Proceedings of the Remembering Who We Are – Human Memory for Artificial Agents Symposium, Mei Yii Lim and Wan Ching Ho (Eds.), at the AISB 2010 convention, 29 March – 1 April 2010, De Montfort University, Leicester, UK

An Object-Based Memory for Supporting Attentive Virtual Agents

Christopher Peters[1]

Abstract. This paper presents an overview of an object-based memory system for supporting attentive behaviours of a virtual agent to its surroundings. The memory model is multi-stage, consisting of successive stages of filtering of objects from the scene database based on the agent's synthetic vision of the environment. Stored information may be out of date and may not necessarily match the actual states of objects in the environment, thus providing the basis for the agent to have its own perspective of the environment. A brief description details the importance of an object-based memory system for supporting inhibition of return in visual attention systems. These allow mobile virtual agent's to continue to move their focus of attention around the scene, thus creating a loop between them and the environment.

1 Introduction

Internal storage, or memory, is crucial for virtual agents expected to act in a plausible manner within their environment. If an agent has no storage mechanism and interfaces directly with the scene database, it will always have perfect information about the state of the environment and, by definition, cannot have a unique perspective of its world. Such access ultimately reduces the credibility of the agent's behaviour with respect to viewers, particularly if it is intended to be simulating a real-world entity that is constrained by senses or must operate based on imperfect or uncertain environment information.

Here, a virtual agent is endowed with a synthetic vision capability for sensing its environment, a 3D virtual environment. A memory-model filters visual information, limiting the amount of environmental information that must be processed and stored by the agent. The model described in Section 3 is based on the popular, but simplified, multi-stage model of memory inspired by Atkinson and Shiffrin [1] (see Figure 1), incorporating modifications from the literature that suggest the existence of a specialised visual short-term memory and allowing latitude [11] from the strict serialisation of information flow between memory suggested in the original model. Visual attention is discussed in Section 4, not only as a method for ensuring a sound filtering strategy, but also in relation to coupling with memory for generating attentive behaviours.

[1] Interactive Worlds Applied Research Group, Coventry University, United Kingdom, email: Christopher.Peters@coventry.ac.uk

2 Background

While memory capabilities of some form are evident in most autonomous agent systems. While in many cases explicit memory models are not defined, some research has considered their implementation for a variety of purposes. Recently, Lim and colleagues [10] have considered memory for virtual and robot companions, particularly for allowing agents to migrate between embodiments and for enabling long-term interactions and relationships with humans. In contrast to work on autobiographical memory systems, other models have been concerned with memory for virtual agents embedded in virtual environments, based on sensory capabilities for assisting *active vision* [2], research sharing similarities with that undertaken in robotics [15]. Kuffner and Latombe [9], for example, presented a flat memory system for objects, which also facilitated forgetting processes. Evers and Musse [5] use a memory model with STM and LTM to study the impact of memory on behaviour using hierarchical finite state machines. The agents in Chopra-Khullar's [3] work store object-related memory thresholds are used to provide a measure of uncertainty for driving the allocation of attentive monitoring behaviours.

Peters and O'Sullivan previously presented a model based on synthetic vision and a staged memory model [13], later applied to visual attention and perspective-taking for theory of mind [12]. Here, further details are provided of how memory and attention mechanisms may be coupled to support attentive behaviours for autonomous virtual agents.

3 Memory Model

Stage theory posits that information progresses through a number of successive, discreet memory stages: A short-term sensory storage (STSS), a short-term memory (STM) and a long-term memory (LTM), as outlined in Section 3.1 and illustrated in Figure 1. An underlying notion in the model is that memories reaching later stages will tend to persist for longer. It represents a great simplification of the way in which information is thought to be stored and associated in humans. Nonetheless, it provides a very useful paradigm for real-time autonomous agents; it suggests a general filtering and storage scheme, free of the computational overhead and technical implications of more elaborate associative models (but also subject to the implied inherent restrictions).

3.1 Memory Structure

Each memory stage consists of a list of memory entries (Section 3.1.1). In order to generalise the description of the stages in the model, here we describe stages parameterised according to *capacity*, *threshold*, *duration* and *rehearsal*. By varying these parameters, the different memory stages elaborated on in the stage theory of memory may be modelled, and new stages or sub-stages defined if necessary. *Capacity* refers to the maximum number of memory entries that a memory stage can store at any one time. The *threshold* parameter is used to decide when memory entries should be removed from a memory stage i.e. forgotten. While the trace strength of a memory entry is greater than this threshold, the item remains in memory and the probability of its recall is presumed to be 1. Memory entries are removed from a memory stage when their trace strength falls below the threshold value. *Duration* is an indication of the amount of time that a memory entry will persist in a memory stage without being rehearsed in any way. *Rehearsal* is another threshold, but is used to decide when an entry should be moved to the next memory stage. Rehearsal may be increased by attending to an object.

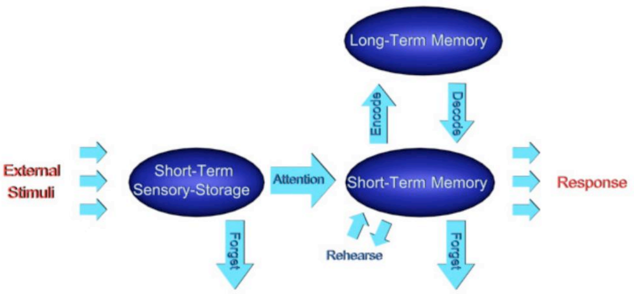

Figure 1. The basic stage theory of memory, a multi-stage model inspired by Atkinson and Shiffrin [1].

3.1.1 Memory Entries

Memories are considered to be composed of discrete constructs called *memory entries*. Memory entries contain information regarding what has been remembered and other information pertinent to the management of the memory system (see Table 1), allowing a recording of how recently the object was perceived and when it should be removed from memory. In the current implementation, there are two different types of memory entry: *proximal stimuli* and *observations*. Proximal stimuli do not contain resolved database object information, but merely information extracted from inspection, such as stimulus false colour (described shortly), retinotopic bounding box, size in pixels and average colour. Observations are elaborations of proximal stimuli, representing a snapshot of object state from the scene database.

Because observations persist in memory, they may often correspond to out-of-date information that is no longer accurate. An agent therefore has an internal representation of the environment that, in a similar manner to humans, may be imperfect. Only by continually sensing and attending to the environment can the agent maintain updates to the internal model.

Basic Memory Entry Structure

Member	Meaning
trace strength	The strength of the memory
decay factor	How quickly the memory decays
activation	Amount of time since the memory was rehearsed
rehearsal	Number of times that a memory has been rehearsed
proximal stimulus	Perceived stimulus
observation	Observation corresponding to the perceived stimulus

Table 1. The memory entry structure contains information for managing memory contents.

3.1.2 Synthetic Vision

The synthetic vision system takes snapshot renderings of the virtual environment from the agent's viewpoint. Unlike normal renderings, here objects are *false-coloured*; they are attributed unique colours and rendered in flat-shaded mode, without textures. Reading the colour values back from synthetic vision false-coloured renderings provides details of those objects that are within the agents field of view and visual range, that are not occluded by other objects. More details are available in [13].

3.1.3 Short-Term Sensory Storage

The short-term sensory storage, or *STSS*, is the first storage stage for stimuli that have passed through initial visual filtering. A memory entry is created for each object that is read back from the false-coloured image, to create a list of *proximal stimuli*. These do not persist for very long between updates of the visual system. Therefore, entries in the STSS could be thought of as corresponding to the 2D optical images deemed to fall on the *retina* of the agent, without containing semantic details. The STSS is updated with each refresh of the agent's viewpoint rendering: When a proximal stimulus is encoded in the STSS, a time-stamp is updated to the current time and the trace-strength is set to the maximum value for the STSS memory stage. The capacity of the STSS is considered large enough to allow storage of all stimuli falling within the agent's view, subject to the process detailed in Section 3.1.2. When a proximal stimulus is encoded in the STSS, a time-stamp is updated to the current time and the trace-strength is set to the maximum value for the STSS memory stage. Memory entries in the STSS are regarded as having a very fast rate of decay. Because of this, entries in the STSS correspond to objects that are currently in view, or were in the agent's view very recently.

3.1.4 Short-Term Memory

Short-term memory is more limited than the STSS in terms of the number of memory entries that can be remembered. However, memory entries generally persist for a longer mean duration. Generally, memory entries are removed, or forgotten, from the STM under two conditions: they are displaced by newer memories when the STM is full and they may also decay over time. Rehearsal describes how entries may remain and persist for longer periods in the STM. As described later, these low-level values are intended to be controlled by other mechanisms, such as attention (Section 4).

3.1.5 Long-Term Memory

Long-term memory, or *LTM*, provides storage for information over a long period of time. For the purposes of the current model, memories that can no longer be retrieved due to lack of sufficient cues are considered to be the same as decayed entries. In LTM, as in other stages, such items decay over time and are completely removed when they fall below the threshold trace-strength.

Long-term memory is considered to have a *back and forth* relationship with short-term memory; items may be moved into short-term memory for processing and then moved back into long-term memory for storage. Information from short-term memory is transferred to long-term memory by means of rehearsal, related to attention mechanisms. In this model, repeated rehearsal of a stimulus will result in it being encoded in long-term memory. Unlike the other memory stores, there are no strict upper bounds on trace-strength in the LTM. If an item is frequently rehearsed in the STM, then it can remain easily accessible in the LTM.

3.2 Memory Dynamics

Given the memory structures described, memory dynamics deals with the movement of information between stages and when information should be removed from memory i.e. forgotten.

Eight different cases for memory updates are possible given the three stage model. Rules are based on the idea that objects in the STSS are, or were recently, visible, that objects in the STM are being attended to and objects in the LTM have previously been attended to and are currently in long-term storage. These rules are as follows:

1. If there is no memory entry for an object in the STSS, STM or the LTM, then it is presumed that the agent has no internal representation of the object whatsoever.
2. A memory entry that exists only in the STSS signifies that an object was recently within the agent's view field, but is not being attended to or rehearsed.
3. When a memory entry is only present in the STM, the agent is attending to it, even though it is no longer in the agent's view.
4. When an item is present both in the STSS and in the STM, then it is or was recently in the view of the agent and is also being attended to by the agent.
5. Memory entries that exist only in the LTM are subject to decay according to the decay function. The rehearsal parameter does not change since the agent is neither receiving new information about the object from the STSS, nor paying attention to it through the STM.
6. Memories that are present in both the STSS and LTM result in normal STSS decay. In the current model, LTM decay is paused while corresponding entries exist in the STSS. This presumes some sort of connection between the STSS and LTM that bypasses the STM. Although such a connection was not part of the original version of the model proposed in [1], the abandonment of strict serial information flow between each memory component is seen as a necessary modification to support experimental evidence (see [11]).
7. If a memory is present in both the STM and the LTM, then trace strength decay takes place in the STM, but not in the LTM. Rehearsal of the entry takes place since the entry is being attended to.
8. When memories exist for an object in all three stages, then the entry trace strength and rehearsal is increased. Object uncertainty is also reduced. This case represents a scenario where the agent is or has recently looked at a stimulus, is paying attention to it and has also previously rehearsed the stimulus and paid attention to it.

These rules form the basis of all updates that take place in the memory system. When used in conjunction with decay thresholds and rehearsal thresholds, they provide a way for items to move between memory stages so that only the most important environmental stimuli are retained in memory.

3.3 Forgetting

Forgetting is of vital importance to a computational model of memory for agents. Items that are considered to be remembered by the agent must be stored in the physical memory of the computer system, which is finite. Operationally, forgetting is implemented as a scheme for the removal of memory entries that are no longer of importance or relevance. In the current system, a simplified model of forgetting is implemented based on the *decay hypothesis* of retention [4]. Power functions illustrate the decline in the ability to recall information from memory as dependent on a *retention interval*, the amount of time since a memory was last entered or rehearsed. In this work, it is assumed that items in memory always have a probability of recall of 1. Trace decay is used only to determine when a memory entry should be forgotten, as oppose to how successfully information is recalled and to what extent. Memory entries are presumed to decay according to an exponential decay function, as used in the decline function in the short-term store of Atkinson and Shiffrin's buffer model [1]. Memory entries are removed from a stage when their trace strengths are reduced below the threshold trace value defined for the respective memory stage.

4 Visual Attention

Attention refers to the way in which some stimuli may receive enhanced cognitive processing, while others may not be processed to the same degree. As a mechanism, a sound attention strategy enables an entity with limited resources to survive in a dynamic and complicated environment by prioritising the consideration of information that is potentially important and relevant, inherently and with respect to current task. The basic visual attention system in use is a bottom-up model that deems areas of high basic visual feature contrast to be of inherent potential relevance to an entity. It has been described elsewhere [14] and is based on [8]. It would be complemented by the integration of a top-down model. Nonetheless, it serves to illustrate important interrelationships between memory and attention, the focus of this paper. In this case, the use of the object-based memory system is described for enabling the focus of attention to move around a scene, even when the viewer is mobile in the environment.

4.1 Inhibition of Return

Maxima in the saliency map, the primary output of the visual attention algorithm, correspond to those areas deemed by the model to grab the focus of attention. One may envisage an undesirable situation where a single location continually maintains the maximal level and therefore holds attention indefinitely. Inhibition of return, or IOR, is therefore usually implemented. It refers to the inhibiting of winning locations in the saliency map in order to ensure that the focus of attention may continue to move around and visit other locations of high value.

Since the saliency map is a spatial representation of a scene, IOR is usually also calculated based on 2D spatial coordinates [8]. In this case, an important problem arises that the spatial coordinates only remain valid if the viewpoint does not change e.g. the scene remains still. In the situation where the viewpoint moves, for example the agent walks forward, IOR locations stored during previous iterations can not be used. In this situation, the object-based memory system may instead be utilised to implement IOR. Rather than keeping track of attended-to *locations* i.e. 2D spatial coordinates, in this model, the synthetic vision system (see Section 3.1.2) resolves such locations to determine attended-to *objects*: a measure of the attention paid to an object is therefore considered, rather than spatial coordinates, for guiding the focus of attention around the scene.

4.2 Uncertainty and Scene Exploration

While inhibition of return operates over small time scales, the method can be extended in order to support more general exploratory behaviours. For example, an uncertainty value is stored with each object. This is a single metric stored with the memory entry describing the degree to which an agent has attended to it and could be thought of as representing the completeness and correctness of the agent's representation of the object in memory with respect to the actual object. High uncertainty values may be used to modulate the output of the visual attention mechanism, thus creating a loop between agent and environment. For example, an object with a high uncertainty value will attract the attention of the agent, causing the agent to orient towards it. This will result in the object being impressed into memory. As it remains there, the uncertainty level will decrease. Eventually, another object will win control of attention, the agent's gaze will be redirected to it and the process can continue. Thus the agent will have acquired the ability to explore the scene. Furthermore, scene uncertainty metrics, considering the mean uncertainty level over all visible objects, can be used to vary the arousal of the agent. This could be used to control, for example, the tendency of the agent towards environment exploration rather than other tasks. Highly uncertain environments, for example, may invoke more looking behaviours than environments that the agent has been familiarised with.

5 Conclusions

An outline of a model for storing incoming sensory information from the environment has been presented. Links between memory and attention have been briefly described: these mechanisms seem complimentary for a variety of functional reasons. Such a coupling allows IOR to be based on stored memory values for attended-to objects. Conversely, memory requires an attention mechanism for defining a sensible strategy for filtering and transferring entries between successive stages. In this respect, attention represents one possible candidate for controlling the values described in Section 3.1.1. While the attention mechanism described here is perceptual in nature and operates only in a bottom-up manner, the examples provided are only a few of the many possibilities that may be achieved. For example, a valid argument that has been made against independent memory systems for virtual agents is the cost of storage of multiple copies of the world database [6]. Robust filtering and forgetting mechanisms, based on a sound visual attention strategy, could greatly alleviate such problems, an important future issue for crowd simulation [7].

REFERENCES

[1] R.C. Atkinson and R.M. Shiffrin, 'Human memory: a proposed system and its control processes', *The psychology of learning and motivation: advances in research and theory*, **2**, 89–195, (1968).
[2] D. H. Ballard, 'Reference frames for animate vision', in *IJCAI'89: Proceedings of the 11th international joint conference on Artificial intelligence*, pp. 1635–1641, San Francisco, CA, USA, (1989). Morgan Kaufmann Publishers Inc.
[3] S. Chopra-Khullar, *Where to look? Automating certain visual attention behaviors of human characters*, PhD dissertation, University of Pennsylvania, January 1999.
[4] H. Ebbinghaus. Uber das gedachtnis, 1913.
[5] T. F. Evers and S. R. Musse, 'Building artificial memory to autonomous agents using dynamic and hierarchical finite state machine', in *Proc. of Computer Animation 2002*, pp. 164–169, (June 2002).
[6] M. Gillies, *Practical behavioural animation based on vision and attention*, PhD dissertation, University of Cambridge Computer Laboratory, 2001.
[7] H. Grillon and D. Thalmann, 'Simulating gaze attention behaviors for crowds', *Computer Animation and Virtual Worlds*, **20**(2/3), 111–119, (2009).
[8] L. Itti and C. Koch, 'Computational modelling of visual attention', *Nature Reviews Neuroscience*, **2**, 1–10, (2001).
[9] J. Kuffner and J.-C. Latombe, 'Fast synthetic vision, memory, and learning for virtual humans', in *Proc. of Computer Animation*, (May 1999).
[10] M.Y. Lim, R. Aylett, W.C. Ho, S. Enz, and P. A. Vargas, 'A socially-aware memory for companion agents', *The 9th International Conference on Intelligent Virtual Agents*, (2009).
[11] *The psychology of attention*, ed., H. E. Pashler, MIT Press, Cambridge, Massachusetts, 1998.
[12] C. Peters, 'Designing synthetic memory systems for supporting autonomous embodied agent behaviour', *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 14–19, (2006).
[13] C. Peters and C. O'Sullivan, 'Synthetic vision and memory for autonomous virtual humans', *Computer Graphics Forum*, **21**, 743–752, (2002).
[14] C. Peters and C. O'Sullivan, 'Bottom-up visual attention for virtual human animation', in *Proceedings of the 16th International Conference on Computer Animation and Social Agents (CASA)*, pp. 111–117. IEEE Computer Society, (2003).
[15] C. Siagian and L. Itti, 'Biologically inspired mobile robot vision localization', *IEEE Transactions on Robotics*, **25**(4), 861–873, (July 2009).