Immersive Communication

Damien Douxchamps, David Ergo, Benoît Macq, Xavier Marichal,
Alok Nandi, Toshiyuki Umeda, Xavier Wielemans
alterface *
c/o Laboratoire de Télécommunications et Télédétection
Université catholique de Louvain
2 place du Levant
B-1348 Louvain-la-Neuve - Belgium
Contact e-mail: marichal@tele.ucl.ac.be

ABSTRACT

An apparatus for communication/entertainment mixing synthetic and natural images in real-time is designed and allows the "user" to be captured through vision-based sensors, like (web) cameras. The composed visual scenes are to be experienced in physical spaces and/or to be viewed through web browsers. The word "transfiction" has been coined to this interactive narrative system where users can interact with narrative machines (devices with computing power and containing databases of meaningful information).

1 Introduction

Contrary to many approaches to virtuality or mixed reality, the designed system does not need any dedicated hardware, nor for computation nor for tracking of real objects/persons. It runs on standard Pentium PCs and cameras are the only used sensors. This vision-based interface approach allows complete freedom to the user, not anymore tied to hardware devices such as helmets and gloves. Various research projects have already adopted such a user-centric approach towards mixed reality. It ranges from the only animation/command of purely virtual worlds, as in the KidsRoom [1], to more mixed worlds where users see a virtually reproduced part of themselves as in N.I.C.E. [2], and goes to the inclusion of the user image within the virtual space in order to fully exploit the potential of mixed reality. In ALIVE [3], "Artificial Life Interactive Video Environment", wireless full-body interaction between a human participant and a rich graphical world inhabited by autonomous agents is used.

The present system of "transfiction" [4] aims at extracting users out of their context when they enter the space of some camera. The image is analyzed, the visual representation of people is automatically extracted and then integrated within a pre-existing story in order to construct the mixed-reality scene, as depicted in figure 1. The users' attitudes and behaviors then influence the narrative, with the explicit intent of making the immersion a rich experience for all users.

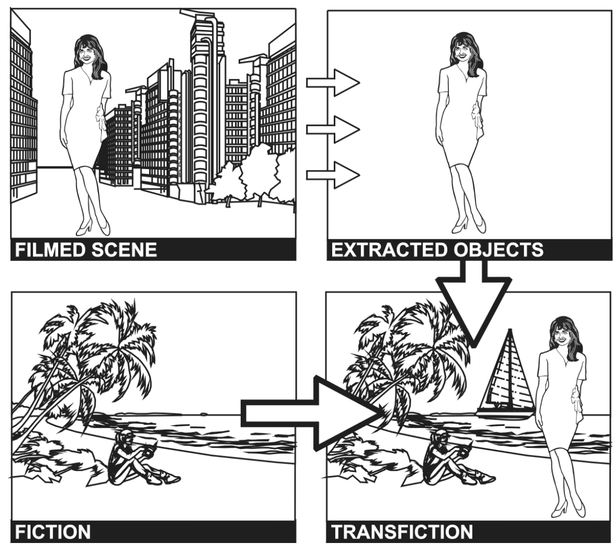

Figure 1: Underlying concept of transfiction

Since it makes use of both real and graphical images, transfiction needs to be positioned in the real-virtual continuum and in the context of the "mixed reality". According to Milgram and Colquhoun [5], 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 2.

The reminder of the present paper is organized as follows. Sections 2 and 3 summarizes some key features of the transfiction [4] system: section 2 elaborates on some of its key concepts while section 3 presents the main as-

* Part of the research presented in the present paper is achieved in the framework of the *art.live* project (http://www.tele.ucl.ac.be/projects/art.live). This project is part of the fifth framework programme of the European Commission: Information Society Technologies (IST, project 10942).

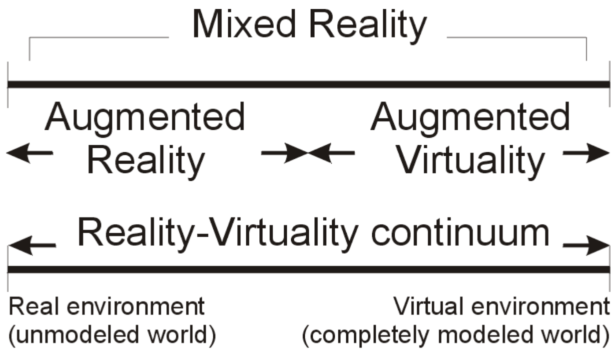

Figure 2: Mixed-Reality framework

pects of the underlying architecture. Finally, section 4 presents some results that are used to explain and illustrate the way interaction makes the scenario progress.

2 Concepts

Considering a human-centric approach, the various "users" depicted on figure 3 are involved within the "transfiction" 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);

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.

It is important to stress that the difference between an interactor and an actor only resides in the degree of awareness of the user itself. Basically, both these users are in front of some cameras and have some influence on the system because of their attitude. Since the interactor is really made aware of the impact of his own behavior thanks to the big screen, it is expected that he will not remain a passive actor anymore but will adopt specific gesture and attitude in order to interact with the system.

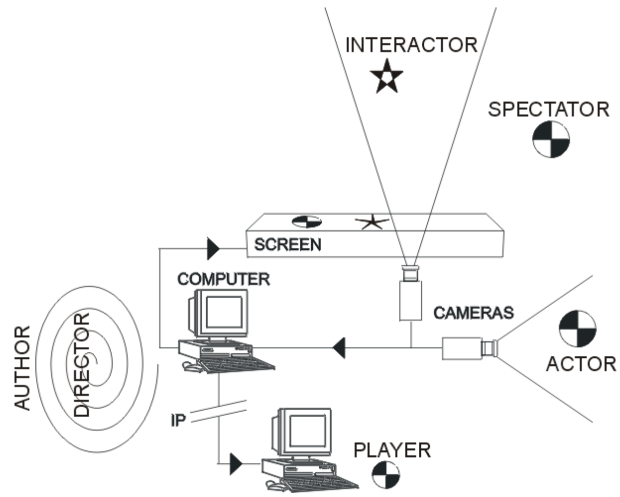

Figure 3: Repartition of users in the system

In figure 3, a systemic approach has been adopted: users are bound to a narrative apparatus which consists of cameras for image capture, computers for image composition, signal processing and scene composition, and screens for image rendering. For the sake of clarity, it is important to provide a praxis-based typology of the different spaces in which the bodies, objects and events are taking place. With this respect, it is interesting to quote Deleuze [6] who already elaborated on such considerations about different types of spaces: "We opposed the virtual and the real: although it could not have been more precise before now, this terminology must be corrected. The virtual is opposed not to the real but to the actual. The virtual is fully real in so far as it is virtual. Exactly what Proust said of states of resonance must be said of the virtual: 'Real without being actual; ideal without being abstract'; and symbolic without being fictional." Relying on such a point of view, the following definitions are used:

- The *Actual Space* is the space in front of the camera. It is the space in which any person becomes an interactor.

- The *Real Space* is the space in which the user is living, be it consumer, spectator, author....

- The *Virtual Space* is the space which is rendered on the screens. It is composed of real-time images of the interactors or other real elements as well as bodies and objects generated from a computer database.

- The *Diegetic Space* (which is more specific for narrative films) refers to the world of a film story. The diegesis includes events that are presumed to have occurred as well as actions and spaces not shown onscreen. The concept of diegesis will take its plain dimension once we will be able to offer to the audience an extended narrative experience similar to viewing a film.

3 Technical Architecture

In order to provide users with such an experience, the technological challenge is to gather all needed subcomponents and issue a real-time implementation of the system. To compose all the visual objects (the "real" and computer-generated ones) within the final mixed-reality scene and to allow for interactivity, the MPEG-4 standard [7] can be used as the transmission layer.

In addition to the composition and transmission facilities, the following issues are addressed:

- Achieving a real time segmentation of moving objects captured by a web camera. As in [8], change detection is combined with automatic background adaptation in order to provide fast but robust object extraction.

- Associating these objects extracted from the real world with predefined synthetic objects in order to compose a coherent mixed-reality scene. Images, with a transparency layer, animations and short movies can be used. However, the *author* must pay extreme attention to the overall coherency of the combination of the different layers.

- Performing the real-time encoding of the arbitrarily shaped objects with the help of various coding techniques that appear within the MPEG-4 framework.

- Establishing a client-server architecture based on the Internet Protocol that allows for ubiquity (simultaneous re-use of the same camera images within various mixed-reality scenes) and composition of scenes from various local or remote sources.

- Automatically extracting (MPEG-7 [9] like) descriptors that are used to describe the behavior of visual objects. Such descriptors are presented with some more details in section 4 and are used to pilot the interactive scenario.

Solutions to these issues have been combined in order to implement the architecture depicted on figure 4.

Thanks to the Internet Protocol, the system is very flexible and allows any screen to access any resource it needs. A phenomenon of ubiquity is therefore provided

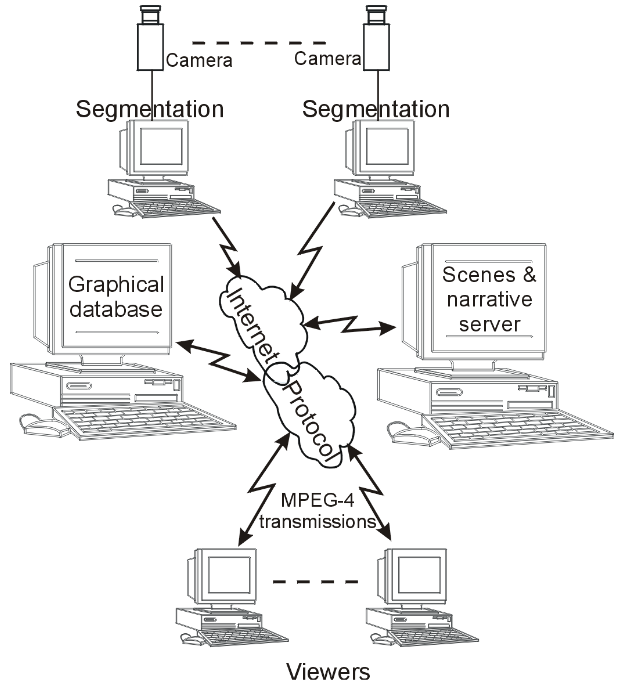

Figure 4: Technical architecture

since two or more screens may simultaneously access the same camera stream. Therefore, the system is very open and any device can be as much reproduced as needed.

4 Interactive Scenario

As already mentioned, descriptors are used for managing the application and offering users the possibility to interact with the scenario. Therefore, descriptors are attached to some of the graphical elements. For instance, on figure 5, one can see that three 'touchzones' are defined at three particular locations of the graphical elements.

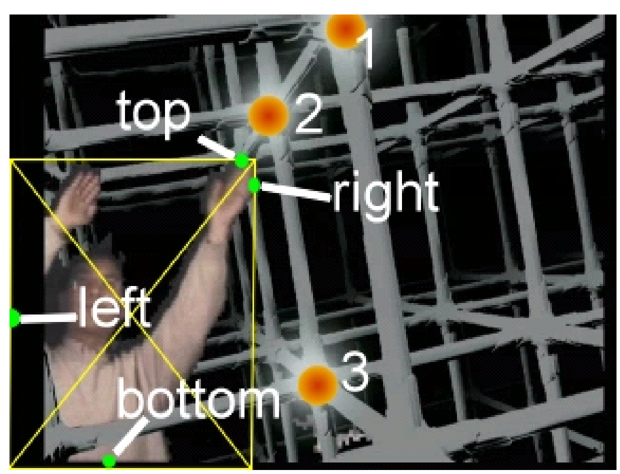

Figure 5: Presence of different descriptors in the current scene

These descriptors are compared in real-time with the ones generated by the interactor: in the present case, his/her image is described in terms of the bounding box (cf. figure 5) and contact points on the side of the box.

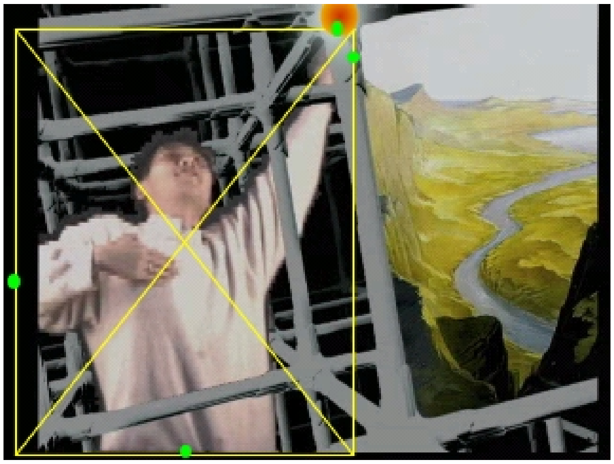

Figure 6: Activation of a scenario event thanks to descriptors

The location of one of these contact points (top - bottom - left - right) with some 'touchzones' causes an event in the scenario. On figure 6, the first 'touchzone' of figure 5 has been activated by the top contact point, provoking the appearance of another layer of graphics. 'Touchzone' number 2 was also to be activated by the top contact point while zone number 3 was targeted for the right contact point.

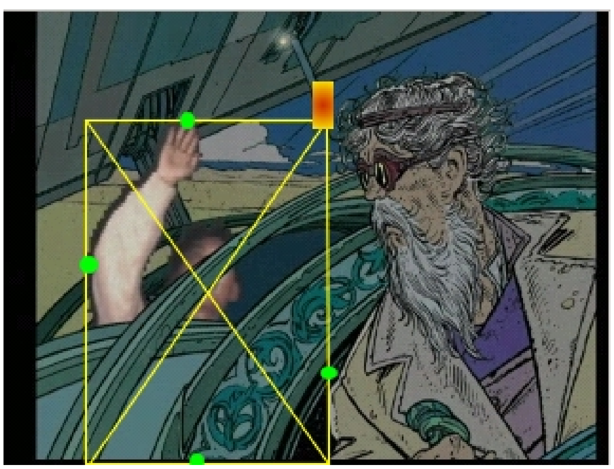

Figure 7: Although the bounding box is in contact with the 'touchzone', no event is activated

As shown on figures 7 and 8 the use of contact points instead of the bounding box itself allows increasing the robustness of the system with respect to noise and segmentation defaults while ensuring that interaction is really driven by the user. Indeed, it is very likely that his/her head and hands (interactive elements *per excellence*)are the elements in contact with the bounding box, therefore driving the application.

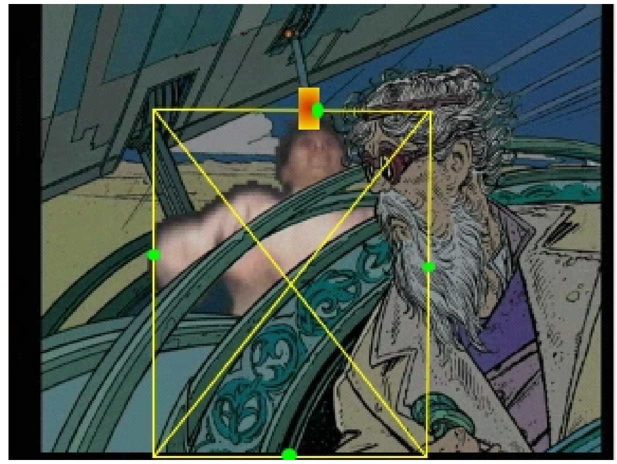

Figure 8: Activated event because of the presence of the top contact point in the 'touchzone'

Of course, intrinsic parameters of the bounding box, as well as other descriptors (like position, texture, motion) can be used to generate more events. For instance, on figure 9, the user (appearing behind the semi-transparent *drosera*, is offered to open or close the *drosera* itself according to the width of the bounding box. If he/she manages to open it completely, another event will occur, i.e. a change of scene and the pursuing of the interactive story.

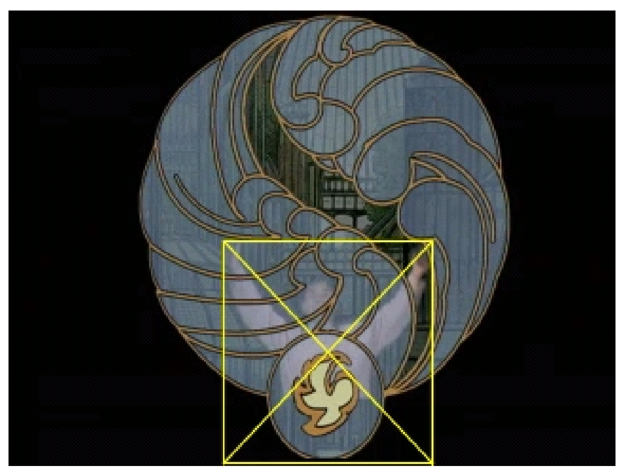

Figure 9: Interactive piloting of some animation (opening of the *drosera*)

References

[1] A. Bobick, S. Intille, J. Davis, F. Baird, C. Pinhanez, L. Campbell, Y. Ivanov, A. Schtte and A. Wilson, "The KidsRoom: A Perceptually-Based

Interactive and Immersive Story Environment", PRESENCE: Teleoperators and Virtual Environments, 8(4), August 1999, pp. 367-391.

[2] M. Roussos, A. Johnson, J. Leigh, C. Vasilakis, C. Barnes and T. Moher, "NICE: Combining Constructionism, Narrative, and Collaboration in a Virtual Learning Environment", Computer Graphics, 31(3), August 1997, pp. 62-63.

[3] P. Maes, T. Darrell, B. Blumberg and A. Pentland, "The ALIVE system: Wireless, full-body interaction with autonomous agents", Multimedia Systems, 5(2), March 1997, pp. 105-112.

[4] Alok Nandi, Xavier Marichal, "Transfiction", Virtual Reality International Conferences 2000, Laval, 18-21 May 2000.

[5] 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.

[6] G. Deleuze, "Difference and Repetition", London, Athlone, 1994.

[7] MPEG Video group, "MPEG-4 Video Verification Model, version 11.0", ISO/IEC JTC1/SC29/WG11 (N2178), Tokyo, March 1998.

[8] X. Marichal, "On-line Web Application using Image Segmentation", WIAMIS'99, Berlin, May 1999, pp. 141-144.

[9] MPEG group, "MPEG-7 description definition language document", version 2.0, ISO/IEC JTC1/SC29/WG11 (N2997), Melbourne, October 1999.