Holographic and Action Capture Techniques

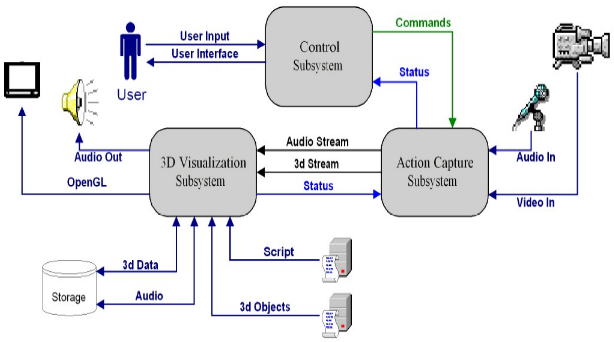
Tomas Rodriguez[1], Adolfo Cabo de Leon[2], Bruno Uzzan[3], Nicolas Livet[4], Edmond Boyer[5], Florian Geffray[6], Tibor Balogh[7], Zoltan Megyesi[8], Attila Barsi[9]

We present an integrated 3D capturing, visualization and user interaction system composed of a computer vision based 3D capturing device, a scene composer and a large scale holographic display. The system performs in real-time and provides the facilities required for capturing realistic human 3D body models, inserting the human representations inside virtual scenarios, detecting 3D interactions between the body models and the virtual objects present in the scene and visualizing the resulting 3D performance on a true 3D holographic display.

Keywords: 3D human action capturing, 3D display, virtual reality, interactive

Introduction

Our Emerging Technology installation will demonstrate the results achieved in the HOLONICS FP6-IST-511318 European project. The installation will demonstrate the technology fusion of a unique large scale holographic display system with a real time computer vision based capturing device capable of producing 3D human body models highly resembling the original geometry, textures and dynamics of a real person. The system will be composed of three main elements: the holographic display, the 3D Capturing Device and the Scene Composer. The complete set-up will allow a human actor to perform in front of the cameras and interact with the virtual objects and scenarios presented in true 3D in front of his eyes in a most realistic and natural way.



1 Schematic Overview

Project Description

The 3D visualization component will consist of a custom 23Mpixel display (1600x900mm screen) capable of recreating the illusion of seeing virtual objects floating at fixed physical locations. The display is truly multi-user and does not require wearing any kind of device. Each viewer will see the scene from its point of view and will enjoy full horizontal parallax. The high luminance of the display will allow the installation to work under normal lighting conditions (see attached video).

The display is based on a scalable holographic system design [BAL+06] targeting multi-user interactive computer graphics applications. It uses a specially arranged array of LED based projection modules and a holographic screen. The light beams are generated by the optical projection modules and the holographic screen makes the necessary optical transformation to create a continuous 3D view. With our advanced software control (operating in our render cluster containing a number of PC-s with off the shelf graphic cards), the light beams leaving the various pixels can be made to propagate in multiple directions, as if they were emitted from physical objects at assigned spatial locations. The display is directly driven by special graphic commands that are internally converted to the images needed to drive the projectors by a dedicated cluster hosting multiple consumer level graphic cards. The display reconstructs the light field of 3D scenes instead of views and can show animated images at 25 frame per second. The distributed image organization makes it different from other known multiview solutions.

The Capturing of human body data is achieved using contact-less Computer Vision technologies. Using a set-up of cameras arranged in the scenario, the 3D Capturing Device will take as input multiple video sequences, will process them by means of the bundled Computer Vision platform and will produce a 3D stream output representation that can be rendered and manipulated using computer animation techniques. The Capturing device produces a complete 3D body model composed of no less than 2000 polygons, supporting 16 million colours, for each time frame. The produced models are highly realistic since the textures are obtained from the images themselves. The system offers scalable camera configurations; from 4 up to 10 cameras. The camera calibration process is computer-assisted and can be completed in a few minutes. Once the system has been correctly calibrated, the body capturing process is fully automatic. No special markers, sensors, clothes, nor specially equipped laboratories will be required. The capturing area is 2x2 meter wide and the capturing environment will be most simple since it will only require a static background. The Capturing device performs at 25 fps, it is person independent and will operate under normal ambience light conditions. No special user training is required.

Computer Vision 3D capturing and holographic display technologies are fused by means of the Scene Composer. The Scene Composer accepts multiple input data streams, one of them the 3D capturing device. The 3D body representations coming from the 3D capturer every 40 msec are inserted into the previously defined virtual scenario as one more object that can be redendered, manipulated and interacted with. The Scene Composer provides the resources necessary to define the animations of the virtual objects and the possible interactions between the virtual and physical objects (i.e. the captured 3D

1. Eptron, tomas.rodriguez@eptron.es
2. Eptron, adolfo.cabodeleon@eptron.es
3. Total Immersion, bruno.uzzan@t-immersion.com
4. Total Immersion nicolas.livet@t-immersion.com
5. Inria, Edmond.Boyer@inrialpes.fr
6. Inria, florian.geffray@inrialpes.fr
7. Holografika, t.balogh@holografika.com
8. Holografika z.megyesi@holografika.com
9. Holografika, a.barsi@holografika.com

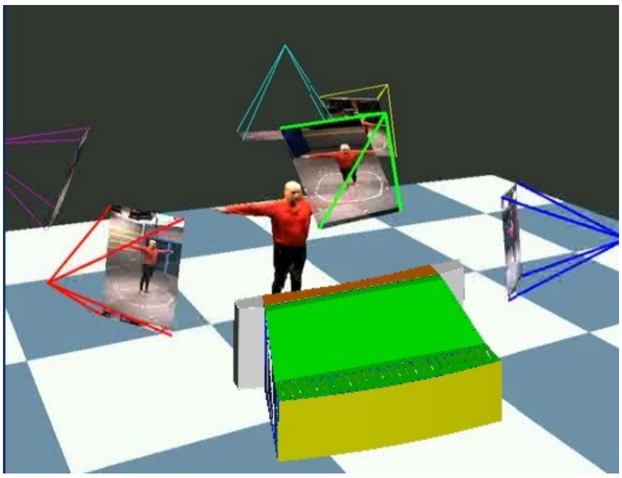
body models). The physical engine bundled with the Scene Composer ensures any 3D object present in the scenario will show correct behaviour according to physical laws. On the output side, the Scene Composer computes in real time the multiple-rendering of the finished scenario and produces an OpenGL stream suitable for the holographic display; by means of the bundled OpenGL Wrapper.

2 The action capture system in operation

Our Vision

Our current work was born with the objective to become the reference project in the field of acquisition, presentation and management of 3D information for the leisure and entertainment sectors. It intends to analyse the complete 3D development chain; a process that includes capturing, displaying, manipulating and interacting with 3D information. A concept supported on highly immersive scenarios where the users can share their experiences with other users and are no longer constrained by: annoying cabling, primitive markers, uncomfortable wearable devices, special lighting, limited freedom to move, time lags and other usability limitations.

The combination of computer vision based 3D capturing technologies and wide-angle holographic displays into a single production chain will enable the emergence of innovative applications; thus revolutionizing existing 3D virtual reality experiences.



3 3D Action Capturing **with 3D Display**

The Goal

The goal of our Emerging Technology installation is to show a working example of an integrated 3D capturing and display system which is able to acquire in real-time the geometry, texturing and dynamics of the body of a human actor, will insert the acquired 3D data stream into a high quality 3D scenario, will present the resulting scene naked eye, in true 3D, on to a large scale holographic display and will permit the human actor to interact naturally and comfortably with the presented virtual objects using his body parts.

Previous Research

Developing a scalable holographic system targeting multi-user interactive computer graphics applications is a large engineering effort that requires advances in a number of technological areas. A full survey is beyond the scope of this short proposal text but a short summary of the existing 3D display technologies can help to see the improvements reached by us.

Autostereoscopic displays implement left/right eye separation using various optical or lens rasters directly on top of LCD or plasma screens (Sharp [EWO+95], IBM, DTI, Samsung) imposing a single static viewing position, which is overcome by head/eye-tracking systems [WEH+98, RS00, PPK00]. However, such a solution cannot support multiple viewers and introduces latency.

Multi-view displays show multiple 2D images in multiple zones in space. They support multiple simultaneous viewers, restricting them, however, to be within a limited viewing angle. At SIGGRAPH 2004, Mitsubishi [MP04] demonstrated a prototype based on this technology and a number of other manufacturers (Philips [vPF96], Sharp [WHJ+00], Opticality [RR05], Samsung, Stereographics, Zeiss) produce monitors based on variations of this technology. But it suffers under cross-talks and discontinuities upon viewer's motion [Dod96].

Volumetric displays project light beams on a semi transparent or diffuse surface positioned or moved in space, that scatters/reflects incoming light [MMMR00, FDHN01]. (SeeReal [RS00], Actuality, Felix, Deep Video Imaging). Portrayed objects appear however transparent, since the light spots addressed to points in space cannot be occluded by foreground voxels.

Pure holographic displays generate holographic patterns to reconstruct the light wavefront originating from the displayed object, using some light manipulation technique [SHLS+95], [SCC+00] or [HMG03]. Compared to stereoscopic and multi-view technologies the quality of the 3D reconstruction is better, but these systems are still confined to research laboratories due to the limitations imposed by their fundamental principle. Our installation significantly eclipses over all above-listed methods in its displaying capability and usability by overcoming the mentioned limitations.

Another very relevant problem in the 3D world today is how to capture or create 3D information in the first place. In this field, we concentrate on the case of understanding both articulated and deformable objects and more specifically human motions. The automatic reconstruction of the human body is inherently difficult. The human body has many degrees of freedom and the individual limbs are deformable. Optical motion capture systems currently use a calibrated camera set-up, place markers on various body parts and are generally deployed in a specially arranged studio. These systems have proved their worth in the computer animation world. Nevertheless highly realistic animations require post-production stages that can take months.

In the recent past, a number of motion capture methods, techniques, and devices have been developed. The general scope of these systems is to provide the motion parameters associated with human-body actions and gestures. A rather complete human body model consists in an articulated kinematic chain with 74 rotational degrees of freedom [FRC+04]. In order to achieve those ambitious objectives, wavelet representations has been investigated in order to extract, describe and represent relevant cues and features independently of their size, resolution and appearance. More recently, a number of researchers addressed the problem of capturing motion directly from videos using neither specially arranged camera set-ups nor body markers [ALL+06]. The general approach to video-based motion acquisition is to consider a simpler kinematic body model and to register these models with the video data. Within the HOLONICS project we propose to recover the geometry of the human body by estimating the 3D Visual Hull [LAU+94] of an articulated shape. The Visual Hull is mathematically represented [FRC+03] as a 3D mesh and it is computed from image silhouettes that are extracted from an arbitrary number of cameras (typically 4 to 10).

Core technical innovations

- This is the first practical demonstration of a hardware and software system capable of showing accurate interactive three-dimensional animated (25fps) images with full, observer independent, continuous parallax within a large workspace.
- The display offers wide viewing angles and high depth of field. The user can move around the displayed objects without perceptible alteration of the quality of the presentation. The system correctly handles occlusions so that when the viewpoint is changed deeper layers are revealed naturally. The 3D display is true multi-person and avoids the use of glasses helmets or tracking sensors.
- The 3D Capturing Device is capable of computing the 3D geometry, texture and dynamics of a complex human body model. The body models are highly realistic since they are computed in real time and the textures are obtained from the images. The Capturing Device is user independent, does not require training and avoids the use of wearable devices. The 3D capturing device only requires a static background. No preparation of the scenario is required. The calibration process is assisted and can be completed in less than 30 minutes.
- The Scene Composer fuses the captured 3D body models and the holographic displays. The body models can be combined with other virtual objects and then the user is given the option to interact with the virtual objects being presented in true 3D in front of his eyes.

The Experience For SIGGRAPH 2007 Attendees

The application used for the demonstration will display the representation of a virtual theatre where a real human actor evolves inside a virtual scenario while he interacts with some of the objects presented in the scene. The representation of the human actor will be highly realistic since the captured 3D geometry of the human body will be high resolution, the textures will be obtained from video images and the output 3D representation will be computed in real-time. The resulting scenarios will be presented naked eye on to the large scale holographic display at 25 frames per second. The number of simultaneous viewers will only be limited by available space (up to 5x6 squared meters in front of the display, see attached technical drawings)

On the other hand, the attendants will be able to experience the interactive capabilities of the system by performing in front of a portable version of the 3D capturing device. The captured 3D models will be acquired in real-time, will be inserted into a virtual scenario and will be displayed on to the holographic display. The complete process will be real-time and interactive. As a result, the users will see how the realistic 3D objects presented in true 3D in front of his eyes can be manipulated and interacted with naturally and in real-time.

The Future Of This Work

The project developments may have important implications for the future of two relevant sectors in the graphics and multimedia sectors: 3D displays and 3D data capturing technologies. In the first case, this emerging technology demonstration will clearly show that it is possible to build large-scale 3D interactive displays that can provide consistent, shared, three-dimensional dynamic information to a group of users. This is clearly a significant step forward in display technology and it paves the way to novel applicative and cooperative tasks, such as the design of intricate manufactured objects or the understanding of complex simulation results.

The second technological result will contribute to give response to an important market demand: the possibility to obtain real-time, highly realistic human body models using contact-less technologies under little constrained scenarios. The new technology clearly outperforms existing motion capturing systems in terms of quality and ease of user while at the same time offering new functionalities not existing before. In that sense the possibility to acquire in real time not only the movement, but also the geometry and textures of the finished model (without post-production) will imply a decisive change in the way 3D information is captured and processed. Important will be also the approach to renounce to any kind of uncomfortable device; such as markers, cabling or sensors; a circumstance that have jeopardised similar attempts in the past.

The move beyond digital tradition and beyond boundaries between creativity, technology and science to transform daily lives

The presented technologies will have interesting applications in many technological and non-technological domains. The 3D displays will find applications in market fields such as: entertainment, media arts, cultural heritage (e.g. museum exhibits), advertising, simulation, scientific visualization, CAD-CAM, gaming, communications (i.e. tele-presence, 3D TV), military applications and in general in any application where high quality 3D graphics is required. On the path of developing displays in the 100+Mpixel range it is also important to show that the technology is already here and future 3D displaying, that will be a boost of 100x-1000x of the capability of current 2D technologies, is not a futuristic dream anymore.

On the other hand, the 3D Capturing device offers interesting exploitation opportunities in market fields as diverse as: communications (i.e. tele-presence), live plays, game development, amusement parks, biomechanics, training and simulation, media arts, special effects, sports, etc. Certainly, these market opportunities will be extended when both technologies are combined together.

At the time of the submission, all the separate components have already being developed and tested. We are currently finalizing the integration. In particular, the display is fully functional, and a full-scale version of the capturing device with 8 cameras is already available. The attached video demonstrates all separate components and provides a visual illustration of the integrated application.

Acknowledgments

Research partially supported by the HOLONICS project (EU-FP6-511318), funded under the European FP6/IST program.

List of referenced works

The display technology is described in the following publications (the papers refer to a smaller scale prototype; a publication on the new display is in preparation):

Full details on the basic technology are given in following patents:

Patent EP 0 693 244 B1 / US Patent 5,801,761; owned by Holografika;
European Patent Application 98 904 307.0 / US Patent 6,201,565 B1; jointly owned by Holografika and Sony;
Patent Application US: 10/276,466 / EP: 01936692.1 / JP: 2001-584933 / CN: 01809764.2 / KR: 2002-7015586 ; owned by Holografika.

White papers and other information on the display technology are also available at the following web sites:

http://www.eptron.es/projects/holonics/holonics.htm
http://www.holografika.com

[BAL+06] The HoloVizio System. Tibor Balogh Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems XIII (2006) Vol. 6055

The referenced competing display technology publications are the following:

[DML+00] DODGSON N. A., MOORE J. R., LANG S. R., MARTIN G., CANEPA P.: Time-sequential multi-projector autostereoscopic 3D display. J. Soc. for Information Display 8, 2 (2000), 169?176.

[Dod96] DODGSON N. A.: Analysis of the viewing zone of the cambridge autostereoscopic display. Applied Optics: Optical Technology & Biomedical Optics 35, 10 (1996), 1705?1710.

[EWO+95] EZRA D., WOODGATE G. J., OMAR B. A., HOLLIMAN N. S., HARROLD J., SHAPIRO L. S.: New autostereoscopic display system. In Stereoscopic Displays and Virtual Reality Systems II (1995), vol. 2409 of SPIE proceedings, pp. 31?40. 2

[FDHN01] FAVALORA G., DORVAL R., HALL D., NAPOLI J.: Volumetric three dimensional display system with rasterization hardware. In Stereoscopic Displays and Virtual Reality Systems VII (2001), vol. 4297 of SPIE Proceedings, pp. 227?235. 2

[HMG03] HUEBSCHMAN M., MUNJULURI B., GARNER H.: Dynamic holographic 3-D image projection. Optics Express 11 (2003), 437?445.

[MMMR00] MCKAY S., MAIR G., MASON S., REVIE K.: Membrane-mirror based autostereoscopic display for teleoperation and telepresence applications. In Stereoscopic Displays and Virtual Reality Systems VII (2000), vol. 3957 of SPIE Proceedings, pp. 198?207.

[MP04] MATUSIK W., PFISTER H.: 3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. ACM Transactions on Graphics 23, 3 (Aug. 2004), 814?824.

[PPK00] PERLIN K., PAXIA S., KOLLIN J. S.: An autostereoscopic display. In SIGGRAPH 2000, Computer Graphics Proceedings (2000), Akeley K., (Ed.), Annual Conference Series, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, pp. 319?326.

[RR05] RELKE I., RIEMANN B.: Three-dimensional multiview large projection system. In Stereoscopic Displays and Virtual Reality Systems XII (2005), vol. 5664 of Proc. SPIE.

[RS00] ROBERTS J. W., SLATTERY O.: Display characteristics and the impact on usability for stereo. In Stereoscopic Displays and Virtual Reality Systems VII (2000), vol. 3957 of SPIE proceedings, p. 128.

[SCC+00] STANLEY M., CONWAY P., COOMBER S., JONES J., SCAT-TERGOOD D., SLINGER C., BANNISTER B., BROWN C., CROSSLAND W., TRAVIS A.: A novel electro-optic modulator system for the production of dynamic images from giga-pixel computer generated holograms. In Practical Holography XIV and Holographic Materials VI (2000), vol. 3956 of SPIE Proceedings, pp. 13?22.

[SHLS+95] ST.-HILLAIRE P., LUCENTE M., SUTTER J., PAPPU R., SPARRELL C. G., BENTON S.: Scaling up the mit holographic video system. In Proc. Fifth International Symposium on Display Holography (1995), SPIE, pp. 374?380.

[vPF96] VAN BERKEL C., PARKER D., FRANKLIN A.: Multiview 3d-lcd. In Stereoscopic Displays and Virtual Reality Systems III (1996), vol. 2653 of SPIE proceedings, p. 32.

[WEH+98] WOODGATE G. J., EZRA D., HARROLD J., HOLLIMAN N. S., JONES G. R., MOSELEY R. R.: Autostereoscopic 3d display systems with observer tracking. Image Communication - Special Issue on 3D Video Technology (EURASIP - 1998) (1998), 131.

[WHJ+00] WOODGATE G. J., HARROLD J., JACOBS A. M. S., MOSELEY R. R., EZRA D.: Flat-panel autostereoscopic displays: characterisation and enhancement. In Stereoscopic Displays and Virtual Reality Systems VII (2000), vol. 3957 of SPIE proceedings, p. 153.

[ALL+06] The Grimage Platform: A Mixed Reality Environment for Interactions. J. Allard, J.-S. Franco, C. Ménier, E. Boyer and B. Raffin. In Proceedings of International Conference on Computer Vision Systems, 2006 (ICVS'06).

[FRC+04] A Distributed Approach for Real-Time 3D Modeling. J.-S. Franco, C. Menier, E. Boyer and B. Raffin. IEEE CVPR Workshop on Real-Time 3D Sensors and their Applications, 2004.

[FRC+03] Exact Polyhedral Visual Hulls. J.-S. Franco and E. Boyer. In Proceedings of British Machine Vision Conference, 2003 (BMVC'03).

[LAU+94] The Visual Hull Concept for Silhouette-Based Image Understanding. A. Laurentini. IEEE Transactions on PAMI, 1994.