Online Event Segmentation in Active Perception using Adaptive Strong Anticipation [1]

Bruno Nery [a] Rodrigo Ventura [a]
[a] *Institute for Systems and Robotics*
Instituto Superior Técnico, TULisbon
Av. Rovisco Pais, 1; 1049-001 Lisbon; PORTUGAL

Abstract. The segmentation of the stream of perceptual inputs a robot receives into discrete and meaningful events poses as a challenge in bridging the gap between internal cognitive representations, and the external world. Event Segmentation Theory, recently proposed in the context of cognitive systems research, sustains that humans segment time into events based on matching perceptual input with predictions. In this paper we propose a framework for online event segmentation, targeting robots endowed with active perception. Moreover, sensory processing systems have an intrinsic latency, resulting from many factors such as sampling rate, and computational processing, and which is seldom accounted for. This framework is founded on the theory of dynamical systems synchronization, where the system considered includes both the robot and the world coupled (strong anticipation). An adaption rule is used to perform simultaneous system identification and synchronization, and anticipating synchronization is employed to predict the short-term system evolution. This prediction allows for an appropriate control of the robot actuation. Event boundaries are detected once synchronization is lost (sudden increase of the prediction error). An experimental proof of concept of the proposed framework is presented, together with some preliminary results corroborating the approach.

Keywords. Event segmentation, anticipative systems, active perception, cognitive robotics.

Introduction

The perception of a robot is grounded on the physical world. Its sensors receive a continuous stream of information, as for instance the light patterns hitting the CCD sensor of a video camera. Cognitive representations, however, are often discrete, as in the case of events and objects. Discretization is commonly performed in fixed, not always adjustable, discretization step (e.g, the frame rate and the pixel resolution of a video camera). The detection of meaningful events from a stream of sensory information is an important challenge, from the point of view of a robot cognitive architecture design,

[1] This work was supported by the FCT (ISR/IST plurianual funding) through the PIDDAC Program funds. Partially funded with grant SFRH/BD/60853/2009, from Fundação para a Ciência e a Tecnologia

contributing to bridge the gap between a continuous time world and discrete time, event-based cognitive representations.

The segmentation of a continuous stream of information into events is often overlooked, being commonly performed in an ad-hoc manner, either recurring to threshold values over heuristic functions, or fixed time triggers. These methods are mostly sensor modality dependent, as well as task specific. This paper addresses the problem of bridging the gap between the time continuous stream of sensory/actuation information, and the discrete time sequence of cognitive representations, proposing a modality and task independent framework for event segmentation.

The Event Segmentation Theory (EST) provides a model of how the human brain segments perception into a sequence of events [15,6]. This model sustains that event segmentation is based on the detection of prediction errors in the sensory stream. In particular, the human brain is permanently making predictions and comparing them with the actual outcome [11]. Events are detected whenever a significant disparity between prediction and outcome is encountered. An event segmentation mechanism can be built following this principle, but the problem of how to make predictions about perceptions has to be addressed first.

Dubois distinguishes between *strong* and *weak anticipation* [3,12]: the latter is based on an explicit model of the world, where the physics is encoded in analytical constructs, that can be mathematically solved given an initial condition. On the contrary, strong anticipation does not rely on a model, but rather on the dynamical evolution of the interaction of the agent with the world, seen as a single system.

Stepp proposes an approach to strong anticipation based on the work developed in the field of chaotic systems concerning synchronization of dynamical systems [12]. Consider two systems, denoted D (drive) and R (response), connected by a unidirectional flow of information from D to R. It is possible to design the system R such that its dynamic evolution synchronizes with the one of D, regardless of the initial condition of each system. More interestingly, if this feedback loop contains a delay, system R is capable, under certain conditions, to anticipate system D [13].

One problem remains to be solved: how to design system R? No system model is assumed *a priori*, since it depends on the coupling involving the robot and the world. A possible approach is to adapt system R during interaction. A solution to the adaptation of response systems in the context of dynamical systems synchronization has been proposed by Chen [1], where the convergence to the solution has been proved using the Lyapunov stability theory. This result does not directly apply, however, to anticipating synchronization.

The contributions of this paper are:

- An event segmentation method based on Stepp's strong anticipation concept [12], cast as an anticipating system synchronization framework;
- The application of Chen's parameter identification method [1] to anticipating synchronization;
- A proof-of-concept implementation of an architecture for event segmentation and active perception, employing these methods.

This paper is organized as follows: after a short section surveying related work, two sections on the theoretical background behind strong anticipation and the adaptation method to learn the response system R follow. Then, the proposed architecture for event segmentation is described, followed by some experimental results of a proof of concept implementation of these ideas. A section presenting some conclusions and open questions closes the paper.

1. Related work

The problem of event segmentation has been studied in the past. See [9] for a review of recent techniques for the formation of event memories in robots. Ramoni *et al.* proposed a method to cluster robot activities using Markov chain models [10]. The spatio-temporal segmentation of video have been studied in [14], applying motion model clustering, and in [2] using hierarchical clustering of the 3D space-time video stream. Gesture segmentation and recognition has been addressed in [5] employing hidden-Markov models (HMM).

2. Strong anticipation

In [12] strong anticipation is modeled using a dynamical system synchronization framework. Consider two continuous dynamical state vectors $x(t), y(t) \in \mathbb{R}^n$ with the following coupled dynamics:

$$\begin{aligned} \dot{x} &= f(x) \\ \dot{y} &= f(y) + k(x - y_\tau) \end{aligned} \tag{1}$$

where $y_\tau = y(t-\tau)$, *i.e.,* a feedback loop with a constant delay τ, and k is a scalar gain. The first system is called the *drive* (D) while the second the *response* (R). This delayed feedback loop in the response system is a fundamental aspect, and is responsible for the response system capability of anticipating the trajectory of the drive.

This delayed feedback loop is neurophysiologically supported by the discovery of forward models in the brain, which predict sensory consequences of motor commands [7, 4]. One important function of this mechanism is to overcome the sensory processing latency in the brain, when the subject is performing quick, controlled movements.

3. Adaptive synchronization

If the drive system corresponds to the world-robot coupled system, its dynamics is not known *a priori*. One way of tackling this problem is to adapt the response system, online, during synchronization.

Chen proposes in [1] an approach to adapt response systems in the context of dynamical system synchronization. It does not account, however, for a delayed feedback.

Consider that the drive system has the form

$$\dot{x} = f(x) + F(x)\theta \tag{2}$$

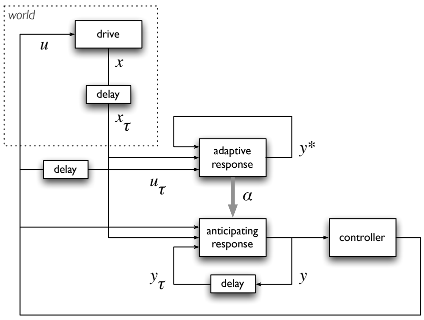

Figure 1. System architecture

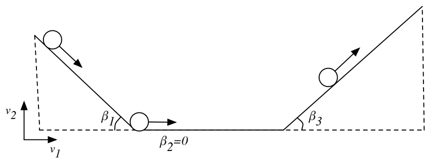

Figure 2. Simulated scenario

where $\theta \in \mathbb{R}^m$ is a vector of (constant) parameters, $f(x) \in \mathbb{R}^n$ and $F(x) \in \mathbb{R}^{n \times m}$. The response system is identical, except for the parameter vector that is unknown, and for the synchronization feedback loop

$$\dot{y} = f(y) + F(y)\alpha + U(y, x, t, \alpha) \tag{3}$$

where α is the response parameter vector, and $U(y, x, t, \alpha)$ is called the controller of the response. Chen *et al.* proved in [1] that, under certain conditions, not only the response system synchronizes with the drive, but also that the response parameters α converge to the ones of the drive θ.

4. Event segmentation

The event segmentation framework we propose in this paper, depicted in Figure 1, consists of a pair of response systems, one performing adaptation (labeled *adaptive response*), and the other anticipation (labeled *anticipating response*). The adaptive response learns the parameter vector α as described in the Adaptive synchronization section, while anticipating response performs anticipating synchronization as explained in the Strong anticipation section. The robot-world coupled system is modeled by the controlled drive system. Note that the access of the architecture to the world state is subject to a delay, modeling for instance the latency of the perceptual channel (image acquisition, processing, and tracking). The controller computes the actuation vector u based on the anticipated world state y.

According to the theory of Event Segmentation [15], perceptual systems continuously make predictions about perceptual input, and perceive event boundaries when transient errors in prediction arise. On the adaptive synchronization framework, the Lyapunov function provides a solid estimate of the prediction error. Event boundaries are detected using a simple hypothesis testing on the statistics of the prediction error. A detailed description of the method can be found in [8].

5. Experimental results

As a proof of concept for the ideas presented here, a simple scenario was simulated: a ball rolling free on a series of inclined planes, with different slopes, is observed by a robot camera which aims to follow it, in order to center it on the image, as depicted in Figure 2. The camera moves parallel to the plane, for simplicity sake. For this proof of concept, we set the response system to be structurally identical, thus employing the same functions f and F, and control input u.

The experiments were conducted after discretizing the system using a simple approximation $\dot{z}(t) \simeq [z(t+T) - z(t)]/T$. The sampling rate was 100Hz. The delay considered was $\tau = 0.65s$ (65 samples). The system is initialized with the ball starting on the top left position of the ramp, and as the ball transverses the scenario there are two events, corresponding to the two changes of the ramp slope. Each simulation takes 100s of simulated time.

Figure 3 (top) shows the evolution of the ball horizontal position in the camera without an anticipating response system. As expected, the delay introduced by perceptual channel jeopardizes the control of the camera. Figure 3 (bottom) shows the ball horizontal position in the camera using the full architecture. In this case, the ball coordinates in the image converge to zero (except for a brief time after each slope change, while the adaptive system learns the new parameters). Also, the anticipating response makes it possible to control the drive system satisfactorily.

Figure 4 (top) attests the performance of the adaptive response system. As can be seen, the synchronization error converges to zero after each event boundary. This only happens because the parameter vector α converges to the true parameters θ. Finally, Figure 4 (bottom) shows the event segmentation results. As expected, each change of plane is detected as an event boundary by the framework.

These results show that the proposed system is capable of correctly (1) detecting the event boundaries that correspond to the change of ramp slope by the ball, (2) controlling the camera movement using anticipation, and (3) learning the correct system parameters.

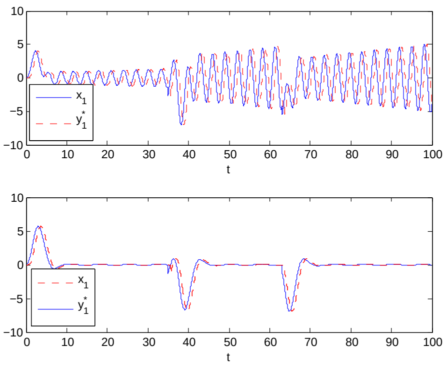

Figure 3. Ball position: response without anticipation (top) and with full architecture (bottom).

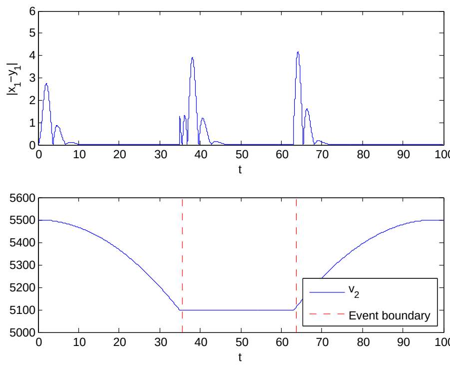

Figure 4. Synchronization error for the ball horizontal position (top) and detected events (bottom).

6. Conclusions and future work

This paper describes an event segmentation framework, targeting active perception in robots, based on the concept of strong anticipation proposed by Stepp *et al.* in [12]. A dynamical system synchronization paradigm is used as theoretical foundation of the proposed architecture, where the robot-world coupled system is identified using an adaptive parametric method proposed by Chen *et al.* in [1], and the control is anticipatory. This anticipation accommodates for the net delay of the perceptual channel. The capability of the architecture to anticipate perception allows the robot to control its actuation based on the prediction of the robot-world state, instead of relying on the delayed perceptual data.

Having the described proof of concept experiments shown that the proposed architecture behaves as expected, future work includes scaling this approach to more complex domains. This involves tackling the issues of the learning rate, which is hidden in the proportionality constant of the Lyapunov function, used in the Chen's learning rule, as well as the automatic design of the controller, given the adapted parameters. Other open questions include dealing with hidden state variables, as well as complex relations among objects (e.g., grasping, occlusion, and so on).

References

[1] Shihua Chen and Jinhu Lü. Parameters identification and synchronization of chaotic systems based upon adaptive control. *Physics Letters A*, 299:353–358, 2002.

[2] D. DeMenthon. Spatio-temporal segmentation of video by hierarchical mean shift analysis. *Language*, 2, 2002.

[3] Daniel M. Dubois. Mathematical foundations of discrete and functional systems with strong and weak anticipations. In *Anticipatory Behavior in Adaptive Learning Systems*, Lecture Notes in Computer Science, pages 107–125. Springer, 2003.

[4] Mitsuo Kawato. Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9(6):718–727, December 1999.

[5] Daehwan Kim, Jinyoung Song, and Daijin Kim. Simultaneous gesture segmentation and recognition based on forward spotting accumulative hmms. *Pattern Recognition*, 40(11):3012–3026, November 2007.

[6] Christopher A. Kurby and Jeffrey M. Zacks. Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2):72–79, February 2008.

[7] R.C. Miall, D. J. Weir, D. M. Wolpert, and J. F. Stein. Is the cerebellum a smith predictor? *Journal of Motor Behavior*, 25(3):203–216, 1993.

[8] Bruno Nery and Rodrigo Ventura. Online event segmentation in active perception using adaptive strong anticipation. Technical Report RT-701-10, Institute for Systems and Robotics, June 2010, **arXiv:** 1006.5829v1 [cs.RO].

[9] Erich Prem, Erik Hörtnagl, and Georg Dorffner. Growing event memories for autonomous robots. In *Proceedings of the Workshop On Growing Artifacts that Live: Basic Principles and Future Trends*, 2002.

[10] Marco Ramoni, Paola Sebastiani, and Paul Cohen. Unsupervised clustering of robot activities: a bayesian approach. In *Proceedings of the fourth international conference on Autonomous agents (AGENTS'00)*, pages 134–135, 2000.

[11] Wolfram Schultz and Anthony Dickinson. Neuronal coding of prediction errors. *Annual Review of Neuroscience*, 23:473–500, 2000.

[12] N. Stepp and M.T. Turvey. On strong anticipation. *Cognitive Systems Research*, 11:148–164, 2010.

[13] Henning U. Voss. Anticipating chaotic synchronization. *Physical review E*, 61(5):5115–5119, 2000.

[14] J.Y.A. Wang and E.H. Adelson. Spatio-temporal segmentation of video data. In *SPIE Proceedings Image and Video Processing II*, volume 2182, pages 120–131, 1994.

[15] Jeffrey M. Zacks, Nicole K. Speer, Khena M. Swallow, Todd S. Braver, and Jeremy R. Reynolds. Event perception: A mind–brain perspective. *Psychological Bulletin*, 133(2):273–293, 2007.