

intelligent multimedia @ VU

www.cs.vu.nl/~eliens/research/note-intelligent-multimedia.html

project leader: dr A. Eliëns
researcher: dr. Z. Huang
programmer: drs. C. Visser
student: M. Hildebrand

project description

We are developing a high-level platform for 3D and rich media virtual environments based on agent-technology, using the languages DLP¹, Java, and VRML. On top of this platform we have developed an scripting language STEP² for specifying humanoid movements and gestures, based on dynamic logic.

Our goal is to study aspects of the deployment and architecture of virtual environments as an interface to multimedia information systems³.

Our platform also supports embodied conversational agents, see Eliens *et al.* (2002). As demonstrators we have developed a distributed soccer-game prototype with intelligent autonomous avatar-embodied agents as players Huang *et al.* (2002), a humanoid animation demonstrating Tai-Chi, Huang *et al.* (2003), avatars presenting a dialog in a mixed media presentation environment, Eliens *et al.* (2003), a domestic agent that can be addressed in natural language, Hildebrand *et al.* (2003), an avatar reaching for objects that uses reasoning and inverse kinematics, Reach, and an avatar conducting music, Ruttkay *et al.* (2004).

As a student project in the Casus Practicum⁴, we have developed a 3D presentation for INCCA⁵, the International Network for the Conservation of Contemporary Art, in cooperation with ICN⁶, the Dutch Cultural Heritage Institute.

funding

Our research has been supported by two NWO projects:

- WASP⁷ – Web Agent Support Program
- RIF⁸ – Retrieval of Information in virtual worlds using Feature detectors

The combined effort of these projects led to the DLP+X3D platform and the development of the STEP language.

In addition, we have one pending proposal with Michiel Hildebrand as candidate researcher:

¹www.cs.vu.nl/~eliens/dlp

²wasp.cs.vu.nl/step

³www.cs.vu.nl/~eliens/research

⁴www.cs.vu.nl/~eliens/casus

⁵www.incca.org

⁶www.icn.nl

⁷www.cs.vu.nl/~eliens/research/wasp.html

⁸www.cs.vu.nl/~eliens/research/rif.html

- (submitted) IMMEDIATE⁹ – Intelligent MultiMEDIA Transactions Environment

cooperation

The work on *embodied conversational agents* is being done in cooperation with dr. Z. Ruttkay from CWI. The work on the *cultural heritage* application is done in cooperation with drs. T. Scholte from ICN.

publications

Apart from publications in international conference proceedings and workshops, we have demonstrated our work on the ICT Kenniscongres 2002 in the theme *intelligent multimedia*¹⁰.

In preparation is a prestigious book on *Life-like Characters*, for which we have contributed a chapter describing our platform and STEP, STEP-book. The editor of this book, Helmut Prendinger from Tokyo University, commented on our chapter: *I really have to say that your chapter is very very well done, very interesting, very comprehensive, and very readable. It says exactly the things people want to know when they are looking for some scripting language to animate their characters. Above that, your STEP system is very well motivated and nicely embedded in other strands of computer science research.* .

a brief history

Although the WASP proposal was written in 1996, long before the RIF proposal (1998), dr. Z. Huang started as a post-doc on the WASP project when the RIF project ran already for more than six months. Since we then felt that the WASP project proposal was slightly outdated, we made an effort to merge the WASP and RIF projects, by focussing on agents in 3D virtual environments which resulted in a paper presenting a taxonomy of Web agents, Huang *et al.* (2000). However, after about a year the continuity of the RIF project was endangered due to a mutation of personell at CWI. So, NWO was asked for permission to utilize the RIF funds for prolonging the WASP project. This was granted, provided that the research goals of the RIF project were sufficiently covered within WASP.

In the RIF project we used the blaxxun Community Server and VRML to realize information retrieval and delivery in multi-user 3D environments. See Navigation. Agents were then conceived as extensions on the server-side using blaxxun's native agents enriched with embedded logic. However, at the same time that the RIF project funds were transferred to WASP, the first prototype of DLP in Java became available, and we decided, somewhat radically, to drop the more low-level blaxxun technology in favor of a unified approach in DLP. Thus, we extended DLP with primitives for manipulating VRML worlds, using the Java External Authoring Interface that is part of VRML. The DLP+VRML framework proved to be surprisingly effective, as testified by the following references:

⁹www.cs.vu.nl/~eliens/research/immediate.html

¹⁰www.cs.vu.nl/~eliens/research/ict.html

- architecture for agents in virtual environments – Architecture
- intelligent avatars in 3D virtual worlds – Avatars
- demonstrator: multi-user soccer game – Huang *et al.* (2002)

Now, the language DLP itself has quite a long history, Eliens (1992). It has also been described in Eliens (2000). As a language, DLP offers an object-oriented extension of (traditional, Edinburgh-style) Prolog, with multi-threaded objects, non-logical instance or state variables, communication by rendez-vous and (distributed) backtracking. After preliminary prototypes in C++, we focussed on an implementation in Java, to be able to use DLP for Web programming, WebComputing. The first Java implementation became available relatively late, but just in time to create the DLP+VRML extension when needed.

As concerns the acceptance of our approach within the Web3D community, we wish to point to the acceptance of our Huang *et al.* (2002) paper for the highly competitive international Web3D Conference 2002 (acceptance: 1 out of 13), and the acceptance of STEPIMP for the Web3D 2003 Conference. More recently, we have made an effort to publish in the ECA (Embodied Conversation Agents) community, as testified by our contribution to the *Life-like Agents* book, STEP-book. We also received an invitation for the Dagstuhl seminar on *Evaluating Embodied Conversational Agents* in March 2004.

embedding in education: focus on multimedia

The *intelligent multimedia* research has a strong impact on the educational activities for the specialisation(s) Multimedia with Computer Science and Multimedia and Culture for Information Science.

In the first year students start with a general Introduction to Multimedia¹¹. This course centers around three themes: the convergence between media, platforms and delivery technology, the availability of broadband communication and its impact on the development of standards such as MPEG-4, and multimedia information retrieval as an essential ingredient of the growing multimedia information repository on the Web.

There are two follow-up courses, which are given in respectively the second and third year:

- Multimedia Authoring I – Web3D/VRML¹²
- Multimedia Authoring II – Virtual Environments¹³

The first of these courses deals with the technology for creating 3D scenes and worlds, whereas the second is more focused on providing intelligent services in virtual environments. Students use the DLP+VRML framework for their assignment in the second course. See Intelligent.

¹¹www.cs.vu.nl/~eliens/imm

¹²www.cs.vu.nl/~eliens/mma1

¹³www.cs.vu.nl/~eliens/mma2

In addition for Multimedia and Culture there is a Multimedia Development Casus Practicum¹⁴ in which the technology is applied in an assignment developed with the Dutch Cultural Heritage Institute (ICN).

For both specialisations, Multimedia and Multimedia and Culture we plan to offer a course on XML-based Multimedia Technology¹⁵, to be developed by dr. Z. Huang, to make students familiar with advanced topics in XML-based information processing. As a remark, our platform does already support the use of XML and XSLT stylesheets, XSTEP, and we are migrating to a DLP+X3D platform, with X3D¹⁶ as the XML-based successor of VRML.

research directions

Apart from the issues involved in the modelling and realization of embodied agents in rich media 3D environments, there are also issues with regard to the architecture and implementation of our DLP+X3D platform.

parallelism and synchronization Complex humanoid gestures are of a highly parallel nature. The STEP scripting language supports a direct way of modelling parallel gestures by offering a parallel construct (*par*), which results in the simultaneous execution of (possibly compound) actions. To avoid unconstrained thread creation, the STEP engine makes use of a thread pool, containing a fixed number of threads, from which threads are allocated to actions. Once the action is finished, the thread is put back in the pool. This approach works well for most examples. However when many threads are needed, as in the conductor example (which requires approximately 60 threads), problems may occur, in particular when there are many background jobs.

modeling and representation Our agent model may be characterized as a BDI-model, extended with *sensors* and *effectors* needed for the interaction in a virtual environment. The STEP scripting language has been developed to facilitate the specification of communicative acts, like gestures. However, we would also like to explore text-to-speech synthesis as an extra modality of communication.

One interesting research issue is how to specify a reusable library of gestures, accomodating for differences in (personal) style. This is currently being investigated by Z. Ruttkay from CWI.

Another interesting issue is the use of inverse kinematics to grasp objects. However, when an object is not within reach, the agent has to reason about the best way to get near to the object, to be able to reach it.

architecture and implementation To solve the problem of reliable timing would require not only a modification of the STEP engine, but also a rather different implementation of the DLP threads supporting the parallelism in STEP.

¹⁴www.cs.vu.nl/~eliens/casus

¹⁵www.cs.vu.nl/~eliens/mm/course-xml.html

¹⁶www.web3d.org

Currently, the implementation only allows for *best effort* parallelism and does not provide the means for *deadline scheduling*.

However, it is our impression that we have reached the utmost efficiency feasible within the Java platform. Therefore we have been considering to redevelop the DLP+X3D platform in a .NET environment¹⁷. An additional advantage of migrating to the .NET environment would be the possible integration of functionality such as *text-to-speech* synthesis which is not readily available in the Java environment.

¹⁷www.microsoft.com