Learning Based Representation of Complex Movement Patterns

Martin Giese & Tomaso Poggio

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Development of theoretical methods for the representation of complex movements based on learned prototypical examples for technical applications, and as basis to understand action recognition in biological systems.

Motivation: Learning-based methods have been successfully applied for the representation of the shape of complex stationary objects. In computer vision and computer graphics representations which are based on learned prototypical example images have been used for the recognition and synthesis of views of 3D objects. The idea of a view-based representation has also provided a valid interpretation of a variety of psychophysical and neurophysiological results on object recognition in humans and monkeys.

Aim of this work is to develop a learning-based approach for the representation of complex movements. The central idea is that classes of complex movements can be represented using only a small number of learned prototypical example movements when representations with suitable generalization properties are defined. Based on this idea, we develop new technical methods for the recognition and synthesis of complex visual movement patterns and neural models for the recognition of biological motion in the brain.

Previous Work: Many existing techniques for the recognition of motion patterns computer vision are either model-based [7], or extract simple spatio-temporal features [6]. Learning-based methods have been proposed for tracking of movements (e.g. [3]). Learning methods have also been applied for the generation of stationary images in computer graphics (e.g. [2]). Only recently, first attempts have been made to apply learning methods for the representation of classes of complex movements (e.g. [4]).

The idea of a representation of the of the stationary shape of three-dimensional objects based on learned two-dimensional prototypical views has been widely investigated (e.g. [12, 1, 14]). Meanwhile the same technique has been generalized for the representation of classes of 3D object views [2]. We try to generalize the same idea for the representation of image sequences.

The concept of a view-based representation of three-dimensional object shape has received strong experimental support from psychophysical and neurophyiological experiments [5, 11]. The combination of view-tuned neurons with the assumption of feature detector hierarchies allows an accurate modeling of pychophysical and neurophysiologiocal results on invariant object recognition in primates [13]. Almost no neural theories exist on the recognition of complex biological motion stimuli in the visual system [9]. We try to explore if a recognition of complex motion patterns can be achieved with similar neural principles as the recognition of stationary objects.

Approach: In the context of technical applications we have developed a new method that we call *spatio-temporal morphable models*. This method allows to represent classes of similar complex movements, like locomotion pattern with different styles (walking, running, limping, etc.), or sports movements with different skill levels. Our technique allows also to morph between different complex movements and to create realistically looking new movement patterns. In this way, we can modify the style of complex movements along easily interpretable dimensions. Morphable models define topologies over classes of similar movements. This makes them suitable for the quantification and parameterization of the spatio-temporal structure of complex movement patterns using abstract parameters that are easily interpretable.

Within our research on theoretical neuroscience, we develop and test a neural model for the recognition of biological motion. This model accurately reproduces different psychophysical results and is based on a neurophysiologically plausible hierarchical structure with two separate pathways for the processing of form and motion. (see Fig. 1 for an overview). The model achieves invariance with respect spatial shifts and stimulus size by pooling the responses of local detectors using maximum operations [13]. Information is associated over time by dynamic recurrent neural networks. We derive predictions from our model about the generalization properties of the learning-based representations and the neural activity distributions over in different parts of the visual pathway. One prediction from the model is the existance of smooth spatio-temporal generalization fields. This prediction was experimentally confirmed in [10].

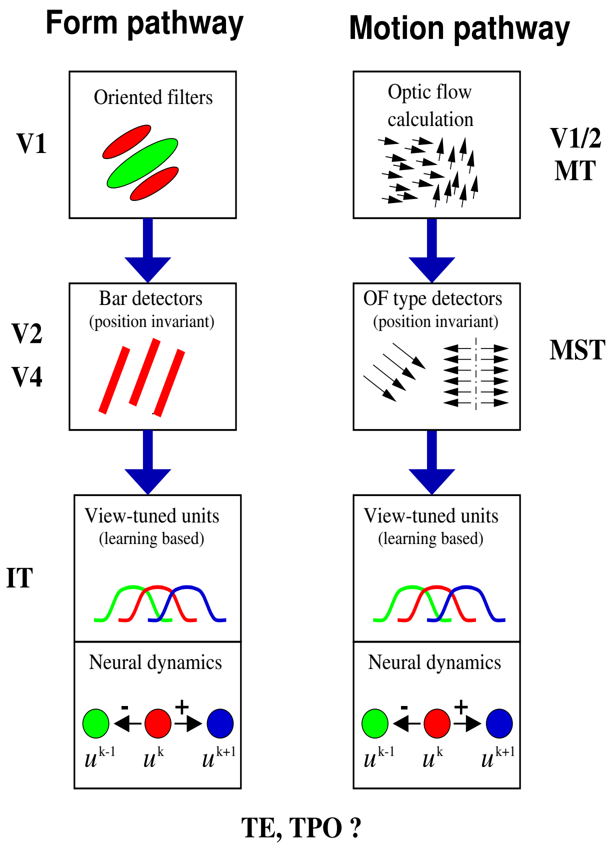
Difficulty: The calculation of correspondences in space-time leads to a special constrained optimization problem because the temporal correspondence shifts must preserve a monotonic temporal order. The developed correspondence algorithm solves this problem by dynamic programming. The biological model requires to represent temporal order of complex patterns. We test different possible neural mechanisms and try to derive predictions that allow to distinguish between them based on neurophysiological data.

Impact: The applied part of the project offers new methods for the generation and analysis of video image sequences by recognition and morphing between different complex movements (e.g. the golf swings of different golfers). For experimental research in biology, the possibility to parameterize the space of complex motion patterns offers new ways for controlled quantitative experiments about the recognition and learning of complex movements. The neural models makes quantitative predictions that can be verified on the basis of FMRI, psychophysical and neurophysiological data.

Future Work: Future work will focus on the development of a robust tracking algorithms that make the method suitable for standard video sequences. In collaborations with different experimentalists, we develop psychophysical and neurophysiological paradigms to test different predictions from the model and to find out how prototypical movement patterns are learned.

Research Support: Research at CBCL is supported by ONR, Darpa, NSF, Kodak, Siemens, Daimler, ATR, ATT, Compaq, Honda. Martin Giese is supported by the Deutsche Forschungsgemeinschaft.

Fig. 1: Neural model for action recognition



References:

[1] Beymer, D & Poggio, T (1996), *Science*, 272, 1905-1909.

[2] Blanz, V & Vetter, T (1999), *SIGGRAPH Proceedings 1999*, 187-194.

[3] Blake, A & Isard, M (2000). *Active Contours*, Springer, Berlin.

[4] Brand M (2000). *SIGGRAPH Proceedings 2000*, (in press).

[5] Buelthoff, H H & Edelman, S (1992). *Proc. Nat. Acad. Sci.*, 89, 60-64.

[6] Davis, J W & Bobick, A F (1996). Technical Report 402, Massachusetts Institute of Technology.

[7] Essa, I A & Pentland, A (1997). *IEEE Trans. PAMI*, 10, 757-763.

[8] Giese, M A & Poggio, T (1999). *Int. J. of Comp. Vis.*, (in press).

[9] Goddard, N H (1992). *The perception of articulated motion: recognizing moving light displays.* PhD Thesis, Univ. Rochester.

[10] Lappe, M & Giese, M A (2000). *Inv. Opht. Vis Sci*, (in press). .

[11] Logothetis, NK, Pauls, J & Poggio, T (1995). *Curr. Biol.* **5**, 552-563.

[12] Poggio, T & Edelman, S (1990), *Nature*, 343, 263-266.

[13] Riesenhuber, M & Poggio, T (1999). *Nat. Nsc.*, 2, 1019-1025.

[14] Vetter, T & Poggio, T (1997), *IEEE Trans. PAMI*, 19, 733-742.