Procedia Computer Science

Volume 88, 2016, Pages 528–533

7th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2016

On Virtual Characters that Can See

Eugene Borovikov[1,2] and Sergey Yershov[1]
[1]*PercepReal, USA*
[2]*National Institutes of Health, USA*
{Borovikov, Yershov}@PercepReal.com

Abstract
A virtual character (VC) acts within its virtual world boundaries, but with vision sensory capabilities, it may be expected to explore the real world and interact with the intelligent beings there. Such a VC can be equipped with algorithms to localize humans, recognize and communicate with them. These perceptual capabilities prompt a sophisticated cognitive architecture (CA), enabling our VC to learn from intelligent beings and perhaps reason like one. This CA needs to be fairly seamless, reliable and adaptive. Here we explore a vision-based human-centric approach to the VC design.

Keywords: cognitive architecture, virtual character, computer vision

1 Introduction

A typical virtual character (VC) is usually confined to its virtual world, but if we equip it with some visual/audio sensors, provide basic algorithms for object recognition and tracking, and add to it some machine learning (ML) based capabilities for it to learn and reason, our VC will be able to evolve into an intelligent virtual being. VC's perception relies on the given sensory capabilities, e.g. video cameras for its eyes or microphones for its ears. Those sensory streams should have enough of the signal resolution to distinguish among the important features of the objects and beings that our VC would need to interact with. Such a VC would need to use its evolving cognitive architecture to decide on a winning combination of the important features characterizing the real-world objects.

Communications between VC and humans are of the most interest to this study, and thus a virtual character should be able to localize and track humans (e.g. via non-rigid 2D or 3D models), recognize them (e.g. by their faces and/or voices) and communicate with them via natural interfaces (e.g. a human-like avatar). Our VC needs to work in visually unconstrained environments, perform its sub-tasks in interactive time, and constantly learn from its experiences with virtual and real worlds. Such interactions should result in a gradual development of that VC, leading in a highly realistic virtual or mixed reality experience for the humans. We propose a vision-based human-centric approach to the virtual character design by equipping the prospective VC with visual sensors and targeting arguably the most visually expressive and natural human real-world manifestations: face and body.

 Selection and peer-review under responsibility of the Scientific Programme Committee of BICA 2016

doi:10.1016/j.procs.2016.07.475

2 Background

A cognitive architecture (CA) is a system that usually comprises multiple computational modules that, working as a whole, attempts to approach human-level intelligence (Goertzel, Lian, Arel, De Garis, & Chen, 2010). It is therefore not surprising that the use of cognitive architectures to generate more humanlike virtual characters has attracted considerable attention.

The biologically inspired (Hubel & Wiesel, 1968) convolutional neural networks (CNN) as introduced as *neocognitron* (Fukushima, 1980) and then improved, generalized and simplified (Simard, Steinkraus, & Platt, 2003), have seen a spectacular renaissance in the recent decade (LeCun, Bengio, & Hinton, 2015) due to the emergence of the affordable GPU computing power, which made the non-trivial image processing tractable for many visual tasks (Abdolali & Seyyedsalehi, 2012; Fan, Xu, Wu, & Gong, 2010) that may be considered quite important for visually capable virtual character adaptive cognitive architecture that needs to learn important features straight from sensors. Modern deep learning (Yue-Hei Ng, Yang, & Davis, 2015) content based image retrieval (CBIR) techniques (Wan et al., 2014) could also help with the VC's robust long-term memory sub-system development, e.g. by transferring deep networks trained on image classification to image retrieval tasks. Deep learning models are an essential part of the approach that we envision for integrating the vision modules described in this paper.

The cognitive architecture ACT-R/E (Trafton et al., 2013) was used to improve human-robot interactions (HRI) where humans and robots share the same virtual or physical environment and where actions of each participants affect those of others. Such HRI often consider the human participant to be more of a perfect machine than a human, i.e. making no mistakes, fully predictable, and is never affected by fatigue or negative emotions. ACT-R/E focused on addressing this gap by developing a CA that is capable of deeper modeling of human cognition to enable the robot participant to recognize when its human counterpart does something wrong. While the focus of our paper is narrower, we believe that the assumption that a physical character from which a virtual one attempts to learn is not perfect is critical to development of robust VC cognition.

For a humanoid VC to be successful in interactions, it is important to understand the human behavioral traits, capturing different personalities for modeling realistic behavior of a VC as suggested by (Saberi, Bernardet, & DiPaola, 2014), where the authors present a hybrid cognitive architecture that combines the control of discrete behavior of the VC moving through states of the interaction with continuous updates of the emotional state of the VC depending on the feedback from the environment. While testing their approach using turn-taking interaction between a human and a 3D humanoid VC, the authors noticed more individualized and believably humanoid artifacts in their VC's behavior.

3 Vision based development of a Virtual Character

Provided with a vision-based cognitive architecture (CA), a VC should be able to utilize its knowledge of the virtual world (e.g. 3D object models and hierarchical infrastructure) for its real-world perception tasks via the given visual sensors. Once aware of the real-world objects and possibly of their inter-relationships and inter-actions, our VC should be able to bring the learned concepts as models to the virtual world, reason about them, emulate some of them, and possibly share them with other intelligent agents, virtual or human.

We argue that the vision-based human-centric approach (Buxton & Fitzmaurice, 1998), involving face and body modeling, can provide the necessary foundation for introducing a sound cognitive architecture to the VC design, especially if it encompasses a human-like avatar, because

- fusing several modalities (texture, color, depth) should result in finer virtual models,
- VC can learn to distinguish between general and person-specific models, and
- real-time interactions promise a natural and incremental VC development.