Towards an optimized 3D broadcast chain.

Marc Op de Beeck*[a], Piotr Wilinski[a], Christophe Fehn[b], Peter Kauff[b], Wijnand Ijsselsteijn[c], Marc Pollefeys[d], Luc Van Gool[d], Eyal Ofek[e] and Ian Sexton[f]

[a]Philips Research, The Netherlands; [b]Heinrich-Hertz-Institut, Germany ; [c]Eindhoven University of Technology, The Netherlands; [d]Katholieke Universiteit Leuven, Belgium; [e]3DV Systems, Israel and [f]De Montfort University, United Kingdom.

ABSTRACT

In this paper we will present the concept of a modular three dimensional broadcast chain, that allows for an evolutionary introduction of depth perception into the context of 2D digital TV. The work is performed within the framework of the European Information Society Technologies (IST) project "**A**dvanced **T**hree-dimensional **Te**levision **S**ystem **T**echnologies" (ATTEST), bringing together the expertise of industries, research centers and universities to design a backwards-compatible, flexible and modular broadcast 3D-TV system. This three dimensional broadcast chain includes content creation, coding, transmission and display. Research in human 3D perception will be used to guide the development process.
The goals of the project towards the optimized 3D broadcast chain comprise the development of a novel broadcast 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 technologies and the design of two new autostereoscopic displays.

Keywords: Keywords: 3D-TV, broadcast, camera, depth, content, video coding, autostereoscopic, display, perception

1. INTRODUCTION.

As early as the 1920s, TV pioneers dreamed of developing high-definition three-dimensional (3D) color TV, to provide the most natural viewing experience possible. During the next eighty years, the early black-and-white prototypes evolved into high-quality color TV, but the hurdle of depth introduction still remains. Nevertheless, 3D is rapidly invading our lives. The steady advancement in 3D gaming is even recognized as one of the main driving forces for the replacement of PC hardware. At the same time we observe a convergence between TV and PC applications in the area of broadcast (e.g. for time shift viewing) and DVD, as well as a convergence in screen format (HD as 720 lines progressive) allowing a high quality visualization of typical PC applications (e.g. internet browsing) on the TV screen. Driven by 3D gaming, the introduction of the 3D displays into the home environment seems to be an obvious evolution.

We believe that 3D will be the next major revolution in the history of TV environment. 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.

Stimulated by popular science-fiction movies 3D-TV is often associated with the futuristic vision of large holographic 3D displays visualizing actors who seem to materialize inside the living room. The viewer would have the opportunity to walk around the scene, or watch it from a distance while locking in on a certain viewpoint or even actor. It is clear

* Corresponding author : Marc Op de Beeck, Philips Research, Prof. Holstlaan 4 (WY8.1), 5656 AA Eindhoven, The Netherlands, E-mail : Marc.Op.de.Beeck@Philips.com

that these game-like applications requiring novel 3D recording and transmission standards and that current bandwidth restrictions currently prohibit such an introduction scenario in the broadcast domain. We believe that 3D-TV can be introduced in the near future, provided that full compatibility with existing 2D oriented recording and transmission broadcast equipment is maintained. 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 focus on the joint optimization of the entire 3D-video chain including content creation, coding, transmission and display, as schematically depicted in Figure 1. Research into human 3D perception will play a central role, and will provide feedback will be given to all individual parts in the video chain to enable an overall optimization of the system.

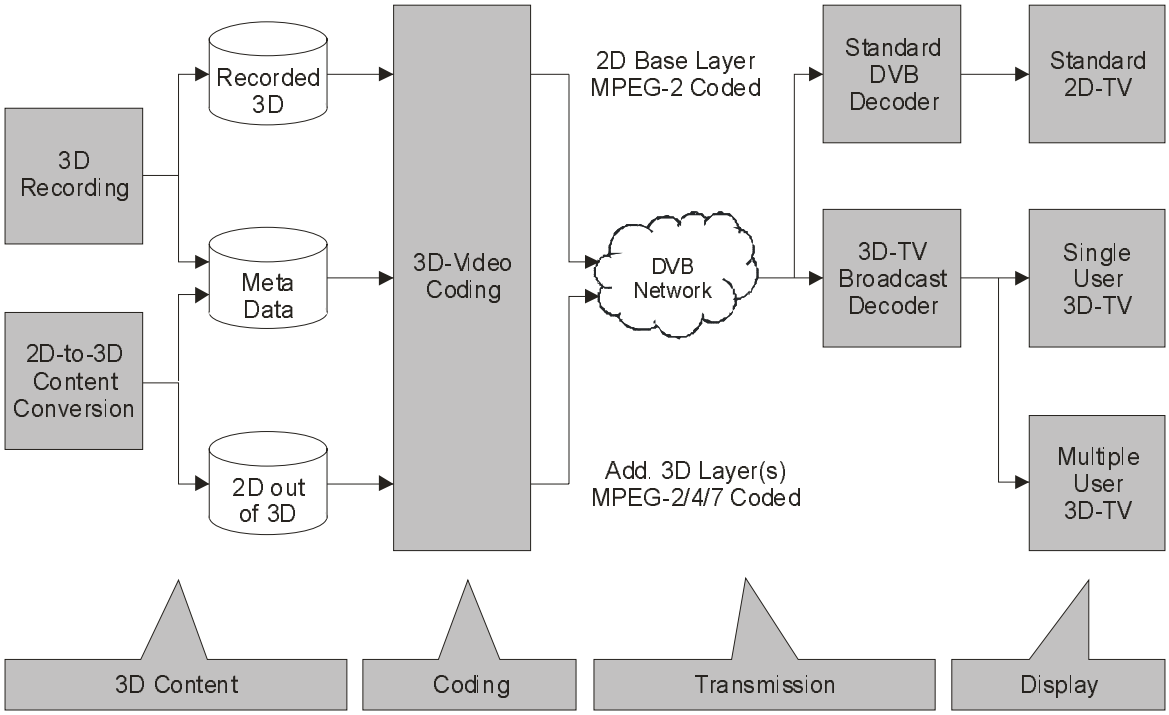

Figure 1 – The ATTEST 3D broadcast chain.

Obviously 3DTV can only be introduced if novel (live) 3D content can be recorded. Hence ATTEST includes the development of a 3D broadcast camera. In contrast to the introduction of color TV, when color broadcast was available only for a limited amount of time per night, consumer expectations have evolved, and it is unlikely that introduction of 3DTV can be driven by novel 3D content only.
Since there is a large repository of existing 2D material, we will also develop algorithms to convert existing 2D-video material into 3D. To some extend, this can be related to the art of re-coloring of black-and-white movies; a technique that became mature too late to assist the transition to color TV. Within ATTEST, we intend to bring the 3D

reconstruction of 2D movies and impressive documentaries to maturity before the actual introduction of 3DTV, as 3D converted 2D material may prove to be an essential factor for the successful transition from 2D to 3DTV.

Compatibility with conventional 2D-TV is of vital importance, as 2D and 3D-TV will co-exist during the introduction 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, while ensuring full compatibility with existing 2D decoders [3]. For transmission, a DVB network will be used.
Worldwide, there is a large development effort resulting in various flavors of 3D displays, some requiring passive or active headgear, some based on various autostereoscopic principals. For gaming, which is a typical single user application, the use of headgear may be acceptable. For TV viewing in a typical family setting in a living room, on the other hand, the use of headgear is considered to be an undesired nuisance. Autostereoscopic displays, i.e. 3D displays not requiring personal visualization aids (active or passive glasses, head mounted displays, …) are already available, but often requiring the viewer to be located in one of the limited number of 'sweet spots'. On positions in between the sweet spots, 3D viewing is impaired. At present, a suitable glasses-free 3D-TV display is not available, ATTEST will develop two autostereoscopic displays (single- and multiple user) that will allow free viewer positioning. Head tracking will be used to drive the display optics such that the appropriate images are projected into the eyes of the viewers.
Finally, as consumer acceptance will ultimately decide on the commercial success of any future 3D-TV system, requirements for optimal 3D enjoyment will be assessed through human perception studies and feedback will by given to all individual parts of the system in an iterative, user-centered design cycle.

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. This will be followed by some initial results in section 6. We will finish the paper with a short conclusion in section 7.

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]. This camera was designed for "depth keying" only, i.e. the segmentation of objects in the scene from objects in different layers according to depth differences, and is intended as alternative to the well-known blue screen techniques. In the project, the camera will be improved to meet the resolution and accuracy demands of 3D-TV.

2.1.1 Depth camera versus multiple 2D cameras

The ability to record accurate depth information is a key requirement for a three dimensional camera. Two approaches can be taken : either record a depth related direct signal, i.e. a radar-like approach [5], or capture signals that allow the calculation of the depth values. A well-known example of the last option is the more conventional approach of a stereo camera consisting of two or more conventional 2D cameras. Both approaches are fundamentally different in many ways.

The multiple camera approach starts from synchronous multiple recordings of the same scene from different angles. Let's for simplicity only discuss the stereo camera approach. The actual 3D-video is then generated through various image processing steps such as camera calibration, correspondence estimation (for features or regions) 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. Moreover, the depth accuracy scales with the distance between the cameras, the so-called base line. When using the cameras in zoomed-in mode, the base line should increase to provide the same depth accuracy putting harsh requirements on camera position, direction and lens calibration. The technique relies on well-defined features or detailed texture. The presence of non or low textured areas often lead to large depth uncertainties. Finally, 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. This type of camera belongs to a broader group of sensors known as scanner-less LIDAR (laser radar without mechanical scanner), see [5].
The operation of the camera is based on generating a "IR 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. The collected light at each of the pixels is inversely proportional to depth of the specific pixel. Since reflecting objects may have any reflectivity coefficient there exist a need to compensate this effect. Hence, a normalization depth is calculated per pixel by simply dividing the front portion pixel intensity by the corresponding portion of the total intensity [4].

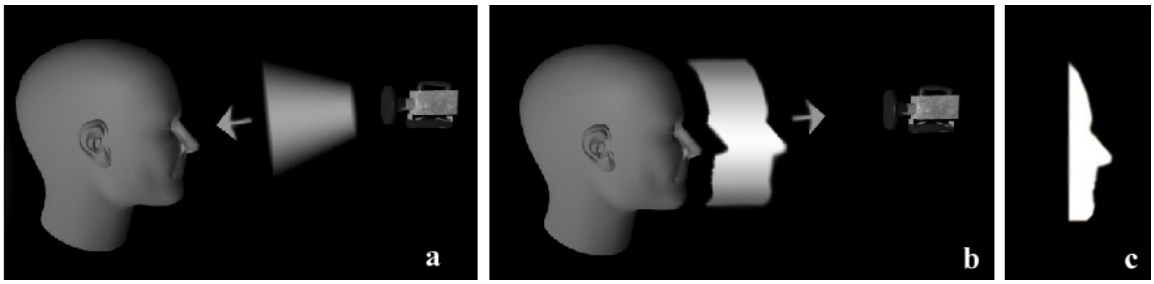

Figure 2: a) *"light wall" moving from camera to the scene,* ***b)*** *Imprinted light wall returning to camera,* ***c)*** *Truncated "light wall" containing depth information from the scene.*

In this set-up, the reflected IR light passes the same lenses as the visual light. Behind the lenses, the IR and visual light are separated and recorded with different sensors. 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. Camera zoom is accounted for in a very natural way as both IR and visual light passes through the same optical path. The depth measurement is independent of the visibly scene content: a depth map is generated even if the scene contains no visible features (e.g. in total darkness) and is independent of local image resolution. 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 designated depth range. This range (minimum and maximum depth distance) can be controlled and modified according to need. This feature allows 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).

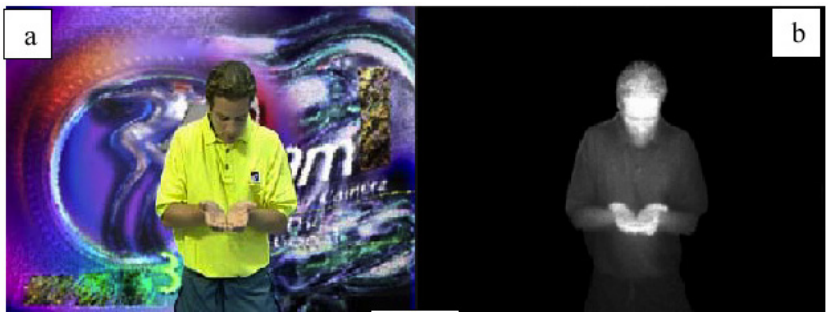

Figure 3: *The images taken by the depth camera,* **a)** *normal RGB image,* **b)** *accompanying depth image: grey level inversely proportional to depth with the [min, max] depth range.*

2.1.2 The technological challenges of the depth camera

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 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 Depth augmentation of existing 2D video content.

The introduction of color TV was fully driven by novel recorded content and in the introduction period the percentage of color broadcast was fairly limited. Today, the consumer's expectations have drastically changed. The timely availability of a sufficient amount of content appears to be a crucial factor in the successful introduction of novel technology. Therefore we will develop algorithms to convert existing 2D video footage to 3D data. This may seem to be impossible at first, but video sequences often include a large set of different depth cues which are implicitly present in the 2D video sequence. In the depth augmentation step, we make these implicit cues explicit.

Within the ATTEST project two types of conversion tools will be developed.
The first set of tools will allow content providers to convert existing 2D video content to 3D before the data are broadcast or put on a distribution medium. These computations can be performed off-line, may include some manual interaction and puts only mild restrictions to computational expenses. It can be considered to be part of the time-consuming (and often costly) post-processing step to provide the best quality possible. In this scenario, all image data from a complete shot is available at once. Hence, all available cues can be integrated over time. If necessary, 3D information obtained in one camera shot of the same scenery 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 or areas are determined for all pixels. When combined with camera calibration, a depth image is computed. An example is shown in Figure 4.

The second approach will be developed to allow on-line depth augmentation using a set-top-box at the receiver end. This should allow a user at home to activate 3D depth augmentation for any suitable 2D video content. 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 hardware capabilities. Within the scope of the ATTEST project the approach described in [6] will be developed further.

(a) **(b)**

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

3. CODING AND TRANSMISSION

Earlier proposals for the introduction of 3DTV were often based on the coding of stereoscopic video information, i.e. one video stream per eye. This format assures backward compatibility with digital 2D video, if the second view is coded in a separate MPEG bit stream. The format itself is very inflexible. The stereo effect is only optimized for a pre-defined screen size, and does not allow for an easy customization of the depth effect.

Within the ATTEST project we focus on 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.

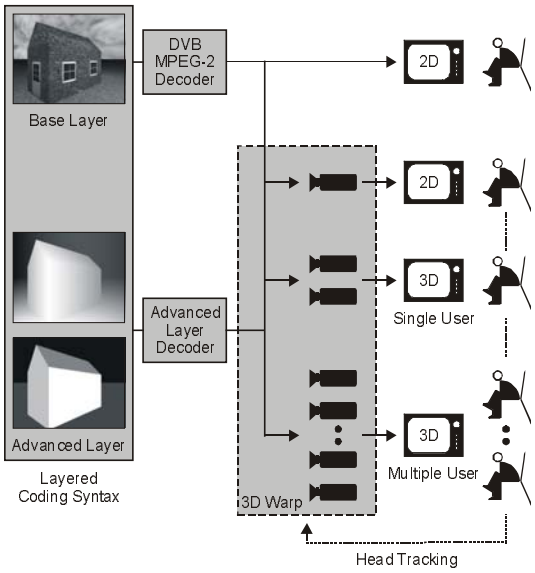

Figure 5: *A layered coding syntax provides backward compatibility to conventional 2D digital TV and allows adapting the view synthesis to a wide range of different 2D and 3D displays.*

To ensure backward compatibility to today's conventional 2D digital TV, the base layer is encoded by using state-of-the-art MPEG-2 and DVB standards. The minimum information transmitted in this enhancement layer is an associated depth map providing one depth value for each pixel of the base layer. Based on the depth information different views (e.g. left and right) can be generated through image morphing techniques. Note that these techniques only provide an approximation of these views. In case a foreground object is observed from different angles, different parts of the occluded background become visible. As this information is not present in the 2D-video view, the occlusion information in the morphed approximation will not be correct. Luckily, the human brain is trained to neglect the information that is only visible to one of the eyes. However, in the case of critical video content (e.g. large scale scenes with a high amount of occlusions and large depth differences) it might be necessary to send further information, for example segmentation masks and occluded texture. Note that the layered structure in Figure 5 is extendable in this sense.

For the transmission of the enhancement layer(s), it is planned to rely as far as possible on already available MPEG-2/4/7 tools. Nevertheless, in case existing tools prove to be inadequate to meet the ATTEST project goals, we are intending to contribute our results to the appropriate standardization bodies. Therefore, ATTEST is already strongly participating in the recently established MPEG Ad-hoc group (AHG) on 3D-video coding [11].

This ATTEST format is very flexible towards the display type. A standard set-top box designed for 2D digital TV broadcast reception, will neglect the enhancement layers and decode only the base layer information, i.e. the 2D video information. A more advanced 3DTV decoder will also decode the enhancement layers, which even can be exploited on a conventional 2D display by providing depth specific rendering. It will even be possible to provide motion parallax by morphing alternative viewing angles depending on the viewer's position. On an autostereoscopic 3D display, different views can be rendered simultaneously, but note that the described layered structure is flexible enough to support alternative forms of depth representation.

The application scenarios above allow for the stepwise introduction of 3D-TV receivers of different complexity. The same 3D bit stream allows the broadcasters to provide the users with a first, limited depth impression through parallax viewing, even on conventional 2D-TV screens, while early adopters can already experience the full 3D solution.
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. In the past, 3DTV has often been referred to as headache TV as viewers were often exposed to largely exaggerated depth effects. 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 that 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. Separating image (texture) and depth information allows for a natural scaling of the depth effect.

4. DISPLAYS

Throughout the world we see a large activity in the area of 3D displays. Opposite to the evolution in 2D displays, where the market seems to evolve into the direction of matrix based displays both for direct view (e.g. LCD) and for projection displays (e.g. LcoS), there is an explosion of different flavors of dedicated 3D displays. Indeed, current 3D displays are based on very different principles: multi-layer (e.g. stacked LCDs), stereoscopic (e.g. head mounted displays, LC shutter glasses), multi-view (e.g. lenticular), holographic…

The whole ATTEST 3D-video processing chains is designed to be flexible towards the use of different 3D displays. We identify two major application domains: single user 3D-TV in the PC domain and multi-user 3D-TV for typical living room environments. Both applications have different requirements, hence the project will develop two types of 3D displays: a single- user and a multiple user autostereoscopic display:

A two view, autostereoscopic, 20" single-viewer display based on lenticular screen optics will be combined with a low cost non-contact infrared head tracker. The head tracker will drive a mechanical displacement device that changes the relative position of the LCD display and the lenticular to ensure correct projection of the appropriate views in the user's eyes. Hence, the viewer has a high degree of movement both laterally and in depth. The display will not suffer from the restriction of confining the viewer to being positioned close to an optimum viewing plane as 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 $\pm 30^{o}$ from the axis. The conventional LCD backlight is replaced by a so-called directional backlight that project the appropriate images in the conjugate eye positions. 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, possibly limiting the extent of the usable viewing field.

5. PERCEPTUAL EVALUATION

The previous sections have discussed 3DTV in terms of technological and economical challenge. 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 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 are often 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 the novel 3D displays, in order to arrive at a perceptually optimal image quality with minimal coding artefacts and negligible side-effects [8-9-10].
Furthermore, ATTEST will also contribute to fundamental insight in 3D-video perception. For example, user control over depth impression has to date received very little systematic experimental investigation. This will be one of the central 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 the 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 environment. 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. ACKNOWLEDGEMENTS

This work has been sponsored by the European Commission (EC) through their Information Society Technologies (IST) program under proposal No. IST-2001-34396. The authors would like to thank their project partners represented by P. Surman (DMU, UK), B. Duckstein, R. de la Barre, B. Quante and I. Feldmann (HHI, D), S. Malassiotis and M. G. Strintzis (ITI, GR), F. Ernst and A. Redert (Philips, NL), D. G. Bouwhuis (TU/e, NL) and P. Gyselbrecht and K. Van Bruwaene (VRT, B) for their support and for their input to this publication.

8. 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] Marc Op de Beeck and André 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] F. Ernst, P. Wilinski, K van Overveld, "Dense structure from motion - An approach based on segment matching", *European Conference on Computer Vision*, May 2002, Copenhagen, Denmark.

[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, pp.14-23, 1999

[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*, 260-265, 1999

[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*, 225-233, 2000

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

[11] Op de Beeck, M., Fert E., Fehn, C. and Kauff, P., March 2002. Broadcast Requirements for 3D Video Coding. *ISO/IEC JTC1/SC29/WG11 MPEG02/M8040*. March 2002.