Keeping in Touch: Agents Reporting from Collaborative Virtual Environments

Brian Logan, Mike Fraser, Daniel Fielding, Steve Benford, Chris Greenhalgh

School of Computer Science and IT
University of Nottingham, Nottingham UK

Pilar Herrero

Languages, Systems and Software Engineering Department
School of Computer Science, Universidad Politécnica de Madrid, Spain

Abstract

Over the past decade, there has been a growing interest in using Collaborative Virtual Environments (CVEs) for on-line gaming, arts, performances and entertainment. One problem facing the users of such environments is that it can be difficult to keep in touch with events when they are unable to directly log in or when real-world demands mean that a user must step into and out of a virtual world during the course of a single session. In this position paper, we report work in progress on building personal news agents for CVEs which observe and report on what is happening in the virtual world. The agents make use of positional information, audio information and information about a user's awareness and focus of attention to reason about encounters between users (and other agents) in the world. We briefly describe a prototype system built using the MASSIVE-3 CVE system and the SIM_AGENT agent toolkit which performs a simple commentary task for a multi-agent scenario, and current work on extending this prototype to produce reports of simple social behaviour.

Introduction

Collaborative Virtual Environments (CVEs) are virtual worlds providing real-time 3D graphics, supplemented with text and/or audio to multiple users over computer networks (Benford et al. 2001). Over the past decade, there has been a growing interest in using CVEs for on-line gaming, arts, performances and entertainment (e.g. Ultima 2001; Benford et al. 2000; Greenhalgh et al. 1999). One problem facing the users of such environments is that it can be difficult to keep in touch with events when they are unable to directly log in. A user might be involved in a long-term persistent game or performance that takes place over several days, or might just wish to maintain contact with an on-line community. Furthermore, studies of social interaction in CVEs have shown that it is common for real-world demands to be interwoven with virtual events over the course of even a single session in a CVE, resulting in users stepping in and out of a virtual world (Bowers et al. 1996a; Bowers et al. 1996b). Thus, even when logged in, users may not be able to constantly keep in touch with events.

In this position paper, we report work in progress on building personal news agents for CVEs which observe and report on what is happening in the virtual world. These agents make use of positional information, audio information and information about user's awareness and focus of attention to reason about encounters between users (and other agents) in the world and to adapt their information gathering behaviour to the current environment. They could also form the basis of companion Non-Player Characters (NPCs) in a game which assist the user by gathering information about what other players are doing in other parts of the game world.

We briefly describe a prototype system built using the MASSIVE-3 CVE system and the SIM_AGENT toolkit for developing agents which perform a simple commentary task for a multi-agent scenario, and current work on extending this prototype to produce reports of simple social behaviour. We conclude with a discussion of open issues and planned future work.

Reporting Agents

Our overall approach is that one or more agents are sent into the virtual world to gather information on the user's behalf. The agents roam the world, locating and observing events of interest such as encounters between other users. An agent can also observe other agents and report on what they are doing. Agents try to infer something of the nature and potential importance of these events and then report back to the user next time the agent is accessed. Users can obtain reports the next time they connect to the environment or can access their agents through personal wireless technologies such as a wireless PDA or mobile phone.

While it would be possible, and perhaps easier, to obtain reports about events in the virtual environment from a disembodied agent or directly from the infrastructure of the environment, we believe it is important that the agent is visibly embodied in the world:

- One reason for embodying the agent is privacy: users may feel more comfortable if they know that they will not be invisibly monitored and the absence of avatars and agents can be taken as an indication that they have privacy.
- If other users or players can see an agent then they can tailor their actions accordingly. They can try to influence it, perhaps deliberately acting-up or directly approaching it if they wish to pass a message on to its user. They can, of course, try to avoid it.
- The presence of the agents themselves might make the world a more interesting and dynamic place.

One approach would be for each user to ‘own’ one or more agents which act on their behalf. For example, an agent might take control of the user’s avatar (graphical embodiment) when the user is logged out. However we can also imagine ‘free-lance’ commentator agents provided by the game itself which publish news reports (for example, on a WAP site) for players who are not currently logged in.

Keeping Up with Events

The simplest approach is for the agent to report what happens at the user’s last location in the world. The agent takes on the point of view (and possibly the avatar) of the user in the virtual world and reports interesting events at this location either periodically or when the user next logs in. However alternative approaches are possible: the agent could follow a particular player in a multi-player game or character in an on-line performance and report on what they do, or the agent could simply wander about, looking for ‘interesting’ events to report.

Events which are directly supported by the environment are relatively easy to report, for example, where a player went, which objects they picked up, which other players they talked to. However in more open-ended environments, many of the more interesting ‘events’ will be interpretations of sequences of the primitive events supported by the world. For example, many social interactions such as two players meeting at a particular location, an argument, or a confidential discussion require a degree of interpretation.

3.1 Inferring Social Behaviour

A key component of managing the interaction between users (or agents) in a CVE is their use of the space itself, i.e., by controlling their position, orientation, distance, etc. users are able to modify their interaction and communication. Awareness models support this process by providing different mechanisms to scope user’s interests and influence in a world (Benford 2001). In our current approach, our agents will not be able to interpret any speech that they overhear. Instead, they will exploit awareness models and other information to guess at the significance of events.

Of particular interest to us are the mechanisms of ‘focus’ and ‘nimbus’ that are provided by an awareness model called the spatial model of interaction (Benford and Fahlén 1993; Benford et al. 1997). *Focus* represents an observer’s interest in a particular medium. ‘‘The more an object is within your focus the more aware you are of it’’ (Benford and Fahlén 1993). Focus may be expressed as a region or as a function defined over space or may be based on the attributes (spatial or otherwise) of the other object. Focus is one half of the basis of awareness, the other half is provided by nimbus. *Nimbus* represents an observed object’s projection in a particular medium. ‘‘The more an object is within your nimbus the more aware it is of you’’ (Benford and Fahlén 1993). Nimbus can be used to distinguish between, for example, interrupting or shouting, normal conversation, peripherally available information and privacy or security restrictions. Like focus, nimbus may be expressed as a region or as a function over space or may be based on the attributes of the other object. Focus and nimbus are symmetric, reflecting the balance of power between speakers and listeners in the spatial model. Figure 1 shows how, even with only simple instantiations of focus and nimbus as simple bounding regions, a user R (receiver) can experience different levels of awareness of another T (transmitter), ranging from maximal awareness, through peripheral awareness to minimal awareness.

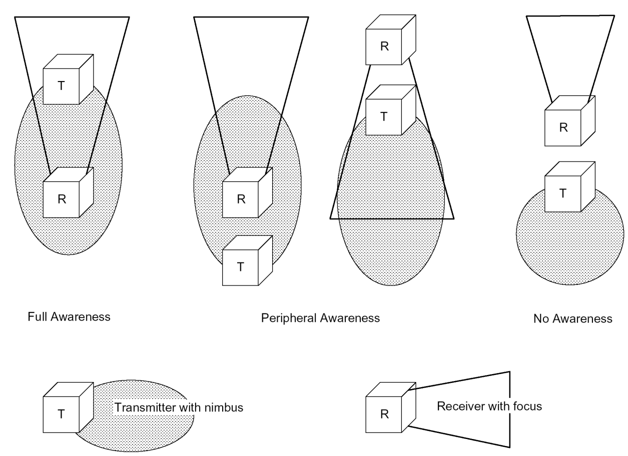

Figure 1: How R's focus and T's nimbus determine R's awareness of T

An agent can use focus and nimbus to reason about awareness and infer social information from interactions between people (via their avatars) in the world. For example, Benford et al’s work on ‘awareness states’ (Benford et al. 1994) considered all possible combinations of focus and nimbus that might arise between two participants and suggested social behaviours that these combinations could represent (e.g., ignoring, interrupting, overhearing, etc.). Using a similar approach, an agent might look at relative levels of focus and nimbus between participants in order to guess whether they are in direct conversation, whether one is overhearing another, trying to interrupt another, and so forth.

Although it cannot understand the semantics of the speech, the agent can also make use of information about audio energy (effectively the volume of the participants' audio), and the knowledge about the avatars' gestures to draw additional inferences about the nature of the interaction (e.g., whether an interaction is a conversation or an argument).

Using Awareness Models to Determine Action

The agent can also use awareness models to determine its own course of action in the world. For example, it might infer whether others are willing to be observed or whether they are trying to avoid being observed, and shape its own actions accordingly. In the former case, it may choose to position itself in among the action, providing maximum opportunity for users to interact with it. In the latter it needs to snoop by maintaining a vantage point such that its awareness of others is maximum while their awareness of it is minimum.

Implementation

To date, our main emphasis has been on the development of tools and techniques for the implementation and evaluation of agent-controlled avatars. This work is based on two existing systems: the MASSIVE-3 CVE system and the SIM_AGENT toolkit. MASSIVE-3 is a distributed multi-user virtual reality system that supports users communicating via a combination of 3D graphics and real-time packet audio. MASSIVE-3 relates to our particular domain of interest, in that it has previously been used in applications such as on-line drama (Craven et al. 2000), performance art (Benford et al. 2000) and virtual, 'inhabited' television (Greenhalgh et al. 1999). MASSIVE-3 has a number of features that make it particularly suitable for our work. One of these is the ability to record and replay events in the virtual world (Greenhalgh et al. 2000). The movements, speech and interactions of all of the avatars in a world can be captured in such a way that they can subsequently be injected back into a live world. The net result is that live participants see the recorded action recreated around them and are able to follow it around the world. Such recordings can facilitate the development of agents by providing standardised environments for testing and evaluation: different versions of an agent can be tested against exactly the same (recorded) social scenario.

SIM_AGENT is an architecture-neutral toolkit originally developed to support the exploration of alternative agent architectures (Sloman and Poli 1996; Sloman and Logan 1999). It can be run both as a pure simulation tool to simulate, e.g., software agents in an Internet environment or physical agents and their environment, and as an implementation language for, e.g., software agents or the controller for a physical robot. In particular, SIM_AGENT makes no assumptions about the embodiment of agents. SIM_AGENT has been used in a variety of research and applied projects, including studies of affective and deliberative control in simple agent systems (Scheutz and Logan 2001) and simulation of tank commanders in military training simulations (Baxter and Hepplewhite 1999) (for this project, SIM_AGENT was integrated with an existing real time military simulation).

The client-server architecture of MASSIVE-3 and the modular design of the sensor and action methods in SIM_AGENT means that integrating the two systems is relatively straightforward. The main tasks are to provide a means for a SIM_AGENT process to connect to MASSIVE-3 as a client, and to override the default SIM_AGENT sensor and action methods so that the sensor methods obtain their information from the MASSIVE-3 environment and the action methods update the MASSIVE-3 environment.

Prototype Commentator Agent

We have implemented a first simple agent that performs a static commentary task for an existing SIM_AGENT application, herding sheep (Carter 1999).[1] The sheepdog simulation consists of five sheep and a dog, all simulated as distinct agents within a single SIM_AGENT process. The agents have 'subsumption-style' architectures, implemented as collections of rules, and the aim is for the dog to herd the sheep into a pen by exploiting their innate flocking and avoidance behaviours. Translating the sheepdog simulation into the MASSIVE-3 environment required defining appropriate embodiments for the sheep, the dog, the pen and the obstacles and splitting the simulation so that sheep and the dog were simulated in different SIM_AGENT processes on different machines, communicating via the MASSIVE-3 environment. With the exception of the interface code which overrides the default SIM_AGENT sensor and action methods, the original sheepdog simulation runs unchanged in the MASSIVE-3 environment. The 'Commentator' agent is also implemented in SIM_AGENT and is embodied in the MASSIVE-3 environment. The Commentator agent observes the actions of the other agents (and indirectly interacts with them via their collision avoidance behaviours), and reports the speed and heading of the

1 Although this may sound like a bizarre entertainment application, it should be noted that for many years in the UK there was a mainstream television program called 'One Man and His Dog' in which an expert commentated on real sheep herding competitions.

sheep and the dog (and anything else moving around in the world). The position and direction of view of the commentary agent can be controlled by the user.

Figure 2 shows a sceenshot of our commentator in action. In this case it is reporting a key event (the presence of a sheep object "sheepwhite.dgl" at a particular coordinate) via a text string that appears above its head like a speech balloon.

Figure 2: Commentating on virtual sheep herding

The following is an extract from a longer externally reported commentary:

```
** [Commentator tracks sheepwhite.dgl]
<sheepwhite.dgl moved from 0.979018 -7.07855 to
-0.269416 -11.4324>
< -- speed is 4.52926, direction 254.0 >
** [Commentator tracks shepherd]
<shepherd moved from -2.15641 -0.186817 to
2.00275 -3.51535>
< -- speed is 5.32708, direction 321.33 >
** [Commentator tracks grid.dgl]
<grid.dgl did not move>
** [Commentator tracks embod_humanoid.dgl]
<embod_humanoid.dgl did not move>
** [Commentator tracks sheepwhite.dgl]
<sheepwhite.dgl moved from 10.1645 -5.88389 to
14.1538 -7.02783>
< -- speed is 4.15016, direction 344.0 >
** [Commentator tracks sky.dgl]
<sky.dgl did not move>
** [Commentator tracks sheepwhite.dgl]
<sheepwhite.dgl moved from 8.20874 -8.58052 to
11.4719 -11.223>
< -- speed is 4.19894, direction 321.0 >
```

Clearly this first news agent is very simple, demonstrating only a basic ability to observe the presence, location and speed of entities in the world, including sheep ("sheepwhite.dgl"), the human-controlled sheepdog, who also captured the screenshot ("shepherd"), the commentator itself ("embod_humanoid.dgl"), the floor ("grid.dgl") and the sky ("sky.dgl").

Reporting on Encounters

Building on this platform, we are currently working on an agent which follows a nominated character in a multi-user CVE and reports on their interaction with objects and other avatars. The agent has basic tracking, collision avoidance and reporting capabilities. It attempts to remain 'behind' and facing towards the target character, and reports any interactions with other characters or objects.

We are particularly interested in situations where the target joins or leaves a group of other characters. Recognition of these events is based on the sociologist Adam Kendon's studies of encounters between small groups of people in open space (Kendon 1990). Kendon argues that people consistently exhibit recognisable spatial and behavioral formations called 'F-formations' when joining, sustaining and leaving groups, and notes that ''the F-formation system provides us with an excellent means of defining a social encounter as a unit for analysis''. We have recently developed a commentary agent that, although incomplete, can estimate the transactional segments of all the users that are in its field of vision, and using this information it can recognise which users form F-formations. To test this approach, we are using 3D recordings of an on-line drama called 'Avatar Farm' that involved eleven participants, spanned two hours and that was staged in MASSIVE-3 (Drozd et al. 2001). Our aim is to build a more sophisticated agent that can report on key scenes and activities in this drama.

We envisage that this basic tracking agent will form a framework for the development of more capable agents able to recognise and appropriately report on more complex social encounters.

Related Work on Recognising Encounters

The general task of reporting 'interesting events' is very broad, and spans a wide range of topics in AI and CSCW, including plan recognition, user modelling, notions of context and so on, and there is a wealth of related work. However, our current objective of reporting on the social interactions of an avatar with other avatars in a CVE is less well studied. In this case, the agent is working with information concerning the pragmatics of social interaction rather than the semantics. Previous work in natural language processing has tended to focus on disembodied agents and simple dialogues, and so has been unable to exploit spatial and/or awareness information. However, some work in the CVE community has addressed these issues and may give us additional ideas as to how we might infer social behaviour from awareness models:

- There is a small body of related work on automating camera views in virtual environments, to find 'interesting' shots. For example, Bowers et al. (Bowers et al. 1999) describe an automated virtual cameras for inhabited television. The 'Puppycam' would try to follow interesting action around a virtual world by looking for areas where there was a high density of focus (i.e., where several participants were focusing at the same point in space). Related work in (Drucker and Zeltzer 1994) describes techniques for providing intelligent and interesting camera views in virtual worlds.
- Related observations by Jeffrey and Mark (1999) describe encounters between avatars in the Alphaworlds graphical/textual CVE, drawing out observations about the role of personal distance in virtual worlds.

Further Work

So far, our main emphasis has been on the development of tools and techniques for the implementation and evaluation of agent-controlled avatars. In future work, we hope to extend the basic framework outlined above in a number of different ways including: reasoning about how to collect information, including the effects of the agent's actions on the perceptions and actions of users; and inferring more complex social behaviour and the beliefs and goals of users and agents. In addition, the fact that agents are embodied raises a number of design issues that we have yet to resolve:

- should the agents be clearly distinguishable from human participants?
- should your agent reveal your identity (assuming a model of personal ownership)?
- should an agent display appropriate reactions to what it observes?

Finally, we require methods for evaluating the effectiveness of different agents as social commentators. With such open-ended tasks, it is difficult to devise precise measures of agent performance. However a number of factors can be used to assess the quality of the reports generated by the agent, including how successful the agent is in obtaining information while satisfying various constraints, how useful or relevant the information it obtains and the inferences it draws about social interactions are judged to be, and how successfully the agent can communicate this information in different output or display formats. We can also perform a comparison of users' categorisation of encounters with those of the agents. One such test might be whether the agent can tell the difference between an autonomous agent and a human-controlled avatar using only awareness based observations. Initial evaluation will exploit the 3D record and replay capabilities of MASSIVE-3, using recordings of specially generated simple social situations to test the tracking and reporting capabilities of the basic agent. These can be replayed in the presence of different agents with different rules for inferring social behaviour and the results compared with each other and with those of expert and non-expert human ratings.

References

Baxter, J. and Hepplewhite, 1999. Agents in Tank Battle Simulations. *Communications of the ACM*, 42 (3), pp. 74-75: ACM Press.

Benford, S. and Fahlén, L. 1993. A Spatial Model of Interaction in Virtual Environments, in *Proceedings of the Third European Conference on Computer Supported Cooperative Work (ECSCW'93)*, Milano, Italy: Kluwer

Benford, S., Greenhalgh, C., Rodden, T. and Pycock, J. 2001. Collaborative Virtual Environments, *Communications of the ACM*: ACM Press.

Benford, S., Bowers, J., Fahlen, L. E., Mariani, J. and Rodden, T. 1994. Supporting cooperative work in virtual environments. *The Computer Journal,* 37(8), pp. 653-668: Oxford University Press.

Benford, S. D., Greenhalgh, C. M. and Lloyd, D. 1997. Crowded Collaborative Virtual Environments, in *Proceedings of 1997 ACM Conference on Human Factors in Computing Systems (CHI'97)*, pp.59-66, Atlanta, Georgia, US: ACM Press.

Benford, S., Reynard, G., Greenhalgh, C., Snowdon, D. and Bullock, A. 2000. A Poetry Performance in a Collaborative Virtual Environment, in *IEEE Computer Graphics and Applications*, 20 (3), pp.66-75: IEEE

Bowers, J., Pycock, J. and O'Brien, J. 1996a. Talk and Embodiment in Collaborative Virtual Environments, in *Proceedings of ACM Computer-Human Interaction (CHI'96)*: ACM Press.

Bowers, J., O'Brien, J., and Pycock, J. 1996b. Practically Accomplishing Immersion: Co-operation in and through Virtual Environments, in *Proceedings of ACM Computer-Supported Cooperative Work (CSCW'96)*: ACM Press

Bowers, J., Jää-Aro, K-M., Hellström, S-O. and Carlzon, M. 1999. Event Management in Electronic Arena by Visualising Participants Activitiy and Supporting Virtual Camera Deployment, in eRENA project Deliverable 4.3/4.4, ESPRIT project 25379, European Commission

Carter, T. 1999. "An architecture for an intelligent sheep-dog in Pop-11", MSc thesis, Cognitive Science Research Centre, University of Birmingham.

Craven, M., Benford, S., Greenhalgh, C., Wyver, J., Brazier, C.-J., Oldroyd, A. and Regan, T. 2000. Ages of Avatar: Community Building for Inhabited Television, in

Proceedings of the Third ACM Conference on Collaborative Virtual Environments (CVE 2000), San Francisco, CA, USA, pp. 189-194: ACM Press

Drozd, A., Bowers, J., Benford, S., Greenhalgh, C. and Fraser, M. 2001. Collaboratively Improvising Magic: An Approach to Managing Participation in an On-Line Drama, in *Proceedings of ECSCW '0*1, Bonn, Germany, pp. 159-178: Kluwer

Drucker, S. M. and Zeltzer, D. 1994. Intelligent Camera Control in a Virtual Environment, in *Proceedings of Graphics Interface '94*, pp. 190-199.

Greenhalgh, C., Benford, S., Taylor, I., Bowers, J., Walker, G. and Wyver, J. 1999. Creating a Live Broadcast from a Virtual Environment, in *Proceedings of ACM SIGGRAPH'99*, pp.375-384: ACM Press

Greenhalgh, C., Purbrick, J., Benford, S., Craven, M., Drozd, A. and Taylor, I. 2000. Temporal links: recording and replaying virtual environments, in *Proceedings of the 8th ACM international conference on Multimedia (MM 2000)*, pp. 67-74: ACM Press

Jeffrey, P. and Mark, G. 1999. Navigating the Virtual Landscape: Co-ordinating the Shared Use of Space. *Social Navigation of Information Space*, A. Munro, K. Höök and D. Benyon (eds.): Springer.

Kendon, A. 1990. Spatial organization in social encounters: the F-formation system. In A. Kendon, *Conducting Interaction: Patterns of Behavior in Focused Encounters*, Cambridge: Cambridge University Press.

Scheutz, M. and Logan, B. 2001. Affective vs. Deliberative Agent Control, in *Proceedings of AISB'01 Symposium on Emotion, Cognition and Affective Computing*, pp. 1-10: AISB.

Sloman, A. and Poli, R. 1996. SIM_AGENT: A toolkit for exploring agent designs, in Intelligent Agents II; in *Proceedings of the Second International Workshop on Agent Theories, Architectures and Languages*, pp. 392-407: Springer-Verlag.

Sloman, A. and Logan. B. 1999. Building Cognitively Rich Agents Using the SIM_AGENT Toolkit. *Communications of the ACM*, 42 (3), pp. 71-77: ACM Press.

Ultima Online, http://www.uo.com, verified 26/09/2001

Authors CVs

Brian Logan is a lecturer in Computer Science at the University of Nottingham. His research interests lie in the area of agent-based systems, and include agent architectures, logics and ontologies for agent-based systems and software tools for building agents. In addition to the work described in this paper, current projects include 'Intelligent Agents for Believable Worlds' sponsored by Sony Computer Entertainment Europe. Recent publications include Communications of the ACM and Proceedings of the IEEE.

Mike Fraser is a lecturer in Computer Science at the University of Nottingham. Recent publications include a chapter in the book 'Collaborative Virtual Environments', and ACM papers in Transactions on CHI, Interactions magazine, and at the CHI, UIST, UBICOMP, CSCW and CVE conferences.

Dan Fielding is a research student at the University of Nottingham where he is studying for a PhD in the area of integrating agents and avatars for CVEs.

Steve Benford is Professor of Collaborative Computing at the University of Nottingham where he is a co-founder of the Mixed Reality Laboratory [www.mrl.nott.ac.uk]. His research interests include collaborative virtual environments and mixed reality. Recent publications include Communications of the ACM, ACM Transactions on CHI and at the CHI and CSCW conferences.

Chris Greenhalgh is a Reader in Computer Science at the University of Nottingham. His research interests include distributed systems support for CVEs and mixed reality and he is the principal architect of the MASSIVE system. Recent publications include Presence, IEEE transactions of Graphics and SIGGRAPH.

Pilar Herrero is an Assistant Professor in the Languages, Systems and Software Engineering Department at the Universidad Politécnica de Madrid. Her research interests are in agents, avatars and the use of awareness models as part of agent perception.