ATTEST:
Advanced Three-dimensional Television System Technologies

André Redert[1], Marc Op de Beeck[1], Christoph Fehn[2], Wijnand IJsselsteijn[3],
Marc Pollefeys[4], Luc Van Gool[4], Eyal Ofek[5], Ian Sexton[6], Philip Surman[6]

[1] Philips Research Laboratories
Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands
{andre.redert,marc.op.de.beeck}@philips.com

[2] Heinrich-Hertz Institute
Einsteinufer 37, D-10587 Berlin, Germany
fehn@hhi.de

[3] Eindhoven University of Technology
Den Dolech 2, 5600 MB Eindhoven, The Netherlands
W.A.IJsselsteijn@tm.tue.nl

[4] Katholieke Universiteit Leuven
Kasteelpark Arenberg 10, 3001 Heverlee, Belgium
{Marc.Pollefeys,Luc.Vangool}@esat.kuleuven.ac.be

[5] 3DV Systems
Industrial Park, Building 7, 20692, Yokneam, Israel
eyal@3dvsystems.com

[6] De Montfort University
The Gateway, Leicester LE1 9BH , United Kingdom
isexton@iee.org, Phil.Surman@btinternet.com

Abstract

We describe the goals of the ATTEST project, which started in March 2002 as part of the Information Society Technologies (IST) programme, sponsored by the European Commission. In the 2-year project, several industrial and academic partners cooperate towards a flexible, 2D-compatible and commercially feasible 3D-TV system for broadcast environments.

An entire 3D-video chain will be developed. We discuss the goals for content creation, coding, transmission, display and the central role that human 3D perception research will play in optimizing the entire chain. The goals include the development of a new 3D camera, algorithms to convert existing 2D-video material into 3D, a 2D-compatible coding and transmission scheme for 3D video using MPEG-2/4/7, and two new autostereoscopic displays.

With the combination of industrial and academic partners and the technological progress obtained from earlier 3D projects, we expect to achieve the ATTEST goal of developing the first commercially feasible European 3D-TV broadcast system.

1 Introduction

As early as the 1920s, TV pioneers dreamed of developing high-definition three-dimensional (3D) color TV, as only this would provide the most natural viewing experience. During the next eighty years, the early black-and-white prototypes evolved into high-quality color TV, but the hurdle of 3D-TV still remains.

We believe that 3D is the next major revolution in the history of TV. Both at professional and consumer electronics exhibitions, companies are eager to show their new 3D products which always attract a lot of interest. Obviously, if a workable and commercially acceptable solution can be found, the introduction of 3D-TV will generate a huge replacement market for the current 2D-TV sets. In this decade, we expect that technology will have progressed far enough to make a full 3D-TV application available to the mass consumer market, including content generation, coding, transmission and display.

This paper describes the ATTEST project that started in March 2002 as part of the Information Society Technologies (IST) programme, sponsored by the European Commission. In the 2-year project, several industrial and academic partners cooperate towards a flexible, 2D-compatible and commercially feasible 3D-TV system for broadcast environments. This goal differs from that of previous 3D-TV projects (e.g. [1,2]), which aimed primarily at technological progress.

In ATTEST, we will design an entire 3D-video chain including content creation, coding, transmission and display, see Figure 1. Research into human 3D perception will play a central role. In an iterative user-centered design cycle, feedback will be given to all individual parts in the video chain to enable an overall optimization of the system.

The need for the 3D-video content will be satisfied in two different ways. First, a range camera will be converted into a broadcast 3D camera that will require a redesign of the camera optics and electronics. Secondly, as the need for 3D content can only partially be satisfied by newly recorded material, we will also develop algorithms to convert existing 2D-video material into 3D. Both offline (content provider) and online (set-top-box) conversion tools will be provided.

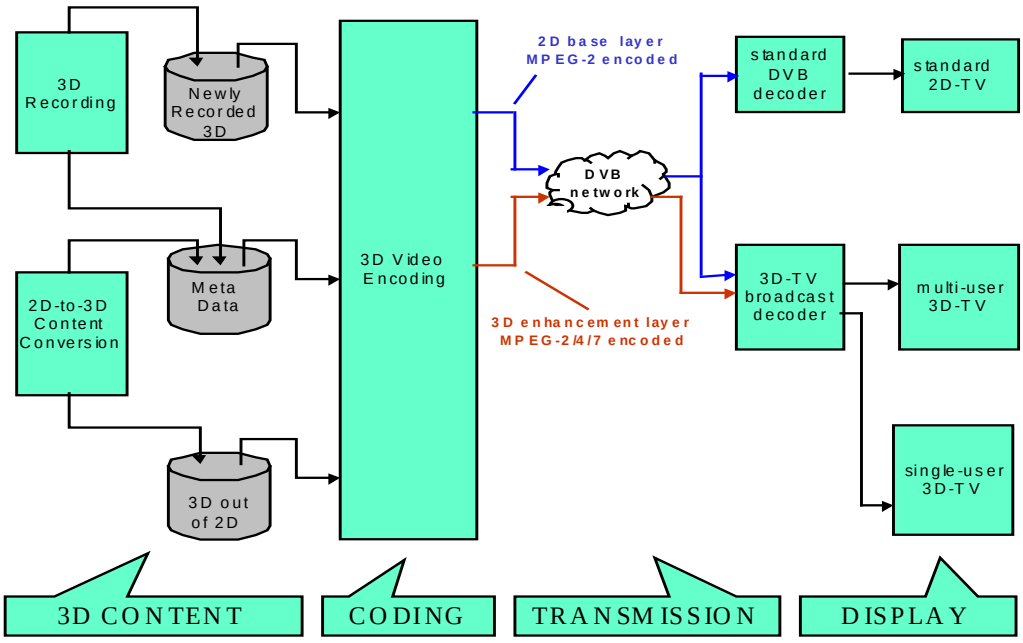

Figure 1: *The ATTEST 3D-video chain.*

Compatibility with conventional 2D-TV is of vital importance, as 2D and 3D-TV will co-exist during the evolution period. Therefore, we will develop coding schemes within the current MPEG-2/4/7 broadcast standards that allow for the transmission of depth information in an enhancement layer [3], while providing full compatibility with existing 2D decoders. For transmission, a DVB network will be used.

As the area of 3D displays is still rapidly evolving, the video chain should be adaptable to a wide range of both 2D and 3D displays. The transmitted video plus depth information allows for rendering images for many such displays [3].

ATTEST will develop two 3D displays, one for a single user and one for multiple users. Both allow free viewing (without stereo-glasses) over a wide viewing angle. Head tracking will be used to drive the display optics such that the appropriate images are projected into the eyes of the viewers.

Next, we will elaborate on individual parts of the 3D-video chain. In section 2, we will discuss the content generation part, followed by coding and transmission parts in section 3. The displays are discussed in section 4. Section 5 deals with the human perceptual evaluation of the 3D-video chain. We conclude the paper in section 6.

2 Content generation

The 3D-video content will be supplied by novel 3D cameras and via conversion from existing 2D-video material.

2.1 Novel cameras for 3D video

The 3D-video camera that will be developed during the ATTEST project is based on Zcam™; an existing depth camera [4]. The latter camera was intended for “depth keying” only, i.e. the segmentation of objects in the scene from objects in the background or foreground according to depth differences. In the project, the camera will be improved to meet the resolution and accuracy demands of 3D-TV.

Next we compare the new approach with more conventional approaches using multiple 2D cameras, followed by a discussion on the used technology and its challenges.

2.1.1 Depth camera versus multiple 2D cameras

The new depth camera will yield conventional 2D-video accompanied with depth-per-pixel via a direct depth-sensing process. This is a major advantage compared to the more conventional approaches with a stereo camera consisting of two or more conventional 2D cameras. The actual 3D-video is then acquired via subsequent image processing such as camera calibration, correspondence estimation and stereo triangulation. However, the accuracy of stereo triangulation deteriorates with the distance from the cameras. As video production requires the ability to shoot both close-ups as well as long-distance shots, the stereo accuracy will thus vary accordingly. The changing production demands require also changing the camera geometry, e.g. zooming via precision setting of two or more independent lenses. This is mechanically impractical and requires reliable calibration tools. Finally, correspondence estimation

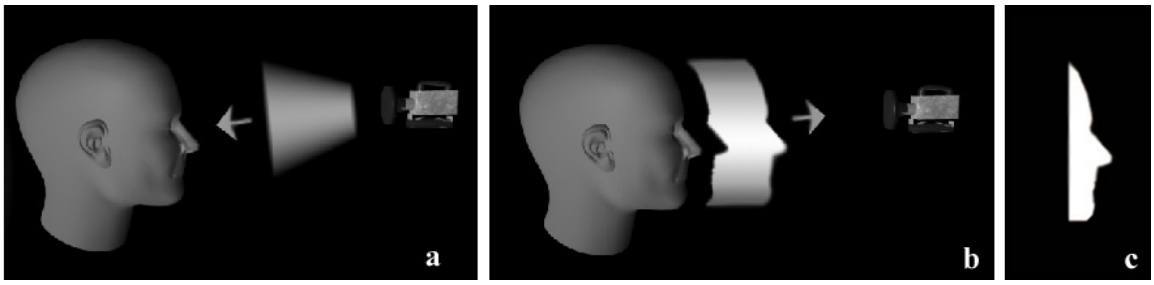

Figure 2: a) *"light wall" moving from camera to the scene,* **b)** *Imprinted light wall returning to camera,* **c)** *Truncated "light wall" containing depth information from the scene.*

depends on the existence of matchable features in the scene content and is still prone to errors due to miss-matches. A scene point must be visible by at least two cameras if its depth is to be recovered. Since the cameras have different positions, it is common that there are areas that are visible by a single camera only. This problem is reduced by increasing the number of 2D cameras, but this will make the stereo camera setup more difficult to handle during production and increase the amount of video streams to be processed.

The camera developed in the ATTEST project overcomes the aforementioned obstacles. The depth camera is based on a single sensor that measures the distance from the camera to the scene at each pixel simultaneously. There are no angular differences between the color camera and the depth sensor, so each pixel of the color camera is assigned a corresponding depth value.

The depth measurement is independent of the visible scene content: a depth map is generated even if the scene contains no visible features (e.g. in total darkness). This assures correct recovery of depth maps for areas of constant color, for example cloths and walls.

The depth accuracy of the camera is independent of the distance from the camera. The camera measures linearly scaled depth values inside a controllable depth range. This enables the camera to handle seamless changes between long distance shots and close-ups, without affecting the quality of the recovered depth and without any change of the camera geometry (such as a change of base line).

2.1.2 The technology of the depth camera

The operation of the camera is based on generating a "light wall" moving along the field of view, see Figure 2. As the light wall hits the objects, it is reflected back towards the camera carrying an imprint of the objects. The imprint contains all the information required for the construction of the depth map. The 3D information can now be extracted from the reflected deformed "wall" by deploying a fast image shutter in front of the CCD chip and blocking the incoming light as shown in Figure 2c. This type of camera belongs to a broader group of sensors known as scanner-less LIDAR (laser radar without mechanical scanner), see [5].

The collected light at each of the pixels is related to depth, but also to the reflectivity of objects. Hence, a normalization step is performed per pixel by simply dividing the front portion pixel intensity by the corresponding portion of the total intensity [4].

The technological challenge of the depth camera is twofold: Fast switching of the illumination source to form the "light wall", and fast gating of the reflected image entering the camera.

In the current depth camera, a cluster of IR laser diodes and corresponding optics is used to generate

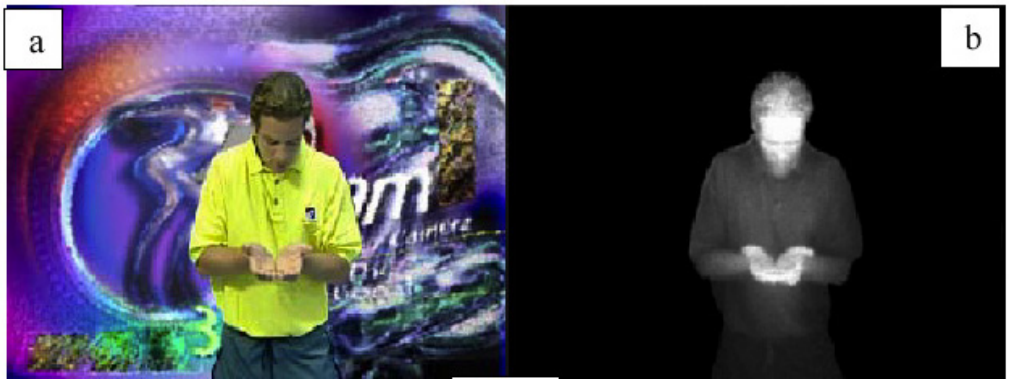

Figure 3: *The images taken by the depth camera,* **a)** *normal RGB image,* **b)** *accompanying depth image (grey level inversely proportional to depth).*

(a) **(b)**

Figure 4: a) *Image extracted from a video and* **b)** *computed depth map.*

homogeneous illumination. The diodes are switched on and off with rise/fall times shorter then 1 nsec. None of the existing fast drivers and switchers was suitable for our extreme application. Hence, super fast driver electronics had to be designed to comply with the fast response, small space and low cost, and yet maintain high efficiency.

The detection of the reflected pulse has to be synchronous with the switched illuminator. For this, a special fast driver has been designed that has rise/fall times shorter then 1 nsec. The current camera uses a fast optical switch on the basis of a so-called gated intensifier. This device is pixelized and contributes a small amount of noise, which limit the depth resolution and accuracy respectively. In the project, we will develop a solid-state shutter, which circumvents both limitations.

Figure 3 shows the video and depth images taken by the current camera [4].

2.2 Conversion from conventional 2D video

In order to provide sufficient 3D content, it is important that one can convert existing 2D-video footage to 3D data. Within the ATTEST project two types of conversion tools will be developed.

The first approach will enable on-line depth augmentation using a set-top-box at the receiver end. This allows a user at home to activate 3D depth augmentation for any suitable 2D-video content that is received. In this case, computations can only be based on video frames that have already been received. Real-time implementations like these require that the developed approach can take full advantage of advanced DSP capabilities. Within the scope of the ATTEST project the approach described in [6] will be developed further.

The second set of tools will allow content providers to convert existing 2D-video content to 3D before broadcasting the data. In this case computations can be performed off-line and computationally more expensive algorithms can be used. Another important advantage is that all image data is available at once. If necessary, 3D information obtained in one camera shot can even be used to provide depth augmentation for another. The ATTEST approach will build further upon techniques developed for 3D modeling from image sequences [7]. First, features are tracked from frame to frame. Using robust statistics and multi-view geometry, we can eliminate wrong matches. The calibration of the camera and the relative motion between scene and camera are computed. Independent object motions will also be computed. In a next stage, corresponding points are determined for all pixels, using a stereo algorithm, and from this a depth image is computed. An example is shown in Figure 4.

3 Coding and transmission

An important issue of the ATTEST project is the development of a novel data representation and a related coding syntax for future 3D-TV broadcast services. In contrast to former proposals, this new approach is based on a flexible, modular and open architecture that provides important system features, such as backwards compatibility to today's 2D digital TV, scalability in terms of receiver complexity and adaptability to a wide range of different 2D and 3D displays. For this purpose, the data representation and coding syntax of the ATTEST system are based on a layered structure shown in Figure 5. This structure consists of one base layer and at least one additional enhancement layer.

To achieve backwards compatibility to today's conventional 2D digital TV, the base layer is encoded by using state-of-the-art MPEG-2 and DVB standards. Thus, this layer can be decoded by standard set-top boxes designed for 2D digital TV broadcast reception.

The remaining enhancement layers deliver the additional information for the 3D-TV receiver. The minimum information transmitted in this enhancement layer is an associated depth map providing one depth value for each pixel of the base layer. However, in the case of critical video content (e.g. large scale scenes with a high amount of occlusions) it might be necessary to send further information, for example segmentation masks and occluded texture. The layered structure in Figure 5 is extendable in this sense.

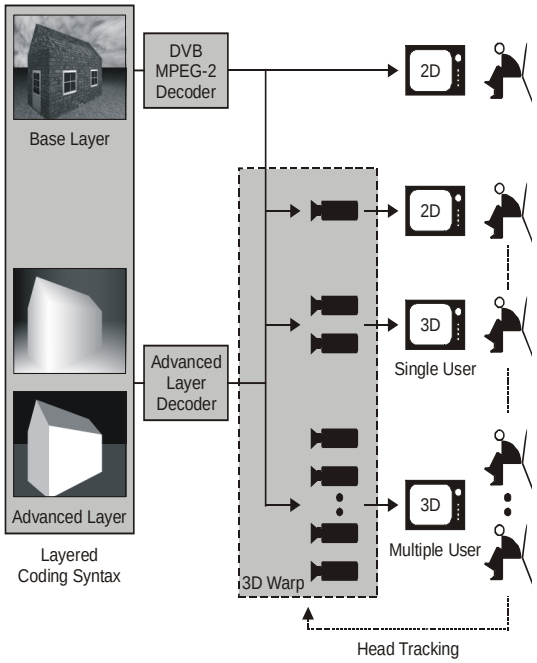

Figure 5: *A layered coding syntax provides backwards compatibility to conventional 2D digital TV and allows to adapt the view synthesis to a wide range of different 2D and 3D displays.*

As stereovision is only one depth cue and other cues such as motion-parallax are of comparable importance, it is significant to note that the described layered structure is flexible enough to support alternative forms of depth representation. This allows for the stepwise introduction of 3D-TV receivers of different complexity. For example, an intermediate low-cost 3D-TV receiver could use the additional depth layer to render individual perspective views according to the head-tracked viewing position of the TV viewer. That way broadcasters could provide the users with a first, limited depth impression through parallax viewing, even on conventional 2D-TV screens.

Another point that should be mentioned is that the proposed syntax will also provide scalability in terms of depth experience. This is particularly important, as perception studies have indicated that there are differences in depth appreciation over age groups. Hence in our view, the TV viewer should be in control of his depth experience. He should be able to set the depth level according to his personal preference – a feature which can also be used for graceful degradation in the case of unexpected artifacts in depth which are usually more annoying in stereovision than in parallax viewing.

4 Displays

We will develop two 3D displays, one for a single user and one for multiple users. A 20" single-viewer display based on lenticular screen optics will use the output of a low-cost non-contact infrared head tracker to ensure that the viewer has a high degree of movement, both laterally and fore-and-aft. The display will not suffer from the restriction of confining the viewer to being positioned close to an optimum viewing plane as is the case in most currently available systems. Within the viewing region the viewer will be provided with a picture with excellent depth quality, good color reproduction and very low crosstalk. Hardware interfaces will enable live video to be displayed in real time.

The multi-viewer display provides 3D for up to four viewers who can occupy a viewing area that is between one and three meters from the screen and ± 30º from the axis. Regions where a left image only, and a right image only, are seen across the complete width of the screen are referred to as the exit pupils. These are formed in pairs as illustrated in Figure 6, and head tracking is employed to ensure that the pupils are always located at the appropriate eyes.

The display operates by having the conventional LCD backlight replaced by steering optics that are able to form the multiple steerable exit pupils. The two images are presented on one screen by displaying the left image on even rows of pixels, and the right image on the odd rows. This halves the vertical resolution, but as a 20.8" QXGA (2048x1536 pixels) LCD will be used, the resolution will still be better than the current 625-line standard.

As in the single-viewer display, an infrared head tracker controls the steering. The steering optics utilize a combination of white LED arrays, a two-dimensional spatial light modulator and a novel optical configuration to control the light. This is an ambitious project and placement of exit pupils over the large region will provide some challenging problems, not least being that of crosstalk which is likely to be the factor limiting the extent of the usable viewing field.

5 Perceptual evaluation

The acceptance, uptake, and commercial success of any advanced technology aimed at the consumer market depend to a large extent on the users' experiences with

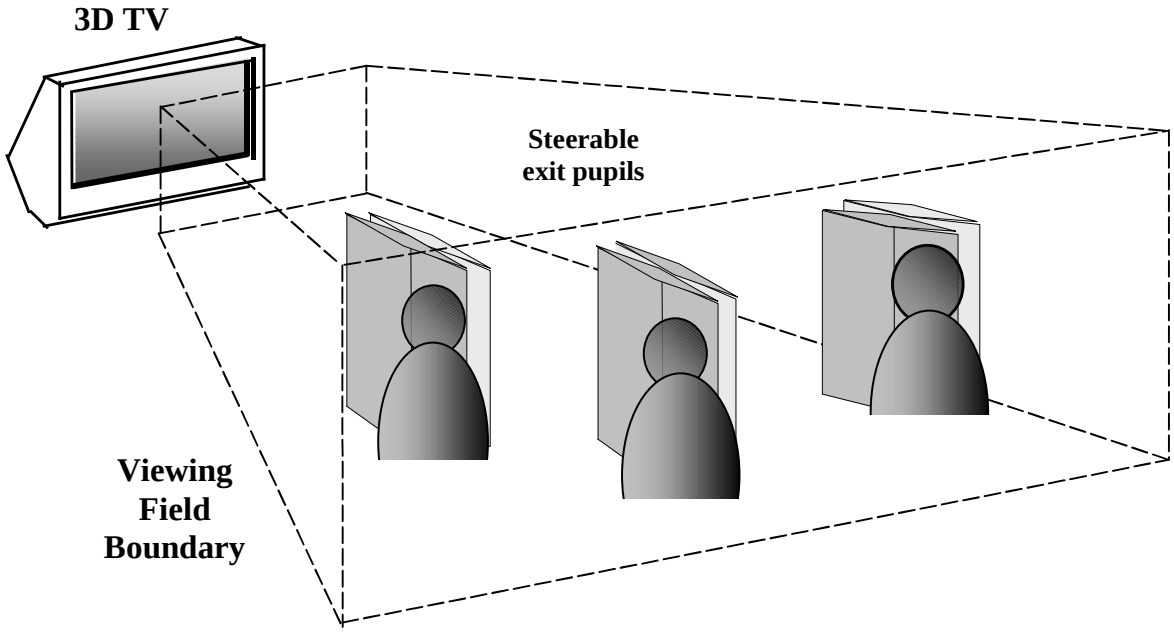

Figure 6: *The multiple-viewer 3D display shows left and right images in the directions corresponding to the positions of the left and right eyes of every viewer.*

and responses towards the system. In the past, 3D video in theme parks, and even in 3D broadcast trials, were often intended to provide the viewers the '3D thrill of their life'. Depth impressions were also exaggerated to enhance the visual impact. Unfortunately, viewers also frequently experienced eye strain, headaches, and other unpleasant side effects. Therefore, it is vital to have a clear understanding of the in-the-home viewing experience of 3D-TV, both looking at the potential added value of the ATTEST 3D-TV systems, as well as the potential drawbacks for users. Our aim is to arrive at a set of requirements and recommendations for an optimal 3D-TV system, and contribute to each individual step in the 3D video chain through perceptual and usability evaluations of the proposed technological innovations.

More specifically, human-factors experiments will be performed to address the depth impression, perception of distortions, eye strain, quality, naturalness, presence, and acceptability of the 3D coding algorithms and novel 3D displays, in order to arrive at perceptually optimal image quality with minimal coding artifacts and negligible side-effects [8],[9],[10]. Additionally, a number of basic and novel areas surrounding 3D video perception will be investigated that will enhance our understanding of the user experience. For example, user control over the depth impression in 3D video has to date received very little systematic experimental investigation. This will be one of the issues that will be addressed in ATTEST, looking at both basic perceptual and cognitive effects as well as ease-of-use. In addition, the fundamental issue of acceptability of 2D production grammars for 3D video will be investigated, requiring a much deeper understanding of how depth perception develops over time - e.g., how tolerant viewers will be to sudden disparity changes - whilst relating these insights to existing 2D and 3D video production grammars.

6 Conclusions

We described the goals of the ATTEST project, which started in March 2002 as a part of the Information Society Technologies (IST) programme, sponsored by the European Commission. In the 2-year project, several industrial and academic partners will cooperate towards a flexible, 2D-compatible and commercially feasible 3D-TV system for broadcast environments.

The 3D-TV system will be an entire 3D-video chain including content creation, coding, transmission and display. All parts will be optimized with respect to the entire chain, guided by research on human 3D perception.

We discussed the specific goals for all system parts. A new 3D camera will be developed that meets the resolution and accuracy requirements of the 3D-TV application. Both real-time and off-line algorithms will be developed to convert existing 2D-video material into 3D. For transmission, we use a 2D-compatible method in which conventional images are accompanied with depth information, coded with MPEG-2/4/7 schemes. This scheme enables addressing of a wide range of 2D and 3D displays. Finally, two autostereoscopic displays will be developed; one optimized for a single viewer, and a second display for multiple viewers.

With the combination of well-established academic and industrial partners, and building upon the technological progress obtained from earlier 3D projects, we expect to achieve the ATTEST goal of developing the first commercially feasible European 3D-TV broadcast system.

7 References

[1] DISTIMA, European RACE 2045 Project, http://www.tnt-uni-hannover.de/project/eu/ distima, 1992-1995

[2] PANORAMA, European ACTS AC092 Project, http://www.tnt-uni-hannover.de/ project /eu/panorama, 1995-1998

[3] M. Op de Beeck and A. Redert, "Three dimensional video for the home", Proceedings of the *International Conference on Augmented, Virtual Environments and Three-Dimensional Imaging (ICAV3D)*, Mykonos, Greece, 2001, pp. 188-191

[4] G.J. Iddan and G. Yahav, "3D Imaging in the studio (and elsewhere…)", SPIE vol. 42983D SMPTE Journal, June 1994

[5] M.W. Scott, "Range imaging laser radar", US patent 4.935.616, 1990

[6] P. Wilinski and K. van Overveld, "Depth from motion using confidence based block matching", in *Proceedings of Image and Multidimensional Signal Processing Workshop*, Alpbach, Austria, 1998, pp. 159-162

[7] M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, "Hand-held acquisition of 3D models with a video camera", *Proc. 3DIM'99 (Second International Conference on 3-D Digital Imaging and Modeling)*, IEEE Computer Society Press, 1999, pp.14-23

[8] L. Stelmach, W.J. Tam and D. Meegan, "Perceptual basis of stereoscopic video", *Proceedings of the SPIE 3639: Stereoscopic Displays and Virtual Reality Systems VI*, 1999, pp. 260-265

[9] W.A. IJsselsteijn, H. de Ridder and J. Vliegen, "Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration", *IEEE Transactions on Circuits and Systems for Video Technology 10*, 2000, pp. 225-233

[10] W.A. IJsselsteijn, J. Freeman, D.G. Bouwhuis and H. de Ridder, "Presence as an experiential metric for 3-D display evaluation", To be presented at the *Society for Information Display 2002 International Symposium*, Boston, MA, USA, May 19-24, 2002