The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence
Symbiotic Cognitive Systems: Technical Report WS-16-14

Artificial Attention at Scale

Alexander M Morison and David D Woods
The Ohio State University
Dept. of Integrated Systems Engineering
1971 Neil Ave, Columbus OH
{morison.6, woods.2}@osu.edu

Abstract

Human-machine systems have expanded in terms of their sensing, communication, and computational capabilities. These capabilities have led to developments of a variety of sensor systems, like robotic platforms. There are benefits to these new sensor systems, however, these benefits have been offset by new difficulties; dynamic data overload, keeping pace with changing tempo, and managing data flows from multiple sensors feeds. One approach to manage data overload from multiple sensor feeds are computational models of attention. These models also address an important aspect of human-machine symbiosis, the need for machines agents to understand attention, manage interaction based on the flow of attention, and anticipate the flow of attention in the future. Unfortunately, existing computational models of attention use assumptions that limit their applicability to human-machine systems. The Artificial Attention Architecture is introduced and demonstrates how computational models of attention can be extended to handle multi-agent, multi-sensor systems. The Artificial Attention Architecture addresses important properties of human-machine systems like the need to build symbiosis between people searching for meaning in extensive data flows and the computational algorithms processing complex and dynamic data flows.

Introduction

Since the 1930s, human-machine systems have expanded to include sensors, computations, and robotic platforms. There is a clear benefit to these new sensor systems that are expanding human range in several important ways. We are able to access previously inaccessible environments, take new vantage points, and explore a scene from multiple vantage points simultaneously. This new access takes the form of image feeds that are captured by local human and machine agents – in the scene of interest – and transmitted to distant problem holders and machine agents. These feeds allow problem holders, assisted by machine agents, to explore distant scenes by navigating over the feeds from multiple sensors (Morison, Woods, and Murphy 2015).

The benefits of expanding human range through sensors have been offset by new difficulties. One is the challenge of dynamic data overload (Woods et al. 2002). For sensor systems, this challenge has grown as the size and diversity of sensor networks has grown. The data overload problem is not new however and systems that manage data overload exist. Previously, one way to escape from data overload has relied on the growth of computational resources and sophisticated inference mechanisms to compute a 'best' answer. An alternative approach to managing data overload utilizes properties and functions of human attention, in particular, builds on advances in computational models of human attention (Woods and Sarter 2010). Neurobiology tackles dynamic data overload by starting with the tracking the flow of events and change, through mechanisms to continually focus and re-focus the perceptual apparatus (Itti et al. 2005, Zachs and Tversky 2001).

A second reason for using computational models of attention to manage data flows from multiple sensor feeds is the problem of pacing as the tempo of events and activities in the world change (Woods and Hollnagel 2006). Computational approaches to dynamic data overload assume processing speed can always outpace the tempo of events. But the true challenge is being sensitive to the change in tempo of activities and events over time, not simply outpacing them. Managing and being sensitive to the varying tempo of activities and events is a basic part of the expertise of experienced human operators.

This means any machine agent that engages in a symbiotic relationship with a human agent will have to have an understanding of attention, manage interaction based on the flow of attention, and anticipate the flow of attention in the future.

Computational approaches have assumed that human attention is a bug not a feature. However, difficulties in finding what is meaningful from sensor networks and sensors

Figure 1. An eye track example. The fixed boundary of the image limits applicability of this data to human-robot systems.

on robotic systems highlights the need for some mechanism that performs a similar function to human attention (in small part because machines are not exempt from bounds on resources). This paper considers how computational models of attention can be extended to create a new form of symbiosis between human and machines to find what is meaningful in the data flows from multiple sensors.

Expanding Symbiosis

Since the publication of Licklider's 1960 paper, research on attention has moved from phenomena at an individual scale, to phenomena at a micro scale and more recently to the underlying neural structures responsible for attentional processes. This is important work and many insights from this work can be applied to the study of human-robot coordination. However, there is little research studying how people and machine agents can coordinate at the new expanded scales that arise from from data flowing from extensive networks of sensors. The important question for the joint system is: *how to explore and discover what is meaningful, in pace with a changing world, when the questions are ill-defined and expectations, context, and priorities change, as more data flows in at rates that hamper available computational resources?*

There are at least two critical gaps in existing approaches to address the above question. The first gap is the ability of machine agents to focus and shift focus given only a partial view of an environment from available sensors. The second gap is the ability to notice deviations from typicality, where typicality is context dependent. Computational models of attention provide a starting point but need to be extended to meet these challenges. For instance, current computational models simulate a saccadic process, but they do not explicitly address the concept of focusing and refocusing, or the idea of a partially observable environment. The underlying assumptions behind current computational models limit their applicability to system that must work at larger scales.

Computational Models of Attention

The starting point is computational models of the neurobiological mechanisms of attention (Itti and Koch 2001). These computational mechanisms can be used to develop Artificial Attention systems, rather than serve as models of a single human's attentional capability. As simulations of attention, these models have the potential to embody the capabilities of human attention, including the ability to function with uncertainty, to shift expectations and priorities, and to focus on what is important now while remaining sensitive to what could be important in the future.

Current computational models of attention are encoded as an algorithm that takes input, which simulates sensory data such as light or sound and then focuses and refocuses over the input array over time. Several versions have been developed including, Koch and Ullman, 1984, which is a precursor to the model developed by Itti and colleagues, Treisman and Gormican 1988, Tsotsos et al. 1995, Le Meur et al. 2006, Frintrop, et al. 2007, and Wickens et al.

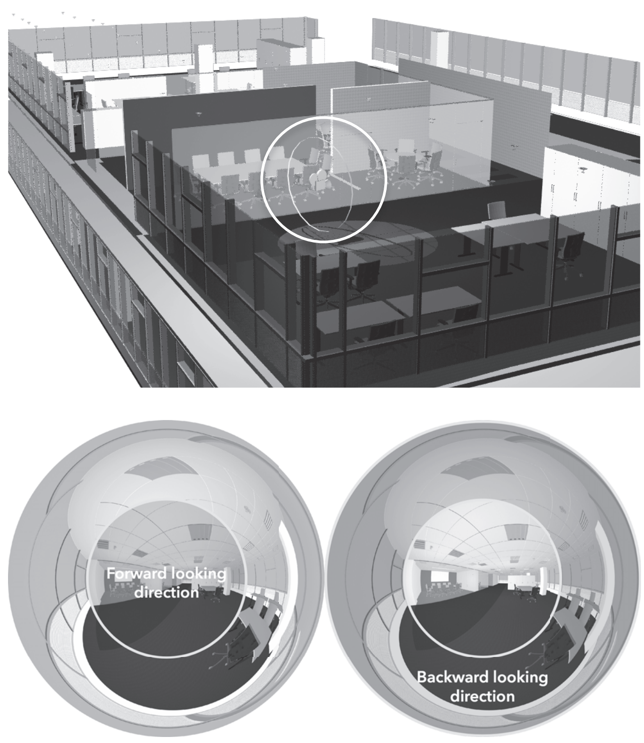

Figure 2. An example of a 3-dimensional virtual environment with sensor (above) and the maximum field of view of the sensor including forward view directions bottom-left image) and backward view directions (bottom-right image).

2003. Figure 1 shows an example of an image of a scene with the simulated eye track from one model superimposed (Itti and Baldi 2005). These model outputs are then used to explain and assess human attentional performance.

Extensions for Sensor Networks

Expanding a computational model of attention to handle multi-agent, multi-sensor systems reveals three requirements to balance limited resources with a complex and dynamic environment. These requirements are fundamental constraints on any system of cognitive agents.

1) *Sampling impacts what is and is not sampled*: The output of an attentional process is a sequence of samples over physical space and time. Importantly, the sample path in the future is affected by past samples, the activity in the world, and the sensitivity of the attention process to particular activities. In past models, the simulated attentional process is not affected by past samples.

2) *Breaking the fixed frame boundary*: Attention is an active sampling of an environment that results in a dynamic panorama with changing shape and extent. In previous computational models the dynamic panorama is narrow and static as shown in Figure 1. A narrow and static panorama can never be expanded to function at the scales required for multi-agent, multi-sensor systems.

3) *Multiple viewpoints*: The input to past computational models of attention is a flat planar representation of a 3-dimensional environment. While a flat planar representation permits a single point-of-observation, the planar frame-of-reference is insufficient for coordinating across multiple viewpoints. As a result, computational models of attention with this simplification cannot function at the scales required for multi-agent, multi-sensor systems.

These requirements place constraints on how an attention process can be simulated. A first constraint that follows from the first two requirements is that the pace of activities in the environment must be well-matched to the sampling process. The pace of activities cannot be too fast or too slow for the sampling process. A second constraint is about the field-of-view of the attention process. Given the second requirement, the field-of-view must be smaller than the maximum field-of-view – a sphere – and must be able to move through the environment. The third requirement means there must be the ability to instantiate multiple sensors simultaneously. These constraints, together, require a high degree of control on the environment of the sampling process. One solution is a simulated 3-dimensional environment that provides control over the pace of activities in the environment and the sensor access. An example of such and environment is shown in Figure 2.

Artificial Attention Architecture

The Artificial Attention Architecture is shown in Figure 3 and combines three components in a unique way to simulate the capability of human attention in a way that can function in a multi-agent, multi-sensor system (Woods and Morison 2014). The first component, bottom-up saliency, is ubiquitous across computational models of attention and is an active productive area of research (Baldi and Itti 2012). The second component is mechanisms for *focusing* and for *reorienting* and how they interact. Initial development and results have demonstrated the importance of how these two processes of focus and reorienting balance. The third component is a computational system to integrate top-

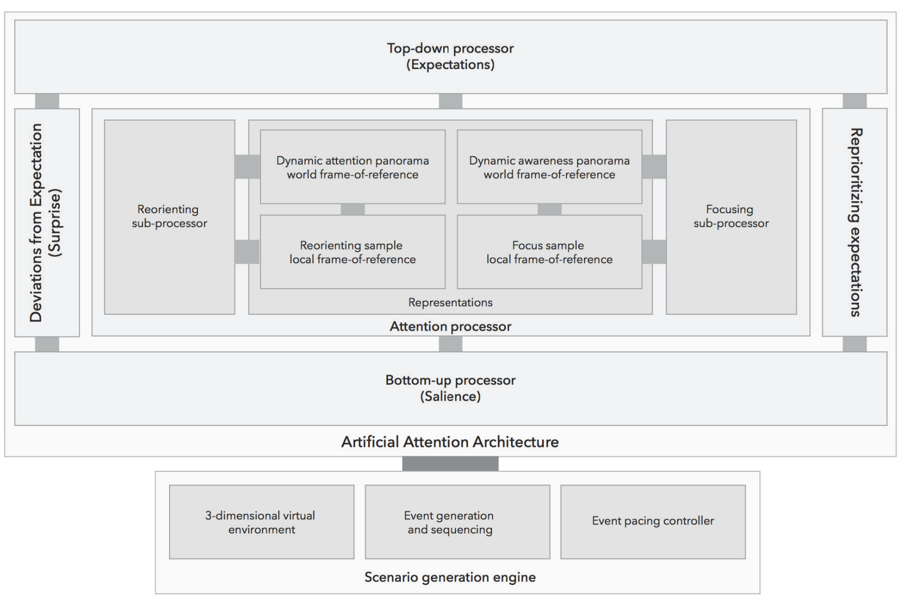

Figure 3. A diagram of the Artificial Attention Architecture with Scenario generation engine.

down expectation in the architecture. Together, these computational sub-processes form an architecture that can continuously output what is most significant as activities, events, objects-of-interest, expectations and priorities change over time.

An example of the output created by the Artificial Attention Algorithm is shown in Figure 4. These snapshots illustrate the three requirements for extending computational models of attention to new scales. First, the sensors sampling the environment do not have complete access to the environment at every instant. Instead, samples are built up over time to provide a view that is partial and incomplete. There are obvious gaps in the visual field. Second, there is no smooth fixed boundary to the attention "panorama", as is the case for other computational models of attention. In fact, the panorama is a complex shape that exists over space and time. The challenge of multiple viewpoints also has been demonstrated with the architecture (but is not illustrated in Figure 4).

This approach for developing the Artificial Attention Architecture for multi-agent, multi-sensor scales uses the functional engineering approach previously advocated by Newell for AI as an experimental science (Newell and Simon 1961). The functional engineering approach stresses the simulation of system function and then using the behavior of the simulated system acting within a representative environment as feedback (e.g., Roberts and Morison 2014). An analysis of the difference between simulated behavior and real system behavior identifies gaps in function and inspires modifications of structural mechanisms to achieve those functions.

Summary

The Artificial Attention Architecture demonstrates how computational models of attention can be extended to function at multi-agent, multi-sensor scales. This architecture extends current computational models of attention by eliminating several hidden assumptions that block their ability to function at scale. These hidden assumptions were identified by contrasting functions from the neurobiology of attention key for explaining attention at the scale of a single individual with constraints that had to be met for any attentional system to function at the larger multi-agent, multi-sensor scale relevant to many systems today. Artificial attention at scale captures Licklider on human-machine symbiosis in several ways.

First, the human-machine joint cognitive system is reframed in terms of the problem of finding what is meaningful in extensive data flows from multiple heterogeneous sensors. Handling this scale of data requires extensive machine processing. It also requires overcoming problems of data overload, which can be accomplished by developing

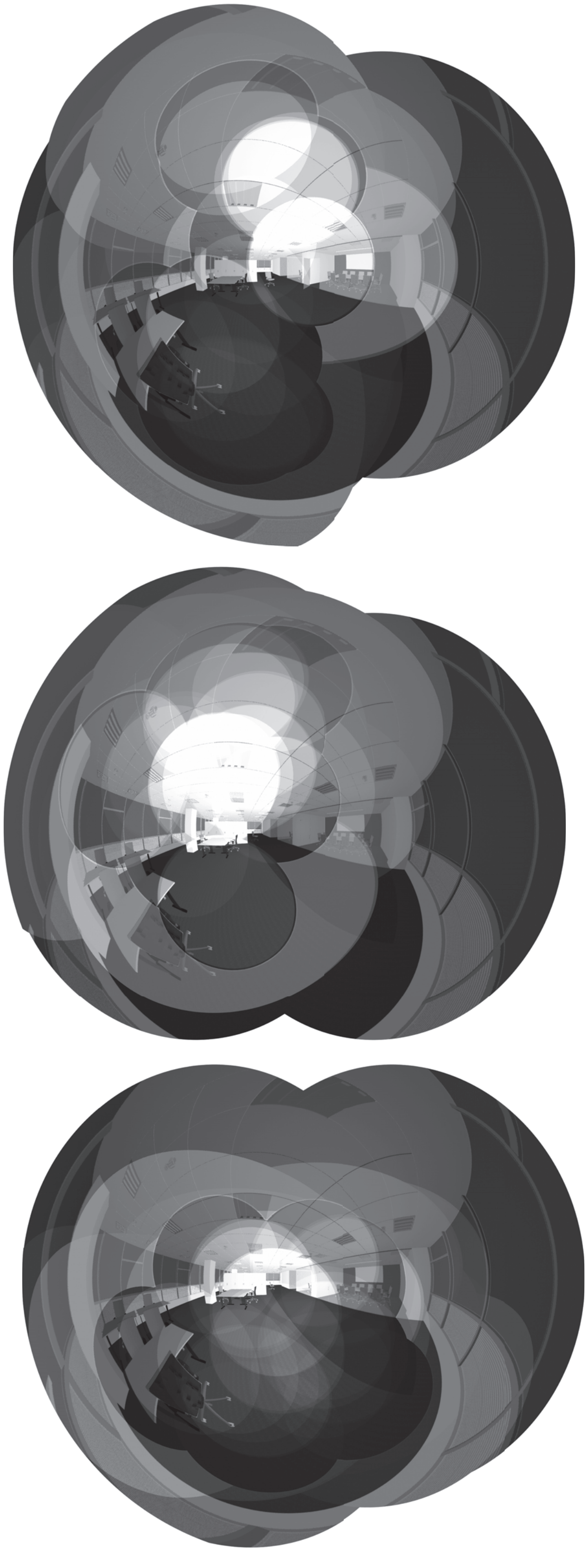

Figure 4. A sequence of snapshots from the output of the Artificial Attention Architecture using the 3-dimensional virtual environment shown in Figure 2(left to right, top to bottom).

attentional capabilities modeled on the latest information about human attention. While inspired by the latest findings and models of attention at an individual level, to get artificial attention mechanisms to work at the new scales required innovations that go beyond explaining the mechanisms behind human attention.

Second, Artificial Attention forms a new approach to build symbiosis between people searching for meaning in extensive data flows and the computational functions processing the data flowing from the network of multiple heterogeneous sensors.

Third, Artificial Attention highlights a performance test for joint human-machine sensor systems: what is the ability of the joint system to re-focus on what might be interesting while keeping pace with the changing events and activities in the scenes of interest made accessible through the network of multiple heterogeneous sensors.

References

Frintrop, S., Klodt, M., and Rome, E. 2007. A real-time visual attention system using integral images. *Proceedings of the 5th International Conference on Computer Vision (ICVS 2007)*.

Itti, L. and Baldi, P. 2006. Bayesian surprise attracts human attention. *Advances in neural information processing systems*, *18*: 547.

Itti, L. and Koch, C. 2001. Computational modelling of visual attention. *Nature reviews: Neuroscience*, *2*: 194–203.

Koch, C. and Ullman, S. 1984. Selecting one among the many: A simple network implementing shifts in selective visual attention. MIT University, Boston, MA.

Le Meur, O., Le Callet, P., Barba, D., and Thoreau, D. 2006. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*: 802–817.

Morison, A. M., Woods, D. D., and Murphy, T. B. 2015. Human-Robot Interaction as Extended Perception. In R. R. Hoffman, P. A. Hancock, R. Parasuraman, J. L. Szalma and M. Scerbo (eds.), *Cambridge Handbook of Applied Perception Research*.

Newell, A. and Simon, H. A. 1961. *Computer simulation of human thinking*. Rand Corporation.

Roberts, D. and Morison, A. 2013. The Artificial Attention Model and Algorithm Baseline Testing over the Parameter Space. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 57(1): 250-254. SAGE Publications.

Treisman, A. and Gormican, S. 1988. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1): 15.

Tsotsos, J. K., Culhane, S. M., Kei Wai, W. Y., Lai, Y., Davis, N., and Nuflo, F. 1995. Modeling visual attention via selective tuning. *Artificial intelligence*, *78*(1-2): 507–545.

Wickens, C., McCarley, J., Thomas, L., Foyle, M. I. D., Goodman, A., and Hooey, B. L. 2003. Attention-situation awareness (A-SA) model. In *NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*.

Woods, D.D. and Hollnagel, E. 2006. *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. Boca Raton FL: Taylor & Francis.

Woods, D.D. and Morison, A.M, 2012. Feasibility of Artificial Attention at Beyond-Human-Scales, Technical Report, FA8650-07-D-1220, Defense Technical Information Center.

Woods, D.D., Patterson, E.S., and Roth, E.M. 2002. Can we ever escape from data overload? A cognitive systems diagnosis. *Cognition, Technology, and Work,* 4(1): 22-36.

Woods, D. D. and Sarter, N. B. 2010. Capturing the dynamics of attention control from individual to distributed systems: the shape of models to come. *Theoretical Issues in Ergonomics Science*, *11*(1): 7–28.

Zacks, J. M. and Tversky, B. 2001. Event structure in perception and conception. *Psychological bulletin*, *127*(1): 3.