THE ART.LIVE ARCHITECTURE FOR MIXED REALITY

The ***art.live*** consortium, IST project 10942[*]
c/o Xavier Marichal
Laboratoire de Télécommunications et Télédétection
Bâtiment Stévin
2 place du Levant
B-1348 Louvain-la-Neuve
Belgium
Email: marichal@tele.ucl.ac.be

Key words: Real-time immersion, video segmentation, mixed-realities, enriched narratives

Abstract

The goal of the *art.live* project was 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: the *art.live* architecture aimed at creating interactive stories that mix graphical elements with inputs from live 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 also be offered to interact with the system.

1   Introduction

The present paper elaborates on the technical system that was designed and implemented by the *art.live* project to 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 summarizes the framework of the project. Section 3 provides the reader with a global view of the system usage, while section 4 outlines the implemented architecture. Section 5 focuses on the management of interactive scenarios.

Finally, section 6 presents some results, with an emphasis on both project trials.

2   The "Transfiction" Concept

According to Milgram and Colquhoun [1], mixed reality covers the whole continuum ranging from reality to virtuality. At the one end is the real environment, made of the real world and image capture of it. On the other end is the virtual environment, i.e. a world completely modeled in terms of shape, location, texture, motion… Mixed reality consists thus of any combination of these

two worlds. According to the relative importance of real or virtual (modeled) elements, one has to deal with augmented reality or augmented virtuality as depicted on figure 1.
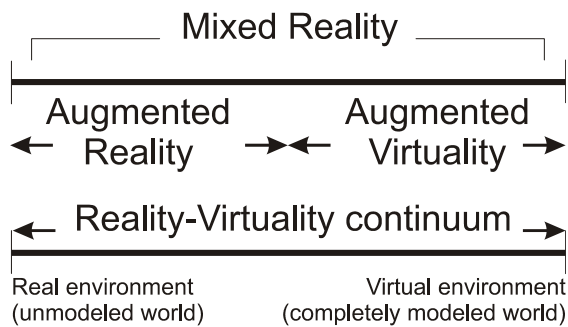

Figure 1: The mixed reality continuum.

Overall, by transfiction, we mean "**trans**portation in **fiction**al spaces" as illustrated on figure 2. Such a transportation occurs thanks to immersion and interaction.

In the framework of mixed reality, transfiction [2] is designed for mixing synthetic and natural images in real time and allows one to interact in these input/output screens. Transfiction systems are intended for intuitive interaction in a non-obtrusive manner, allowing one to develop a novel media, accumulating the representation knowledge of previous media, such as cinema, performance, theatre.
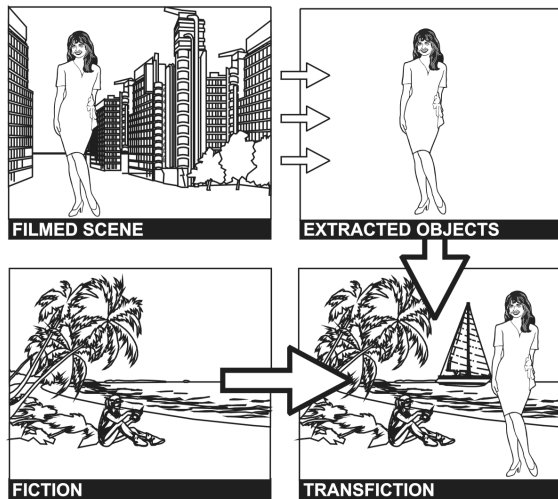

Figure 2: Transfiction concept.

The augmented (or magic) mirror effect, which creates the illusion of **immersion**, is provided thanks to image segmentation followed by re-composition. According to the context (indoor/outdoor, still/moving objects) different techniques can be used to allow the system to separate the user(s) from its (their) environment. The user layer(s) can then be composed with other

layers in order to produce the final rendering. These layers can be stills (JPG, PNG…), animations (SVG), movie excerpts (MPEG)… External effects (e.g. lights) commanded by the system can also reinforce the sense of immersion.

The **interaction** aspect is provided thanks to basic body/gesture analysis that allows users to use their hands and head to 'touch' or 'grasp' objects and participate through their body attitudes in complete interactive scenarios: "the body is the joystick".

3   System Overview

Considering a human-centric approach, the various users depicted on figure 3 are involved within the designed architecture.
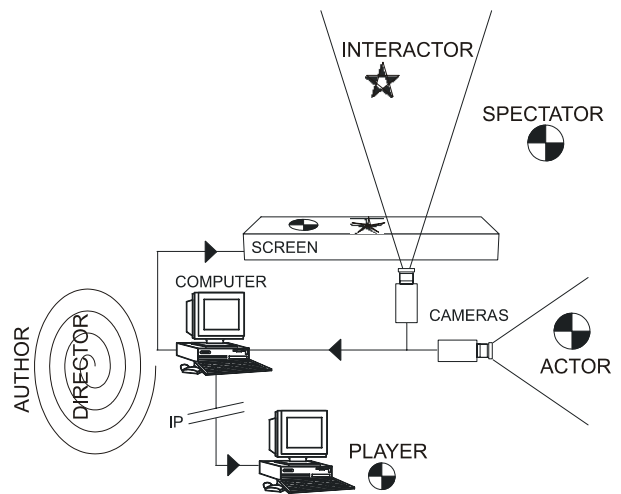

Figure 3: Repartition of users in the system

They are here 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);

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 made of the previously mentioned graphical elements.

In order to be able to offer a rich environment to authors, the developed system 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;

- a tentative authoring tool to pilot the system.

4 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 consisted therefore in combining them to create meaningful mixed-reality scenes. The main assumption was 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 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): a robust image segmentation algorithm based on change detection has been implemented to extract objects of interest. Filters are applied to remove noise while smoothing segmentation borders. However, such a fast detection is not able to tackle shadows that remain around the objects: Invariant techniques [3] are used to remove them. Finally, due to potential illumination changes in the environment, the reference image is regularly updated [4].

- Low-Level Description: basic position information of real objects along with bounding box and contour is computed. Such basic descriptors are then coupled with more elaborated techniques that mostly rely on skin color detection and tracking of detected patches in order to provide the system with the hands and head positions for the different users [5]. 3D description of objects can be added to obtain more information about the object. For instance, a simple calibrated camera can provide metric information like the position and the height of objects on a ground plane [6]. These metric measures are also interesting for the author of the scenario who can set thresholds in metric values, being thus more independent on the camera pose.

- Interpretation and Triggering of events according to scenario state and Scenario Management accordingly (cf. section 5).

- Scene Composition according to the scenario status & Rendering (visualization with possible interaction).

These tasks are complemented by some "horizontal" ones in order to obtain a fully integrated system:

- Information Management for the Storage & Retrieval of all elements (objects, scenarios...);

- Compatibility with MPEG-4 for transmission and Coding of natural and synthetic images;

- MPEG-7 descriptors and combination of these for triggering (cf. section 5);

- (Secure) Communications and Coordination within a multi-agents platform that is FIPA-compliant;

- Camera Integration (MMX optimization, or DSP-based cameras) for the sake of the real-time constraints.

5   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 trigged. 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 4 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).
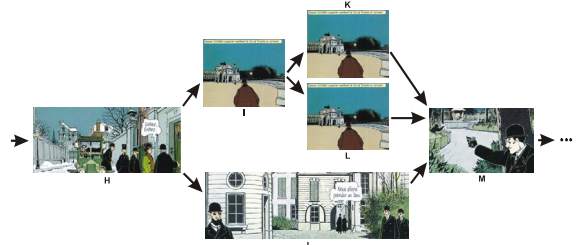

Figure 4: Typical scenario graph (images are © Casterman-Tardi).

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.

6   Results

The implementation of the presented architecture performs real-time on standard PCs. The achieved

framerate varies from 15 to 25 fps according to the complexity of the scene to compose and the precise CPU performance. Latency is inferior to 1/25 s, which suits usability constraints.

Figure 5 introduces a typical usage of the transfiction where an interactor is in front of a large screen: he sees himself immersed into a graphical environment that reacts to his own gestures and attitudes.

Figure 5: typical Transfiction situation (background image is © Casterman-Schuiten).

In order to investigate the various possibilities of the developed system, the *art.live* projects conduced two trials in public places in order to offer end-users a real test and criticism of the system.

6.1 Trial 1

art.live trial 1 has been held at the heart of an exhibition about gardens in comics ("Les jardins de la BD") organized by the "Mairie de Paris" within the Bercy Park. During the three weeks of duration, about a thousand people did try the system that aimed at investigating the interaction modalities between *actors* and *players*. Actors were moving and behaving in front of three cameras (two in the park, one in Louvain-la-Neuve, Belgium) while up to three simultaneous players were able to interact thanks to the dedicated computers.

The so-called "Adèle@Bercy" scenario was proposing visitors to help comic strip hero Adèle, from Jacques Tardi series "Les aventures extraordinaires d'Adèle Blanc-Sec" to investigate some mysteries in Paris. Nine small investigations were proposed to them, corresponding to nine different places on a map (cf. figure 6): there were every time proposed with a succession of mixed-reality scenes into which they had to look for the 'bad' robots or find (i.e. click) real people at

specific locations. It meant that they had to send a friend in front of the ad hoc camera in order to have somebody to click on.

At that time, three regular Pentium-II PCs were used for the players (respectively 200, 200 and 266 MHz) and three higher-end ones for the cameras: a PIII-800, a PII-350 and Fastcom PowerSight platform with a Celeron 533MHz. One additional machine was used as a server for the images, scene descriptions and some FIPA-related agents. All machines were communicating through the Internet Protocol.
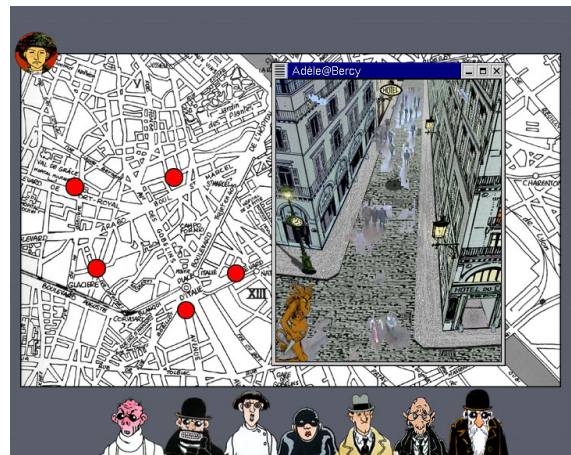

Figure 6: Interaction screen for players of trial1 (graphics are © Casterman-Tardi).

6.2 Trial 2

The second trial took place in the "Saline Royale" at Arc et Senans, France, for two weeks by the end of 2001. The installation was allowing visitors to learn about the selected theme of utopia and 300 kids from schools of the Franche-Comté region were invited to experiment and criticize the system.
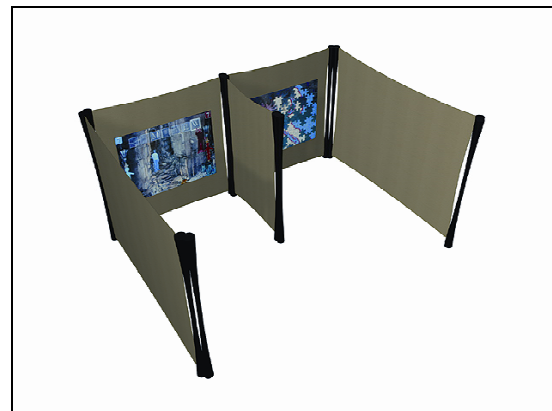

Figure 7: Trial2 settings.

Thanks a double magic mirror (camera+screen), two interactors were simultaneously involved and invited to investigate the barriers of collaboration but also competition (and *coopetition*, an activity at the edge of cooperation and competition, or looked through a temporal axis, when cooperation is first needed to find the competition terrain where the opponents can then fight for their due).

Additionally, descriptors of the real objects were used to animate graphics or avatars. For instance, as illustrated on figure 8, the global position of the avatar on the screen was the reflection of the real user's position, while attitudes will be basically mimicked. Users do not see themselves on the screen anymore but control the moves of fictional characters: the moves and gestures become the identity of the users [7].

Figure 8: One interactor 'chasing' another one, appearing as a butterfly (background image is © Casterman-Place).

7   Conclusion

Through public demonstrations and trials, the project aimed 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 and requires the definition of ad hoc narrative patterns.

One of the major conclusions of the project is the strong need for inputs from artists, authors, creators… for mixed-reality applications to emerge. Then technological tools can become mature enough for creating new experiences for the masses.

References

[1] A. Nandi and X. Marichal, "Transfiction", Virtual Reality International Conference 2000, Laval, May 2000, pp. 76-88.

[2] P. Milgram and H. Colquhoun, "A Taxonomy of Real and Virtual World Display Integration", in *Mixed Reality, Merging Real and Virtual Worlds*, Y. Ohta and H. Tamura editors, Ohmsha-Springer, 1999.

[3] A. Cavallaro, E. Salvador and T. Ebrahimi, "Knowledge-based Analysis for Shadow Recognition", submitted to IEEE Transactions on Image Processing.

[4] X. Marichal, "On-line Web Application using Image Segmentation", Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Berlin, May 1999, pp. 141-144.

[5] V. Girondel, L. Bonnaud and A. Caplier, "Hands detection and tracking for interactive multimedia applications", International Conference on Multimedia and Expo (ICME), 2002.

[6] O. Faugeras, "Three-Dimensional Computer Vision: A Geometric Viewpoint", MIT Press, Cambridge, Massachusetts, 1993.

[7] A. Nandi and X. Marichal, "Senses of Spaces through Transfiction", to appear in "Entertainment Computing: Technologies and Applications" (Proceedings of the International Workshop on Entertainment Computing, Makuhari, Japan).