



DELIVERABLE D7.1

Scene Modelling and 3D Presentation of Learned Activities - Design Issues

November 15, 2002

Authors: Vladimír Štěpán, Václav Hlaváč

Czech Technical University, Faculty of Electrical Engineering

Department of Cybernetics, Center for Machine Perception

121 35 Prague 2, Karlovo náměstí 13, Czech Republic

{stepanv,hlavac}@cmp.felk.cvut.cz, <http://cmp.felk.cvut.cz>

Project acronym: **ACTIPRET**

Project full title: **Interpreting and Understanding Activities of
Expert Operators for Teaching and Education**

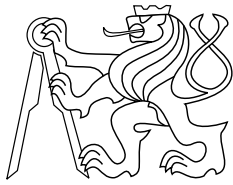
Action Line IV.2.1: **Real Time Distributed Systems (Cognitive Vision)**

Contract Number: **IST-2001-32184**





CENTER FOR
MACHINE PERCEPTION



CZECH TECHNICAL
UNIVERSITY

RESEARCH REPORT

ISSN 1213-2365

Scene Modelling and 3D Presentation of Learned Activities – Design Issues

Deliverable 7.1
of the project IST-2001-32184 ActIPret

Vladimír Štěpán, Václav Hlaváč

{stepanv,hlavac}@cmp.felk.cvut.cz

CTU–CMP–2002–17

November 15, 2002

Available at
<ftp://cmp.felk.cvut.cz/pub/cmp/articles/hlavac/Stepan-TR-2002-17.pdf>

Authors were supported by the project IST-2001-32184 ActIPret,
by the project GAČR 102/00/1679, and by the Czech Ministry of
Education under the project MSM 212300013.

Research Reports of CMP, Czech Technical University in Prague, No. 17, 2002

Published by

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: <http://cmp.felk.cvut.cz>

Abstract

This Deliverable 7.1 document describes the ActIPret Work Package 7 (WP7), its relationship to the framework of the ActIPret project and the functionality this Work Package provides. WP7 tasks are divided in two parts, the part of scene modeling and reconstruction and the part of presenting the activity by animating virtual human model. In corresponding two sections of the document the problems and challenges in these two areas are discussed and the possible solutions are suggested.

Contents

1	Introduction	2
2	Interfaces to the ActIPret Framework	3
3	Building the VR model of the ActIPret scene	6
3.1	Task specification	6
3.2	State-of-the-art in scene reconstruction for VR and available tools	6
3.2.1	VRML	7
3.2.2	Manual creation of the VR model	7
3.2.3	Computer vision-based 3D reconstruction	8
3.3	Design issues relevant to creating the VR model	9
3.4	Our initial experiments	10
3.5	Summary of experience gained	11
3.6	Outlook	13
4	Activity Presentation By Animating Virtual Humanoid Model	13
4.1	Task specification	13
4.2	State-of-the-art	14
4.2.1	Humanoid body modeling	14
4.2.2	Humanoid body animation	15
4.3	Available tools	17
4.4	Our choice	17
4.5	Initial experiments	17
4.6	Summary of experience gained	18
4.7	Outlook	18
5	Conclusions	19

1 Introduction

This document is a deliverable 7.1 related to the Work Package 7 *Scene Modelling and Presentation* (WP7) of the IST-2001-32184 project ActIPret. The WP7 constitutes a presentation module of the project. The human activity learned in other work packages will be presented using virtual reality (VR) tools in a realistic 3D environment. This includes (a) 3D reconstruction of the real scene from images; (b) actual VR presentation of the learned human activity from varying viewpoint.

This report summarizes design issues of the WP7, discusses relations to other WPs, specifies interfaces of the WP7 to other other components of the *ActIPret Demonstrator*. The specification of the VR 3D presentation module is given too. The question how to reconstruct a 3D scene of the demonstrator scenario and convert it to the form usable by the VR module is also addressed. The related state-of-the art is briefly mentioned. The report also summarizes the work which has been carried out under the WP7 so far.

The VR presentation includes realistic visualization of the learned human activity in order to easily verify activity plan, i.e., a sequence of actions. The activity can be displayed to the human operator from the varying viewpoint. This goal implies that the 3D scene has to be reconstructed and an animation of the learned activity has to be created.

The WP7 is understood as a service WP for remaining parts of the project. It works offline with respect to the ActIPret framework. Thus it can be developed independently and its performance is not time critical provided that the demonstrated activity plan can be viewed by human observer without annoying delays.

The plan is to apply standard tools for 3D scene reconstruction, both for VR modelling, and for presentation of the learned activity to the user. We do not expect to carry out much of the research in this work package. However, the interesting research pathes will be followed as they appear.

It was supposed that the WP7 will start in month 7 of the ActIPret project. However, we were able to recruit the researcher (Mr Vladimír Štěpán) with MSc level education in computer graphics and experience with VR modeling in sport scenarios. Thus the work on the WP7 started ahead of the schedule in month 3. We were able to demonstrate preliminary design of the presentation module to reviewers at the project meeting in Heraklion in June 2002. The plan for WP7 was approved by reviewers.

The aim is to use standard tools for:

- To define the geometry of the scene and functionality of the scenario,

the VRML (Virtual Reality Modelling Language, ISO/IEC 14772) language will be used.

- User interaction with the VRML scene as well as the enhanced control functions can be provided via External Authoring Interface (EAI), which is also part of ISO/IEC 14772 standard.
- The humanoid body model used for the activity presentation will correspond to the H-Anim 1.1 standard that defines the hierarchical structure of the human body model and also provides typical anthropometric data.

The research issues within WP7 might arise in the area of creating and handling the animations for an H-Anim 1.1 humanoid model. There is a connection between the computer vision and computer graphics in the scene reconstruction. The interesting question is to study how the motion of a human can be recovered from real sequences having a priori the model of a generic human.

The rest of this deliverable is structured as follows. In Section 2 the relationship to the framework and other modules is discussed. Section 3 specifies the use of standard VR tools in scene modelling and reconstruction and points out some of the problems in area of scene reconstruction. Section 4 deals with the aspects of activity presentation using the virtual humanoid model. We conclude the deliverable in Section 5.

2 Interfaces to the ActIPret Framework

The WP7 is relatively self-standing WP in the ActIPret project. The overview of the ActIPret framework is shown in Fig. 1. The position of the WP7 in ActIPret functionality shows Fig. 2 and its relationship to 3 ActIPret phases is depicted on Fig. 3.

The presentation module will work offline. The module will be used in ActIPret learning, expert and tutor phases. There are two subtasks involved. The first subtask performs reconstruction of the real scene from images or video. The aim is to represent the scene and manipulated objects in VRML. Such data can be used for the VR presentation. The subtask will cooperate with the Object Recognition module and will also use images provided by the Image Server.

The second subtask has to create the animation to present the activity in the current state of learning. Displaying (and animation) of the activity will use activity plan as input. No output to another module is expected.

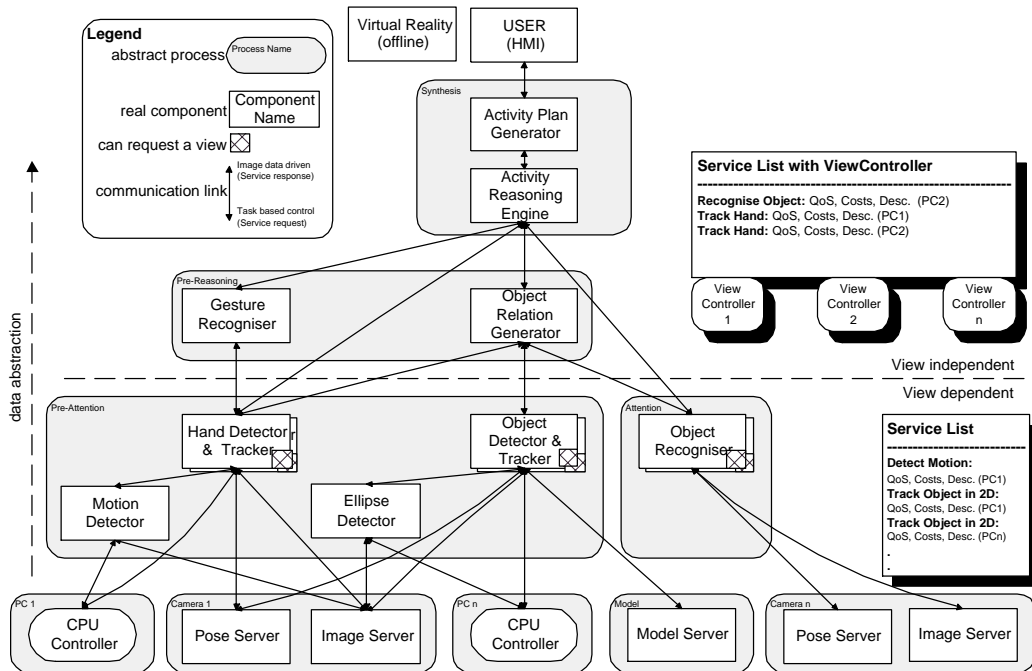


Figure 1: Diagram of the ActIPret framework.

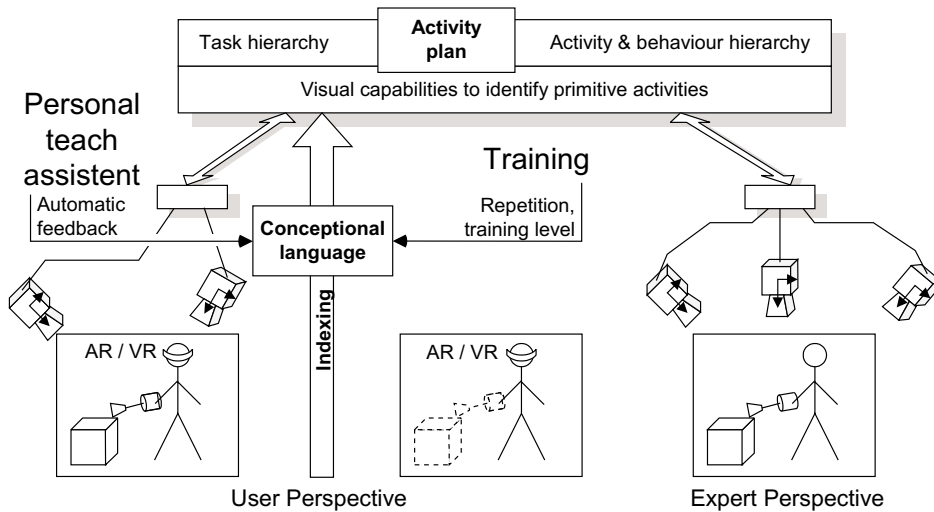


Figure 2: Using the ActIPret system to record and retrieve activities.

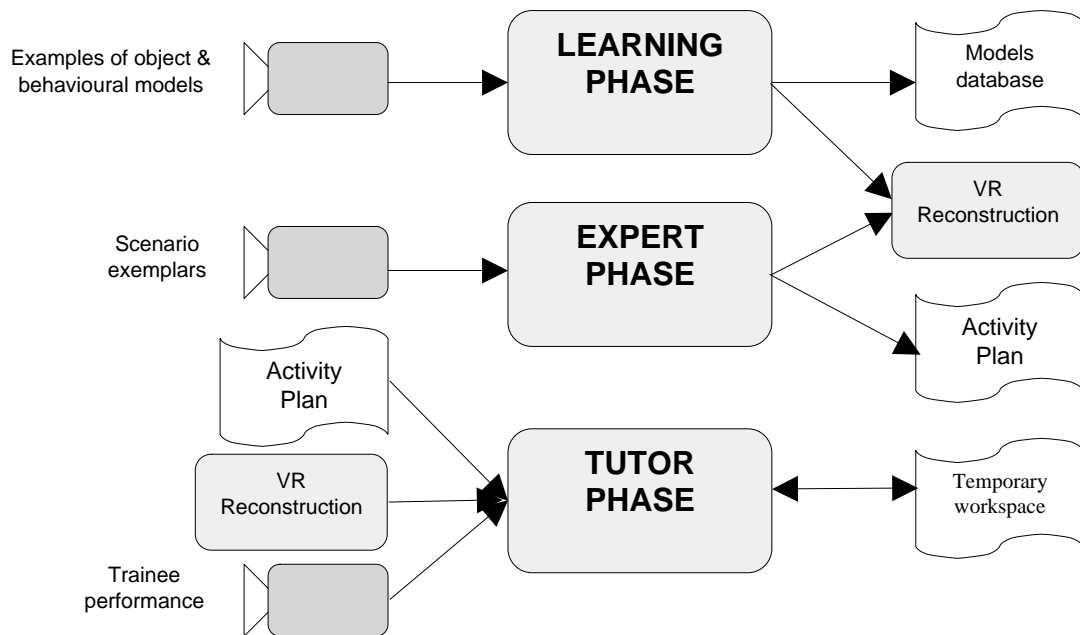


Figure 3: The relationship of the VR module to the 3 ActIPret phases.

Relationship To Other Work Packages

The WP4 and WP5 are related to the presentation module WP7 and provide needed functionality.

WP4 Recognition of objects is related to our WP7 because it provides information about the objects in the scene which are relevant to the activity plan and that are to be reconstructed and presented using VR techniques.

We are likely to use the same images for the 3D reconstruction of the scene as the WP4 uses for recognition purposes.

WP5 Synthesis of task and behavior representation provides the activity plan represented in the conceptual language which constitute input data of WP7. The aim is to create a VR animation of the humanoid body that will demonstrate the learned activity.

3 Building the VR model of the ActIPret scene

3.1 Task specification

One of the tasks entrusted to WP7 is to *build the VR model of the ActIPret scene* in which the test scenario is carried on. A certain degree of realism is expected.

If the learned activity has to be presented in VR environment similar to that one in which the training sequences were captured then the 3D model of scene has to be created. This implies that 3D geometry and appearance of the scene has to be reconstructed. The VR model of the scene can be created either by hand, i.e., by building the scene manually from simpler geometric building blocks, or the model building can be aided from 2D images or image sequence with the help of computer vision techniques. The latter technique has twofold advantages: (a) The created VR model resembles the actual scene; (b) The process is faster. It is foreseen that both approaches – manual and aided by real scene images – will be pragmatically combined in the ActIPret project.

The WP7 virtual reality will be *based on the VRML standard*, which is suitable for creating the geometry and functionality of the models. The VRML standard also defines an interface to Java applet or an application that allows us to control and influence the VRML scene from outside.

3.2 State-of-the-art in scene reconstruction for VR and available tools

The VR model building and its reconstruction from images of a real scene is a quite broad area, so this part of the deliverable will be a little sketchy, dealing with just the most relevant aspects. However, we can build on the previous experience of the Czech Technical University team in ActIPret project in 3D geometry reconstruction from multiple images [7, 8] and in VR modeling.

The Computer Graphics Group of the Czech Technical University's Department of Computers works on a Virtual Old Prague project [11, 12, 13] that is concerned with creating virtual cities. The reconstruction part is not the real focus of this project though and is done manually.

The synthesis of 3D images of the scene is usually based on a geometric model, material properties and lighting of the scene. These parameters have to be acquired from the real scene.

The relevant state-of-the-art is reviewed in several areas:

- VRML standard and relevant tools.
- Tools helping manual creation of the scene.
- Computer vision-based methods which help to capture and reconstruct 3D geometry and appearance of the 3D scene.
- Overview of the approaches of similar projects which attempt to represent and present a human activity to a trainee.

3.2.1 VRML

VRML is a standard widely used for 3D scenes on the www. The VRML standard (ISO/IEC 14772) [27] defines the textual description of virtual world. This description includes static and dynamic elements such as the geometry, appearance (material or texture) of the static object, illumination, sounds in the scene, and the animations provided by linear interpolators of attributes (position, rotation, coordinates, etc.) related to the objects. The standard also defines an interface (External Authoring Interface - EAI) that can be used to access the VRML world and its attributes from an external program. Since the VRML is www oriented, the EAI interface has been implemented in the Java language. Therefore it is possible to access the VRML scene from a Java applet or a stand-alone application.

The common VRML browsers are the plug-ins for the www browsers (mainly Microsoft Internet Explorer or the Netscape Navigator). Several companies [18, 19, 15] have created their own VRML browser. The Cosmo Player by Cosmo Software [19] used to be popular and widespread, but lately Parallel Graphics [18] Cortona VRML Client has replaced it. Also the Java implementation of the EAI interface is distributed with this browser.

Since the VRML is a textual format it is easy to edit in any text editor. A specialized VRML editor has been created by Parallel Graphics (VrmlPad 2.0 - free trial version) [22]. Many 3D scene editors (e.g., 3DStudio Max) can export to the VRML format.

3.2.2 Manual creation of the VR model

In the process of building a virtual model, we need the information about objects in the scene, their size, mutual position, etc. The manual approach to VR model creation implies that this information is provided by the user. Most of the systems used for 3D modeling belong to this category. The reconstruction software based on computer vision and image processing that will be described later can be considered a tool for the manual creation too,

because it needs a lot of human work to produce the model. Such programs usually require the user to define the correspondences.

Let's now discuss the software tools and approaches that serve to create the VR model of the scene fully manually. Typically the 3D editors, such as 3D Studio Max, are used for this task. The professional 3D modeling and animation tools can usually export the modeled scene to the VRML format although it means simplification of results these tools are capable of. Also the simpler editors specialized at visual editing of VRML scene (Spazz3D - free trial [24], ParallelGraphics Internet Space Builder and Internet Scene Assembler - free trial [22], etc.). These editors often utilize the possibilities of VRML quite well.

Another approach to the VRML scene editing is textual editing of the VRML code. This is a true manual approach, but still it is to be taken into account as a very useful method thanks to the VRML code editor mentioned above.

3.2.3 Computer vision-based 3D reconstruction

Real 3D objects/scenes acquisition methods can be divided into two broad categories: active and passive ones. The active methods cast some known energy onto the scene, examples being the Lidar or the structured light illumination. The approach relies on a special and typically expensive hardware. In CMP, we have experience, the hardware and tools for capturing 3D models using the laser structured light [3]. The acquisition procedure is designed for small objects which can be placed on the turn table. The method produces high precision results. The acquisition process is time consuming and might require modifying the scene (painting objects opaque white).

Passive methods just observe the scene by a camera and derive a 3D scene description from images or video sequences. The key issue of these methods is the solution to the correspondence problem which is ill-posed in general and is often computationally expensive.

The simplest methods use techniques known in photogrammetry for more than hundred years. The 3D geometry is reconstructed from still images captured by the fully calibrated camera. The uncalibrated cameras can be used too [6, 7, 8].

The most pragmatic approaches let the human operator to solve the correspondence problem. Currently, the user clicks the correspondences on the screen. Examples of this approach are: (a) our own RECx (research toolbox in Matlab) [9], (b) PhotoModeler program [23] which is available as a free-ware, and (c) Canoma by Meta Creations [17]. Canoma can reconstruct a scene from a single image using additional constraints such as the predefined

primitive shapes and topological relationships. Neither of these tools works automatically. Considerable amount of user interaction is necessary.

We have bought Canoma and we like to use it in the ActIPret project for reconstruction.

The richer source of information, a video sequence, can be used for scene reconstruction too. It can provide views captured from several viewpoints. The spatial resolution of video sequences is typically lower compared to still images captured by a still digital camera.

We have tentatively tested the Boujou software [16] manufactured by the 2d3 company from the U.K. The program is designed to find a trajectory of the camera by analyzing a video sequence. The prime customer is the movie industry. Knowing the position and orientation of the camera, the VR objects can be augmented into a real movie. Boujou provides also coordinates of corresponding points detected automatically throughout the sequence. Also it is possible to edit manually selected points and, e.g., assign them to individual objects in the 3D scene. Thus we can get coordinates or trajectories of coordinates belonging to a certain object. If the camera taking the sequence was static, the primary output (i.e., camera position and orientation) loses its sense but the secondary output of the selected points coordinates can be used.

3.3 Design issues relevant to creating the VR model

We have selected the Parallel Graphics Cortona VRML browser [18] for virtual reality presentation in the ActIPret project. The reason is that this tool seems to be under the most rapid development. We also use the EAI interface Java classes which are distributed with Cortona.

Using EAI interface we are able to control the scene from outside, from an applet or an application. We have implemented both platforms, an applet and a stand-alone application for communication with the VRML scene. Controlling the VRML scene via EAI interface is a technique that has been handled well. Objects can be added into the scene, any arbitrary attribute can be changed, any dynamic event can be caught or simulated by the external program.

For the object with simple geometry (table, CD player, CDs) the manual reconstruction is easy and usable. This means to measure an object with a ruler to obtain the data for defining the geometry of the model. Besides the geometry, also the surface characteristics of the model are to be defined. This is usually done by applying a texture. The texture is obtained by editing an image of an object in a bitmap editor.

More automated ways of reconstruction are suitable and will be further investigated. The use of some of the software for reconstruction from images mentioned in this document (e.g., Canoma) can bring at least the advantage of automatic texture application. The geometry definition still needs a significant amount of manual work.

3.4 Our initial experiments

The aim of initial experiments was to evaluate selected approaches and relevant tools on the simplest ActIPret scenario (placing a CD into a CD player). We like to prepare tools for routine 3D scene reconstructions, converting it into the VRML format, and preparing for the 3D presentation together with the animated humanoid performing the ActIPret activity plan.

We have tested the suitability of ISO/IEC 14772 standard (VRML and EAI interface) for our purpose.

The testing scene for the CD scenario was set up as a part of our initial experiments and reconstructed. A video of a person carrying out the CD scenario was taken. Several still images were captured too for the purposes of the 3D reconstruction by Canoma software. The scene was then reconstructed in VR. The set of models of the objects from the scene (CD player, CD, table?) was created. The reconstruction was done completely manually ? objects were measured and photographed to get their dimensions and surface textures. Functionality of some objects was implemented in the VRML code (CD player can accept an event that opens/closes the CD rack).

Then the program in a form of the Java applet was created to test and demonstrate the possibilities of the EAI interface (Fig. 7). Finally this program was implemented in a form of the stand-alone Java application, which is generally much less used and documented part of EAI capabilities.

Several tests of 3D reconstruction SW were also performed in order to find a way to make the reconstruction process more automated. We tested Canoma (Fig. 5) and Photo Modeler (Fig. 6) for creating the models of objects and 2d3 Boujou (Fig. 4) as an image sequence processing tool with various outputs.

We have used Canoma and Photo Modeler to create model of the scene from images taken from the video sequence of the CD scenario. Using Canoma we were able to build fairly good model just from one image, because the user can define additional constraints concerning the shapes of the objects and their topological relationships. On the other hand the Photo Modeler works solely with points and correspondences defined in several images. In this case, when we worked with the images taken from not very different

viewpoint, the resulting model was not usable.

Basically we can say that we used Canoma and Photo Modeler for the creation of static model. On the contrary, with Boujou we tried to use some of its functions to obtain the data for the dynamics of the scene. Boujou is a tool that can retrieve the trajectory of the camera in the scene out of the video sequence. The basic idea of the experiments was: If the camera is static, we might be able to retrieve the trajectories of moving objects. The results proved that process of making use of Boujou would be more time consuming than expected, it was not completely forsaken though.

At this point our acknowledgement belongs to Tomáš Pajdla, Jiří Žára and the student Michal Jůza, for consult or direct help in performing these experiments.

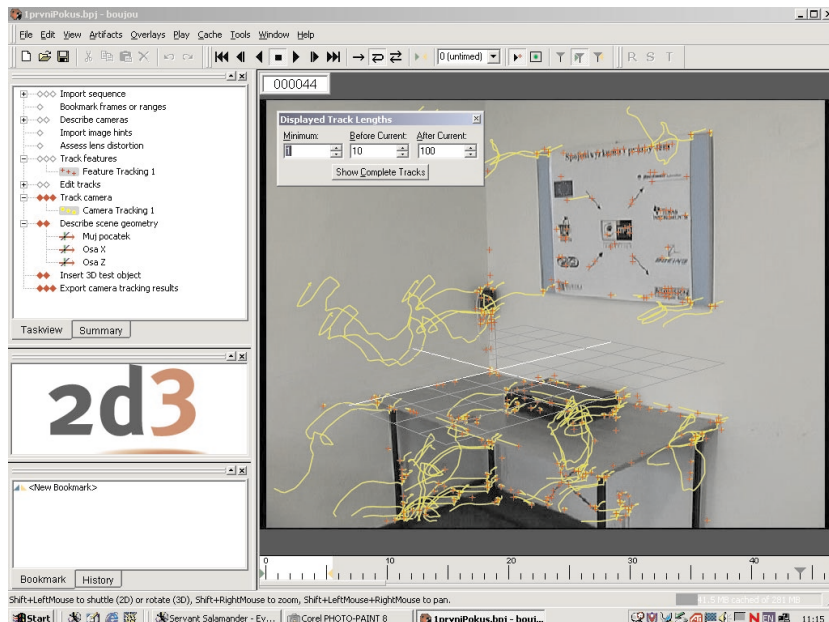


Figure 4: Example of the tested SW: Boujou.

3.5 Summary of experience gained

We found the ISO/IEC 14772 standard well suitable for the purposes of WP7. We have learned to use the EAI interface to communicate with the VRML scene and we obtained the know-how of creating an applet or a stand-alone application tool for this task.

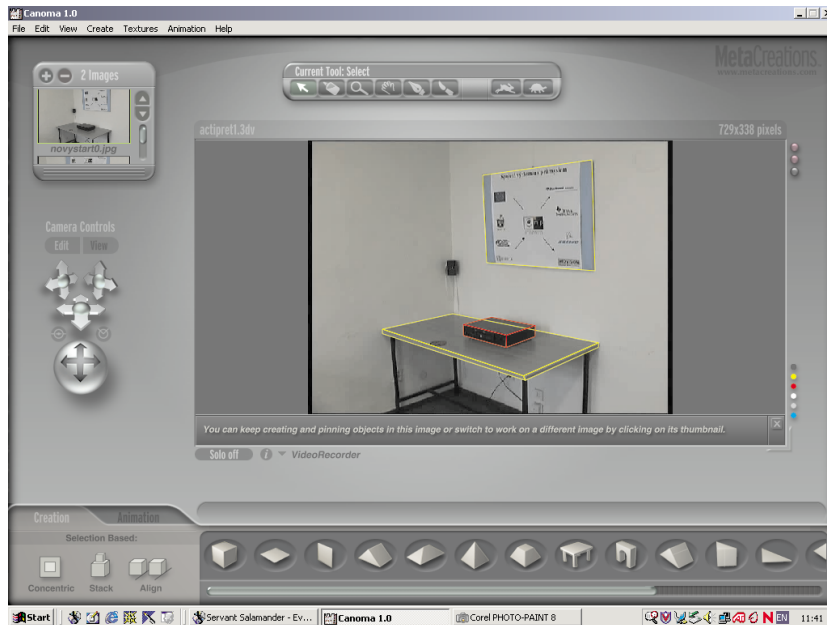


Figure 5: Example of the tested SW: Canoma.

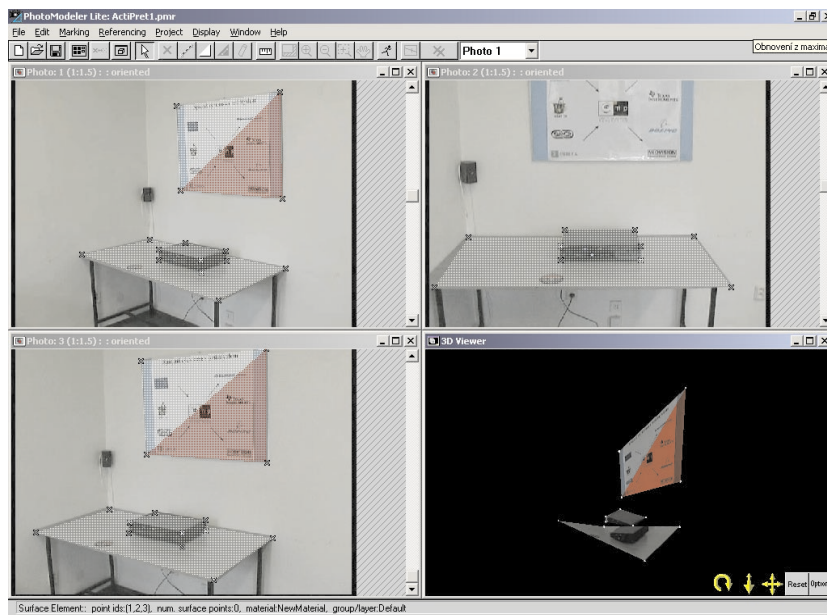


Figure 6: Example of the tested SW: Photo Modeler.

Creating the models of the scene has been manual so far and there probably always will be the necessity of direct human intervention in the process. Some SW tools (Meta Creations Canoma, Photo Modeler, 2d3 Boujou) however, have a potential to make the process easier and more automated as was mentioned in section dedicated to the experiments performed. The automation of the reconstruction process will not go then to the tools mentioned above. Most likely we will stay with Canoma in combination with the direct VRML code editing.

3.6 Outlook

The virtual reality presentation will be able to show the model of the scene in which the activity is going on and the humanoid model animated to perform the observed activity. This will be enhanced by a possibility of a text overlay that can be used for displaying the activity plan when viewing the activity animation. The text overlay was suggested by reviewers in the Heraklion review meeting in June 2002.

We intend to define the structure of the scene and create the prototypes of the objects. The object prototype should accommodate the information on the kind of manipulation that is possible with it. This will be the part of creating functionality of the reconstructed scene. The object geometry reconstruction is to be partially automated with the help of the software tools that have been tested.

4 Activity Presentation By Animating Virtual Humanoid Model

4.1 Task specification

The WP7 provides Virtual Reality module that will work offline. There is no need for any real time simulation. The learned activity can be thus presented by animating the simple humanoid avatar (a virtual human model that looks realistic, but does not necessarily have to function in a realistic way). This approach to create and perform training is common in passenger airplanes equipped with the TV screen. Security instructions are often presented by an avatar. The avatar used in the project deliberately does not resemble any real person.

The module will display virtual humanoid body model that will perform the animation of the learned activity in the environment provided by the 3D reconstruction of the scene. The term *animation* will appear in two

contexts, referring to the motion of avatar itself and more often to the data (spatial coordinates, angles and time) describing the motion. The activity plan represented by the conceptual language from WP5 will be the base for creating these animations.

The usable VRML models of humanoid body that correspond to the H-Anim 1.1 standard are available in the computer graphics community. The task is to define the format of animation easily connectable to the H-Anim 1.1 humanoid body and to design a process which creates the animation data describing the learned activity.

4.2 State-of-the-art

In context of the specified problem of off-line animation creating and viewing it on a simple humanoid model, we will mention the problems of the humanoid body modelling and more interestingly several issues concerning its animation.

4.2.1 Humanoid body modeling

The standard for the humanoid body modeling already exists, designed by H-Anim, the Humanoid Animation Working Group [20] of the Web3D Consortium [28]. This specification of the standard mainly describes the hierarchical structure of the humanoid body, defines the names of the joints in the hierarchy and provides typical anthropometrical data. It also suggests a possible VRML implementation.

The implementation of the humanoid body geometry can basically be a segmented one or a single mesh one [1].

1. The segmented implementation uses separate geometry for each body segment such as thigh, calf, etc. It is less realistic, because the segments, although they are moved by the joint transformations, remain of the constant shape. On the other hand, this approach is easy to implement and is not computationally demanding.
2. The single mesh implementation is much more realistic. The body is created as a single polygonal mesh and each segment carries information (indexes) of the points belonging to it for resolving the joint transformations. This is more demanding for the computational power, but it can provide realistic effect of body deformation when moving.

The single mesh implementation can be of more levels of realism. Variations come from the way how the body deformation is simulated. An example

of highly realistic muscle deformation [2] is the multilayered implementation. The first layer is the skeleton, the joint hierarchy carrying the information of the body pose. Upon it a layer that simulates the behavior of muscles and soft tissues is built. It is basically the shape of a body consisting of ellipsoidal metaballs, the implicit surfaces that are not visible otherwise but via its influence on the surface of the model. Finally, the outermost and the only visible layer is the skin constituting the mesh of points that are finally rendered. The process of creating this mesh is not trivial. The implicit surface of the metaball layer is sampled. The sampled points become the control points of the B-spline patch for each body part. These patches are tessellated for seamless joining of the skin parts. Their polygonal representation is created to become the actual skin layer. Such models are usually enhanced by a physically simulated hair and cloths.

Moreover, there is more to human modelling than just the body geometry. So far we have discussed basically just a simple avatar, the entity driven by predefined animations. Also the models, that can act realistically and follow the laws of physics have been created. Currently a lot of research [21, 25] has been performed in the field of autonomous agents and even an artificial life. The models respond realistically to events coming from the environment and their inner states. The biomechanically correct models are becoming also psychologically functional, thus allowing to extend the area of human simulation as far as to simulation of an individual and even a crowd behavior.

4.2.2 Humanoid body animation

The VRML humanoid animation is based on usage of rotation interpolators. The interpolator is a special VRML node that is responsible for the dynamical behavior of the scene and linearly interpolates between key values given to it. The linear interpolation can be a drawback but it can be reduced by appropriate selection of key-frames. A set of these interpolators is connected to the set of body joints and sends the current rotation values to the joints. The semantics of the interpolators data are key-frames and relevant key values. Each animation requires its own set of interpolator data. The more complicated the animation is the more data it needs.

Creating the animation. The first problem concerning the animations is how to obtain the animation data itself.

The first option is to create animations manually using some animation SW. The animation can be composed either by defining key-frames and relevant key values directly or by using inverse kinematics algorithms to find the key value rotations from a user-defined trajectory of the end-effector. This

approach does not require special HW, but can cost a lot of time. It requires the skilled human animator.

The second option is the motion capture technology. The motion capture is a process of recording real actor's motion in order to animate a virtual character. We shall not discuss various technologies in this report. It has become a vivid commercial area recently, namely in the movie and computer games industry. An important thing is, that the motion capture gives us the trajectories of selected points during the recorded motion. Positions of these points can be transformed into the joint rotations. Using the inverse kinematics algorithms for doing this can also reduce the necessary number of tracked points. Recording motion requires an expensive special HW and a 3D animation SW for treating the recorded data. On the other hand, it provides highly realistic animation without much of manual intervention in the process.

An interesting project was presented [5] in the conference Eurographics 2002. A tool for creating the animation by a virtual puppetry (performance animation) was presented. This technique interactively controls the character motion. This project is a good example of combining a physical simulation with an interactive user input that is based on the use of two motion trackers moved by the user (the puppeteer).

Storing the data. Another problem is how to store created animation data.

Avatars (avatar is the users arbitrary projection to the VR world) from multi-user VR environments [10] are usually animated with the limited set of simple gestures. These animations do not use many interpolators (also the structure of such an avatar is simple and implements just a few joints) and the interpolators do not have many key-frames. In this case the animations can be stored in the same file with the geometry of the avatar.

However the application might require a larger set of animations or a possibility to add or remove the animation. In our ActIPret project, it is better to separate the animations from the humanoid definition. Although there has been an attempt to create a standard interface between the H-Anim humanoid and the animation [14], it still remains a problem to be solved.

Vector representation of motion data. The motion data can be viewed as a vector. Therefore we can treat simple motions as a vector space. This is possibly the way to motion synthesis and parametrization based on the example set of sample motions. There has been research on this topic performed by the VRLab (Lausanne, Switzerland) [4]. They describe how the

motions are synthesized from the set of similar, but distinct motions.

4.3 Available tools

The area of virtual humanoid animation is rapidly developing. It reflects the growing demand in entertainment industry and other areas. Many professional and semi-professional tools for creating such animations using the sophisticated inverse kinematics algorithms exist (Character Studio, Maya, etc.). Also simpler programs for editing gestures of VR avatars (Avatar Studio) are available. These programs usually use their own humanoid body model, not corresponding to the H-Anim standard.

We have found only one program that is able to work with an H-Anim compliant humanoid body. It is Parallel Graphics Internet Character Animator [22]. It is a tool for creating the animation manually.

Since we do not need to create the animation data, we have no need for any tools mentioned above. Our task is to transfer data from other ActIPret modules (e.g., a hand tracker) to the form of the VRML rotation interpolator set. For this specialized tools have to be implemented.

Concerning the presentation of the animation, a suitable basis for development of an H-Anim compliant animation viewer is available.

4.4 Our choice

We use an H-Anim 1.1 compliant humanoid body. At this stage it is a model with segmented structure, downloaded from LIG (Computer Graphics Lab, EPFL, Lausanne, Switzerland) [26] web site, where it is available for non-commercial use. This model is fully sufficient at its level of realism for the purpose of WP7, though we are planning on creating our own deformable mesh humanoid body.

The animation data contained in a special prototype VRML node are separated from the humanoid geometry definition.

The animations for the ActIPret CD scenario will be created using the hand tracker data and inverse kinematics algorithms. As a subject of research, there is a way of creating animations based on some sort of a conceptual language and a database of parameterized gestures or poses.

4.5 Initial experiments

The experimental program tool for viewing the VRML based animations has been created (Fig. 7). We have tested possibilities of interactive connecting and disconnecting the animation to the H-Anim 1.1 humanoid.

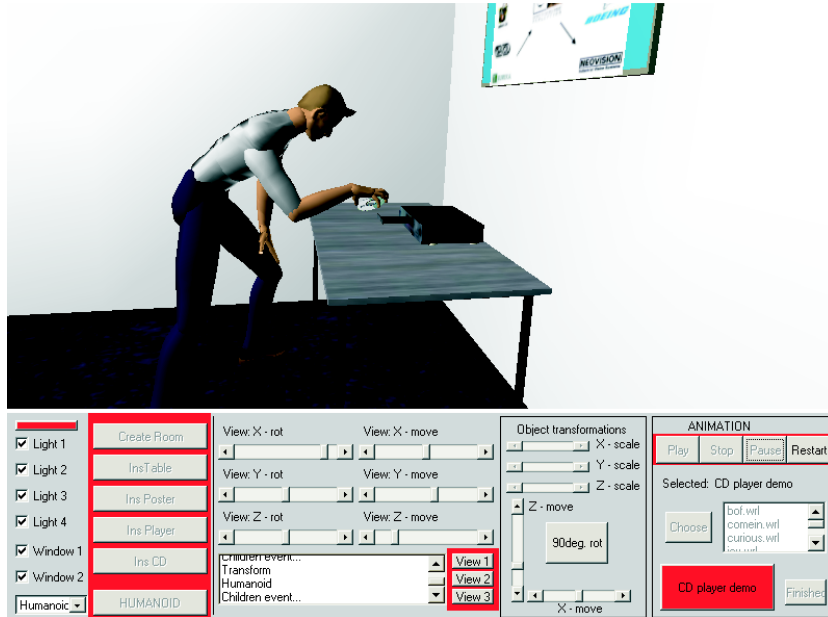


Figure 7: User interface of the experimental program.

A CD scenario animation for demonstration purposes was created. The process of creating it was fully manual again. The work on it was aimed to verify first ideas of creating animation of reusable (possibly parameterized) poses or gestures. The implementation of humanoid's influence on the other objects (displacing the CD) is temporary.

4.6 Summary of experience gained

We have created an interface between the H-Anim 1.1 compliant standard humanoid and an animation. We are able to connect and disconnect the animation to the humanoid and view it.

4.7 Outlook

We plan for the future to put the animations together from shorter animations, possibly poses. This should reflect the fact that an activity consists of simpler actions. Synthesis of the animations to be put together will be a subject of research. The humanoid model can be animated with the trajectories from the tracker, but the possibility of using the motion vector space representation for synthesis is not to be neglected.

We also work on remaking animation prototype to accommodate the possibility of humanoid influencing other objects in the scene. Subsequently the

format of animation would be not just a recorded sequence of joint angles but also a record of events caused by the activity.

5 Conclusions

The interfaces to the framework and other WPs of the ActIPret project and their modules have been suggested. Standard VR tools (VRML - ISO/IEC 14772 - and Java language) have been selected. The possible background scene modelling procedure has been described as well as an approach to create the 3D models of objects in the scene. The articulated 3D model of the human already exists and has been standardized, as is mentioned in this document too.

The WP7 task has been divided into two subtasks - the scene reconstruction and visualization and the activity presentation by animating the virtual humanoid body model. In both directions several experiments have been done.

The scene reconstruction and creation of realistically functional models has not been possible without a significant amount of user intervention. We have performed experiments with various kinds of program tools for 3D reconstruction from images in order to make the scene reconstruction process automated. However, we doubt the possibility of fully automatic scene reconstruction at least in the ActIPret project time span.

In the area of animating the virtual humanoid body model, we use the VRML H-Anim standard humanoid body. We are able to connect the animation (as a set of rotation interpolators) to the model of the body and thus view it. We have defined the interface between the model of body and the animation.

As a part of these experiments the Java applet for demonstration the VRML capabilities and carrying out our tests have been created. For the demonstration purposes, the scene and animation of ActIPret CD scenario viewable by the applet mentioned above have been created.

Month 18 demo plan

For the review in month 18 we shall attempt for the month 24 goals, marked in Annex 1 - "Description of Work" as *M2 - Presentation of the scenario "placing the CD in player"*. This includes creating the 3D models of the background scene possibly extracted from image data and demonstration of learned activity (may be in a simplistic way).

As the WP 7 does not lie on project's main road, we choose pragmatic approach to solving its problems. Most of the reconstruction is done and will be done manually or by combining existing software tools.

There are two task that are being worked on and are promising to be finished for the month 18 meeting. One is to create the specification of data format to describe the learned activity to the presentation module. The second is creating the animation of the activity as a sequence of short animations of actions connected together.

The scenario presented by the demo will be the CD scenario again.

References

- [1] Thalmann D. Babski C. A seamless shape for HANIM compliant bodies, 1999. <http://ligwww.epfl.ch/~babski/StandardBody/Deformation/>.
- [2] R. Boulic, P. Fua, L. Herda, M. Silaghi, J. Monzani, L. Nedel, and D. Thalmann. An anatomic human body for motion capture. In *Proceedings of EMMSEC '98, Bordeaux*, 1998.
- [3] Pavel Krsek, Tomáš Pajdla, Václav Hlaváč, and Ralph Martin. Range image registration driven by a hierarchy of surface differential features. In M. Gengler, M. Prinz, and E. Schuster, editors, *22nd Workshop of the Austrian Association for Pattern Recognition*, pages 175–183, Piaristengasse 19, Wien, Austria, May 1998. Osterreichische Computer Gesellschaft.
- [4] Ik Soo Lim and Daniel Thalmann. A vector-space representation of motion data for example-based motion synthesis. *Deformable Avatars edited by Nadia Magnenat-Thalmann and Daniel Thalmann, Kluwer Academic Publishers*, pages 169 – 179, 2001.
- [5] Sageev Oore, Demetri Terzopoulos, and Geoffrey Hinton. Local physical models for interactive character animation. In *Computer Graphics Forum, The International Journal Of The Eurographics Association*, volume 21, pages 337 – 346, September 2002.
- [6] Tomáš Pajdla and Václav Hlaváč. Camera calibration and Euclidean reconstruction from known observer translations. In *International Conference on Computer Vision and Pattern Recognition*, pages 421–426, Santa Barbara, California, USA, June 1998. IEEE.

- [7] Martin Urban, Tomáš Pajdla, and Václav Hlaváč. Consistent projective reconstruction from multiple views. In Aleš Leonardis, Franc Solina, and Ruzena Bajcsy, editors, *Proceedings NATO Advanced Research Workshop Confluence of Computer Vision and Computer Graphics*, NATO Science Series ; Ser. 3 : High Technology, pages 49–67, Dordrecht, Germany, August 2000. Kluwer Academic Publishers.
- [8] Martin Urban, Tomáš Pajdla, and Václav Hlaváč. Projective reconstruction from n views having one view in common. In Bill Triggs, Richard Szeliski, and Andrew Zisserman, editors, *Vision Algorithms: Theory & Practice*, LNCS, pages 116–131, Berlin, Germany, September 2000. Springer.
- [9] Tomáš Werner, Tomáš Pajdla, and Martin Urban. REC3D: Toolbox for 3D Reconstruction from Uncalibrated 2D Views. Technical Report CTU-CMP-1999-4, Czech Technical University, FEL ČVUT, Karlovo náměstí 13, Praha, Czech Republic, December 1999.
- [10] S. K Wilcox. *Web developer.com Guide to 3D Avatars*. John Wiley and Sons, Inc., Canada, 1998.
- [11] J. Zara. Concise tour to the virtual old Prague. In *EUROGRAPHICS 2002, Short Presentations*, pages 191–198, September 2002.
- [12] J. Zara. On the complexity of web-based presentations of large urban scenes. In *Proceedings of East-West-Vision 2002, Graz*, pages 99–108. Osterreichische Computer Gesellschaft, September 2002.
- [13] J. Zara, P. Chromy, Cizek J., Ghais K., Holub M., Mikes S., and Rajnoch J. A scalable approach to visualization of large virtual cities. In *Proceedings Fifth International Conference on Information Visualisation*, pages 639–644. IEEE Computer Society Press, 2001.
- [14] Brief information on the animation library. <http://ligwww.epfl.ch/~babski/StandardBody/Animation.html>.
- [15] Blaxxun VRML browser. <http://developer.blaxxun.com/>.
- [16] 2d3 Boujou homepage. <http://www.2d3.com/2d3/products/boujou.shtml>.
- [17] MetaCreations Canoma homepage. <http://www.metacreations.com/products/canoma/>.

- [18] Parallel Graphics Cortona VRML browser. <http://www.parallelgraphics.com/products/cortona/>.
- [19] Cosmo Player VRML browser. <http://ca.com/cosmo/>.
- [20] H-Anim (Human Animation Working Group). <http://www.hanim.org>.
- [21] MIRALab homepage (University of Geneva). <http://www.miralab.unige.ch/newMIRA/MIRALabHtml.htm>.
- [22] Paralel Graphics VrmlPad page. <http://www.parallelgraphics.com>.
- [23] PhotoModeler homepage. <http://www.photomodeler.com/>.
- [24] Spazz3D editor homepage. <http://www.spazz3d.com>.
- [25] Demetri Terzopoulos home-page. <http://www.cs.toronto.edu/~dt/>.
- [26] Virtual reality lab (former LIG), EPFL Lausanne. <http://vrlab.epfl.ch>.
- [27] VRML 97 (ISO/IEC 14772) specification. <http://www.vrml.org/technicalinfo/specifications/vrml97/index.htm>.
- [28] Web3D Consortium. <http://www.web3d.org>.