Human motion primitive discovery and recognition

Marta Sanzari, Valsamis Ntouskos, Simone Grazioso, Francesco Puja, Fiora Pirri

Abstract—We present a novel framework for the automatic discovery and recognition of human motion primitives from motion capture data. Human motion primitives are discovered by optimizing the “motion flux”, a quantity which depends on the motion of a group of skeletal joints. Models of each primitive category are computed via non-parametric Bayes methods and recognition is performed based on their geometric properties. A normalization of the primitives is proposed in order to make them invariant with respect to anatomical variations and data sampling rate. Using our framework we build a publicly available dataset of human motion primitives based on motion capture sequences taken from well-known datasets. We expect that our framework, by providing an objective way for discovering and categorizing human motion, will be a useful tool in numerous research fields related to Robotics including human inspired motion generation, learning by demonstration, and intuitive human-robot interaction.

I. INTRODUCTION

Human motion primitives are those movements that span an interval of time in which a change in position of a limb or body part takes place. Examples are rotate the shoulder, turn right the head, lean forward the torso, flex the knee.

Human motion underlying every natural activity shows an extraordinary variation of movements, which are also not easy to denote. Gestures to interact with others, primitive actions and objects manipulation are all composed of a number of primitive motions whose kinematics is restrained by the limbs articulation and body structure. This tremendous variety of human movements has led the research community to study human motion under very different purposes, including tracking [1], [2], temporal segmentation [3], [4], [5], motion parsing [6], articulated motion discovery [7], [8], [9], motion pattern inference [10], primitive interactions [11] and in many other fields, especially in robotics [12], [13], [14] and neurophysiology [15], [16].

Still to the best of our knowledge just quite few works, among which we recall [17], [18], [19], have directly faced the problem of discovering motion primitives in video or motion capture (MoCap) sequences and quantitatively evaluate the ability to recognize them.

The difficulty in treating motion primitive generation lies on the lack of ground truth or exemplars against which quantitatively establishing the extent of a motion primitive. Likewise the difficulty to recognize the difference between action primitives, which are stages of actions such as “pick-up”, “walking”, or “running”, and motion primitives is due to the lack of a vocabulary to denote motion primitives.

Yet recognizing human motion primitives is crucial for understanding human activities, for simulating human behavior, modeling robot behavior, and any interaction between humans, computer systems and robots.

In this paper we present three novel contributions to the research on human motion primitives:
1. We introduce a method to discover motion primitives. The method discovers the primitives relying on the variation of the velocity of a group of joints. It proves to give very similar results to manual segmentation.
2. We introduce a hierarchical method for the modeling and recognition of the motion primitives. The method is based on the Dirichlet process mixture model which selects a set of hypotheses for each primitive query, based on the curvature and torsion of the 3D curve of the motion primitives. Further, basing on a set of geometric features, the proposed method chooses the hypothesis that is more likely, finding the curve minimizing a suitable cost function.
3. We have created a new dataset of human motion primitives from three widely-used public MoCap databases ([20], [21], [22]).
4. We provide a detailed evaluation of our method, presenting quantitative results on how our method performs both with respect to motion primitive discovery and recognition.

The paper is organized as follows. In the next section we explore related work. In Section III-A we discuss notation and preliminaries. Section III-B introduces motion primitives discovery following our definition based on the concept of maximal flux. In Section III-C we introduce the normalization applied to motion primitives, the features used and how models for each primitive category are obtained using Dirichlet process mixture models. We also introduce a cost function based on the primitive geometric properties and discuss how recognition of a query primitive is achieved. Section IV presents an experimental evaluation of primitive discovery and recognition together and introduces the motion primitives dataset. Finally, conclusions and future work are discussed in Section V.

II. RELATED WORK

While human motion in its generality is a vast research area, the paradigm of motion primitives has mainly been explored from the point of view of action primitives and temporal segmentation of actions. Many approaches have explored video sequences segmentation to align similar action behaviors [4]. As a matter of fact, many of the earliest more relevant approaches share the paradigm that understanding
human motion requires view independent representations and that a fine-grained analysis of the motion field is paramount to identify primitives of motion. In early days this required a massive effort in visual analysis to obtain the poses, the low-level features, and segmentation. Nowadays, scientific and technological advances have made it possible to exploit several methods to measure human motion, such as the availability of a number of MoCap databases [20], [22], [7]. Furthermore, recent findings result in methods that can deliver 3D human poses from videos if not even from single frames [23], [24], [25].

Many research areas have taken advantage from the use of motion primitives, in robotics for learning by imitation studies [26], [27], [28] or learning task specifications [29] where primitives are analyzed from a dynamical point of view, or represented as hidden Markov models [40], [41], [32]. In Neurophysiology by [33], [15], [16], [34], [35], [36], where the idea that kinetic energy and muscular activity are optimized in order to conserve energy is commonly employed.

Our view on motion primitive shares this hypothesis of energy minimality during motion, likewise the idea to characterize movements using the proper geometric properties of the skeleton joints space motion. However, for primitive discovery, we go beyond these approaches capturing the variation of the velocity of a group of joints using this as the baseline for computing the change in motion by maximizing the motion flux.

Besides these works, only [17], [18], [19], [36] have targeted motion primitives, to the best of our knowledge. [17] focuses on 2D primitives for drawing, on the other hand [18] does not consider 3D data and generate the motion field considering Lukas-Kanade optical flow for which Gaussian mixture models are learned. Furthermore, they do not provide quantitative results for motion primitives, but only for action primitives, which makes their method not directly comparable with ours. To the best of our knowledge, only [19], [36] use 3D data and explicitly mentions motion primitives, providing quantitative results. In [19] the authors also consider optical flow to account for the velocity field and focus on the recognition of motion primitives basing on harmonic motion context descriptors. In particular, primitive discovery is contextual to recognition. Since [19] deal only with upper torso gestures we compare with them only the primitives they mention. And we note, finally, that in our method start and end are unknown, and primitive discovery is modeled by motion flux. In [36] the authors achieve motion primitives segmentation from wrist trajectories of sign language gestures following the principles of minimization of jerk and acceleration during motion, focusing on unsupervised segmentation obtained with Bayesian Binning.

III. METHODOLOGY

A. Preliminaries

The human skeleton we consider, shown in Fig. 2, is composed by a set of joints $J = \{j_i\}_{i=1}^J$, with $J=17$. We partition the skeleton into 6 groups, namely $G_m \subset J$, with $m = \{1, \ldots, 6\}$ corresponding to head, left and right arm,
torso, left and right leg, respectively. Each group has one parent joint \( p_m \in G_m \), which is the group joint closest to the root joint \( \mathcal{J} \) according to the skeleton hierarchy.

The pose of the \( j \)-th joint at the \( n \)-th frame of a MoCap sequence is defined as \( T^n_{j} \in SE(3) \). The pose of each group is normalized at each frame of the sequence. More specifically, joints \( j \in G_m \) are transformed first according to the transformation \( T^n_{p_m} \) of their parent at the current frame and then by transformation \( T^n_{p_m} \) of their parent at frame 1. The normalized pose of joint \( j \) at frame \( n \) is:

\[
\hat{T}^n_{j} = (T^1_{\text{root}})^{-1}T^n_{p_m}(T^n_{p_m})^{-1}T^1_{j}, \quad \forall j \in G_m.
\]

The above transformation maps each joint to the root of the initial MoCap frame, denoted as \( T^1_{\text{root}} \).

**B. Motion Primitive Discovery**

Let \( \gamma_j : I \subset \mathbb{R} \mapsto \mathbb{R}^3 \) be the trajectory of the \( j \)-th joint and \( C = \{ \gamma_j(t) | t \in I \} \) the corresponding curve. A smooth trajectory is obtained by cubic spline interpolation over the MoCap data of the most out joint of a given group, according to the skeleton hierarchy. This interpolation provides a continuous-time trajectory with a smooth velocity and continuous acceleration, satisfying natural constraints of human motion.

**Definition III.1.** The motion flux with respect to the time interval \( I = [t_1, t_2] \) is

\[
\Phi(t_2, t_1) = \sum_{j \in G} \int_{t_1}^{t_2} |\dot{\gamma}_j(\tau) \cdot \mathbf{g}| \, d\tau,
\]

with \( G \) a specific group of joints and \( \mathbf{g} \) the unit vector corresponding to the direction of the gravity vector. The motion flux captures the variation of the velocity of a group of joints with respect to the rest pose of the group.

Let \( s_j(t) \) be the arc length function of \( \gamma_j \), for joint \( j \). Given the motion flux for a group of joints, as defined in eq. (2), we discover motion primitives based on an optimization principle. In particular, we define motion primitives as motions corresponding to maximal flux between time instances where the average group velocity tends to be minimal.

**Definition III.2.** Let us assume without loss of generality that \( t_0 \) corresponds to a frame of the MoCap sequence. A motion primitive of a group of joints corresponds to the time interval \( [t_0, \rho] \) which maximizes the following energy-like function:

\[
P(\rho; t_0) \doteq \Phi(\rho, t_0) - \frac{\alpha}{2} \sum_{j \in G} (\|\dot{\gamma}_j(\rho)\|^2 + \|\gamma_j(t_0)\|^2) - \beta \sum_{j \in G} |s_j(\rho) - s_j(t_0)|.
\]

Here \( \alpha \) and \( \beta \) are parameters, which depend on the sampling schema used to collect the joint trajectories, given the varying frame rates of the sequences taken from the chosen MoCap databases ([21], [20], [22]). Given a starting time instant \( t_0 \), a motion primitive is extracted by identifying the time instant \( \rho \), which corresponds to a local maximum of (3). The optimality condition of (3) gives:

\[
\sum_{j \in G} |\dot{\gamma}_j(\rho) \cdot \mathbf{g}| - \alpha \sum_{j \in G} \|\dot{\gamma}_j(\rho)\|^2 - \beta N = 0,
\]

with \( N \) the number of joints in group \( G \). Given the one-dimensional nature of the problem, finding the zeros of (4) and verifying whether they correspond to local maxima of (3) is a trivial task.

A set of primitives is extracted from a sequence by sequentially finding the set of time instances which maximize (3), namely \( T \doteq \{ \rho \in \mathbb{R} | P(\rho; \rho) \to \max, \rho \in T \} \cup \{t_0\} \).

**C. Motion Primitive Recognition**

We describe in the following, primitive normalization, modeling of each motion primitive category, and primitive recognition. For primitive recognition we expect that a certain motion sequence, which we call a primitive motion query, is observed and the task is to recognize which primitive is, among all the known primitives of a specific group.

**Normalization** In order to make the primitives invariant with respect to anatomical variations such as skeleton limb length and data sampling rate, we apply a normalization. More specifically using the notation of Sections III-A and III-B we define a trajectory corresponding to a primitive as \( \gamma_Q : I_Q \subset \mathbb{R} \mapsto \mathbb{R}^3 \) with \( I_Q \) the interval where the primitive is defined. Normalization is achieved by scaling \( \gamma_Q \) by a normalization constant \( k_G \).

We test different normalization constant candidates. Let \( ||\dot{\gamma}_Q|| \) be the average velocity along \( \gamma_Q \). \( \ell_G \) the length of the limb corresponding to group \( G \) and \( A_Q \) the area covered by group \( G \) during its motion. The boxplots in Fig. [3] show the arc length distribution of primitives of a specific category for groups \( G1 \) and \( G4 \). The first box corresponds to the original distribution while the following boxes correspond to the distributions resulting by scaling with \( ||\dot{\gamma}_Q||, ||\dot{\gamma}_Q||/A_Q, 1/\ell_G, \) and \( 1/A_Q \), respectively. The plots suggest that normalization according to \( 1/\ell_G \) gives the best results. We observed that this is consistent across the groups and primitive categories, hence, we chose \( k_G = 1/\ell_G \).

**Model Generation** We use curvature and torsion as features for training models for each motion primitive category, computed by differentiation of the analytical form of the interpolated spline. We choose these features as they suffice
to specify a curve up to a rigid transformation. Additionally, we use the corresponding positions of the curve to avoid fragmentation. For each group of joints we consider only the two most external joints. Hence an observation related to group $G$ is a vector $y_i \in \mathbb{R}^d$, with $d=10$.

Given a dataset of annotated primitives (see Section IV) we use the features described above for training a model for each motion primitive category. Let us assume that for a fixed group of joints the dataset contains $M$ primitive categories. Let $D_k = \{y_{k_1}, \ldots, y_{k_q}\}$ be the observations of the $k$-th category of motion primitives. Hence $y_{k_i} \in \mathbb{R}^d$ is the observed value of a stochastic $d$-dimensional variable $Y_{k_i}$, with $Y_1, \ldots, Y_{Mq}$ conditionally independent and normally distributed multivariate random variables. Here $2 \le Mq \le \sum_{k=1}^M |D_k|$, with $|\cdot|$ indicating cardinality. Namely, we have $M>1$ sets of training data, with each set holding a number $q>2$ of samples.

We make the assumption that the distribution of an unseen feature vector $y_{Mq+1}$ has as reference density an infinite mixture of normals with posterior predictive distribution $\Theta_{Mq+1} \mathbb{P}_{k} \{ D \}$, where $\mathbb{P}_k \{ D \}$ indicates the $s$-th dataset used in estimating $\Theta_{Mq+1}$. In particular, we consider the well known DP mixture (DPM) [37], [38], [39] for each set of data $D_k, k = 1, \ldots, M$. We refer to the model of primitive category $Q$ as $M_Q$ and $c_Q$ denote the components of $M_Q$. Each model generates a number of manifolds gathering segments of the original curves which share the same statistics, as a realization of such a primitive, though collected in different sequences and executed by different subjects. Fig. 4 shows two such manifolds as specific components of two DPM models. Algorithm 1 summarizes the steps of model generation.

**Input:** Annotated motion primitives $Q$ for each group $G_m$  
**Output:** Hierarchical DPM models

```plaintext
foreach $Q$ do
    parametrize $\gamma : I \mapsto \mathbb{R}^3$, for joints $j_1, j_2 \in G_m$
    and positions $x_{j_1:j_2}$ compute $\kappa_{j_1:j_2}$ and $\tau_{j_1:j_2}$
    $y \leftarrow (x_{j_1:j_2}, \kappa_{j_1:j_2}, \tau_{j_1:j_2})^T$
end

foreach $G_m$ do
    foreach $Q$ do
        Build feature matrix $(y_1, \ldots, y_q)^T \in \mathbb{R}^{(q \times 10)}$.
        Estimate DPM $M_Q$ according to [40]
    end
end

foreach $M_Q, c \in M_Q$ do
    Compute ruled surface $R = \frac{n \times n'}{|n \times n'|}$.
    Associate data to $M_Q$
end
return Models $M_Q$
```

**Recognition** For recognition we formulate a query motion primitive from which we can estimate the best model for $y$. Note that, at this level of the hierarchy, we do not want to restrict to a single hypothesis. As the query is generated by maximizing the motion flux there is a certain amount of noise, especially around the start and end of the curve segment. It follows that a better set of hypotheses can be chosen by thresholding the expectation, the higher the threshold the less are the number of hypotheses.

To deal with a set of hypotheses we extend the statistics generated by each component with more structure resorting to additional geometric features. To this end, for all segments of each DPM model component $c$ we compute the tangent $t$, velocity $n$, normal $n$ and binormal $b$ vectors. Based on these vectors, we compute the ruled surface $R = \frac{n \times n'}{|n \times n'|}$ forming a ribbon of tangent planes.

The recognition proceeds as follows. We consider the query as a feature set $D$, and the parametrized curve $\gamma$...
maximizing the flux on a subset of $\mathcal{D}$. Let $\mathbb{H}$ be the set of hypotheses collecting the best fitting models $\mathcal{M}_Q$ for $D_\gamma$, that is, for the feature set restricted to the points in $\gamma$, namely

$$
\mathbb{H} = \{ \mathcal{M}_Q \mid \mu < p(y) = \sum_{Q} \pi_Q f(y|\theta_Q), \theta_Q \in \Theta_Q \},
$$

with $\mu$ a threshold value. The objective is to find, among the hypotheses in $\mathbb{H}$, the one with the highest number of manifold components that are more similar to $\gamma$. Let $y_0$ be the features specifying position in $\gamma$ and let $y_{r_k}$ be the projection of $y_0$ on the ruled surface $R_k$ of the $k$-th curve. Moreover, let $\hat{w}$ be the closest point of $\gamma$ to $y_{r_k}$. We denote the curvature and torsion distances as $\delta_c(y_{r_k}, \hat{w})$ and $\delta_t(y_{r_k}, \hat{w})$, respectively. Additionally, given the Frenet frames at $\gamma$ and at $y_0$ we compute the rotation $R$ between the two frames. Let $F(R) = \text{trace}((I - R)(I - R)^\top)$, with $I$ the identity matrix, the query curve belongs to the motion primitive category, which corresponds to the manifold minimizing the following cost function among all DPM models $\mathcal{M}_Q \in \mathbb{H}$

$$
\mathcal{L}(\gamma) = \sum_{\gamma \in \mathcal{M}_Q} \|y_{r_k} - \hat{w}\| + \delta_c(y_{r_k}, \hat{w}) + \delta_t(y_{r_k}, \hat{w}) + F(R).
$$

Algorithm 2 summarizes the motion primitive recognition.

**Input:** A query as the trajectories of two external joints of group $G$

**Output:** Category of query primitive

**foreach** $\mathcal{M}_Q \in \mathbb{H}$, $c \in \mathcal{M}_Q$ do

Compute $\mathcal{L}$ according to (6) and find the segment of $c \in \mathcal{M}_Q$ minimizing the cost

**end**

**return** Primitive category $\hat{Q}$.

**Algorithm 2:** Recognition of motion primitives

### IV. EXPERIMENTS

In this section we evaluate the accuracy of the motion primitives discovered using the motion flux. We examine also the accuracy of primitive classification based on the DPM models established for each primitive category. The evaluation is performed on MoCap sequences taken from three widely used publicly available MoCap datasets.

**Datasets** The datasets we consider in this work are the Human3.6M dataset (H3.6M) [20], the CMU Graphics Lab MoCap database (CMU) [21] and the KIT Whole-Body Human Motion Database (KIT-WB) [22]. The pose of the joints specified in Fig. 2 are extracted for each frame of the sequences considered as described in section II-A. For KIT-WB the trajectories of the joints are computed from the marker positions taken from the C3D files. Table I shows for each dataset the total number of motion primitives discovered based on the motion flux for each group $G_m$.

**Motion Primitive Discovery** For each MoCap sequence we asked 18 participants to observe a specific limb of the subject performing the activity and manually segment the video sequences into motions they consider as meaningful and not further dividable to shorter motions. The participants were able to freely navigate through the videos and place ‘start’ and ‘stop’ markers for manually defining motion primitives. As a result each group $G_m$ has been segmented by three different participants. Note that the manual labeling of the primitives is independent of the automatic process described in section III-B. The manually segmented primitives, although subjective, form a baseline to assess the accuracy of the automatic primitive discovery based on the motion flux in the lack of a ground truth.

Table II shows the average difference in frames between the manually assigned ‘start’ and ‘stop’ markers and the corresponding endpoints of the automatically discovered primitives. It shows as well the percentage of outliers, namely those endpoints which are distant more than 5 frames from the ones obtained by manual segmentation. The results show that the proposed method discovers action primitives whose endpoints are close to the endpoints of the manually extracted ones.

**TABLE II**

|     | G1 | G2 | G3 | G4 | G5 | G6 |
|-----|----|----|----|----|----|----|
| Avg. difference | 3.1 | 3.8 | 2.6 | 2.9 | 4.7 | 4.6 |
| % of outliers     | 5.9 | 6.2 | 4.8 | 5.0 | 7.1 | 6.9 |

**Motion Primitive Recognition** The manually extracted primitives of each group have been labeled providing the categorization shown in Fig. 7. We split the resulting primitives dataset keeping 20% of the primitives for validation and using the rest for training. More specifically, a DPM model is computed for each set of primitives sharing the same label as described in Section III-C. Using these models we assess the accuracy of primitive recognition based on Algorithm 2 both with respect to the validation set and with respect to the primitives discovered based on the motion flux. The overall accuracy for each group of joints in these two cases are given in Table III while Fig. 5 and Fig. 6 show the corresponding confusion matrices. The results suggest that the DPM classification together with the proposed recognition algorithm capture the main characteristics of each motion primitive category. Table IV presents an ablation study showing the accuracy of primitive recognition based on Algorithm 2.

**TABLE I**

| Number of discovered motion primitives |
|---------------------------------------|
| G1 | G2 | G3 | G4 | G5 | G6 |
|----|----|----|----|----|----|
| H3.6M | 185 | 89 | 203 | 186 | 128 | 112 |
| CMU | 146 | 63 | 203 | 137 | 106 | 74 |
| KIT | 159 | 51 | 148 | 115 | 87 | 63 |

**Comparison** We compare our results with [19], where performance is evaluated for a subset of 4 primitives of the arms (called gestures), namely *Point right*, *Raise arm*, *Clap* and *Wave*. The authors perform two tests, one without noise in the start and end frames of the primitives and one where
Fig. 5. Confusion matrices for the recognition of primitives taken from the validation set of the manually extracted primitives. In order the matrices for G1, G2, and G3 are shown at the top, and G4, G5 and G6 at the bottom.

Fig. 6. Confusion matrices for the recognition of motion primitives discovered based on the motion flux. In order the matrices for G1, G2, and G3 are shown at the top, and G4, G5 and G6 at the bottom.
the primitves are affected by noise. In the noise-free case their overall accuracy is 94.4% while in the presence of noise the accuracy drops to 86.9%. Our results are not directly comparable with the ones of [19] as the definition of the primitives is different and the dataset built by the authors is not publicly available. We notice though that the overall accuracy of our framework for the arms (Groups 3 and 4) is 94.4% for the validation set and 85.1% for the discovered primitives. These results are very similar to those of [19] in the noise-free and noisy cases, respectively.

**Motion Primitives Dataset** The dataset of manually annotated motion primitives extracted from the MoCap sequences of the H3.6M [20], CMU [21] and KIT- \(^\text{\textsuperscript{1}}\) a part of the dataset is public at the moment
The results show that our framework discovers and recognizes motion primitives with high accuracy with respect to the manually defined baseline while providing competitive results with respect to [19] the only work, to the best of our knowledge, providing quantitative results on similarly defined motion primitives.

In the future, we plan to examine the rules of human motion primitive composition and coordination as well as to further extend the dataset of human motion primitives by considering additional MoCap datasets.

REFERENCES

[1] H. Sidenbladh, M. J. Black, and L. Sigal, “Implicit probabilistic models of human motion for synthesis and tracking,” in ECCV, 2002, pp. 784–798.

[2] E.-J. Ong, A. S. Micilotta, R. Bowden, and A. Hilton, “Viewpoint invariant exemplar-based 3D human tracking,” Computer Vision and Image Understanding, vol. 104, no. 2, pp. 178–189, 2006.

[3] F. Zhou, F. D. I. Torre, and J. K. Hodgins, “Hierarchical aligned cluster analysis (haca) for temporal clustering of human motion,” IEEE Trans. Pattern Anal. Machine Intell., vol. 35, no. 3, pp. 582–596, 2013.

[4] D. Gong, G. Medioni, and X. Zhao, “Structured time series analysis for human action segmentation and recognition,” IEEE Trans. Pattern Anal. Machine Intell., vol. 36, no. 7, pp. 1414–1427, 2014.

[5] B. Kruger, A. Vogele, T. Willig, A. Yao, R. Klein, and A. Weber, “Efficient unsupervised temporal segmentation of motion data,” IEEE Trans. Multimedia, 2016.

[6] H. Shen, S.-I. Yu, Y. Yang, D. Meng, and A. Hauptmann, “Unsupervised video adaptation for parsing human motion,” in ECCV, 2014, pp. 347–360.

[7] L. Sigal, A. P. Balan, and M. J. Black, “HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion,” JVC, vol. 87, no. 1, p. 4, 2009.

[8] L. Del Pero, S. Ricco, R. Sukthankar, and V. Ferrari, “Articulated motion discovery using pairs of trajectories,” in CVPR, 2015, pp. 2151–2160.

[9] J. Won and J. Lee, “Shadow theatre: Discovering human motion from a sequence of silhouettes,” ACM Trans. Graph., vol. 35, no. 4, pp. 147:1–147:12, 2016.

[10] X. Zhao and G. Medioni, “Robust unsupervised motion pattern inference from video and applications,” in ICCV, 2011, pp. 715–722.

[11] R. Filipovych and E. Ribeiro, “Recognizing primitive interactions by exploring actor-object states,” in CVPR, 2008, pp. 1–7.

[12] 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 Trans. on Robotics, vol. 32, no. 4, pp. 796–809, 2016.

[13] J. F.-S. Lin, M. Karg, and D. Kulić, “Movement primitive segmentation for human motion modeling: A framework for analysis,” IEEE Trans. on Human-Mach. Syst., vol. 46, no. 3, pp. 325–339, 2016.

[14] W. Takano, H. Imagawa, and Y. Nakamura, “Spatio-temporal structure of human motion primitives and its application to motion prediction,” Robotics and Autonomous Systems, vol. 75, pp. 288–296, 2016.

[15] T. Flash and B. Hochner, “Motor primitives in vertebrates and invertebrates,” Current Opinion in Neurobiology, vol. 15, no. 6, pp. 660–666, 2005.

[16] T. Flash, Y. Meirovitch, and A. Barliya, “Models of human movement: Trajectory planning and inverse kinematics studies,” Robotics and Autonomous Systems, vol. 61, no. 4, pp. 330–339, 2013.

[17] D. D. Vecchio, R. M. Murray, and P. Perona, “Decomposition of human motion into dynamics-based primitives with application to drawing tasks,” Automatica, vol. 39, no. 12, pp. 2085–2098, 2003.

[18] Y. Yang, I. Saleemi, and M. Shah, “Discovering motion primitives for unsupervised grouping and one-shot learning of human actions, gestures, and expressions,” IEEE Trans. Pattern Anal. Machine Intell., vol. 35, no. 7, 2013.

[19] M. B. Holte, T. B. Moeslund, and P. Fihl, “View-invariant gesture recognition using 3D optical flow and harmonic motion context,” Computer Vision and Image Understanding, vol. 114, no. 12, pp. 1353–1361, 2010.

[20] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, “Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments,” IEEE Trans. Pattern Anal. Machine Intell., vol. 36, no. 7, pp. 1325–1339, 2014.

[21] CMU Mocap Database. http://mocap.cs.cmu.edu/.

[22] C. Mandery, O. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, “The kit whole-body human motion database,” in ICAR, 2015, pp. 329–336.

[23] M. Sanzari, V. Ntouskos, and F. Pirri, “Bayesian image based 3d pose estimation,” in ECCV, vol. 8, 2016, pp. 566–582.

[24] X. Zhou, M. Z. Zhu, S. Leonards, K. Derpanis, and K. Daniilidis, “Spariness Meets Deepness: 3D Human Pose Estimation from Monocular Video,” in CVPR, 2016.

[25] D. Tome, C. Russell, and L. Agapito, “Lifting From the Deep: Convolutional 3D Pose Estimation From a Single Image,” in CVPR, 2017.

[26] A. Gans, T. PETRIČ, M. DO, B. NEMEC, J. MORIMOTO, T. ASFOUR, and A. UDE, “Adaptation and coaching of periodic motion primitives through physical and visual interaction,” Robotics and Autonomous Systems, 2016.

[27] P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal, “Learning and generalization of motor skills by learning from demonstration,” in ICRA, 2009, pp. 763–768.

[28] D. Kulić, C. Ott, D. Lee, J. Ishikawa, and Y. Nakamura, “Incremental learning of full body motion primitives and their sequencing through human motion observation,” The Int. J. of Robotics Research, vol. 31, no. 3, pp. 330–345, 2012.

[29] A. L. P. Ureche, K. Umezawa, Y. Nakamura, and A. Billard, “Task parameterization using continuous constraints extracted from human demonstrations,” IEEE Trans. Robot., 2015.

[30] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, “Embodied symbol emergence based on mimesis theory,” The Int. J. of Robotics Research, 2004.

[31] T. Asfour, P. Azad, F. Gyarfas, and R. Dillmann, “Imitation learning of dual-arm manipulation tasks in humanoid robots,” Int. J. of Humanoid Robotics, 2008.

[32] A. G. Billard, S. Calinon, and F. Guenter, “Discriminative and adaptive imitation in uni-manual and bi-manual tasks,” Robotics and Autonomous Systems, 2006.

[33] E. Bizz and F. A. Mussa-Ivaldi, “Toward a neurobiology of coordinate transformations,” in The Cognitive Neurosciences, 1995, pp. 495–506.

[34] P. Viviani and T. Flash, “Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning,” Journal of Experimental Psychology: Human Perception and Performance, vol. 21, no. 1, p. 32, 1995.

[35] T. Flash and A. A. Handzel, “Affine differential geometry analysis of human arm movements,” Biological Cybernetics, vol. 96, no. 6, pp. 577–601, 2007.

[36] D. Endres, Y. Meirovitch, T. Flash, and M. A. Giese, “Segmenting sign language into motor primitives with bayesian bining,” Frontiers in computational neuroscience, vol. 7, 2013.

[37] T. S. Ferguson, “A bayesian analysis of some nonparametric problems,” Ann. Stat., pp. 209–230, 1973.

[38] C. E. Antoniak, “Mixtures of dirichlet processes with applications to bayesian nonparametric problems,” Ann. Stat., pp. 1152–1174, 1974.

[39] R. M. Neal, “Markov chain sampling methods for Dirichlet process mixture models,” J. of Comp. and Graph Stat., vol. 9, no. 2, pp. 249–265, 2000.

[40] S. Jain and R. M. Neal, “A split-merge markov chain monte carlo procedure for the dirichlet process mixture model,” J. of Comp. and Graph. Statistic., 2012.