Representation of Complex Movement Sequences based on Hierarchical Spatio-temporal Correspondence for Imitation Learning in Robotics

W. Ilg[1], G.H. Bakir[2], M.O. Franz[2], M.A. Giese[1]

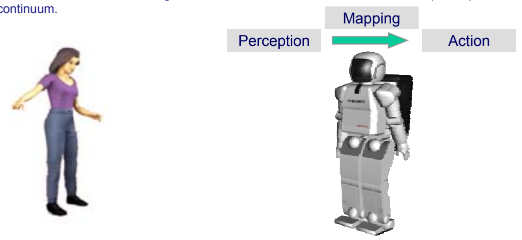
[1] *Cognitive Neurology, University Clinic Tübingen*
[2] *Max Planck Institute for Biological Cybernetics, Tübingen*

Introduction

Imitation learning of complex movements has become a popular topic in neuroscience, as well as in robotics. The goal of Imitation learning in robotics is to teach robots by observation of movement sequences. Imitation learning has to address two fundamental problems concerning the representation of movement sequences.

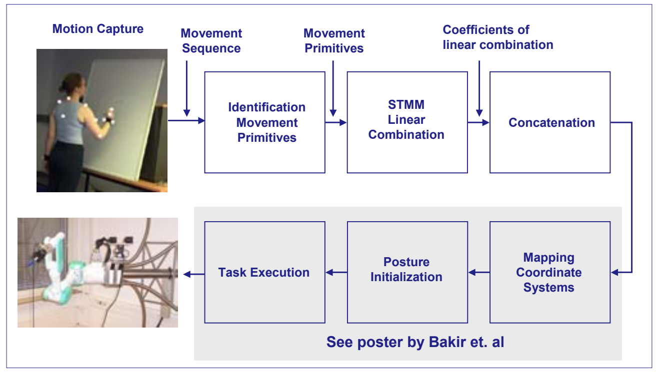
(1) The movement characteristics of observed movements have to be transferred from the perceptual level to the level of generated actions.

(2) The representation should be able to generalize from a small number of learned example sequences to a movement continuum.

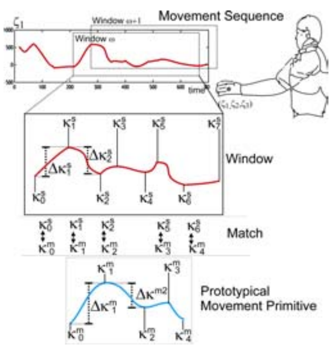


For the transfer of complex movements from perception to action we exploit a learning-based method that represents complex action sequences by linear combination of prototypical examples (Ilg & Giese, 2002). The method of hierarchical spatio-temporal morphable models (HSTMM) decomposes action sequences automatically into movement primitives. These primitives are modeled by linear combinations of a small number of learned example trajectories. The learned spatio-temporal models are suitable for the analysis and synthesis of long action sequences, which consist of movement primitives with varying style parameters.

Imitation Learning of Writing Movements



Identification of Movement Primitives



- Elementary key features κ based on velocity zeros
- Sequence alignment problem
  - Window κ^s_0 κ^s_2 κ^s_3 κ^s_4 κ^s_5 κ^s_6 κ^s_7
  - Template κ^m_0 κ^m_1 κ^m_2 κ^m_3 κ^m_4 κ^m_5
- Robust Identification based dynamic programming

Recursive cost function

$$\begin{aligned} D(i,j)=\min(&D(i-1,j-1)+\|\Delta \kappa_i^s-\Delta \kappa_j^m\|,\\ &D(i-2,j-1)+\|\Delta\kappa_{[i-1,i]}^s-\Delta\kappa_j^m\|,\\ &D(i-3,j-1)+\|\Delta\kappa_{[i-2,i]}^s-\Delta\kappa_j^m\|,\\ &D(i-1,j-2)+\|\Delta\kappa_i^s-\Delta\kappa_{[j-1,j]}^m\|,\\ &D(i-2,j-2)+\|\Delta\kappa_{[i-1,i]}^s-\Delta\kappa_{[j-1,j]}^m\|,\\ &D(i-3,j-2)+\|\Delta\kappa_{[i-2,i]}^s-\Delta\kappa_{[j-1,j]}^m\|) \end{aligned}$$

i and j denote the indices for the key feature κ_i^s respectively κ_i^m. The starting value D(1,1) is given by

$$D(1,1)=\|\Delta \kappa_1^s-\Delta \kappa_1^m\|$$

The resulting difference vector between the two successive key features κ_i^{s-3} and κ_i^s is determined by

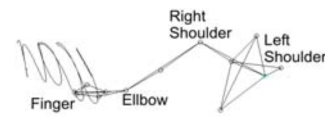
$$\Delta \kappa_{[i-2,i]}^s=\Delta \kappa_{i-2}^s+\Delta \kappa_{i-1}^s+\Delta \kappa_i^s$$

The minimal costs δ of movement primitive m for window ω

$$\delta(m_\omega)=\min_k(D(k,q))$$

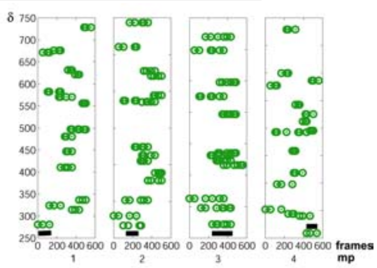
Motion Capture

We recorded writing movements of two human actors who wrote the word ``ICAR'' using a commercial motion capture system (VICON 612, Oxford) with 6 cameras. We used 10 (passive) markers that included the shoulders, 2 front and one rear torso, upper arm, elbow, front arm, hand and index finger of the writing arm.



Identification of Movement Primitives

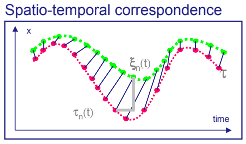
Individual letters were defined as movement primitives. The automatic segmentation of the movement primitives was based on the index finger trajectories. The segmentation algorithm was trained with one example for each movement primitive that was obtained by manual segmentation of the trajectory of one of the actors.



Spatio-Temporal Morphable Models

STMM represent the spatio-temporal characteristics of complex movement sequences by linear combination of example trajectories with different characteristics. Linear combinations of space-time patterns can be defined efficiently by exploiting spatio-temporal correspondence, by weighted summation of spatial and temporal displacement fields that morph the prototypical movement trajectories into a reference pattern.

Spatio-temporal correspondence



Shifts: $x_2(t) = x_1(t') + \xi(t)$

$t' = t + \tau(t)$

Minimize: $$E = \int \left[\left|\mathbf{x}_1(t') - \mathbf{x}_2(t)\right|^2 + \lambda (t-t')^2 \right] dt$$

Synthesis of new movements

By linear combination of spatial and temporal shifts the STMM allows to interpolate smoothly between motion patterns with significantly different spatial structure, but also between patterns that differ with respect to their timing.

[Giese&Poggio2000]

Linear combination of spatial and temporal shifts:

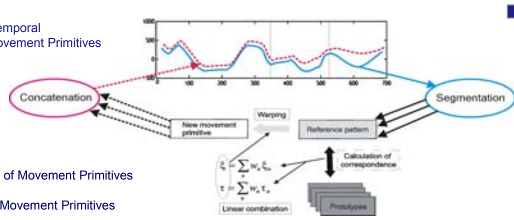
Spatial displacements: $$\xi(t) \cong \sum_{n=1}^{N} \boxed{\omega_n} \xi_n(t)$$

Temporal shifts: $$\tau(t) \cong \sum_{n=1}^{N} \boxed{\omega_n} \tau_n(t)$$

Linear combination

Morphing of Movement Primitives

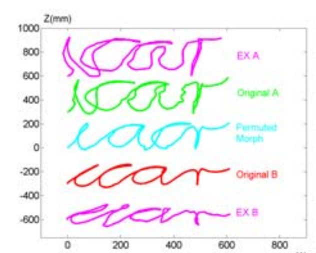
Hierarchical Spatial-Temporal Correspondence of Movement Primitives

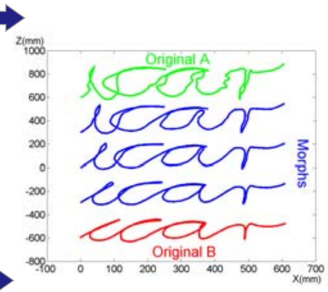


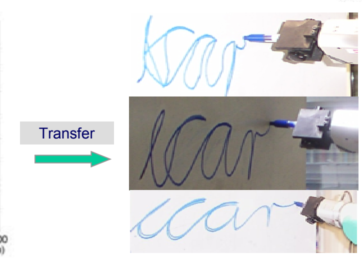
- coarse level
  ⇒ Identification of Movement Primitives
- fine level
  ⇒ Morphing of Movement Primitives

Syntheses of Writing Movements

Continuous spaces of individual movements are generated by linear combinations of the segmented movement primitives. These movements are then automatically concatenated into longer sequences including multiple movement primitives. The method allows to morph continuously between the writing sequences of the two actors. In addition we can synthesize caricatures of the specific writing styles of each actor. Also, the individual movement primitives can be reassembled in a different sequential order, e.g. in order to write the word "'IACR'' . All movement sequences were synthesized based on only two prototypical example trajectories.







Discussion

The learned spatio-temporal models are suitable for the analysis and synthesis of long action sequences, which consist of movement primitives with varying style parameters. The method of HSTMM has the advantage that it works with very small sets of training data (Giese00, Ilg02, Ilg03). Many popular methods for the representation of trajectories, e.g. HMMs or unsupervised learning of manifolds (Fod02, Jenkins02, Brand00) typically require substantial amounts of training data. Another advantage of HSTMMs is the relatively intuitive interpretation of the weights of the linear combinations that specify the style characteristics of the individual prototypes. This shows that our method is suitable for building models for continuous movement spaces from a small amount of example data that can be used for analysis and synthesis. Therefore, our method is a suitable representation for the transfer from perception to action in imitation learning

Acknowledgements Supported with a grant by the Deutsche Volkswagen Stiftung. Thanks to the Max Planck Institute for Biological Cybernetics for providing support.