

IMMERSIVE INTERACTIVE NARRATIVES

*The art.live consortium, IST project 10942**

c/o Xavier Marichal

Laboratoire de Télécommunications et Télédétection - Université catholique de Louvain

2 place du Levant, B-1348 Louvain-la-Neuve, Belgium

E-mail: marichal@tele.ucl.ac.be

ABSTRACT

The goal of the *art.live* project is to develop **an architecture and a set of tools**, both generic and application dependent, for the enhancement of narrative spaces thanks to the production of a mixed-reality environment. Having in mind the mixed-reality framework (as defined by Milgram in "A Taxonomy of Real and Virtual World Display Integration", in *Mixed Reality - Merging Real and Virtual Worlds*, 1999, Ohmsha Ltd. and Springer-Verlag), the *art.live* architecture aims at **creating narrative spaces**, e.g. interactive stories, that mix graphical elements with live inputs of cameras. The real impact of the system occurs when some persons are in front of the cameras: they get themselves immersed within the visual ambiance and they are therefore involved within the narrative, which they are able to **interact** with through their behavior. At the opposite side of the architecture, other persons are looking at screens where the mixed images are rendered and they might be offered to interact with the system.

Keywords: Real-time immersion IP-compliant, video segmentation, interactive stories, MPEG-4, MPEG-7, FIPA

1 INTRODUCTION

The present paper elaborates mainly on the technical system that will, through an authoring tool, offer multimedia authors the capacity of designing immersive interactive narratives involving real people into their own universe of pictures, graphics and associated designs. Section 2 provides the reader with a global view of the system, while section 3 outlines the implemented architecture. Section 4 focuses on one of the novel aspects of the *art.live* approach: the management of such interactive scenarios, whose

implementation strongly relies on standards as exposed in section 5.

2 SYSTEM OVERVIEW

Considering a human-centric approach, the various "users" depicted on Figure 1 are involved within the designed architecture.

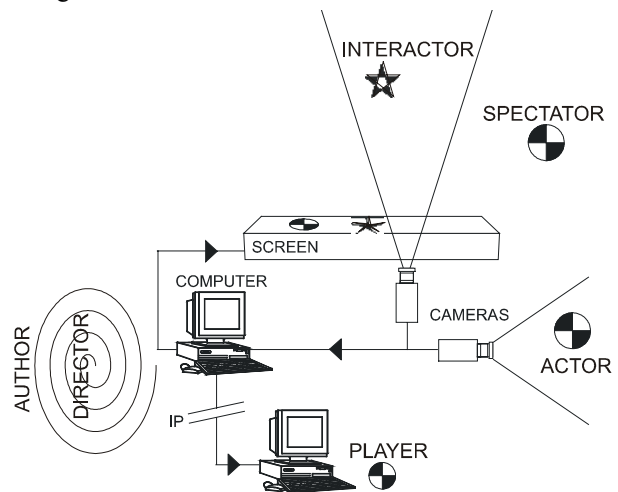


Figure 1: Repartition of users in the system

They are ranked here by their degree of influence on the overall system:

1. The **Author**, who designs the whole narrative system, i.e. the scenario, the related story and its elements (musical analogy to composer or theatre/film analogy to the scriptwriter);
2. The **Director**, who can modify (via the authoring tool) some aspects that the author prepared to be modifiable (musical analogy to performer or theatre/film analogy to the director achieving the mise-en-scene);

* The eight following partners constitute the art.live consortium: *Université catholique de Louvain*, Louvain-la-Neuve, Belgium; *Casterman Edition S.A.*, Bruxelles, Belgium; *Associação para o desenvolvimento das telecomunicações e técnicas de informática*, Lisboa, Portugal; *Ecole Polytechnique Fédérale de Lausanne*, Switzerland; *Fastcom Technology S.A.*, Lausanne, Switzerland; *Association pour le Développement de l'Enseignement et de la Recherche en Systématique Appliquée*, Paris, France; *Université Joseph Fourier Grenoble 1*, France; *Centro Studi e Laboratori Telecomunicazione S.p.A.*, Torino, Italy.

3. The **Consumer-Interactor**, who is captured by some live camera, and can directly interact with the system via its gesture. The Interactor is aware of his/her role in the narrative thanks to some (large) screen where s/he sees himself/herself within the mixed-reality environment;
4. The **Consumer-Player**, who interacts with the system through a mouse on a Web browser (clicking on some MPEG-4 hypervideo);
5. The **Actor**, who is any person in front of some live camera. The Actor is not aware of his/her status within the system;
6. The **Spectator**, who is any person looking at the images without interacting or being captured by the cameras.

On the content point of view, the *art.live* system considers various types of objects:

- **Artificial Objects (AO)**. They have been artificially created by an artist: typically they are produced by a graphical designer thanks to a computer, e.g. pictures, drawings, graphical animations, and so on;
- **Real Objects (RO)**. They exist in the real world: typically they are extracted from a real scene and separated from the background;
- **Virtual Objects (VO)**. They are perceived through a media, being:
 - **Synthetic virtual Objects (SO)**, when they represent some artificial objects in the virtual space (e.g. a particular picture becoming the background of the mixed-reality scene),
 - **Natural virtual Objects (NO)**, when they represent some real objects that entered the actual space (typically the persons, actors and interactors, in front of the system cameras).

In this framework, all the users are proposed to participate (at their own level) into a scenario, a narrative, made of the previously mentioned graphical elements.

In order to be able to offer a rich environment to authors, the system currently under development relies on the following choices:

- an open (notably through the use of standards) and scalable architecture, open to emerging and future technologies like, for instance, new cameras;
- a WWW-compliant system, notably with communications based on the Internet Protocol;
- the inclusion of database aspects that facilitate the work of the creator/author;
- an authoring tool to pilot the system.

Moreover, through public demonstrations and trials, the project aims not only at technological developments but also at tackling the narrative aspects, where the sole goal is to offer a “meaningful experience to the user”. This means allowing the consumer to have some pleasure in following a story that is to be provided to him/her. This requires the definition of ad hoc narrative patterns.

3 SYSTEM ARCHITECTURE

At a high level of abstraction, the *art.live* objective of capturing real-life objects and including them into mixed-reality worlds can be described as a combination of two generic processes: Computer Creation of Natural Objects, and Human Creation of Synthetic Objects.

In the Computer Creation of Natural Objects, one or more sensors capture the reality, typically under the form of audio and video. “The computer” then creates a virtual reproduction of the real objects and characterizes them, i.e. it **creates** an automatic **simulation** and interpretation of the real world. Finally, some **visualization** of this simulated world is provided. This visualization can motivate or not some changes in the reality that is caught by the sensors. Then this process starts again. It is important to note that this process is data driven in the sense that from a set of real data (implicit model of the world), it attempts to discover the explicit model.

In the Human Creation of Synthetic Objects, the reality is also captured by one or more sensors, like keyboards and mouse, specially designed to interact with individuals. Here the **simulation** begins with a set of editors that help humans to reproduce and characterize their creations and dreams. These synthetic objects can then be digitized into the computer, **visualized** and can stimulate, or not, some desired interaction. This makes the human creation process start again. Such a process is model driven, in the sense that it transforms a set of human modeled objects into computer data. These models are implicitly (ex. set of draws) or explicitly defined by **describing** their essential properties.

Both processes are relatively classical and well mastered in their own. The major challenge consists therefore in combining them to create meaningful mixed-reality scenes. The main assumption is to consider that these processes are very complementary: they are both directly involved into a generic and repetitive process of

- 1) sensors **capture** of reality (which includes human and computer creations),
- 2) **simulation** (virtual or mixed realities, objects representation and characterization),
- 3) **visualization**.

When it is considered as an endless iterative process, one obtains a combined process of a specific **living art** creation permanently inspired by reality.

The concrete implementation of the system architecture relies on the following dedicated tasks:

- Human Creation and its Edition through Authoring Tool (artistic creation of SO and design of the narrative);
- Computer Creation: Natural Objects Extraction (automatic segmentation and tracking of natural objects);
- Low-Level Description (2D & 3D features extraction, notably through camera calibration);
- Interpretation and Triggering of events according to scenario state and Scenario Management accordingly;
- Scene Composition according to the scenario status & Rendering (visualization with possible interaction).

These tasks are complemented by some “horizontal” ones to obtain a fully integrated system:

- Information Management for the Storage & Retrieval of all elements (objects, scenarios...);
- MPEG-4 Transmission and Coding of natural and synthetic images (cf. section 5);
- MPEG-7 descriptors and combination of these for triggering (cf. section 5);
- (Secure) Communications and Coordination within a multi-agents platform (cf. section 5);
- Camera Integration for the sake of the real-time constraints

4 SCENARIO MANAGEMENT

Any *art.live* narrative is established as a (narrative) **graph**. Every node of the graph provides a **scene** (composed by various virtual objects) along with a list of **events** to be triggered. For instance, the author may decide to look for a NO being a person, to ‘touch’ a particular SO which is a door, to detect two NO moving ‘fast’ in opposite directions, or simply to wait for 15 seconds. According to the detected **trigger**, an **action** occurs. The action is a move to some next node of the graph, where another scene is depicted and other triggers are searched for. The scene can be a completely new one or the same one from the previous node with just some additional (or suppressed) graphical element (scene refresh or update).

Figure 2 presents an excerpt from a simplified typical scenario (where one should imagine that more elements like moving people and graphical objects are added to the scenes).

It is crucial to note that the evolution of the narrative is different for every screen of the system, i.e. for any player or for any set of interactors in front of the same big screen. Narrative graphs are thus managed in an autonomous way in order to allow different users to enjoy the same story at different moments and with different interactions.

5 USE OF STANDARDS TO CREATE MEANINGFUL MIXED-REALITY SCENES

As it has already been stated previously, standards are used in order to implement such a scenario management in an open and flexible way. Figure 3 shows how the **MPEG-4**, **MPEG-7** and **FIPA** standards are integrated within the architecture.

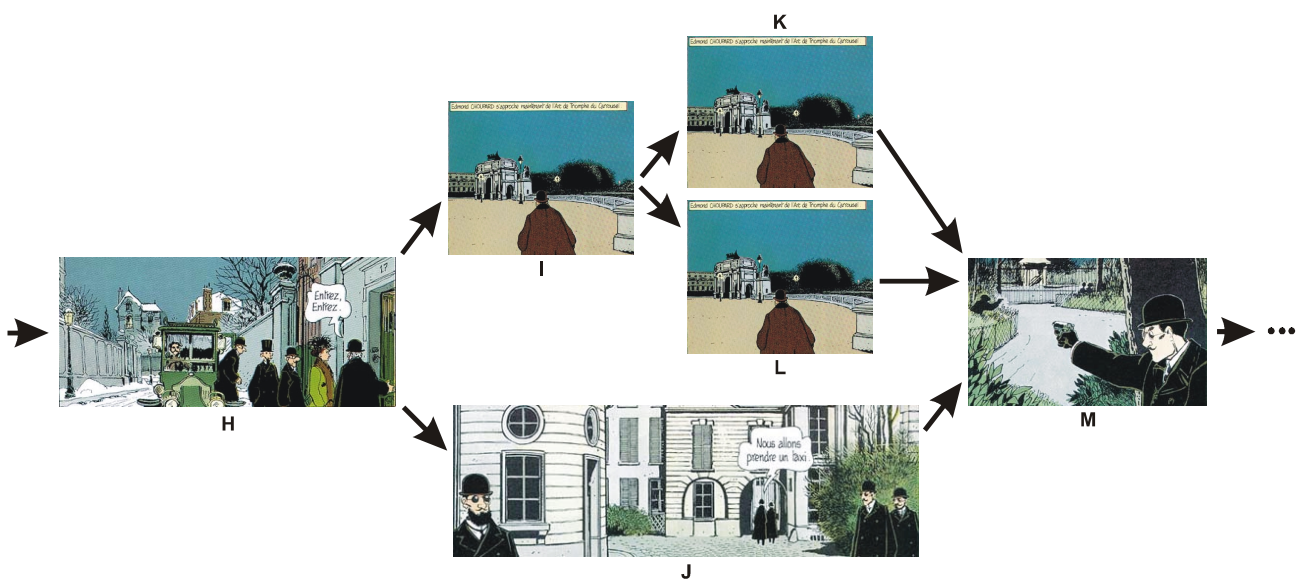


Figure 2: Example of narrative graph (images are © Casterman-Tardi)

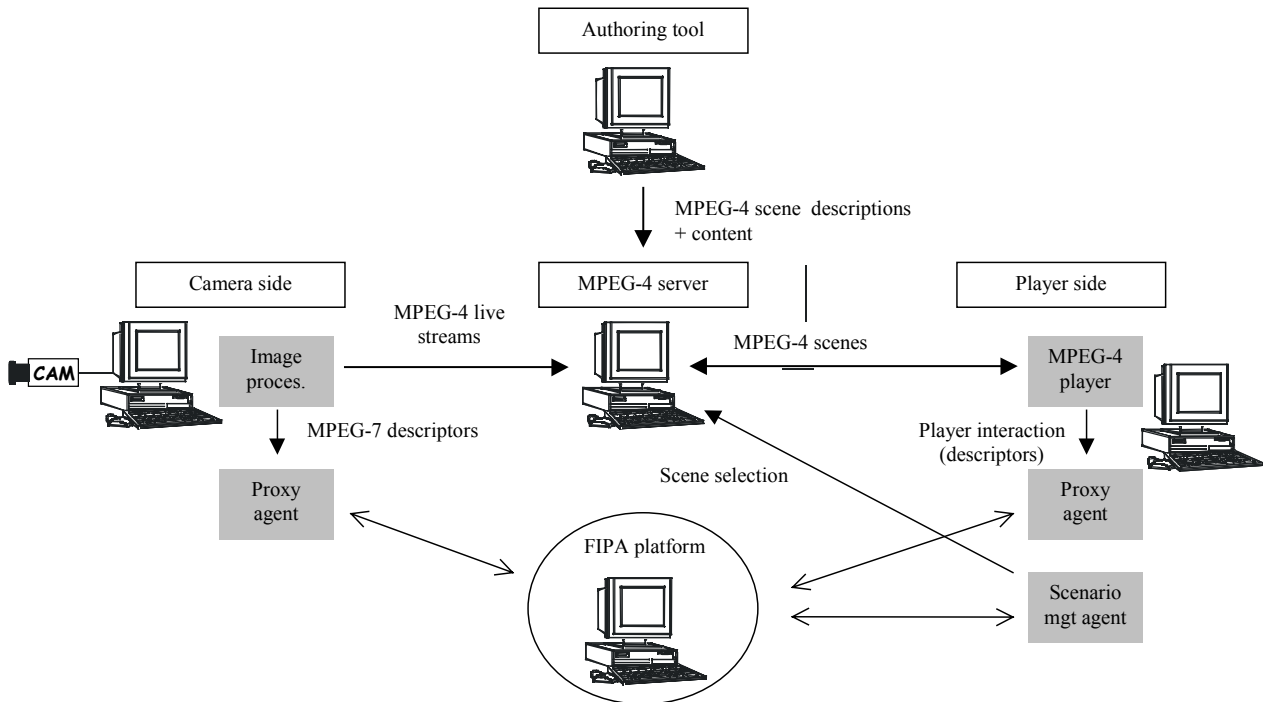


Figure 3: Implications of standards within the art.live architecture

The distributed management of the application is ensured through a real-time implementation of the *FIPA* standard that focuses on an abstract architecture providing meaningful messages exchanges between agents. The involved agents may use different messaging transports, different agent communication languages, or different content languages and are behaving in a totally autonomous way. They are therefore particularly well suited to manage local scenarios for the different screens of the system.

MPEG-7 is used to provide standardized content-based description for the various types of audiovisual information existing in the system. In particular, (segmented image appearances of) actors and interactors are described in terms of their bounding boxes (BB):

- location of the BB;
- motion of the BB;
- main color/texture of the BB;
- number of BB.

In parallel, player interaction (typically mouse clicks) are also described.

Joining the *FIPA* and *MPEG-7* standards allows agents in charge of the scenario management to manipulate well-defined triggers that are combination of standardized descriptors.

MPEG-4 is used for the coding and transmission of the various contents: live and automatically segmented camera inputs, the (inter)actors, as well as author-

prepared graphical material and scene descriptions. An MPEG-4 server allows for many screens to exploit the same content (typically the live streams) within different scenes at the same time. Synchronization is ensured through the 'commanding' agents.

6 RESULTS & CONCLUSION

A typical scene appearing on screens of the *art.live* system is shown in figure 4.



Figure 4: User is filming himself on the MR screen

As a conclusion, one may summarize the position of the project as an MPEG-4 content creator using MPEG-7 descriptors to design interactive & immersive multimedia stories that are to be experienced in an autonomous way on many different screens.