3-D Devices and Virtual Reality in Human Animation

Nadia Magnenat Thalmann
University of Geneva
Geneva, Switzerland

Daniel Thalmann
Computer Graphics Lab
Swiss Federal Institute of Technology
Lausanne, Switzerland

ABSTRACT

In this paper, we show how 3-D input devices used in virtual reality may drastically change the way of designing animation scenes. We analyze the various stages of animation production and show when and how these new concepts/devices may be involved into the creation process. We illustrate this new approach by describing the *5th Dimension* animation system developed in our laboratories. We emphasize typical applications in surface sculpting, facial animation, cloth design, camera path design, gesture input and animator-actor communication.

1. INTRODUCTION

The first 2D animation systems like ANIMATOR (Talbot et al. 1971) were command-oriented and not graphics-based systems. However, very soon real-time animation systems have been developed allowing interactions to define path and motion. Now, paint systems incorporate facilities to define motions using a graphics tablet and several 2D animations systems are available on the market with real-time capabilities and real-time mouse-based interaction, e.g. Macromind DIRECTOR.

In the area of 3-D animation, the evolution is much more different. Early 3-D computer-generated films have been produced using programming. First 3-D interactive animation systems like BBOP (Stern 1983) or TWIXT (Gomez 1984) were keyframe systems. These keyframe-based tools are now standard possibilities in commercial products like Alias, Wavefront, Explore or SoftImage. The animator could easy define simple motions, however it was rather limited for implementing complex motions, especially when they are basically physical. For this reason, systems were developed to allow the definition of choreographies involving cameras, actors, lights, based on interactive commands. Then, many authors introduced physics concepts into animation in order to improve the quality of motion. Most physics-based systems are menu-driven without 3-D graphics interaction.

Now, with the existence of graphics workstations able to display complex scenes containing several thousands of polygons at interactive speed, and with the advent of such new interactive devices as the SpaceBall, EyePhone, and DataGlove, it is possible to create applications based on a full 3-D interaction metaphor (Balaguer and Mangili 1991) in which the specifications of deformations or motion are given in real-time. In this paper, we try to explain how 3-D input devices used in virtual reality may drastically change the way of designing animation scenes.

We analyze the various stages of animation production and show when and how these new concepts/devices may be involved into the creation process. We illustrate this new approach by describing the *5th Dimension* animation system developed in our laboratories at the University of Geneva and the Swiss Federal Institute of Technology in Lausanne.

2. THE ANIMATION PROCESS

Designing an animation sequence (Magnenat-Thalmann and Thalmann 1990) consists of creating a scene, characterized by a description, called a script. Each scene contains static objects grouped into a decor and animated objects that change over time according to motion laws. Moreover, in a 3-D space, scenes are viewed using virtual cameras and they may be lit by synthetic light sources. These cameras and lights may evolve over time as though manipulated by cameramen. In order to create all the entities and motions, coordinate and synchronize them, known collectively as choreography, it is necessary to know the appearance of the scene at this time and then Computer Graphics techniques allow us to build and display the scene according to viewing and lighting parameters. The problems to solve are how to express time dependence in the scene, and how to make it evolve over time.

In our case, our scenes involve synthetic actors which means more complex problems to manage. Synthetic actors have a very irregular shape hard to built especially for well-known personalities. Once the initial human shape has been created, this shape should change during the animation. This is a very complex problem to ensure the continuity and the realism of the deformed surfaces. The human animation is very complex and should be split into body motion control and facial animation. Basically a synthetic actor is structured as an articulated body defined by a skeleton. Skeleton animation consists in animating joint angles. There are two main ways to do that: parametric keyframe animation and procedural animation based on mechanical laws. The face is a small part of a synthetic actor, but it plays an essential role in the communication. People look at faces for clues to emotions or even to read lips. It is a particular challenge to imitate these few details. An ultimate objective therefore is to model human facial anatomy exactly including its movements to satisfy both structural and functional aspects of simulation. This however, involves many problems to be solved simultaneously. Some of these are: the geometric representation must be very close to the actual facial structure and shape, modeling of interior facial details such as muscles, bones, tissues, and incorporating the dynamics and movements involved in making expressions etc. Each one of these is, in itself, a potential area of research.

Finally, animation scenes may be rendered using techniques like ray-tracing and radiosity.

3. THE NEW HARDWARE FOR ANIMATION

3.1. The real-time capabilities

Until recently, the greatest obstacles in the elaboration of intuitive human-machine interaction methods for 3-D graphical applications came from basically insufficient computing power, slow frame rates and the lack of adequate multi-dimensional input devices. Highly interactive applications that require uninterrupted interaction in 3-D space rely on fast display rates in order to assure that the user may view the result of his actions without any perceived time delays.

Now superworkstations like the IRIS 4D VGX series are able to display up to 1000000 shaded polygons. From an animation point of view, it corresponds to a maximum value of 40000 polygons/frame at the European normal video frame rate. In fact, this value is theoretical and cannot be reached using an animation software. However, 3-D interaction don't really need 40000 polygons per frame and can work at a 12 frame/sec rate or even less.

3.2. 3-D Real-time Multi-dimensional input devices

In the next sections, we present four popular 3-D devices for virtual reality: the DataGlove , the SpaceBall, the 3-D Polhemus digitizer and the EyePhone. More details may be found in (Balaguer and Mangili 1991; Brooks 1986; Fisher et al. 1986).

3.2.1. DataGlove

Hand measurement devices must sense both the flexing angles of the fingers and the position and orientation of the wrist in real-time. Currently, the most common hand measurement device is the DataGlove from VPL Research. The DataGlove consists of a lightweight nylon glove with optical sensors mounted along the fingers. In its basic configuration, the sensors measure the bending angles of the joints of the thumb and the lower and middle knuckles of the others fingers, and the DataGlove can be extended to measure abduction angles between the fingers. Each sensor is a short length of fiberoptic cable, with a light-emitting diode (LED) at one end and a phototransistor at the other end. When the cable is flexed, some of the LED's light is lost, so less light is received by the phototransistor. Attached to the back is a 3Space Isotrack system (see 3.2.3) to measure orientation and position of the gloved hand. This information, along with the ten flex angles for the knuckles is transmitted through a serial communication line to the host computer. Knuckle data can be sampled at up to 60 Hz.

3.2.2. SpaceBall

Some people have tried to extend the concept of the mouse to 3-D. In order to address this problem, Spatial Systems designed a 6 DOF interactive input device called the SpaceBall. This is essentially a “force” sensitive device that relates the forces and torques applied to the ball mounted on top of the device. These force and torque vectors are sent to the computer in real time where they are interpreted and may be composited into homogeneous transformation matrices that can be applied to objects. Buttons mounted on a small panel facing the user control the sensitivity of the SpaceBall and may be adjusted according to the scale or distance of the object currently being manipulated. Other buttons are used to filter the incoming forces to restrict or stop translations or rotations of the object.

3.2.3. Polhemus 3Space Isotrack

Magnetic tracking devices have been the most successful and the Polhemus 3Space Isotrack, although not perfect, is the most common one. A source generates a low frequency magnetic field detected by a sensor.

3.3. The stereo displays and the head-mounted displays

Binocular vision considerably enhances visual depth perception. Stereo displays like the StereoView option on Silicon Graphics workstations may provide high resolution stereo real-time interaction.

The EyePhone is a head-mounted display system which presents the rich 3-D cues of head-motion parallax and stereopsis. It is designed to take advantage of human binocular vision capabilities and presents the general following characteristics:

- headgear with two small LCD color screens, each optically channeled to one eye, for binocular vision.
- special optics in front of the screens, for wide field of view
- a tracking system for precise location of the user's head in real time; it is the Polhemus 3Space Isotrack composed of a source generating a low frequency magnetic field detected by a sensor.

3.4. Real-time video input with the Living Video Digitizer

Input video is now a standard tool for many workstations. However, it generally takes a long time (several seconds) to get a complete picture, which makes the tool useless for real-time interaction. For real-time interaction and animation purpose, images should be digitized at the traditional video frame rate. One of the possibilities for doing this is the Living Video Digitizer (LVD) from Silicon Graphics. With the LVD, images are digitized at a frequency of 25 Hz (PAL) or 30 Hz (NTSC) and may be analyzed by the animation program.

4. THE *5TH DIMENSION* ANIMATION SYSTEM

4.1. The architecture of the system and the *5th dimension* toolkit

The 5th Dimension Project is a large research project in three-dimensional animation and visualization. The main objective of the project is the animation of synthetic actors in their environment, which involves a number of related areas of computer animation and scientific visualization. In particular, the following applications are being developed:

- animation of articulated bodies based on mechanical laws
- vision-based behavioral animation
- hair rendering and animation
- object grasping
- facial animation
- personification in walking models
- synchronization in task-level animation
- deformation of flexible and elastic objects
- cloth animation with detection of collision

To coordinate efforts and allow good communication between the various applications, a toolkit of high-level dynamic graphical classes, both two and three dimensional, has been constructed. This toolkit, called the *5th Dimension Toolkit* uses a uniformly object-oriented design for all its data structures, resulting in a high degree of integration between various applications.

The *5th dimension* animation system is intended to offer to the animator a full 3-D interaction including the possibility to enter into the virtual world and to communicate with synthetic actors. The hardware used is composed of 21 Silicon Graphics IRIS workstations including three Powervision (VGX). Most *5th Dimension* applications take advantage of visual 3-D interfaces using the various 3-D devices available in our laboratories: two datagloves, several SpaceBalls, an EyePhone, a 3-D Polhemus digitizer, a living video digitizer, a StereoView station and a synthesizer keyboard controlled by a NeXT Cube workstation.

Fig.1 shows the architecture of the 5th Dimension animation system. In the current version, 6 applications provide a user interface based on 3-D devices:

- the sculpting program SURFMAN
- the Muscle and Expression editor in the SMILE Facial Animation system
- the cloth design software
- the hand gesture recording system in GESTURE LAB
- the program to create 3-D paths for cameras, objects and light sources
- a communication program animator-actor (in development)

The first three programs, are mainly based on the ball and mouse metaphor described in the next Section. SURFMAN may also take advantage of StereoView and the 3-D Polhemus digitizer. Hand gestures are recorded using the DataGlove and 3-D paths are mainly generated using the SpaceBall. We are developing a way of creating camera paths based on the EyePhone. The communication program animator-actor uses the Living Video Digitizer to capture the animator face. Other applications in the *5th Dimension* system are only based on mouse interaction.

In the next sections, we only describe the applications already working and using a full 3-D real-time interaction.

4.2 The Space and Ball metaphor

Visual feedback, in a typical computer graphics application that requires items to be positioned or moved in 3-D space, usually consists of a few orthogonal and perspective projection views of the same object in a multiple window format. This layout may be welcomed in a CAD system where, in particular, an engineer might want to create fairly smooth and regular shapes and then acquire some quantitative information about his design. But in 3-D applications where highly irregular shapes are created and altered in a purely visual and esthetic fashion, like in sculpting or keyframe positioning, this window layout creates a virtually unsolvable puzzle for the brain and makes it very difficult (if not impossible) for the user of such interfaces to fully understand his work and to decide where further alterations should be made.

In essence, *motion parallax* consists of the human brain's ability to render a three-dimensional mental picture of an object simply from the way it moves in relation to the eye. Rotations offer the best results because key positions located on the surface move in a larger variety of directions. Furthermore, in a perspective projection, depth perception is further accentuated by the speed in which features flow in the field of view ¾ points located closer to the eyes move faster than the ones situated in back. In a 3-D application, if motion parallax is to be used effectively, this implies the need for uninterrupted display of object movements and thus the requirement for hardware capable of very high frame rates. To acquire this depth perception and mobility in a 3-D application, we make use of a SpaceBall.

When used in conjunction with a common 2-D mouse such that the SpaceBall is held in one hand and the mouse in the other, full three-dimensional user interaction is achieved.

The SpaceBall device is used to move around the object being manipulated in order to examine it from various points of view, while the mouse carries out the picking and transformation work onto a magnifying image in order to see every small detail in real time (e.g. vertex creation, primitive selection, surface deformations, cloth panel position, muscle action). In this way, the user not only sees the object from every angle but he can also apply and correct transformations from every angle interactively.

In order to improve our approach using stereo display, we also use “StereoView”.

4.3 Sculpting human faces

Our first application based on the ball and mouse metaphor consists of sculpting and animating human faces based on highly irregular polygon mesh surfaces. For example, a realistic human character may be produced (Paouri et al. 1991) with a method similar to the modelling of clay, work which essentially consists of adding or eliminating parts of the material, and turning around the object when the principal form has been set up.

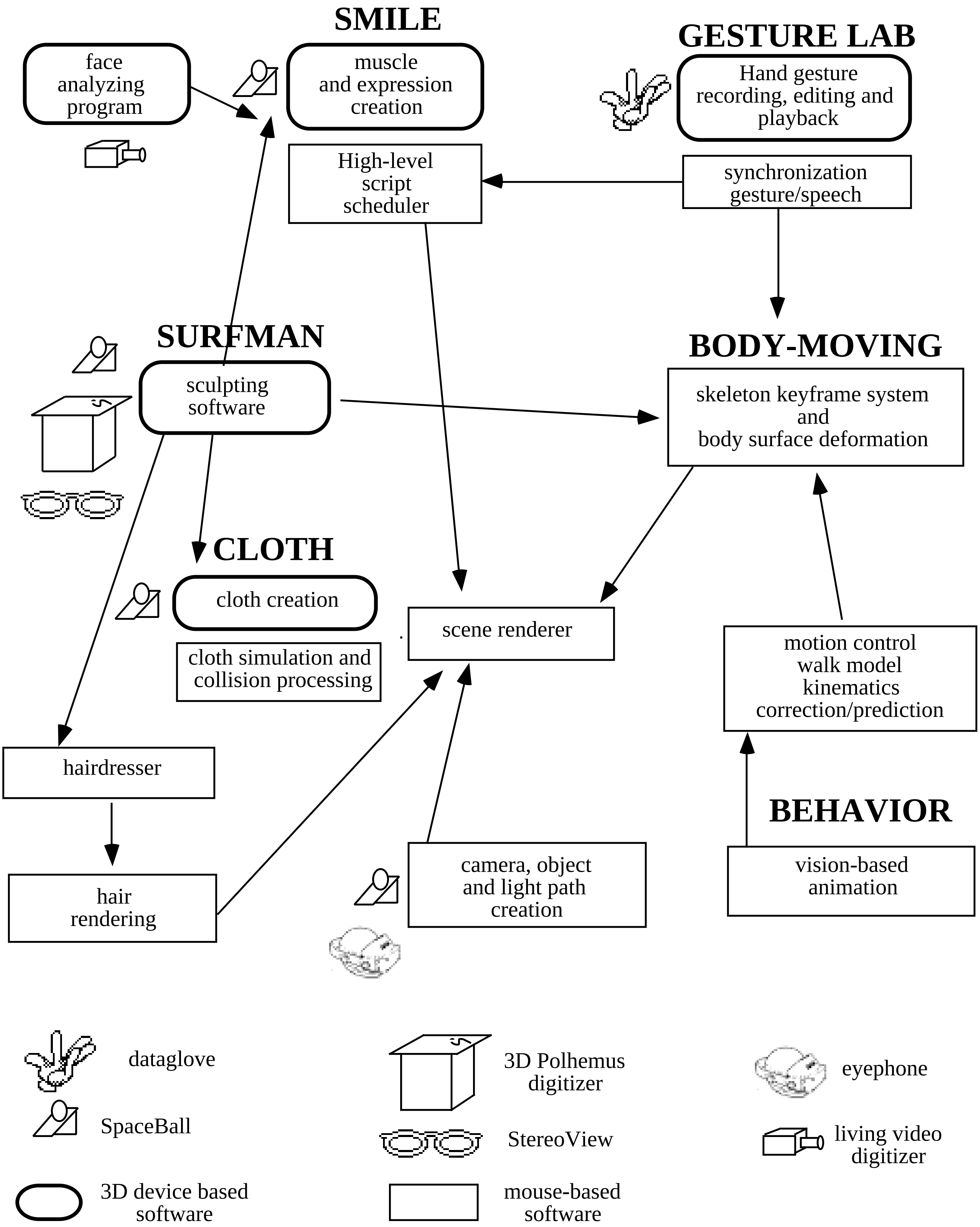

Fig.1 The Architecture of the *5th Dimension* Animation System

The operations conducted in a traditional sculpture can be performed by computer for computer generated objects. A sculpting software (LeBlanc et al. 1991) which is based on the Ball and Mouse methodology. This allows the user to create a polygon mesh surface. With our ball and mouse approach, the operations performed while sculpting an object closely resemble traditional sculpting. The major operations performed using this software include: creation of primitives, selection, local deformations.

Figure 2 shows a head created from a sphere using local deformations.

Fig.2 Head created from a sphere

4.4 Muscle and Expression Edition for Facial Animation

Ensuring synchronization of eye motion, expression of emotion and word flow of a sentence, as well as synchronization between several actors, are the heart of the SMILE facial animation system (Kalra et al. 1991). This system is based on a methodology for specifying facial animation based on a multi-layered approach. Each successive layer defines entities from a more abstract point of view, starting with muscles deformations, and working up through phonemes, words, sentences, expressions, and emotions.

At the lowest level, to simulate the muscle action on the skin surface of human face, we developed a 3-D interactive system to define regions on the face mesh which correspond to the anatomical description of the facial region on which a muscle action is desired. In this system based on the Ball and Mouse methodology, a parallelepiped control unit can be defined on the region of interest. The deformations which are obtained by actuating muscles to stretch, squash, expand and compress the inside volume of the facial geometry, are simulated by displacing the control point and by changing the weights of the control points of the control-unit. The region inside the control-unit deforms like a flexible volume, corresponding to the displacement and the weights of the control points. For the facial deformations, we use Free Form Deformation (FFD) which is a technique for deforming solid geometric models in a free form manner (Sederberg and Parry, 1986). It can deform surface primitives of any type or degree, for example, planes, quadrics, parametric surface patches or implicitly defined surfaces. FFD involves a mapping from R^3 to R^3 through a trivariate tensor product Bernstein polynomial. Physically, FFD corresponds to deformations applied to an imaginary parallelepiped of clear, flexible plastic in which are embedded the object(s) to be deformed. The objects are also considered to be flexible so that they are deformed along with the plastic that surrounds them. The high level layer allows the manipulation of these entities, ensuring synchronization of the eye motion with emotions and word flow of a sentence. A language for synchronizing speech, emotions and eye motions is developed to provide a way to naturally specify animation sequences.

4.5 Cloth Design

In our new cloth animation software (Magnenat Thalmann et al. 1991), an article of clothing consists of several panels seamed together, so its data structure contains the panel data and the information about seaming the panels, together.

The creation of clothes may be divided into four steps:

1. design of the shape of each cloth panel in 2D space
2. definition of the physical parameters
3. transfer of the 2-D panels into 3-D space on ruled surfaces
4. seaming panels and attaching them to the actor's body by internal forces.

As a tailor, we cut the cloth panel from a 2D rectangular mesh cloth. The panels can be various polygonal shapes. It is as convenient and easy to design the clothes as to buy a dress pattern and cut the cloth according to the pattern. At this point, we also decide which panel edges will be seamed together, and which edges will be attached to the actor's body.

Using a SpaceBall we rotate and zoom on the human body or other objects and, using the mouse, we first identify key points on the human body or on the objects, and then create cloth panels passing each of these points, finally we seam these panels together to form the garment or cloth things, such as skirts, T-shirts and pants. For fashion design, the 3-D cloth panels should be transformed into 2D polygonal panels in order to produce cutting patterns. With the elastic surface model the 2D panels can be put on the human body and deformed to 3-D panels by applying external forces, and seamed by using local constraints. With the 3-D interactive cloth definition interface we can conveniently design and create various kinds of clothes. We can also set the physical parameters of the clothes with respect to their texture and materials. Plate 1 shows a cloth created using this 3-D interface.

4.6. Physically-Based Interactive Camera Motion Control Using 3-D Input Devices

In this application (Turner et al. 1991), naturalistic interaction and realistic-looking motion is achieved by using a physically-based model of the virtual camera's behavior. The approach consists to create an abstract physical model of the camera, using the laws of classical mechanics to simulate the virtual camera motion in real time in response to force data from the various 3-D input devices. The behavior of the model is determined by several physical parameters such as mass, moment of inertia, and various friction coefficients which can all be varied interactively, and by constraints on the camera's degrees of freedom which can be simulated by setting certain friction parameters to very high values. This allows us to explore a continuous range of physically-based metaphors for controlling the camera motion. A physically-based camera control model provides a powerful, general-purpose metaphor for controlling virtual cameras in interactive 3-D environments. When used with force-calibrated input devices, the camera metaphor can be reproduced exactly on different hardware and software platforms, providing a predictable standard interactive "feel". Obviously, pressure-sensitive input devices are usually more appropriate because they provide a passive form of "force-feedback". In our case, the device that gave the best results is the SpaceBall.

The relationship between device input and virtual camera motion is not as straightforward as one might think. Usually, some sort of mathematical function or "filter" has to be placed between the raw 3-D input device data and the virtual camera viewing parameters. Several recent papers have proposed and compared different metaphors for virtual camera motion control in virtual environments using input devices with 6 degrees of freedom (Ware and Osborne 1990, Mackinlay et al 1990). These metaphors are usually based on a kinematic model of control, where the virtual camera position, orientation, or velocity is set as a direct function of an input device coordinate.

The interactive camera control metaphor is based on physical modeling of the virtual camera, using forward dynamics for motion specification. The important mechanical properties of this model which affect its motion are its mass, its moments of inertia, and the coefficients of friction and elastic forces imposed by the camera mount. The general motion of a rigid body such as a camera can be decomposed into a linear motion of its center of mass under the control of an external net force and a rotational motion about the center of mass under the control of an external net torque. Plate 2 shows the virtual camera driven by a force and a torque.

4.7. A gesture-oriented animation system based on the use of the DataGlove

This gesture-oriented animation system (Mato Mira 1991) consists basically of two programs: GESTURE LAB, which enables an animator to record, playback and edit real hand movements using a VPL Research DataGlove; and VOGE (VOice + GEsture), a program which accepts a monologue script consisting of phrases, emphasis and rhythm parameters, and gesture names to generate an animation sequence in which lip and hand movements are synchronized according to the specification. The output of VOGE is fed into the BODY-MOVING system to obtain the final animation scene.

The DataGlove is used in GESTURE LAB to measure the angles from the metacarpophalangeal joints. The user can select a previously recorded sequence to be played, and insert a part of this sequence in a new one, or perform a live recording using the DataGlove. There is a possibility of setting timers to have precise control over the duration and starting point of a playback or live performance.

Even if the current version of GESTURE LAB only allows the recording of the performance of a hand, nothing would disallow its extension for full body movement sampling, given the availability of the corresponding devices (e.g. DataSuit).

Fig.3 shows the GESTURE LAB interface.

Fig.3 The GESTURE LAB Interface

4.8. A Behavioral communication system based on video input

At the behavioral level, we may consider emotional communication between the actor and the animator. We may restrict emotions to a few, such as happiness, anger, and sadness and consider only the facial expressions as manifestations of these emotions. In such an emotional communication system, the animator may smile, his face is recorded in real-time using the Living Video Digitizer and the emotion is detected using an image processing program. The dialog coordinator decides which emotion should be generated in response to the received emotion. This emotion is translated into facial expressions to be generated by our facial animation system SMILE. The complete process is explained in Fig.4.

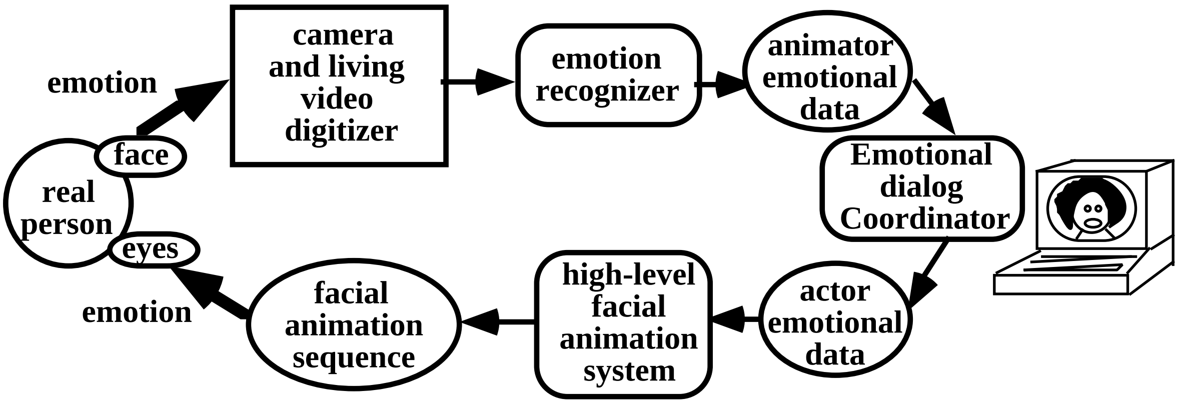

Fig.4 Principle of the communication animator-actor

Consider the example where Marilyn smiles when the animator smiles. The difficulty in such a process is to decide whether the animator is smiling based on the analysis of the image captured by the living video digitizer. At present, our experience is developed based on snakes (Terzopoulos and Waters 1991). This is of course a limitation to the system. The animator should be known from the system and must in a preprocessing step adjust the snakes to its face image captured by the video camera. Only after this preprocessing step, the image processing program is able to detect the facial expressions by analyzing the snake transformations. Only small images with a rather limited processing is possible with the actual hardware; this implies that the detection of subtleties of the face is not yet feasible.

5. CONCLUSION

We have shown the importance of the new 3-D devices and the virtual reality approach in human animation. It considerably improves the animator-computer interface. The *5th Dimension* animation system will certainly take more advantage of 3-D input devices and video input in the future in particular for motion control and behavioral animation. The advent of force-feedback devices and image processing programs to analyze images coming from video cameras will bring new dimensions to such systems.

ACKNOWLEDGMENT

The research was partly sponsored by "Le Fonds National Suisse de la Recherche Scientifique"

6. REFERENCES

Balaguer F, Mangili A (1991) Virtual Environments in: N.Magnenat-Thalmann and D.Thalmann (Eds) New Trends in Animation and Visualization, John Wiley and Sons, pp.91-106.

Brooks F.P. Jr (1986) Walkthrough - A Dynamic Graphics System for Simulating Virtual Buildings Proceedings 1986 Workshop on Interactive 3-D Graphics ACM, pp.9-22

Fisher S.S., McGreevy M., Humphries J., Robinett W.,(1986), "Virtual Environment Display System", Proceeding 1986 Workshop on Interactive 3-D Graphics, ACM, pp 77-87

Forrest AR (1986), User Interfaces for Three-Dimensional Geometric Modelling, Proceedings 1986 Workshop on Interactive 3-D Graphics, ACM Press, pp. 237-249.

Gomez JE (1984) Twixt: a 3-D Animation System, Proc. Eurographics '84, North Holland, pp.121-134.

Kalra P, Mangili A, Magnenat-Thalmann N, Thalmann D (1991) SMILE: a Multilayered Facial Animation System, Proc. IFIP Conference on Modelling in Computer Graphics, Springer, Tokyo, Japan

LeBlanc A, Kalra P, Magnenat-Thalmann N, Thalmann D (1991) Sculpting With the "Ball & Mouse" Metaphor - Proc. Graphics Interface '91, Calgary, Canada

Mackinlay J.D., Card S.K, Robertson G. (1990) Rapid Controlled Movement Through a Virtual 3-D Workspace, Computer Graphics 24(4) : 171-176

Magnenat-Thalmann N, Thalmann D (1990) Computer Animation: Theory and Practice, 2nd edition, Springer-Verlag, Tokyo

Magnenat Thalmann N, Ying Y, Thalmann D (1991) The Problematics of Cloth Animation, 2nd Conf. on CAD and CG, Hangzhou, China (to appear)

Mato Mira F (1991) ICSC World Laboratory LAND-5 Project, Computer Graphics Lab, Swiss Federal Institute of Technology, Lausanne, Switzerland

Paouri A, Magnenat Thalmann N, Thalmann D (1991) Creating Realistic Three-Dimensional Human Shape Characters for Computer-Generated Films, Proc. Computer Animation '91, Geneva, Springer-Verlag, Tokyo

Sederberg TW, Parry SR (1986) Free-form Deformation of Solid Geometric Models, Proc.SIGGRAPH'86, Computer Graphics, Vol.20, No4, pp.151-160.

Stern G (1983) Bbop - A Program for 3-Dimensional Animation, Proc. Nicograph '83, pp.403-404

Talbot PA, Car III JW, Coulter Jr. RC, Hwang RC (1971) Animator: An On-Line Two-Dimensional Film Animation System, Communications of the ACM, Vol. 14, No4, pp.251-259

Terzopoulos D, Waters K (1991) Techniques for Facial Modeling and Animation, in: Magnenat Thalmann N, Thalmann D (eds) Computer Animation '91, Springer, Tokyo, pp.59-74

Turner R, Gobbetti E, Balaguer F, Mangili A, Thalmann D, Magnenat-Thalmann N (1990) An Object-Oriented Methodology Using Dynamic Variables for Animation and

Scientific Visualization, in: Chua TS, Kunii TL, CG International '90, Springer, Tokyo, pp.317-328.

Turner R, Balaguer F, Gobbetti E, Thalmann D (1991), Physically-Based Interactive Camera Motion Control Using 3-D Input Devices, in: Patrikalakis N (ed.) Scientific Visualization of Physical Phenomena, Springer, Tokyo, pp.135-145.

Ware C, Osborne S, “Exploration and Virtual Camera Control in Virtual Three Dimensional Environments”, Computer Graphics, 24 (2), pp. 175-183.

Plate 1. Cloth created using a 3-D interface

Plate 2. Camera control using 3-D devices