic hand was presented by Wimböck et al.¹¹. In addition to a simplification of the postural hand control, Gabbicini and Bicchi showed that synergies can also be applied to estimate and monitor grasp forces taking into account a dynamic model of the humanoid robotic hand¹².

Besides easing the coordinated control of individually actuated joints, the mechanical adaptive coupling of several degrees of freedom in underactuated mechanisms is widely applied in humanoid hands. Brown and Asada first presented a tendon-based, underactuated distribution based on a synergy representation. It allows the exact implementation of eigenpostures as a representation of the synergy based coupling of joints¹³. This was later extended with additional actuation capabilities for adjusting movements¹⁴. The concept of *adaptive synergies* by Catalano et al. transfers soft synergies to a mechanically implementable, space-optimized mechanism¹⁵ and paved the way for several robotic¹⁶ and prosthetic¹⁷ hands synergistically driven by a single motor. Further synergy-inspired underactuation mechanisms include the findings of synergy manifold optimization⁹ and adaptive, parametric coupling of finger closing behaviour¹⁸. Exploiting the force controllability of synergy-based posture descriptions, Gabbicini et al. developed a kinostatic characterization for underactuated power grasping¹⁹, which was later extended to cover precision grasps²⁰. The existence of individual force synergies describing hand contact forces independent of finger posture was shown by Santello and Soechting²¹ followed by a broad overview on synergy-based approaches for human grasping description and the underlying correlations²².

While the merit of postural grasp synergies is well documented, the synthesis of new grasp configurations with a specific grasp type remains challenging. The main contribution of our work is the ability to synthesize human-like grasping configurations which (1) are derived from human grasping data, (2) contain implicitly encoded grasping parameters, (3) enable explicit parameterization of the grasp aperture, and (4) have a high grasp success probability. To our best knowledge, such a representation is novel and has not been proposed in the literature before. In detail, the four main aspects of the contribution can be broken down as follows:

(1) We collect 2250 human grasping examples involving 15 subjects that are tasked with grasping 35 diverse objects from the KIT²³ and YCB²⁴ object sets.

(2) Taking into account the non-linear nature of hand joint configurations, we train a deep autoencoder network for synergy extraction and additionally constrain it to provide an intuitive encoding of grasp classification.

(3) We introduce a constraint to the deep autoencoder by passing the diameter of the target object as an explicit additional parameter to the decoder part of the network. This allows the encoder to focus on encoding a grasp descriptor independent of object size within the latent space.

(4) Compared to our previous work²⁵, we show that the decoder of the network can synthesize high-quality grasp configurations while respecting a given grasp type and object size.

Table 1: Information about the data collection

Characteristic	Value
grasps	2250
subjects	15 (9 male, 6 female)
age	27.0 years \pm 2.0 years
average hand length	180.2 mm \pm 16.9 mm
objects	35
grasp types	5

2. Human Grasping Study

For the extraction of a universally applicable, general synergy model, a large-scale study of human grasping postures was performed. Detailed information on the conditions of data collection are listed in Table 1. Its characteristics and execution are described hereafter, followed by fundamental observations extracted from the gathered human grasping data.

2.1. Data Acquisition

A description of human grasping postures is gathered by recording angle data of individual finger joints within the human hand while performing static grasps. The angles are measured with a CyberGlove III (Cyber Glove Systems LLC, USA) containing 18 joint angle sensors recording angular positions of metacarpophalangeal and proximal finger joints as well as adduction/abduction, thumb circumduction and the curvature of the palm. As hand orientation in space is only depending on the relative object pose and not relevant for the shape of the grasp itself, the two angles describing wrist motion are omitted within this study.

The remaining 16 joint angle parameters, whose notation is depicted in Fig. 1, are calibrated with an adjusted version of the protocol proposed by Gracia-Ibáñez et al.²⁶. All finger angles are calibrated assuming a linear dependency by measuring two reference positions captured with the help of wooden blocks. Contrary to the cited protocol, we position the reference blocks on the palmar side of the finger to avoid interference with the dorsally attached glove sensors. The angle of palm curvature is measured for each participant according to their individual range of motion. For visualization purposes, the acquired joint angle data is transferred to the human hand model of the *Master Motor Map*^{2,3} with 23 degrees of freedom, scaled according to a body height of 1.70 m and a mass of 70 kg. The exact derivation of segment lengths and widths is described in the *Master Motor Map*^{2,3}. However, the trained algorithm for synergy extraction described in the following section is designed only on the mere joint angle data.

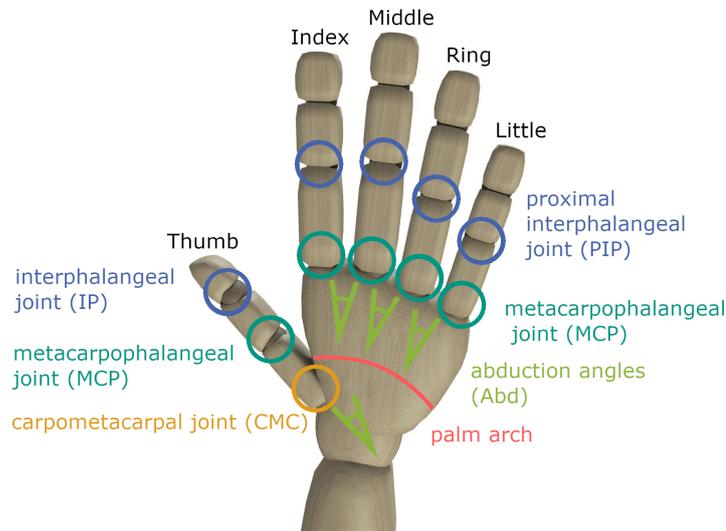


Fig. 1: Annotation of the considered joint angles; the abbreviations used in all following graphs are comprised of the first letter of the finger followed by the associated joint name (e.g. I-PIP for the index proximal interphalangeal joint)

2.2. Grasp Procedure

Within the presented study, 2250 grasps by 15 subjects carried out on 35 objects were gathered. Some exemplary human grasps are presented in Fig. 2. All subjects gave informed signed consent before participating in the presented recordings. While a central aim of the study was to gather natural, intuitive grasping behaviour, the resulting grasps needed to be classified according to descriptive categories of hand posture. Based on Cutkosky's grasp taxonomy²⁷ five different grasp types were defined as listed in Table 2. While this subsumes grasps with similar finger configurations, it is still flexible enough to allow for individual adjustments of grasp characteristics by including power as well as precision grasps. This enables subjects to grasp the presented objects in a natural, intuitive manner according to their estimate of necessary grasp force and contact properties. The choice of recorded grasp types was taken based on their frequency of occurrence in activities of daily living^{28,29}.

For each grasp type the subjects performed grasps on ten different objects taken from the *KIT Object Database*^{23a} and the *YCB Object Set*^{24b}. A list of all used objects and their association to the considered grasp types can be found in Table 3. At the beginning of each data acquisition set, the defined grasp type was explained

^a<https://h2t-projects.webarchiv.kit.edu/Projects/ObjectModelsWebUI/>

^b<http://www.ycbbenchmarks.com/>



Fig. 2: Human grasps performed for data acquisition; cylindrical grasp on a banana (a), disk grasp on a tuna can (b), pinch grasp on a spoon (c) and lateral grasp on a bowl (d)

Table 2: Classification of applied grasp types within Cutkosky’s taxonomy²⁷

Grasp Type	Cutkosky’s Taxonomy Class ²⁷
cylindrical	prismatic power large diameter (1)
	prismatic power small diameter (2)
spherical	circular power sphere (11)
	circular precision sphere (13)
disk	circular power disk (10)
	circular precision disk (12)
pinch	prismatic precision thumb-index finger (9)
	prismatic precision thumb-2 finger (8)
lateral	power lateral pinch (16)

to the participant with the help of the corresponding drawings from Cutkosky’s taxonomy²⁷. The recording then started with the subject positioning their hand flat on the table as the reference pose. In the following, each of the objects was grasped three times in a row and the recording ended with the hand positioned in the flat reference pose again. All five grasp types were taken subsequently in one session. The complete data acquisition including glove calibration lasted roughly half an hour for each participant. All subjects performed the grasping procedure with their dominant hand. The study contains data of 14 right- and one left-handed participant. A comparison of grasping data for the purpose of this work showed no significant difference regarding the variation of handedness. The resulting grasping data, accompanied by videos of the study’s procedure, are publicly available on the *KIT Whole Body Human Motion Database*^{30c}.

The mean intra-subject variability calculated by comparing all three demonstrations on the same object by the same subject accounts to $2.94 \pm 1.32^\circ$ without taking into account adduction of the middle finger. The inter-subject variability between

^c<https://motion-database.humanoids.kit.edu/>

Table 3: List of objects grasped with the five chosen grasp types

Object	Grasp Type	Database
Spoon	cylindrical, pinch	YCB
Sweetener	cylindrical, disk	KIT
Mug	cylindrical, disk, lateral	YCB
Flat Screwdriver	cylindrical	YCB
Small Green Cup	cylindrical, disk, lateral	YCB
Small Orange Cup	cylindrical, disk	YCB
Hammer	cylindrical	YCB
Tomato Soup Can	cylindrical, disk	YCB
Ravioli Large	cylindrical, disk	KIT
Plastic banana	cylindrical	YCB
Plastic orange	spherical	YCB
Golf Ball	spherical	YCB
Raquetball	spherical	YCB
Plastic apple	spherical	YCB
Tennis Ball	spherical	YCB
Plastic pear	spherical	YCB
Mini soccer ball	spherical	YCB
Baseball	spherical	YCB
Soft ball	spherical	YCB
Plastic lemon	spherical	YCB
Washer 51 mm	disk, lateral	YCB
Starkist Tuna Fish can	disk	YCB
Amicelli	disk, pinch	KIT
Wine glass	disk, pinch	YCB
Blue Salt Cube	pinch	KIT
Hot Pot	pinch	KIT
Small marker	pinch, lateral	YCB
Dice	pinch, lateral	YCB
Bolt	pinch	YCB
Nail big	pinch	YCB
Coloured Wood Block	pinch, lateral	YCB
Bowl	lateral	YCB
Large marker	lateral	YCB
Credit Card blank	lateral	YCB
Nut	lateral	YCB

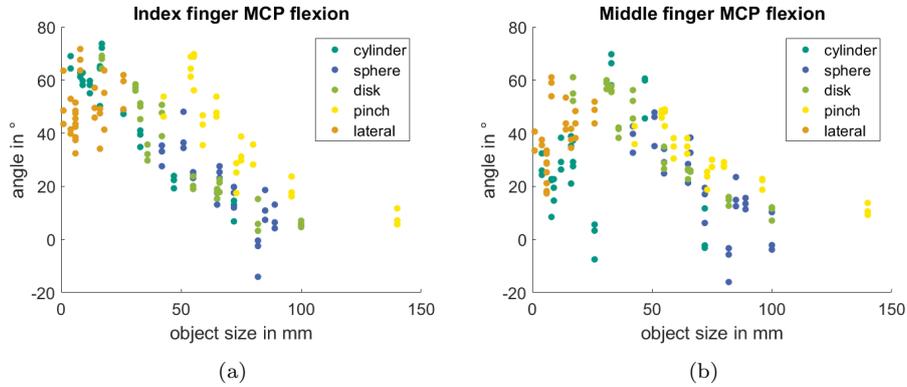
8 *Starke et al.*

Fig. 3: Interdependency between the object size and finger flexion in the metacarpophalangeal joints for all five grasp types and the index (a) and middle finger (b)

the demonstrations on the same object by all subjects stays below 5° for all long fingers, demonstrating the complexity in the performance of a reliable calibration of thumb base joint motions by the higher overall variability in this finger.

Having a close look at the interdependency of human grasps compared to the size of the object they were performed on, a strong correlation becomes apparent in most joints. In finger flexion joints of the index and middle finger, a negative correlation is strongly notable as depicted in Fig. 3. Over all subjects and objects this is proven by a Pearson's Correlation Coefficient (PCC) of -0.52 for the metacarpophalangeal joint of the index finger and -0.31 for the middle finger.

3. Synergy Grasp Representation

To develop a synergy representation of human grasping postures, the hand's parameter space containing 16 joint angles for our representation needs to be significantly reduced while preserving crucial grasp information in an understandable manner. As demonstrated by Romero et al., this can be notably improved by nonlinear dimensionality reduction techniques, since the underlying human grasps exhibit a strongly non-linear behaviour⁸. We apply an autoencoder network for synergy extraction due to its ability to encode complex, non-linear correlations while offering a high flexibility to influence the underlying reasoning of the dimensionality reduction. The flexibility of such networks allows the elaborate design of synergies inherently containing intuitive information on high-level grasp parameters. In addition, the decoder automatically offers an independent tool to generate new synthetic grasp postures directly derived from human grasping behaviour. The autoencoder is implemented by a network comprising five fully connected layers. A three-fold loss function ensures an ordered latent space and makes use of parallel execution of the encoder part. The detailed design of the applied autoencoder network and the

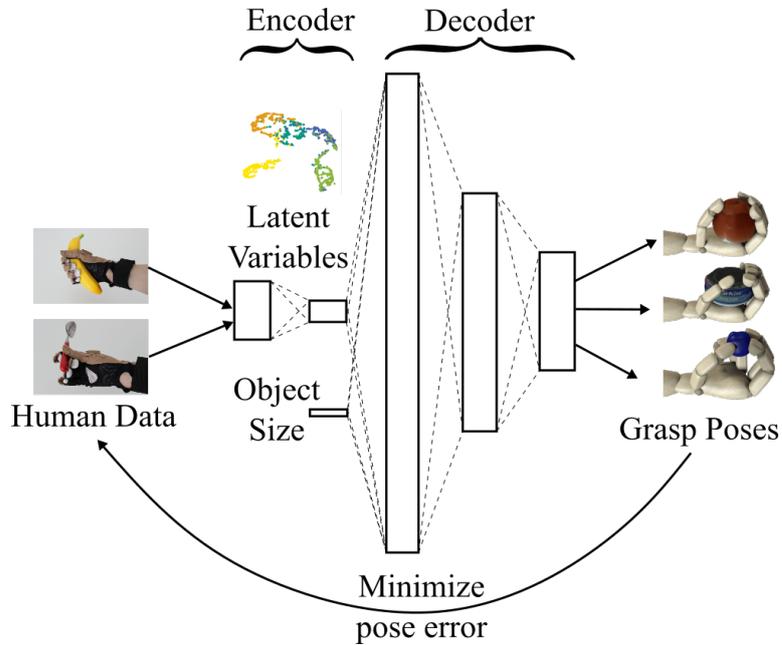


Fig. 4: Architecture of the proposed deep autoencoder network

choice of used hyperparameters are described in the following.

3.1. Autoencoder Design

The structure of our autoencoder is depicted in Fig. 4. The encoder consists of two fully connected layers with eight and three neurons respectively. The latter represent the latent synergy space thereby providing a low-dimensional description of the trained human grasps. The decoder is given more flexibility by allocating three fully connected layers of decreasing size including 64, 32 and 16 neurons respectively. The 16-dimensional output space describing a humanoid grasp pose can therefore be accurately shaped based on the chosen synergy values and allow for adjustment according to the size of the object based on an additional input parameter.

While we started with two synergy parameters according to the findings by Santello et al., an increase to three latent variables proved to be necessary to enforce the additional restructuring of the synergies accounting for high-level meta-information on the acquired grasp type. In addition, this third latent parameter helps to describe the fine-granular final pose adjustment necessary for the fingers to adapt to the specific object surface.

A hyperbolic tangent activation function is applied. While we tested different activation functions including a sigmoid function, a hyperbolic tangent function and a rectified linear unit function (RELU), the hyperbolic tangent provided the best

Table 4: Summary of used symbols and values

Symbol	Description	Dimensionality
\mathbf{y}	real data sample	16
$\hat{\mathbf{y}}$	decoded data sample	16
\mathbf{a}	latent representation	3
$\text{Enc}(\cdot)$	encoder network	
g, h	grasp type	
$g_i, g_j \in g$	different samples from type g	
$\alpha = 1.0$	weighting parameter	1
$\beta = 0.5$	weighting parameter	1
$\gamma = -0.15$	weighting parameter	1

results for this problem.

3.2. Loss Function and Training

To ensure a precise reproduction of grasps learned from human demonstration, we apply the conventional autoencoder loss function represented by the Mean Squared Error (MSE) calculating the difference between the input y and the corresponding output \hat{y} .

$$\text{MSE}(\mathbf{a}, \mathbf{b}) = \frac{1}{N} \sum_{i=1}^N (a_i - b_i)^2 \quad (1)$$

$$\mathcal{L} = \text{MSE}(\mathbf{y}, \hat{\mathbf{y}}) \quad (2)$$

A summary of all symbols used within the description of the loss term is listed in Table 4. However, to additionally enforce an intuitive encoding of the applied grasp type in our synergy representation, we foster an aggregation of grasps with the same type within the latent representation calculated as

$$\mathbf{a} = \text{Enc}(\mathbf{y}), \quad (3)$$

with \mathbf{a} being the latent representation resulting from the execution of the encoder network on input \mathbf{y} . The aggregation is done by adding a second term to the autoencoder loss function penalizing the distance between two grasps of similar types in the latent representation

$$\mathcal{L}_{\text{similar}} = \text{MSE}(\text{Enc}(\mathbf{y}_{g_i}), \text{Enc}(\mathbf{y}_{g_j})) \quad (4)$$

By the same means, a separation of grasps with different types is promoted. Quantified by the distance between two samples of different grasp types, this is described by a comparable term

$$\mathcal{L}_{\text{different}} = \text{MSE}(\text{Enc}(\mathbf{y}_{g_i}), \text{Enc}(\mathbf{y}_{h_i \notin g})) \quad (5)$$

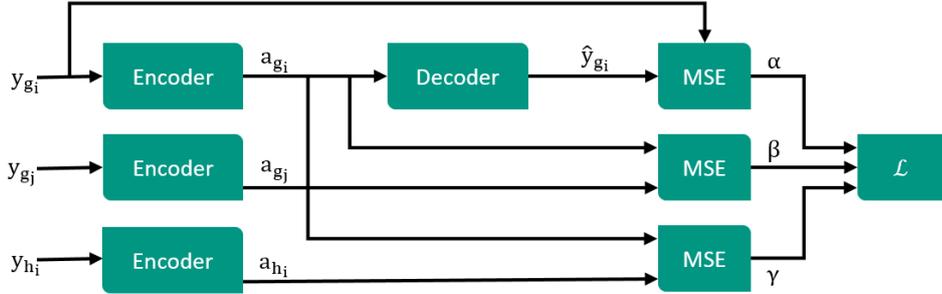


Fig. 5: Loss function for training the deep autoencoder network, taking into account grasp reproduction, aggregation of similar grasps and separation of different grasp types

Altogether, the loss function is comprised of three individual optimization criteria, thereby taking into account three executions of the encoder throughout each training step. All three loss terms are combined with weighting factors empirically determined according to Table 4. The complete loss \mathcal{L} applied for training of the proposed deep autoencoder network results to

$$\mathcal{L} = \alpha \cdot \text{MSE}(\mathbf{y}_{g_i}, \hat{\mathbf{y}}_{g_i}) + \beta \cdot \mathcal{L}_{\text{similar}} + \gamma \cdot \mathcal{L}_{\text{different}}. \quad (6)$$

The choice of weighting parameters according to Table 4 is based on a thorough empirical evaluation. The reproduction of human-like grasps according to the given demonstrations is emphasized by choosing a high α . The clustering of grasps in latent space as an additional functionality is promoted by enlarging β , while its value is capped to preserve the predominance of functional grasp reproduction. The segregation of distinct grasp types in latent space caused by γ is maximized within the range, where a dominance of the basic grasp reproduction is still guaranteed. This ensures the general validity of the resulting synthetically generated grasps, while still allowing a sufficient separation to enable an intuitive latent encoding. A visualization of the complete loss term is presented in Fig. 5.

For training an Adam optimizer is used³¹. As we aim to define a comprising, general synergy description enabling the generation of human-like grasps, we focus on the broad representation of given grasp demonstrations covering as much of the synergy space as possible with sensible, seen grasps. Experiments with variational autoencoders showed that they were able to encode grasping postures in a low-dimensional space but could not include additional meta-information of the given grasps in an intuitive, readable manner. Therefore, we opted for a conventional autoencoder design while adding noise to the latent grasp representation. This normally distributed blur enlarges the space a single human demonstration is mapped to in the synergy representation enforcing a fluent, comprising encoding in the latent space.

In addition, the decoder is provided with the diameter of the grasped object

as described in subsection 3.1. However, it has to be noted that the finger flexion angles required to stably grasp an object do not only depend on the diameter of the object itself, but also on the hand's kinematics. As our study contains subjects with a wide range of hand sizes, the joint angles for the same object vary notably according to the lengths of the subject's finger phalanges. To consider this factor in the proposed grasp generation, the decoder is fed with a parameter describing the object diameter normalized over the size of the hand. By these means, the resulting joint angles can be directly related to an object-to-hand ratio independent of the individual finger sizing compared to the object surface.

The data was split into training and test set by a proportion of 90 % to 10 %. The high proportion of training data is chosen to ensure a thorough coverage of distinct possible grasp postures within the same grasp type. By performing cross-validation, we ensured the network was tested on all available data.

4. Evaluation and Insights

The non-linear nature of the synergies derived by the presented neural network shows a complex interaction between the complicated system of joints contained in the human hand. Nevertheless, their interpretation gives interesting insights into the way humans grasp and allows to transfer knowledge inherently contained in human grasp poses to robotic applications. Here, we show a thorough evaluation of the presented autoencoder as well as an evaluation of grasps generated with the synergy representation. In addition, the deduced synergies and the resulting implications are discussed in detail.

4.1. Autoencoder Validation

As the PCA is still the most applied method to extract grasp synergies, we evaluate the grasps reproduced by our autoencoder against the first two principal components extracted from PCA. To enable a fair comparison of both methods, an autoencoder comprising a two-dimensional latent space is implemented. Its reproduction error as depicted in Fig. 6a already proves the advantages of this non-linear, continuous dimensionality reduction for the given problem, as the autoencoder outperforms PCA on the same set of data.

The most important impact on reproduction quality, however, is achieved by adding information on the object diameter as a scalar to the decoder. This additionally lowers the overall reproduction error by 26 %. While these two-dimensional autoencoders perform better than a linear synergy extraction, they still do not allow for a discernible representation of high-level grasp information within the latent space apart from the object size. Therefore, a third parameter was added to the synergy space.

As shown in Fig. 6b, this additional parameter allows the spatial separation of all five grasp types in latent space by the methods presented in section 3. While the synergy representation exhibits a clear separation of individual grasp types, a

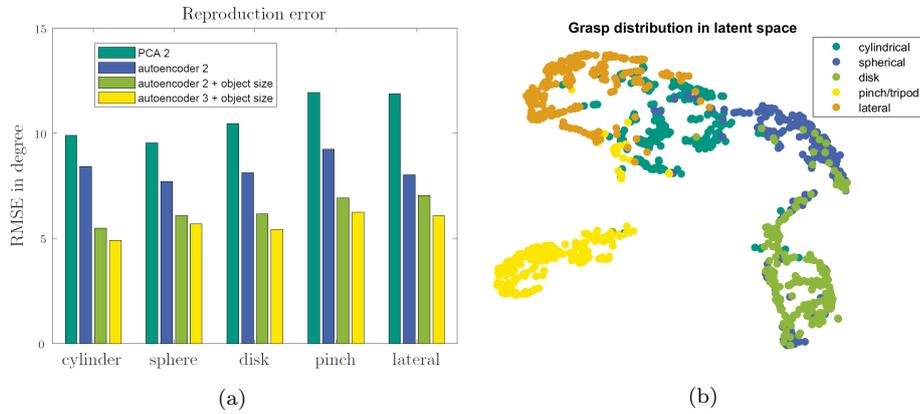


Fig. 6: Autoencoder evaluation in terms of the reproduction error compared to PCA (a) and the latent representation (b) visualized in two dimensions calculated by *t-Distributed Stochastic Neighbor Embedding* (tSNE)

meaningful arrangement of the grasp clusters within the space is also notable. Thus grasps with related configurations in a subset of joints are positioned close together while fundamental differences result in a wide distance to other grasp clusters. Spherical and disk grasp, which mainly differ in the flexion of metacarpophalangeal finger joints, transition fluently in the latent space while the lateral grasp following an essentially different grasp strategy is positioned farther away from all other clusters.

4.2. Quality of Generated Grasps

Considering the relation between finger flexion and object size, a strong dependency is mainly notable in the flexion joints with a negative PCC. The similarity of this measure in the human grasping data proves the successful reproduction of grasps for arbitrarily sized objects. The PCC of human grasp demonstrations is positive for the adduction joints of the middle and little fingers with PCCs of 0.28 and 0.10 respectively, which is also reflected in the generated grasps with a PCC of 0.42 and 0.56 respectively. Interestingly, the human ring finger adduction's PCC is negative with -0.22 while the presented decoder expects a positive correlation of 0.55 similar to the aforementioned fingers.

The general capability of the presented decoder to generate grasps for a given object size is clearly notable. However, the hand aperture is still subject to small deviations due to different hand kinematics and uneven objects surfaces both in the human grasp examples and the target objects to be grasped. Complex object shapes allow only a rough definition of the correct object diameter and while the subject's hand length is already taken into account, their individual hand kinematics including finger lengths and palm sizes are not. Thereby the resulting hand aperture

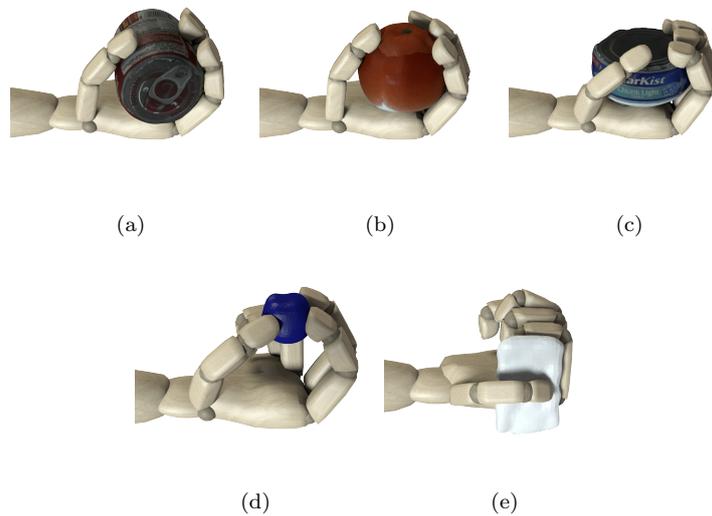


Fig. 7: Grasps generated from the synergy representation: cylindrical grasp on a tomato soup (a), spherical grasp on an orange (b), disk grasp on a tuna can (c), tripod grasp on a wooden cube (d) and lateral grasp on a credit card (e)

described by an angular finger configuration slightly differs compared to a unified human hand model.

To overcome those slight deviations in hand aperture for generated grasps and make their execution more robust regarding prerequisites like the relative hand-to-object pose or friction conditions, we combine our approach for human-like grasp generation with the concept of soft synergies⁷. By grasping 24 objects in simulation, we prove the quality of grasps generated with our approach. The grasps are applied on the human hand model contained within the *Master Motor Map*³⁰, which is scaled to a body height of 1.7 m and a weight of 70 kg. A grasp is generated by the decoder for a target object by providing a latent sample positioned one standard deviation around the mean of the desired grasp type. The relative object size provided to the decoder is calculated by dividing the object diameter by the hand length of the scaled reference model being 180 mm. This grasp is applied to the hand with the finger angles opened by 10 % to perform a preshape around the simulated object. By applying a soft synergy approach with the grasp as attracting synergy configuration the object placed in the middle of the palm is grasped, making contacts between all fingers and the object surface. For the lateral grasp type, the object is positioned on the inner lateral side of the palm and above the index finger. Exemplary grasps formed by this procedure are depicted in Fig. 7.

The quality of the accomplished grasps is evaluated by calculating the mean ϵ -metric over 50 grasps with the hand pose perturbed in a range of 10 mm and

10° . The ϵ -metric describes the radius of the largest sphere $S(\epsilon)$ positioned at the object's center of mass that can be fitted into the convex hull of all grasp contact wrenches W ^{32,33}.

$$\epsilon_{GWS}(W) = \max_{\epsilon} [S(\epsilon) \subseteq \text{convexhull}(W)] \quad (7)$$

To get a stable, reproducible measure for grasp quality, we calculate the mean ϵ for 50 perturbed grasp poses thereby diminishing the influence of small disturbances in the object's mesh. For each object ten grasps are individually generated by sampling the latent input for the decoder within one standard deviation around the mean for the desired grasp type. The calculated grasp quality thereby results in

$$\epsilon_{mean} = \frac{1}{500} \sum_j^{10} \sum_i^{50} \epsilon_{GWS_{i,j}}. \quad (8)$$

Simulation of grasp and grasp quality calculation including the described perturbation is executed with the tools of the grasp planner *Simox*³⁴. The yielded grasp quality for the simulated 270 grasps is presented in Fig. 8 (b). While cylindrical, spherical and disk grasps all range in a stability of $\epsilon = 0.3 - 0.5$, the evaluation of pinch grasps reveals a median of 0.15. This is mainly due to pinch grasps relying on less contact points and therefore being much more dependent on a good positioning of the object. The pinch and tripod grasps performed on the centrally positioned object achieve grasp qualities of up to 0.40. In Fig. 8 (a), a comparison of the perturbed ϵ -metric for three known and three similar objects, which were not included in the presented grasping study is shown. Apart from the grasp type it can be seen that the object size also seems to be an important parameter for grasp success. At the same time, the model copes well with unknown, but similar objects. The general comparison over all grasp types reveals an overall force closure rate including all perturbed grasp poses of 86.9% with a mean of $\epsilon = 0.35$.

4.3. Control Implications

Considering the three synergies resulting from the presented latent representation, overall implications for the mutual actuation of the human joints can be derived. While every synergy is controlling all degrees of freedom, there is a notable diversity in the influence the synergies have on each joint. As depicted in Fig. 9, the first extracted synergy is mainly controlling all metacarpophalangeal joints of the long fingers, while little influence is notable on the proximal interphalangeal joints. The opposite applies to the third synergy, which mainly affects PIP joints. Interestingly, in both synergies a strong correspondence between the index, middle and little finger is notable, whereas the metacarpophalangeal joint of the little finger behaves in a noticeably distinct manner.

16 *Starke et al.*

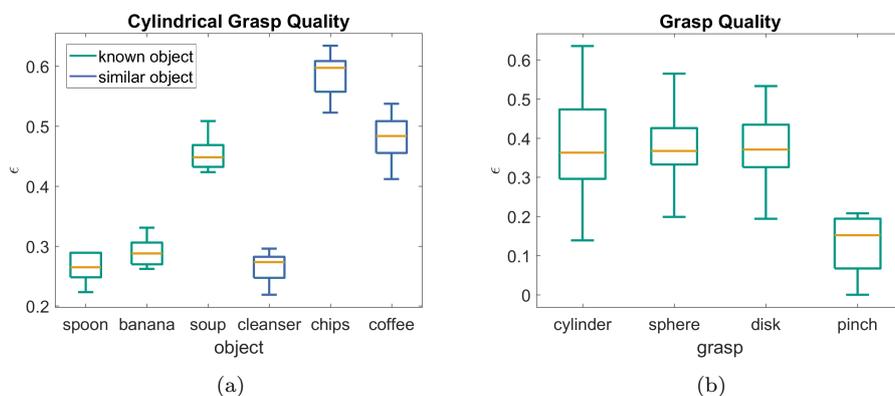


Fig. 8: ϵ -metric over perturbed poses for several cylindrical objects (a) and the mean for each grasp type (b)

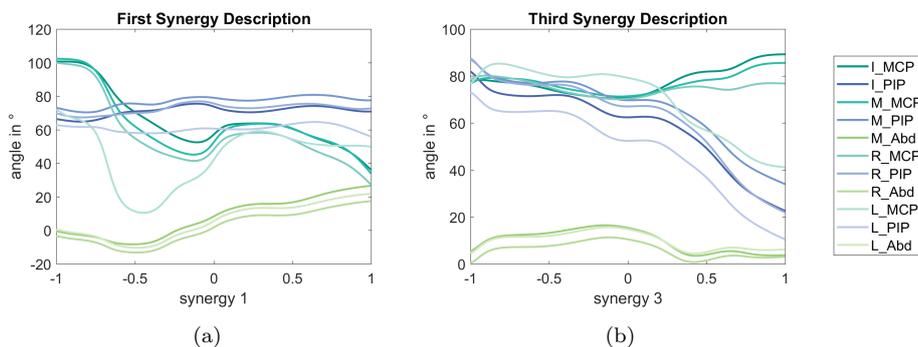


Fig. 9: Joint angle configuration for the four long fingers sampled along the parametric range of the first and third synergy

In addition, the first synergy is the only one allowing to fully open the thumb interphalangeal joint, thereby enabling direct thumb opposition in pinch grasps. The second synergy is mainly responsible for controlling the adduction joints of fingers and thumb. The fine granular control of interphalangeal finger joints is enabled by the third synergy. It highlights the similarities of cylindrical and spherical grasps in finger flexion and reveals the main difference between spherical and disk grasps, which are clearly detached in this third dimension of the latent space. Overall, the extracted synergies demonstrate, that static grasping postures are less dependent on individual finger motions and rely more on the relative deflection of similar joints over all fingers.

5. Conclusion

In this paper we presented an integrated approach to synthesize new human-like grasp configurations given an initial set of observed human grasps. The grasp synthesis process is learned from human grasping data and allows the generation of a distribution of grasp configurations given a grasp aperture and the grasp type.

To acquire the necessary grasping data we performed a study of human grasps, which is publicly available on the *KIT Whole Body Human Motion Database*. The acquired dataset is comprised of 2250 grasps, performed by 15 subjects grasping 35 diverse objects.

Inspired by the linear dimensionality reduction concept of grasp synergies extracted by a PCA⁴, we define a non-linear synergy representation. We train an autoencoder implemented as a deep neural network. The encoder maps the human grasp demonstrations to a 3D latent space. The decoder is presented with the latent representation of the grasp and the diameter of the target object. This constrained autoencoder architecture has several key advantages over the PCA approach. First, it allows the separation of object size and grasp encoding. Second, by introducing additional terms to the loss function, discrete clustering of different grasp types in the latent space can be enforced. Third, the presented approach outperforms the PCA in terms of the reproduction error by 26%, as shown in the evaluation. Finally the autoencoder approach allowed us to synthesize new grasp configurations with defined grasp type and object size. To this end, we first used the encoder to map all grasp configurations to the latent space, while preserving the associated grasp type. This allowed us to obtain a statistical description in terms of mean and standard deviation for each grasp type. By sampling from this latent distribution, the decoder can synthesize new grasp configurations conforming to the given grasp type and aperture.

The evaluation showed that the synthesized grasps are stable in 86.5% of the tested cases evaluated under pose uncertainty. The grasp configurations are applied to a given object using a soft synergy approach. In particular cylindrical, spherical and disk grasps show a high robustness to perturbations of the hand pose.

In future work we want to extend our approach by a temporal dimension to encode not only the final grasp configuration but also the entire grasping sequence, including the hand position and orientation. In addition we want to transfer the acquired grasp model to humanoid robotic and prosthetic hands.

Acknowledgements

This work has been supported by the German Federal Ministry of Education and Research (BMBF) under the project INOPRO (16SV7665).

References

1. B. León, A. Morales, and J. Sancho-Bru, “The Model of the Human Hand,” in *From Robot to Human Grasping Simulation*, vol. 19, ch. 5, pp. 123–173, Springer International Publishing, 2014.
2. O. Terlemez, S. Ulbrich, C. Mandery, M. Do, N. Vahrenkamp, and T. Asfour, “Master motor map (mmm) - framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots,” in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, (Madrid, Spain), pp. 894–901, November 2014.
3. 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, pp. 796–809, August 2016.
4. M. Santello, M. Flanders, and J. F. Soechting, “Postural Hand Synergies for Tool Use,” *The Journal of Neuroscience*, vol. 18, no. 23, pp. 10105–10115, 1998.
5. 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.
6. C. Castellini and P. Van Der Smagt, “Evidence of muscle synergies during human grasping,” *Biological Cybernetics*, vol. 107, no. 2, pp. 233–245, 2013.
7. A. Bicchi, M. Gabbicini, 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.
8. 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.
9. T. Chen, M. Haas-Heger, and M. Ciocarlie, “Underactuated Hand Design Using Mechanically Realizable Manifolds,” in *IEEE Int. Conf. on Robotics and Automation*, pp. 7392–7398, 2018.
10. 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.
11. T. Wimböck, J. Reinecke, and M. Chalon, “Derivation and Verification of Synergy Coordinates for the DLR Hand Arm System,” in *IEEE Int. Conf. on Automation Science and Engineering*, pp. 454–460, 2012.
12. M. Gabbicini, A. Bicchi, D. Prattichizzo, and M. Malvezzi, “On the Role of Hand Synergies in the Optimal Choice of Grasping Forces,” *Autonomous Robots*, vol. 31, no. 2-3, pp. 235–252, 2011.
13. 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*, pp. 2877–2882, 2007.
14. J. B. Rosmarin and H. H. Asada, “Synergistic Design of a Humanoid Hand with Hybrid DC motor-SMA Array Actuators Embedded in the Palm,” in *IEEE Int. Conf. on Robotics and Automation*, pp. 773–778, 2008.
15. M. G. Catalano, G. Grioli, A. Serio, E. Farnioli, C. Piazza, and A. Bicchi, “Adaptive Synergies for a Humanoid Robot Hand,” in *IEEE-RAS Int. Conf. on Humanoid Robots*, pp. 7–14, 2012.
16. 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.
17. 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.
18. P. Weiner, J. Starke, F. Hundhausen, J. Beil, and T. Asfour, “The kit prosthetic hand: Design and control,” in *IEEE/RSJ International Conference on Intelligent Robots and*

- Systems (IROS)*, (Madrid, Spain), pp. 3328–3334, October 2018.
19. M. Gabiccini, E. Farnioli, and A. Bicchi, “Grasp and Manipulation Analysis for Synergistic Underactuated Hands Under General Loading Conditions,” in *IEEE Int. Conf. on Robotics and Automation*, pp. 2836–2842, 2012.
 20. D. Prattichizzo, M. Malvezzi, M. Gabiccini, and A. Bicchi, “On Motion and Force Controllability of Precision Grasps with Hands Actuated by Soft Synergies,” *IEEE Transactions on Robotics*, vol. 29, no. 6, pp. 1440–1456, 2013.
 21. M. Santello and J. F. Soechting, “Force synergies for multifingered grasping,” *Experimental Brain Research*, vol. 133, no. 4, pp. 457–467, 2000.
 22. 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.
 23. A. Kasper, Z. Xue, and R. Dillmann, “The KIT object models database: An object model database for object recognition, localization and manipulation in service robotics,” *Int. Journal of Robotics Research*, vol. 31, no. 8, pp. 927–934, 2012.
 24. B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. Dollar, “Benchmarking in Manipulation Research : The YCB Object and Model Set and Benchmarking Protocols,” *IEEE Robotics & Automation Magazine*, vol. 22, no. 3, pp. 184–185, 2015.
 25. J. Starke, C. Eichmann, S. Ottenhaus, and T. Asfour, “Synergy-based, data-driven generation of object-specific grasps for anthropomorphic hands,” in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, (Beijing, China), pp. 327–333, November 2018.
 26. V. Gracia-Ibáñez, M. Vergara, J. H. Buffi, W. M. Murray, and J. L. Sancho-Bru, “Across-subject calibration of an instrumented glove to measure hand movement for clinical purposes,” *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 19, no. August, pp. 1–11, 2017.
 27. M. R. Cutkosky, “On Grasp Choice, Grasp Models, and the Design of Hands for Manufacturing Tasks,” *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 269–279, 1989.
 28. I. Bullock, J. Z. Zheng, S. De La Rosa, C. Guertler, and A. M. Dollar, “Grasp Frequency and Usage in Daily Household and Machine Shop Tasks,” *IEEE Transactions on Haptics*, vol. 6, no. 3, pp. 296–308, 2013.
 29. M. Vergara, J. L. Sancho-Bru, V. Gracia-Ibáñez, and A. Pérez-González, “An introductory study of common grasps used by adults during performance of activities of daily living,” *Journal of Hand Therapy*, vol. 27, no. 3, pp. 225–234, 2014.
 30. C. Mandery, O. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, “The KIT whole-body human motion database,” in *International Conference on Advanced Robotics (ICAR)*, (Istanbul, Turkey), pp. 329–336, July 2015.
 31. D. P. Kingma and J. L. Ba, “Adam: A Method for Stochastic Optimization,” 2014.
 32. C. Ferrari and J. Canny, “Planning Optimal Grasps,” *1992 IEEE International Conference on Robotics and Automation*, pp. 2290–2295, 1992.
 33. J. Weisz and P. K. Allen, “Pose Error Robust Grasping from Contact Wrench Space Metrics,” *2012 IEEE International Conference on Robotics and Automation*, pp. 557–562, 2012.
 34. N. Vahrenkamp, M. Kröhnert, S. Ulbrich, T. Asfour, G. Metta, R. Dillmann, and G. Sandini, “Simox: A robotics toolbox for simulation, motion and grasp planning,” in *International Conference on Intelligent Autonomous Systems (IAS)*, pp. 585–594., 2012.



Julia Starke received her M.Sc. degree in mechatronic engineering from the Leibniz University Hannover, Germany, in 2016. She is currently a Ph.D. student at the Institute for Anthropomatics and Robotics at Karlsruhe Institute of Technology, Germany, and a member of the High Performance Humanoid Technologies Lab. Her major research interests include human-inspired grasp control and hand prosthetics.



Christian Eichmann is studying towards his M.Sc. degree in Computer Science at Karlsruhe Institute of Technology, Germany. He specializes in the fields of Robotics and Automation. His research interests include human-inspired grasp control and the application of service robotics in palliative care.



Simon Ottenhaus received his diploma degree in electrical engineering from the University of Stuttgart, in 2014. He is currently a Ph.D. student at the Institute for Anthropomatics and Robotics at KIT where he is a member of the High Performance Humanoid Technologies lab. His major research interests are haptic exploration for grasping and data-driven grasping of unknown objects.



Tamim Asfour received the Diploma degree in electrical engineering in 1994 and the Ph.D. degree in computer science in 2003, both from the University of Karlsruhe, Germany. Currently, he is full professor at the Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT) where he is head of the High Performance Humanoid Technologies Lab.

His current research interest is high performance 24/7 humanoid robotics which focus on engineering humanoid robots that can act and interact in the real world, learn from human observation and experience. He is developer of the ARMAR humanoid robot family.